Masterthesis

Henk Jan Reinders

MEASURING INNOVATION AND THE PROPENSITY TO PATENT

TU Delft

TITLE PAGE

Title	Measuring Innovation and the Propensity to Patent
Author	Henk Jan Reinders (1239620)
Date	July 29th, 2011
Phone	+31 6 10448423
Email	h.j.reinders@student.tudelft.nl
University	Delft University of Technology
	Faculty of Technology Policy & Management
Program	Management of Technology
Section	Economics of Innovation
Graduation supervisors	Alfred Kleinknecht (Economics of Innovation) - Chairman
	Petra Heijnen (Energy and Industry)
	Zenlin Kwee (Economics of Innovation)

Abstract

Patents are often taken as an indicator to measure innovativeness, because they are a lot easier to obtain than most others. There is a complication though, which is that patents are not a direct measure of innovativeness: the fit between patents and innovativeness is biased by differing propensities to patent. The propensity to patent was analyzed in the 1980's and 1990's, but results vary and more recent research focuses mostly at specific case studies. With the availability of the CIS (Community Innovation Survey) databases of 2000 and 2004 there is a good opportunity to test again, in a structured way, which factors are important for the propensity to patent. For this research CIS data from three North-Western European countries were analyzed: Belgium, Norway and Germany. Descriptive results show that the propensity to patent varies greatly among different types of innovation, and especially when looking at the difference between goods, services and processes. Process and service innovations turn out to be rarely patented, with average patent propensities of respectively 7.6% and 9.6%. Moreover, descriptive results confirm that the propensity to patent differs across industries.

Next to that a logistic regression analysis was performed. This analysis tested existing hypotheses and explored new factors that came available in the CIS questionnaires. By only including those enterprises in the analysis that actually had innovative output, factors could be determined that increase the propensity to patent these innovations. Factors that were found significantly relevant include EU funding (+), having a new to market innovation (+), cooperation arrangements with universities (+), having a local/regional market (-), having a market outside Europe (+) and using private R&D institutions as information sources for innovation (-). On top of those it was confirmed that patent propensity is higher in specific industries and countries, as well as for different types of innovation, even when correcting for the aforementioned factors. Results provide a comprehensive overview of the importance of, and some correlations between, firm level, industry level and country level factors of the propensity to patent. These results can be used to improve patent based innovation measures, as well as to provide additional insights into appropriability conditions between sectors.

Keywords: innovation, propensity to patent, innovation indicators, CIS, patents

PREFACE

This document represents my final thesis for the Master of Science in Management of Technology at the TU Delft, faculty Technology, Policy and Management.

I started this program two years ago, with the goal to learn more about the societal and managerial context of technological progress, as well as developing my skills as a researcher and as a professional. This cumulated in a specialization in innovation economics, which is the field this thesis hopes to make a contribution to. In my opinion there are few things more interesting than how technological progress can lead to sustained economic growth and improved welfare. It is at least a field which I feel very connected to, and which also resulted in my decision to pursue an additional master's degree in economics at the University of Amsterdam next year.

Here I would also like to take the opportunity to show my appreciation to my graduation committee, to Alfred Kleinknecht for the many meetings and discussions about the topic and to Petra Heijnen and Zenlin Kwee for their critical and very useful feedback to draft versions of the thesis. Furthermore to Dr. Rammer for providing quick and extensive information on the German CIS questionnaire. Also I would like to thank Martin de Jong for his flexibility, as he provided me with the opportunity to combine the preparations for this thesis with an internship at Roland Berger at the start of this semester.

I greatly enjoyed the master's program, as well as the process of writing this thesis.

Henk Jan Reinders

Utrecht, July 2011

TABLE OF CONTENTS

INTRODUCTION

Scientific and technological advancements are a main driver of economic growth. To capture the economically beneficial part of these technological advancements, new products and processes are being developed. It is important to understand the factors which influence this process of innovation, in order to optimally allocate the societies resources in creating new products and processes. The overall importance of this process can hardly be understated – just imagine how the world would look like without commercial aircraft, computers or today's medicine.

To make informed policies aimed at improving the innovation process it is important to have measures for innovative output. Within the field of understanding the innovation process, the objective of measuring innovativeness is a long recognized problem. (Kuznets, 1962, Griliches, 1998, Kleinknecht et al., 2002) Often then, the amount of patents applied for (or granted) is taken as a proxy for innovativeness. The correlation between patents and innovativeness is not perfect, however. To explain this discrepancy, several factors have been identified which moderate the relationship between R&D expenditures and the probability that a company will apply for a patent. The propensity to patent is a term coined to capture some of these factors. (Scherer, 1983)

RESEARCH PROBLEM

Appropriability and patents

For private organizations a main goal is to make a return on invested capital. For the private sector to produce innovation, then, it is important that the benefits of this innovation can be – at least partially – appropriated by the innovating firm. If this would not be the case, firms would not have the necessary incentives to invest in R&D to produce those innovations, since a competitor could copy them at little or no cost. Because of this public good nature of knowledge, many countries have patent systems that provide monopoly rights to inventors to capitalize on their R&D investment. (Teece, 1986)

The idea that patents are a necessary tool to provide incentives to innovate is challenged in some cases, since other means of appropriation exist. Levin et al. (1987) identify several of these means, including lead time, learning curves, sales & service and, especially, secrecy. Moreover it was found that the importance of these means differs between industries. An important implication of this finding is that patents may not accurately measure innovative output, depending on whether or not patents are a good means of appropriation. This especially goes for process innovation, where secrecy is often preferred over a patent. (Levin et al., 1987)

From a societal perspective, a high propensity to patent – compared to for example secrecy – may not necessarily be a bad thing for more than that it provides the necessary incentives. This is because patents also have the property of diffusing knowledge to the public domain, either when the patent expires or even before that, on the basis of the contents of the patent. This content may then give direction to the research efforts of other companies. (Cohen et al., 2002)

Insight into the propensity to patent is needed to determine the functioning of the patent system, especially when comparing between sectors. Levin et al. (1987) mention in this respect the difference between the production of aircraft and the production of new drugs. For the first, patents are of minor importance because lead time will provide the necessary appropriation conditions. For the latter however, patents are of major importance to appropriate benefits from R&D efforts. It may therefore be beneficial to discriminate across sectors with different patent propensities / appropriability regimes when implementing new policy.

Measuring the propensity to patent

Besides providing insight into appropriability conditions, patent propensity has important implication for patent based innovation measures. Measuring innovative output is, as shortly stated in the introduction, a problem that has no definitive answer yet. To estimate innovative output then, several proxies exist of which R&D and patent applications are the ones that are most often used. (Kleinknecht et al., 2002) The idea behind the use of patents as an indicator is that both patents and innovative output have a common causal factor, namely R&D expenditures. However in order to use patents as an indicator for innovation there must be a strong correlation between patents and innovative output. From the conceptual model in figure 1 it can be seen that the propensity to patent plays a major role in biasing this patent measure.

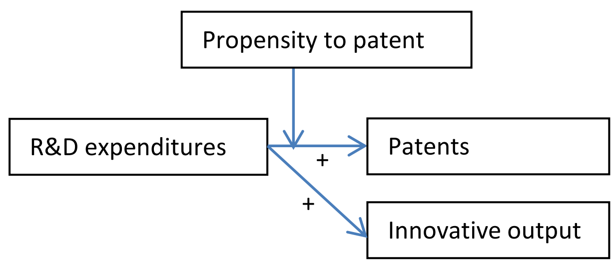

Figure 1 – Conceptual model

Since public policy is increasingly concerned with promoting innovation and economic growth there is then a clear need to have more robust innovation indicators. However, the factors which determine the propensity to patent have only been evaluated partially and mostly more than at least a decade ago. (Griliches, 1990, Brouwer and Kleinknecht, 1999, Arundel and Kabla, 1998) With the availability of new Community Innovation Survey (CIS) data, it is then possible to test again, in a structured and comprehensive way, the role of the propensity to patent. Moreover, it will be possible to look at the combined impact of country, sector and firm level variables.

RESEARCH OBJECTIVE

So, based on the previous paragraph, in order to know whether patents are a good indicator for innovativeness, it is important to know to which extent patents are biased by the concept of patent propensity. In fact, when the factors which determine the propensity to patent are better understood, this will be very valuable to evaluate and improve the use of patents as a tool in measuring innovation. Moreover, the propensity to patent includes the concept of appropriability.

A sound understanding of the propensity to patent will therefore provide insight into the biases of patent measures as well as providing insights into the importance of different forms of appropriability. The fourth version of the CIS (Community Innovation Survey) provides the information necessary to get better answers to these questions. The objective of this research will therefore be to explore the determining factors for the propensity to patent, based on the most recent available CIS database.

RESEARCH QUESTIONS

Based on the research problem and objective the main research question will focus at the propensity to patent. Moreover, since the goal is to find discrepancies between innovative output and patent measures (in order to improve upon them), this thesis will look at the propensity to patent an <u>innovation</u>. Hence, the main research question is:

- **Which factors affect an innovating firm's propensity to patent?**

To further clarify this question, it first has to be clear what is meant by an innovating firm. Different types of innovation exist, which may not have equal propensities to patent. CIS for example makes a distinction between product and process innovations. If these different types of innovation turn out to have different patent propensities (or different factors determining that propensity) the answer to the research question will then depend on the definition of an innovating firm. Therefore the following sub question is defined:

- Are there any differences in propensities to patent among innovating firms, according to the type of innovation they introduced?

Then, apart from the differences in types of innovation, it will also be useful to look at different levels of the propensity to patent. At this point it is mainly shown that the propensity to patent differs between sectors. (Hagedoorn and Cloodt, 2003) However, both at country and at firm level, differences in patent propensities may also be present. (Brouwer and Kleinknecht, 1999, van Pottelsberghe de la Potterie and de Rassenfosse, 2008) The CIS questionnaires focus at the firm level. Findings may then be linked to existing literature at the sector level and, in a more limited way, to existing literature at the country level. Further questions then are:

- Do patent propensities differ across sectors, and moreover, are these in line with earlier results?
- Do patent propensities differ across countries, and moreover, are these in line with earlier results?
- Can these differences be explained by firm level factors?

Combined, these questions are expected to give insight into the relevant[1] factors linked to the propensity to patent and thereby contribute towards building a better understanding of the concept of propensity to patent.

[1] In the following econometric analyses, relevance will be based on statistical significance. In this context, unless otherwise stated, significant will mean significant at the 0.05 level for p-values. This is equivalent to a z-score equal to or higher than 1.96, which will sometimes be used to avoid too many zeros in shown results.

Scientific and management relevance

A sound understanding of the propensity to patent is important for several reasons. Based on the research problem, two important policy issues have been raised that link directly to the propensity to patent. First, the propensity to patent is a major factor in biasing patents as an innovation measure. This is even more important since European and national policies are more and more aimed at stimulating innovation and economic growth. These policies can only be correctly evaluated with patent measures if the drawbacks of these measures are known. Moreover, the propensity to patent provides some of the underlying factors that are causing the bias in patent measures. If known, these can then be used to measures of innovative output.

In this light it may be worth mentioning that the 7^{th} Framework Program of the European Commission distributes an amount of €32bn to stimulate collaboration and innovation. (EC, 2006) To use this money wisely, an understanding of the innovation process and its measurement are essential.

Besides that, the propensity to patent has strong ties with the concept of appropriability of innovation efforts. If other means of appropriation (compared to patents) become more important, the propensity to patent is expected to drop. Hence, the propensity to patent provides valuable insight into the relevance of patents, especially between sectors. This information can be used to with regard to improving patent regulation.

From a more theoretical perspective, measuring innovation is still an important topic in the innovation literature, and no conclusive innovation indicator has been found (van der Panne, 2007). So, from a more scientific point of view, this research will contribute towards understanding the advantages and disadvantages of patents as an innovation indicator. Besides that, identified factors can then be accounted for when using patents as an indicator, making it a more valid and robust tool to measure innovation. This is especially important because patents are (maybe too) often very easily used as a general indicator of innovativeness, even across industries.

Research boundaries

This thesis will be mainly based on data provided by the CIS4 questionnaire. Since this data does not provide unique identification of firms, it is not possible to link the data to other databases. All factors included in the research will therefore be restricted by availability in the CIS database. In terms of timespan, CIS4 focuses on the years 2002-2004. In total three years are covered by this questionnaire. If necessary, a comparison can be made with CIS3, which was also available for this research. CIS3 covers the years 1998 to 2000.

This research focuses on three countries: Belgium, Norway and Germany. A main reason for this is the availability of high quality data for these countries. An additional advantage is that these (Northern-European) economies are quite comparable and as such are expected to provide clear results that are less dependent on country level differences. However, one negative consequence hereof is that this research will mainly be able to draw conclusions for a limited geographical region, and additional research may be needed to generalize results to other countries.

To further narrow down the research population, a focus will be placed on product innovation due to its natural link with patents. In the more descriptive results process innovations will also be included. However, from these descriptive results it will be seen that product innovation is (by far) the most important source for patent applications. The subsequent econometric analysis will therefore focus on product innovation.

THEORY

PREVIOUS APPROACHES IN THE LITERATURE

Measuring the relation between R&D and patents

In order to use patents as an indicator for innovation there must be a strong correlation between the two. Scherer (1983) investigates the relationship between R&D expenditures and patenting. It is shown that industry differences can account for a large part of the variation in patenting compared the R&D expenditures. So within individual industries, patenting commonly rises proportionately with R&D effort. (Scherer, 1983) Moreover, it is plausible that appropriability conditions may account for part of the differences between industries, since appropriability is likely to be an industry phenomenon. (Levin et al., 1987)

In itself the research of Scherer (1983) shows clearly that there is no absolute ratio of R&D input versus patent output, especially not between industries. In other words, the amount of patents per dollar of R&D input differs substantially across industries. Although this can be an indication that patents are not directly suitable to measure innovativeness, this is not necessarily true. This is because innovativeness itself may also be expected to have different R&D input to innovative output ratios across industries.

Later on several authors have argued for a more refined investigation of the propensity to patent. Both Arundel and Kabla (1998) and Brouwer and Kleinknecht (1999) identify the problems with R&D efficiency and account for this. In the same way, a moderator variable between R&D expenditures and innovativeness is identified as *research productivity* which is conceptually different from the propensity to patent (de Rassenfosse and van Pottelsberghe de la Potterie, 2009). This is depicted in the conceptual model in figure 2.

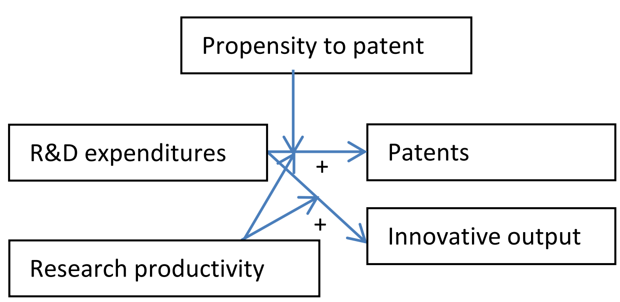

Figure 2 – Extended conceptual model

This conceptual model implies that to be able to filter out the moderator variable research productivity (efficiency) there is at least a direct measure for innovativeness needed. When assumed that there are no other relevant concepts to be included in the model above, it can be seen that the propensity to patent can be found by comparing patents to innovative output. Any factors that can explain a difference between patents and innovative output can then be said to belong to the concept of propensity to patent.

For example, when it is found that there are industry differences in the amount of patents applied for, while the amount of innovation is the same, then it can be said that these sectors have different propensities to patent. This is the approach that Arundel and Kabla (1998) take. In this way, the propensity to patent provides the factors that should be taken into account when using patents as indicators for innovativeness.

The important conceptual point which is made here is that the efficiency variable does not only impact the R&D input to patent output relationship, but also the R&D input to innovative output relationship. In order to find the systemic biases in patent statistics as indicators for innovative output, the latter approach to propensity to patent is likely to be a more fruitful one. In the remainder of this thesis this approach will be termed *propensity to patent innovations*, which is in contrast to the *propensity to patent* which links to earlier works.

FACTORS DETERMINING THE PROPENSITY TO PATENT

From here an overview will be given of earlier work on the subject. The main goal is to identify relevant factors for the propensity to patent as well as identifying important issues that rose along the way. Worth noticing is that several approaches exist, and great care has to be taken in interpreting and comparing results. This is not only so for the discussed approaches above (between propensity to patent an innovation versus propensity to patent an invention), but also for the different research populations that form the basis of those analyses. For example, Arundel and Kabla (1998) only focus on large firms. Representativeness of the samples discussed below is therefore an important issue to take into consideration when comparing earlier results.

Based on short reviews of relevant papers, summary tables are provided to give an overview of relevant factors and their significance. The factors included always focus on product innovation and only those factors are given that were directly linked to the propensity to patent. For example Acs and Audretsch (1988) and (1989) have analyzed multiple factors, but only those two factors are provided where a discrepancy between patenting and innovation was found, since it is exactly that discrepancy which indicates differences in the propensity to patent a product innovation. Only for Scherer (1983) all factors are provided, since he did not yet disentangle the efficiency factor from his analysis and therefore a split cannot be made.

Industry level and firm level factors

Kuznets (1962)

One of the first authors to write on the subject was Kuznets (1962). Although he does not directly test relevant factors impacting the propensity to patent, he does state several difficulties with using patents as an output indicator for inventive activity. Many of these difficulties may actually be hypothesized to impact the propensity to patent, such as firm size and government support. Next to that, Kuznets also identifies problems with the economic value of patents, since there is no ground to assume that all patents (or inventions for that matter) have equal value. (Kuznets, 1962)

Scherer (1983)

Two decades later, Scherer (1983) looks at both industry as well as firm level factors when analyzing the propensity to patent. Worth noticing immediately is that Scherer looks at the propensity to patent by relating it to R&D input. Therefore he looks at patent propensity as the number of patents per unit of expenditure on R&D. Hereby it is not possible to filter out efficiency effects.

Scherer finds that there is a relationship between R&D input and number of patents. Also, he finds that the explained variance increases by including dummy variables for industrial sectors. This implies that there are differences in the R&D input – patent output relationship between industrial sectors. Moreover, the explained variance increases further as the level of detail of industrial sector classification increases. So, within broad sectors there are again differences in patent propensity that can be explained by including narrower sector dummies.

In between a sector level analysis and firm level factors, Scherer devotes some attention to the Schumpeterian hypothesis, which states that large firms are technologically more progressive than small firms. On the basis of this, one could expect that large firms have more patents per dollar R&D than small firms do. By looking at increasing and decreasing (patent) returns to R&D investment, Scherer finds that there is an indication that there are decreasing returns to the amount of R&D investment. This is a finding contrary to the earlier stated Schumpeterian hypothesis. Two different explanations can be given for this: that larger firms generate fewer patentable inventions per dollar of R&D or that larger firms choose to patent fewer inventions.

At the level of individual firms, Scherer looks at several factors that may impact the propensity to patent. Quite robust evidence is found for increased numbers of obtained patents per firm for firms that have predominantly overseas sales. Also, diversification seems to unambiguously stimulate patenting. Other variables have a more mixed message. Federal funding is found to be significant, but changes the sign for its coefficient when a more detailed sector classification is used. (Scherer, 1983)

Scherer, 1983	*Regression with number of patents as dependent variable, no selection of firms and thereby includes efficiency effects*			
Factors	**Significant / sign**		**Present in CIS**	**Level**
Overseas sales	✓	positive	✓	Firm
Diversification	✓	positive	✖	Firm
Federal funding	✓	mixed	✓	Firm
Scope of invention use	✖		✖	Firm
Invention type	✖		✖	Firm
Other issues raised				
Differences across industries				
Schumpeterian hypothesis				

Table 1 – Factors and issues identified by Scherer (1983)

Bound et al. (1984)

As part of a book edited by Griliches (1984), Bound et al. (1984) ask the question of who performs R&D and who patents. The main thing of interest here is the finding that small firms patent more per dollar of R&D investment than large firms do. This is in line with earlier findings of Scherer; although this time the observations are more conclusive, since more small firms were included in the sample. This again seems to run against the Schumpeterian hypothesis. However, Bound et al. do question the validity of this conclusion, since there may be a selection bias for small firms. This is because small firms were only included in the analysis if they were somehow successful and thereby registered. For large, mostly publically traded, firms this selection plays almost no role. (Bound et al., 1984)

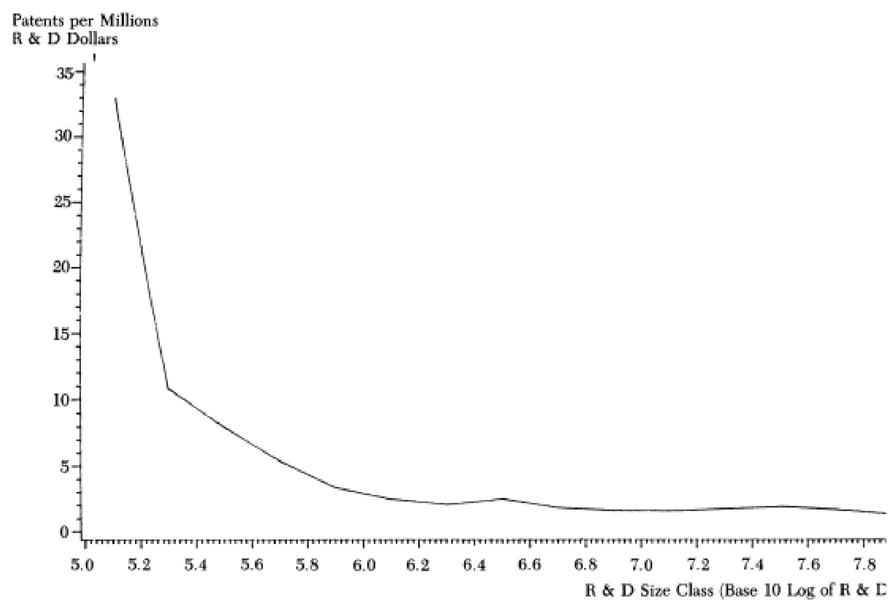

Figure 3 – Patents per Million R&D dollars by the base 10 log of R&D for firms with both R&D and patents (Bound et al., 1984)

Pavitt (1985)

In his 1985 review paper, Pavitt identifies several subjects on which more research is required. For the propensity to patent this includes sector specific factors related to the effectiveness of patenting as a barrier to imitation, when compared to alternatives. Next to that it includes factors that are related to the perceptions of cost and benefit of patenting, both at a firm level and at a country level. This implies that firms that have many alternatives to patenting as a means of protection would have a lower propensity to patent. Also, for inventions (and resulting innovations) with small economic value it may economically not be viable to apply for a patent, since it incurs additional costs for application as well as enforcement of the patent. (Pavitt, 1985)

Mansfield (1986)

Mansfield (1986) compares the percentage of patentable inventions that were patented between two groups: all firms and firms with 1982 sales exceeding $1 Billion. It is found that large firms patent generally more of these, whilst this higher patent propensity *an invention* is significant for at least the pharmaceuticals, chemicals and petroleum industries. Combined with the finding that small firms generally have more patents per dollar R&D invested than large firms, this gives an indication that smaller firms have a higher number of patentable inventions per dollar of R&D compared to the larger firms. (Mansfield, 1986) However, this does not mean that there are no advantages to large firm size, because of the idea of R&D cost spreading. (Cohen and Klepper, 1996)

Acs and Audretsch (1988 &1989)

In their 1988 paper, Acs and Audretsch use a direct measure of innovative output to find differences between 247 different industries according to the SIC classification. The impact of several factors on this measure is analyzed using a regression analysis. They find that the numbers of innovations increases with industry R&D expenditures, however at a decreasing rate. Moreover they find that industry innovation tends to decrease when the level of concentration or the level of unionization in that industry rises. (Acs and Audretsch, 1988) This may be due to a size effect: firms in concentrated industries are generally bigger in terms of turnover.

