2015•2016
FACULTEIT BEDRIJFSECONOMISCHE WETENSCHAPPEN
master in de toegepaste economische wetenschappen: handelsingenieur

Masterproef

Innovation and firm performance: An econometric analysis on the Belgian pharmaceutical sector

Promotor :
Prof. dr. Mark VANCAUTEREN

Copromotor :
Prof. dr. Wim VANHAVERBEKE

Sander Vandevenne

Scriptie ingediend tot het behalen van de graad van master in de toegepaste economische wetenschappen: handelsingenieur

universiteit hasselt
KNOWLEDGE IN ACTION

Universiteit Hasselt | Campus Hasselt | Martelarenlaan 42 | BE-3500 Hasselt
Universiteit Hasselt | Campus Diepenbeek | Agoralaan Gebouw D | BE-3590 Diepenbeek

universiteit hasselt
KNOWLEDGE IN ACTION

2015•2016
FACULTEIT BEDRIJFSECONOMISCHE WETENSCHAPPEN
master in de toegepaste economische wetenschappen: handelsingenieur

Masterproef

Innovation and firm performance: An econometric analysis on the Belgian pharmaceutical sector

Promotor :
Prof. dr. Mark VANCAUTEREN

Copromotor :
Prof. dr. Wim VANHAVERBEKE

Sander Vandevenne

Scriptie ingediend tot het behalen van de graad van master in de toegepaste economische wetenschappen: handelsingenieur

universiteit hasselt
KNOWLEDGE IN ACTION

Preface

The writing and presentation of this master thesis is the final milestone of my career at Hasselt University. It allows me to bundle the competences that I have learned in the fields economics, econometrics and the dynamics of innovation over the past five years, and for which I am grateful. It was with joy that I set out to complete a study involving patents in the pharmaceutical industry. I have always been intrigued by the pros and cons of intellectual property rights and their implications for society. Furthermore, it is econometrics that is my passion. To combine both was therefore a challenge happily accepted.

In order to accomplish this project successfully, I was lucky enough to be able to rely on several other people for valuable feedback and input. As such, I would like to take the opportunity to thank them. First of all, I would like to thank my promotor, Mark Vancauteren, and my co-promotor, Wim Vanhaverbeke, to provide me with the opportunity of tackling this subject. Moreover, professor Vancauteren gave useful directions, feedback and aided to resolve the encountered problems.

Other than both professors, I would like to express special gratitude to my parents and grandparents. First of all, my parents have always supported me financially and especially my father invested a lot of time guiding me throughout my academic career. Likewise, my grandparents have always been ready to take me in at times exams had to be prepared. It is to their merit that I am on the brink of graduating.

Furthermore, I would like to thank Shreosi Sanyal. Although I was never personally in contact with dr. Sanyal, her doctoral dissertation was a good guide and academic source of inspiration. Correspondingly, I want to acknowledge the efforts of several people from my personal space. The contributions of Stephanie Jans, Vincent Compagnie and Dieter Tomsin were fruitful. Throughout my progress they aided by proofreading and giving feedback.

To conclude, my experience at Hasselt university has formed me in many ways: I developed intellectually and grew as a person. I believe that the past five years have built a good foundation for the future. Moreover, I got inspired by financial risk management in the last two years of my education. As follows, I took the decision to deepen my understanding on that subject in an additional year of education. I look forward to starting a second master on the subject 'Banking and Finance' at Gent University upcoming academic year.

Sander Vandevenne,

August 22th, 2016

Summary

The past century has shown macro-economic growth is the most important driver of increasing living standards. Establishing a framework that facilitates firms in their development and helps them exploiting their innovative capabilities is therefore very relevant. Pharmaceuticals in that sense is an industry of special interest. Alongside the semi-conductor industry it is one of the large industries that dedicates a very high percentage of its funds to research efforts. Firms rely heavily on bringing new products to the market in order to survive and make a profit.

However, accruing the forthcomings of medicines is not self-evident. Knowledge, new innovations and chemicals compounds in particular are non-excludable and non-rivalrous in nature. Hence, an inventor cannot prevent others from copying his ideas and using them to their own benefit. It justifies cause of action by the government. As a consequence, the latest decades have seen a tremendous development of intellectual property rights. Those allow inventors to make a living by making inventions temporarily excludable. In this master thesis we focus on patents. In contrast to other industries, pharmaceutical firms turn out to value them a lot more as a means of appropriation. In addition, the literature points out patents are a good fit to innovation in product market environments.

The conceptual framework that was adopted is inspired by Crépon, Duguet & Mairesse (1998) that disentangle innovation and performance applying three structural equations. We test whether firms with higher innovative efforts have more patents and study whether this means that they are more (labor) productive. We are interested in the transformation of resources in the knowledge generation process. In that respect we formulate determinants of R&D (section 8.2). Subsequently, the fitted R&D intensities are imputed in the innovation (section 8.3) and productivity estimation (section 8.4).

In addition to testing the conceptual framework, an attempt is made to address the contemporary topic of open innovation (OI). We look into extramural collaboration using information about the inventor(s) concealed in patents. An indicator of OI is constructed using inventor nationalities, assuming firms are more engaged in OI activities if more people with different nationalities are included in the innovative process (SHII-indicator: Guellec & van Pottelsberghe de la potterie, 2001; Nagaoka, Motohashi & Goto, 2010). This indicator is included in all regressions as an explanatory variable in order to explore potential entries of open innovation.

The dataset used, was compiled from three different sources. It combines R&D information extracted from Belgian annual reports (Kruispuntbank van Ondernemingen) with patent data from the European patent office and financial information obtained from Bureau van Dijk. A panel dataset was constructed for the period 2006 to 2012. It allows us to track entities at different time intervals and use methods of estimation that can exploit this feature.

To investigate the determinants of R&D (section 8.2), we used a sample selection model to cope with differences between observed, reported R&D and true R&D intensities. We find size and competition influence the R&D intensity negatively. In other words, less innovate efforts are

undertaken if competition intensifies and R&D expenditures do not rise proportionally to size. R&D expenditures are found to increase between 38% and 73% if the corporate size doubles. On the other hand open innovation and fewer budgetary constraints are found to affect the R&D intensity positively. In fact, equation one is where open innovation seems to enter the series of structural equations (section 8.5). Whether this means open innovation causes the R&D intensity to be boosted because of expanded technological opportunities, or, whether rising R&D expenditures as an important cost caused firms to look beyond the borders of the own organization could not be verified. It is probably not the one or the other explanation, but a combination of both. Additionally, we found that including information on tax regimes and the financial health of a firm have potential to augment the quality of the results for similar studies in the future.

In the second part (section 8.3) of the structural equations model the fitted R&D intensities were used as an explanatory variable for the number of patent application counts. It translates the concepts of innovative input and output into practice. Both count models and GLS estimation were fitted to the data. Under these specifications a positive impact could univocally be confirmed. Still, results suggest pronounced decreasing returns to scale (10-22.5%) are present. Those are in line with soaring R&D costs and the difficulties of sustaining the product line encountered in modern-day pharmaceuticals. Other than R&D, firm size and previous patenting activities entered the regressions with a positive beta. Larger firms are known to rely more heavily on the patenting system although frequently not because of desirable economic motives, but rather strategically. In addition maturity, as measured by firm age was associated negatively with patent counts. It suggests more mature firms make more use of alternative routes to appropriate returns, or that they are less innovative.

In section 8.4 the link between innovative output and firm performance was investigated. Firm performance which is a multidimensional construct was measured in monetary terms. It was represented by labor productivity – the log value added per employee. Here, a positive relationship between innovation and firm performance is suggested by the results but not unambiguously supported. Using fixed effects regression, patent counts do not seem to have an effect on labor productivity. Otherwise, the elasticity is found to lie between 4% and 7%. We suspect another operationalization of the variables and the usage of citation weighted patents could build a stronger case considering the findings of other studies in the literature. Further, a direct link of innovative efforts to labor productivity did not seems present. Fitted R&D intensity has no explanatory power. Profit margin did have a positive impact on labor productivity suggesting market power increases financial results. Moreover, labor productivity was also clearly autoregressive.

To conclude we revisit the conceptual framework. In general, the structural equations were found to be adequately representing the way innovation is translated in financial prosperity. Combining equations two (H2) and three (H1) results in concluding that firms making a higher R&D effort will perform better, ceteris paribus. Nevertheless, the interlinkage is rather weak and the relationship between innovative success and labor productivity was not always established. The effect of competition is ambiguous, but overall the results hint less competition might harbor more innovation. Finally, open innovation did not seem to influence innovative output or firm performance directly. It enters as a determinant of R&D intensity.

Table of Contents

List of Figures

List of Tables

1 Introduction

"A competitive economy, in essence, is one in which institutional and macroeconomic conditions allow productive firms to thrive. In turn, the development of these firms supports the expansion of employment, investment and trade" (Mario Draghi, 2012). Furthermore, economic growth is essential to increasing living standards and reducing poverty. The government, therefore, has a responsibility of establishing a framework that facilitates firms in their development and helps them exploiting their capabilities.

Pharmaceuticals in that sense is an industry of special interest. It relies heavily on research and development (R&D) and depends on bringing new products to the market to survive and make a profit. However, accruing the forthcomings of medicines is not self-evident. Knowledge, new innovations and chemicals compounds in particular are non-excludable and non-rivalrous in nature. An inventor cannot prevent others from copying his ideas and using them to their own benefit. Hence, the government and its institutions have established intellectual property rights (IPR) to meet these concerns. They give inventors protective rights to their creations of mind and are thereby creating incentives to innovate (WIPO, 2015).

In this master thesis the relationship between innovative success and productivity in the research-based Belgian (bio-)pharmaceutical sector is investigated. Moreover, the focal point are patents on a firm level. Patents are an intellectual property right giving the owner a temporary monopoly for an innovation. We test whether firms with higher innovative efforts have more patents and study whether this means that they are more (labor) productive. We are also interested in the transformation of resources in the knowledge generation process, we formulate determinants of R&D (section 8.2) and subsequently impute fitted R&D in the innovation (section 8.3) and productivity estimation (section 8.4). This master thesis is based on a framework established by Crépon-Duguet-Mairesse (see Crépon, Duguet & Mairesse, 1998; Vancauteren, Melenberg & Plasmans, 2014).

In addition to testing the conceptual framework, an attempt is made to address the contemporary topic of open innovation (OI). We look into extramural collaboration using information about the inventor(s) concealed in patents. An indicator of OI is constructed using inventor nationalities, assuming firms are more engaged in OI if more people with more different nationalities are included in the innovative process (SHII-indicator: Guellec & van Pottelsberghe de la potterie, 2001; Nagaoka, Motohashi & Goto, 2010). This indicator is included in all regression as an explanatory variable the explore potential entries of open innovation. We continue with a recap of the contributions of pharma to healthcare and the quality of life, the economy and growth. We situate its importance in Belgium and reflect upon the position of the industry in Belgium compared to the world. Thereafter, the problem statement comes to the point and the most important topics of the literature are discussed. Next, the conceptual model is built and empirically tested. Finally, conclusions are draw and the most important limitations framed

2 Context

2.1 The Pharmaceutical Industry

Pharmaceuticals have been improving the quality of life ever since the last century. Nowadays humanity is already able to tackle cardiovascular diseases, many cancers, AIDS and multiple sclerosis. In Europe people can expect to live, on average, 30 years longer than one hundred years ago. This rapid progress was only possible through heavy research and development investments (R&D). Correspondingly the industry is one of the top performing high-tech sectors and an important source of macroeconomic growth (EFPIA, 2014).

The continuous flow of product innovations that originate from the R&D investments are a trademark characteristic of the pharmaceutical industry. For non-generic firms, innovation defines competitiveness. They face a constant urge to identify new molecular entities (NMEs) which is progressively becoming more difficult (Charles River Associates, 2004; CRA onwards). The major challenge for a pharmaceutical firm is therefore consolidating, strengthening and enlarging their knowledge base. Accordingly, they foster the ability to provide the market with new and improved medicines (Petrova, 2014).

The majority of the world's pharmaceutical sales, production and innovation activities are located

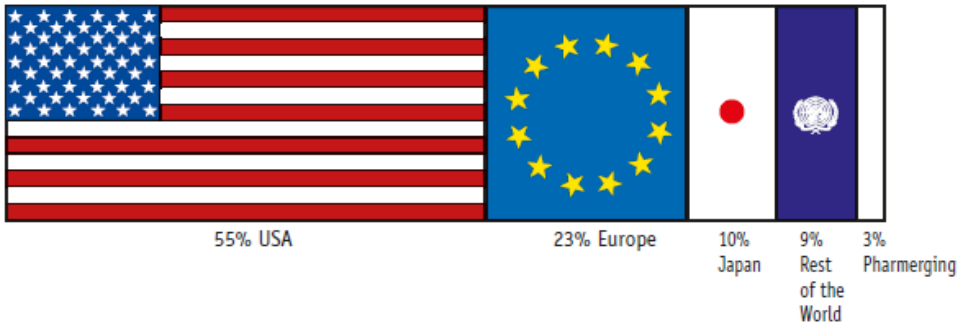


Figure 1: Geographical Breakdown (by main markets) of sales of new medicines launched during the period 2009 – 2013. Source: (EFPIA, 2014)

within the most developed traditional western countries (Figure 1). The USA adopts a dominant role with Europe and Japan also contributing significantly (Sanyal, 2014, p23). Furthermore, the Brazilian, Chinese and Indian sales and research environments are growing rapidly (EFPIA, 2014).

The worldwide pharmaceutical industry is the industry with the highest R&D intensity. The sector also tops the value-added-per-employee ranking (EFPIA, 2014). Although, preluded by sales market dominance of US firms, the R&D center of gravity has gradually shifted towards the USA in the past twenty years (Figure 2; EFPIA, 2014). In spite of Europe's best efforts most of the R&D activity occurs in the USA nowadays. This is accompanied by an ongoing trend to move research laboratories towards to USA (CRA, 2004, p7) and imposes the threat of a brain drain in Europe. Nevertheless, gross R&D expenditures continue to grow in Europe.

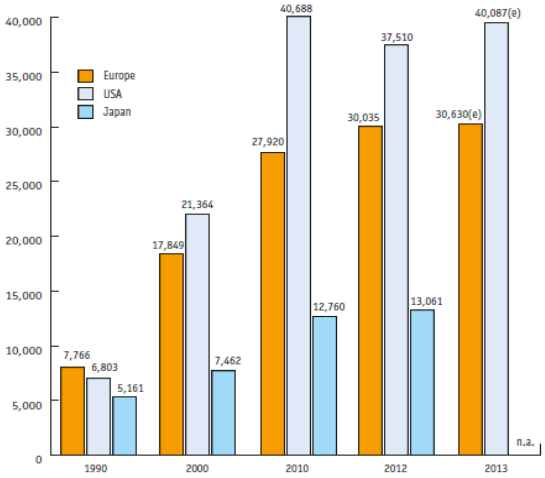


Figure 2: Pharmaceutical R&D expenditure in Europe, USA and Japan (Millions of national currency units *), 1990-2013 (EFPIA, 2014)

In order to sustain the product pipeline, the industry relies on its employees and collaborative networks. Firms require a highly skilled workforce to provide the necessary underpinning of their knowledge base. As for 2012, EFPIA (2014) estimates that pharmaceutical firms had about 700 000 people on their payroll in Europe, of which approximately fifteen percent directly related to R&D. Furthermore, the production and management in pharma also requires highly trained employees. In total 60 percent of the workforce has enjoyed higher education (EFPIA, 2014). Moreover pharma.be (2016) states the sector creates a manifold of jobs in related sectors throughout its supply chain.

