Doctoral Thesis

Firm Renewal in the Regional Economy

Tina Wallin

Jönköping University
Jönköping International Business School
JIBS Dissertation Series No. 118, 2017
Acknowledgement

“The consequences of our actions are always so complicated, so diverse, that predicting the future is a very difficult business indeed” (Rowling, 1999, p. 311).

Which is probably why I had no idea when I came to Jönköping 11.5 years ago that I would not only stay for my bachelor’s degree, as I planned, but I would add a master’s and a doctoral degree to that. In retrospect, I’m glad my original forecast was wrong. These years have given me so much – apart from a Ph.D. I’ve learned new skills, met scholars from all over the world, seen places all over the world, and I’ve met a lot of bright and kind colleagues that I’m happy to call my friends. For having had those opportunities I am very grateful. Indeed, if one would use a regression attempting to explain education level from my background characteristics (no parents with higher education, low-income, and raised in a neighbourhood with high segregation) the predicted outcome would never have placed me here. I have to admit that it feels rather good to be that outlier.

I would like to start with thanking my supervisors Börje Johansson, Lars Pettersson and Johan Klaesson for the encouragement and support throughout my doctoral studies. An extra thank you to Johan who agreed to jump in towards the end and supervise me, despite already supervising half of the PhD candidates in the discipline. As Johan was also my supervisor when I wrote my bachelor thesis I rather feel that the loop has been closed. And who likes an unclosed loop?

Great advice has also been given in other places, for example through our ‘tortures’ (or feedback sessions as we tried to rename them for public use to make them sound less scary) and Friday seminars. Over the years I have been a Ph.D. candidate a generational shift has taken place from a time when the quartet of current emeritus professors Åke E. Andersson, Charlie Karlsson, Börje Johansson and Per-Olof Bjuggren were among those attending the tortures, doing their best to keep us on track, to a new era with a younger, but still very knowledgeable group, consisting of Scott Hacker, Agostino Manduchi, Kristofer Månsson, Johan Klaesson and others who have taken over the baton. Considering that I made it this far I think they did a good job. During the Friday seminars I have usually had most colleagues from the department there, generously providing suggestions and going over nitty-gritty things in my papers. Your support over the years has been invaluable and I am very grateful for it. Someone who also provided highly useful comments that greatly improved my work is Dieter Pennerstorfer, who agreed to be my discussant at my final seminar earlier this year. Thank you very much for making such an effort in providing constructive comments!

Not only has advice been generously given, but I have been lucky to having had the opportunities to take part in projects of different kinds. For example, I have been part of research projects with many colleagues, such as Börje

---

Johansson, Johan Klaesson, Lars Pettersson, Sara Johansson, Lina Bjerke, Mikaela Backman, Kristofer Månsson, Sofia Wixe, Peter Warda, and Therese Norman-Monroe. I have also had the opportunity to work closer with Sara, Börje, Mikaela through co-authoring two of my papers. Thank you for being part of them and this thesis! I am grateful for the opportunity to having learnt more about research and the writing process from you. Together with Charlie Karlsson and Andreas P. Cornett I have also had the opportunity to be a book editor. The process has been very interesting, and I am grateful for having been part of it. A person who has shown me another very different side of academia is Charlotta Mellander who, with some assistance of Emil Danielsson, has taught me that journal articles alone will not make you a force to be recognised in non-academic contexts. Through her encouragement and support I have written some smaller popular pieces through the vertikals blog, I have been interviewed on the radio, and I have been part of the construction and presentation of the yearly index ‘bäst-att-bo’ (with the direct translation being ‘the best place to live’) Thank you for giving me the opportunity to be part of the group!

Something that I have really come to appreciate over the years is the great environment we have in the department and how people really love what they do and don’t mind helping others. If you run into a problem, there is always someone who’s willing to lend you an ear and provide suggestions. We help each other out and look after each other, a little bit like a family. Slightly odd and dysfunctional from time to time, but which family does not have its quirks? If comparing us to a family, then Kerstin Ferroukhi, Katarina Blåman, Monica Bartels and Marie Petersson have taken it in turns to be our mothers. We have not always been the easiest bunch to raise, but thank you for taking such good care of us!

Many families also come with siblings. I have had plenty of those over the years, and as opposed to my ‘real family’, here I have not been the oldest one. Rather I have been something like a middle child. When I started this Ph.D. adventure there was a bunch of slightly older siblings (Özge Öner, Johan P Larsson, Peter Warda, Sofia Wixe, Viroj Jienwachtaramongkhol, and Therese Norman-Monroe) who seemed very wise and cool – like any older siblings do. After getting to know them I realised that it was not just the appearance, they really were (and still are) wise and cool. Özge is an inspiring person who is always full of ideas and she doesn’t mind discussing and sharing them (or her snacks and food) with her younger siblings. Johan is the cool brother that you would want with you if you’re in trouble and want someone to argue your case. I don’t think I’ve ever seen him lose a discussion. Peter is the neat one who’d help you colour coordinate you books in your book shelf and keep your room tidy so that your parents don’t get angry. Our lunch room on the 5th floor, and I, really miss him! Sofia is the organised sister who makes sure that you did your homework on time and brought you gym kit to school for PE. She is a very caring and smart person that I feel lucky to having had the opportunity to study and travel with. Viroj is the artistic brother who lives a double life, economist by day and photographer by night, and he excels at both. Therese is a special sister who is closest to me by age in this odd bunch of siblings. She was actually one of the first persons I met when
I moved to Jönköping as she was one of my ‘fathers’ during the introduction week. Then we were both became research assistants and later PhD candidates, and it has been great to have her support though both good and rough times. I am very grateful for having all of these special siblings in my life over the last 6 years.

As all families, also this one has changed its shape over the years. After a few years I also got a bunch of younger siblings, or cones (as in fir cones) as they are sometimes referred to. Among them is the little brother Toni Duras who is crazy about two things in life; football and statistics. I usually don’t understand either of them that much. Although I did beat him in the intra-department bet we had during the European Championships last year, which has to count as one of my major achievements during my Ph.D. candidate years. The family has also become a very multicultural group with the newer PhD candidates having different experiences from many different countries. Thank you all for making our department an enjoyable and dynamic place to work, you have your special places in the department too! For example, Orsa Kekezi is my go-to girl in any matter related to Harry Potter, Mark Bagley is always sharing nice tips about books to read or series to watch (not to mention fitting large tables into this thesis template), Helena Nilsson is my partner in crime when it comes to crosswords and wine (not necessarily in that combination), and Jonna Rickardsson has been a great listener when I have needed to ventilate both working related issues and private ones. I am very happy to see that the open and kind spirit in our family continues to blossom with this new generation.

If this has been my nearest family I have also had some cousins, Lina Ahlin, Sam Tavassoli, and Trudy-Ann Stone, that I have met through courses, summer schools, and conferences. Thank you for all the good times we had together and let’s hope there are more to come!

Naturally, life as a Ph.D. candidate revolves a lot around the thesis, but it’s important to be able to do something else as well. Luckily, I have brilliant friends also outside academia who have been able to keep me grounded and give me a different perspective on life when I have needed it. Sofia Forsberg is like my natural science oriented twin that also has a slight obsession with books, including Harry Potter, and who almost always is willing to travel to London with me when my abstinence symptoms start to kick in. Teresa Thörn is one of my oldest friends who also happens to be my neighbour, which has proved very convenient for walks, dressing advice, and great conversations. Johan Alvarsson is the friend that I stole from my husband (and haven’t returned), who used to be my gym buddy, and whom I can always count upon has a space on his sofa and his computer loaded with movies for when a tired Ph.D. candidate needs to reboot. Lovisa Skyborn is the friend who sort of knows what I am doing as we did our bachelor studies together. This has proved very useful during tough periods as her insight allows her to give great advice. She also re-introduced me to knitting and crocheting which has been very relaxing as it allows the hands to do some work after the brain has already done its share for the day. All of you have helped keeping me sane throughout the process and for that I am very grateful!
Lastly, but definitely not least, I would like to thank my husband Linus Wallin for standing by my side throughout these years. Apart from being my husband he is my friend, my therapist, my IT support, my conscience, and also my human radiator during the winters. When I was most stressed out with my thesis work he cooked me lunch boxes to bring to work, and when I barely left my computer, he rubbed my shoulders and brought tea. He even helped me reconstruct tables for this thesis template when Word was driving me crazy, which was very lucky for the computer whose life otherwise would have ended below my office window. You are my rock and I would be lost without you!

Jönköping, 5 December 2017
Tina Wallin
Abstract

This thesis consists of four independent papers. In each paper the purpose is to analyse firm renewal, with special consideration to regional characteristics. Highlighted here are the knowledge available in the region where a firm is located as well as the local supply of financial intermediaries.

The first paper analyses how the interaction of firms’ internal and external knowledge, in terms of knowledge intensive business services (KIBS), is related to the intensive and extensive margins of export flows. The second paper can be considered a continuation of the first, as the focus is solely on the export of new products. Also here, the interaction of internal and external knowledge is the main focus, analysed through education types. The results of these papers suggest that the regional contexts are indeed relevant for firm renewal, as firms that benefit from being located in areas with high external knowledge already have high internal knowledge.

The third paper shifts focus to the innovation process itself by analysing how the access to the local supply of banks is related to firms’ perceived obstacles for innovation activities. The results indicate that lower access to banks increases the probability to experience problems obtaining external capital, which could be detrimental for firms’ renewal possibilities in the long run. The fourth paper focuses on the creation of firms and analyses whether individuals are more or less likely to be self-employed after having children in a context where an extensive welfare system exists. The results show that individuals are less likely to be self-employed after having children than before, thus lowering the rate of firm renewal in the economy.
# Table of Contents

1. Introduction and summary of the thesis .......................................................... 1

2. Firm renewal and the relevance of space ....................................................... 5
   2.1. Conceptual foundations ........................................................................... 5
      2.1.1. The theory of the firm ...................................................................... 5
      2.1.2. The multifaceted meaning of renewal ............................................. 7
   2.2. Drivers of renewal .................................................................................. 9
      2.2.1. Cost effectiveness through process innovation as an attempt to
              maximise profits ............................................................................. 9
      2.2.2. Global competition pushing firms to renew themselves through
              product innovation ......................................................................... 10
      2.2.3. Technological shifts pushing firms to renew themselves through
              product innovation ......................................................................... 12
   2.3. The innovation process ......................................................................... 13
      2.3.1. The linear model of innovation ....................................................... 13
      2.3.2. The chain-linked model .................................................................. 14
      2.3.3. Institutions and innovation systems .............................................. 16
   2.4. Bringing innovations to the market ....................................................... 17
      2.4.1. New firm formation ........................................................................ 18
      2.4.2. Innovations in firms ....................................................................... 19

3. Data and empirical issues ............................................................................. 23
   3.1. Sweden at a glance .............................................................................. 23
   3.2. Measuring innovative behaviour ............................................................ 25
      3.2.1. Unidimensional measures ............................................................... 26
      3.2.2. Multidimensional measures ............................................................ 27
      3.2.3. The Community Innovation Survey .............................................. 28
   3.3. Measuring entrepreneurship ................................................................. 29
   3.4. Knowledge ............................................................................................ 30

4. Summary and contributions of the individual chapters ............................... 32

References ....................................................................................................... 37
Paper 1 - Internal and external knowledge and introduction of export varieties .......................................................... 53