Then, building on their 1988 paper, Acs and Audretsch perform the same regression analysis but then to a measure of patenting (number of patents.) By comparing both outcomes they find that both capital intensity and unionization show differences between the models. In the model that predicts patented inventions, capital/output is significant. Also in that same model, unionization is not significant. When looking at the model that predicts innovative output, however, capital/output is not significant, while unionization is significant and negatively related the total number of innovations. These differences can then be thought of as indicating factors that influence the propensity to patent innovations. (Acs and Audretsch, 1989)

Acs and Audretsch, 1988 & 1989	*Regression analyses with both number of patents as well as innovative output as dependent variables, at industry level*		
Factors	**Significant / sign**	**Present in CIS**	**Level**
Capital/output	✓ positive	Industry variable	Industry
Unionization	✓ positive	Industry variable	Industry
Other issues raised			
Schumpeterian hypothesis			

Table 2 – Factors and issues identified by Acs and Audretsch (1988; 1989)

Arundel and Kabla (1998)

Arundel and Kabla (1998) use a joint survey by MERIT (Netherlands) and SESSI (France) to provide empirical estimates of the propensity to patent both product and process innovations, by industrial sector. Descriptive results show large differences between sectors. Especially in sectors such as textile and clothing, petroleum refining, basic metals, transport and telecom services the large majority of innovations are <u>not</u> patented.

Regression analysis then shows several factors to have a significant impact on the propensity to patent. These include the logarithm of sales (as a control measure of firm size; positive), the importance of patenting (positive), the importance of secrecy (negative) and sales in the US and/or Japan (positive). Besides that, the fact that a firm is based in Germany is also found to have a significant and positive impact on the propensity to patent. One serious drawback of the MERIT (PACE) data is that it only includes the European Union's 500 largest R&D performing industrial firms. These factors thus cannot be automatically taken to be (significantly) important for smaller firms too. (Arundel and Kabla, 1998)

Arundel and Kabla, 1998	*Regression on propensity to patent product innovations*		
Factors	**Significant / sign**	**Present in CIS**	**Level**
Firm size (sales, logarithm)	✓ positive	✓	Firm
Opinion about importance of patents	✓ positive	✖	Firm
Opinion about importance of secrecy	✓ negative	✖	Firm
R&D intensity	✖	✓	Firm
Sales in markets outside of Europe	✓ positive	✓	Firm
Opinion about importance of earning license fees	✖	✖	Firm
Located in Germany	✓ positive	✓	Firm
Other issues raised			
Differences across industries			

Table 3 – Factors and issues identified by Arundel and Kabla (1998)

Brouwer and Kleinknecht (1999)

In their 1999 paper, Brouwer and Kleinknecht start out by comparing qualitative results from the 1992 CIS questionnaire to the differences in propensity to patent between sectors found by Arundel and Kabla (1998). Although not entirely comparable due to its qualitative nature, results show a consistent picture and thereby provide support for the finding that differences in the propensity to patent an innovation exist between sectors.

Thereafter a multivariate analysis is performed, where four factors are included: sector, firm size, cooperation and R&D intensity / share of new products in total sales. It is worth noticing that Brouwer and Kleinknecht take a more narrow approach to the propensity to patent, by excluding innovations that are not new to the market (e.g. only new to the firm.) Therefore they look at the *propensity to patent new to the market product innovations*. Both a model for the probability of at least one patent application and a model for the number of patent applications are estimated.

Firm size (as a control variable) and high technological opportunity sectors were found to be significant and positive. R&D collaboration and R&D intensity had less robust t-values (with minimum values for one of the models at 1.3 and 0.4 respectively.) Moreover, it was found that smaller firms have a lower propensity to apply for a first patent, but given that they patent their patent propensity is generally higher than for other firms. (Brouwer and Kleinknecht, 1999)

Brouwer and Kleinknecht, 1999	*Regression on propensity to patent new to market product innovations*		
Factors	**Significant / sign**	**Present in CIS**	**Level**
R&D collaboration	◆ positive	✓	Firm
High technological opportunity sectors	✓ positive	✓	Firm
R&D intensity	◆ positive	✓	Firm
Other issues raised			
Differences across industries			
Undercounting of small firm R&D			
Smaller firms have a lower propensity to apply for a (first) patent			

Table 4 – Factors and issues identified by Brouwer and Kleinknecht (1999)

Country level factors

Building on the idea that patents not only reflect the propensity to patent, but also research productivity, Rassenfosse and Pottelberghe de la Potterie (2009) look at the differences between countries in these two variables. The propensity to patent at a country level is shown to depend on several factors, including the number of patentable subject matters, restrictions, enforcement mechanisms and – especially – its fees. The first three of these factors were combined into Ginarte and Park's IP index of patent rights (IPI). (de Rassenfosse and van Pottelsberghe de la Potterie, 2009)

Rassenfosse and Pottelberghe de la Potterie, 2009	*Regression analysis on propensity to patent on a country level*		
Factors	**Significant / sign**	**Present in CIS**	**Level**
Patent fees	✓ positive	Country variable	Country
Index of patent rights (IPI)	✓ positive	Country variable	Country

Table 5 – Factors and issues identified Rassenfosse and Pottelberghe de la Potterie (2009)

OTHER ISSUES

The importance of patenting versus other means of appropriation

A factor that may have a very direct impact on the propensity to patent is the substitutes that are available for it. One factor that is thought of as especially relevant is the importance of secrecy. Others include lead times, learning curves, brand and complementary investments. (Levin et al., 1987) The decision whether or not to patent then depends on the relative value of a patent towards other means of appropriation.

Worth noting here is that patents have drawback for a firm, too. The period of the granted monopoly (patent life) is always limited. In this way patents make sure that, after the patent expires, the information covered by the patent will be in the public domain. Moreover, even before the patent expires, it also provides information to competitors about what the patenting firm is up to and may in that way stimulate competitors to do the same, but in a way that is not covered by the patent. Information disclosure is thus a side-effect of patents that may have negative value to a firm when compared to secrecy. This especially goes for process innovations.

On top of that, patents also bring costs in terms of application and enforcement. This makes the patent choice a cost / benefit trade-off. This is also seen in qualitative results by Brouwer and Kleinknecht (1999), where it is found that patent protection is only the fourth most important means of appropriation, in terms of firms' judgment about their effectiveness. Time lead on competitors, keeping qualified people in the firm and secrecy are all seen as more effective means of appropriation.

The importance of the propensity to patent for high tech industries

Hagedoorn and Cloodt (2003) ask the question how important the propensity to patent – or other moderator variables for that matter – actually is. For four high tech industries (aerospace and defense, computers and office machinery, pharmaceuticals and electronics and communications) they compare the main concepts of R&D, new products and patents as indicators for innovative performance. For high tech industries, they find that the statistical overlap between these three indicators is very strong. The main implication here is that – for the four high tech industries covered – the average propensity to patent innovation is roughly constant between these industries.

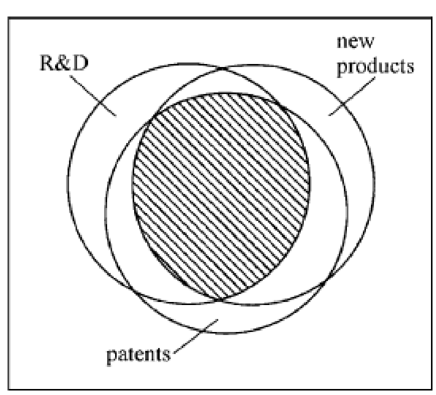

Figure 4 – Venn diagram representing the relationship between R&D, patents and new products. (Hagedoorn and Cloodt, 2003)

HYPOTHESES FROM LITERATURE

Based on the previous review of literature regarding the propensity to patent, several hypotheses can be tested through analysis of the CIS databases. In the overview below only those factors have been included whose concepts are somehow included in the CIS questionnaires.

Factor	Hypothesis
Overseas sales	Firms that have international sales have a higher propensity to patent
Sales in markets outside of Europe	
Federal funding	Firms that receive government funding have a higher or lower propensity to patent
Firm size (sales, logarithm)	Firms with higher sales/turnover numbers have a higher or lower propensity to patent
Firm size (employees, logarithm)	Firms with more employees have a higher or lower propensity to patent
R&D intensity	Firms with a higher R&D intensity have a higher propensity to patent
R&D collaboration	Firms that have R&D collaboration on innovation projects have a higher propensity to patent
Located in Germany	German firms have a higher propensity to patent than Norwegian or Belgian firms
Industrial sector	There are differences between industrial sectors with respect to the propensity to patent

Table 6 – Hypothesis based on previous research

For some of these factors, high correlations amongst them can be expected. For overseas sales and sales in markets outside of Europe a combined hypothesis has been formulated, since both directly aim at essentially the same thing which is international sales. For firm size, two different hypotheses have been stated, since there is a more relevant conceptual difference between the two. It cannot a priori be stated that for example labor-intensity does not play any role. However, it can be expected that for these two measures collinearity issues will arise. The variable that fits best in the model will then be kept. Another solution can be to use for example labor-intensity as a separate variable, by dividing number of employees by turnover.

CONCEPTUAL FRAMEWORK

To get a conceptual idea of where all these hypothesized factors fit in, a top-down approach has been adopted and depicted in figure 5. From a business perspective, patents should be only applied for when the (perceived) benefits outweigh its (perceived) costs. Going from this approach, several sub-dimensions have been identified based on an integration of ideas that were identified in this chapter. By no means is this framework thought to be exhaustive, nor is it the only way to think about the propensity to patent. However, it may provide a useful framework to structure thinking about factors that impact this concept.

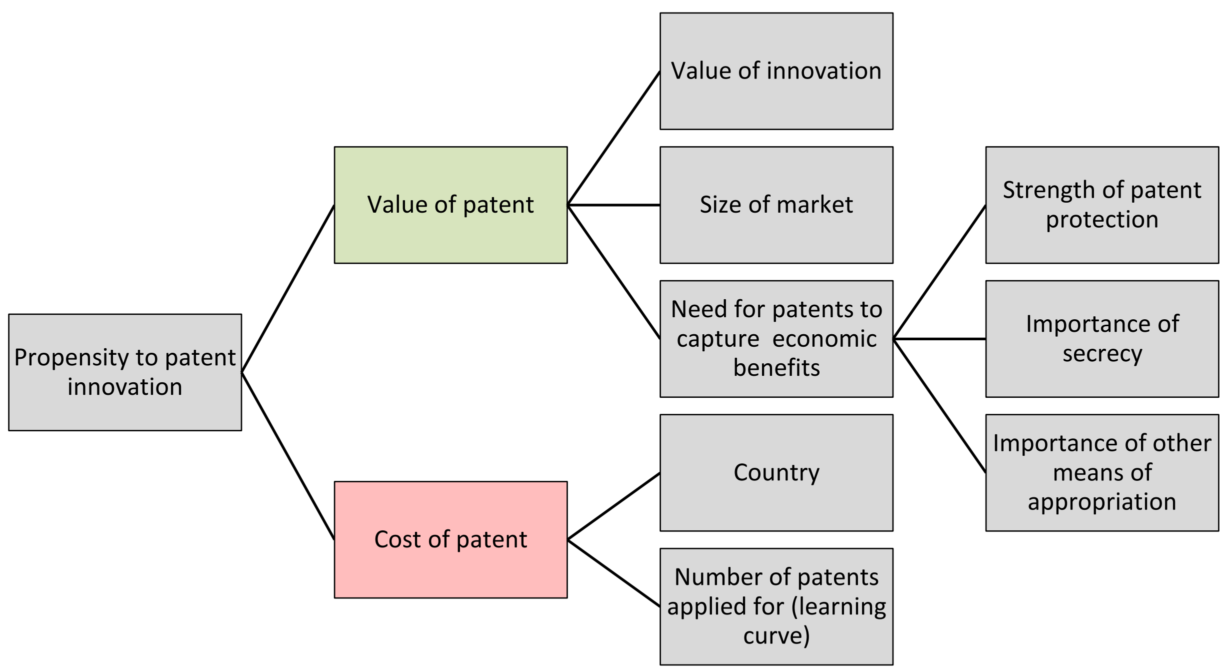

Figure 5 – Proposed conceptual framework

Both the value of the innovation and the number of innovations that can be sold (market size) have an impact on the potential economic benefit of a patent. Besides that, it is then important that there is a need for patents to capture that economic benefit. Factors that are important here include lead time, learning curves, importance of secrecy (as a substitute to patents) and firm size (larger firms may have other means of appropriation available, such as established reputations and strong sales and service.) (Levin et al., 1987) Costs of patents are mostly determined by the national patent fees (as well as costs of enforcement) and thus play at a country level. Furthermore, the cost of a patent may decline when more patents have been applied for. This is based on the finding of Brouwer and Kleinknecht (1999) that small firms tend to have a higher propensity to patent given that they already applied for their first patent.

METHOD

APPROACH

Probably a main reason why for example Scherer (1983) did not account for efficiency – he rather assumed that the factor would be constant – is the difficulties in separating it from the propensity to patent factor. When only R&D input data and patent data are available, this distinction is almost impossible to make. However, with the availability of a direct measure of innovative output it becomes possible to directly compare that measure with a patent application measure. In that way, the discrepancies between the direct innovation measure and the patent measure are the result of the propensity to patent.

Definitions

Based on the previous, this thesis will look at the propensity to patent as the factor that is the main cause of the discrepancy between patents and innovativeness. Defined in such a way it will be most suitable for answering questions of the applicability of patents as an innovation indicator. It is therefore that this thesis will explicitly separate effects of research productivity and propensity to patent. (See the conceptual model in figure 2.) This is in contrast to Scherer (1983), who does recognize the research productivity effect, but assumes it to be a more or less constant factor.

Defined in this way, the propensity to patent is the number of patents per unit of R&D input, compensated for any research productivity effects. This comes close to the definition of Mansfield, who defines propensity to patent as the percentage of patentable inventions that are actually patented. (Mansfield, 1986) Even more close is the extended version of this definition which uses innovation instead of invention. (Arundel and Kabla, 1998)

One major advantage of this definition is, as is also put forward by Arundel and Kabla (1998), that it is not influenced by any R&D productivity factors such as the previously stated efficiency. It thereby directly links the patent measure to a measure of innovative output.

Firm level definition

Another approach can be not to take the absolute number of patents but rather whether or not an innovative firm has applied for at least one patent. This approach may be preferred since the CIS questionnaire is a firm level based survey in which individual innovations and their associated patent applications are almost impossible to identify. In that case, the definition of propensity to patent is applied to the firm level, meaning that it will say something about the percentage of innovative *firms* which have applied for a patent. (Even though the firm can have more than one innovative product and/or more than one patent.) Thus, for this thesis, the definition used will be:

➔ The percentage of innovative firms which have applied for a patent.

This definition is directly measured in the CIS questionnaire with the binary variable asking whether or not the firm has applied for a patent during the years 2002 to 2004.

Data analysis

Since we are looking at the influence of several factors on the propensity to patent and since a large database (CIS) is available, a quantitative analysis of this data is possible. Independent variables will include those factors that have a hypothesized impact based on the review of earlier research on the subject. On top of those, independent variables will be included that are available in the CIS questionnaires and which may have a relationship to the propensity to patent as well. This part of the analysis will then be more explorative in nature. For all independent variables bivariate analyses will be performed to check their individual relationship to the propensity to patent. (van der Velde et al., 2004)

Since the dependent variable for patent applications in the CIS database is binary, analyses will mainly be done using a (binary response) non-linear probability model. For this, a logit model is the most suited and SPSS and/or Stata will be the main tool to use. (Heij et al., 2004, Rice, 2007) The product of this research will be an empirical model, with measures of significance attached to tested factors. Moreover, the regression coefficient gives an indication of how important these factors are in determining the propensity to patent. If many significant and important factors can be found, this will be an indication that patents are not (always) a good indicator for innovativeness.

CIS QUESTIONNAIRES

Every four years, the Community Innovation Survey is carried out, providing information on multiple aspects of innovation on a firm-level basis. This thesis will be based on CIS4 and – to a much lesser extent – on CIS3. The fourth version of CIS covers the period of 2002 to 2004, while CIS3 covers the period 1998 to 2000.

The surveys themselves consist of several main categories, of which the first few aim to classify firms into several types of innovators. These types can be product innovation, process innovation, ongoing innovation, abandoned innovation or a combination of these. CIS4 also distinguishes several subcategories within product and process innovation, such as goods and services.

Other sections of CIS4 cover innovation activities and expenditures, sources of information and co-operation for innovation activities, effects of innovation, factors hampering innovation activities, intellectual property rights and organizational and marketing innovations. Also, basic information on the enterprise is available, including turnover rates, employees, geographic location of markets and the firms' main activity. (OECD, 2004a)

CIS uses product and process innovation in the following way:

- Product innovation: market introduction of a **new** good or service or a **significantly improved** good or service with respect to its capabilities, user friendliness, components or sub-systems.
- Process innovation: implementation of a **new** or **significantly improved** production process, distribution method or support activity for the firm's goods or services.

Product innovation includes two subcategories: new or significantly improved goods and new or significantly improved services. Moreover, for product innovation, CIS4 distinguishes between "new to market" and "new to firm" innovations. Also, information is available about whether the product innovation was developed 1) mainly by the enterprise itself or the enterprise group, 2) by the enterprise together with other enterprises or institutions, or 3) mainly by other enterprises and institutions.

CIS research population

CIS4 classifies firms according to the NACE classification. Its target population covers mainly NACE activities C to K. More specifically, the core target population of CIS4 consists of the following industries: (OECD, 2004b)

- Mining and quarrying (NACE 10-14)
- Manufacturing (NACE 15-37)
- Electricity, gas and water supply (NACE 40-41)
- Wholesale trade (NACE 51)
- Transport, storage and communication (NACE 60-64)
- Financial intermediation (NACE 65-67)
- Computer and related activities (NACE 72)
- Architectural and engineering activities (NACE 74.2)
- Technical testing and analysis (NACE 74.3)

Based on this classification, 25 individual sectors have been defined. An overview of these can be found in appendix I.

For this thesis the focus will be on three countries: Germany, Belgium and Norway. Since the CIS surveys are the responsibility of each country individually, three separate databases had to be combined in order to generate the database that forms the basis of further analysis. In table 7 the share of observations are given for each country. In general, compared to population, Germany is underweighted in the amount of observations. This is not necessarily a problem and will be resolved by including dummy variables per country in the regression analysis. However, especially in the descriptive analyses, it has to be taken into account that Germany is somewhat underrepresented.

Country	**CIS4**	
	Frequency	**Percentage**
Belgium	3322	27.63
Germany	4054	33.71
Norway	4649	38.66

Table 7 – Share of observations per country for CIS4

VARIABLES AND VALIDITY

To be able to test factors that are identified within the literature, there must be a measure for that factor in the CIS questionnaire. Since the CIS questionnaire is anonymous at the firm level, it will be impossible to link it to other databases. Therefore, to build the model, there is a requirement that the hypotheses to be tested can be measured by variables which exist in the CIS database.

Dependent variable (patent application)

The dependent variable measures whether or not the firm applied for a patent. For CIS4 this is measured as the – binary – statement whether or not the enterprise applied for at least one patent during the years prior to the questionnaire, 2002 to 2004 (*propat*).

Two – probably relatively minor – validity issues are apparent from this definition. First, this definition implicitly assumes that if there was an innovation in the period from 2002 to 2004, that the patent also would be applied for in that same period. This is not necessarily the case. Patents may especially be applied for in the period before.

A second validity issue with the way the dependent variable is measured in CIS is the fact that it looks at patent *applications.* Because of this there is a need to refine the definition of propensity to patent to the propensity to *apply* for a patent (linked to some sort of innovation.) That this issue is not just a theoretical issue, but may actually be of practical importance is shown by the fact that only about half of patent applications are actually granted. This goes for the EU as well as other regions. (Patentlens, 2003)

Unfortunately, CIS does not measure the fact whether a patent was actually granted. This has to be kept in mind when interpreting the results of the regression analysis. An advantage of this approach is that it limits the problems with time delay between patents and innovations that is described above.

Patent office	Patent applications	Granted patents	Percentage granted
EPO	116613	59992	51%
JPO	413092	122511	30%
USPTO	342441	169028	49%

Table 8 – Patent applications versus granted patents for Europe, Japan and the US according to Patentlens (2003)

Hypothesized independent variables

The hypotheses contain several concepts that need to be operationalized in order to test for their impact on the propensity to patent. These operationalized concepts will form part of the independent variables of the logit model. Below, an overview will be given of all relevant variables present in CIS4 that can be used to operationalize these hypothesized concepts.