Summarizing, the overwhelming presence of pharmaceuticals in developed countries for R&D, manufacturing and sales is a logical conclusion: the markets are the largest and the resources available. Innovation holds a cornerstone position in a pharmaceutical firm's competitiveness. It determines whether a firm will perform in the long run. Understanding the dynamics between innovation and firm performance is hence mandatory (to be unfolded within the literature review).

2.2 The Pharmaceutical Industry in Belgium

Belgium is a highly developed economy with strong institutions and a transportation gateway to Europe. It has a knowledge-intensive industry and qualitative educational system to support it (Abrahamsen et al., 2011). The presence of highly skilled labor, a profitable sales market and its strategic logistic location make the country an excellent operating ground for pharmaceutical firms. As a consequence, big pharma is omnipresent. Moreover, a unique Belgian ecosystem of venture capital, research facilities, networks, universities and incubators (such as Bioville, Diepenbeek) allow for biopharmaceutical SMEs to thrive (Pharma.be, 2016).

Whereas Belgium clearly has something to offer the pharmaceutical industry returns the favor. It is on the forefront of pharmaceuticals in Europe (and the world) on almost all levels. It beats almost every industry standard. In 2015 it directly employed almost 35000 people (≈ 1% of the workforce) in Belgium, accompanied by 100 000 jobs created in other industries through its activities (Pharma.be, 2016). In addition, more than 60% of the workforce has had higher education. Besides the sector is known to have good remuneration standards (Pharma.be, 2016). The industry is exceptional with respect to R&D. It is reinvesting two to three billion euros in R&D on a yearly basis, which is over half of its turnover (52,4% in 2014). This is more than three times the industry average in Europe (16,5%) and roughly one third of annual R&D investments in Belgium. As a consequence bio-pharmaceuticals attain a value-added per capita of €200 000 (Pharma.be, 2016).

Furthermore production and export volumes are above the standards. The total value of production has been growing seven percent on average since 2000. Meanwhile the share of pharmaceutical exports in total exports for Belgium has climbed from 3,5% in 2000 to fifteen percent in 2014. According to *de Nationale Bank van België* (NBB) this strong upsurge of the export figures can largely be accredited to increased distributive activities which confirms Belgium's role as an open economy. Both evolution also influence the trade balance positively (Pharma.be, 2016).

In July 2015 the Belgian government formally underwrote the importance of the industry by closing on a *Toekomstpact*. It recognizes the elicit position Belgium attains in bio-pharmaceuticals and aims at strengthening this position. It involves a gamma of engagements such as: updating the legal framework, facilitating faster lead times by updating testing procedures, stimulating private-public cooperation relating to R&D, financial R&D support and transportation related issues (Pharma.be, 2016).

Most of these efforts somehow relate to innovation, which is one of the most important drivers of competitiveness. A lot of innovative efforts happen in Belgium, but are shifting away from Europe towards the USA and cheaper alternatives. The question why Belgium is so resilient is interesting. The threat of the industry crumbling is frightening (relocation of R&D centers, erosion of distributive activities, …). Studying the dynamics between innovation and firm performance can therefore prove useful insights.

3 Definitions

Many concepts that are used in this thesis are open to slightly different interpretations. Also the specific context calls narrower formulations. It is therefore useful to delineate the contours of each concept in a way that fits the framework of this thesis. The terms biopharmaceuticals, innovation and firm performance are elaborated.

3.1 (Bio-)pharmaceuticals

Based on a definition formulated by Charles River Associates (2004, p4) in a study undertaken for the European commission we define biopharmaceuticals as *the industry concerning the development of medicinal products for human or animal consumption*. This particularly includes new chemical entities (NMEs) and biologics (NBEs). Moreover, medical devices are part of our sample too, to the extent that they are part of the portfolio of firms involving in the development of chemical/biological drugs. This definition was used when checking the firms on the longlist of the biopharmaceutical industry in Belgium (see data construction).

3.2 Innovation

Innovation is characterized by progress. In the context of this paper it is defined as *the invention of a new or significantly improved product or process* which reduces production costs or increases therapeutic value (CRA, 2004). It is non-obvious and useful (requirements to patent). In addition to product- and process innovation the OECD's Oslo Manual (2005) also defines organizational- and marketing innovation but they are less relevant to (bio-)pharmaceutical firms. Inasmuch product innovation is the most prevailing type of innovation in pharmaceuticals, process innovations tend to relate to the production process of these product innovations (Charles River Associates, 2004).

Furthermore, innovation has many facets to it and innovative activity can be gauged at several stages in the production chain (CRA,2004). It is a common approach to break the concept up into an input/effort side and an output side. Whereas R&D expenditures and patent application counts respectively are conventional tools to operationalize both sides. Besides, monitoring patent information to examine innovative activity suits product innovation well (Nagaoka et al., 2010).

3.3 Firm performance

Firm performance is profoundly multidimensional. Nowadays it is popular to evaluate firm using the triple bottom line, disentangling ecological, economic and social performances. In this thesis we confine the concept to economic measures. They are available and reasonably objective. As a consequence, performance will be represented by financial productivity measures. It is defined as *the ratio of outputs over inputs* (Tangen, 2005). A special interest is taken in labor productivity measures, and data reported in the balance sheet are extracted.

4 Problem Statement & Research Question

"The biggest challenge facing the industry is to be able to sustain the flow of innovative medicines from our pipelines"

-Eli Lilly CEO John Lechleiter (2010)

Innovation has always been at the core of the pharmaceutical industry. It is atoned for by the level of R&D expenditures. Because of the nature of a medicine and its direct impact on human wellbeing (effectivity, consistency and quality) the development of new drugs is very resource consuming. The 1990s have been a very successful period with a lot of breakthrough drugs making it through the pipeline. Yet nowadays the firms in the industry are struggling to keep up the pace. This trend has been accompanied by a marked decline in productivity. The return to R&D expenditures is under pressure and does no longer promise a proportionate level of innovative output and subsequent returns. Whereas in the past couple of decades this was assumed to be the case for manufacturing firms (Khanna, 2012; Crépon et al., 1998).

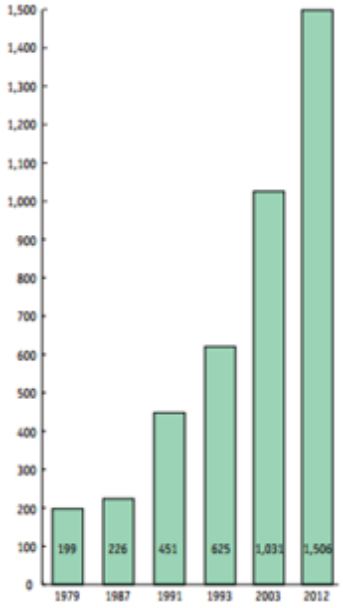


Figure 3: Estimated full cost of bringing a new chemical or biological entity to the market ($ million - year 2011 $), 1979-2012 (EFPIA, 2014)

Two explanations are advocated. The first is the fact that cost of bringing a new chemical or biological entity to the market has vastly risen (Figure 3). Khanna (2012) principally attributes this to increasingly stringent regulations which are accompanied by ever lengthier clinical trials. The latter in turn are necessary to safeguard human health. The second explanation is a significant deterioration in the number of blockbuster drug discoveries, which are drug discoveries with a high sales potential (>1 billion dollar annually). Khanna (2012) points out that the number of approved

NMEs or NBEs in the US has come down from record setting 56 and 45 in 1996-97 to between 20 and 25 in the 2005-10 period. Yet others still doubt the rate of newly approved drugs is actually stagnating given a wider time perspective (Munoz, 2009). Therefore, the most reasonable explanation for the declining return to R&D and return on assets, in general, is that the latest generation of newly developed drugs does not have as much potential to generate revenues. This is in accordance with Khanna (2012) noting there is a shift away from landslide drugs development towards more specialized drug development. Furthermore, it is harder to discover new blockbuster drugs to the extent that the stock is not unlimited and it is hard to make a living out of incremental innovations since development costs are so high.

With respect to the product pipeline it is projected that only 2,5% of the investigated NMEs or NBEs enter the pre-clinical testing phase. At the same time an estimated 5 out of 10000 entities actually undergo clinical trials (0,05%). Where after 0,01% of the initially investigated NMEs or NBEs eventually convert into an approved new medicine (PhRMA 2008 in Hannigan, Mudambi & Sfekas, 2013; Scherer, 2010, p25). The completion of this process involves an estimated cost in excess of 1 billion euros, on average (Figure 3; Khanna, 2012). It implies larger firms have a competitive advantage since they have sufficient revenues over which they can spread the costs (Schumpeter, 1942). It means big enough firms enjoy R&D portfolio diversification benefits. However, it observed radical innovation often originate from smaller specialized firms, operating in niche markets (Foxon, 2007), which calls for a dynamic interplay between firms of different sizes. All with all the literature tends to suggest solely relying on large-scale in-house R&D operations is not the way to address pharmaceutical innovation (Hannigan, Mudambi & Sfekas, 2013). In practice big pharma – the dozens of largest pharmaceutical firms, also recognizes the *closed innovation model* is no longer suited and explore external options to improve ROA within their ecosystems (Chesbrough, Vanhaverbeke & West, 2006).

To appropriate returns from their innovative activity, pharmaceutical firms heavily depend on patents (Scherer, 2010). Moreover, environments that are more technology intensive and directed towards product innovation favor patents. Due to the high investment costs pharmaceutical firms also need more time to recover their developing efforts. In addition, a rather clear definability of molecular structures when compared to non-drug innovations enables a more extensive use of patents in pharma. Yet, the pharmaceutical industry is an exception with respect to the importance of intellectual property rights for appropriation. Most industries value mechanisms based on lead time, secrecy or complementary assets more. Furthermore, patents are not only used as an instrument to extract economic value but increasingly they are used for strategic purposes such as the creation of leverage for negotiations with competitors (Nagaoka et. al., 2010; Lopez, n.d.).

As mentioned, pharma is a dynamic industry with a distinct technology characteristic and heterogeneous product market (Khanna, 2012). Throughout the process of revenue generation, the ability to sustain a flow of innovation will therefore be essential for firms to make profits convicting operational excellence to being of secondary concern. A dominant positive relationship between innovativeness and above-average financial performance may therefore be expected (Artz et al., 2010).

Contemporary challenges for the industry are the ability to embrace increased competition within and from newly emerging technological fields as well as decreasing product life cycles. Balancing suffering profits with engaging to take on less profitable diseases is another difficulty (for example: ebola, snake bites). Additionally, marketing techniques are subject to social constraints which deviate from traditional capitalist sales-intensive communication (Khanna, 2012).

As mentioned, the Belgian pharmaceutical cluster has been strongly export-oriented. It has known strong growth last couple of decades and consequently its importance both to the Belgian economy and the worldwide pharmaceutical industry has been reinforced (EFPIA, 2014; Figure 21 in appendix). A relatively high amount of research and development efforts are undertaken in Belgium when compared to other clusters in the US, Switzerland and neighboring countries. Though innovative output, as measured by the number of patents per inhabitant, seems to lack behind. But there is evidence this conclusion is mitigated by the corporate ownership structure which causes patent applications to be filed in other jurisdictions (Abrahamsen et al., 2011).

With respect to the regions Flanders, Brussels and Wallonia, the latter is less established in traditional pharmaceutical activities. However, Wallonia does account for almost 70 percent of employment in biotech. Moreover, traditional pharma is dominated by a few big pharma firms (J&J, GSK, UCB; Figure 20 and Figure 19 in appendix). Belgium being a federal state has its pros and cons. On the one hand bureaucratic complexities increase transaction costs and lead times. On the other hand, the institutional structure allows for more adequate facilitating policy. In this sense Belgium succeeded in creating a knowledge ecosystem (universities, incubators) and established fiscal regimes for fostering innovation (research grants, tax subsidies; Pharma.be, 2016). Concurrently the number of clinical trials is very high and their approval processes for the first two phases are very fast (18 to 26 days; Figure 22 in appendix; Abrahamsen et al., 2011, p21). Re-imbursement procedures, on the contrary, tend to be lengthy. Furthermore, Belgium has a problem of high labor costs. The country has invested a great deal in its human capital, though within pharmaceuticals there are signs of a scarcity both in quality and quantity of personnel which aggravates the cost issue (Abrahamsen et al., 2011, p23-24). Notwithstanding, cost and scarcity problems, hourly labor productivity turns out to be high (Abrahamsen et al., 2011).

Both from the previous discussion as from theory (Solow, 1957) it is clear innovation is a key contributor to economic growth. It is through technological progress that total factor productivity is enhanced (Mohnen & Hall, 2013). Therefore, unsurprisingly previous research has been able to confirm a positive and robust relationship between innovation and productivity empirically (for example: Vancauteren, Melenberg & Plasmans, 2014). Furthermore, innovative efforts have proven to reflect positively on output and firm performance in a manufacturing context (R&D elasticity). Although it is hard to pinpoint the relationship exactly as it depends on the measures and method used. Still, elasticity estimates (productivity measures - R&D) have been between 3 and 50%. In order to capture the relationship between innovation and productivity in a more comprehensive way, different conceptual models further disentangle it (Sanyal, 2014; Vancauteren, Melenberg & Plasmans, 2014).

One important school of thought has been the Crépon-Duguet-Mairesse (CDM) structural equations model (Crépon, Duguet & Mairesse, 1998). CDM recognizes the sequential nature of research (R&D), inventions (patents applications/innovative sales percentage) and productivity. The model estimates these relationships in a recursive fashion. In this study, in a similar way, firm performance (labor productivity) will be linked to innovation output (patent counts), innovation output to R&D, and R&D to its determinants. In fact, this approach means using an estimated dependent variable from the previous equation as an input in the next logical equation, effectively breaking up the general relationship between innovation and productivity into more tangible parts (Crépon, Duguet & Mairesse, 1998).

On broad terms one argues successful research means (more) innovative output is produced (which is frequently patented). Innovations, in turn, are deemed to boost the financial prosperity of the inventor (Grilliches, 1990). Therefore, firms that manage to be better innovators will outperform their peers, ceteris paribus. The main objective for this master thesis is to check whether these relationships can be demonstrated for the pharmaceutical industry in Belgium. The following research question is formulated:

"Can a positive relationship between innovation and firm performance (productivity) be confirmed in the Belgian pharmaceutical industry?"