1. Introduction .......................................................................................................................... 55
2. Theoretical framework ......................................................................................................... 58
   2.1. Developing and introducing export varieties .................................................................. 58
   2.2. Combination of internal and external knowledge .......................................................... 60
   2.3. Hypotheses ....................................................................................................................... 63
3. Empirical strategy .................................................................................................................. 64
   3.1. Extensive and Intensive Margins of export flows .............................................................. 65
   3.2. Regression model ............................................................................................................. 66
   3.3. Independent variables ....................................................................................................... 68
4. Data and descriptive statistics .............................................................................................. 70
5. The influence of internal and external knowledge on the scope of export varieties in local industries .................................................................................................................. 73
   5.1. Regression results for the total stock of variety triplets 2006 ........................................ 74
   5.2. Regression results for the gross introduction of variety triplets 2002-2006 .................. 77
6. Concluding remarks .............................................................................................................. 79
References .................................................................................................................................... 82

Paper 2 - Exporting firms’ absorptive capacity and innovative behaviour ... 89

1. Introduction .......................................................................................................................... 91
2. Theoretical framework ......................................................................................................... 92
   2.1. Sources and types of knowledge ..................................................................................... 93
   2.2. Absorptive capacity and relatedness .............................................................................. 94
   2.3. Previous studies and hypothesis .................................................................................... 95
3. Data and method ................................................................................................................... 96
   3.1. The dependent variable .................................................................................................. 97
   3.2. The independent variables ........................................................................................... 98
   3.3. Descriptive statistics .................................................................................................... 102
   3.4. Model and estimation method ...................................................................................... 106
4. Empirical results ................................................................................................................ 107
   4.1. Regression results ......................................................................................................... 107

References .................................................................................................................................... 82
4.2. Sensitivity analysis ................................................................. 114
5. Conclusions .............................................................................. 115
References ................................................................................... 117

Paper 3 - Access to banks and external capital acquisition: .............. 125
1. Introduction ............................................................................. 127
2. Why is geographical proximity to financial intermediaries important? ................................................................. 129
  2.1. The case of innovation ....................................................... 130
  2.2. The Swedish financial sector ............................................. 131
3. Data, variables and method ...................................................... 133
   3.1. Method ............................................................................... 134
   3.2. Variables ............................................................................. 135
   3.3. Descriptive statistics ........................................................... 138
4. Empirical results ...................................................................... 139
  4.1. Access to banks ............................................................... 142
  4.2. Control variables ............................................................... 144
  4.3. Sensitivity analysis ............................................................. 146
5. Conclusions .............................................................................. 148
References ................................................................................... 149

Paper 4 - Self-employment and parenthood ....................................... 157
1. Introduction ............................................................................. 159
2. Background ............................................................................ 160
   2.1. The Swedish welfare system ............................................. 160
   2.2. A self-employed career choice ....................................... 162
   2.3. Self-employment among parents ................................... 164
3. Empirical strategy .................................................................. 165
   3.1. Method ............................................................................... 165
   3.2. Data ................................................................................... 166
   3.3. Dependent variable ........................................................... 167
   3.4. Main variables ................................................................... 167
   3.5. Control variables ............................................................. 168
1. Introduction and summary of the thesis

“...all things are in motion and nothing at rest...”

In Plato’s work *Cratylus*, a reference is made to the philosopher Heraclitus’s idea of the persistence of change. Over 2000 years later, the phrase is still relevant in many areas of society. One such relevant area is in the free and competitive market, where this motion is represented by ever-changing firm dynamics, as market actors struggle to keep ahead of competitors.

This motion is often observed in our society, as new products are constantly released on the market, and their advertisements compete for our interest and money. One example is Apple, which releases new versions of the iPhone frequently, allowing it to continuously have a new product for sale with the latest features to attract customers to their mobile phone and away from Samsung’s. Similarly, some clothing shops advertise that new garments can be found in their stores at frequent intervals to lure the customers into visiting them rather than a competitor. Another example is the structural change in the media sector, as a result of lower demand for printed media. This has forced firms within this sector to reinvent the way they organise, their business models, and the content of their products in order to survive.

An influential strand of the literature in economics discussing this process of keeping ahead of competitors started with Schumpeter (1934) over 80 years ago. He highlighted the entrepreneur as the main driver of innovations by combining existing resources in new ways. When a disruptive innovation reaches the market, it upsets the status quo through creative destruction. The new and improved product attracts new customers and increases the firm’s market share. Consequently, firms selling old varieties of the product lose customers and sales – some may even be driven out of business. Creative destruction can also be more comprehensive, leading to changes in development blocks. These blocks rely upon several complementary innovations that together have the ability to revolutionise entire industries and even to transform society (Dahmén, 1988; Schön, 2012). Constant motion thus creates dynamics on the market, where only the most efficient and relevant firms survive. Through this competitive structure, natural, human, and financial resources are used in an efficient way that contributes to economic growth.

The process of motion, considering pressure from competitors, also indicates the relevance of geography. If competitors are located at a distance, chances are that they are not competing in the same geographical market for customers. Not

---

2 Plato (2013, p. 79)
unless there is a geographic proximity will two comparable products act as substitutes for customers³. The same relevance attributed to geographic proximity can arguably be applied to the development of ideas; not until ideas meet and are questioned can they be developed further. Returning to Plato, he was influenced by great scholars and influenced others, such as Aristotle and Socrates, by meeting them and engaging in discussion in person in Athens. Thereby Athens became an important metropolis in ancient Greece for knowledge exchange and the development of philosophy, democracy, epistemology, mathematics, etc. This development would most likely never have occurred if Athens had not been an agglomeration of academic institutions, other public arenas, and intelligent individuals, thereby attracting these scholars.

“No man is an island, entire of itself; every man is a piece of the continent, a part of the main”⁴

This quote, from a poem by the English poet, priest, and politician John Donne, emphasizes the importance of the interdependencies between individuals just described in relation to Athens. Instead of individuals living and working separately, as if they were islands, individuals within a geographic area do interact with others, intentionally and unintentionally, through their upbringing, work life, and leisure time. These interactions allow for the exchange of goods, services, ideas, and knowledge.

As the exchange of knowledge is facilitated by geographic proximity, any formal model relying upon the creation and dissemination of knowledge inherently also relies upon geography. One example of this relation is the development of endogenous growth models in the 1980s and 1990s, which introduced the role of space, perhaps unintentionally, into the literature on economic growth, as several of its components require proximity. The most apparent component is knowledge accumulation, which Romer (1986) claimed could be generated by investing in R&D, but not perfectly patented by the firm conducting the investment, as some knowledge would spill over and be useful to firms located in the vicinity. Knowledge spillovers have since been considered a prominent factor in firms’ ability to acquire new knowledge.

Knowledge spillovers have been discussed in various ways in the literature on externalities. One approach was proposed by Marshall (1890), who discussed how specialised industrial districts were beneficial for firms due to their ability to utilise the knowledge being disseminated in the area. These benefits arise when firms can take advantage of pooled labour markets and lower transport costs, this latter being dependent upon the specialised cluster being located close to

³ Unless you count online shopping. However, even here, products are not perfect substitutes, as varying transport costs, delivery times, and tariffs make products from different countries appear more or less appealing.
⁴ Donne (1999, p. 101)
suppliers. These positive externalities are usually labelled localisation economies in the literature. Similar approaches were later advocated by Arrow (1962) and Romer (1986), resulting in their frequent grouping into MAR (Marshall-Romer-Arrow) externalities.

A second approach was suggested by Porter (1990), who also discussed the benefits of specialised clusters. However, he offered a different perspective, suggesting that competition in these clusters helps foster knowledge externalities. The MAR externalities, by contrast, favour a monopoly structure, as that facilitates the internalisation of benefits from the knowledge externalities.

A third approach was proposed by Jacobs (1969). She claimed that the diversity and scale of economic activity that is seen in cities is a great breeding ground for new knowledge. Consequently, most innovation and most growth should occur there. These positive externalities are part of the external effects usually denoted urbanisation economies in the literature, as they do not simply occur in any locality, but in urban localities. Together these two concepts constitute agglomeration economies. Regardless of whether the MAR externalities or Jacobs externalities are the major sources of growth, the idea is that knowledge spreads between people in dense areas, which implies that firm performance cannot be considered in isolation – the firm is part of a larger context.

With that in mind, the purpose of this thesis is to analyse the ability of firms to renew their businesses with special consideration given to the regional and local context. This purpose is accomplished by analysing renewal both through the creation of new firms and through the creation of new products within these firms. The regional and local contexts enter the analyses through the regional endowments of useful resources found in the geographical proximity of the firms. More specifically, the first chapters of the thesis examine the connection between firm renewal and regional supplies of knowledge and financial capital, whereas the last chapter focuses on the creation of firms and the local context in terms of family for those individuals starting new firms. It should be noted that the thesis is delimited to the supply side of the economy and does not regard the demand driving the introduction of new products on the market or the establishment of new firms.

Altogether, this thesis consists of five chapters; this introductory chapter and four empirical chapters, which can all be read and understood independently. Chapter 2, Internal and external knowledge and the introduction of export varieties (co-authored with Börje Johansson and Sara Johansson) analyses how the interaction of internal and external knowledge is related to the intensive (the degree of market penetration) and extensive (the degree of market reach) margins of export flows. This chapter also conducts a deeper analysis of the general patterns observed by Swedish exporting firms. Chapter 3, Exporting firms’ absorptive capacity and innovative behaviour can be considered a continuation of Chapter 2, where the focus is solely on the export of new products. Also in this chapter, the interaction of internal and external knowledge is the main focus, albeit analysed in a different way. Chapter 4, Access to banks and external capital acquisition – perceived innovation obstacles (co-authored with Mikaela
Backman) analyses how access to local bank branches is related to firms’ perceived obstacles to conducting innovation activities. Last, Chapter 5, *Self-employment and parenthood* changes focus to the creation of firms and analyses whether individuals are more likely to be self-employed after having children, also in an institutional context with an extensive welfare system. It also examines whether not controlling for unobserved heterogeneity biases the results.

The results of these chapters suggest that the regional and local contexts are indeed relevant to firms’ ability to renew their businesses. The main conclusion from the second and third chapters regarding the interaction of internal and external knowledge is that firms with high internal knowledge are those that benefit from being located in areas with high external knowledge. The implication of these results is that firms do not necessarily develop more new products in regions with a high level of knowledge that potentially could spill over; it depends on their internal ability to absorb and utilise it. Apart from the regional knowledge base being relevant to firms’ innovation activities, chapter four concludes that the regional supply of financial intermediaries is also important. The results indicate that lower access to bank branches increases the probability of firms experiencing obstacles to their innovation processes, which could be detrimental to firms’ renewal possibilities in the long run. The results of the last chapter showed that the welfare system also has a potential impact on firm renewal, as individuals under a strong welfare system, like Sweden’s, are less likely to be self-employed after having children than they are before.