Hypothesized concept	**Variable(s) in CIS4 and their definition**
International sales	*marloc, marnat, mareur, maroth:* binary variables measuring whether or not a firm was present in the respective geographic markets (local/regional, national, EU/EFTA/EU candidate countries, all other countries)
Government funding	*funloc, fungmt, funeu, funrtd:* binary variables measuring whether or not a firm received any public financial support for innovation activities (local/regional authorities, central government, EU and the EU 5^{th} framework programme respectively)
Sales	*turn02, turn04:* enterprise's total turnover for 2002 and 2004 respectively
Employees	*emp02, emp04:* enterprise's total number of employees in 2002 and 2004 respectively
R&D intensity	*turn04, rrdinx:* enterprise's total intramural R&D in 2004 divided by the enterprise's total turnover for 2004
R&D collaboration	*co, co[x][y]:* binary variable for co-operation on innovation activities with other enterprises and institutions, separated along type of co-operation partner [x] (other enterprises within your enterprise group, suppliers, clients/customers, competitiors, consultants/commercial labs/private R&D institutes, higher education institutes, government/public research institutes) and location [y] (national, other Europe, United States and all other countries.) Co is the broad binary variable for any type of co-operation on innovation activities
German firm	*country:* nominal variable, DE indicating Germany, NO indicating Norway and BE indicating Belgium
Industrial sectors	*Nace_pro:* nominal variable giving the NACE classification code for the main activity the enterprise is involved in

Table 9 – Operationalization of hypothesized concepts using CIS4 variables

Explorative independent variables

Besides the hypothesized factors, CIS4 provides a question about the importance of different information sources for innovation activities. These sources may be especially relevant since they say something about the possibility for the enterprise to patent their innovation. A plausible hypothesis may for example be that when suppliers are an important information source for innovation, this may actually indicate that the innovation was bought from those suppliers. It is then likely that a patent would be applied for by the supplier and not by the responding enterprise. All these variables are measured on a 4 point ordinal scale, where 0 stands for not used, 1 for low degree of importance, 2 for medium degree of importance and 3 for high degree of importance.

Explorative concept	Variable(s) in CIS4 and their definition
Internal information sources	*sentg:* importance of an internal (within enterprise group) information source for innovation activities
Suppliers as information source	*ssup:* importance of suppliers (equipment, components or software) as information source for innovation activities
Clients as information source	*scli:* importance of clients as information source for innovation activities
Competitors as information source	*scom:* importance of competitors as information source for innovation activities
Consultants & commercial labs as information source	*sins*: importance of consultants & commercial (R&D) labs as information source for innovation activities
Higher education institutes as information source	*suni*: importance of higher education institutes as information source for innovation activities
Government as information source	*sgmt*: importance of government and public institutes as information source for innovation activities
Conferences as information source	*scon*: importance of conferences, trade fair & exhibitions as information source for innovation activities
Publications as information source	*sjou*: importance of scientific journals & trade/technical publications as information source for innovation activities
Professional associations as information source	*spro*: importance of professional and industry associations as information source for innovation activities

Table 10 – Definition of explorative independent variables of CIS4

CIS innovation output measure

By using the approach of analyzing only those firms that actually have innovative output, it is important to consider if it captures exactly the output variable that is of interest. Of course, this depends on what one wants to measure. (Griliches, 1990) However, two issues are important to identify and keep in mind when using the – binary – CIS classification of innovation output.

Value of innovation (and the importance of the new to market variable)

The CIS questionnaire provides data about innovative output on firm level with questions about new product and process introductions. Since these are binary variables, they do not directly measure the value of the innovative output.

It is possible, though, to look at the value of an innovation using CIS. Two approaches can be taken. The first is to look at changes in firm turnover. However, as will be discussed in a few pages, there are serious reliability issues with the CIS turnover variable. Therefore a second approach may be to use the distinction between "new to the market" and "new to the firm only" innovations. When assumed that new to market innovations generally have a higher economic value, this binary variable (*newmkt*) might be included in the regression analysis to account for some part of the value dimension. In turn this dimension is expected to be relevant since it directly impacts the cost / benefit decision of an enterprise when choosing whether or not to patent.

Over counting of propensity to patent for big firms

Another issue to consider is that the CIS questionnaire does not ask for the number of new or improved products and services. For big firms – which might have more than one innovation – this distorts the picture of the actual amount of inventions that are patented. In that respect it may over count the propensity to patent of big firms. This is the main reason that firm size was found to be a relevant control variable, especially in the research of Brouwer and Kleinknecht (1999).

PROCEDURE

First a descriptive analysis will be performed using CIS4. The main aim of this analysis will be to look at different types of innovation and their respective propensities to patent. Also some descriptive insights may be used to structure the subsequent econometric analysis. The main goal of that econometric analysis will be to build a model that provides insight in the relevant factors impacting the propensities to patent. Several different versions of the model will be put next to each other in order to check for the stability of factors and their impact on the propensity to patent. At least one version will include industry dummies.

Beforehand, though, a bivariate analysis will be performed on all individual variables. In that way, more robust conclusions can be drawn as well as insights in possible interrelationships between independent variables (when combined with regression results.) Also beforehand a more descriptive industry analysis will be performed, where results will be compared to earlier work by Arundel and Kabla (1998) and Brouwer and Kleinknecht (1999).

Logistic regression

For ordinary OLS regression analysis, one major drawback is that the regression equation will not have boundaries between one and zero. In this case, where the dependent variable is binary, it will then produce results which do not have a real world meaning – such as values higher than 1. With logistic regression this problem is dealt with by predicting logits, which are natural logarithms of odds. Odds are in turn based on probabilities, which run from zero to one.

With logits, which can run from minus infinity to infinity, it is then possible to predict probabilities ranging from zero to one. Note here that for binary variables, strictly speaking, only values of zero and one would have meaning. However, with logistic regression, values in between one and zero can be interpreted as the probability that the final result will be either one or zero (de Vries and Huisman, 2007).

There are several assumptions underlying the logistic regression equation. By using this equation it is assumed that the link function (in case the logit function) is the correct function to use. Besides that, it is assumed that all relevant variables are included, that no variables are included that should not be in the model and that the logit function is a linear combination of the predictors (so a linear combination of the independent variables has to be sufficient.) (ATS, 2011)

To make the model as relevant and clear as possible, some additional checks will be performed, including checks for outliers as well as tests for multicollinearity. Also, to check the robustness of the model several nonlinear combinations of independent variables may be identified and included. Several versions of the model will be estimated and compared.

For logistic regression it is not possible to define an objective measure for explained variance. A pseudo r-squared measure will thus be used to assess the explanatory power of the model. However, because this pseudo r-squared measure does not have a direct meaning (as it does have in OLS regression) an additional analysis will be performed based on the ability of the model to correctly predict different outcomes.

RELIABILITY ISSUES

Micro-aggregation

For CIS4 an anonymisation method was used to make sure that an individual enterprise can no longer be identified as such. This method is based on a micro-aggregation process (MAP) and results in somewhat distorted data for variables such as turnover and R&D investment. Companies with similar values for those variables are pooled together with a common (average) value for them. In each pooled category, at least three companies are present. In general, the reliability of data is lowered by this micro-aggregation process.

For most statistical analyses in this thesis, this micro-aggregation is not so problematic. However, trouble does arise when defining a new variable for R&D intensity. This variable is defined as the ratio between R&D investment and turnover. Since both these variables are manipulated, the error in measurement is expected to increase rapidly. For R&D intensity, numbers around 3% would be common. However, in some preliminary analysis it can be seen that 656 firms have values of more than 10% for their R&D intensity, while 120 firms have an R&D intensity higher than 50% and 42 have a number higher than one (which makes no sense conceptually.) Moreover, since the amount of firms that has (extremely) high values for R&D intensity is very substantial this is also an indicator that for firms with R&D intensity below 10% that their values may not accurately represent their real situation.

R&D Intensity	**Observations**
10% or more	656
50% or more	120
100% or more	42

Table 11 – Number of firms that have R&D intensities of respectively more than 10%, 50% and 100%

DESCRIPTIVE STATISTICS

CIS4

Overview

A first step to approach the CIS databases is to narrow down the research population. In order to make statements about the propensity to patent, the main interest goes to the group of firms which actually have innovative output. In that way, statements can be made about the propensity to patent innovations. The CIS questionnaire defines 4 groups of firms which are marked to have innovative output. This can be either new or significantly improved products, new or significantly improved processes, ongoing innovation activities, or abandoned innovation activities. Combinations are also possible. The timeframe for these activities is 3 years (2002-2004). These descriptive statistics are based on the combination of the German, Norwegian and Belgian CIS database and include all sectors that were originally present (i.e. manufacturing and services.)

New to CIS4 is the subdivision of product innovation into innovation of goods and innovation of services. Also for process innovation, several subcategories have been made. The graph in figure 6 shows that out of all firms with some innovative activity, about one quarter of firms (25%) do actually apply for a patent.

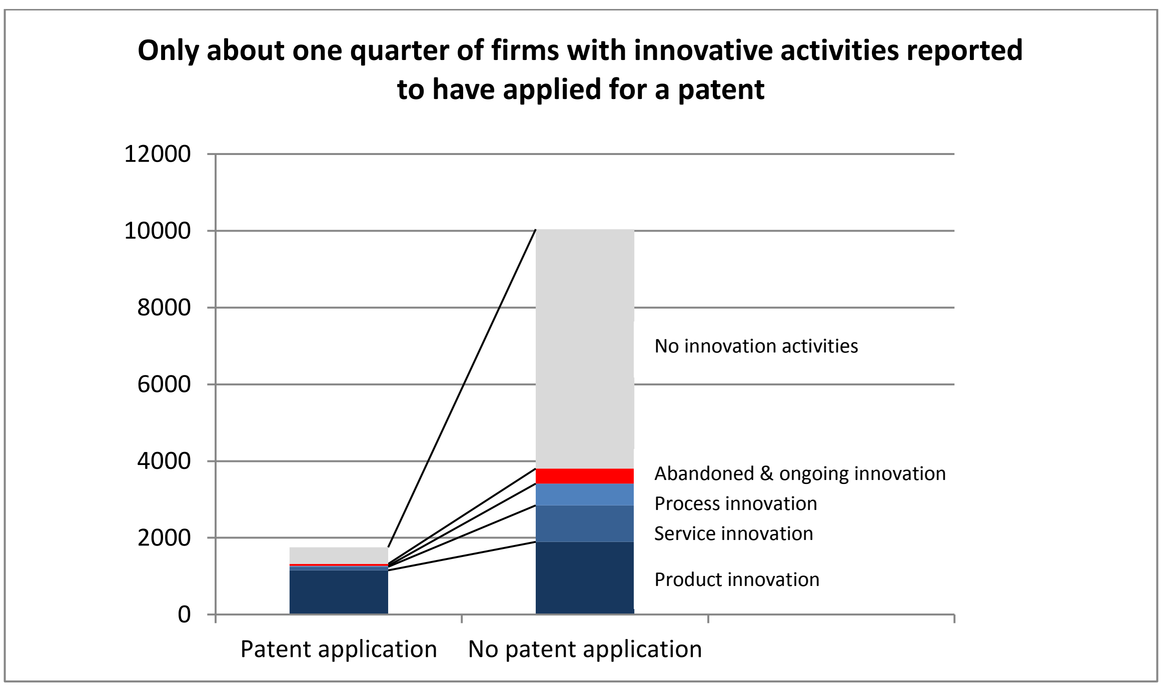

Figure 6 – patent propensity rates for product and process innovators

In the figure above it can be seen that firms with no innovative activities rarely apply for a patent – which is to be expected by definition. Also it can be seen that the overall propensity to apply for a patent for a firm with innovative activities is about 25%.

A problem that arises with a classification such as the one above is that the CIS questionnaire allows firms to be present in multiple groups. It is, for example, possible to have both product and process innovation. This slightly distorts the picture in figure 6, since product innovation can also include some process innovation. Since product innovation came first in the questionnaire, it is here taken as the main group.

Moreover, this is a first indication that care has to be taken into account when talking, for example, about the propensity to patent a process innovation. It may well be that patent applications of firms which do have process innovation actually are based on a product innovation in the same firm. This will be elaborated upon further on.

What is interesting about the graph above is the fact that the "no innovative activities" group is still quite large for firms that *did* apply for a patent. Without innovative activities it is hard to actually have patentable ideas. After some checking of the CIS database, it was found that Germany had used a slightly different approach to the filtering questions. Before continuing, a short overview will be given of this new categorization.

New categorization of CIS4

In CIS4 a slightly different classification system is used compared to older versions. Both the product innovation and process innovation categories have been split into several constituent categories. There are some (validity) issues with the level of exhaustion these groups have toward the bigger groups of product and process innovations. Several countries have therefore opted to additionally include broader categorization questions for product and process innovation.

It turned out that the German questionnaire has a slightly different structure compared to the Norwegian and Belgian one, and that it also asks for the broad product and process groups in general. This classification then follows the structure of CIS3. The reasoning behind doing so is that it was being felt that the more narrow categories of CIS4 did not cover the whole spectrum of the broader category (so that it wasn't exhaustive.) The interesting thing is that when only the answers to the new CIS4 categories are considered, Germany has a far bigger group of firms that did patent but did not fall into one of these innovation categories. See table 12.

Country	Observations
Belgium	17
Germany	365
Norway	50

Table 12 – Number of patenting firms that fall outside the four CIS4 categorizing groups

From this it can be inferred that there is actually an impact of the way the German CIS is structured compared to how the narrower categorization questions (goods and services) are answered. If it would be the same, one would expect to find equal shares of firms that patent but do not have any of these questions ticked with "yes". Yet table 12 shows otherwise. This may have to do with the rather forced choice that respondents have to make while answering the Norwegian and Belgian questionnaire, while the narrower categories are more presented as an "option" in the German questionnaire.

To make a fair comparison between countries, the firms that answered "yes" to the broader German categorization questions (inpd and inps, which were only present in the German CIS) are excluded from the no innovation activities group and divided amongst process and goods innovations respectively. This then results in a more equal distribution of firms that patent but nonetheless fall outside the categorizing innovation groups. See table 13.

Country	Observations	Percentage of total
Belgium	17	7,1%
Germany	82	7,7%
Norway	50	11,2%

Table 13 – Number of patenting firms that fall outside the four CIS categorizing groups, as well as outside the broader categorizing groups for the German questionnaire

Process and service innovation

When zooming in on the group of firms that applied for a patent (which is the left bar in figure 6), it can be seen that most patent applications are based on *at least* a new or significantly improved good.

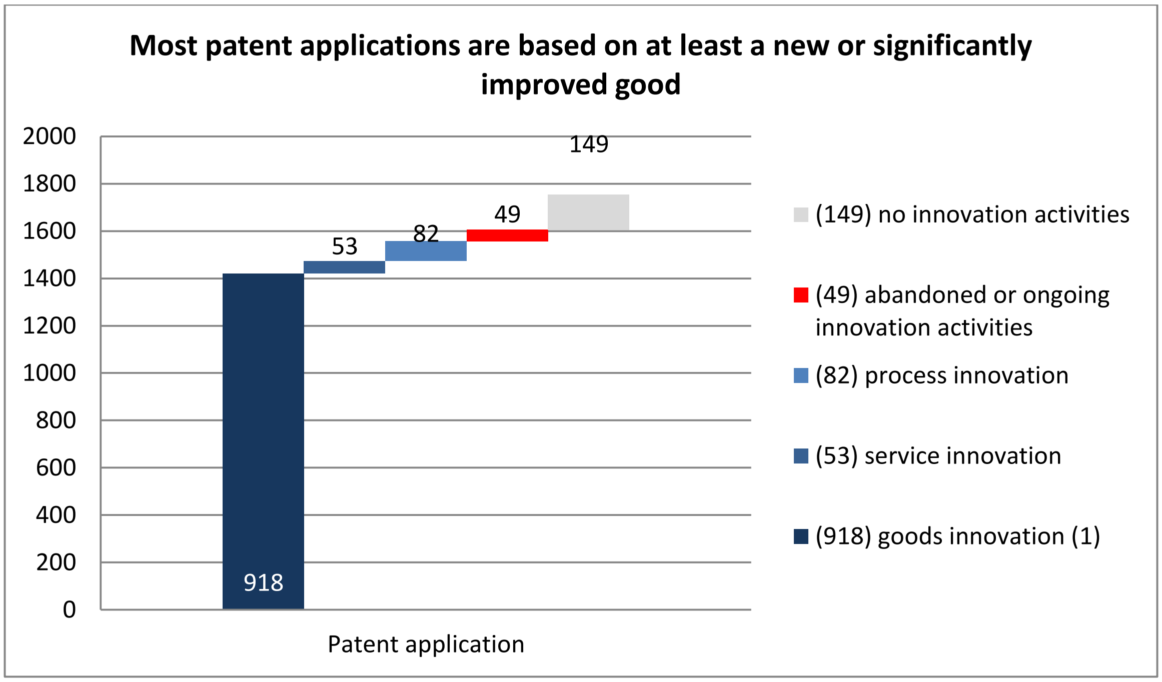

Figure 7 – Number of firms that have a specific type of innovative output, for patent applicants only

Note: (1) The German CIS4 questionnaire has been extended with categories for product and process innovations in general. These are respectively added to the goods and process categories.

From this figure it can be seen that process innovations *alone* are rarely patented. Moreover, the 82 observations in figure 7 include a possible overlap with ongoing and abandoned innovation activities. To check the relevance of including process innovation in further analyses, a new variable is defined which has a value of one only for those firms that had process innovation and not product innovation, ongoing innovation activities and/or abandoned innovation activities.

When this variable is tabulated against the question whether or not the firm has applied for a patent, it can be seen that out of the 82 patent applications in figure 7, only 36 can be attributed to process innovation *only.* On top of that, there are 524 firms which had process innovation only but did not apply for a patent. This equals a propensity to patent process innovations of 6.4%. The same argument goes for service innovation only, which has a resulting propensity to patent of 5.6%

Type of innovation	Total firms	Patent applicants	Percentage applicants
Process innovation only	524	36	6.4%
Service innovation only	180	10	5.6%

Table 14 – Process and service innovation only versus patent applications in 2002-2004

The finding that the propensity to innovate for process only innovators might be explained by the realization that process innovations are harder to patent because they are less tangible, as well as that they have less need to be patented (secrecy may often be preferred.) (Levin et al., 1987)

Although it could have been expected that some types of innovation have a somewhat lower propensity to patent than others, it is quite remarkable that the propensity to patent process innovation only is that low. This is graphically depicted in the Venn diagram below. These findings contradict earlier work by Arundel and Kabla (1998), whom find that the propensity to patent process innovations (within sectors) generally lies in between 20 and 30%. It is shown here that for process innovation only, patent propensity rates are substantially lower.

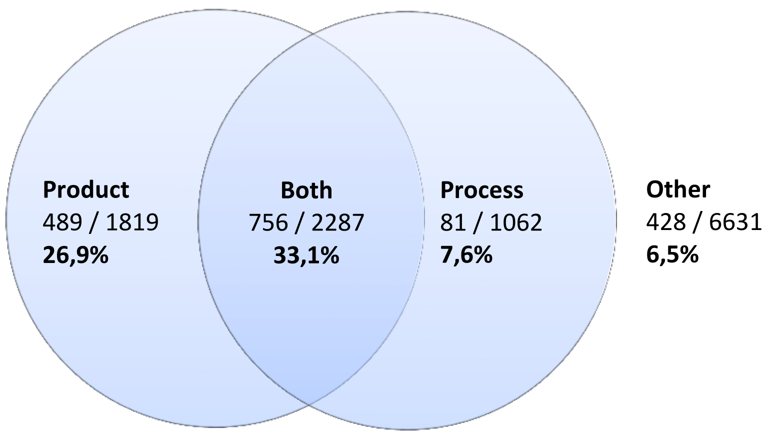

Figure 8 – Patent propensity rates for product and process innovators

Even more interesting is an extended analysis where product innovations are split up into goods and services, as is shown in the figure below. First, it shows that goods have an almost 4 times higher propensity to patent than services. However, maybe even more surprisingly, it also shows that firms which have good *and* service innovations have a lower propensity to patent (24.2%) than firms that have innovation of goods *only* (38.6%). Service innovation thus seems to be correlated with a lower propensity to patent a goods innovation. An explanation might be that new products in service oriented industries are more difficult to patent, or that other means of appropriation are more important. See figure 9.

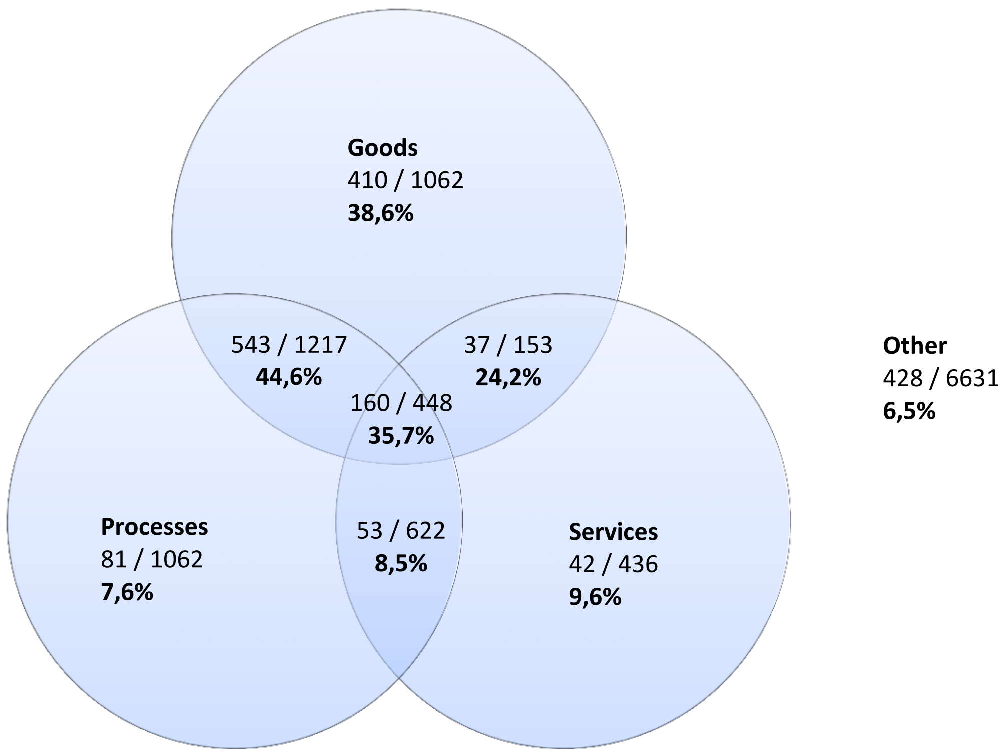

Figure 9 – Patent propensity rates for goods, services and process innovators

New to market innovations

Besides defining 4 groups of innovative activities, the CIS questionnaire splits the group of product innovations in products that are new (or significantly improved) only to the firm, or innovations that are new to the market. One can make an argument that innovations which are "new to the firm" are mainly innovations that were already introduced earlier by others, and that the more patentable innovations are represented by the group of "new to the market" innovators. Brouwer and Kleinknecht (1999) use this line of reasoning when choosing to only include "new to market" innovators in their analysis. Based on figure 10 on the next page, only about two fifths of the firms with product innovations can be seen as innovative to the market, and about three fifths would be introducing a product innovation that is only new to the firm.