Throughout the literature review we will provide the underpinnings for the disentanglement of the innovation-performance paradigm (hypothesis formulation). Also, relevant empirical and theoretical literature will be revisited. Whereby the aim is to identify robust measures and methods. Furthermore, the context will be sketched adequately with respect to recent trends (open innovation, section 5.7) and the Belgian environment (above). Finally, an overview will be given wherein the hypotheses are summarised and the conceptual framework is established. This framework will be distilled from the literature, while considering limitations encountered in the data construction process and limitations to our own capabilities

5 Literature Review

5.1 Innovation, technical change and Productivity

Relating to the definition of firm performance, recall productivity growth means more output is produced with a similar amount of inputs. Or stated differently, the ratio of output over input increases (Tangen, 2005; Mohnen & Hall, 2013). It is the most prominent source of a comparative advantage and can be pronounced cost and quality wise (Tangen, 2005, p35; Figure 4 below). On a firm level, productivity is therefore a key driver of competitiveness. Being more 'productive' than your competitors is important. Productivity growth can be achieved in a number of ways: through economies of scale, increased efficiency, increased input quality, reduced idle time or reduced rigidities in input markets and by an outward shift of the production possibility frontier (Mohnen & Hall, 2013, p50). It is this outward shift that is accomplished through technological progress and it has spurred our standards of living (Solow, 1956; Solow,1957).

At the core technical change is economic growth not resulting from changes in the level of one or more factor inputs of the production function. Given the assumptions of constant returns to scale and competitive factor pricing, it is the 'residual' growth as first defined by Robert Solow (1957) using a growth accounting approach. Because this residual cannot be accounted for by input factor changes, it must represent improved technology (Gärtner, 2013). In this study a more direct approach is adopted in trying to link output directly to innovation rather than accrediting unexplained growth to technical change which allows for a gamma of errors such as bias introduced by omitting variables.

The most tangible and relevant expressions of technological progress in pharmaceuticals are product and process innovations. Although the OECD's Oslo Manual (2005, in Mohnen & Hall, 2013) defines organizational and marketing innovations as well. Product innovation is defined as a new or significantly improved good or service with respect to its characteristics and intended uses. In the context of this study it usually adopts the form of new or improved drugs, and has a good chance of being patented (Nagaoka, Motohashi & Goto, 2010; Lopez, n.d.). Process innovation is defined as a new or significantly improved production process or delivery method, and the former seems the most relevant. Note, production process innovations are less likely candidates for patents because they disclose information on which infringements are much harder to detect and litigate. They are frequently filed as trade secrets (Nagaoka, Motohashi & Goto, 2010, p1106).

The literature recognizes difficulties with capturing all aspects of innovation, productivity and the relationship between them. Tangen (2005) signals three reasons: discrepancies between the verbal definitions and their mathematical translation (especially true for innovation), a lack of defining the concept by authors in their studies and a lack of awareness about the multi-interpretability of productivity. Also both concepts have been utilized in dissimilar contexts and at various levels of aggregation (Tangen, 2005, p35). The definitions constructed in section 3 are an attempt to meet those concerns.

While defining innovation we established it as a process. A process which can be assessed at different points in the production chain relating to efforts or outcomes. In the literature innovative efforts, or inputs, are mostly related to R&D expenditures (Rogers, 1998). These expenditures are often transformed to allow for more robust estimations (log-transformation), represent knowledge stock or R&D capital intensity. From time to time R&D expenditures are found categorized according to funding source or research goal (Vancauteren, 2014).

In addition to R&D expenditures, Rogers (1998) states statistics about the acquisition of technology from others, marketing expenditures for new product launches, training costs, R&D staff, intellectual property rights, composite indicators and questionnaires may be used to measure innovative input. Whereas, relating to the pharmaceutical context, other clear-cut tools such as the number of clinical trials and their relative speed become relevant as well. Withal, the fact that so many measurement tools have been tried, supports the statement that innovation is a concept with many facets. It is hard to represent adequately using just one tool. Hence, in this study standard practice of R&D expenditure transformations will be carried out.

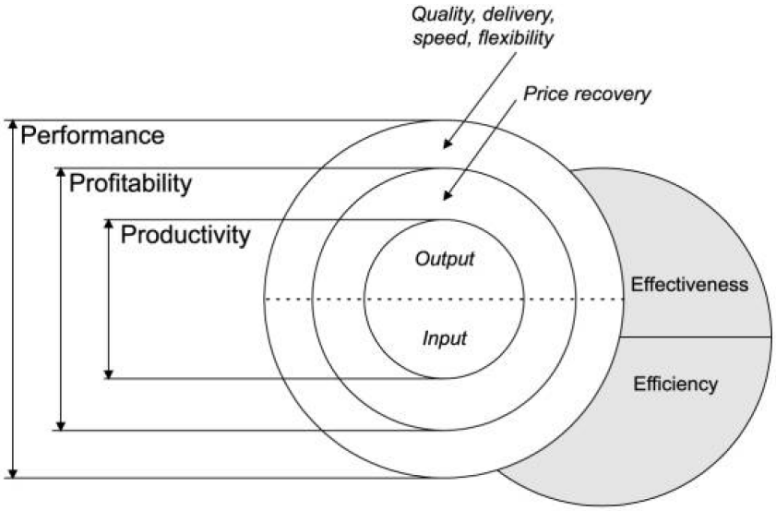


Figure 4: The triple P-model (Tangen, 2005, p43)

Innovative output is also embodied in various ways. Rogers (1998), in the same paper, sums up four categories of output measurements: (counts of) the introduction of new products and processes, sales percentage of new products, firm performance (indirectly) and intellectual property rights. Specifically, patent statistics will be used as a proxy, and this is the subject of sections 5.4– 5.6.

Subsequently innovation can be linked to firm performance. As mentioned, firm performance will be represented by (financial) productivity measures. These are at the core of a for-profit firm's performance (Figure 4). Moreover, they are available and reasonably objective due to reporting standards. Increasing the output-on-input ratio is associated with effectiveness (doing the right things) and efficiency (doing things right). It is meeting customer requirements, thereby creating value (effectiveness, output) while simultaneously using resources in an economical way (efficiency, input) during the transformation process (Tangen, 2005).

Building on the discussion above, multifactor factor productivity (growth) and labor productivity measures are identified as the main dependent variables (Vancauteren et al., 2014; Sanyal, 2014; Mohnen & Hall, 2013). Reviewing the existing evidence Mohnen & Hall (2013) summarize value added, sales, gross output and (Solow residual) TFP as the most prevalent ways of operationalization. Subsequently these measures are commonly log-transformed and corrected for the level of activities (divided by the number of employees). Also the time-differences of the former are frequently used thereby linking the independent variables dependent variable growth directly. In accordance with sketching the importance of the pharmaceutical industry for labor opportunities in the Belgian economy this study will employ labor productivity measures to represent firm performance. The natural logarithm of value added per employee and revenues per employee will be used in the empirical part.

Concerning the empirical evidence, different conceptual models and operationalizations have been tried to capture the relationship between innovation and productivity. As mentioned, one important school of thought has been the Crépon-Duguet-Mairesse structural equations model (Crépon, Duguet & Mairesse, 1998). CDM recognizes the sequential nature of research (R&D), inventions (patents applications/innovative sales percentage) and productivity simultaneously. In this sense the authors assume innovative output adequately and fully represents the effect of the innovative activities on firm performance. The principal difference with other schools of thought is not directly linking R&D and firm performance (Sanyal, 2014, p219-220). In this thesis, the significance of R&D on labor productivity will be checked empirically to assert the robustness of our model in representing the underlying relationships. Although, the CDM model will still be the model underlying our approach (Sanyal, 2014; Mohnen & Hall, 2013).

Outlining the CDM script, we first note, Crépon et al. (1998) used three different kind of equations: research, innovation and productivity equations. The research equation targets the engagement in R&D. In two phases a firm's probability to perform R&D is modelled and its R&D-intensity estimated. CDM control for size, market share, diversification, demand conditions and technological opportunity. Next, the estimated values entry the innovation equations as an independent repressor representing innovative input. Both patent applications and innovative sales percentage are subsequently estimated, and in a similar way they enter the productivity equations. Productivity is represented by the natural logarithm of value added per employee.

CDM also included several other variables. Concerning the demand pull and technology push indicators Crépon et al. (1998) were able to conclude that market demand is perceived as the stronger incentive to innovate. Nonetheless, they confirmed positive effects on a firms R&D probability and intensity for both forces. Their direct effects as far as innovative output could also be established. Notably, the authors controlled for the skill composition of labor in the productivity equation. They did this by identifying the percentages of engineers and administrators, which were tested significantly with coefficients reflecting their marginal pay. Yet, usage of technology push, demand pull and skill decomposition will not be followed through in the empirical part due to lack of the appropriate data.

Returning to the equations market share, firm size and diversification are diagnosed as the most important determinants of R&D engagement and R&D intensity, though firm size does not seem to affect intensity more than proportionally. Regarding the innovation equation, the results do not seem to agree Schumpeterian conjectures since the equations between innovative input and output exhibit constant returns to scale rather than increasing returns. Even so the influence of R&D intensity on innovative output is quite strong. Furthermore, CDM also establishes the positive influence of innovation on productivity with the elasticity of the logarithm of value added per employee to patent counts being 0.13 (Crépon et al., 1998). Later studies generally confirm CDMs results, depicting a positive and robust relationship between innovation and productivity (Vancauteren, Melenberg & Plasmans, 2014).

In addition, the empirical literature identifies a direct positive influence between the efforts made to innovate and productivity (for example: Klette and Grilliches, 1996 or Smolny, 2000, in Sanyal, 2014, p203). Yet the strength of this relationship is unclear. The estimated elasticity of productivity to R&D for manufacturing firms hoovers between 0.03 and 0.50 and possibly even larger in R&D-intensive industries such as pharmaceuticals (Sanyal, 2014, p203). Also, a large portion of this variance is thought to be the outcome of the studies including different (proportions of) industries and considering different geographical areas. Explaining why there is a direct effect between R&D and productivity two prominent arguments are: first, patents or other proxies of innovative output do not succeed in representing all underlying aspects of the relationship between innovation and productivity whereas some of these aspect can be represented by innovative input measures relating to R&D expenditures and second, there can be knowledge spillovers, for example a stock a knowledge can be built which makes other innovative efforts more efficient and productive because of gained complementary insights.

Revisiting the research question of whether or not we can identify a positive relationship between innovation and firm performance, the underpinning for disentanglement are now provided. Similar to the CDM framework a model three sequential equations is applied. Referring to firm performance we can therefore formulate the following research question:

"Is there a positive relationship between innovative output and firm performance?"

And subsequently formulate to following hypothesis:

H1: There is a positive relationship between innovative success (as measured by patent application counts) and labor productivity (as measured by the natural logarithm of value added/revenues to the number of employees).

Mathematically this translates to:

$$\ln\left(\frac{Value\ added\ or\ turnover}{\#\ employees}\right) =$$

$$\beta 0 + \beta 1 * Predicted[\ln(R\&D\ intensity)] + \beta 2 * predicted[\ln(patent\ application\ counts)] + \beta 3 * X + \varepsilon$$

For each identifiable i-th firms in each t-th year (subscripts omitted). Whereby the beta coefficients β1 and β2 represent the elasticities (% to % change) of labor productivity to innovative input and innovative output respectively. We leave the proof of this to econometric textbooks such as Basic econometrics by Gujurati and Porter. Furthermore, X represents the vector of control variables such as open innovation, size and competition controls. ε, the error term is assumed to be independently and identically distributed conforming to standard OLS procedures. One of the control variables in the vector X is competition. It is the subject of the next section.

5.2 Innovation and Competition

The relationship between innovation and competition is ambiguous and might contain a problem of causality. It can be questioned whether the competition faced by an individual firm influences the decision to take up innovative activities or that in fact innovative activity determines the level of competition faced by the firm. Furthermore, the directionality of effect of competition is found to be dependent on the context in which it is investigated (Garcia-Fontes, 2011; Gilbert, 2006). Differences with respect to technological intensity, industry characteristics and the legal framework can play an important role. Here, the assumption is made that competition acts as an exogenous variable mediating decision making throughout the process of innovation (Garcia-Fontes, 2011).

In academic literature a gap exist with respect to the effect of product market competition on innovative output for the targeted at the pharmaceutical sector specifically (Sanyal, 2014). Therefore, we turn to literature that applies for manufacturing industries in general. Early work already posits contradicting theory and empirical evidence. Schumpeter (1942) argues a more concentrated market fosters innovation, implying less competition and more market power. He reasons market power allows firms to set prices above (perfectly) competitive levels, make a profit and deliver return on R&D investments. In a less competitive industrial environment one may therefore expect greater incentives to innovate. In the context of pharmaceuticals, the importance of patents to appropriate returns are supporting Schumpeter's claim. Yet, subsequent research has also countered Schumpeter's rationale with a lot of critique.

In fact, contrarily to a negative association, Arrow (1962) proposed a positive effect competitive environment on innovation. Implicitly he argues the problem of appropriation can be dealt with (for example: licensing). And he posits a competitive firm has a greater incentive to look for a comparative advantage. Whereby especially cost-cutting process innovations are deemed relevant. Similarly, urban knowledge dictates that making no progress really means getting behind. However, this study comprises the Belgian pharmaceutical sector which is mainly product-oriented and draws on patent protection (less suited for process innovations). A positive (linear) relationship therefore seems less likely. But referring to economic activity as a whole (including all other industries) a positive relationship could still hold. A view which turns out to be endorsed by the European Unions' policy, as they strive for unification of internal markets and more competition (European Commission, 2016).

Other than a linear relationship between competition and innovation, non-linear relationships have been proposed as well. Yet, there still is no consensus about the functional form to be fitted. Prominently, Scherer (1967, in Aghion et al., 2005) demonstrated a positive association between firm size and patenting but it reversed at larger firm sizes, thereby underpinning a (quadratic) inverted-U relationship (for a theoretical explanation see: Aghion et al., 2005). Other studies such as Sanyal (2014) even found support for a U-relationship, using citation-weighted patents as dependent variable. The effect might even be indirect as Gayle (2003, in Sanyal, 2014) argues, because more concentrated markets are associated with higher R&D expenditures in turn associated with higher innovative output, ceteris paribus. Nonetheless, most authors agree results vary according to the operationalization used (Garcia-Fontes, 2011).

Typically, competition is represented by a concentration index, entry and exit statistics or based on the price-cost margin (Aghion et al., 2005; Mohnen & Hall, 2013). In this study a concentration index based on domestic market share (domestically reported revenues) will be constructed in the philosophy of the Herfindahl-Hirshman index.