The policy implications of these results are several. First, it is not enough to establish policies enhancing the knowledge base in regions to increase the number of dynamic firms. The internal capabilities must be able to bridge the cognitive distance from the external knowledge to fruitfully take advantage of it. Thus, firms are not lone islands but are reliant upon interdependencies with the rest of the region, as were scholars in ancient Greece. Second, a continued closure of bank branches, similar to the trend seen in many parts of Sweden over recent decades, may harm firms’ innovation activities due to a lack of funds. In the long run, this could affect firms’ profitability and even their survival, as innovation processes are required for firms to keep ahead of the competition and, as Plato wrote, to stay in motion. Third, the welfare system for parental benefits in its current form may discourage individuals from starting their own firms. In theory, parental leave is equally accessible for employees and self-employed individuals. However, the value of the monetary compensation received, both during parental leave and the subsequent absences for sick children, etc., is based upon the salary received, which unintentionally puts self-employed parents at a disadvantage.

The remainder of this introductory chapter aims at providing a context for the subsequent chapters, and it starts off by discussing firm renewal, which is central to all four papers. Section 2 first covers some important micro foundations of this process and connects them to a spatial framework before discussing the ways in which firm transformation can occur: by creating completely new firms, which is analysed in Chapter 2 and Chapter 5, or by creating new products within existing firms, which is analysed in Chapters 2-4. Section 3 provides an overview of the
Swedish geography and administrative system before discussing empirical measurement issues that arise in the later chapters related to innovation, entrepreneurship, and knowledge. Finally, Section 4 summarises the remaining chapters and their contributions.

2. Firm renewal and the relevance of space

In economic theory, renewal is not an aim in itself – the larger objective is to explain economic growth and wealth, which often occurs through the development of new technologies. The emphasis on technological change is evident in models of economic growth when observed over the past 60 years.

In the 1950s, neoclassical (or exogenous) growth models were developed (Solow, 1956, 1957; Swan, 1956). Albeit highly focused on capital and labour, there was room for a technology component – often referred to as the Solow residual. Solow (1957) empirically estimated this component’s contribution to the growth in the US between 1909 and 1949. He found that 87.5 per cent of productivity growth was caused by technological change, captured by the residual. Later, the role of technology and innovation was endogenised in different ways: one suggestion was to let technology be affected by the labour input in the education sector (Uzawa, 1965); another was to assume a degree of non-excludability in the knowledge generated by firms’ R&D efforts, thus leading to knowledge spillovers and the possibility of increasing returns to scale (Romer, 1986, 1990).

However, these growth theories focus on the macro level, implying that the behaviour of individual firms cannot be understood based on them. Consequently, we must change perspective to the firm level to understand the drivers of change.

2.1 Conceptual foundations

2.1.1 The theory of the firm

Before entering any discussion about how and why firms renew themselves, it is useful to first consider what a firm is and why it exists. Several views and theories have been suggested over time and I will present six of the more prominent ones here.

One of the oldest is the neoclassical view of the firm, in which the firm is merely one of many actors, and its manager combines various inputs to obtain outputs whilst attempting to maximise the current profits and present value of the
future stream of profits (Chandler, 1992; Hart, 1989). In that sense, the firm is a black box, and we only observe its contacts with the outside world.

Another view of the firm is the principal-agent theory developed mostly in the 1970s by Alchian and Demsetz (1972), Jensen and Meckling (1976), and Fama (1980) among others. This strand of literature focuses on the separation between ownership and control as manifested by firms with separate owners and managers. Principal-agent problems arise because managers may have incentives to act in their own self-interest rather than in the interest of the owners. Their ability to do so is a consequence of information asymmetry, where they have access to more knowledge than the owners.

A third view that arose mainly in the 1970s and 1980s is transaction cost theory. This view was heavily pushed by Oliver Williamson, who returned to the ideas of Coase (1937) and suggested that the key to why firms exist is that the transaction costs arising when producing goods and services is lower if the transaction is internalised rather than left to the market (Riordan & Williamson, 1985; Williamson, 1975; Williamson, 1981; Williamson, 1985). This is particularly true for transactions including investments in human capital or in specialised equipment that is difficult to use for other purposes, a phenomenon called asset specificity (Riordan & Williamson, 1985). However, it is also related to the frequency of the transaction. If a transaction is recurring over a long time, it becomes difficult for the parties to obtain all information necessary to write a satisfactory contract that eliminates opportunism. The cost-effective solution is then to internalise the transaction within a firm (Williamson, 1979).

The evolutionary theory suggested by Nelson and Winter (1982) can be considered a fourth strand of the literature. At any point in time, they assume, firms have certain capabilities, decision rules and routines, and the process of changing these routines is mostly performed by R&D departments. The search policy of the individual firm determines the probability distribution over how much, or what type of, new technology will be discovered. Which of the technologies from the given set is discovered depends on both the technology embedded in the innovation and the technologies already discovered and assimilated by the firm. Thus, this theory assumes a path dependency, where the organisation learns from past experiences and uses that knowledge to branch out to new markets (Chandler, 1992).

A rather different approach is the resource-based view of the firm, which developed in parallel with many of the aforementioned theories. This field disregards the firm’s accomplishments, but highlights which inputs, or resources, it uses. The implication of this theory is that the firms with the best resources will be the most successful and thus able to earn profits on those inputs (Penrose, 1959; Rubin, 1973; Wernerfelt, 1984). Consequently, one large issue within this field is how to preserve these resources and maintain the comparative advantages with respect to other firms (Nonaka, Toyama, & Nagata, 2000; Wernerfelt, 1984).

The newest of these theories is the knowledge-based view of the firm, which was developed in the 1990s and early 2000s. This theory perceives the firm as a knowledge-creating entity, where activities, strategies and culture are key
ingredients fostering new knowledge and thus maintaining the firm’s competitiveness (Grant, 1996; Kogut & Zander, 1992; Nickerson & Zenger, 2004; Nonaka et al., 2000). Since knowledge can be considered one of the resources of the firm, this whole stream of literature could be considered a sub-field of the resource-based view. However, this field is much more explicit about the process for transforming inputs into outputs.

As the topic of this thesis is firm renewal, and the focus is on knowledge utilisation and the dynamics of firms’ production in terms of innovative activities, the evolutionary perspective and the knowledge-based view are especially relevant. The evolutionary perspective highlights path dependency in the search for new innovations, where previous knowledge matters for the creation of new knowledge, as the models in Chapter 2 and Chapter 3 indicate. The knowledge-based view is similar in the sense that the firm’s knowledge is highlighted as key to the firm maintaining competitiveness. Nevertheless, Chapter 4 discusses information asymmetries between agents, which is one of the key features of the principal-agent theory. However, in my case, the asymmetry arises between the firm and financial institutions rather than between owners and managers, as the principal-agent theory suggests.

### 2.1.2. The multifaceted meaning of renewal

Entrepreneurs, or business owners, contribute to the renewal of industries and markets by new firm formation, but existing firms can also contribute to renewal by innovating. However, it is by no means a simple task to define the term innovation and thereby make sense of what it is these innovating firms do. A multidisciplinary study based on a content analysis of articles by Baregheh, Rowley, and Sambrook (2009) yielded over 60 definitions. Clearly, innovation can be defined in many ways.

A simple definition from the Merriam-Webster English dictionary suggests that an innovation is “the introduction of something new”. This appears straightforward, but the meaning of several of these words can be discussed. What is this something that is new? How different does it have to be to be new? For whom should it be new? Should it be new to the world market, the local market or is it enough to be new for the producing firm? Additionally, what is meant by the introduction of this something? Does it occur on a market? Can it be in terms of a patent? Is the introduction of an idea sufficient? All these questions show that the simple dictionary definition is not sufficient for research purposes.

An economics dictionary provides a slightly more specific definition of innovation, which is twofold: (1) an improvement of the production process and (2) a change of characteristics in marketable products (Pearce, 1992). This definition coincides with the common division of innovation into processes and products seen in large parts of the literature (M. Andersson & Karlsson, 2006; Cohen & Klepper, 1996; Schmookler, 1966; Utterback & Abernathy, 1975).

Others widen this definition by discussing three versions of innovation: product innovations, process innovations, and organisational innovations (Boer &
During, 2001; Greenhalgh & Rogers, 2010). Nevertheless, the most elaborate definition and discussion of innovation was made by Schumpeter (1934) almost 80 years ago. He proposed five definitions of innovation.

1. The introduction of a new good or an old good with improved quality
2. A process innovation
3. An organisational innovation
4. The entry into new markets, both in terms of product markets and geographical markets
5. The introduction of a new raw material or supplies of raw material

The question about what an innovation is thus has many answers, and researchers need to be clear about what they do and do not include in the concept. Another dimension that can be discussed is how large the difference is between the new and the old products. A common approach is to distinguish between incremental innovations and drastic, or radical, innovations. An incremental innovation is a smaller improvement of a previous innovation, while a radical innovation is very different (Fagerberg, 2005). Often, radical innovations are considered to be more important (e.g., Schumpeter (1934) highlighted this aspect), although incremental innovations are often those that make an innovation useful for a wider audience. Hence, the cumulative impact of incremental innovations is likely to be even larger than that of the first radical innovation (Fagerberg, 2005; Lundvall, 1992).

Another issue raised is for whom the product, process, etc., should be new for it to be considered an innovation. According to Fagerberg (2005), a common view is that a product should be new to the market for it to be called an innovation, and if another firm introduces the same product in a different context, that firm should be regarded as an imitator. However, the commonly used Oslo Manual compiled by the OECD suggests that it is enough for a product to be new to the firm to be considered an innovation (OECD, 1997). Their compiled definition is the following:

“A technological product innovation is the implementation/commercialisation of a product with improved performance characteristics such as to deliver objectively new or improved services to the consumer. A technological process innovation is the implementation/adoption of new or significantly improved production or delivery methods. It may involve changes in equipment, human resources, working methods or a combination of these” (OECD, 1997, p.9).

This definition by the OECD also raises the issue of the commercialisation of the new product. In the 1960s and 70s, innovations were considered to be a process that brought about change or simply the actual generation of ideas (Cumming, 1998). Over time, this notion has changed, and the idea generating process is considered to be a form of creativity; for it to be called an innovation,
the new product must also be introduced to the market. The latest development, according to Cumming (1998) is to also require success from an idea for it to be a proper innovation. This distinction between idea generation and commercialisation is analogous to the distinction between invention and innovation. As explained by Lundvall (2007), an invention only becomes an innovation when the entrepreneur introduces the new product to the market. This shows that the idea generation process results in invention until the moment when the product is successfully produced and offered for sale in appropriate markets. Consequently, a patent is more closely related to inventions than to innovations, and to discuss firm innovations, one must consider markets (Acs, Anselin, & Varga, 2002).

Although I agree with all five definitions suggested by Schumpeter (1934), the nature of the questions to be analysed and data availability implies that focus in this thesis lies on definition 1, the introduction of new products, and definition 2, the entry into new markets. The Oslo Manual is also adhered to, as my definition is that the innovation should be new to the firm and introduced on a market.

2.2. Drivers of renewal

2.2.1. Cost effectiveness through process innovation as an attempt to maximise profits

From textbook microeconomics, based on the neoclassical view of the firm, we know that firms want to maximise their profits $\pi$. The profit maximisation problem can be described by Equation 1.

$$\pi (q) = \max q \left( p q - c(q) \right) = p q^* - c(q^*) = \int_0^{q^*} \left( p - c'(q) \right) dq$$

The price $p$ multiplied by the quantity sold $q$ yields the total revenue of the firm, whereas $c(q)$ is the total cost of producing and selling this quantity $q$. The optimal quantity to produce, $q^*$, is found when the price $p$ equals the marginal cost, $c'(q)$, which implies that the total profit $\pi$ is equal to the integral under the supply curve.