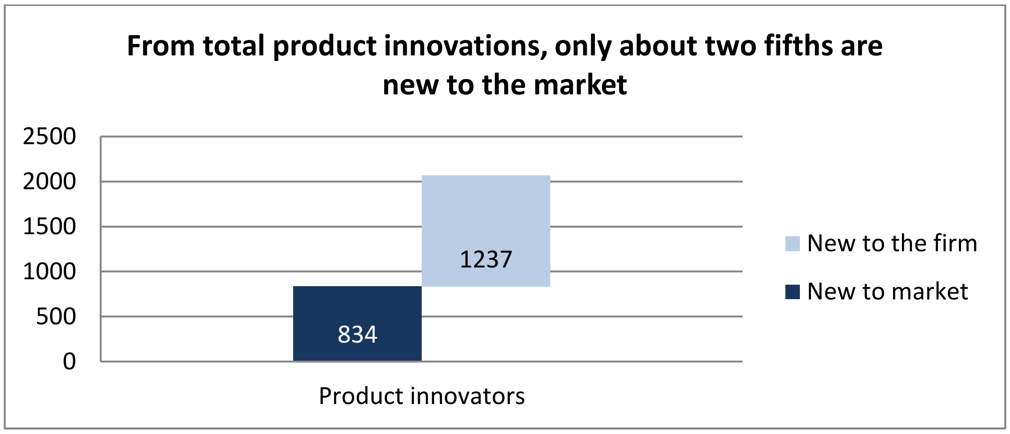

Figure 10 – Product innovators, split by "new to market" and "not new to market" (new to the firm)

Continuing from this line of reasoning it would be expected that the majority of patent applications will fall in the "new to market" category, with only a very small number of patent applications in the "new to firm" category. In other words: the propensity to patent a not new to market product innovation is expected to be low.

However, when we look at firms with patent applications for product innovations it can be seen that still one third of this group consists of firms which have product innovations that are not new to the market. (See figure 11.) This is remarkable, since patent applications require an original inventive step and, moreover, no prior demonstration.

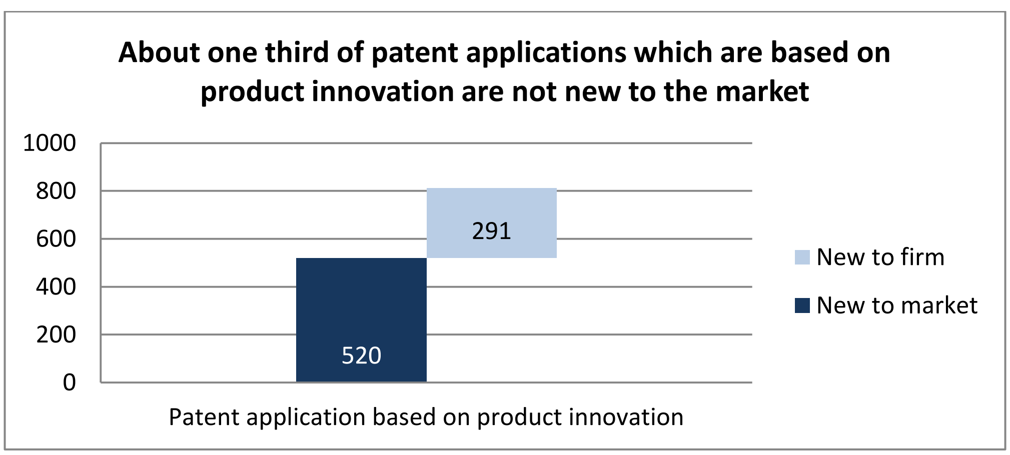

Figure 11 – Patent applications by product innovators, split to "new to the market" and "new to the firm"

Because of these patent requirements, it is hard to conclude that the group of "not new to the market" product innovators is constituted by pure imitators only. In this respect, one might think of devices that have common functionalities, but perhaps with some new or improved components, such as cars and mobile phones. In any case it is not possible to neglect the group of "not new to market" product innovations as not being important. For the rest of this thesis it will therefore be included in the analyses.

However, interesting to notice out of figures 10 and 11 is that although originally only two fifths of all product innovations are "new to the market", about two thirds of the patent applications relate to product innovations that are "new to the market." This means that – other things being equal – product innovations that are "new to the market" have a higher propensity to be patented than product innovations that are "not new to the market."

INDUSTRY AND COUNTRY LEVEL COMPARISON

Differences amongst industries

In the table below a comparison is made between propensities to patent product innovations, by sector of industry and by different sources.

Propensity to patent across industrial sectors		*According to Arundel and Kabla*	*According to Brouwer and Kleinknecht*	All product innovations CIS 3	All product innovations CIS 4
NACE					
10-14	Mining	**27,7%**		**43,5%**	**37,5%**
15-16	Food, Beverages and tabacco	**26,1%**	**24,6%**	**12,1%**	**13,7%**
17-19	Textiles and clothing	**8,1%**	**22,1%**	**11,1%**	**18,9%**
20-22	Wood, paper and printing		**25,1%**	**16,7%**	**19,2%**
23	Petroleum Refining	**22,6%**			**55,0%**
24	Chemicals and pharmaceuticals		**36,3%**	**48,7%**	**55,0%**
25	Rubber and plastic products	**33,7%**	**36,4%**	**32,5%**	**40,6%**
26	Glass, clay and ceramics	**29,3%**	**11,8%**	**32,6%**	**28,3%**
27	Basic metals (iron and steel)	**14,6%**	**9,9%**	**34,7%**	**37,1%**
28	Fabricated metal products	**38,8%**	**23,6%**	**27,1%**	**41,9%**
29	Mechanical engineering	**52,4%**	**26,4%**	**50,5%**	**54,6%**
30-31	Electrical equipment (including computers)	**43,6%**	**27,7%**	**44,5%**	**50,4%**
32	Communication equipment	**46,6%**	**27,7%**	**45,6%**	**50,4%**
33	Precision instruments	**56,4%**	**35,8%**	**49,5%**	**50,4%**
34	Automobiles	**30,0%**	**29,6%**	**39,2%**	**50,0%**
35	Other transport equipment	**31,2%**		**34,8%**	**50,0%**
36	Manufacture of furniture			**30,1%**	**33,1%**
37	Recycling			**15,8%**	**33,1%**
40-41	Electricity, gas, steam and (hot) water supply			**10,2%**	**7,7%**
51	Wholesale trade and commission trade (except motor vehicles)			**22,1%**	**21,0%**
60-62	Land, water and air transport			**4,7%**	**9,1%**
63	Auxiliary transport activities; travel agencies			**5,3%**	**2,5%**
64	Post and telecommunication			**22,4%**	**7,1%**
65-67	Financial intermediation			**0,5%**	**1,6%**
72-74	Business activities (e.g. consultancy; r&d)			**22,9%**	**25,6%**

Table 15 – Average patent propensity rates per industry, for product innovations

Notes: (1) Arundel and Kabla (1998) use a sales weighted propensity to patent product innovations
(2) Brouwer and Kleinknecht (1999) use a qualitative measure, here the percentage is given of Dutch firms that appreciate the value of patent protection for product and process innovation as very important or crucial.
(3) The colored cells for CIS4 are combined groups and the average propensity to patent is taken for these industries together, since no distinction is made within CIS4 for these industries.

Although the approach – as well as the results – differs substantially per author, a fairly clear pattern emerges of industries where patent propensities are higher than in others. Especially industries with NACE codes[2] ranging from 24 to 36 show high propensities to patent. These industries are characterized by manufacturing industries that focus on relative high tech products when compared to, for example, manufacturing of food and textiles or the more service oriented industries. This is also graphically depicted in the figures below.

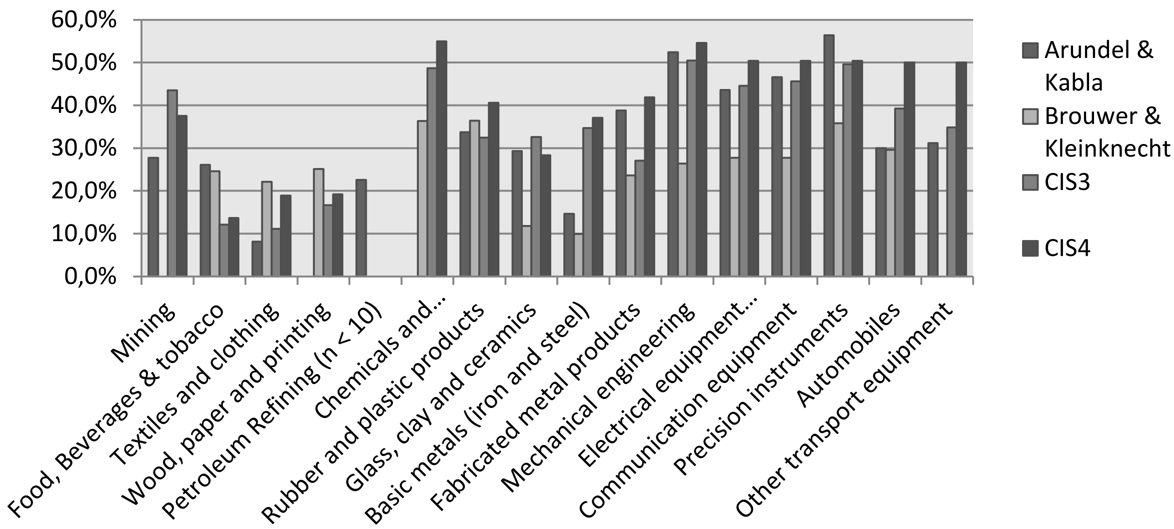

Figure 12 – Average patent propensity rates per sector

Note: (1) The qualitative measure of Brouwer & Kleinknecht does not measure absolute patent propensity.

To make the absolute percentages in the figure above better comparable, the graph below shows the same data but indexed to its overall average. This is especially relevant for the qualitative measure of Brouwer and Kleinknecht (1999) since that percentage does not correspond to an absolute propensity to patent.

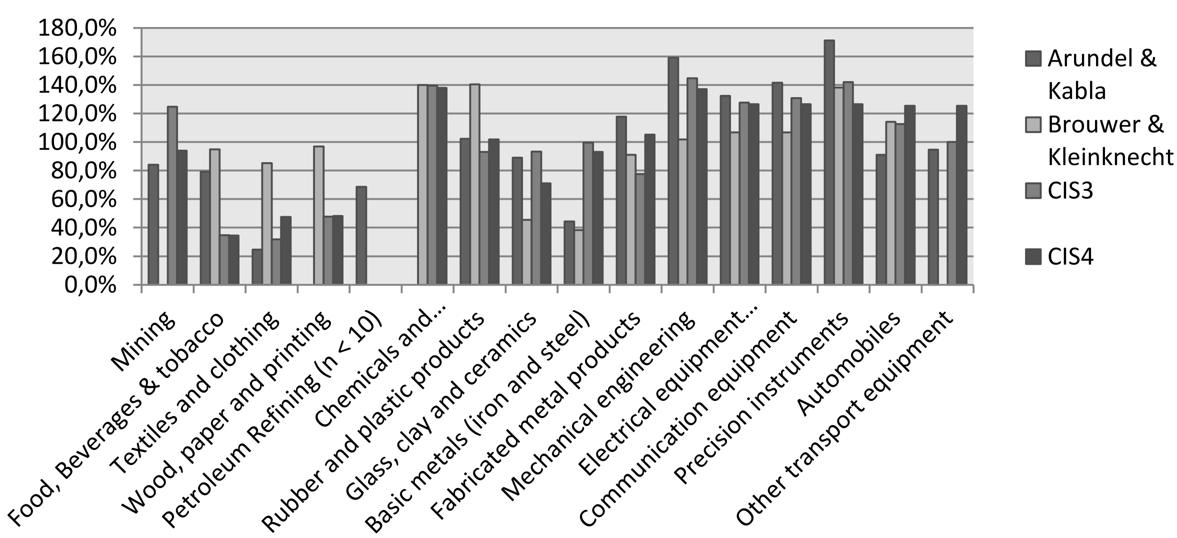

Figure 13 – Indexed average patent propensity rates per sector, where total average is 100%

[2] For a legend of sectors, NACE codes and their corresponding numbering in the sector variable, see appendix I.

Differences amongst countries

There is evidence to believe that there exist country-level factors that have an impact on the propensity to patent. (de Rassenfosse and van Pottelsberghe de la Potterie, 2009) From a simple analysis it can be seen that indeed Germany has – on average – a higher propensity to patent than Belgium or Norway do.

However, this does not automatically mean that country-level factors are at work here. It may very well be that Germany has relatively more firms in industries that have high propensities to patent, making industry-level factors the cause of the observed higher propensity to patent. Indeed, this may be plausible. Moreover, this argument also goes for firm-level factors. To see whether firm level factors are of importance, Germany will be taken into account in the logistic regression model as a dummy variable.

Propensity to patent across countries	All product innovations CIS 3	All product innovations CIS 4
Country		
Belgium	**23,4%**	**21,2%**
Norway	**23,9%**	**24,8%**
Germany	**32,9%**	**39,6%**

Table 16 – Average patent propensity rates per country

An interesting observation from table 16 is that the propensity to patent increased quite substantially for Germany between CIS3 and CIS4, while values for Belgium and Norway remained roughly constant. Several explanations might be given for this, including the fact that the German questionnaire for CIS4 has been slightly adjusted. This higher propensity to patent is also reflected when comparing the CIS3 and CIS4 propensity to patent rates across industries. See the figure 14.

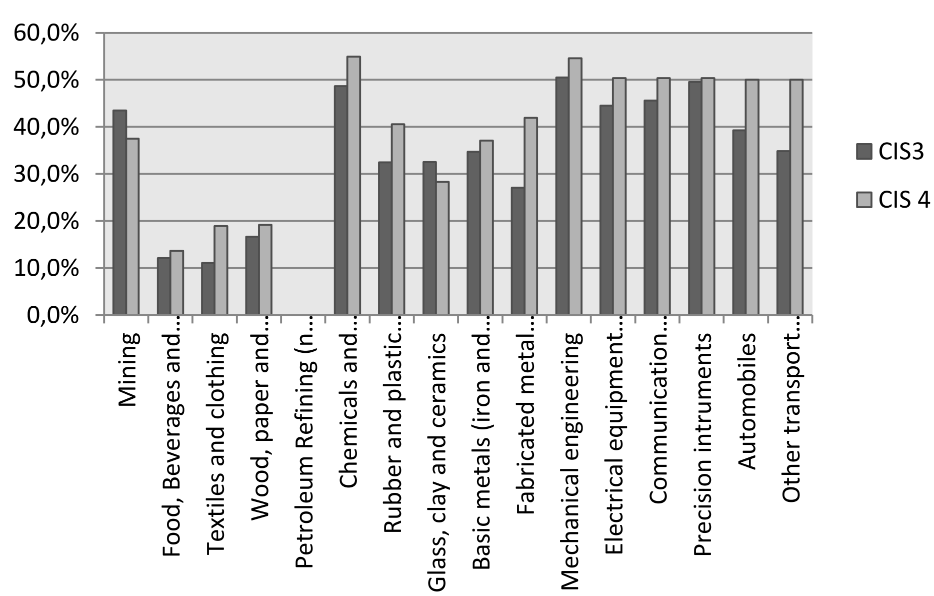

Figure 14 – Patent propensities for different sectors in CIS3 and CIS4

Up until this point, descriptive analyses suggest that there are multiple factors that may have an impact on the propensity to patent, including type of innovation, its newness to the market as well as industry and country level factors. To see whether these factors stand alone or are correlated to other factors is a question that can be answered by multiple regression analysis.

ECONOMETRIC ANALYSIS OF CIS4

REGRESSION APPROACH

The goal of this section is to determine the relevance of several factors that can impact the propensity to patent product innovations. In existing literature several factors have already been identified. These – theory based – factors will be included and again be checked for their relevance as well as the magnitude of their coefficients. One limitation here is that not all hypothesized factors are represented by variables in the CIS4 questionnaire.

Besides literature based factors, other factors will be included that are measured by variables in the CIS4 questionnaire. This will make the analysis more explorative in nature. For all factors a bivariate analysis will be performed to check for individual relationships with the propensity to patent. Thereafter, the factors will be jointly included in a logistic regression model.

Here it may be that factors that have no direct individual impact on the propensity to patent will be found to have an effect when corrections are being made for other factors. On the other hand, variables which have a significant individual relationship with the propensity to patent may turn out to be correlated with other factors and thereby do not have a significant impact in the logistic regression model anymore.

In the resulting model only those variables will be kept that are significant at the .05 level. This will be done by stepwise deletion of variables which have a z-value closest to zero (and are thus least significant.) Various versions of the model will then be estimated to check for robustness of both the z-values as well as the odd-ratio coefficients. Collinearity issues are mostly automatically detected by Stata, but if necessary checked by additional collinearity diagnostics.

Model research population

From descriptive analyses it showed that product innovation is the most important category in terms of propensity to patent. This analysis will therefore focus on the propensity to patent a product innovation. CIS4 includes a question that asks firms if they have product innovations. Based on this question, a research population can be defined that had a product innovation introduced in the period 2002-2004. When a logistic regression analysis is then performed on this group with patent application as the dependent variable, it can be checked what factors have an impact on the propensity to patent a <u>product innovation</u>. These factors are then represented by the (significant) independent variables in the model.

In the remainder of this section several variables will be discussed and tested for their individual relationship with the propensity to patent. In CIS4, patent propensity can be measured by looking at the <u>propat</u> variable. This – binary – variable has value one when the firm has applied for a patent in the period 2002-2004. Otherwise, the value of this variable equals zero. When this variable is averaged over a larger group of firms it gives an indication of the percentage of firms that applied for a patent. When this value between zero and one is interpreted as a chance, it then represents the propensity to patent of that group of firms.

VARIABLES

Firm size

For firm size, the most direct measure available in the CIS4 questionnaire is the turnover in 2002. To make the numbers smaller and easier to interpret, the log of turnover is taken. This is in line with earlier approaches. Another, methodological, note here is that this variable has been manipulated in order to make the participating firms unrecognizable. This slightly decreases the reliability of the sample. From a two-group mean comparison test it can be seen that there still is, as expected, a significant difference in turnover between firms that had applied for a patent and firms that did not. This is entirely as expected, since firm size is a measure for the amount of innovations a firm produced. The CIS database does not include measures on the amount of innovations and therefore firm size will be used as a control variable to account for the earlier described multiple innovations problem.

Dependent variable			
Patent application	*propat*		
Independent variable		**Test**	**Significance**
2002 turnover	*turn02log*	t-test, unequal variances	Pr (T > t) = 1.0000

Table 17 – T-test for 2002 turnover

R&D intensity

R&D intensity is not directly measured by CIS4, but can be calculated from total in-house R&D in 2004 and turnover in 2004.

As discussed before it turned out that several firms had values for R&D intensity exceeding one, which is conceptually troublesome. Two explanations are possible. The first is that due to the financial reporting method of the firm in-house R&D is broader than what is included in revenues. This can for example be the case with R&D institutes and commercial labs. The second explanation is that, due to the adjustments made by the mico-aggregation process, turnover values are lower than they are in reality which causes a higher R&D intensity value.

Sector	Frequency
2	1
7	4
8	1
9	1
11	1
13	7
24	2
25	11
26	21

Table 18 – Number of firms that have an R&D intensity higher than 1, split by sector

From the above shown table it can be seen that most outliers are present in sectors 25 and 26. These sectors both represent business activities and include R&D firms (these are sectors 72 to 74 in the NACE classification.) This then gives plausibility to the thought that most outliers are represented by special cases of financial reporting. Since these outliers have a very big impact on the means of the groups (almost all observations fall in between 0 and 1, while the outliers went up to values of 2000) the t-test has also been performed with an additional criterion that the R&D intensity should be below one. This may result in a small bias in the research population since 49 observations are excluded, however this will also limit the range of values for R&D intensity to between zero and one, so that it can be included in the model more easily.

Dependent variable			
Patent application	*propat*		
Independent variable		**Test**	**Significance**
R&D intensity	*rrdinxturn*	t-test, unequal variances	Pr (T > t) = 0.9984
R&D intensity < 1	*rrdinxturn if rrdinxturn <1*	t-test, unequal variances	Pr (T > t) = 1.0000

Table 19 – T-tests for R&D intensity and for R&D intensity excluding R&D intensities > 1

R&D intensity compared to industry average

Another interesting hypothesis concerning R&D intensity is that the impact of it on the propensity to patent does not depend so much on its absolute value, but rather on the comparison of the value with the average of the firm's sector of principal activity. Hypothesized here is that firms with an above average R&D intensity for their industry are the technological "front runners" and therefore patent more. To check for this, the R&D intensity was calculated for 25 sectors within CIS4 and after that subtracted from each firm's individual R&D intensity figure.

Sector		R&D intensity	Sector		R&D intensity
1	Mining	0.57 %	14	Automobiles and other transport equipment	3.13%
2	Food, Beverages and tobacco	0.87 %	15	Manufacture of furniture and recycling	1.38 %
3	Textiles and clothing	1.15 %	16	Electricity, gas, steam and water supply / treatment	4.74 %
4	Manufacture of leather and leather products	0.64 %	17	Construction	0.11 %
5	Wood, paper and printing	2.14 %	18	Sale, maintenance and repair of motor vehicles	0.17 %
6	Wood, paper and printing (publishing)	0.65 %	19	Wholesale trade and commission trade (except motor vehicles)	0.46 %
7	Petroleum Refining + chemicals and pharmaceuticals	7.00 %	20	Retail trade	0.01 %
8	Rubber and plastic products	2.64 %	21	Land, water and air transport	0.22 %
9	Glass, clay and ceramics	1.27 %	22	Auxiliary transport activities; travel agencies	0.18 %
10	Basic metals (iron and steel)	0.82 %	23	Post and telecommunication	2.84 %
11	Fabricated metal products	1.08 %	24	Financial intermediation	1.65 %
12	Mechanical engineering	2.97 %	25	Computer and related activities	11.47 %
13	Electrical equipment (including computers, communication equipment, precision instruments)	10.89 %	26	Research and development and other business activities	0.65 %

Table 20 – Average R&D intensity per sector

Deviation of R&D intensity from the sector average is found to have a significant relationship with the propensity to patent. Because this variable is based on the R&D intensity, an additional t-test has been performed where all firms with intensity ratios larger than one are excluded. See table 21.