Herfindahl-Hirshman index: $HHI_i = \sum_k b_{i,k}^2$

Lerner index: $LI_i = \frac{\text{price} - \text{marginal cost}}{\text{price}}$

Its robustness will be compared with a similar index constructed based on corporate employment and a profit margin measure (~Lerner). Because the relationship between competition and innovation is of secondary interest in this thesis, it will be included as a control variable in various regressions. It means focusing only on the direction and significance of its first order coefficient. Furthermore, a measure of size (the natural logarithm of employees) will also be included not to confuse (dis)economies of scale with the competition effect.

On a personal note, we argue competition is desirable, as it allows the economy to reach short term equilibria and allows factors to be optimally priced (=marginal cost) from a private point of view. Whereas in the presence of a negative associating between competition and innovation 'perfect' competition might not be optimal. In this case, short run optimal pricing can be offset by enhanced growth prospects. Effectively a trade-off exists.

5.3 Incentives to innovate: Determinants of R&D Investments

The pharmaceutical industry provides an interesting setting to study the economics of innovation. The invention of new NMEs and NBEs requires an unusually high up-front investment and a lot of time. Accordingly, it is a very knowledge-intensive process requiring extensive coordination and collaboration to bring about a new product. Moreover, product innovation is the most important driver of revenues streams for pharmaceuticals. Equivalently (expected) patent protection is important to predict the ability to appropriate the returns (Nagaoka et al., 2010; Scherer, 2010; Khanna, 2012).

For individual firms an R&D investment is evaluated in a similar way to other capital investment projects. If a firm is profit maximizing this implies it will embark on the investment if it expects positive net cash flows (net present value), considering the opportunity cost of capital. Incentives to innovate are therefore closely related to the determinants of R&D investments. Conjointly, because of the lack of perfect information and because of the amount of resources invested (relative to total assets: R&D intensity), R&D projects are above-average risky investments (Arrow, 1962). Financial portfolio theory (Capital Asset Pricing Model & Efficient Market Hypothesis) therefore states they should be discounted a higher rate (see for example Bodie, Kane & Marcus, 2013).

As shown in section 5.1, investing in R&D, propelling innovation, is nonetheless desirable from a macro-economic point of view. Consequently, the government is a stakeholder. Ergo, it is no surprise that the regulation (intellectual property rights, quality assurance) and fiscal regime pursued by different governments are important determinants for the level of R&D investments. This study, however, is confined to Belgium and on this ground we do not set out to proxy (changes in) those regimes and regulations. To that end, the assumption is made all companies are affected in a similar time-constant fashion and regional differences in policies are insignificant. Note, however, substantial regulation and fiscal policy is involved (Scherer, 2010; Pharma.be, 2016).

An important argument for pharmaceutical mergers and acquisitions (M&A) in recent decades, has been increasing diversification possibilities of R&D (CRA, 2004). It is the ability to increase the expected risk-adjusted return on investment because the outcomes of several R&D projects, are not perfectly positively correlated (Bodie et al., 2013). Effectively, this argument boils down to existing economies of scale (Ciftci & Cready, 2011). Likewise, the fact that R&D costs have been disproportionately rising will undoubtedly also have urged executives to resort to M&A coping with the budgetary constraint. By the same token, economies of scope have also been frequently advocated (Schumpeter, 1942; Graves & Langowitz, 1993; Henderson & Cockburn, 1996; CRA, 2004).

Contrarily, the increasing size of corporations may also result in diseconomies of scale. Mainly because of rising organizational costs and increasing complexity of managerial tasks a higher level of activities could lead to a less-than proportionate return. Whereas potential (dis)economies of

scale are principally subject of the R&D-patenting investigation, it is clear that corporate size might also influence R&D-intensity. Moreover, hypothesizing a positive association between the level of R&D expenditures and the level of corporate activities is evident. In the regressions size will be approximated by the natural logarithm of employees (Graves & Langowitz, 1993; Artz et al., 2010).

With size introduced as a first plausible determinant, an exploration of other firm-specific factors that influence in the R&D investment decision imposes itself. Therefore, the following research question can be formulated:

"What tangible firm-specific factors are determinant for firms to invest in R&D?"

Classically, this research question also is the starting point of the pioneering framework introduced by Crépon et al. (1998). Subsequently CDM set out the estimate R&D-intensity in two stages by first determining whether or not resources are allocated to R&D [1] (probit equation) and thereafter the (natural logarithm of the) level of R&D expenditures per employee [2] (OLS: two step Heckman).

[1] $Dummy\_(non-)performing\ R\&Dit\_hat = \beta 0 * X0it + u0it$ with,

i = firm, t = year, u0it = i.i.d. error, X0it = vector of explanatory variables

left hand side = estimated [0:1] (above threshold yes/no)

[2] $\ln(R\&D/employee)it\_hat = \beta 1 * X1it + u1it$ if [1] = 1

$\ln(R\&D/employee)it\_hat = 0$ if [1] = 0 with,

i = firm, t = year, u1it = i.i.d. error, X1it = vector of explanatory variables

left hand side = estimated true (latent) R&D intensity

In the philosophy of Crépon et al. (1998), the estimation method employed will also be a sample selection model: the two stage tobit model (Heckman, 1979). Tobit II first estimates the R&D-probability that a firm (~probit equation) performs R&D to surrogate the observed (reported) values and account for innovative firms that do not report R&D. This is appropriate because we may hypothesize non-reporting but R&D performing firms to have systematical differences with respect to the ultimate dependent variable R&D-intensity (mills lambda; Huang et al., 2011). R&D-intensity [2] is subsequently estimated (OLS in two step Heckman) if the left-hand side is above some threshold. Otherwise, it is assumed to be zero. Furthermore, a technical note referring to the regression method is that the vector of independent variables in the two phases need not be the same. For proper estimation via the two step method, the probit equation even has to include at least one additional independent variable (Vancauteren et al., 2014; Crépon et al., 1998).

Elaborating on the vector of independent variables, CDM (1998) included other Schumpeterian determinants such as competition (market share and diversification of activities), together with the number of employees as a size estimate. Moreover, demand conditions and technological

opportunity also entered their augmented equations. Yet, these estimates were not retrieved for this thesis. Nevertheless, they appeared to have a consistent (though not always significant) positive effect on R&D intensity. Determinants of competition will be represented by the measures of competition established in the section 5.2 (Crépon et al., 1998; Garcia-Fontes, 2011).

More recent academic research on innovation-productivity intricacies theorized the inclusion of other variables. As a first, dynamics of entry and exit together with firm age enable to investigate maturity and the stability in the industry. Withal, the relative factor input mix is likely to change as the firm ages. Also the innovative success between survivor and busted firms is likely to be systematically different. Analogously, this might be the case for the market value of innovations between juvenile and established firms. Besides, since, R&D (sunk) cost take very high proportions, this is also likely to deter new entrants. Preliminary empirical regressions show firm age, entry and exit are unreliable repressors in our dataset. This might be because of the limited timeframe (Sanyal, 2014, p71-72).

Another frequent determinant of R&D activities is the overall capital intensity that characterizes a firm. The argument can again be related to financial portfolio theory. Whereas firms that exhibit a higher capital intensity can be assumed to face less of a budgetary constraint. Therefore, firms that are less impeded financially can adjust their R&D expenditures more flexibly to level deemed optimal. Moreover, a greater pool of funds allows for more diversification. The expected coefficient on capital intensity is subsequently positive (Hottenrott & Peters, 2011).

Other determinants confer on patent stock and knowledge capital stock. Patents as an instrument for the protection of successful innovations are included because previous success may indeed spur firms to invest more in R&D than they would otherwise ("success breeds success hypothesis"). Similarly, it is likely that previous commitment to R&D influences contemporary and future commitment to R&D. In econometrics these relationships are said to be autoregressive (dependent on its lagged values) and can be underpinned in various ways. For example because of an irrational attitude towards sunk costs or because of a built up knowledge stock. Empirically, stock measures or dummies can be introduced to represent the existence of (past) innovative activities (Vancauteren et al., 2014).

Finally, the plausible influence of open innovation is explored in this thesis. With a primary concern to the sign of the coefficient of the OI indicator. Do open innovating firms have more or better technological opportunities and therefore invest more in R&D? Are they able to mitigate soaring cost through OI collaboration, therefore exhibiting a lower R&D intensity? Or are different specifications inconclusive on the sign? Consequently indicating both effects could be present, none of them could be present, measurement is inadequate, model specification is biased or a more complex relationship exist? Section 5.7 will investigate OI in the pharmaceutical sector more thoroughly and propose a measurement. The discussion is therefore postponed. We now turn to appropriability mechanisms and patents in particular.

5.4 Appropriability: Intellectual Property Rights and Patents

After being more direct in approaching the determinants of R&D, the ins and outs of intellectual property rights will be discussed here. They are tied to appropriability mechanisms as a means to generate return on R&D-investment. Both this section and the next will provide useful insights into the innovation-productivity relationship. This section sketches a more broad perspective linking types of appropriability mechanisms to inventions, where in fact those appropriability mechanisms closely relate to business models. It links innovation as a black-box process outcome to profitability. The next section focusses on patent statistics as a proxy for innovation output such as frequently used in academic research. Furthermore, it provides background info on patent law and the filing procedure.

Starting out with appropriability, economic theory differentiates four types of goods based on excludability and rivalry: private, common, club and public goods. Rivalry means a good can only be consumed once, or at least not simultaneously. Excludability means individuals can be prevented from consumption, and therefore a positive price can be asked. Private goods are both excludable and rivalrous, common goods non-excludable, club goods non-rivalrous and public goods both non-excludable and non-rivalrous (Lipsey & Chrystal, 2011). Whereas excludability is a prerequisite to adequate (private) market pricing, non-rivalry can also pose appropriation problems because of heterogeneous individual utility within differentiable customer segments (club goods).

As outlined in section 3 this master thesis adopts a narrow definition of innovation. It was described as the invention of a new or significantly product or process. More conceptually one could state an invention is in essence a new idea. Hence, regarding inventions as novel ideas, they can be labelled public goods. Namely, distinctive properties of ideas are their non-rivalry and non-excludability. The consumption, for example, of the idea by an individual does not prohibit others to simultaneously consume it. An idea can be used multiple times (non-rivalry) and an idea is also not easily locked away for outsiders by its owner (non-excludability). Left unaddressed, economic theory learns inventions will therefore be underproduced, if produced at all.

From the paragraph above it is evident that unrestricted consumption is not always to the liking of the intellectual owners of innovating ideas. They would want to accrue the forthcoming profits (for example: Arrow, 1962). In order to provide enough incentives, they should be allowed so. On these ground commercial exploitation of a novel idea can only be accomplished if potential consumers and competing firms are somehow prevented from unrestricted and free consumption for at least some time. A business model needs to be built (Vanhaverbeke, 2016, guest lecture). In general, this means at least one appropriation mechanism is needed. What is more, usually multiple are combined (Lopez, n.d.). In pharmaceuticals intellectual property rights dominate the field, thereby providing an excludability mechanism similar to temporary monopoly rights.

Although intellectual property rights dominate the appropriability landscape in pharmaceuticals, this is rather exceptional. Patents in particular are found no practical means to an end in many other industries. Many different industries use a combination of many different mechanisms of whom the perceived effectiveness varies a great deal among those different settings. Among the

more formal mechanisms innovating firms use, Lopez (n.d.) mentions copyright and industrial designs next to intellectual property rights (IPRs). Intellectual property is defined by the world intellectual property organization (WIPO, 2015) as creations of the mind, such as inventions; literary and artistic works; designs; and symbols, names and images used in commerce (WIPO, 2015). Although, here, IPRs are interpreted as legal appropriation mechanisms for process and product innovations which contribute to the value stream of a pharmaceutical firm. Other than IPRs informal appropriation mechanisms are commonly deemed very effective. The most prominently cited pathway is lead time (innovation to market entrance). It is followed by learning-curve experience, simple secrecy and complementary capabilities (knowledge, manufacturing, marketing, ...) (Lopez, n.d.).

Notwithstanding the practical importance of the informal mechanisms, the importance of intellectual property rights has steadily increased throughout the last decade. Subsequently, they have been gradually attracting most academic attention. Lopez (n.d.) argues this upsurge can be attributed to the increased patenting practices, strengthening IPRs both nationally and internationally, and rapid development of innovation-intensive industries such as pharmaceuticals. They are an important (measurement) tool nowadays. Additionally, innovation surveys burst onto the scene recently, to meet the imperfection of the existing innovation measurement tools. They supplement information extracted from patents and balance sheet information. Moreover, they enable academics depict a more robust blueprint of the dynamics of innovation (Lopez, n.d.; Mairesse & Mohnen, 2010).

Concerning the effectiveness of the different appropriation mechanisms, research has shown that in general the perceived effectiveness depends on the context. For service firms and non-innovating firms IPRs are not valuable. In general, this is also true for industrial sectors. Lopez (n.d.) states there is a limited importance of patents for innovative manufacturing firms. Whereas especially informal mechanisms such as lead time and secrecy (first mover mechanisms) seem to be more highly valued. But also complementary assets (size, market access, employees) are mentioned. If these assets are indeed important they help explaining the inclusion of size into the R&D determinants regression. Moreover, other studies could also consider to include the skill composition of labor and other measurable complementary assets (Crépon et al., 1998; Lopez, n.d., p9).

Regardless of the informal pathways, early studies on patents and appropriability already confirmed patents are essential for converting innovations in profits, though only in the pharmaceutical industry (Scherer et al., 1959, in Lopez, n.d., p9; Silberton 1973, in Lopez, n.d., p9). They confirm that the pharmaceutical industry is an exception to the rule. Later studies also indicated the key role of the patenting system to the product stream in the pharmaceutical industry and added it plays a significant role in the chemical industry as well (Mansfield, 1986 in Lopez, n.d.). This might be attributable to the peculiarities of chemical compounds, which lean themselves to quite transparent formulations. Furthermore, the temporary exclusive rights which establish a monopoly-like situation, are more important because the initial investment which has to be recovered far exceeds those in other industries. As described in the problem statement,

investments typically accumulate to more than one billion euro, thereby needing ten to fifteen years for a new product to be propelled through the pharma-pipeline.

Readdressing manufacturing firms in general the question of "what influences the preferred appropriation strategy a firm chooses?" imposes itself. It is a question that hints at the perceived effectiveness of the different options. Of course the industrial context in which the firm operates is of importance. Moreover, also the strength of the legal mechanism and the extent to which they are enforced are crucial. Yet, those factors of context aside, it appears R&D-performing firms are more likely users of IPRs (Hanel, 2005). Moreover, R&D-intensive firms even more likely to patent as part of their strategy (Lopez, n.d.). A probable explanation is that these firms are part of industries where (product) innovation is a cornerstone capability. On top of that firms engaging in cooperative R&D perceived a higher effectiveness of patenting system. Patents are therefore also a viable instrument in an open innovative climate (Lopez, n.d.).