The profit function is determined by the properties of the cost function. Assuming that the firm’s production function is a function of capital $K$ and labour $L$, as displayed in Equation 2, the short run cost function is displayed in Equation 3.

$$y = f(K, L)$$

$$\text{cost}(q|K) = \min_L rK + wL = rK + wL_s(q, K)$$

The cost function for quantity $q$, given the fixed amount of capital $K$ in the short run, implies that the firm only minimises the labour costs $wL$, whilst the cost of capital $rK$ is constant. The amount of labour needed is thus determined by the
production function, and its properties are determined by the complementarity and substitutability of the inputs as well as by the firm’s economies of scale. The production function can also become more efficient through process innovations, as these improvements change the proportions of inputs needed to produce the same output. For example, the automatisation of a labour-intensive stage in the production process may reduce production costs, as it allows for a reduction of labour when the cost of hiring is high.

Another way to increase cost effectiveness is to take advantage of economies of scope. This concept implies that producing a wide range of products is more cost efficient than producing one, and thus relying upon economies of scale only (Panzar & Willig, 1977, 1981). This could be one explanation as to why we observe that many firms produce more than one good each. In fact, I observe and analyse multi-product firms in several of the chapters in this thesis.

Producing several goods together is cost efficient because some inputs can be shared between the production processes. Examples are tangible inputs in terms of indivisible physical assets or proprietary knowledge (Teece, 1980). Goldstein and Gronberg (1984) add to this discussion by highlighting the relevance of space for cost effectiveness. Sometimes it is more cost efficient for a firm to produce several goods in one location than to produce them in multiple locations.

2.2.2. Global competition pushing firms to renew themselves through product innovation

Apart from cost efficiency and low product prices, another way to attract customers is to develop new products, in terms of goods and services. Through natural selection, non-profitable firms will be driven out of business. Those surviving will be those who manage to keep an advantage over others through investing in research and discovering (and applying) new technologies (Nelson & Winter, 1982).

The key behind this strategy is that commercialised innovations temporarily yield monopoly profits, which enable the survival of the firm. However, as imitations appear, monopoly profits decrease and eventually disappear (Aghion & Howitt, 1992; Schumpeter, 1934). As that occurs, pressure on the firm to introduce a new product to the market increases.

According to the trade literature, this process is further compounded by globalisation and trade, as competition is higher in international contexts, and only the most productive firms survive (Baldwin & Krugman, 1989; Bernard & Jensen, 2004; Das, Roberts, & Tybout, 2007; Melitz, 2003). To date, highly developed countries have managed to keep their advantage over less developed countries by constantly generating new knowledge and new technological innovations (Grossmann & Helpman, 1991).

This imbalance between developed and less developed countries is a component in product life cycles. The famous article *International Investment and International Trade in the Product Cycle* by Vernon (1966) explains how the firm renewal processes differ depending on the stage in the product life cycle. During
Introduction and summary of the thesis

In the first stage, as new products are introduced, there are many competitors with differentiated products that tend to locate in technologically advanced countries. This location pattern occurs because firms need to be close to both customers (and those with high purchasing power are located in these countries) and suppliers to communicate and make amendments to the production process. As the product matures and some product standards are developed, the focus shifts increasingly towards process innovation and increasing cost efficiency. In this stage, the product has become standardised, and the major issue for the firm is cheap labour to effectively produce high amounts of the product; this is why the production process at this stage is often outsourced to less technologically advanced countries, and a new cycle starts in the advanced country. The model was later extended to consider, e.g., tariffs (Segerstrom, Anant, & Dinopoulos, 1990) and the heterogenous capabilities of firms (Klepper, 1996).

The reasons for the increased role of international competition are threefold. First, the development of new technologies, and especially information and communication technologies (ICT), has allowed for a more rapid dissemination of information that extends beyond national borders. Not only has this contributed to and increased trade in high-tech production inputs and the outsourcing of stages of the production process (Baldwin, 2012), but it has also facilitated the trade and outsourcing of services (Busi & McIvor, 2008). Second, decreased transportation costs, as well as increased transportation speed, have further enabled firms to trade both intermediate and final goods. Potentially the largest reason for the lowered transportation costs is the development of large container ships (Hummels, 2007; Levinson, 2016; Rodrigue & Notteboom, 2009). The development of the jet engine for airplanes has also made it much easier to achieve rapid delivery (Hummels, 2007). These advances in containerisation and freight air traffic have also helped the development of new organisational models such as just-in-time production (Iammarino & McCann, 2013). Third, institutional changes such as reductions of tariffs through trade agreements and free trade areas (e.g., EU, NAFTA & Mercosur) have also made it cheaper and easier to trade goods and services (Levinson, 2016; Subramanian & Wei, 2007). Institutional changes can also occur on a national level, whilst still having a large impact on the international arena. Examples of this are the export-focused policies observed in some countries in East Asia, which have increased the education levels of the population and made it more attractive for foreign firms to establish themselves in these countries and for foreign investors to invest (Krugman, 2009; Lall & Teubal, 1998; Stiglitz, 1996).

Since these developments have made trade easier and made competition a world phenomenon, rather than a national one, it follows that firms that operate on global markets are of interest to study. In a small and open country like Sweden, these firms are highly relevant, as approximately 10% of all firms are exporters. Because these firms face higher competition than domestic ones, their

---

7 Based on figures from 2013, excluding firms that only consist of 1 person.
ability to renew themselves to stay ahead of the competition is even more important and so is our understanding of what facilitates this renewal.

2.2.3 Technological shifts pushing firms to renew themselves through product innovation

In addition to national or international competition pushing firms to develop new products, technological shifts can introduce large discontinuities and push firms to develop new products and services. Bresnahan and Trajtenberg (1995) call such shifts general purpose technologies (GPT). These technologies are “characterized by the potential for pervasive use in a wide range of sectors and by their technological dynamism. As a GPT evolves and advances it spreads throughout the economy, bringing about and fostering generalized productivity gains. Most GPTs play the role of enabling technologies, opening up new opportunities rather than offering complete, final solutions” (p. 84). This type of radical shift in technology is equivalent to the development blocks mentioned earlier (Dahmén, 1988; Schön, 2012), and they have also been called macro innovations (Kander, Malanima, & Warde, 2013), logistical revolutions (Å. E. Andersson, 1986), technology systems or techno-economic paradigms (Dicken, 2015) in other strands of literature. Due to the large discontinuities these GPTs introduce in substantial parts of society, they can be thought of as large-scale examples of Schumpeter’s creative destruction.

The current revolution in ICT is one example of a GPT (Lipsey, Bekar, & Carlaw, 1998). ICT, like previous technologies such as the waterwheel, electricity, steam power and combustion engines, have emerged over the years within the field of power. At their time of introduction, most depended on several existing physical components and/or previous technologies, implying that they were not independent innovations that could function in isolation of other technologies or products. Commonly, they can also be grouped into technology systems, meaning that they are used together to produce related goods.

Aghion and Howitt (1998) discusses the macroeconomic effects of the GPTs. They build upon their previous article (Aghion & Howitt, 1992) by incorporating intermediate products into the model. In the presence of GPTs, the economy will experience two stages that together create a business cycle. In the first stage, the GPT is discovered, seemingly like a by-product of something else or by accident. Whilst devoting R&D towards inventing complementary intermediate goods that will increase efficiency, productivity goes down, creating a slump. When the complementary goods have been developed, the full potential of the new GPT can be explored and productivity goes up, creating a boom. This creates a cyclical pattern for the economy. However, they also suggest a third stage in this model: a stage with technological spillovers. Firms not only learn how to adopt the new GPT not only due to their own commitment to R&D but also through the ways other firms handle that transition.

A prerequisite for the introduction of these GPTs is knowledge-intensive inputs, as new technologies often require large amounts of resources being spent
on research within both universities and firms. The dissemination is then mediated by geographical proximity (Aghion & Howitt, 1998). Torsten Hägerstrand, in his famous book *Innovation Diffusion as a Spatial Process*, describes how an innovation diffuses in a community through private communication (Hägerstrand, 1967). He says that “on the average, the density of the contacts included in a single person’s private information field must decrease very rapidly with increasing distance” (p. 235). This implies that diffusion is a spatial process, where the core learns about and adopts an innovation faster than the periphery. This mechanism is identical to that for knowledge spillovers, which explains why firms like to be “in the thick of things”. It enables them to, if at all, benefit more quickly from spillovers. Arzaghi and Henderson (2008) portray this behaviour in their study of the marketing industry in Manhattan, where they note that the benefits one marketing firm gets from other neighbouring marketing firms disappears after 750 metres.

2.3 The innovation process

Not only is it very difficult to define what an innovation is, but it is also difficult to explain how they occur. Over time, several models have been suggested, where the older linear model and the more recent idea of innovation systems are the most common. However, one should notice that they explain the context in which innovations are said to occur, but they do not provide exact recipes for the creation of innovations that could easily be adapted to other contexts.

2.3.1 The linear model of innovation

The linear model of innovation has an uncertain origin; some claim that it is a type of “folk model” that developed loosely almost on its own for decades (Balconi, Brusoni, & Orsenigo, 2010; Godin, 2006). Others claim that its roots originate in a report by Bush (1945). No proper model actually exists in Bush’s report, but rather some statements that can be considered building blocks for the linear model. For example, he separated basic from applied research in the discussion by claiming that basic research is done without consideration for specific applications. Additionally, he claimed that scientific progress was essential to technological innovations, which in turn fosters economic growth. From that perspective, the government ought to support basic innovations (Bush, 1945).

Godin’s view suggested a slower development over three stages and several decades. The first stage stretches from the early 1900s to the end of WWII and was a period with heavy focus on basic research and applied research. The second stage started in the early 1930s and ended in the 1960s. It highlighted technological development as another step in the model, and thus deviated from the view of pure science held previously. The third, and last, stage started in the 1950s, when economists added the importance of diffusion and the actual use of products to the model. The final model is displayed in Figure 1.
Figure 1. The linear model of innovation, adapted from Godin (2006).

Much criticism has been issued over the years of this model. One well-cited example is Kline and Rosenberg (1986), who criticise the model for being overly simplistic and thus distorting reality. One of their main points is the lack of feedback in the model, which they claim is crucial for planning ahead. Based on the success of an idea, future steps are decided, whether it means going back to the drawing stage or moving ahead to larger production volumes, etc. The linear model also overemphasises the role of science and underestimate the role of design, in their view, and ignores the importance of small incremental innovations or process innovations, which may both have large impacts on economic development.

Other criticisms commonly put forward are the clear division between basic and applied research, the unclear relation between basic research and technological improvements, the smooth process from research to diffusion, which is portrayed as being without impediments, the focus on only universities and large firms as relevant actors in the process, etc. (Balconi, Brusoni, & Orsenigo, 2010).