Dependent variable			
Patent application	*propat*		
Independent variable		**Test**	**Significance**
R&D intensity compared to industry average	*rrdinxturnsd*	t-test, unequal variances	Pr (T > t) = 0.9964
R&D intensity compared to industry average, with R&D intensity < 1	*rrdinxturnsd if rrdinxturn <1*	t-test, unequal variances	Pr (T > t) = 0.9999

Table 21 – T-tests for R&D intensity compared to industry average

Co-operation

Co-operation is quite extensively measured in CIS4. Besides asking a yes/no question about the existence of any co-operation agreements on innovation activities, there are refinements along two dimensions: type of partner and location (National, Europe, US and all other countries.) When looking only at the binary variable co, it can be seen that there is a significant difference between the group that applies for a patent and the group that does not. See table 22. Since it is hypothesized that cooperation leads to a higher propensity to patent, a McNemar test is performed.

Dependent variable			
Patent application	*propat*		
Independent variable		**Test**	**Significance**
Co-operation	*co*	McNemar chi-square	Pr > chi2 = 0.0000

Table 22 – McNemar's chi-square test for co-operation

For the cooperation variable, CIS4 provides an additional question to probe into the type of partner and the location of the cooperation. For both national cooperation and international cooperation only, there is a significant difference in propensity to patent between firms that cooperate and firms that do not. See table 23.

Dependent variable			
Patent application	*propat*		
Independent variable		**Test**	**Significance**
National co-operation	*conatbin*	Standard chi-square	Pr = 0.000
International co-operation	*cointbin*	Standard chi-square	Pr = 0.000

Table 23 – Chi-square test for national and international co-operation

To look further into the possible relationship of cooperation and patent applications, a split can be made according to the type of partner. Since no direction can be assumed beforehand, for each type of partner a chi-square test has been performed, with an additional split being made for national and international cooperation. The results are given in the table below.

Type of co-operation partner	National	International
Other enterprises within enterprise group	0.000 (199)	0.000 (227)
Suppliers	0.000 (332)	0.000 (305)
Clients or customers	0.000 (361)	0.000 (280)
Competitors (or same industry)	0.025 (131)	0.000 (155)
Consultants, commercial laboratories and/or private R&D institutes	0.000 (263)	0.000 (145)
Universities or other higher education institutes	0.000 (371)	0.000 (92)
Government or private non-profit	0.000 (309)	0.000 (54)

Table 24 – Chi-square tests for subgroups based on type of partner and location (1)

Notes: (1) In between brackets is the number of observations in the smallest category of the original 2x2 table

From the chi-square tests in table 24 all factors have a significant relationship with the propensity to apply for a patent. However, since not all of these factors are expected to have a direct relationship with the propensity to patent, it can be expected that not all of them will be significant when included in the logistic regression model.

Since table 24 contains quite a lot of variables, an attempt has been made to group some of these individual variables together. This may simplify the model, while not only using the most general cooperation variable. To define those groups a split will be made according to the type of cooperation partner as well as to the location of the co-operation partner.

Note that in CIS4, consultants are included with commercial laboratories and private R&D institutes and do not form a separate category anymore (which was the case in CIS3, where it was found that cooperation with consultants has a negative impact on the propensity to patent.) All four newly defined variables have a significant relationship with the propensity to patent, as can be seen from the chi-square test in table 25 below.

Dependent variable			
Patent application	*propat*		
Independent variable		**Test**	**Significance**
Co-operation within enterprise group	*cooe*	Standard chi-square	Pr = 0.0000
Co-operation with suppliers	*cosup*	Standard chi-square	Pr = 0.0000
Co-operation with clients	*cocli*	Standard chi-square	Pr = 0.0000
Co-operation with competitors	*cocom*	Standard chi-square	Pr = 0.0000
Co-operation with commercial labs, R&D institutions and consultants	*cornd*	Standard chi-square	Pr = 0.0000
Cooperation with universities	*couni*	Standard chi-square	Pr = 0.0000
Cooperation with governments	*cogov*	Standard chi-square	Pr = 0.0000

Table 25 – Additional chi-square tests for the newly defined variables

Sources of information for innovation

For the sources of information for innovation the CIS questionnaire provides questions on different sources of information and their importance. Per source there is one question. The answers to these questions are measured at an ordinal scale (0 for not used; 1 for low; 2 for medium and 3 for high.) To test for a relationship with the propensity to patent two-sample Wilcoxon rank-sum (Mann-Whitney) tests have been performed. All tests, except for professional associations as information source for innovation, show a significant difference between the group that applies for a patent and the group that does not.

In the regression analysis, these variables will have to be transformed into multiple dummy variables to account for their ordinal level of measurement. This would lead to three dummy variables per variable. Another approach might be to make one dummy variable which has value zero if the information source is not used and value one if it is used (independent of its importance.) Besides that, a third approach might be to make the assumption that the differences between these numbers are of equal magnitude and that they thus provide information on an interval scale rather than at an ordinal scale.

Dependent variable			
Patent application	*propat*		
Independent variable		**Test**	**Significance**
Internal information sources	*sentg*	Wilcoxon rank-sum	Pr = 0.0000
Suppliers as information source	*ssup*	Wilcoxon rank-sum	Pr = 0.0001
Clients as information source	*scli*	Wilcoxon rank-sum	Pr = 0.0000
Competitors as information source	*scom*	Wilcoxon rank-sum	Pr = 0.0000
Consultants & commercial (R&D) labs as information source	*sins*	Wilcoxon rank-sum	Pr = 0.0000
Higher education institutes as information source	*suni*	Wilcoxon rank-sum	Pr = 0.0000
Government as information souce	*sgmt*	Wilcoxon rank-sum	Pr = 0.0000
Conferences as information source	*scon*	Wilcoxon rank-sum	Pr = 0.0000
Journals as information source	*sjou*	Wilcoxon rank-sum	Pr = 0.0000
Professional associations as information source	*spro*	Wilcoxon rank-sum	Pr = 0.0742

Table 26 – Wilcoxon rank-sum tests for different information sources

Preliminary runs of the model showed that most of these information sources for innovation turn insignificant when other factors were included into the model. Moreover, it turned out that for the variables that kept their significant impact, the relationship between low, medium and high importance was not linear. This was the case for both universities and private R&D institutes as information sources. For these variables three separate dummies have been created (suni1, suni2, suni3, sins1, sins2 and sins3) where suni1/sins1 represent low importance and suni3/sins3 represent high importance. For clients as information source all coefficients for low, medium as well as high were approximately equal. Therefore the variable clients is represented by a single dummy variable: sclibin (where 1 stands for either low, medium or high importance of this factor.)

Market location

CIS4 offers 4 binary questions about the selling of goods or services during the years 2002 to 2004 at a local/national/EU/other level. Each of those variables has a significant relationship with the propensity to patent, see figure 10. Notice that these categories are not mutually exclusive and firms can be present in multiple regions at the same time.

Dependent variable			
Patent application	*propat*		
Independent variable		**Test**	**Significance**
Local market	*marloc*	Standard chi-square	Pr = 0.000
National market	*marnat*	Standard chi-square	Pr = 0.002
European market	*mareur*	Standard chi-square	Pr = 0.000
Market outside of EU	*maroth*	McNemar's chi-square test	Pr > chi2 = 0.0000

Table 27 – Chi-square tests for market location variables

Countries

To include country level effects, such as patent legislation, a dummy variable has been created for each country: countryde, countryno and countrybe. Two of these will be included in the model to account for country level effects. All variables have a significant relationship to patent application. Since Germany is expected to have a positive relationship towards the propensity to patent, a McNemar's chi-square test is performed there.

Dependent variable			
Patent application	*propat*		
Independent variable		**Test**	**Significance**
Germany	*countryde*	McNemar's chi-square test	Pr > chi2 = 0.0000
Norway	*countryno*	Standard chi-square	Pr = 0.000
Belgium	*countrybe*	Standard chi-square	Pr = 0.000

Table 28 – Chi-square tests for country variables

Sectors

CIS4 uses the NACE system to classify firms into sectors. This classification system uses numbers for narrow industry definitions and letters for broader industry definitions. Each letter generally has multiple constituent numbers. In table 29 it is shown how the classification for CIS4 is done. Some industries have been taken at "letter level", while some are separated in more detail by using numbers. No overlap between these broader and narrower categories exists.

Sector	#	NACE	NACE
Mining	1	C	10
			11
			12
Mining of metal ores		CB	13
Other mining and quarrying			14
Food, Beverages and tabacco	2	DA	15
			16
Textiles and clothing	3	DB	17
			18
Manufacture of leather and leather products	4	DC	19
Wood, paper and printing	5	DD	20
		DE	21
Wood, paper and printing (publishing)	6		22
Petroleum Refining	7	DF	23
Chemicals and pharmaceuticals		DG	24
Rubber and plastic products	8	DH	25
Glass, clay and ceramics	9	DI	26
Basic metals (iron and steel)	10	DJ	27
Fabricated metal products	11		28
Mechanical engineering	12	DK	29
Computers	13	DL	30
Electrical machinery			31
Communication equipment			32
Precision instruments			33
Automobiles	14	DM	34
Other transport equipment			35
Manufacture of furniture	15	DN	36
Recycling			37
Electricity, gas, steam and hot water supply	16	E	40
Collection, purification and distribution of water			41
Construction	17	F	45
Sale, maintenance and repair of motor vehicles and motorcycles	18	G	50
Wholesale trade and commission trade (except motor vehicles)	19		51
Retail trade	20		52

Hotels and restaurants		H	55
Land, water and air transport	21	I	60
			61
			62
Auxiliary transport activities; travel agencies	22		63
Post and telecommunication	23		64
Financial intermediation	24	J	65
			66
			67
Real estate activities		K	70
Renting of machinery and equipment			71
Computer and related activities	25		72
Research and development	26		73
Other business activities			74

Table 29 – Sector classification

Notes: (1) Grey shaded areas show which level of detail is being used for that category
(2) Dark grey shaded areas identify multiple letter or numbers that were combined into one category
(3) Rows that are not shaded represent sectors that are not present in the CIS database
(4) The significance level is shown for a chi-square test for the relation between the sector and the propensity to patent

In total CIS4 distinguishes 26 sectors/industries. To easily account for them each sector is given a unique number. Each of these categories can be represented in a dummy variable, since all categories are mutually exclusive. For each of the variables, the significance of a chi-square test was calculated which tested the relationship of the sector variable to the propensity to patent. Notice however that for some categories, such as textiles and clothes (4), electricity, gas, steam and (hot) water supply (18) and wholesale trade and commission trade (20), expected cell counts are below five, making the chi-square test unreliable. For most others , the tests showed significant relationships with the propensity to patent.

Most other industries show significant differences with respect to the propensity to patent, which is in line with results of Arundel and Kabla (1998) who state that there are indeed differences in the propensity to patent between sectors.

The interesting thing now will be to check what part of these relationships can be explained by other factors, such as firm size. Since it is expected that several variables differ amongst industries it can thus be expected that the significance as well as the impact of several industry variables will change in the logistic regression model.

After several model runs, the sectors that scored lowest on their odds ratio coefficients have been combined into two more general variables, lowsector2 and lowsector3. The criterion to enter a variable into lowsector2 was a significant odd ratio between 0.1 and 1, while for lowsector3 this criterion was a significant odd ratio lower than 0.1. See table 35 further on for the odds ratios per industry.

Types of innovation

Within the product innovation category, CIS4 asks additional questions to refine product innovation into subcategories. Two splits are made: between goods and services and between new to the market and new to the firm only. Descriptive analysis has shown that substantial differences exist in the propensity to patent between these groups.

For service versus goods innovation it can be expected that goods have a higher propensity to be patented due to their more tangible nature. For “new to market” innovations it can be expected that they have a higher propensity to be patented than “new to firm” innovations, because patents require an inventive step. For new to the firm only innovations there is a larger chance that this requirement is not met.

Dependent variable			
Patent application	*propat*		
Independent variable		**Test**	**Significance**
Newness to market	*newmkt*	McNemar chi-square	Pr > chi2 = 0.0000
Goods innovation	*Inpdgd*	McNemar chi-square	Pr > chi2 = 0.0000

Table 30 – McNemar chi-square tests for newness to market and goods innovation

Both the new to market and the goods innovation variable have a (positive) significant impact on the propensity to patent. See the table above.

Funding of innovation activities

CIS4 also includes variables on public financial support for innovation activities, which have been hypothesized by Scherer (1983) to have an impact on the propensity to patent. Since Scherer does not separate efficiency and propensity to patent effects as is done in this thesis, it cannot be automatically assumed that funding will have the same (positive) relationship with the propensity to patent. However, as can be seen from the chi-squared tests below, all funding activities do have some relationship with the propensity to patent which is significant.

Dependent variable			
Patent application	*propat*		
Independent variable		**Test**	**Significance**
Local funding	*funloc*	Standard chi-square	Pr = 0.000
Government funding	*fungmt*	Standard chi-square	Pr = 0.000
EU funding	*funeu*	Standard chi-square	Pr = 0.000
EU funding – framework programme	*funrtd*	Standard chi-square	Pr = 0.000

Table 31 – Chi-square tests for several public financial support activities

Outsourced development

In the product innovation sub-part of CIS4 an option is presented to indicate that the innovation is mainly developed by other enterprises or institutions. It can be easily hypothesized that if others developed the innovation, the propensity to patent that innovation is generally lower. A McNemar chi-square test showed that a negative relationship to the propensity to patent exists with a prob > chi2 = 0.0000.

CIS4 MODELS

Below is an overview of variables that have been discussed in the previous section. All these variables may have an impact on the propensity to patent.

propat	Patent application during 2002-2004	**sentg**	Importance of an internal (within enterprise group) information source for innovation activities
turn02log	Log of turnover in 2002	**ssup**	Importance of suppliers (equipment, components or software) as information source for innovation activities
rrdinxturn	R&D intensity; in-house R&D in 2004 divided by turnover in 2004	**scli**	Importance of clients as information source for innovation activities
rrdinxturnsd	R&D intensity as deviation from sector average in 2004	**scom**	Importance of competitors as information source for innovation activities
co	Cooperation on innovation activities during 2002-2004	**sins**	Importance of consultants & commercial (R&D) labs as information source for innovation activities
conatbin	Cooperation on innovation with a partner in same country	**suni**	Importance of higher education institutes as information source for innovation activities
cointbin	Cooperation on innovation with a partner in a country abroad	**sgmt**	Importance of government and public institutes as information source for innovation activities
cooe	Cooperation on innovation with other enterprises within your enterprise group	**scon**	Importance of conferences, trade fair & exhibitions as information source for innovation activities
cosup	Cooperation on innovation with suppliers of equipment, materials, components or software	**sjou**	Importance of scientific journals & trade/technical publications as information source for innovation activities
cocom	Cooperation on innovation with competitors or other enterprises in you sector	**spro**	Importance of professional and industry associations as information source for innovation activities
cocli	Cooperation on innovation with clients or customers	**countrybe**	The firm is Belgian
cornd	Cooperation on innovation with consultants, commercial labs, or private R&D institutes	**countryde**	The firm is German
couni	Cooperation on innovation with universities or other higher education institutions	**countryno**	The firm is Norwegian
cogov	Cooperation on innovation with government or public research institutes	**mareur**	Firm sold goods in the EU during 2002-2004
marloc	Firm sold goods locally/regionally during 2002-2004	**maroth**	Firm sold goods in all other countries during 2002-2004
marnat	Firm sold goods nationally during 2002-2004	**funloc**	Firm received public financial support for innovation activities from local authorities
inpdgd	The firm introduced a new of significantly improved product during 2002-2004	**fungmt**	Firm received public financial support for innovation activities from central government
newmkt	The firm introduced a product innovation onto the market before competitors did (new to market)	**funeu**	Firm received public financial support for innovation activities from the European Union
sector*[x]*	The firm belongs to sector x according to the NACE classification	**funrtd**	Firm participated in the EU's fifth or 6th framework programme
lowsector2	Binary variable with value one if sector equals 2, 3, 5, 6, 10, 19, 21, 23, 25 or 26.	**lowsector3**	Binary variable with value one if sector equals 16, 22 or 24
inpdtwonly	The innovation is developed mainly by other enterprises or institutions		**For sector/NACE classifications, see appendix I**

Table 32 – Overview of variables

Except for professional associations as information source and some sector variables all variables have a significant relationship with the propensity to patent individually. The variables for universities and commercial R&D as information source have been split into three dummy variables, representing low importance (sins1 & suni1), medium importance (sins2 & suni2) and high importance (sins3 & suni3) for their respective innovation information source. For clients as information source a binary dummy has been created that has value one for all three levels of importance and value zero if not used (sclibin). This was done based on the preliminary model covered on the next page.

FIRST MODELS

To take a look at the interplay between those variables, model one will consist of all variables except the sector dummies, which will be added later on. Moreover, only the most general variable for cooperation (co) will be included (this variable will be included with more detail later on.) Other important conditions for the model are:

- A backward selection of variables, based on a significance level of 0.05
- The research population consists of all product innovators

Preliminary model run

Based on a preliminary model run, several observations can be made. First, sources of information for innovation, clients, universities, R&D institutes and industry associations are all significantly relevant. Second, all dummies for clients as information source have odds ratios that are about equal. Therefore, these three dummies will be combined into one that encompasses all three degrees of importance: sclibin. For the dummies that represent universities and research institutes as information sources, odd ratios increase and decline as expected. This means that all these relationships increase in strength as the degree of importance to that factor increases. Moreover, these increases are not linear but rather exponential for universities as information source. See table 33.

Number of observations		**2361**	**Pseudo R2**		**0.3051**
Variable	**Odds ratio**	**Z-score**	**Variable**	**Odds ratio**	**Z-score**
Turnover	1.34	9.69	Clients as information source (scli1)	2.43	2.60
Co-operation	1.76	4.69	Clients as information source (scli2)	2.48	2.80
Local funding	1.43	2.11	Clients as information source (scli3)	2.37	2.68
EU funding	1.57	2.20	Research institutes as information source (sins1)	0.63	-3.59
Goods innovation	4.85	9.82	Research institutes as information source (sins2)	0.53	-3.78
New to market	2.14	6.95	Universities as information source (suni1)	1.38	2.35
Innovation by others	0.04	-3.12	Universities as information source (suni2)	1.76	3.46
Local market	0.67	-3.29	Universities as information source (sun3)	2.95	4.54
European market	1.48	2.92	Professional associations as information source (spro2)	0.62	-3.12
Other markets	2.12	6.06	German firm	1.70	3.81

Note: (1) Countrybe was excluded due to between-term collinearity

Table 33 – Preliminary model, without sector dummies

Model one

We can now turn to the first model, that has the client as information source dummies (scli1, scli2 and scli3) combined into one dummy variable (sclibin). See the first column in table 35 for a summary of the model, and table a in Appendix II for the complete Stata output.

Several funding activities are not significant, which also goes for presence in the national market. Moreover, R&D intensity and R&D intensity relative to sector average are not significant. Especially for R&D intensity it could be that it mostly works at the industry level and therefore that industry dummies need to be included in order to see the effects of R&D intensity in the model. Model two will do exactly this, and will include dummy variables for all sectors defined by CIS.

Model two

Results of the regression analysis that included sector dummies are again tabulated in table 35 and in table b in Appendix II. In trying to perform this analysis with Stata, the calculation of the model stopped when including "R&D intensity relative to sector average" into the model as well. This is an indication of collinearity between "R&D intensity relative to sector average" and other variables. Most likely this is R&D intensity, since both variables are based on the same data. Moreover, both variables are influenced by the micro-aggregation process and may therefore include the same errors. A collinearity analysis shows that this is indeed the case:

Variables	VIF / Square root VIF	Tolerance	R-squared
R&D intensity			
R&D intensity minus sector average	246.42 / 15.70	0.0041	0.9959

Table 34 – Collinearity diagnostics for R&D intensity and R&D intensity minus sector average

Since VIF values are extremely high, collinearity is an issue here, even though the two variables conceptually measure quite different things (R&D intensity versus deviation from average sector R&D intensity.) However, since sector dummies are now included in the analysis it may well be that both concepts do measure more or less the same thing when R&D intensity is being adjusted for different sectors. Therefore, R&D intensity relative to sector average will be excluded in the next analysis.

Observations from the second model run include an increase in pseudo r-squared from 30% to 36%, which can be attributed to the adjustments made for sectors in the second model.

Furthermore local funding, European markets and universities as information source (suni1) are not significant anymore when including sector dummies into the model. This may indicate that these variables differ per sector and are thus (partly) explained by some of the included sectors. Even more important are the variables that have remained significant in the second model. Even when adjusted for industries, they have an impact on the propensity to patent.

Also, Germany is still present as a variable in the model. This indicates that the higher propensity to patent that was observed for German firms cannot be adequately explained by differences in the industries that are present in that country compared to industries present in Norway and Belgium.