Continuing to build understanding of perceived effectiveness, the relative importance of secrecy mechanisms declines with the increasing size of a firm. This means that more inventions are patented, readily disclosing information (Lopez, n.d.). On the other hand, it should be cautioned increased strategical patenting practices with increasing size may heavily inflate their patent counts (Nagaoka et al., 2010). Moreover, Lopez (n.d.) does state secrecy a frequent substitute for IPRs when dealing with process innovation rather than product innovation. With process innovation, infringements are harder to detect and enforce. In addition, uniqueness (as required for patent eligibility) is harder to prove. Hanel (2005) notes that larger firms therefore tend to be heavier involved IPRs on average to protect their knowledge, shielding it off from competitors.

When dealing with product innovation the perceived effectiveness of product patents proves to be positively correlated with the frequency of those innovations. Similarly, the count of applications also correlates positively with the perceived effectiveness of patenting. It is indicative of innovative persistence: "success breeds success", at least if adequately represented by patent counts. Other potential factors such the length of the product life cycle or the extent to which a firm's products were a novelty, had almost no explanatory power to the propensity to patent (Vancauteren et al., 2014; Lopez, n.d.).

Looking into the motives for patenting on the level of an individual firm, we notice that appropriating positive returns from an innovation is not the only reason firms patent. Cohen et al. (2000) gave, in order of importance: prevention of copying, patent blocking (protective portfolio generation), prevention of suits and bargaining power within negotiations as patenting motives. Notably licensing out was not perceived important: directly valorizing knowledge by selling it appears not to be common. Conversely, torchbearers of the open innovation paradigm, which is a relatively new concept, do label licensing as a potential for strengthening the internal knowledge base and expand pathways to the market (Chesbrough et al., 2006). The importance of licensing nowadays may therefore be upcoming or at least different from its stature at the time Cohen et al. (2000) conducted their research.

The study by Cohen et al. (2000) further hints motives differ according to the nature of the product market, according to the homogeneity of its products. In a more differentiated product market portfolio patenting is a useful tool to getting competitors to negotiate. Whereas in a more homogenous product market being first with an innovation and subsequently blocking out rivals and their potential substituting innovations appears to be the primary concern. Regarding the motives not to file for a patent the disclosure of information and the ease of inventing around are cited. For smaller firms the lack of financial resources or the inability to detect and litigate infringements also impact the decision (Lopez, n.d.).

The most important individual motives to patent all seem to have a strategic connotation to them. This explains the "patent paradox": the coexistence of a rising trend w.r.t. patent applications and the limited (perceived) effectiveness of the patenting system in terms of appropriation. While appropriation (relating to incentives to innovate) is the economic reason for maintaining a patenting system, the evidence of a patent paradox should at least ought policy makers to rethink the current regulatory body. The mere existence of hold-ups by patent trolls, for example, only destructs welfare. This at least hints at contra productiveness for economic development as a whole.

In summary, the patent paradox implies society might not be acknowledging the virtues of a self-organizing market adequately. Yet, governments do have reasons to acknowledge and facilitate IP, installing intellectual property rights (IPRs). First, the 'social justice' of enabling inventors to be rewarded for their efforts is a viable argument. Secondly, IP are intended to correct private market imperfections induced by externalities within markets for the supply innovation. Relating to the previous argument the most important reason for a government to intervene is macro-economic growth. Growth is propelled by the rate of innovation. (Solow residual; Solow, 1956; Solow, 1957). Short term economic sacrifices can therefore be defended by enabling increased innovation and more prosperity within a wider horizon. Policy makers already proof to acknowledge macro-economic growth is desirable since it allows societies to flourish. Furthermore, the legislator must actively seek for ways to encourage the technological progress, effectively updating the framework to the contemporary macro-economic environment (incentive based) and industry specific context. Moreover, the government can add value by providing stability which reduces the uncertainty of investment projects (Bodie et al., 2013).

5.5 Patent law, the procedure and patent statistics

Centuries ago no intellectual property rights or patents existed. Thenceforth the first intellectual property rights and were subject to national authorities. Then, towards the second half of the 20th century governments felt an urge for harmonization, standardization and simplification. This meant significant progress could be made on the convergence of national patent systems in supranational frameworks. Although both still exist simultaneously. In 1970 the foundation was laid by establishing the Patent Cooperation Treaty (PCT) which made it possible to apply for one patent internationally instead of having to file at multiple (national) patent offices. In recent decennia minimum protection standards to be attained by national patent laws were promoted (European Commission, 2014).

In Belgium a lawful patent can be administered by the *Dienst voor de Intellectuele Eigendom* (DIE) for national validity. Typically, they are also filed for international validation at European Patent Office (EPO). As a rule of thumb a patent is granted for 20 years starting from the application date (≠first filing), both for Belgium and internationally (FOD Economie, K.M.O. Middenstand en Energie, 2016). For the pharmaceutical industry a prolongation up to five years can be provisioned (supplementary protection certificates or SPCs), if it is demanded by market conditions (European Commission, 2008). However, the coexistence of multiple authorities has the important drawback of "forum shopping" (European Commission, 2014). It allows firms to address themselves strategically to the most benevolent jurisdiction. Consequently, the level playing field is destroyed (European Commission, 2014).

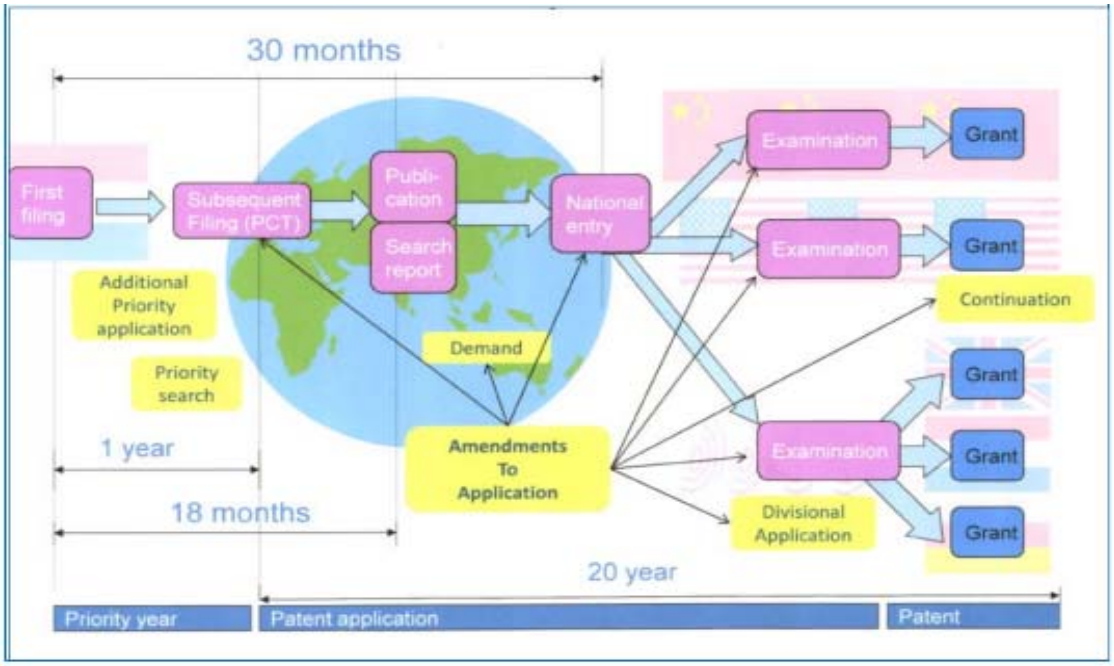


Figure 5: The Patenting Process

The entire life cycle of a patent can be represented by a multi-phased timeline (see Figure 5 above). The patenting procedure starts while R&D is still ongoing. The applicant claims to have discovered a nonobvious and useful novelty or a significant improvement of an existing product (Thomson Reuters, 2016). Subsequently, the innovation can be filed for a patent. In the mean while the product continues its route from the R&D-pipeline to the market.

EPO (2016) breaks down the procedure in nine stages. Besides, the applicant usually addresses an expert attorney to follow through the filing on his behalf. First, the applicant hands in documentation with information on the patent conforming to all official requirements. If the filing appears legitimate and is accepted one's application is given a priority date (Figure 5). During the following year, the patent authority will examine the formalities more thoroughly and you are allowed to file for the same patent in with different authorities (countries). In accordance, these filings will be treated as if they were filed for at the priority date.

12 to 18 months from this first filing a search report is completed (3rd stage). It is assembled by an experienced examiner on behalf of the patent office and lists all prior art documents that were found and deemed relevant. It reflects on the claims for novelty, the drawings and the description as filed by the applicant. Also, a first opinion on the patentability is included. In the 4th stage the documentation is published (after the 18th month). From this point on protection is established (granted or not) and the patent lifetime starts. Before the end of the second year after the assigned priority date the applicant is requested to decide whether to continue the application or not. If the patent is pursued a substantive examination will follow (5th stage). A panel of three examiners eventually decide whether your application meets the official requirements and the decision is made publicly. If the patent is granted and all required payments have been made the patent is instated and is enforceable from the application date onward (EPO, 2016).

The final three stages embrace legal issues. The validation of the patent in the different jurisdictions (7th stage). This has to be done in a timely manner. The penultimate stage (8th) is opposition by a third party. This is the last possibility to challenge the patent on a supranational level. And finally a 9th stage is identified as all EPO decisions being open to appeal (EPO, 2016). All of the patents can be consulted and they are aggregated in databases. These, usually paying, since databases are valuable. They can serve as information logbook (for authorities), for technological benchmarking (corporate) and for data analysis (by researchers).

The global patent index (GPI), a database governed by the EPO, was used in this thesis (see section 7: "data construction"). It contains over 80 million patents from EPO's worldwide patent data. Moreover, EPO's databases contain patents from more than 90 granting authorities (GPI introductory guide; EPO, 2013). It is frequently updated and offers expert services in term of the level of detail that can be made in its search queries. On top, it allows for visualization and descriptive statistics of the data to be extracted.

Patent documentation is a rich source of information, qualitative on the specifics of the invention: claims, drawings descriptions, bibliography, and quantitative as they hold accounts of useful details such as: among others, the applicants, inventors, their nationalities, application date, citations and technological class. Also the increasing availability and globalization of patent data has contributed to them being used more often in research. Patent statistics are a useful tool and lend themselves to econometric analysis, unfortunately they are not stored in a user-friendly way in terms of quantitative analysis. Furthermore, the usage of (forward) citations has provided more profound insights in the value of patents and allow to control for their quality (Nagaoka, Motohashi & Goto, 2010).

Different from most studies concerning patent information this thesis addresses special interest towards inventors and their nationality. We attempt approximating open innovation (section 5.7) or collaborative intensity with external parties by constructing an indicator based on the inventor nationalities. Inspiration was found in the indicators for technological internationalization proposed by Guellec & Van Pottelsberghe de la Potterie (2001). A higher propensity of different nationalities (on an applicant's average patent in a certain year) is assumed to converge with more collaborative practices and more open innovation. Henceforth, this indicator will be used to explore the role of open innovation in the dynamics between innovation and firm performance.

Revisiting patent statistics, they are frequently used in quantitative analysis on innovation activities. Most commonly patents are treated as the outcome of the innovative process by academics. Innovation which is the motor of firm performance (here: labor productivity). In those studies, measures using patent counts have long been the dependent variable in the knowledge production function, and as an independent variable in the general production function (Grilliches, 1990). Moreover, knowledge stock measures have been constructed from them in similar fashion (Nagaoka et al., 2010).

Patent statistics as a proxy for innovative success are far from perfect though, and can possibly lead to misleading conclusions (Nagaoka et al., 2010). As Rogers (1998) states the ultimate key measurements of innovative success are those represented by firm success, such as (increased) profits, revenue growth and share performance which can be attributed to certain innovations. It already hints at large discrepancies in the market value of different patents representing innovations (Hall, Jaffe & Trajtenberg, 2005). Furthermore, the discussion of the patent paradox in the appropriability subsection learns that patents are not only used to protect valuable innovation, but also for strategic purposes. At least to some degree simple patent counts are therefore inadequate proxies of innovative success.

Consequently, researchers improved upon the simple patent counting practice by correcting for the quality of those patents (Hall et al., 2005). In fact, the quality of patents is represented by their market value which appears to be distributionally quite similar to the respective citations of a patent. Many patents are worth (almost) nothing and a few are highly valuable (Nagaoka et al, 2010). Similarly, Figure 6 illustrates many patents have very few citations while only a small fraction of patents is heavily cited (Aghion et al., 2002). Hence unsurprisingly, empirical research

has been able to confirm a positive association between technological importance, market value and (forward) citations of a patent (Nagaoka et al., 2010; Vancauteren et al, 2014).

Forward citations are not the only patent quality indicator. Other indicators involve backward citations, patent family size, and composite indices and are proven relevant. Although forward citations are found to be the most important indicator for drugs (Lanjouw & Schankerman, 2004, in Nagaoka, 2010). It is therefore the preferred measure to be used here. Forward citations nevertheless suffer from truncation problems. With fairly recent data, patents from earlier years will have to much relative weight because they simply have been around longer. Whereas younger patents suffer from truncation problems because more of their citations are yet to be made in the future (Nagaoka et al., 2010).

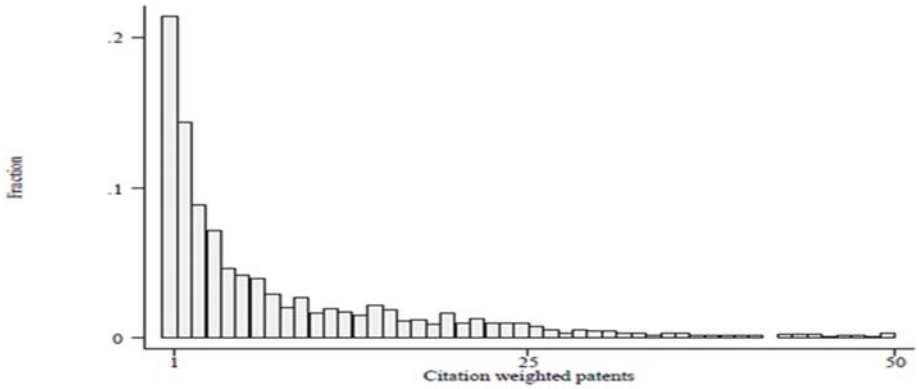


Figure 6: Citation weighted patent fractions. Aghion et al., (2002).

Concerning this master thesis simple patent counts will be used. Because of the relatively recent data set (2006-2012) this can be defended. Though honesty requires stating truncation problem can be dealt with. Furthermore, patents were not citation weighted because we were unable to extracted them in a practical manner from EPO's GPI database (2016), which is unfortunate. They could be extracted individually on a per patent basis but not in an aggregated fashion. Yet, this might not be an insurmountable hurdle. For example, Sanyal (2014) proved results were similar and robust interchanging simple count with citation weighted counts in a similar study for the Netherlands. Furthermore, using simple count was standard practice for a long time. The limitation remains, though, that result might not be as accurate as they could have been. The next section sets out to formulate the third hypothesis discussing the relationship between innovative input and innovative output.