2.3.2. The chain-linked model

After the criticism against the linear model of innovation, an alternative was presented by Kline and Rosenberg (1986) – the chain-linked model. Instead of having one unidirectional movement through the model, as in the linear model, this model contains five different paths, shown in Figure 2. The path that is more or less equivalent to the linear model is the one shown by arrows going from left to right through the boxes “potential market” to “distribution and marketing”; this path can be called the central chain of innovation. What differs from the linear model is the allowance for feedback between all steps, as well as feedback from the last step to the previous ones. That is, input from customers and the results of marketing efforts provide information that is utilised for new designs, products, etc.
Introduction and summary of the thesis
Another important difference is that research, and the knowledge base it expands, is not relevant at only one point in the model but is here seen as a resource available throughout the whole process. This is represented by the vertical lines with bullet points connecting the steps at the bottom of the figure to knowledge and research. The two last paths are the arrow RI to the left and the arrow NR to the right. RI denotes radical innovations, which sometimes are enabled by new scientific discoveries, and NR denotes new research, which sometimes is enabled by innovations. Kline and Rosenberg (1986) use the telescope and microscope as two examples of innovations that enabled further scientific advances within fields such as astronomy and biology.

2.3.3. Institutions and innovation systems

“Institutions are the rules of the game in a society or, more formally, are the humanly devised constraints that shape human interaction” (North, 1990, p. 1)

The purpose of imposing these institutions is to reduce uncertainty and achieve desired outcomes. North divides institutions into three areas of interest: formal institutions, informal institutions, and the effectiveness of enforcement. Formal institutions, which are easier to observe, consist of rules, laws and contracts and only make up a small part of all the constraints that make us act the way we do. Informal institutions, such as norms, conventions, and codes of behaviour, have a much larger influence, although they often rest upon formal institutions. Thus, only observing formal institutions will not explain the diverging development between countries (North, 1990).

To understand the process of economic change, North claims that we must understand the change of these institutions. Effectively, what occurs when institutions change is that the incentive structure of the economy changes (North, 2005). North provides examples of how introducing patent laws increased firms’ incentives to innovate and how lowering the cost of enforcing contracts also lowered the transaction costs overall. As institutions affect the incentive structure, one could also say that they essentially affect the costs of production and trade.

If “institutions are the rules of the game, organizations are the players; it is the interaction between the two that shapes institutional change” (North, 2005, p. 59). Regions contain a variety of these organisations, such as economic organisations, political organisations, social organisations, and educational organisations. Economic organisations mostly consist of firms, which are embedded in regions, as are, to some extent, political institutions, since they not only exist nationally but also locally. Many countries, including Sweden, have local governments shaping the business climate for firms through regulations concerning permits, licences and tax levels. Social organisations can be religious bodies and clubs of various kinds, which bring together individuals with shared interests and beliefs. Educational organisations consist of schools and universities and thus increase the human capital of the region.
These types of organisations are also highlighted as important for economic development in the literature on innovation systems. Through the distribution of two books in the 1980s by Freeman (1987) and Freeman and Lundvall (1988), the concept of a national innovation system (NIS) became widely spread. This system describes the components involved in the dynamic process leading to knowledge development and innovations nationally (Lundvall, 2007). The concept is widely used in international contexts, both in research and in various types of strategies produced by the EU, the OECD, and the World Bank.

A regional innovation system (RIS) builds upon the same ideas and components as an NIS but is geographically concentrated in a smaller region (Cooke, 1996; Cooke, Uranga, & Etxebarria, 1997; Malmberg & Maskell, 1997). Although many organisations and institutions are national, collaboration and interaction between the different actors in the system is geographically bounded. Thus, a regional system coordinating regional resources is in many contexts more useful.

Based on earlier empirical studies, Hekkert, Suurs, Negro, Kuhlmann, & Smits (2007) propose seven basic functions that should exist within an innovation system: (1) Entrepreneurial activities, (2) Knowledge development, (3) Knowledge diffusion through networks, (4) Guidance of the search to optimise investment opportunities, (5) Ability of the innovator to enjoy temporary monopoly profits, (6) Mobilisation of human and financial capital, (7) Creation of legitimacy/counteract resistance to change.

Although their study was not focused solely on RISs, but innovation systems in general, many of these functions can be connected to organisation on a regional level. For example, knowledge development is dependent upon the availability of schools, universities, and research institutes that are embedded in regional contexts. Some of these functions, e.g., resources mobilisation, highlight the relevance of another type of organisation—financial organisations. Whilst stock markets usually exist in the capital markets and operate nationwide, the regional context, where the firms often prefer to search for financial capital, contains banks, venture capitalists, and business angels (Backman, 2015; Fritsch & Schilder, 2008; Pollard, 2003).

### 2.4. Bringing innovations to the market

With the foundations in place, new products and services can reach the market through two different channels, and both are relevant in the context of this thesis. The first channel is through new firm formation, which is Schumpeter’s classic way of describing the introduction of innovation (Schumpeter, 1934). This way of thinking, characterised in his earlier work, is often called Schumpeter Mark I. The second channel is through innovation activities in established firms. Later in his career, Schumpeter’s view on the individual entrepreneur as a requirement for innovation softened, and he also recognised large firms as important sources of
innovations (Schumpeter, 1943). This concept is often referred to as Schumpeter Mark II.

2.4.1. New firm formation

Any textbook in microeconomics will explain that a competitive industry is characterised by zero profits due to the entry and exit of firms in a neoclassical fashion (c.f. McDowell, Thom, Pastine, Frank, and Bernanke (2012) and Varian (2014)). When assuming free entry into the market, new firms enter an industry if the incumbent firms enjoy positive profits, i.e., when the maximisation problem in Equation 1 yields a positive value. As more firms enter, the price falls due to higher competition and greater industry supply – implying that each firm’s profits are pushed towards zero. As firms differ in productivity and thus face different cost structures, those firms that see the price falling below their average variable cost will exit the market. Mansfield (1962) showed how these factors could be aggregated to create a simpler entry function as shown in Equation 4.

\[ E_{it} = f(\Pi_{it}, C_{it} \ldots) \quad (4) \]

where the entry \( E_{it} \) is a function of \( \Pi_{it} \), the average rate of return the firm gets from entering industry \( i \) compared to all available industries; and \( C_{it} \), the amount of investment required to start a firm of the minimum efficient scale. The entry rate is thus positively affected by increases in \( \Pi_{it} \) and negatively affected by increases in \( C_{it} \).

However, this is a simplified version of market dynamics. As explained by Pakes, Ostrovsky, and Berry (2007), a potential entrant in a market needs to estimate the discounted sum of all future net profits. If this exceeds the entry cost, the firm will indeed enter the market and otherwise not. Even though a market has no barriers to entry, it does not mean that it is costless to start a business. Moreover, incumbent firms do a similar calculation, where their future profits must exceed the sell-off value of their capital for them to stay in business.

Geroski (1995) highlights some stylised facts about the entry and exit of firms, suggesting that this baseline model is also too simplistic. For example, he notes that real entry rates are difficult to explain using these conventional models, that the incumbent firms in the market seldom take explicit action as new entrants appear and that high entry rates not only have a displacement effect by crowding out less efficient firms but also introduce many innovations to the market. Many scholars have worked on extending this baseline model, attempting to make it more compatible with observed firm behaviour. One way to adapt the model is to change the way entry costs are measured, which appears to have large implications for the outcome of the model (Cavallari, 2015). Another way to adapt the model is to assume that firms face idiosyncratic shocks in their production, which leads some firms to exit the market (Hopenhayn, 1992). Other scholars have worked towards explaining why incumbents do not necessarily strongly react to the entry of new firms, as suggested by Geroski (1995). A common theme
is to focus on learning by doing (Jovanovic & Lach, 1989; Pakes & Ericson, 1998) or learning through other plants of the firm (Disney, Haskel, & Heden, 2003). The consequence of incorporating these into the model is that incumbents gain a productivity advantage due to their time on the market, which allows them to sit back when new entrants appear.

Klepper and Miller (1995) describe how in many sectors, it can be observed that the number of entrants increase quickly but that the number of firms then falls as the output increases – so-called shakeout periods. They adapt the model to incorporate the possibility of overshooting due to an inability of the firms to coordinate entry among each other, which helps explain why there might initially be too many firms in the market. Another way of modelling this is to connect the model more closely to product life cycles. This connection implies that early in the cycle, many firms enter the market, but as customers form their demand and a dominant design (Abernathy & Utterback, 1978; Suárez & Utterback, 1995) emerges, many of the firms become obsolete, and the survivors increase their production and focus on process innovation (Acemoglu & Cao, 2015; Agarwal & Gort, 1996; J. R. Campbell, 1998; Klepper, 1996).

A different way of thinking about new firm formation is to consider the maximisation problem of the entrepreneur regarding his/her career options. One way to do that is to estimate the net present value of various career options and choose the one with the highest value (C. A. Campbell, 1992). The expected net present value of a self-employed career option in comparison to a salaried position is displayed in Equation 5.

\[
NPV_e = E\left[ \int_0^t \left[ P_t(t) * Y_e(t) - P_t(t) * Y_w(t) \right] e^{-rt} dt - C(0) \right]
\]  

where \( P \) represents the probability of success, \( Y \) represents the average income, the subscripts \( e \) and \( w \) represent the case of the entrepreneurial option and the wage labour option respectively, \( r \) represents the discount rate and \( t \) represents time. The term \( C(0) \) is added to represent the fixed costs, both monetary and psychological, of starting up a new business. A positive net present value indicates that the self-employed career option is the most beneficial and a negative net present value that a salaried position is preferable. Similar modelling can be done comparing the utility the individual gains from different career options (c.f. Eisenhauer (1995), and Douglas and Shepherd (2000)). In such models, it is common to not solely focus on monetary gains but also incorporate the utility gained from the work tasks, flexibility, etc. enjoyed in each career option.

### 2.4.2. Innovations in firms

As mentioned previously in relation to Schumpeter Mark II, new products in terms of goods and services do not require the establishment of new firms to reach the market. Established firms also renew their portfolio with new products. One major reason to do this is to stay ahead of the competition and temporarily enjoy a monopoly position on the market, which implies greater profits. This is a common
behaviour, and several theories of the firm include this perspective of renewal. However, the innovations do require effort from the firms. As displayed in Figure 1 and Figure 2, there are plenty of inputs needed in the R&D process to eventually be able to introduce innovations on the market.

Griliches (1979) attempted to formalise this process with a knowledge production function, which is displayed in Equation 6. This function showed how a firm’s production function $Y$ depends on the inputs of technological knowledge ($K$), an index of regular inputs such as labour and capital ($X$), and other unmeasured inputs ($u$).

$$Y = f(K, X, u) \quad (6)$$

The focus of his study was on the development of input $K$, the technological knowledge of the firm. This input depended on a vector of the relative contribution ($W$) of a vector of previous and current R&D expenditures ($R$), as well as on other unmeasured inputs ($v$), as displayed in Equation 7.

$$K = g(W'R, v) \quad (7)$$

The implication of Equation 7 is that R&D expenditures are highlighted as the key to developing new technologies and innovations. To a certain extent, that is true. Conducting these R&D activities to develop new products does not occur for free in a black box. Firms need human capital, financial capital, and perhaps also physical capital, but that can be purchased with the financial capital.

R&D activities are carried out by highly educated and skilled individuals, either within a specific R&D department of a firm or through collaborations between skilled workers in firms, research institutions, and universities. These workers do not work for free just because the product they are developing has not yet been sold. Hence, firms need to pay out wages to produce a product that has yet to reach the market. When the product does reach the market and the firm enjoys monopoly power, it hopes to generate enough revenue to cover the cost of development (Aghion & Howitt, 1992). This implies that firms initially need to acquire financial resources elsewhere, either in terms of internally generated profits from previously sold products or in terms of externally acquired investments or loans from financial intermediaries. Due to the uncertain nature of innovation, acquiring this external capital may be difficult (Hall, 2002; Hall, Moncada-Paternò-Castello, Montresor, & Vezzani, 2016).