Logistic regression		**Model one** No sector distinction		**Model two** Sectors included		**Model three** Split cooperation		**Model four** Sectors combined R&D intensity < 10%	
	Conditions	Inpdt=1		Inpdt=1		Inpdt=1		Inpdt=1	
		Rrdinxturn<1		Rrdinxturn<1		Rrdinxturn<1		Rrdinxturn<0.1	
		Backwards (0.05)**		Backwards (0.05)		Backwards (0.05)		Backwards (0.05)**	
Label	**Variables**	Odds-ratio	SE	Odds-ratio	SE	Odds-ratio	SE	Odds-ratio	SE
Turnover	TURN02LOG	1.33 (**9.57**)	0.040	1.53 (**11.49**)	0.057	1.51 (**11.03**)	0.057	1.55 (**11.05**)	0.061
R&D intensity	RRDINXTURN	n.s.		10.30 (**3.80**)	6.333	7.32 (*3.14*)	4.639	2.59 (**3.54**)*	0.697
Type=goods	INPDGD	4.86 (**9.85**)	0.781	2.85 (**5.32**)	0.561	2.91 (**5.33**)	0.582	2.59 (**4.94**)	0.499
New to market	NEWMKT	2.14 (**6.99**)	0.233	2.01 (**6.06**)	0.231	2.05 (**6.23**)	0.236	2.00 (**5.63**)	0.245
Cooperation	CO	1.76 (**4.68**)	0.211	1.87 (**4.93**)	0.236	-		-	
national	CONATBIN	-		-		n.s.		n.s.	
international	COINTBIN	-		-		1.99 (*2.84*)	0.485	1.83 (*2.26*)	0.490
enterprise group	COOE	-		-		n.s.		n.s.	
suppliers	COSUP	-		-		n.s.		n.s.	
clients	COCLI	-		-		n.s.		n.s.	
competitors	COCOM	-		-		n.s.		n.s.	
R&D institutes	CORND	-		-		n.s.		n.s.	
universities	COUNI	-		-		2.13 (**5.04**)	0.320	2.21 (**4.87**)	0.359
government	COGOV	-		-		n.s.		n.s.	
Local funding	FUNLOC	1.42 (*2.10*)	0.240	n.s.		n.s.		n.s.	
EU funding	FUNEU	1.56 (*2.20*)	0.320	1.72 (*2.62*)	0.359	1.63 (*2.31*)	0.348	1.89 (*2.65*)	0.456
Local market	MARLOC	0.66 (*-3.41*)	0.081	0.74 (*-2.35*)	0.095	0.74 (*-2.27*)	0.097	0.72 (*-2.37*)	0.100
EU market	MAREUR	1.50 (*3.05*)	0.198	n.s.		n.s.		n.s.	
Other market	MAROTH	2.14 (**6.19**)	0.264	1.85 (**5.21**)	0.219	1.86 (**5.26**)	0.221	1.99 (**5.46**)	0.252
Clients	SLCIBIN	2.39 (*2.77*)	0.751	2.05 (*2.16*)	0.068	1.96 (*2.00*)	0.663	2.43 (*2.50*)	0.860
Commercial labs	SINS	-		-		-		-	
low importance	SINS1	0.63 (**-3.68**)	0.080	0.69 (*-2.87*)	0.089	0.74 (*-2.37*)	0.095	0.68 (*-2.83*)	0.093
medium importance	SINS2	0.50 (**-4.06**)	0.085	0.52 (**-3.70**)	0.092	0.58 (*-3.11*)	0.101	0.49 (**-3.78**)	0.092
high importance	SINS3	n.s.		n.s.		n.s.		n.s.	
Universities	SUNI	-		-		-		-	
low importance	SUNI1	1.35 (*2.19*)	0.186	n.s.		n.s.		n.s.	
medium importance	SUNI2	1.68 (*3.19*)	0.273	1.38 (*2.23*)	0.200	n.s.		n.s.	
high importance	SUNI3	2.80 (**4.34**)	0.663	2.04 (*3.10*)	0.470	n.s.		n.s.	
Innovation by others	INPDTWONLY	0.04 (*-3.14*)	0.041	0.04 (*-3.23*)	0.037	0.04 (*-3.23*)	0.038	0.05 (*-3.02*)	0.047
Germany	COUNTRYDE	1.76 (**4.06**)	0.244	1.90 (**4.40**)	0.277	2.89 (**5.83**)	0.526	2.18 (**4.39**)	0.389
Sectors odd <1	LOWSECTOR2	-		-		-		0.26 (**-9.65**)	0.036
Sectors odd <0.1	LOWSECTOR3	-		-		-		0.04 (**-6.67**)	0.019
Pseudo R-squared			**0.30**		**0.36**		**0.37**		**0.38**
Model significance			**0.0000**		**0.0000**		**0.0000**		**0.0000**
Number of observations			**2361**		**2355**		**2356**		**2075**

Logistic regression					
	Model one	Model two		Model three		Model four
SECTOR1	-	n.s.		n.s.		-
SECTOR2	-	0.11 (**-6.93**)	0.034	0.10 (**-6.91**)	0.034	-
SECTOR3	-	0.18 (**-4.28**)	0.073	0.18 (**-4.32**)	0.072	-
SECTOR4	-	n.s.		n.s.		-
SECTOR5	-	0.42 (*-2.99*)	0.122	0.42 (*-3.01*)	0.121	-
SECTOR6	-	0.13 (**-4.33**)	0.062	0.13 (**-4.35**)	0.061	-
SECTOR7	-	n.s.		n.s.		-
SECTOR8	-	n.s.		n.s.		-
SECTOR9	-	n.s.		n.s.		-
SECTOR10	-	0.30 (*-2.90*)	0.124	0.28 (*-2.95*)	0.121	-
SECTOR11	-	- reference		- reference		-
SECTOR12	-	n.s.		n.s.		-
SECTOR13	-	n.s.		n.s.		-
SECTOR14	-	n.s.		n.s.		-
SECTOR15	-	n.s.		n.s.		-
SECTOR16	-	0.07 (*-3.43*)	0.054	0.06 (*-3.40*)	0.048	-
SECTOR17	-	n.s.		n.s.		-
SECTOR18	-	n.s.		n.s.		-
SECTOR19	-	0.45 (*-2.50*)	0.144	0.47 (*-2.28*)	0.155	-
SECTOR20	-	n.s.		n.s.		-
SECTOR21	-	0.27 (*-2.48*)	0.143	0.28 (*-2.42*)	0.147	-
SECTOR22	-	0.07 (*-2.98*)	0.065	0.07 (*-2.89*)	0.067	-
SECTOR23	-	0.16 (*-2.62*)	0.111	0.17 (*-2.40*)	0.125	-
SECTOR24	-	0.02 (**-5.23**)	0.014	0.02 (**-5.08**)	0.016	-
SECTOR25	-	0.15 (**-6.87**)	0.041	0.16 (**-6.64**)	0.044	-
SECTOR26	-	0.59 (*-2.31*)	0.134	0.58 (*-2.34*)	0.134	-

Notes: Bold z-values have a significance <0.001,
Italic z-values have a significance <0.05 (Z-values in between brackets)
"n.s." means that the factor was excluded from the model based on p>0.05
"-" means that the factor was excluded beforehand
* R&D intensity has been multiplied by ten here.
** For models 1 & 4 backwards and forwards selection methods lead to the same result.

Table 35 – Overview of models 1 to 4

Model three

The third model goes into more detail on co-operation arrangements. Nine subcategories have been defined, based on location of the cooperation (national/international) and type of cooperation partner. See the results for this model in table 35.

From these results it can be seen that most types of cooperation arrangements do not have a significant impact on the propensity to patent. Only international cooperation and cooperation with universities tend to have a significantly increased propensity to patent. Moreover, the significance of cooperation with universities goes at cost of the presence of all universities as information source dummies. This could of course be expected, since cooperation almost immediately also means that the cooperation partner – the universities in this case – is also seen as an information source for innovation.

Nonetheless, the pseudo r-squared increases a bit, which may be an indication that the split up of cooperation variables increases explained variance. However, especially with small increases it has to be kept in mind that this pseudo r-squared measure is not a perfect indicator of the actual explained variance.

Model four

To further refine the present model, two additional adjustments can be made. First, many sector variables are present at the time in the model. To decrease the number of variables the significant industries will be combined into two generalized categories represented by lowsector2 and lowsector3 (for an operationalization, see table 32 at the start of this chapter.)

Another adjustment made is based on observation that many firms currently have values for their R&D intensity that are above 10%. This is extremely high, since the overall average R&D intensity lies around 3%. By excluding all observations with R&D intensities higher than 10%, the reliability of the sample will probably increase, since most of these observations are likely have errors in the underlying variables of turnover and/or R&D expenditures. This is then likely a result of the micro-aggregation process.

Most importantly though, excluding them will give a better picture of the variable R&D intensity, since outliers for this variable will be excluded. Next to that, by setting the maximum R&D intensity to 10% it is possible to get a more clear and insightful figure for the odds ratio of the R&D intensity. By multiplying R&D intensity by 10, its values will then lie in between one and zero. In models five and six, the assumption that most observations with R&D intensities above 10% are erroneous will be loosened to R&D intensities above 50%, by excluding some sectors that have relatively many of these outliers.

PREDICTIVE POWER OF THE MODELS

One drawback of logistic regression is the fact that there is no objective measure for how well the model fits the data. As a standard, Stata provides each logistic regression model with McFadden's R2. (These are the values for pseudo R-squared in the logistic regression models above.) However, this measure cannot be interpreted as the proportion of explained variance in the model. Several pseudo R-squared measures have been defined by different authors. An additional package has been downloaded with Stata that provides most of these (fitstat.) Results are shown below for model four.

Measure	Score	Measure	Score
McFadden's R2	0.376	McFadden's Adj R2	0.363
Maximum likelihood R2	0.386	Cragg & Uhler's R2	0.531
McKelvey and Zavoina's R2	0.628	Efron's R2	0.432
Variance of y*	8.852	Variance of error	3.290
Count R2	0.810	Adj Count R2	0.463
AIC	0.828	AIC*n	1717

Table 36 – Fitstat results for model 4

From this it can be seen that pseudo R-squared measures range from 0.376 for McFadden's R2 to 0.628 for McKelvey and Zavoina's R2. Since these values are so much apart and clear interpretation is almost impossible, some additional tests will be performed by comparing predicted values for the propensity to patent to the values that are actually observed.

A first possibility is to predict the outcomes of model 4 into the new variable pmodel4. Then a criterion is set that values of pmodel4 that are higher than 0.5 will be predicted as successes (patent applications) and values lower than 0.5 will be predicted as failures (no patent application.) When these values are tabulated against the observed patent applications one can see that 1206 + 518 cases were predicted successfully, while 178 + 232 cases yielded another result than predicted. See figure 37. This means that the model predicts a correct result for **80.8%** of all observations. For models one to three these numbers are respectively 77.5%, 80.5% and 78.7%.

Predicted / Actual	0	1	Total
0	1,206	232	1,438
1	178	518	696
Total	1,384	750	2,134

Table 37 – Predicted versus actual outcomes for model 4

To get an idea of how well this model performs these results can be compared to what would have been an "educated guess" without any model. Since the majority of cases yield no patent application, a good guess would have been to predict only zero outcomes. In that case, 1206 + 178 = 1384 cases would have been predicted correctly. Compared to this approach, model 4 predicts ((1206 + 518 – 1384) / 1384) = **24.6%** more correct outcomes.

Stata can also perform a test based on this idea, which is the Hosmer and Lemeshow's goodness-of-fit test. Instead of using two categories, it is common practice to use the test by creating 10 groups and form a contingency table of 2 by 10 for a test with about 700 observations (ATS, 2011). The results are shown on the next page for each model, as well as an example for model 4. The model can be seen as a good fit when it passes Hosmer and Lemeshow's goodness-of-fit test, which is the case when prob > chi2 is bigger than 0.05. The test does not say anything about the extent of the fit.

Group	Probability	Observed 1	Expected 1	Observed 0	Expected 0	Total
1	0.0131	3	1.2	205	206.8	208
2	0.0411	7	5.2	200	201.8	207
3	0.0879	17	13.1	191	194.9	208
4	0.1608	29	25.2	178	181.8	207
5	0.2731	33	43.8	175	164.2	208
6	0.4012	71	70.3	136	136.7	207
7	0.5471	88	99.3	120	108.7	208
8	0.6845	129	127.9	78	79.1	207
9	0.8395	163	158.5	45	49.5	208
10	0.9964	195	190.7	12	16.3	207
Number of observations			**2075**	**Hosmer-Lemeshow chi2(8)**		**12.91**
Number of groups			**10**	**Prob > chi2**		**0.1151**

Table 38 – Hosmer and Lemeshow's goodness-of-fit test for model 4

In the table below the Hosmer and Lemeshow's results are shown, as well as the percentage correctly predicted outcomes and McFadden's R2:

Measure	Model 1	Model 2	Model 3	Model 4
McFadden's R2	0.302	0.364	0.368	0.376
Hosmer and Lemeshow's prob > chi2 (8)	0.77	0.09	0.21	0.12
Percentage correctly predicted outcomes	77.5%	80.5%	78.7%	80.8%

Table 39 – Quality measures for model 1 to 4

According to these statistics, model four has the highest quality. All models pass the Hosmer and Lemeshow goodness-of-fit test. Model one scores the lowest on both measures, which is not surprising given the fact that it is the only model that does not include sector variables. Quality measures for model two and three show numbers close to those of model four. However, model four has an additional advantage in that it has only two sector variables instead of 26. It is therefore simpler and needs fewer variables to explain differences in the propensity to patent. Model four will thus provide the basis for further analysis.

ANALYSIS OF WRONGLY PREDICTED OBSERVATIONS

To further refine model four into a fifth model an analysis will be performed of firms that find themselves at extremely high or low predicted propensities to patent, but that nonetheless do not patent (or do.) This will be done by separating all firms that were included in the analysis into 10 groups that are ordered by their predicted chance to apply for a patent (1 = 0-10%; 2 = 10-20% etc.)

Group with predicted chance of:	**# no patent**	**# patent**	**Group with predicted chance of:**	**# no patent**	**# patent**
0%-10%	672	37	50%-60%	85	73
10%-20%	252	44	60%-70%	79	117
20%-30%	141	37	70%-80%	47	104
30%-40%	133	62	80%-90%	30	141
40%-50%	98	65	90%-100%	35	189

Table 40 – Patent applications, split to groups with equal predicted chances (per 10%)

Both the highest category (90%-100%) and the lowest category (0%-10%) of table 40 are tabulated against sector and probability to patent. On the left side of table 41, the numbers of firms are shown that had a very low predicted probability to patent, but nonetheless do apply for one. On the right side of table 41, the numbers of firms are shown that had a very high predicted probability to patent, but nonetheless don't apply.

Number of firms with predicted probability <10% that do apply for a patent.				**Number of firms with predicted probability >90% that don't apply for a patent.**			
Sector	**#**	**Sector**	**#**	**Sector**	**#**	**Sector**	**#**
2	2	16	1	2	0	16	0
3	1	17	1	3	0	17	0
5	2	19	4	5	0	19	0
6	1	21	2	6	0	21	0
7	0	22	1	7	5	22	0
12	2	23	0	12	2	23	1
13	0	24	2	13	9	24	0
14	0	25	2	14	1	25	14
15	0	26	15	15	1	26	2

Note: (1) Sectors that don't have any observations in one of the groups are not shown.

Table 41 – Number of firms per sector that have wrongly predicted patent propensities

From these tables it can be seen that sector 26 has an extreme high proportion of firms that are predicted a very low chance to apply for a patent but nonetheless do. The opposite goes for sector 25. It thus seems as if sector 25 and 26 are rather special sectors. Sector 25 represents NACE sector 72, which consists of computer and related activities (hardware consultancy, software consultancy, data processing, database activities and maintenance/repair.) Sector 26 represents NACE sectors 73 and 74 which are research and development and other business activities (including legal, advertising and technical testing and analysis.)

Model 5

For model five, sectors 25 and 26 have been excluded from the research population. Since also most outlying values for R&D propensity were present in these sectors, the requirement that firms should have an R&D intensity lower than 10% has been increased to R&D propensities lower than 50%. When looking at the results for the fifth model (table 43 on the next page) no real surprises show in the impact and z-value of variables. Only international co-operation becomes insignificant compared to model four. The model passes Hosmer and Lemeshow's goodness-of-fit test and the other quality measures show numbers which are about equal to previous models. Therefore, the most important observation here is that predicted variables and their odds-ratios are very stable between the models.

ANALYSIS OF MISSING OBSERVATIONS

When performing a logistic regression analysis, Stata automatically excludes all firms that have missing values for one or more of the dependent and independent variables. When looking at these missing values, two observations can be made. The first observation is that for most excluded firms, only a few values are missing. This means that there is still a lot of information present in the data that is not taken into account when performing the logistic regression analysis. A second observation is that missing values are especially common for firms of the Belgian database. (A note here might thus be that the quality of the Belgian database is somewhat less than the quality of the Norwegian and German ones.) This might also explain the between term collinearity of the Belgian country variable in the first model.

To increase the number of observations, and to include more Belgian firms into the regression model, a new model will be generated. For this model it will be assumed that all missing values have the value of zero. It is immediately clear then that an increase in number of observations will go at cost of reliability, since it is not entirely sure that these missing values are indeed equal to zero. For most of them the assumption is likely to hold though, since many questions in the CIS questionnaire (especially about most significant markets) might be interpreted as not needing an answer when the question does not apply. In this way, model 6 will be able to provide more clarity about the robustness of earlier found factors since it now includes approximately 50% more observations (3037 compared to 1913.) Moreover, it provides more insight into differences between countries, since both the German and Norwegian country variable are present in the model.

Model 6

Model 6 is summarized in the table on the next page. At first sight, there are no major differences compared to model five. Most z-values are slightly higher because of the higher number of included observations. Moreover, the Norwegian country variable is now included with a z-value of just below 6, which is very high. There are no direct indications that there are indeed problems with reliability, although the McFadden's R2 measure dropped to 33%. The percentage of correctly predicted outcomes, however, increased comparing to model five. Model six will mainly be used to analyze sectorial and country differences, because of the high number of included observations (which is essential for having as many observations as possible per sector.)

Measure	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
McFadden's R2	0.302	0.364	0.368	0.376	0.371	0.333
Hosmer and Lemeshow's prob > chi2 (8)	0.77	0.09	0.21	0.12	0.32	0.15
Percentage correctly predicted outcomes	77.5%	80.5%	78.7%	80.8%	78.7%	80.0%

Table 42 – Quality measures for model 1 to 6

Logistic regression		**Model five** R&D intensity < 50% Excluding sector 25, 26		**Model six** R&D intensity < 50% Excluding sector 25, 26 Including missing obs.	
	Conditions	Inpdt=1 Rrdinxturn<0.5 Backwards (0.05)		Inpdt=1 Rrdinxturn<0.5 Backwards (0.05)	
Label	**Variables**	Odds-ratio	SE	Odds-ratio	SE
Turnover	TURN02LOG	1.61 (**11.50**)	0.067	1.45 (**11.73**)	0.047
R&D intensity	RRDINXTURN	1.53 (**3.76**)*	0.174	1.47 (**4.35**)*	0.133
Type=goods	INPDGD	2.98 (**4.79**)	0.692	2.67 (**5.09**)	0.516
New to market	NEWMKT	2.15 (**6.16**)	0.268	2.07 (**7.16**)	0.209
Cooperation	CO	-		-	
national	CONATBIN	n.s.		n.s.	
international	COINTBIN	n.s.		1.61 (*3.03*)	0.252
enterprise group	COOE	n.s.		n.s.	
suppliers	COSUP	n.s.		n.s.	
R&D institutes	CORND	n.s.		n.s.	
universities	COUNI	2.55 (**6.12**)	0.389	2.01 (**5.43**)	0.257
government	COGOV	n.s.		n.s.	
Local funding	FUNLOC	n.s.		n.s.	
EU funding	FUNEU	1.97 (*2.80*)	0.480	1.80 (*3.06*)	0.345
Local market	MARLOC	0.70 (*-2.49*)	0.099	0.75 (*-2.26*)	0.096
EU market	MAREUR	n.s.		n.s.	
Other market	MAROTH	1.94 (**5.19**)	0.248	1.78 (**5.61**)	0.184
Clients	SLCIBIN	2.07 (*2.08*)	0.728	2.89 (**3.91**)	0.786
Suppliers	SSUPBIN	n.s.		0.60 (-2.61)	0.117
Commercial labs	SINS	-		-	
low importance	SINS1	0.68 (*-2.72*)	0.095	n.s.	
medium importance	SINS2	0.50 (**-3.64**)	0.095	n.s.	
high importance	SINS3	n.s.		n.s.	
Universities	SUNI	-		-	
low importance	SUNI1	n.s.		n.s.	
medium importance	SUNI2	n.s.		n.s.	
high importance	SUNI3	n.s.		1.64 (*2.52*)	0.324
Innovation by others	INPDTWONLY	0.05 (*-2.96*)	0.049	0.09 (**-3.92**)	0.056
Germany	COUNTRYDE	1.91 (**4.18**)	0.297	5.96 (**11.11**)	0.960
Norway	COUNTRYNO	-		2.39 (**5.78**)	0.362
Sectors odd <1	LOWSECTOR2	0.23 (**-9.16**)	0.037	0.34 (**-8.90**)	0.041
Sectors odd <0.1	LOWSECTOR3	0.04 (**-6.71**)	0.018	0.05 (**-7.20**)	0.021
Pseudo R-squared			**0.37**		**0.33**
Model significance			**0.0000**		**0.0000**
Number of observations			**1913**		**3037**

Notes: Bold z-values have a significance <0.001,
Italic z-values have a significance <0.05 (Z-values in between brackets)
"n.s." means that the factor was excluded from the model based on p>0.05
"-" means that the factor was excluded beforehand
* R&D intensity has been multiplied by ten here.

Table 43 – Overview of models 5 and 6

Odds ratios and factor impact

Odds ratios are interpreted as if they are a relative risk. For example, an odds ratio of 2 indicates that a firm for which the factor applies has odds of two to one of applying for a patent compared to a firm for which the factor does not apply. Still it can be hard to directly intuitively interpret an odds ratio. To get a grip on the impact a change in one variable has, the change in predicted patent propensity is calculated from a base case. Both the base case propensity and the new propensity are based on the regression equation that underlies model five. The base case consists of a German goods innovator with a turnover of one million Euros and an R&D intensity of 3%. In that case, predicted patent propensity is 27%. See the table below. Additionally, an example of the cumulative effect of changes in variables is shown in figure 15.