5.6 The Effect of R&D Investments on Patenting

Discussing the effect of R&D investments on patenting pertains to the relationship between innovative input and innovative output (the knowledge production function). As outlined in the previous subsections both will be represented by R&D intensity (R&D per employee) and simple patent counts respectively. Modelling their relationship is the last stone to building the innovation-performance framework similar to the likes of the celebrated framework by Crépon et al. (1998). In the literature consensus exist about a positive and significant elasticity patents and R&D, yet estimates of the strength diverge (Sanyal, 2014). In accordance, the following research question is formulated:

"Is there a positive relationship between innovative input and innovative output?"

And subsequently this research question is met by the following hypothesis:

H2: There is a positive relationship between innovative efforts and innovative output

Mathematically this translates to:

$$(patent\ application\ counts) = \beta 0 + \beta 1 * Predicted[\ln(R\&D\ intensity)] + \beta 2 * X + \varepsilon$$

Where R&D intensity is estimated based on its determinants rather than using its observed values, X is the vector of independent control variables and ε the error under assumption of normality. If this relationship is supported empirically together with hypothesis 1, then the link between innovative efforts and firm performance can be established for the Belgian pharmaceutical industry. Other than channeling innovative efforts through innovative output, R&D intensity will also be imputed in the performance equation (hypothesis 1) to test for a direct linkage, as outlined in the that section. Yet, a significant positive coefficient could also indicate simple patent counts are an incomplete measure of innovation (Artz et al., 2010).

Concerning the vector of independent variables, it includes the usual control variables for size, competition and firm age for reason similar to those explained in the determinants of R&D subsection. It also includes the open innovation indicator. Furthermore, specifications with distributed lag modelling of R&D expenditures and autoregressive modelling of historical patent counts will be attempted. These stock variables are theorized to represent persistence of innovation. Persistence that can be explained by building up expertise, complementarity of innovations, broader technological opportunity and increased barriers of entry (Sanyal, 2014).

Returning to the patent citation discussion the distribution of the market value of patents has to be elaborated, since it is the dependent variable. Hall et al. (2005) and Aghion et al. (2002) noted the distribution of the market value of patents is highly skewed, and therefore inadequately represented by a normal distribution. Also patent data make up a discrete dependent variable with a significant proportion of zero observations, meaning no filings occurred (patent population is fully observed in our dataset). Together these properties suggest linear OLS regression is flawed. This means the resulting bèta estimates are inefficient and biased. Linear OLS regression should therefore be replaced by a more fitted estimation model.

Studying innovation and productivity of Dutch firms, Vancauteren et al. (2014) examine using a Poisson model or Negative Binomial model which are frequently used when count data make up the dependent variable. Based on the log likelihood ratio test (H0: mean = dispersion) outcomes produced by the Negative Binomial distribution are to be preferred. Yet, taking into account the large proportion of zero observations, Sanyal (2014) elaborates by expanding this model into a Zero-Inflated Negative Binomial model (ZINB). It gains credit by succeeding to improve on similar Poisson models by correcting for overdispersion (conditional variance > conditional mean) which it does by introducing unobserved heterogeneity for each observation (Sanyal, 2014). In fact, this heterogeneity relates to different propensities to patent and different abilities to come up with inventions. A zero inflated model assumes excess zero observations are created by a second underlying process. Inasmuch, ZINB improves upon regular negative binomial estimation by acknowledging the excessive proportion of zero observations which are caused by a variety of reasons (for elaboration on mathematical implications see Sanyal, 2014, p118 onwards).

Concerning the empirical results Sanyal (2014, Dutch pharmaceuticals) and Vancauteren et al. (2014, Dutch firms in general) establish a robust positive relationship between R&D expenditures and patenting. Nevertheless, the strength indicated by the bèta coefficients is less than unity, suggesting decreasing returns to scale concerning expected innovative output. Though Crépon et al. (1998) found an elasticity close to one, which is occasionally supported by others. Furthermore, Lopez (n.d.) also ascertains decreasing returns to scale with respect to firm size, implying a less than proportionate increase. He posits smaller market participants make more use of the patenting system in order to compensate for their lack in market power. Though this argument is in conflict with portfolio theory and strategic patenting practices favoring larger firms as predominant users of patents. The beta coefficient of the size variable will hence be interesting.

Nevertheless, in the same study Lopez (n.d.) referred to other studies arguing that corporate size does in fact influence the propensity to patent (Hussinger, 2005). In addition, Hussinger (2005) showed the perceived effectiveness of secrecy as an appropriation mechanism, previous patenting practices (persistence), and the industry context are significant predictors of the propensity to patent. Moreover, patents are more frequently used in product innovation contexts. Other studies cited in Lopez (n.d.) found capital intensity and exporting activities also contributed to the predicting power. If one is allowed extrapolate the significance of exporting activities to collaboration and open innovation an argument can be made for firms engaging in open innovation being more likely to file for patents. Yet, a true underlying relationship could be mitigated by ownership-related issues formulating IPRs for a collaborative invention. The next subsection looks at open innovation in more detail.

5.7 Literature: Open Innovation

In the past it was common practice to rely on in-house research and development to come up with new ideas and products. Beliefs were that company researchers were perfectly qualified to provide sustained input to the product pipeline. They were even deemed best at doing so. Yet, in the last two decades diminishing returns to R&D and subsequent budgetary impediments urged firms to explore other options (Gilbert, 2006). More creativity was necessary. Subsequently, individual firms started collaborating more and did no longer only look for ideas internally but started to incorporate external sources of knowledge more and more (Enkel, Gassmann & Chesbrough, 2009).

Open innovation (OI) therefore is a relatively new concept and it was first introduced in the early 2000s by professor Henry Chesbrough. "It is the purposive inflows and outflows of knowledge to accelerate internal innovation and expand the markets for external use of innovation. As a consequence, technology is advanced at an increased pace and value is created". (Chesbrough, Vanhaverbeke & West, 2006).

There is an increasing range of situations where the OI-concept is regarded as applicable, of which established examples are bio-tech or crowdsourcing. Basically, three distinct underlying processes exist: the outside-in, the inside-out and the coupled process (Enkel et al., 2009), of which the first two are adequately represented by Figure 7. The outside-in process can be defined as expanding the firm's own knowledge base through external sourcing, integration of the supply chain and innovation networks. It acknowledges not all smart people work for the firm (Chesbrough, 2003 in Enkel et. al., 2009). On the contrary the inside-out process tries to valorize the internal knowledge base through non-traditional channels, for example, by licensing IP or the creation of a spin-off. This means the locus of exploitation is shift outside the firm. Additionally, market access is fastened and improved. The third process, the coupled process stresses collaboration. Partnering firms want to promote their both their knowledge base and commercial abilities. Typical constructions are cooperative research contracting, joint ventures and alliances (Enkel et. al., 2009).

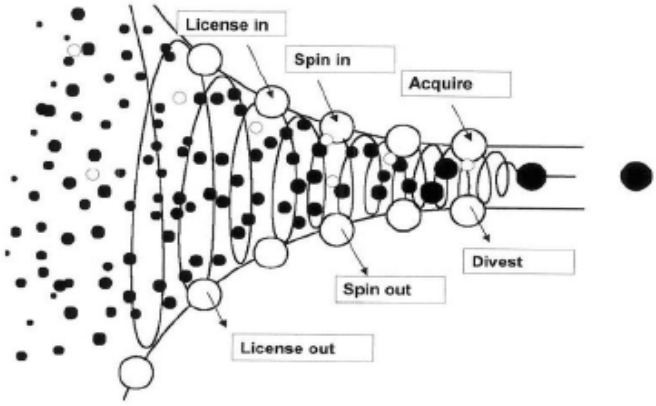


Figure 7: Open Innovation according to Kirschbaum

The last couple of decades, communication became easier and less costly, specialization increased and consequently benefits combining forced became more attractive (Enkel et. al., 2009). Moreover previous sections outlined the properties of ideas and knowledge as a public good (Lipsey & Chrystal, 2011). It therefore diffuses across borders easily which allows firms to tap information from outside their direct environment. It strengthens the case for cooperation (Guellec & van Pottelsberghe de la Potterie, 2001). Nevertheless, cultural, scientific and regulatory barriers are capable of posing real problems and for pharmaceuticals they do so in practice (Munos, 2010).

Clearly, open innovation is not all red roses. Its practices have to be managed properly. Next to the mentioned barriers, internal resistance to change is also plausible (Enkel et al., 2009). Particularly outsourcing activities can be sensitive. Not to mention there are definite risks to open innovation. The eventual loss of core competencies, for example, threatens the survival of the firm. But also loss of knowledge, loss of control, higher managing complexity and subsequent rising coordination costs are frequently problematic (Enkel et al., 2009).

Notwithstanding those risks the fact that several big pharmaceutical firms such as Eli Lilly, Johnson & Johnson, Novartis, and GlaxoSmithKline engage in OI, proves the industry sees potential in looking beyond corporate borders (Munos, 2010). Chesbrough et al. (2006) theorize OI is a plausible source of a competitive advantage. In this master thesis we therefore formulate the following research question:

"Referring to the relationship between innovation and firm performance, are firms that incorporate an open innovative approach as part of their strategy more productive?"

To accomplish answering this research question in an exploratory fashion, an open innovation indicator is inserted in all three stages of the regression analysis. The focus will be both on its significance and the direction, which is expected to be most prominent in the innovative input-innovative output equation. As a consequence, the following hypothesis is formulated.

H4: The coefficient of OI indicator entering the three equations is positive?

Because OI is a relatively new concept, only limited measurement approaches are around and only key performance indicators are known (Enkel et al., 2009). Besides, the literature shows mostly a qualitative approach is used. We therefore turn to Guellec & van Pottelsberghe de la Potterie (2001) who analyzed the internationalization of technology with patent data in order to construct a similar indicator with the available data.

Internationalization means more resources cross borders. This involves knowledge, the people generating inventions and the ownership of these inventions (Guellec & van Pottelsberghe de la Potterie, 2001). On top, Belgium is a small country with an open economy and a good portion of multinationals. Inasmuch it probably experiences this exposure more intense. Consequently, the extent to which a firm engages in open innovation can therefore be approximated by some cross border movements. Patents, for example, do not only convey technical details but among others list the applicants and inventors. In this study we follow the approach of Guellec and van Pottelsberghe de la Potterie (SHII indicator, 2001) and focus on the nationalities of the inventors.

OI, or else collaborative intensity external parties, will be approximated with by constructing an indicator based on the inventor nationalities. A higher propensity of different nationalities is assumed to converge with more collaborative practices and more open innovation. Henceforth, this indicator will be used to explore the role of open innovation in the dynamics between innovation and firm performance. It is constructed by calculating the propensity of nationalities other than the Belgian nationality on an applicant's average patent in a certain year for a specific firm (panel data):

$$\mathrm{OTHinventorPct\ i,t} = \frac{\left[\sum \mathrm{p} \frac{\#\ \text{foreign inventors}}{\text{Total}\ \#\ \text{inventors}}\right]}{Total\ \#\ patents}$$

Where p represents a specific patent, i a firm and t the year. Thereby, constructing an observation for each combination i,t: for each firm each year.

Before continuing with empirical research we summarize these research question and hypotheses, depict those relations in a conceptual framework and elaborate on the data construction process.

6 Hypotheses

6.1 Summary

Pharmaceuticals have been improving the quality of life drastically over the past couple of decades. Heavy knowledge-intensive research and development investments made it possible to constantly propel new innovations through the pipeline. Together with the United States, Belgium is an important innovative hub. It has a unique ecosystem combining the virtues of big pharma with biotech SMEs, academic research, an open economy and government facilities (Abrahamsen et al., 2011). Yet, recently progress seems to be stagnating associated with a falling return on dollar invested in R&D (Khanna, 2012).

Nevertheless, the ability to sustain a flow of innovation is essential for the profits of pharmaceutical firms, causing many of them nowadays to switch to survival mode (John Lechleiter, CEO Eli Lilly, 2010). Furthermore, innovation enhances long run total factor productivity and it is a key contributor to economic growth (Solow, 1957; Mohnen & Hall, 2013). This positive relationship between innovation and productivity has already been demonstrated for manufacturing firms (e.g. Crépon et al., 1998; Vancauteren et al, 2014) and for the pharmaceutical industry in particular (e.g. Sanyal, 2014).

In similar fashion to those empirical studies this master thesis will also investigate this relationship. We therefore formulated following central research question:

"Is there a significant positive relationship between innovation and firm performance (productivity) in the Belgian pharmaceutical industry?"

Crépon et al. (1998) have shown a good attempt to modelling the intricacies between innovation and productivity is breaking the process up in a recursive manner. This disentanglement is achieved by estimating innovative input, innovative output and productivity sequentially. Many would follow to do the same. Moreover, their approach will also be pursued here.

First, the tangible firm specific factors that are determinant for firms to invest in R&D will be investigated. Subsequently, R&D expenditures will be estimated and further used in the empirical research to overcome bias introduced by non-R&D-reporting firms. The first question to be addressed will therefore be:

"What tangible firm-specific factors are determinant for firms to invest in R&D?"

This corresponds to empirically testing whether the coefficients on the hypothesized determinants are statistically different from zero. The statistical null hypothesis is therefore that those coefficients are in fact zero (H3). As a dependent variable research capital intensity will be used. As outlined in section 5.3, particular interested is directed towards the coefficient on firm size because of potential (dis)economies of scale are theorized. Furthermore, the coefficients on variables measuring competition, capital intensity and firm age will be investigated. Besides, Heckman estimation (1979) is the econometric estimation method of choice since it has been

concluded robust, unbiased and efficient in preceding empirical literature (Vancauteren et al., 2014; Sanyal, 2014).

After estimation R&D-intensity, it will subsequently be imputed as main regressor testing to next hypothesis (H2):

"There is a positive relationship between innovative input and innovative success?"

Here, patent applicant counts are used to construct the dependent variable, and R&D-intensity is regressor along with control variables similar to its determinants. A consensus exists elasticity of innovative output with respect to innovative input is positive. Nevertheless, estimates vary substantially and are deemed to depend on the method, the measures and the industrial context. Most of the time decreasing returns to scale are concluded (e.g. Sanyal et al., 2014). Yet, the assumption of constant returns to scale cannot always be rejected statistically (e.g. Crépon et al, 1998).

To conclude the CDM framework innovative output is linked to firm performance (H1). New products eventually lead to growing revenues, profits and return on equity. In that respect simple patent application counts surrogating those new products are linked to labor productivity (value added per employee). The goal is to confirm a positive relationship as suggested by the empirical literature (e.g. Crépon et al., 1998; Sanyal, 2014; Vancauteren et al., 2014). It translates into the following hypothesis:

"There is a positive relationship between patent application counts and labor productivity as measured by value added per employee?"