The other important input required for the R&D process is knowledge or human capital, which can be considered to be what Griliches (1979) modelled as technological knowledge at the firm level. As most important technological knowledge is embodied in people, and not within firms, it makes sense to consider the knowledge of the workers instead. There are three ways for a firm to acquire and utilise their knowledge. First, the firm can hire people with knowledge suitable for their needs. Second, the firm can cooperate with knowledgeable people at other firms, universities, or research institutes. Third, the firm can
absorb knowledge through spillovers. All these are dependent upon the regional context in which the firm is embedded, in one way or another.

Figure 3 displays the distribution of individuals with higher education, at least a bachelor’s degree, and the distribution of financial intermediaries in terms of bank branches, venture capitalists and public capital through the state-owned organisation Almi.

Figure 3: The share of individuals with at least a bachelor’s degree in 2013, and the location of universities and university colleges (left), and the location of financial institutions in 2013 (right). Source: Own maps based on self-collected data and data from Statistics Sweden and Almi.

The left-hand map shows that the variation between municipalities is very large; with some municipalities having less than 10 per cent of the population with a university degree and some municipalities having above 40 per cent with a university degree. What can also be observed is that there is a clear correlation between the location of the universities and university colleges and the education level in the municipalities where they are located. In the right-hand map, we can observe that bank branches are widely available all over Sweden, but a large number of the offices are found in large metropolitan areas. Venture capital is highly concentrated in Stockholm, with only a few offices scattered in other
Regarding the first point, the hiring of workers has a clear spatial dimension. When individuals look for jobs, their mobility is usually restricted to the labour market where they live (Johansson, Klaesson, & Olsson, 2002; Johansson, Klaesson, & Olsson, 2003). To illustrate this phenomenon, Johansson et al. (2002) show that the willingness to commute decreases drastically as the single journey extends beyond 15 minutes and comes to a halt at a low level of approximately 45 minutes. Therefore, firms that employ similar skills tend to co-locate to take advantage of labour market pooling (Henry & Pinch, 2000; Marshall, 1890; Pinch & Henry, 1999), thereby creating a denser labour market for specific skills with more potential for interfirm knowledge transfers (Almeida & Kogut, 1999; Franco & Filson, 2000; Møen, 2005; Song, Almeida, & Wu, 2003). This aspect is usually considered to be one of the sources of agglomeration economies (Duranton & Puga, 2004; Rosenthal & Strange, 2001). The larger the region is, the better are the chances of successful matches on the labour market (Helsley & Strange, 1990).

Regarding the second point, it is becoming increasingly common for firms to collaborate in R&D investments and to conduct open innovation (Chesbrough, 2003; Chesbrough & Crowther, 2006; Enkel, Gassmann, & Chesbrough, 2009; Rothwell, 1994). With these strategies, not only are the costs and risks shared between several firms, but the knowledge embodied by those involved can be shared, combined and utilised in the innovation process. This trend is driven partly by the increased complexity of modern technology but also by globalisation trends: a firm can generate new knowledge by being part of a multinational firm (Narula & Zanfei, 2006); invest in locations with specialised knowledge (Castelli & Castellani, 2013); and connect to global networks to compensate for a lack of high quality collaborations regionally (Chaminade & Plechero, 2014). Firms can also gain competitive advantages in foreign markets (Lavie & Miller, 2008). This also suggests that this point has a spatial dimension. Regarding the third point, which has the most evident connection to geography, one does not necessarily have to formalise collaborations or to hire people to be able to absorb new knowledge, but knowledge spillovers may be sourced in the local environment (Glaeser, 1999; Glaeser, Kallal, Scheinkman, & Shleifer, 1992; Rosenthal & Strange, 2001; Teece, 1992). Glaeser (1999) constructed a model of skill accumulation that led him to the conclusion that dense areas will have higher skill levels, which is made possible by the larger number of interactions that occur there with possibilities for knowledge sharing. Thus, these dense areas will contain a broader set of knowledge.

---

8 Although co-location is not the rule (Combes & Duranton, 2006).
3 Data and empirical issues

3.1 Sweden at a glance

All empirical chapters in this thesis use data on Swedish firms and individuals together with regional characteristics. As not all readers may be acquainted with the Swedish geography and administrative system, a brief outline together with some graphical representations of key characteristics are presented here.

Considering the area, Sweden is the fifth largest country in Europe, with only Russia, Ukraine, France and Spain being larger. However, according to the World Bank’s World Development Indicators, Sweden is only the 17th largest based on population. This implies that the population density is not very high compared to other countries. The map presented in Figure 4 shows how the population and population density are distributed over the Swedish geography.

The maps indicate that most of the population lives in the southern part of the country, and particularly in the three major regions, Stockholm (the capital), Gothenburg, and Malmö (which is part of the Öresund region just across the bridge from Copenhagen, Denmark). The population density is calculated based on the area in km² and the population of each municipality. In total, there are 290 municipalities (kommuner), and each one is self-governing, with a publicly elected government. Municipalities collect their own taxes from individuals and firms, and they are responsible for schools and pre-schools, elderly care, social services, waste collection and other public works, emergency services, and urban planning.

Each municipality belongs to one of the 21 counties (län), which are equivalent to NUTS 3 regions in the EU nomenclature. Each county has a county council (landsting) and a county administrative board (länsstyrelse). The county council is also publicly elected every four years, and its major responsibilities include health care other than elderly care, public transportation, and regional planning. The administrative board is appointed by the national government and is supposed to monitor regional development and coordinate regional policies with national policies. Nevertheless, the different policies and tax levels enacted by the local governments can lead to large variations in the conditions for businesses on a local level.

---

9 One exception is the police force, which is centrally governed.
10 The other NUTS- classifications are
NUTS0=Sweden (SE)
NUTS1= East (SE1), South (SE2), and North (SE3).
NUTS2= Stockholm (SE11), East Middle (SE12), Småland and the Islands (SE21), South (SE22), West (SE23), North Middle (SE31, Middle Norrland (SE32), and Upper Norrland (SE33).
Some of these differences can be seen in Figure 5. The left-hand map is based on an index conducted yearly by the Confederation of Swedish Enterprise and is based on a combination of register data concerning, e.g., self-employment, and surveys to businesses where they describe how they experience, e.g., local politicians’ attitudes towards business owners and access to human capital. The right-hand map shows the growth in establishments in the municipalities from 2003 to 2013, normalised by the number of inhabitants in thousands. What is interesting is that the map showing the highest ranked business climate largely does not coincide with the map showing the actual growth of establishments over the preceding 10-year period. The best business climate according to the index is found in the three metropolitan areas and central parts of southern Sweden, whereas the greatest growth in establishments can be observed in the northern part of the country. Important to note is that these figures only reflect the number of establishments and does not incorporate their size or their performance.
3.2. Measuring innovative behaviour

Not only are innovations difficult to define, as we observed previously in Section 2.1.2., they are also difficult to measure, which was highlighted as early as the 1960s (Kuznets, 1962). The issues Kuznets highlighted were the inability to identify the innovation and the difficulty of valuing it. An additional difficulty is noted by Griliches (1979), who said that many innovations may occur in industries where we have difficulty even measuring the output, such as health, defence, space research, etc.

Nevertheless, many different variables have been used as proxies in the literature over the years to attempt to measure innovation. They can be divided into three subcategories: (1) measures based on one variable only; (2) measures
based on a multitude of variables – indices; and (3) measures based on surveys, such as the commonly used Community Innovation Survey (CIS).

3.2.1. Unidimensional measures

One of the most commonly used proxies for innovation is R&D. As noted by, e.g., Acs et al. (2002), R&D is actually an input to innovation; nevertheless, in search for better proxies, many authors have used it to proxy also for innovation output. The specific form of R&D being used varies substantially between studies. Some studies use absolute numbers (Cassia, Colombelli, & Paleari, 2009; Varga & Schalk, 2004), the growth of the absolute numbers (Deschryvere, 2014), either of these normalised to another variable such as the population of firms or individuals (Akcakaya & ter Weel, 2009; Frenken, Van Oort, & Verburg, 2007; Love, Roper, & Bryson, 2011; Rodríguez-Pose & Crescenzi, 2008), or R&D expenses directed at specific areas (Brouwer, Kleinknecht, & Reijnen, 1993).

One of the most common criticisms of using R&D to measure innovation is that R&D only measures one input in the innovation process and not the actual innovations being achieved. Specifically, it only accounts for the budgeted resources a firm devotes to innovative activities, regardless of their success (Acs et al., 2002; Acs & Audretsch, 1987). Another critique that could be put forward is that the productivity of R&D, in terms of the innovations being generated by the R&D expenses, varies greatly between regions (Fritsch & Franke, 2004). Hence, using it as a proxy of innovation output between regions may be very misleading.

Another very common proxy for innovation output is patents, although they rather measure the intermediate step in the innovation process of discovering an invention (Acs et al., 2002). Various forms of this measure have also been commonly used, such as the number of patents in absolute numbers (Bilbao-Osorio & Rodríguez-Pose, 2004; Bottazzi & Peri, 2003; Lööf, Nabavi, Cook, & Johansson, 2013; Marrocu, Paci, & Usai, 2013), the number of patents normalised by another variable such as population or GDP/GRP (Cassia et al., 2009; Jalles, 2010; Salgado-Banda, 2007; Wong, Ho, & Autio, 2005), or the percentage of patents that are granted in several countries (Hasan & Tucci, 2010). There is also a difference between patents granted, which is preferred by some researchers, and patent applications, which is preferred by others.

Much criticism has been put forward for this proxy as well (Acs et al., 2002; Acs & Audretsch, 1987; Griliches, 1990; Pakes & Griliches, 1980). The most common is that not all patents result in innovations and vice versa, that not all innovations stem from patents. Consequently, what is being measured does not necessarily overlap much with what one intends to measure. Also, as Shepherd (1979) discussed, some patents are filed simply to prevent further development and competition, which means that these new products never develop into innovations. Another criticism from Griliches (1990) is the varying level of technology and economic significance between various patents.
Considering that neither R&D nor patents actually measure outputs from the innovation process, some attempts have been made to find proper output variables to be used as alternatives. Two such measures that are commonly used are literature-based measures and survey answers. The former refers to announcements of new products being made in technical journals and the latter refers to surveys where firms are asked, e.g., whether they innovate or not. This is the case for the CIS described in Section 3.1.3.

Literature-based measures were developed mainly in the 1980s and 1990s\(^\text{11}\) and have several strengths compared to the previous measures (Acs et al., 2002). Since the firms are not contacted, there is no problem with response rates, and the data search can be extended back in time and yield time series data sets (Kleinknecht & Reijnen, 1993). Coombs, Narandren, and Richards (1996) also claim that this measure is superior for tracking innovations in small firms, since they are less likely to apply for patents; it may also be good for international comparisons as long as a standard classification system is used. Because of these strengths, several studies have used this method to find a measure for innovation (Acs & Audretsch, 1987; Coombs et al., 1996; Corsino & Gabriele, 2010; Pavitt, Robson, & Townsend, 1987).