Base case: German goods innovator, €1M turnover, 3% R&D intensity			**0.27**
Variable	**New patent propensity**	**Variable**	**New patent propensity**
Turnover €10M	0.49 (+81%)	Turnover €1B	0.90 (+233%)
R&D intensity 10%	0.88 (+226%)	Clients as source	0.44 (+63%)
Service innovation	0.11 (-59%)	Commercial labs low	0.20 (-74%)
New to market	0.45 (+66%)	Commercial labs medium	0.16 (-41%)
University co-operation	0.49 (+81%)	Innovation by others	0.02 (-93%)
EU funding	0.43 (+59%)	Not German	0.16 (-59%)
Local market	0.21 (-22%)	Low propensity sector	0.08 (-70%)
Market outside EU	0.42 (+56%)	Lowest propensity sector	0.01 (-92%)

Table 44 – Change in predicted patent propensity from base case when a variable changes

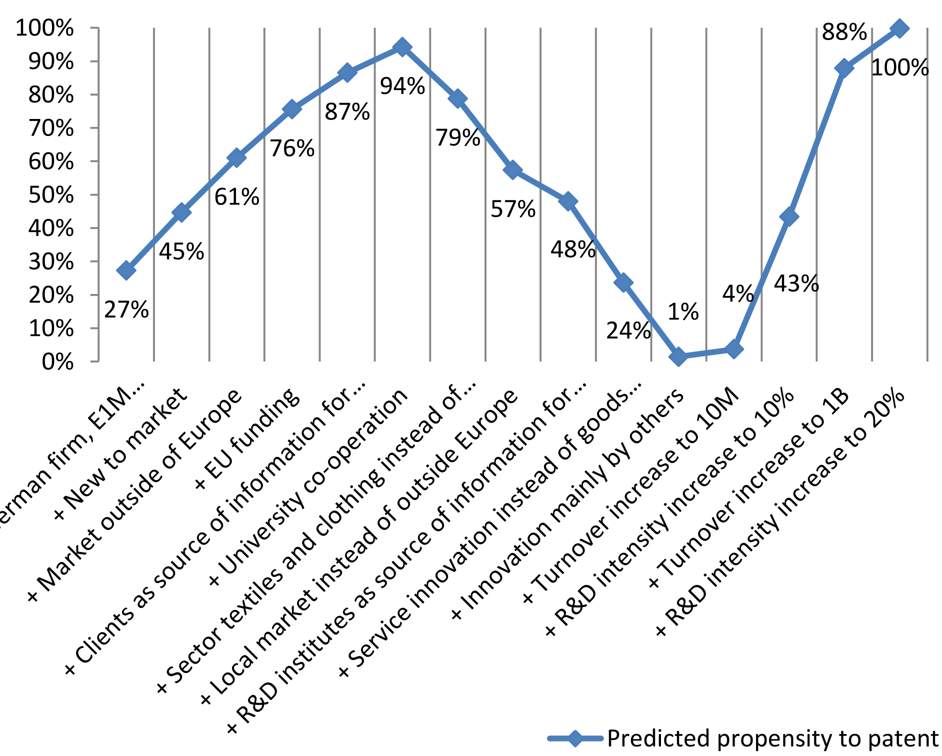

Figure 15 – Example of cumulative impact of variable changes on the propensity to patent

Sectors

To look at differences in propensity to patent between sectors, model six has been extended with all sector variables. To define a high and a low propensity to patent group, the reference sector has been taken as textiles and clothing (sector 3) with an average propensity to patent of 9%. Including all sectors again, instead of lowsector2 and lowsector3, raised the pseudo R-squared measure to 35% for model 6. With backwards selection of variables, the remaining – significant – sectors have been included in table 45.

Worth noticing is that three categories arise from this table. The first category consists of the variables that are not included because of their significance. For sectors 4, 18 and 20 this is likely to be because of the few observations present in these categories (less than 12.) All other sectors have at least 50 observations. It is thus likely that all other sectors that were not included actually have patent propensities close to those of textiles and clothing. This idea is confirmed by looking at the descriptive patent propensities that have been analyzed before. The other two categories consist of either sectors that have lower propensities to patent or higher propensities to patent relative to textiles and clothing. (All sectors with odd-ratio's higher than one have significantly higher propensities to patent than textiles and clothing.) Worth noticing is that the 95% confidence intervals overlap for both the all the high and all the low sectors. The given odds-ratios should thus not automatically be interpreted as ranking the sectors on patent propensities within these three categories.

Sector	Label	Odds-ratio
SECTOR1	Mining	- ref
SECTOR2	Food, Beverages and tobacco	**0.32**
SECTOR3	Textiles and clothing	**- ref**
SECTOR4*	Manufacture of leather and leather products	- ref
SECTOR5	Wood, paper and printing	- ref
SECTOR6	Wood, paper and printing (publishing)	**0.35**
SECTOR7	Petroleum Refining + chemicals and pharmaceuticals	**1.80**
SECTOR8	Rubber and plastic products	**2.32**
SECTOR9	Glass, clay and ceramics	- ref
SECTOR10	Basic metals (iron and steel)	- ref
SECTOR11	Fabricated metal products	**2.68**
SECTOR12	Mechanical engineering	**3.06**
SECTOR13	Electrical equipment (computers, communication, precision instr.)	**2.06**
SECTOR14	Automobiles and other transport equipment	**2.31**
SECTOR15	Manufacture of furniture	**2.14**
SECTOR16	Electricity, gas, steam and (hot) water supply	**0.19**
SECTOR17	Construction	- ref
SECTOR18*	Sale, maintenance and repair of motor vehicles and motorcycles	- ref
SECTOR19	Wholesale trade and commision trade (exept motor vehicles)	**1.77**
SECTOR20*	Retail trade (exept motor vehicles)	- ref
SECTOR21	Land, water and air transport	- ref
SECTOR22	Auxiliary transport activities; travel agencies	**0.15**
SECTOR23	Post and telecommunication	**0.19**
SECTOR24	Financial intermediation	**0.06**

Notes: (1) The original reference group was taken as textiles and clothing (sector 3)

Table 45 – Industry odds-ratios toward patent propensity, based on model six

Countries

Country	Odds-ratio	Z-value (p-value)	S.E.	95% lower	95% upper
Germany	5.96	**11.11** (0.000)	0.960	4.35	8.18
Norway	2.39	**5.78** (0.000)	0.362	1.78	3.22
Belgium					*- Reference*

Table 46 – Country odds-ratios toward patent propensity, based on model six

From model 6 also some interesting observations can be made about the differences in propensity to patent between countries. First, a very high odd-ratio is observed for the variable indicating that the firm originates from the German CIS. (See table 46.) Since the Norwegian country variable is also present in the model, the odds-ratio indicates that, other things being equal, the odds of a German firm patenting compared to a Belgian firm are almost 6 to 1. Immediately it should be realized that this figure is a best guess, based on the logistic regression model, and that standard error is quite large. It is therefore likely that in reality the German-firm patenting-odds are not 6 times those of Belgian firms. (The 95% confidence interval for the German variable runs from 4.35 to 8.18.)

Also, when looking at the 95% confidence intervals it can be seen that the upper bound of the Norwegian variable is lower than the lower bound of the German variable. Moreover, the Norwegian lower bound is still above one. An intermediary conclusion that can thus be drawn here is that the patent propensity of German firms is higher than the patent propensity of Norwegian firms, which is in turn higher than the propensity to patent of Belgian firms. This is then after other factors, such as industry differences and firm level factors, are taken into account.

In appendix III an overview is given of results from model six for each individual country. Interesting to see are the quality measures of the model, where McFadden's R2 is substantially higher for Germany than for Norway and Belgium. Moreover, the model for Belgium seems not to be a good fit since it does not pass Hosmer and Lemeshow's goodness-of-fit test. (Prob > chi2 is lower than 0.05.) See table 47.

Measure	**Germany**	**Norway**	**Belgium**
McFadden's R2	0.405	0.171	0.287
Hosmer and Lemeshow's prob > chi2 (8)	0.12	0.60	0.04
Percentage correctly predicted outcomes	80.5%	83.6%	76.1%

Table 47 – Quality measures model 6 for individual countries

CONCLUSION AND DISCUSSION

Main factors that are found to be relevant in determining the propensity to patent include turnover, R&D intensity, EU funding, geographical location of the firms market, universities as cooperation partner / source of information and R&D institutes as a source of information. Moreover, the propensity to patent depends on the characteristics of the innovation, such as its newness to market or the fact that it is a good instead of a service. Innovation that was developed mainly by other enterprises or institutions was – logically – found to be far less patented by the firm in question. On top of that, clear differences exist between sectors and countries since these variables occur in all models. Especially Germany has a higher propensity to patent than Norway or Belgium. These findings will be elaborated upon below to answer the research (sub) questions.

Differences between type of innovation

A clear conclusion can be drawn from the descriptive results, which is that goods innovations are far more likely to be patented compared to services or processes. For firms that have only a goods innovation, average patent propensity equals **38.6%.** For firms that have only a service innovation that same propensity equals only **9.6%.** For process innovation the propensity to patent is even lower: **7.6%** of firms that only have process innovation apply for a patent. This figure is remarkably low compared to earlier work by Arundel and Kabla (1998), which does not look at process innovation *only*. For an overview, see the figure below.

Another interesting observation from this figure is that patent propensity increases to 44.6% for firms that have both process and goods innovations. For the combination of services and goods the opposite is true: average propensity to patent a goods innovation declines when the firm also has a service innovation. Having a process innovation thus correlates with a higher propensity to patent goods innovation, while having service innovation correlates with lower propensity to patent goods innovation.

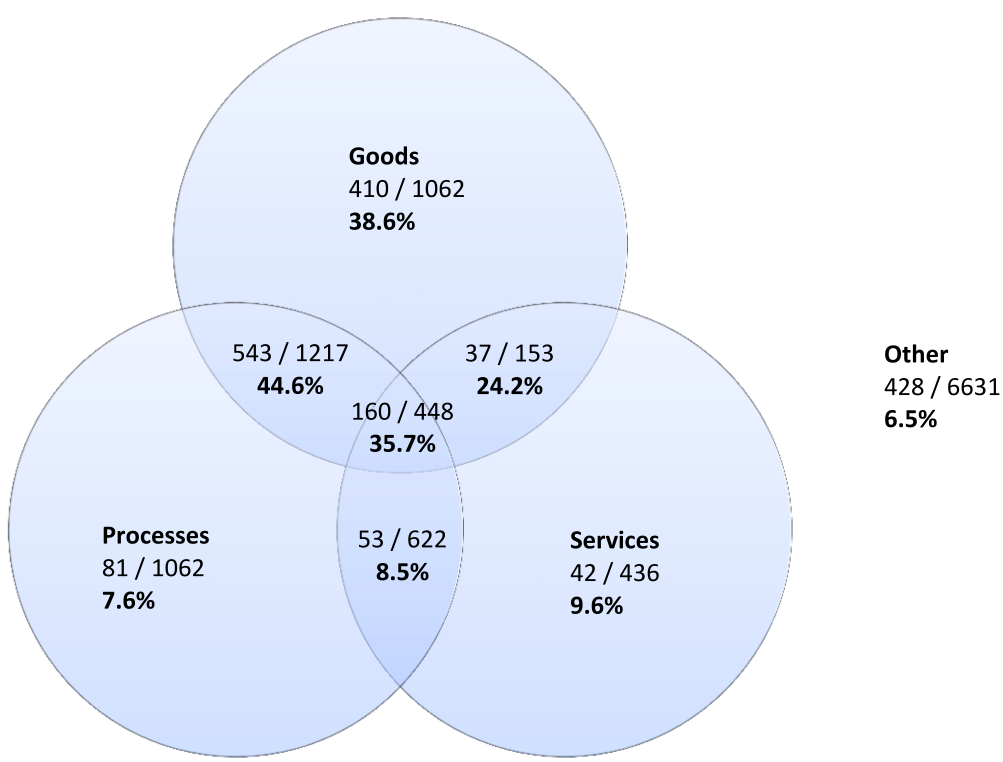

Figure 16 – Patent propensity rates for goods, services and process innovators

Based on these descriptive results, the odds of applying for a patent based on a goods innovation, compared to service and process innovation, are about 4 to 1. Regression analysis focused at product innovation and thereby included goods and services into the logistic regression model. The variable that distinguishes goods innovation is found to have a highly significant odds ratio of about 3 across different models. Combining these observations, there is an indication that other factors are correlated with goods innovations that explain part of their higher propensity to patent. Especially different sectors and size may play a role here. A quick check confirms that the average turnover of goods innovators is about 30% higher than the average turnover of others. This finding is important since the inherent difference in patent propensity between goods and service innovation is probably less than descriptive results may suggest.

Differences between sectors

Besides differences in propensity to patent based on the underlying type of innovation, regression results also showed significant differences between industries. Moreover, these differences persist when other firm and country level factors are taken into account. Based on table 45, a list can be compiled of sectors that have a structurally higher propensity to patent than others and sectors that have a structurally lower propensity to patent than others. This list is adjusted for other factors that are present in model six, and provides an exhaustive view of sectors present in CIS, except NACE 19 (Manufacture of leather and leather products), NACE 50 (Sale, maintenance and repair of motor vehicles) and NACE 52 (Retail trade.)

Compared to results of Arundel and Kabla (1998), the manufacture of food, beverages and tobacco is found to have a lower propensity to patent compared to textiles and basic metals. Other sectors show roughly the same pattern, implying that results are consistent and that firm and country level factors do not impact these results much. It is also confirmed that especially manufacturing sectors show high propensities to patent.

Low propensity to patent		**Medium propensity to patent**		**High propensity to patent**	
Sector	**NACE**	**Sector**	**NACE**	**Sector**	**NACE**
Manufacture of food products, beverages and tobacco	15-16	Mining and quarrying of energy producing materials	10-12	Manufacture of coke, refined petroleum products and nuclear fuel Manufacture of chemicals and chemical products (including pharmaceuticals)	23-24
Wood, paper and printing (publishing)	22	Manufacture of textiles and textile products	17-18	Manufacture of rubber and plastic products	25
Electricity, gas, steam and hot water supply Collection, purification and distribution of water	40-41	Manufacture of wood and of products of wood (including straw, plaiting materials, pulp and paper)	20-21	Manufacture of fabricated metal products, except machinery and equipment	28
Supporting and auxiliary transport activities (including cargo handling and travel agencies)	63	Manufacture of other non-metallic mineral products (including glass, clay and ceramics)	26	Manufacture of machinery and equipment	29
Post and telecommunication	64	Manufacture of basic metals (including iron and steel)	27	Manufacture of electrical and optical equipment (including computers, communication equipment and precision instruments)	30-33
Financial intermediation (including banking, insurance, pension funding and security broking)	65-67	Construction	45	Manufacture of motor vehicles, trailers, semi-trailers and other transport equipment (including cars, ships and aircraft)	34-35
		Transport (including land, water and air transport, also including pipelines)	60-62	Manufacture of furniture and recycling	36-37
				Wholesale trade and commission trade	51

Figure 48 – Propensity to patent product innovations, classification of sectors

Differences between countries

Results show a clear difference between countries when looking at the propensity to patent. Especially Germany has a high propensity to patent compared to Norway and Belgium, even when other factors – such as differences in industry composition – are accounted for. Even more surprising, the odds-ratio of Germany is higher than descriptive results suggest (about 6 to 1 instead of 2 to 1.) Part of this may be due to an error in the model, altough the lower bound of the 95% confidence interval suggests that the odds-ratio of Germany compared to Belgium is at least 4 to 1.

Moreover, there seems to be an ordinal ranking between these three countries in terms of their propensities to patent. Since none of the 95% confidence intervals overlaps, this implies that Germany has a higher propensity to patent than Norway, which in turn has a higher propensity to patent than Belgium. One note here is that quality measures for both Norway and Belgium are lower than those for Germany. Results may therefore be influenced by the possible difference in quality of the data provided by each of the countries.

Country	Odds-ratio	Z-value (p-value)	S.E.	95% lower	95% upper
Germany	5.96	**11.11** (0.000)	0.960	4.35	8.18
Norway	2.39	**5.78** (0.000)	0.362	1.78	3.22
Belgium					*- Reference*

Figure 49 – Propensity to patent product innovations, differences between countries

Since country variables persist in through the models, country differences are not (fully) explained by other factors in the model.

Discussion of firm level factors

CIS only measures whether a company applies for at least one patent. It may therefore be that one company has multiple innovations and/or multiple patents applied for. For firms with large innovation spending, this means that the propensity to patent may be over counted by the method employed by CIS. Moreover, from the regression results this seems to be the case, since turnover and R&D intensity both have odds-ratios higher than one as well as very robust z-scores. Although other explanations are possible, these variables are interpreted as control variables for the multiple innovations problem. Turnover and R&D intensity are then complementary indicators for the amount of innovations produced by the firm. With this method it is not possible to disentangle any other size effects on the propensity to patent. To answer more size related questions it would be a suggestion to include the amount of patents applied for as well as the amount of innovations that were introduced in the next CIS.

The relevance of newness to the market and goods innovation is quite straightforward. Both variables indicate that the innovation in question possesses characteristics that are patentable. Descriptive analysis showed that goods are almost 4 times more likely to be patented than services. This is not so surprising given the fact the current patent system places much emphasis on goods as patentable subjects. Besides that, patents require an inventive step which is likely to be correlated with newness to market. New to market goods are therefore generally more patentable, which increases the firms propensity to patent those.

Co-operation with universities was found to be highly relevant. The mechanism by which this variable increases patent propensity is not entirely clear though. One explanation might be that in co-operation with universities, secrecy is a less attractive means of appropriability because of publication of results. To appropriate the returns of the innovation, a patent is then applied for instead. Note that the causality may also run the other way around: firms that don’t need secrecy much may be more willing to co-operate with universities.

Besides this, other explanations may include that firms that co-operate with universities have higher quality innovation processes, which correlate with higher propensities to patent. However, although the exact mechanism is open to debate, firms that co-operate with a university are about two times more likely to apply for a patent than others. For EU funding the same lines of reasoning can be taken, although this variable is only significant at the 0.05 level and not at the 0.001 level. Especially the higher quality innovation process argument may apply here, since for example the EU Framework Programs seem to attract the "elite" of European innovators. (Dekker and Kleinknecht, 2008)

Label	**Variables**	Odds-ratio (z-value)	SE
Turnover	TURN02LOG	1.61 (**11.50**)	0.067
R&D intensity	RRDINXTURN	1.53 (**3.76**)*	0.174
Type=goods	INPDGD	2.98 (**4.79**)	0.692
New to market	NEWMKT	2.15 (**6.16**)	0.268
Cooperation	CO	-	
universities	COUNI	2.55 (**6.12**)	0.389
EU funding	FUNEU	1.97 (*2.80)*	0.480
Local market	MARLOC	0.70 (-*2.49)*	0.099
Other market	MAROTH	1.94 (**5.19**)	0.248
Clients	SLCIBIN	2.07 (*2.08)*	0.728
Commercial labs	SINS	-	
low importance	SINS1	0.68 (-*2.72*)	0.095
medium importance	SINS2	0.50 (-**3.64**)	0.095
Innovation by others	INPDTWONLY	0.05 (-*2.96*)	0.049

Figure 50 –Significant firm level factors in model five

In terms of relevant markets, it was found that firms that sold goods outside of the EU on average had a higher propensity to patent than other firms. Moreover, firms that said to sell goods in a local market on average had a lower propensity to patent than other firms. Combined, it looks like the scope of the firm's market influences its propensity to patent. Firms that sell outside of Europe probably face more fierce competition, increasing the need to secure a competitive position by holding a patent. A second explanation by Levin et al. (1987) is that to gain access to certain foreign markets it is required to license technology to local firms. In order to allow for such licensing agreements, patents are filed. Besides that, just the size of the market may determine if the costs of a patent outweigh its benefits.

Besides differences between countries and sectors, one last group of included variables measured outsourced development. If the innovation was mainly developed by other enterprises or institutions, it is not likely that the firm in question would apply for a patent. The same goes for commercial (R&D) labs as information sources for innovation. Moreover, another interesting observation is that using "suppliers as information sources for innovation" decreases patent propensity, while using clients as information sources for innovation increases patent propensity. Based on this an argument could be made that CIS actually double counts some innovations if there are multiple firms in the value chain. A patented innovation by a supplier might also lead to an innovation at their clients (which might even have been involved in that process.) Both may then report an innovation, but likely only the supplier applies for a patent. This is confirmed by the high importance of the variable that measures whether or not the innovation was developed "mainly by other enterprises or institutions."

When these results are compared to Arundel and Kabla (1998) this study confirmed the positive relationship of sales in markets outside of Europe with the propensity to patent, as well as the higher propensity to patent for German firms. As an addition to the importance of sales in markets outside of Europe, it is also found that sales in local markets decrease a firm's propensity to patent. It thus seems that "market scope" is a relevant variable in determining a firm's propensity to patent. Comparing to Brouwer and Kleinknecht (1999) this study confirms the importance of collaboration on innovative activities on the propensity to patent. More specifically firms that have co-operation arrangements with universities tend to patent relatively more innovations.

Principally new finding include the importance of goods versus services, newness to market, EU funding and clients / commercial labs as information sources for innovation. The last is, together with innovation by others, probably an indicator of outsourcing of R&D activities. All factors mentioned above are robust between different versions of the regression model, except for commercial labs as information source for innovation, which is not significant in model six.

Limitations and further research

As has been discussed, one major limitation of CIS4 was the absence of a measure for numbers of innovations and patents, causing an over-counting of large firms' propensity to patent. Another limitation was found to be the broad classification of firms into sectors. For example, it was not possible to distinguish pharmaceuticals from other manufacturers of chemicals. The same goes for several other sectors, while there is enough reason to assume that large differences exist within these (especially also in terms of their appropriability regimes.) A recommendation would therefore be to use narrower sector classifications in newer versions of CIS, to make a more detailed analysis possible.

Other open questions include the extension of the model to other countries in Europe to generalize results and to further look into differences at the country level. These country level factors were found to be very substantial when looking at Germany, Norway and Belgium. Also it would be very interesting to find out which exact mechanism drives the increased propensity to patent of firms that collaborate with universities and/or receive EU funding.

REFERENCES

ACS, Z. J. & AUDRETSCH, D. B. 1988. Innovation in Large and Small Firms: An Empirical Analysis. *The American Economic Review,* 78**,** 678-690.

ACS, Z. J. & AUDRETSCH, D. B. 1989. Patents as a Measure of Innovative Activity. *Kyklos,* 42**,** 171.

ARUNDEL, A. & KABLA, I. 1998. What percentage of innovations are patented? Empirical estimates for European firms. *Research Policy,* 27**,** 127-141.

BOUND, J., CUMMINS, C., GRILICHES, Z. & HALL, B. H. 1984. Who Does R&D and Who Patents? *In:* GRILICHES, Z. (ed.) *R&D, patents and productivity.* Chicago: University of Chicago Press.

BROUWER, E. & KLEINKNECHT, A. 1999. Innovative output, and a firm's propensity to patent.: An exploration of CIS micro data. *Research Policy,* 28**,** 615-624.

COHEN, W. M., GOTO, A., NAGATA, A., NELSON, R. R. & WALSH, J. P. 2002. R&D spillovers, patents and the incentives to innovate in Japan and the United States. *Research Policy,* 31**,** 1349-1367.

COHEN, W. M. & KLEPPER, S. 1996. A Reprise of Size and R & D. *The Economic Journal,* 106**,** 925-951.

DE RASSENFOSSE, G. & VAN POTTELSBERGHE DE LA POTTERIE, B. 2009. A policy insight into the R&D-patent relationship. *Research Policy,* 38**,** 779-792.

DE VRIES, R. M. & HUISMAN, M. 2007. Logistische Regressianalyse.