It means assessing the direction of the coefficient on patent application counts and its statistical significance. Though patent counts as a measure of innovative output is subject to a lot of critique. Especially its unintended strategic uses (inflating counts) and inability to represent the underlying value of the invention are cumbersome. Consequently, they do not fully represent innovative output in an adequate way (Lopez, n.d.). Part of this process might therefore be captured by other variables. A likely candidate to absorb this effect is the proxy of innovative efforts. Yet, should this proxy be statistically positive and statistically significant from zero it could also indicate a direct channel through which efforts increase productivity (Sanyal, 2014).

To conclude this master also tries to meet contemporary developments in the industry. Because of declining blockbuster drugs discoveries and increasing costs firms have increasingly been looking at open innovation for repowering their product pipeline (Khanna, 2012). At first (big) pharmaceutical corporations adopted an approach of buying knowledge to strengthen their base. Nonetheless, more and more collaborative efforts are witnessed throughout the whole industry nowadays. Therefore, an exploratory investigation is followed through, testing whether OI efforts effectively enter innovation-performance dynamics positively. The fourth hypothesis (H4) is therefore:

"Firms engaging in OI are more productive?"

The open innovation proxy is constructed based on inventor nationalities. It adopts values between zero and unity reflecting the percentage of foreign nationalities in patent of a certain firm in a certain year. It will be imputed in all regressions, and is expected to enter the innovative input-output and the innovative output-labor productivity regression positively.

In the next section we will progressively enter all information in a conceptual framework. CDM methodology will make up its core with the interlinkages tested by the hypotheses elaborated in the literature and summarized above. For reasons of transparency all four hypotheses are again listed in Table 1 below.

Table 1: Summary of hypotheses

H1: There is a positive relationship between patent application counts and labor productivity as measured by value added per employee?
H2: There is a positive relationship between R&D intensity and patent application counts?
H3: What tangible firm-specific factors are determinant for firms to invest in R&D?
H4: Firms engaging in OI are more productive?

6.2 Conceptual Framework

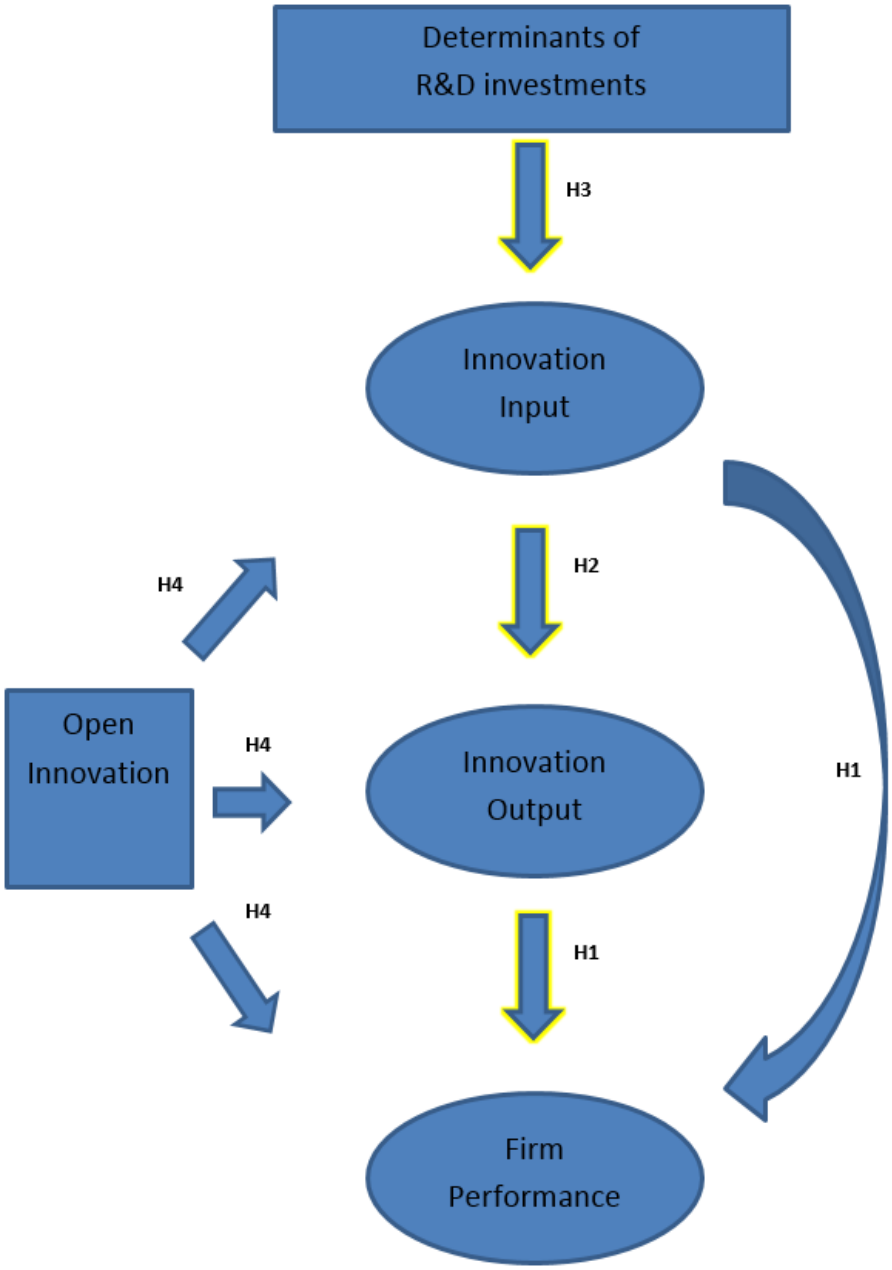


Figure 8: Conceptual Framework

7 Data construction

Before advancing to the empirical section the data construction process is elaborated. It provides the sources of the data and holds potential insights with respect to explaining several empirical results. The whole population of Belgian pharmaceutical firms is covered. Furthermore, the data is constructed by combining two databases (Global Patent Index & Bureau van Dijk), consulting legal documents filed by individual firms and checking several annual accounts. After describing the how the population was defined, patent & R&D gathering are discussed.

7.1 List of firms

The field of corporations used in this study is restricted to the (bio-)pharmaceutical firms geographically located in Belgium. It may already introduce bias because pharmaceutical firms can compete in the Belgian market without administering a legal entity in the country, which for example might be due to tax reasons (CRA, 2004). Nevertheless, boundaries have to be defined. After visiting pharma.be, the Belgian pharmaceutical representative in Brussels (November 2015), it subsequently became clear the population would be built using the members of three industry-representative groups: biowin.org, flandersbio.be and pharma.be. In all three cases they list their members on their websites. Accordingly, they were consulted and the members compiled.

This process resulted in a 'raw' list summing up 586 firms with some double counting. First of all, these double counts were eliminated to until only unique firms were retained. Next, all of them were checked for their engagement to pharmaceutical activities. Firms needed to have a reasonable amount of overlap with drug- and medical device markets. The latter was achieved by consulting the website of each respective firm often complemented by searches via google when in doubt.

In most cases prospecting the firm website was sufficient. Entities like Ernst & Young (EY), Vlaamse Instelling voor Technologisch Onderzoek (VITO) or Bank Degroof Petercam were eliminated this way. Because of the lack of a clear, delineated and objective criterion as to what is a medical device or drug, personal subjective evaluation was required. Which is, to some extent, likely to introduce bias in the sense that the 'true' industry is misrepresented. Although the direction or magnitude of this bias is ambiguous, we expect the error to be small. Eventually 215 firms were retained and subsequently data construction proceeded by compiling balance sheet variables for the firms surviving selection criteria. The sample consists mainly of large pharmaceutical firms (and their daughters), some diversified chemical and healthcare firms and a lot of (smaller) biotechs.

7.2 Balance sheet variables

In the second phase of the data construction process, the 215 firm names were run through the Bureau van Dijk (BvD) database. BvD is a global firm specializing in private firm information including financials and corporate structures. It uses data extracted from balance sheets and annual reports to construct a wide variety of financial information and firm specific data. After running the firm names through the database, unfortunately, not all names could be matched against BvD equivalents. This, in turn, narrowed the firm list down to 191 firms. In addition, another two firms were identified Luxembourgian by BvD. Consequently, these were deleted from the sample. As a result, the final sample consists of 189 firms.

For the remaining 189 firms a time window of seven years was investigated: 01/01/2006 – 12/31/2012. Observations were made on a yearly basis. Although a longer observation period could potentially be valuable the BvD student version accessible at Hasselt university is constrained to data dating ten years back. As a result, 2006 is a lower bound. Notwithstanding more years being a statistical trump, the transformation of R&D into innovative output and profit is a lengthy process, definitely in pharma. Recall, a typical business cycle for the development of a new drug from the discovery of a new active compound (active pharmaceutical ingredient, API) and patent application is 10-15 years (Khanna, 2012).

Among the extracted information there was ownership information such as the number of stockholders and their percentage of shares, which are subsequently used to consolidate the financials. These financials are constituted by profit (margin), revenues, debt, cost of sold good, cash flows, taxation and so on. Other data involve the amount of employees, classification of economic activities by EU standards (NACE-code), liquidity and solvency ratios and value added.

7.2.1 Ownership structure

To consolidate the financials the practical rule of thumb being that a mother-daughter relationship was recognized if a parent firm possessed at least 50 percent of the shares of the daughter was adopted. Only BvD non-paying access is limited. Which means it was not possible to export the ownership structure. Nevertheless, the ownership structure was available in a view-only format. Consequently, we went through the process of manually copying those figures. Ultimately this led to the consolidation of just one firm being the incorporation of Janssen Cilag's financials into Janssen Pharmaceutica effectively reducing the unique set of firms to 188.

7.3 Patent data

In the first and second phases of the collection process the foundation for econometric analysis was established. They will provide the input to for the determinants of R&D and labor productivity measures used in the empirical part. Progressively data to proxy innovative efforts and innovative success was retrieved. Starting with the patent information, these had to be accumulated for each company each year. Patent data were acquired via the European Patent Office (EPO), by consulting their Global Patent Index (GPI) database. GPI contains over 80 million patents from EPO's worldwide patent data. Moreover, EPO's databases contain patents from more than 90 granting authorities (GPI introductory guide; EPO, 2013). It is frequently updated and offers expert services in term of the level of detail that can be made in its search queries. Also it allows for visualization and descriptive statistics of the data to be extracted.

Via their expert services interface it was possible to conduct advanced search queries (https://data.epo.org/expert-services/index-2-2-5.html). For each firm (the applicant, 188) the BvD matched firm name was used. Subsequently the query was run 188 times, replacing the applicant name (Figure 9: search query for EPO's Global Patent Index.). The query withdraws information about the application date, inventor country of residence, applicant(s), applicant country of residence and whether or not the patent was granted. Thereafter the content was extracted and saved in an excel format for each firm. The individual sheets were then compiled and organized to resemble an entry for each firm each year.

Eventually the information used would be inventor nationality percentages (OI) and patent counts for all combination of entity (i) and year (t). Unfortunately compiling forward patent citations in a similar way was not successful. Though, it was possible to extract them on a per patent basis. But extracting and subsequently converting this information in a useful format was infeasible in practice. Nevertheless, their intended inclusion caused observations of other variables were only

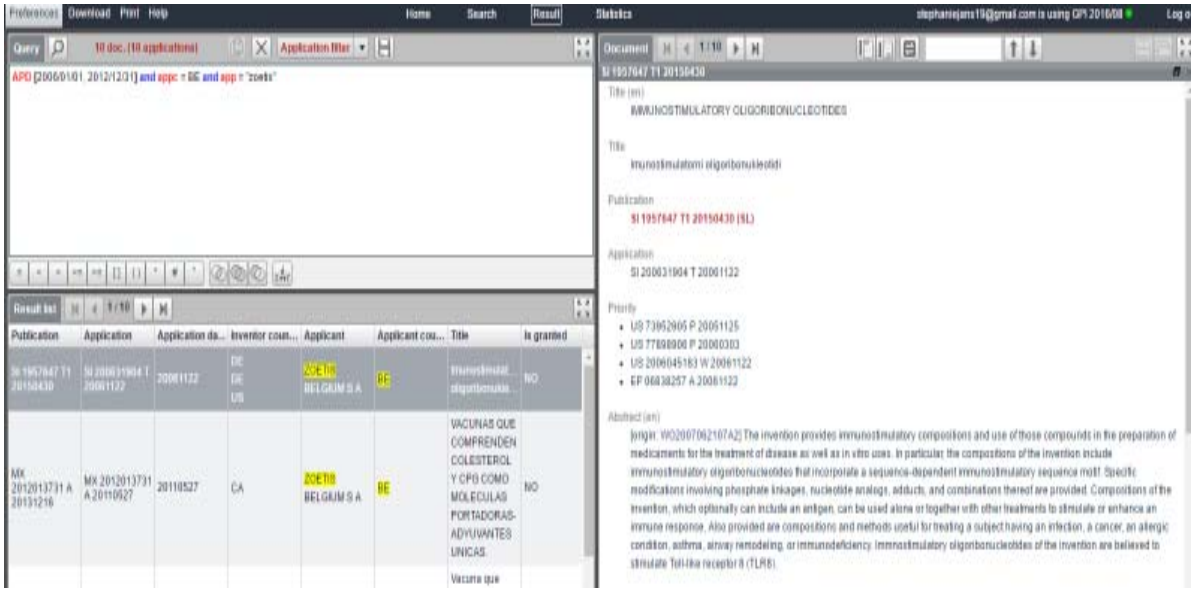


Figure 9: search query for EPO's Global Patent Index.

made up until 2012. The justification is the truncation error of the forward citations of young patents would be hard to deal with.

7.4 R&D data

Finally, the dataset was supplemented with the reported R&D expenditures of the firms applying for at least one patent within the 2006-2012 timeframe. This was done by going through the annual reports of the individual firms for each of the seven years. In Belgium these are compiled at the 'Kruispuntbank van Ondernemingen' (KBO) by the 'FOD Economie', a governmental institution. One can consult these annual reports freely by the following website: http://kbopub.economie.fgov.be/kbopub/zoeknummerform.html .

The entry for code 8021 (mutations during the accounting year, including acquisitions and produced capital) was used to search for R&D. As a result, 17 companies we effectively observed to report R&D. Since the population was not fully observed due to non-reporting issues and not investigating non patenting firms, the regressions will use an estimated figure of R&D (based on its determinants) rather than the observed values.

8 Empirical analysis

8.1 Descriptive statistics

As introduction to the econometric analysis, we start with describing the data and summary statistics of the main variables of interest. The purpose is to depict their distribution as well as unfolding a context in which we can interpret the data. As outlined in the data construction part, 189 registered firms are retained in the sample (188 consolidated). Which should be approximately the full sample of biopharmaceutical firms in Belgium during the period 2006-2012. It means a maximum of 1316 (= 7 x 188) observations are made for each variable. First, overall descriptive statistics are discussed for the variables. Afterwards special interest is taken in the dispersion of the variable size and those at the core of innovation analysis: R&D and patent count.