Nevertheless, there are also several weaknesses with this measure. Kleinknecht and Reijnen (1993) highlight that one must assume that firms have an incentive to report their innovations in these technical journals and that the selection of journals examined is sufficient and relevant for the measure to be useful. An additional weakness is that only substantial product innovations are counted with this measure, although, e.g., process innovations or smaller incremental innovations could also be economically beneficial (Coombs et al., 1996; Kleinknecht & Reijnen, 1993). Regarding firms of different sizes, the superiority in tracking small firms (Coombs et al., 1996) may turn into a disadvantage if large firms are underrepresented. This could be the case if they feel less need to report their innovations in these journals than small firms do (Acs et al., 2002).

Surveys of different kinds are superior if one wants to examine both product and process innovations (Coombs et al., 1996; Kleinknecht & Reijnen, 1993) but also for those researching service innovations (Love et al., 2011). Either the researchers can create their own surveys and send them out to a group of firms, as was done, in e.g., Love et al. (2011), or one could rely on a survey conducted by a statistical agency if such are available.

### 3.2.2. Multidimensional measures

Some organisations attempt to construct indices based on several variables to get a broader picture of innovativeness, mostly on the national level. Two such examples are the Global Innovation Index (GII) and the Innovation Union Scoreboard (IUS).

\(^{11}\) Although there are earlier examples of these measures such as Mansfield (1962).
The GII is constructed by the World Intellectual Property Organization (WIPO)\(^{12}\), the Johnson School of Management at Cornell University, and INSEAD Business School. The first edition of the index was published in 2007, and it has been conducted annually or semi-annually since then. The last version from 2017 includes 127 countries and uses in total 81 indicators divided between innovation inputs (institutions, human capital and research, infrastructure, market sophistication, and business sophistication) and innovation outputs (knowledge and technology outputs, and creative outputs). Half the weight of the main index comes from the innovation input sub-index and the other half from the innovation output sub-index. The main goals of this index are to improve the understanding of innovations and to help to identify important areas for policy improvement (Cornell University, INSEAD, & WIPO, 2017).

The IUS has been published annually by the European Commission since 2001. The 2017 edition covers the 28 EU members, as well as Norway, Iceland, Israel, FYR Macedonia, Serbia, Switzerland, Turkey, and Ukraine. In total, 27 variables are used and divided between ten groups of indicators, such as firm investments, finance and support, human resources, and attractive research systems. The goal of this index is to make a comparative assessment between the member countries and thus identify potential weak areas where specific efforts are needed to increase performance (European Commission, 2017a). A regional version is also conducted on the NUTS 1 and NUTS 2 level\(^{13}\), although on a smaller set of indicators due to data availability (European Commission, 2017b).

### 3.2.3. The Community Innovation Survey

One commonly used survey is the CIS, which is conducted in all EU member countries\(^{14}\) every second year. The survey asks firms about the innovation activities they have participated in during the last three-year period and their success in producing innovations. The definition of innovation is built upon the definition in the Oslo Manual (OECD, 2005) and asks about new product, process, marketing, and organisational innovations. For these to be considered new, the processes or products must be totally new or majorly improved and new to the firm. The same survey is sent out in all member countries so that cross-country comparisons are possible. The selected sample is stratified by sector and size, so that firms with at least 10 employees within selected NACE 2-digit level industries in the manufacturing and service sectors are included in the sample. Among firms with 250 and more employees, the survey covers the entire population, and for firms with 10-249 employees, it is a random sample with known probabilities (Statistics Sweden, 2012). Only micro firms, i.e., firms with fewer than 10 employees, are excluded.

---

\(^{12}\) The WIPO is part of the United Nations.

\(^{13}\) The regional IUS includes 55 NUTS 1 regions and 135 NUTS 2 regions.

\(^{14}\) The CIS also includes some non-EU countries such as Norway, Serbia & Turkey, which participated in the 8th wave conducted in 2012.
Due to the richness of questions regarding the innovation process, the survey has been rather popular, and it has been used to study a variety of research questions in several countries, e.g., Sweden (Lööf & Broström, 2008; Lööf & Heshmati, 2006; Tavassoli & Karlsson, 2015; Wixe, 2016), Finland (Deschryvere, 2014; Leiponen, 2005; Rouvinen, 2002), the UK (Battisti & Stoneman, 2010; Freel & Robson, 2004; Frenz & Ietto-Gillies, 2009), and Italy (Evangelista, Iammarino, Mastrostefano, & Silvani, 2001) or in a combination of several countries (Antonucci & Pianta, 2002; Hashi & Stojčić, 2013; Mairesse & Mohnen, 2002).

3.3. Measuring entrepreneurship

Another tricky issue to define and measure is entrepreneurship. As noted in Section 2.4.1., many scholars discuss entrepreneurship synonymously with self-employment. However, considering the larger body of entrepreneurial research, the two are not necessarily equivalent. Several studies have attempted to define what is meant by entrepreneurship by, e.g., surveying articles and books (Davidsson, 2004; Morris, 1998) or sending questionnaires to researchers, business leaders, etc. (Gartner, 1990). Such studies usually find a multitude of suggestions that are difficult to reconcile into a concise definition. For example, Gartner (1990) compiled a list of 90 attributes that 44 respondents considered to be central to their perception of entrepreneurship.

A common trend among many researchers is to include the creation of a firm or organisation as one important characteristic of entrepreneurship, among others such as innovation, value creation, ability to exploit opportunities, growth aspirations, and self-gain (Baumol, 1990; Bull & Willard, 1993; Gartner, 1988, 1990; Kirzner, 1973; Morris, 1998; Schumpeter, 1934), although it should be noted that even these characteristics can have different interpretations. One such example is the difference in the views of opportunity seeking seen in Schumpeter (1934) and Kirzner (1973). Whilst Schumpeter believed that the entrepreneur takes the opportunity to push market boundaries and start the process of creative destruction by exploiting completely new possibilities, Kirzner had the view that the entrepreneur discovers unexploited possibilities that allow society to become more efficient. When using a production possibilities curve to illustrate this example, the differences can be explained by the initial position of society: Schumpeter believed that society started on the productions-possibilities frontier pushing the boundary outwards, whilst Kirzner believed that society started below the curve, and the entrepreneur helped move the economy towards the boundary.

Despite this variation in characteristics and definitions, Davidsson (2004) suggests that one can combine these into two underlying processes: 1) the tendency of some people to choose self-employment over regular employment (cf Gartner (1988), and Gartner (1990)) and 2) the micro-level initiatives that help advance society (cf Shane and Venkataraman (2000)). A similar suggestion is made by Landström (2010), who suggests that scholars are either discussing the
development of new organisations or new opportunities. Both Davidsson and Landström note that the second process is broader than the first, as it may encompass the creation of new firms, but it is not limited to it.

If agreeing upon the distinction of the two processes, another problem arises, as one moves forward attempting to conduct empirical analyses of entrepreneurship. The most common measures of entrepreneurship are the self-employment of individuals (Evans & Jovanovic, 1989; Hamilton, 2000; Hurst & Lusardi, 2004; Van Stel, 2006) and the creation of new firms/start-ups (Acs & Storey, 2004; Audretsch & Keilbach, 2004; Feldman, 2001; Lee, Florida, & Acs, 2004; Van Stel, 2006). Many authors argue that these types of measures should not be considered to constitute entrepreneurship (Henrekson & Sanandaji, 2014; Hurst & Pugsley, 2011; Shane & Venkataraman, 2000). However, concrete alternatives are rather scarce. As written by Salgado-Banda (2007) “Although there are excellent concepts and ideas in the literature, there are not sufficiently good measures to proxy for entrepreneurship in the Schumpeter-Baumol context” (p. 8). As a measure of productive entrepreneurship, he suggests using patent information to identify individuals with innovative ideas who are interested in reaping financial benefits from them. Henrekson and Sanandaji (2014) provide another alternative measure by identifying founders of new ventures who are present in the list of billionaires presented yearly by Forbes Magazine. Henrekson and Sanandaji (2014) argue that these individuals have managed to grow and have a large impact on the creative destruction of the market and can be considered entrepreneurs in a “true” sense.

3.4. Knowledge

Knowledge was highlighted as important for the creation of innovations through the R&D process and through spillovers between firms. This latter feature is enabled due to two distinguishing features of knowledge (Romer, 1990): 1) the value of knowledge does not diminish as it is used, and it is not restricted to the use of a certain number of individuals at a time – which makes it a nonrival good; 2) to some extent, the use of a certain piece of knowledge cannot be restricted – which makes it partially nonexcludable. The implication of these peculiarities is that knowledge has the ability to remove the restriction of constant returns to scale in production functions, which makes it a very powerful input. As knowledge is the explicit focus of the first two papers and appears in the others as well, there is a need to discuss what it is, where it is found and how it can be measured.

Several taxonomies have been created in the literature to attempt to define knowledge. One such division is that into codified knowledge (information) and tacit knowledge (know-how) (Å. E. Andersson & Beckmann, 2009; Gertler, 2003;
The distinguishing feature between these categories is that codified knowledge can be easily transferred between individuals, and tacit knowledge cannot. Due to the complexity of tacit knowledge, face-to-face interactions are required to successfully transmit it\textsuperscript{16}. Consequently, this type of knowledge is much more distance sensitive, and the ability to successfully transmit it through network effects or knowledge spillovers is one of the reasons we observe agglomerations in general (Duranton & Puga, 2004; Glaeser, 1999; Storper & Venables, 2004) as well as sectoral agglomerations.

Another way of classifying knowledge is by dividing it into product specific information and more general information. Product specific information makes it easier for firms to exclude others through intellectual property rights. General information is more difficult to appropriate and therefore contributes to a large extent to the knowledge spillovers that enable firms collectively to avoid diminishing marginal returns (Grossmann & Helpman, 1991).

Another issue is where to find knowledge, and Warda (2015) suggests four potential sources. First, knowledge can be embodied in individuals (Å. E. Andersson & Beckmann, 2009). Second, knowledge can be embodied in goods and thus transferred through domestic and international trade between firms (Grossmann & Helpman, 1991; Romer, 1990), which is why we observe product life cycles as suggested by Vernon (1966). Third, knowledge can be disembodied in R&D activities and disseminated through citations of patents and other publications (Griliches, 1979; Jaffe, 1986, 1989; Xu & Wang, 1999). Fourth, knowledge can be found in a disembodied form in licensing (Arora, 1995; Cassiman & Veugelers, 2000).

As the main focus in this thesis is the knowledge embodied in people, the relevant measurements of knowledge are measures of individuals’ knowledge. In the literature, many different measures have been used, and they all represent different parts of individuals’ knowledge, which is usually obtained through schooling and work experience.

Among the measures based on education, the most common is an individual’s education level, measured either in number of years spent in school or as degrees achieved, indicating the length of schooling – e.g., whether an individual has a high school degree, bachelor’s degree, etc. (Faggian & McCann, 2008; Glaeser & Maré, 2001; Glaeser, Scheinkman, & Shleifer, 1995; Rodríguez-Pose &

\textsuperscript{15} Alice Lam further divides the knowledge types in her typology by saying that codified knowledge in terms of abstract, scientific knowledge is embrained in an individual and encoded in the collective. The tacit dimension, consisting of know-how that is more action oriented, is similarly divided between embodied knowledge in the individual and embedded knowledge at the collective level (Lam, 2000).