DEKKER, R. & KLEINKNECHT, A. 2008. The EU Framework Programs: Are They Worth Doing? *the Comparative Analysis of Enterprise Data (CAED) conference.* Budapest.

GRILICHES, Z. 1990. PATENT STATISTICS AS ECONOMIC INDICATORS - A SURVEY. *Journal of Economic Literature,* 28**,** 1661-1707.

GRILICHES, Z. 1998. *R&D and Productivity: The Econometric Evidence,* Chicago, The University of Chicago Press.

HAGEDOORN, J. & CLOODT, M. 2003. Measuring innovative performance: Is there an advantage in using multiple indicators? *Research Policy,* 32**,** 1365-1379.

HEIJ, C., DE BOER, P., FRANSES, P. H., KLOEK, T. & VAN DIJK, H. K. 2004. *Econometric Methods with Applications in Business and Economics,* New York, Oxford University Press.

KLEINKNECHT, A., VAN MONTFORT, K. & BROUWER, E. 2002. The Non-Trivial Choice between Innovation Indicators. *Economics of Innovation and New Technology,* 11**,** 109 - 121.

KUZNETS, S. 1962. Inventive activity: Problems of definitions and measurement. *In:* RESEARCH, N. B. O. E. (ed.) *The rate and direction of inventive activity.* Princeton University Press, Princeton, NJ.

LEVIN, R. C., KLEVORICK, A. K., NELSON, R. R., WINTER, S. G., GILBERT, R. & GRILICHES, Z. 1987. Appropriating the Returns from Industrial Research and Development. *Brookings Papers on Economic Activity,* 1987**,** 783-831.

MANSFIELD, E. 1986. Patents and innovation: an empirical study. *Management Science,* 32**,** 173-181.

PAVITT, K. 1985. Patent statistics as indicators of innovative activities: Possibilities and problems. *Scientometrics,* 7**,** 77-99.

RICE, J. A. 2007. *Mathematical Statistics and Data Analysis,* Belmont, Thomson Higher Education.

SCHERER, F. M. 1983. The propensity to patent. *International Journal of Industrial Organization,* 1**,** 107-128.

TEECE, D. J. 1986. Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research Policy,* 15**,** 285-305.

VAN DER PANNE, G. 2007. Issues in measuring innovation. *Scientometrics,* 71**,** 495-507.

VAN DER VELDE, M., JANSEN, P. & ANDERSON, N. 2004. *Guide to Management Research Methods,* Oxford, Blackwell Publishing Ltd.

VAN POTTELSBERGHE DE LA POTTERIE, B. & DE RASSENFOSSE, G. 2008. Policymakers and the R&D-patent relationship. *Intereconomics,* 43**,** 377-380.

OTHER SOURCES

ATS. 2011. *Logistic Regression Diagnostics* [Online]. Available: http://www.ats.ucla.edu/stat/stata/webbooks/logistic/chapter3/statalog3.htm [Accessed].

EC 2006. FP7 - Tomorrow's answers start today. http://ec.europa.eu/research/fp7/pdf/fp7-factsheets_en.pdf

OECD 2004a. The Fourth Community Innovation Survey. http://www.oecd.org/dataoecd/52/35/40140021.pdf

OECD 2004b. The Fourth Community Innovation Survey (CIS 4) Methodological recommendations. http://www.oecd.org/dataoecd/53/11/40140079.pdf

PATENTLENS 2003. Reason for lower numbers of granted patents / number of patent applications. http://www.patentlens.net/daisy/bios/2666.html

APPENDIX I – SECTOR CLASSIFICATION

Sector	#	NACE	NACE
Mining	1	C	10
			11
			12
Mining of metal ores		CB	13
Other mining and quarrying			14
Food, Beverages and tabacco	2	DA	15
			16
Textiles and clothing	3	DB	17
			18
Manufacture of leather and leather products	4	DC	19
Wood, paper and printing	5	DD	20
		DE	21
Wood, paper and printing (publishing)	6		22
Petroleum Refining	7	DF	23
Chemicals and pharmaceuticals		DG	24
Rubber and plastic products	8	DH	25
Glass, clay and ceramics	9	DI	26
Basic metals (iron and steel)	10	DJ	27
Fabricated metal products	11		28
Mechanical engineering	12	DK	29
Computers	13	DL	30
Electrical machinery			31
Communication equipment			32
Precision instruments			33
Automobiles	14	DM	34
Other transport equipment			35
Manufacture of furniture	15	DN	36
Recycling			37
Electricity, gas, steam and hot water supply	16	E	40
Collection, purification and distribution of water			41
Construction	17	F	45
Sale, maintenance and repair of motor vehicles and motorcycles	18	G	50
Wholesale trade and commission trade (except motor vehicles)	19		51
Retail trade	20		52
Hotels and restaurants		H	55
Land, water and air transport	21	I	60
			61
			62
Auxiliary transport activities; travel agencies	22		63
Post and telecommunication	23		64

Financial intermediation	24	J	65
			66
			67
Real estate activities		K	70
Renting of machinery and equipment			71
Computer and related activities	25		72
Research and development	26		73
Other business activities			74

APPENDIX II – STATA MODEL OUTPUT

Logistic regression

Number of obs = 2361
LR chi2(17) = 931.01
Prob > chi2 = 0.0000
Log likelihood = -1076.6622
Pseudo R2 = 0.3019

propat	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
turn02log	1.334027	.0401782	9.57	0.000	1.257559	1.415146
funeu	1.567619	.320225	2.20	0.028	1.050417	2.339481
co	1.756236	.2114246	4.68	0.000	1.38711	2.223591
inpdgd	4.864764	.7811413	9.85	0.000	3.551266	6.664081
newmkt	2.138714	.2325276	6.99	0.000	1.728254	2.646659
sclibin	2.387136	.7509598	2.77	0.006	1.288549	4.422352
sins1	.6260222	.0797016	-3.68	0.000	.4877751	.8034519
sins2	.5033307	.0850661	-4.06	0.000	.3614056	.7009902
funloc	1.425195	.2402426	2.10	0.036	1.024209	1.983169
suni1	1.352717	.1864788	2.19	0.028	1.032439	1.772349
suni2	1.681051	.2734762	3.19	0.001	1.222098	2.312362
suni3	2.797037	.6628234	4.34	0.000	1.757868	4.450516
countryde	1.756408	.2436832	4.06	0.000	1.338228	2.305265
marloc	.6589605	.0806513	-3.41	0.001	.5184172	.8376053
inpdtwonly	.0402827	.041246	-3.14	0.002	.0054145	.2996917
mareur	1.497052	.1978868	3.05	0.002	1.155371	1.93978
maroth	2.143316	.2639544	6.19	0.000	1.683677	2.728436

Table a – first model, without sector dummies

Logistic regression

Number of obs = 2355
LR chi2(28) = 1119.34
Prob > chi2 = 0.0000
Log likelihood = -979.81772
Pseudo R2 = 0.3635

propat	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
turn02log	1.529589	.0565924	11.49	0.000	1.422596	1.644629
rrdinxturn	10.30713	6.332703	3.80	0.000	3.091428	34.36499
co	1.865493	.2358042	4.93	0.000	1.456126	2.389946
inpdgd	2.848086	.5606242	5.32	0.000	1.936424	4.188955
newmkt	2.008063	.2308788	6.06	0.000	1.602912	2.51562
sclibin	2.052206	.6827973	2.16	0.031	1.06909	3.939377
sins1	.6920413	.0887306	-2.87	0.004	.5382631	.8897529
sins2	.5213746	.0917841	-3.70	0.000	.3692344	.736203
sector23	.1559967	.1106857	-2.62	0.009	.0388289	.6267236
sector21	.2687268	.1425167	-2.48	0.013	.0950352	.7598668
suni2	1.381235	.1997929	2.23	0.026	1.040262	1.833969
suni3	2.041095	.4702587	3.10	0.002	1.299422	3.206092
countryde	1.899319	.2769613	4.40	0.000	1.427166	2.527675
marloc	.7376943	.0953159	-2.35	0.019	.5726569	.9502949
sector10	.2895686	.1239291	-2.90	0.004	.1251571	.6699577
sector19	.4488485	.1437418	-2.50	0.012	.2396106	.8408013
maroth	1.852593	.219083	5.21	0.000	1.469329	2.335829
sector26	.5903232	.1344235	-2.31	0.021	.3777995	.9223981
sector24	.0184312	.0140761	-5.23	0.000	.0041256	.0823422
funeu	1.723728	.3585939	2.62	0.009	1.146542	2.591478
sector22	.0737761	.0646181	-2.98	0.003	.0132549	.4106346
inpdtwonly	.0364173	.0373105	-3.23	0.001	.0048891	.2712608
sector16	.069429	.053967	-3.43	0.001	.0151323	.3185499
sector2	.1060686	.0343419	-6.93	0.000	.0562331	.2000696
sector3	.1838713	.0728134	-4.28	0.000	.0846129	.3995686
sector5	.4224815	.1215929	-2.99	0.003	.2403402	.7426581
sector6	.1313893	.0615522	-4.33	0.000	.0524561	.3290971
sector25	.1494097	.0413343	-6.87	0.000	.0868747	.2569593

Table b – second model, including sector dummies

Logistic regression		Number of obs	=	2356
		LR chi2(28)	=	1133.36
		Prob > chi2	=	0.0000
Log likelihood = -973.25703		Pseudo R2	=	0.3680

propat	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
turn02log	1.511985	.056656	11.03	0.000	1.404921	1.627208
rrdinxturn	7.321012	4.638639	3.14	0.002	2.114679	25.34532
sector24	.0210926	.0160327	-5.08	0.000	.0047547	.0935703
cointbin	1.994337	.4850478	2.84	0.005	1.238157	3.21234
sector23	.168084	.125147	-2.40	0.017	.0390629	.7232502
sector25	.1554669	.0435623	-6.64	0.000	.0897694	.2692448
sector2	.1027075	.0338329	-6.91	0.000	.0538527	.1958831
sector22	.0742731	.0668804	-2.89	0.004	.0127161	.4338199
sector3	.1837894	.072003	-4.32	0.000	.0852798	.3960909
couni	2.129935	.3196878	5.04	0.000	1.58711	2.858417
sector26	.5845167	.1341578	-2.34	0.019	.3727605	.9165665
inpdgd	2.905855	.5818805	5.33	0.000	1.962573	4.302512
newmkt	2.050513	.2364292	6.23	0.000	1.63575	2.570444
sclibin	1.962755	.6627819	2.00	0.046	1.012589	3.804515
sins1	.738324	.0945886	-2.37	0.018	.5743774	.9490666
sins2	.5846346	.100804	-3.11	0.002	.4169832	.8196917
sector5	.4221283	.1208228	-3.01	0.003	.2408857	.7397379
inpdtwonly	.0369024	.0377503	-3.23	0.001	.0049693	.2740397
sector10	.2829136	.1211717	-2.95	0.003	.1222038	.6549721
sector16	.0577237	.0484658	-3.40	0.001	.0111345	.2992526
countryde	2.888681	.5260086	5.83	0.000	2.021631	4.127599
marloc	.7449797	.0966977	-2.27	0.023	.5776426	.9607926
sector6	.1291682	.0607479	-4.35	0.000	.0513848	.3246958
sector21	.2792267	.1472004	-2.42	0.016	.0993637	.7846685
maroth	1.864197	.2207791	5.26	0.000	1.478029	2.35126
sector19	.4748476	.1554368	-2.28	0.023	.2499893	.9019594
fungmt	1.376441	.2083498	2.11	0.035	1.023087	1.851836
funeu	1.635928	.3480734	2.31	0.021	1.078092	2.482406

Table c – third model, including sector dummies and detailed cooperation variables

Logistic regression		Number of obs	=	2075
		LR chi2(16)	=	1013.72
		Prob > chi2	=	0.0000
Log likelihood = -841.92682		Pseudo R2	=	0.3758

propat	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
turn02log	1.546798	.0610562	11.05	0.000	1.431642	1.671217
rrdinxturn	13594.6	36561.37	3.54	0.000	69.84608	2646006
inpdtwonly	.0456136	.046709	-3.02	0.003	.0061299	.33942
cointbin	1.831193	.4902846	2.26	0.024	1.08351	3.094819
maroth	1.994478	.2520232	5.46	0.000	1.556937	2.55498
lowsector3	.038474	.0187799	-6.67	0.000	.0147802	.1001508
countryde	2.184305	.3885493	4.39	0.000	1.541344	3.095473
funeu	1.892157	.4561447	2.65	0.008	1.17966	3.034991
marloc	.7202515	.0997122	-2.37	0.018	.5490894	.9447682
couni	2.206584	.3588883	4.87	0.000	1.60427	3.035034
lowsector2	.2605077	.0363264	-9.65	0.000	.1982097	.3423861
inpdgd	2.589575	.4987155	4.94	0.000	1.775412	3.777094
newmkt	1.996982	.2452491	5.63	0.000	1.569778	2.540447
sclibin	2.42508	.8599015	2.50	0.012	1.210323	4.859045
sins1	.6777591	.0931941	-2.83	0.005	.5176453	.8873979
sins2	.4902488	.0923462	-3.78	0.000	.338906	.7091758

Table d – fourth model, with two sector dummies and excluding observations with R&D intensities > 10%

```
Logistic regression                               Number of obs   =       1913
                                                  LR chi2(15)     =     957.94
                                                  Prob > chi2     =     0.0000
Log likelihood = -811.59775                       Pseudo R2       =     0.3711
```

propat	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
turn02log	1.614484	.0672677	11.50	0.000	1.487882	1.75186
rrdinxturn2	1.533417	.1742092	3.76	0.000	1.227316	1.915861
maroth	1.94096	.2478394	5.19	0.000	1.511219	2.492905
inpdtwonly	.0476287	.049061	-2.96	0.003	.0063251	.3586476
sins2	.5036445	.0949849	-3.64	0.000	.34801	.7288809
funeu	1.974741	.480198	2.80	0.005	1.226093	3.180511
lowsector2	.226563	.0367079	-9.16	0.000	.1649213	.3112442
marloc	.7048099	.098887	-2.49	0.013	.5353599	.9278937
sins1	.6845813	.0953427	-2.72	0.007	.521047	.8994421
lowsector3	.0359197	.0178004	-6.71	0.000	.0135991	.0948756
couni	2.547066	.3891219	6.12	0.000	1.887992	3.436216
countryde	1.91431	.2971694	4.18	0.000	1.412132	2.595071
inpdgd	2.977404	.6916077	4.70	0.000	1.888496	4.694178
newmkt	2.1534	.2682213	6.16	0.000	1.686947	2.74883
sclibin	2.076963	.7283893	2.08	0.037	1.044511	4.129947

Table e – model 5, based on model 4 but excluding sectors 25 and 26

```
Logistic regression                               Number of obs   =       3037
                                                  LR chi2(17)     =    1298.99
                                                  Prob > chi2     =     0.0000
Log likelihood =  -1299.591                       Pseudo R2       =     0.3332
```

propat	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
turn02log	1.459747	.0470684	11.73	0.000	1.370349	1.554977
rrdinxturn2	1.479132	.1332527	4.35	0.000	1.239719	1.764779
inpdtwonly	.0910781	.0556862	-3.92	0.000	.0274776	.3018898
countryde	5.965301	.9592879	11.11	0.000	4.352608	8.175517
cointbin	1.6064	.2515657	3.03	0.002	1.181827	2.183503
lowsector3	.0495956	.020688	-7.20	0.000	.0218967	.1123333
funeu	1.798557	.3445795	3.06	0.002	1.235508	2.618202
suni3	1.644488	.3240974	2.52	0.012	1.11757	2.419838
marloc	.7462794	.096481	-2.26	0.024	.5792364	.961495
countryno	2.396739	.3624842	5.78	0.000	1.781905	3.223718
couni	2.005015	.2569851	5.43	0.000	1.559618	2.577609
lowsector2	.3442455	.0412546	-8.90	0.000	.2721823	.4353884
inpdgd	2.670165	.5156751	5.09	0.000	1.828732	3.898758
newmkt	2.065413	.209232	7.16	0.000	1.693472	2.519045
sclibin	2.892966	.7858084	3.91	0.000	1.698767	4.926663
ssupbin	.6004858	.1174212	-2.61	0.009	.4093131	.8809471
maroth	1.783694	.1840586	5.61	0.000	1.457086	2.183512

Table f – model 6, based on model 5 but including missing observations

```
Logistic regression                               Number of obs   =       3017
                                                  LR chi2(29)     =    1344.81
                                                  Prob > chi2     =     0.0000
Log likelihood = -1268.3006                       Pseudo R2       =     0.3465
```

propat	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
turn02log	1.501288	.0507804	12.01	0.000	1.404987	1.604189
rrdinxturn2	1.505498	.1424903	4.32	0.000	1.250596	1.812356
cointbin	1.556696	.2474974	2.78	0.005	1.139914	2.125864
couni	2.054161	.2670708	5.54	0.000	1.592084	2.650349
inpdgd	2.665348	.5277415	4.95	0.000	1.80807	3.929096
newmkt	2.132785	.2197702	7.35	0.000	1.742754	2.610106
sclibin	2.884285	.7946087	3.84	0.000	1.680876	4.949265
suni3	1.654954	.331539	2.51	0.012	1.117541	2.450802
ssupbin	.6496126	.1289397	-2.17	0.030	.4402526	.9585326
countryde	5.904997	.9807317	10.69	0.000	4.264301	8.176953
countryno	2.426892	.374646	5.74	0.000	1.79328	3.284375
marloc	.758603	.1003121	-2.09	0.037	.585407	.98304
maroth	1.687636	.1797243	4.91	0.000	1.369716	2.079349
funeu	1.861331	.364494	3.17	0.002	1.268054	2.732182
inpdtwonly	.0757142	.0464975	-4.20	0.000	.0227216	.2522993
sector16	.1929445	.1368844	-2.32	0.020	.0480339	.7750272
sector2	.3211369	.0837178	-4.36	0.000	.1926588	.5352929
sector24	.0563768	.0356818	-4.54	0.000	.0163065	.1949127
sector19	1.769174	.4106665	2.46	0.014	1.122502	2.788395
sector6	.3597913	.1498487	-2.45	0.014	.1590507	.8138903
sector7	1.79879	.3714353	2.84	0.004	1.200092	2.696166
sector8	2.321544	.5578452	3.51	0.000	1.449576	3.718028
sector23	.1936368	.1391179	-2.29	0.022	.0473635	.791647
sector22	.1548644	.1294023	-2.23	0.026	.0301093	.7965302
sector11	2.68431	.5451593	4.86	0.000	1.802857	3.996723
sector12	3.055668	.5624409	6.07	0.000	2.130247	4.38311
sector13	2.058728	.362375	4.10	0.000	1.458047	2.906876
sector14	2.231371	.5158772	3.47	0.001	1.418339	3.510456
sector15	2.144671	.5418741	3.02	0.003	1.30706	3.519053

Table g – model 6, including significant sector variables

APPENDIX III – COUNTRY MODELS

Logistic regression		Model Germany		Model Norway		Model Belgium	
		R&D intensity < 50% Excluding sector 25, 26 Germany only		R&D intensity < 50% Excluding sector 25, 26 Norway only		R&D intensity < 50% Excluding sector 25, 26 Belgium only	
	Conditions	Inpdt=1 Rrdinxturn<0.5 Backwards (0.05)		Inpdt=1 Rrdinxturn<0.5 Backwards (0.05)		Inpdt=1 Rrdinxturn<0.5 Backwards (0.05)	
Label	**Variables**	Odds-ratio	SE	Odds-ratio	SE	Odds-ratio	SE
Turnover	TURN02LOG	1.62 (**10.42**)	0.075	1.28 (**3.81**)	0.083	1.32 (**3.98**)	0.092
R&D intensity	RRDINXTURN	1.77 (**3.82**)*	0.265	n.s.		2.23 (*3.38*)*	0.529
Type=goods	INPDGD	3.09 (**3.59**)	0.969	2.13 (*2.84*)	0.570	4.01 (*2.22*)	2.501
New to market	NEWMKT	2.12 (**5.20**)	0.306	1.97 (**3.72**)	0.209	2.00 (*2.77*)	0.498
Cooperation	CO	-		-		-	
national	CONATBIN	n.s.		n.s.		0.54 (*-2.01*)	0.166
international	COINTBIN	n.s.		2.09 (*3.27*)	0.475	n.s.	
enterprise group	COOE	n.s.		n.s.		1.95 (*2.53*)	0.516
universities	COUNI	2.09 (**3.80**)	0.407	1.58 (*2.02*)	0.357	3.21 (**3.81**)	0.979
government	COGOV	n.s.		n.s.		2.70 (*3.46*)	0.775
Local funding	FUNLOC	n.s.		n.s.		n.s.	
EU funding	FUNEU	1.85 (*2.43*)	0.469	n.s.		n.s.	
Local market	MARLOC	0.69 (*-2.40*)	0.107	n.s.		n.s.	
EU market	MAREUR	n.s.		n.s.		n.s.	
Other market	MAROTH	1.82 (**3.98**)	0.273	1.96 (**3.71**)	0.354	n.s.	
Clients	SLCIBIN	3.04 (*2.39*)	1.417	2.94 (*2.74*)	1.165	n.s.	
Suppliers	SSUPBIN	n.s.		n.s.		n.s.	
Commercial labs	SINS	-		-		-	
low importance	SINS1	n.s.		0.58 (*-2.46*)	0.128	n.s.	
medium importance	SINS2	n.s.		0.61 (*-2.04*)	0.148	n.s.	
high importance	SINS3	n.s.		n.s.		n.s.	
Universities	SUNI	-		-		-	
low importance	SUNI1	n.s.		n.s.		n.s.	
medium importance	SUNI2	n.s.		n.s.		n.s.	
high importance	SUNI3	n.s.		1.64 (*2.52*)	0.324	n.s.	
Innovation by others	INPDTWONLY	0.05 (*-2.90*)	0.052	n.s.		n.s.	
Germany	COUNTRYDE	-		-		-	
Norway	COUNTRYNO	-		-		-	
Sectors odd <1	LOWSECTOR2	0.20 (**-8.22**)	0.039	0.44 (**-4.05**)	0.089	0.62 (*-2.04*)	0.144
Sectors odd <0.1	LOWSECTOR3	0.02 (**-6.67**)	0.013	n.s.		0.08 (*-2.37*)	0.084
Pseudo R-squared			**0.41**		**0.17**		**0.29**
Model significance			**0.0000**		**0.0000**		**0.0000**
Number of observations			**1465**		**783**		**774**

APPENDIX IV – CIS4 QUESTIONNAIRE