\textsuperscript{16} Some would argue that face-to-face interactions are no longer needed, as ICT develops, e.g., Cairncross (1997) with the book \textit{The Death of Distance} and O’Brien (1992) with the book \textit{The End of Geography}. However, several studies argue the opposite case (McCann, 2007; Storper & Venables, 2004), e.g., Gaspar and Glaeser (1998) argue that ICT should be considered a complement to face-to-face interaction instead of a substitute, and McCann (2008) argues that we observe the opposite in that the world is actually becoming more curved.
Instead of focusing on the number of years spent in school, one could consider what subject was learned during those years by looking at the type of education (Wixe & Andersson, 2016). All of these are used in one way or another in the empirical chapters.

Another strand of the literature focuses on knowledge obtained through working rather than schooling. One such example is to consider work experience or tenure (G. S. Becker, 1994; Kambourov & Manovskii, 2009; Mincer, 1962); another is to focus on what occupation individuals have (Gabe, 2009; Wixe & Andersson, 2016). Much focus has been put on creative occupations (Florida, 2002; Florida, Mellander, & Stolarick, 2008). Another way to work with individual occupations is to classify them based on the type of skills most frequently used in each occupation. Such examples are the division of routine/non-routine skills (M. Andersson, Klaesson, & Larsson, 2014; Autor, Levy, & Murnane, 2003; S. O. Becker, Ekholm, & Muendler, 2013; Spitz-Oener, 2006) or the division between cognitive skills, people skills, and motor skills (Bacolod, Blum, & Strange, 2009). These classifications are not used as much in this thesis, although work experience is proxied by age minus the number of years spent in school and before school.

4. Summary and contributions of the individual chapters

The rest of the thesis consists of four independent papers that all study firm renewal either in terms of innovative activities in firms or in terms of the self-employment of individuals. All four papers also highlight and study the local context in which the observed firms and individuals find themselves and how that is related to their renewal processes.

The first paper *Internal and external knowledge and introduction of export varieties* combines the renewal of firms and that of products, as it studies the joint influence of internal and external knowledge on both the number of exporting firms in a local industry and the number of product-destination links introduced on the export market (the extensive margin). These are together referred to as variety triplets – based on the three dimensions of firm, destination, and product. Furthermore, the influence of internal and external knowledge on the average unit price and the average quantity of product-destination links is assessed (the intensive margin).

These analyses use detailed export data as well as information about firms and their employees. In total, we identify 22 manufacturing industries in 290 municipalities and examine their situations in 2002 and 2006, as well as the

---

17 Some researchers have also added a fourth category to this classification – management and administrative skills (Backman, 2014; Johansson & Klaesson, 2011; Wixe, 2015).
change between these years, to obtain an indication of the role of internal and external knowledge both statically and dynamically. Internal knowledge is defined as the mass of knowledge found in each local industry, measured in terms of the years of schooling of employees. External knowledge potential is defined as the access to knowledge intensive business services (KIBS) and measured as the distance-discounted access to these knowledge intensive sectors in all locations in Sweden. A linear model is estimated by ordinary least squares for each component of the extensive and intensive margin.

The results suggest that internal knowledge is positively related to both measures of the extensive margin and to the average unit price, but not to the average quantity. Regarding external knowledge, it is positively related to the number of exporting firms and the average unit price, but not to either the number of product-destination links or the average quantity. However, in these latter cases, the interaction of internal and external knowledge has a coefficient that is positive and significantly different from zero. This suggests that external knowledge may not be beneficial for all aspects of the export flows for the average firm but that it does add a benefit for local industries that have high internal knowledge.

The study contributes to the trade literature by extending the understanding of export structures by not simply considering the total export volume but instead disentangling two aspects of the extensive margin and two aspects of the intensive margin. The study also contributes to the literature on firm renewal and regional development by utilising the accessibility of KIBS as a measure of the knowledge milieu surrounding these local industries and thus of the potential access to them in terms of knowledge mediation. A third contribution is related to the in-depth study of the micro-dynamics of the export sector in comparison to the aggregated flows. For example, the total number of variety triples changed only very slightly between 2002 and 2006, whereas approximately 71 per cent were replaced during that period, indicating that intensive entry-exit processes are present on the micro-level, whilst the macro level appears stable and static.

The second paper *Exporting firms’ absorptive capacity and innovative behaviour* builds heavily on the first paper and can be considered an extension. This longitudinal study covering the period 2003-2013 has a heavier focus on the firm level; a narrower focus on the export structure, as only the number of new products are included; and a deeper focus on the type of external knowledge available in the firms’ surrounding milieu. The purpose of the study is to examine firms’ internal knowledge in combination with their external knowledge and to examine their joint relation to firms’ ability to introduce new export products. The interest is to further explore the mediating role of internal knowledge in the process of absorbing external knowledge, as hinted at in the results of the first paper.

To accomplish this purpose, a data set with the full population of Swedish exporters is used. For each firm, the goods exported from year to year can be traced, implying that one can identify the introduction of new products, and this is used as the dependent variable. Internal knowledge is measured as the average
years of schooling among the firm’s employees, and external knowledge is measured as the related diversity of individuals’ type of education in the municipality where the firm is located. To enhance the focus on the relatedness of topics, this measure is restricted to individuals with engineering-related educations. A negative binomial model is estimated with maximum likelihood for four different groups of manufacturing firms, which are defined based on Eurostat’s classification of technology intensity.

The results show that most sectors only benefit from external knowledge diversity through the interaction effect. These results provide empirical support for the importance of the firm having absorptive capacity, making it able to fruitfully absorb external knowledge. These results also reinforce the findings in Chapter 2, where the access to KIBS was not necessarily beneficial on its own but required local industries with higher levels of internal knowledge.

This study contributes to the literature on firm renewal and regional development in three ways. First, the measure for product diversification can also be considered a proxy for innovative behaviour. The number of new products exported is based on commercialised products, which is often considered a requirement in the definition of innovation. That is not the case with other proxies commonly used in the literature such as patents or R&D expenditures, as explained in Section 3.1.1. The number of new products exported is also more in line with the requirement that an innovation be new to the firm, as suggested by, e.g., the Oslo Manual (OECD, 2005). Second, the main interest of the study is the interaction between internal and external knowledge, where external knowledge is based on the type of education of individuals. The interaction highlights the relevance of the firm’s absorptive capacity and how it determines the possibility of assimilating and utilising external knowledge in internal production processes. Third, the external knowledge available to firms is measured in terms of related diversity in technologically related subjects instead of overall diversity. This has two advantages: 1) the minimum value will indicate a heavy specialisation within one area, but this would still be an area that is relevant for manufacturing firms; 2) the maximum value is achieved by an even distribution between all fields, and if that is the case, these areas are also relevant for manufacturing firms; and 3) it reduces the issue of having too much diversity, which could be disadvantageous for the firms (Boschma, 2004).

These two papers both address the output of firms, which is the end product after many innovative projects, some successful and some not. The third paper, Access to banks and external capital acquisition – perceived innovation obstacles, moves the focus to an earlier stage in the process – the stage where financial resources are needed to initiate such innovation projects. The purpose of the study is to analyse how access to external financial institutions is related to firms’ perceived difficulties in obtaining external financial capital for their innovation processes.

To fulfill this purpose, geo-coded firm level data are combined with the CIS wave from 2010. In this particular wave of the CIS, firms were asked to rank how the seriousness of the problems they have experienced when attempting to acquire
external financial capital for their innovation activities, which allows this question to be further examined. The access to financial institutions is measured in several ways: 1) the Euclidian distance to the nearest bank branch and 2) the number of bank branches within 5 km of the firm. A logit model is estimated with maximum likelihood separately for manufacturing and service firms and separately for firms located in rural and urban areas.

The hypothesis is that lower access to banks is associated with a higher probability of experiencing problems acquiring external capital for innovation activities. The results show that a longer distance is indeed associated with a larger probability of experiencing external capital problems. In addition, the local supply of bank branches appears to be important, as a larger number of local bank branches is associated with a lower probability of experiencing external capital problems. For manufacturing firms, the number of bank branches within 5 km appears to be the main issue, whereas the distance to the nearest bank branch is the main issue for service firms. Considering that many bank branches have closed in some parts of Sweden during recent decades, the implications of these results are that firms may have difficulties conducting innovation activities due to lack of financial funds, which in the long run could affect firms’ profitability and survival unevenly across Sweden.

This study contributes to the literature on access to financial intermediaries for firms in how it measures the financial supply. Previous studies have usually focused on the total supply in a region, with the disadvantage being that intra-regional locations are not considered. By directly calculating the Euclidean distance to the bank branches, that issue is considered. Another advantage is that administrative borders do not restrict the measure. This is especially relevant for firms located close to the administrative border between two regions, where it is not necessarily the financial institutions within the same region that are the closest ones. Another contribution of the study is that it controls for the purpose of the bank loans, which reduces the heterogeneity associated with the use of the acquired funds, especially since innovations are associated with a high degree of uncertainty and result in high monitoring costs for the banks.

The last paper *Self-employment and parenthood* returns to the issue of industry renewal through self-employment and shifts the focus to the individuals and a more local context – the family situation. The purpose of this study is to examine whether individuals are more likely to be self-employed after having children, as well as to examine whether it matters if one controls for unobserved heterogeneity.

This is accomplished using micro data for the total population of Swedish individuals covering the period 2001-2012. This data set contains not only information about individuals’ career choices and family situation but also allows for matching individuals to their partners and parents. A linear probability model is estimated in both a pooled version and a fixed-effects version to see whether controlling for unobserved heterogeneity results in biased estimates. To account for the different job market opportunities available to individuals with university
degrees and individuals without, as well as for possible gender differences, regressions are run separately for these groups.

The results for the pooled model are consistent with most previous studies, indicating that individuals are indeed more likely be self-employed after having children (Beutell, 2007; Ekinsmyth, 2013a, 2013b; Gimenez-Nadal, Molina, & Ortega, 2011; Kirkwood & Tootell, 2008; Rønsen, 2014; Wellington, 2006). However, controlling for the possibility of unobserved heterogeneity, the results indicate that individuals are less likely to be self-employed when having a family. These results are in direct contrast to most previous studies, one exception being a Danish study concluding that the difficulties with combining self-employment and parenthood are not just attributable to the welfare system but also to the loyalties and trust that businesses build up with their customers (Neergaard & Thrane, 2011). This issue could be a contributing factor to my results as well. Another explanation could be the terms and conditions of the Swedish welfare system, which in its current form calculates monetary compensation in a way that favours employment over self-employment.

This study contributes to several aspects of the literature on self-employment. First, a quantitative method based on longitudinal micro data of the population is used rather than surveys or interviews based on smaller samples. This allows controlling for unobserved heterogeneity and reduces potential selection issues that could arise in survey data. Second, both genders are included in the study due to the high level of gender equality in Sweden, which has generally not been the case in these studies. Third, the study is based in Sweden, which has a generous welfare system supporting working parents; this is different from, e.g., Anglo-Saxon countries, where a large number of studies have been conducted. Fourth, in addition to considering the individuals’ own characteristics, it also considers individual’s relatives and others in the geographical vicinity as potential influences on self-employment attitudes.
References


Introduction and summary of the thesis