8.1.1 Overall Descriptive Statistics

In order to choose the best functional specification and estimation method underlying the equation, it is useful to describe the data underlying the sample. Although, note, that the literature review already described those by consulting relevant literature. In Table 2 (below) the amount of observations, minimum, maximum, the four statistical moments and quartiles are described for the non-transformed variables. It allows to grasp their average level, consistency and overlap with the standard normal curve. Furthermore, their pairwise correlation is displayed in table 3.

stats	FirmAge	Patent~n	Compet~t	OTHInv~t	Gemidd~s	NettoT~e	RnDExp~s
N	1316	1316	712	220	905	967	427
max	10	1236	2.79e+12	1	7426	2.82e+09	1.22e+09
min	0	0	13.89186	0	1	-2.01e+07	0
mean	7.68617	7.297872	1.46e+10	.433005	156.2011	3.99e+07	3.45e+07
sd	3.337469	63.95151	1.57e+11	.3546199	663.8288	2.25e+08	1.64e+08
variance	11.1387	4089.796	2.46e+22	.1257553	440668.7	5.07e+16	2.70e+16
skewness	-1.133247	14.86787	13.88588	.1968986	7.764238	8.026502	5.56113
kurtosis	2.798373	253.6801	211.4948	1.662302	68.26919	71.77779	35.15678
p25	5	0	42482.45	.0189	11	546950	0
p50	10	0	248481.5	.43095	29	2528071	0
p75	10	0	2163157	.75	67	9808415	0

Table 2: Descriptive statistics for non-transformed variables.

As illustrated in Table 2 the variables that are expressed in monetary quantities are very large, on average. Moreover, their sample distributions do not resemble a symetrical standard normal. Rather they exhibit a substantial positive skewness (substantial: >1 as a rule of thumb) and kurtosis (=3 for normal distribution). This indicates the underlying data has an asymmetrical distribution with a long and fat right-end tail. Stated differently, there is more probability density to the tails (kurtosis) and more extreme values in the right-end tail (skewness) than one would normally expect. It is a conclusion that can be extended for the number of patent applications, the average number of employees and the inverse of the Herfindahl-Hirshman index which represents

competition. On the other hand the assumption of normality turns out to hold relatively well for firm age and the open innovation proxy.

Next, there are a lot of missing values. These influence the empirical results to the extent that those missing values exhibit systematically different features that influence the regressions. Using a panel data set already allows to control for some of the resulting (time constant) heterogeneity by tracking individual firms througout the year. It enhances information over the cross-sectional approach. Furthmore not all of these seemingly missing values result from the same process. For example, Net Added Value is not observed 1316 times because late entry or early exit of firms in the BvD dataset caused their entries for different years to be empty. On the other hand the 220 observation for our open innovation proxy mean not all firms filed for a patent. For those firms, not filing for patents, it could not be computed.

Subsequently, pairwise correlation between the individual regressors was examined. It sheds a light on possible multicollinearity problems. High pairwise correlations between regressors means the variables have very similar patterns to them. As a consequence variation in the dependent variables can not adequately be assigned to either one. This should not cause inconsistent estimates (repeated sampling). Yet it causes the standard errors to become large and hypothesis test scores to become very low and unreliable. As a result, high pairwise correlations threaten (OLS) regression analysis to become pointless. From Table 3 it is evident such dangers exist as indicated by the asterixes (sig < 0.001). Especially figures that tend to increase with firm size turn out to correlate highly. Similar results are obtained by regressing all variables in Table 3 on RnD expenditures, thereafter calculating the variance inflating factor (Mean VIF = 4.41, with Value Added and Employees major contributors).

	FirmAge	Patent~n	Compet~t	OTHInv~t	Gemidd~s	NettoT~e	RnDExp~s
FirmAge	1.0000						
PatentAppl~n	0.0743	1.0000					
Competitio~t	-0.3648*	-0.0131	1.0000				
OTHInvento~t	0.3724*	0.1625	-0.0141	1.0000			
Gemiddelda~s	0.0858	0.6802*	-0.0163	0.0834	1.0000		
NettoToege~e	0.0740	0.7186*	-0.0195	0.1243	0.9508*	1.0000	
RnDExpendi~s	0.1298	0.8380*	-0.0254	0.2113	0.8020*	0.8953*	1.0000

Table 3: Pairwise correlation of non-transformed variables

As explained multicollinearity should not be a problem for linear estimation per se. It renders signifiicance levels unreliable (e.g.: adjustment for SEs needed) and interpretability of the hypotheses less feasible if it is troublesomely high. But the estimates of population coefficients should not be inconsistent. Provided the pairwise correlations in table 3, regression results should therefore be interpreted with caution (Gujarati, 2009).

The more important problem for consistency is the violation of normality assumptions. They can in fact cause estimation to be inconsistent if $n \longrightarrow \infty$. To deal with this problem common practise

dictates a log transformation (for example, Vancauteren, (2014), course: Time-series analysis, Part II). Table 4 illustrates log-transforming certain variables does indeed improve their normality since skewness (=! 0) and kurtosis (=! 3) which have diminished. Nevertheless, firm age turns out to be better represented by its standard formulation and therefore we will proceed with the logarithms of all variables except that of firm age and the open innovation proxy which is interpreted as a percentage.

stats	LN_Fir~e	LN_Pa~en	LN_Com~t	OTHInv~t	LN_Gem~s	LN_Val~d	LN_RnD
N	1246	230	712	220	905	835	64
max	2.302585	7.119636	28.65752	1	8.912743	21.76081	20.92376
min	0	0	2.631303	0	0	7.488853	6.907755
mean	1.974445	1.839308	12.84199	.433005	3.402154	15.10346	16.08184
sd	.5871458	1.630645	4.054901	.3546199	1.556258	2.069361	3.6059
variance	.3447402	2.659004	16.44223	.1257553	2.42194	4.282257	13.00252
skewness	-2.015148	1.069206	.711293	.1968986	.4914075	-.0767509	-.2330981
kurtosis	6.322616	3.844683	5.04539	1.662302	3.837271	4.529502	1.995423
p25	1.94591	.6931472	10.65684	.0189	2.397895	13.80617	13.25076
p50	2.302585	1.609438	12.42309	.43095	3.367296	15.11084	16.20304
p75	2.302585	2.484907	14.58707	.75	4.204693	16.28641	19.93411

Table 4: Descriptive statistics for log-transformed variables.

Aiming to exploit the virtues of the normal distribution even further, a possibility exists to delete the upper and lower percentiles from the dataset. Namely, these most extreme values are usual suspects to colouring the third and fourth moment. Another feature of log transformation is that it generally introduces more linearity to the dataset. Here, this also holds true which can be asserted comparing Table 3 and Table 5. A higher collinearity might cause problems later on.

	LN_Fir~e	LN_Pa~en	LN_Com~t	OTHInv~t	LN_Gem~s	LN_Val~d	LN_RnD
LN_FirmAge	1.0000						
LN_Patenten	0.2384*	1.0000					
LN_Competi~t	-0.3683*	-0.4627*	1.0000				
OTHInvento~t	0.3250*	0.0899	-0.2877*	1.0000			
LN_Gemidde~s	0.3297*	0.5983*	-0.7668*	0.3003*	1.0000		
LN_ValueAd~d	0.2769*	0.6295*	-0.9081*	0.3211*	0.9225*	1.0000	
LN_RnD	0.2446	0.8755*	-0.5688*	0.2692	0.7069*	0.6653*	1.0000

Table 5: Pairwise correlation for log-transformed variables

To conclude the overview of summary statistic we look at the natural logarithm of three variables. which are mainly a combination of the basic set described above: capital intensity, labor productivity and R&D-intensity (Table 6). They all seem to fit the desired values for the statistical moments fairly well, with only kurtosis possibly out of line. Furthermore, calculating their pairwise correlations with the log-transformed specification (Table 5) similarly leads to some high but generally acceptable observations. Additionally, a visual depiction of those variables confirms their approximate normality (Figure 23 & Figure 24 in appendix). Yet, the histogram of R&D-intensity (Figure 10) appears out of line. Even though, its statistics do not immediately indicate such, two different R&D intensities seemingly exist.

stats	LN_Mat~N	LN_Val~N	LN_RnD~N
N	877	801	64
max	13.91917	17.03253	15.23187
min	3.415382	7.569309	1.418818
mean	9.23113	11.73421	10.84331
sd	1.793918	.7487928	2.554801
variance	3.218143	.5606906	6.527009
skewness	-.2723182	-.6053684	-1.688199
kurtosis	2.817179	9.429754	5.963645
p25	7.975526	11.30576	9.841098
p50	9.45246	11.81477	11.48704
p75	10.48251	12.19254	12.49433

Table 6: Descriptive statistics for selected composed variables.

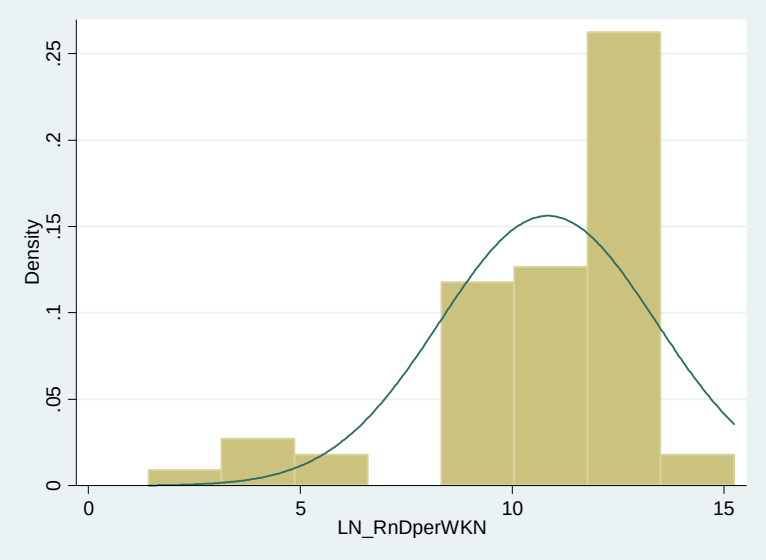


Figure 10: Histogram of R&D intensity

8.1.2 Firm Size

Firm size is one of the most important control variables. Sections 2 and 1 already reveal the pharmaceutical industry is an interesting ecosystem that is roughly comprised of a few large companies together with a lot biotechnological SMEs. Big pharma is said to have very large internal R&D projects while SMEs have been accredited with some radical innovation (Foxon, 2007). The influence of firm size as a determinant of R&D investments, hence, is heavily studied. Section 8.2 will accordingly investigate whether it influences the decision to perform R&D activities (probit part) and whether decreasing, constant or increasing returns can be supported for the Belgian context. Through R&D scale and intensity, firm size will inevitably also influence developments further downstream the innovation-performance equations. It is most likely to correlate substantially with several regressors and the respective regressands. It is therefore included in all three equations.

The literature identified two important proxies for firm size: the average number of employees and turnover. In accordance to Crépon et al. (1998), the number of employees is used in this thesis, as it seems to be the mainstream proxy. It can be argued less susceptible (in the short run) to several sources of variation and therefore superior. Nonetheless, we check whether both proxies plausibly lead to a similar conclusion here.

Summarizing statistics for non-transformed average number of employees confirms there are a lot

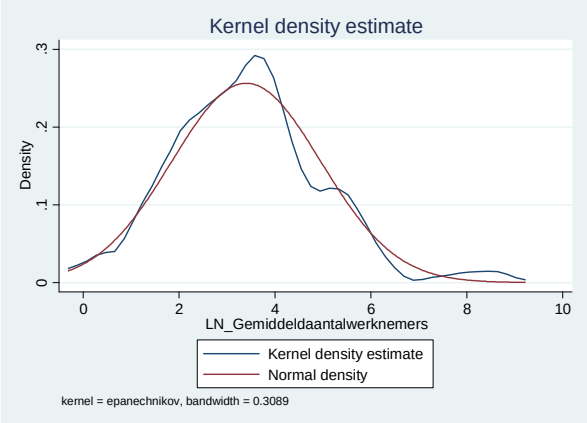


Figure 11: Kernel density of log-transformed average number of employees.

of relatively small firms and some large corporation. Moreover, tabulating its percentiles and surveying its four statistical moments authenticate this statement (Table 2). Coincidentally, the existence of distinct pools of firm sizes is illustrated by plotting the (smoothening) Kernel density of log-transformed average number of employees against the standard normal distribution (Figure 11). Not only do small versus large firms exist, but seemingly there is a third firm size. The smallest group has an average of 35 employees ($\approx e^{3.5}$), medium-sized firms 250 employees ($\approx e^{5.5}$), and large (multinational) corporations 3000 to 5000 employees on average ($\approx e^{8}$- $e^{8.5}$). These big pharmas are identified as Janssen Pharmaceutica, GlaxoSmithKline and UCB pharma.

Summary statistics of non–transformed turnover support the finding using the average number of employees. Stata (statistical package) reports a positive skewness of 6.25 and positive kurtosis of 44.7 which can largely be explained by the inclusion of big pharma in the sample. After their exclusion skewness and variance already reduce to 3.3 and 15.7 respectively. Subsequently log-transforming the complete data set (JNJ, GSK and UCB included) holds a skewness of -0.7 and kurtosis of 4.9 which is very much in line with statistics for the employee-based proxy (Table 4). Although the latter seems marginally better.

In the remaining discussion of the descriptive statistics section the attention will be shift to innovation. Since R&D data were only assembled for those firm reporting at least one patent over the course of seven years, patents will be treated first, and subsequently R&D will be elaborated.

8.1.3 Patents

In this master thesis simple patent counts are representing innovation output. The literature identifies its pros and cons, pointing out they might not only and fully represent innovative succes. Furthermore using forward citations as weights allows to incorporate the market value of the inventions underlying the patents (for example: Nagaoka et al., 2010). Still, similar studies using simple patents counts accomplished to establish robust result using them in a pharmaceutical environment (Sanyal, 2014). Missing out on forward citations should therefore not be an insurmountable problem.

Via EPO's GPI database patent application counts were obtained for all 188 consolidated firms in the sample. Table 7 shows 58 of them were found to apply for at least one patent during the 2006-2012 period. It lists these application counts and reported R&D expenditures for these firms cumulatively over those seven years. In line with the literature and the discusion on firm size above there are only a few companies registering a lot of patents (> 100, Table 7). Furthermore innovative succes as proxied by simple patent counts turns out to coincide with the level of innovative efforts, with six firms reporting over 100 million euros spent on R&D in the observed timeframe. A scatterplot confirms this conclusion. The three largest firms, Janssen Pharmaceutica, GlaxoSmithKline and UCB reported between 1 and 10 billion, and Ablynx, Galapagos and Bayer between 100 million and 1 billion euros spent on R&D.

Count of UniqueID	**RnD Exp**						
Nbr of Patents	**<1M**	**<10M**	**<100M**	**<1B**	**<10B**	**Not Rep**	**Grand Total**
no patents						130	130
1 patent		1				4	5
2-10 patents	1	2				15	18
11-100 patents		4	1			24	29
101-1000 patents				3	1		4
>1000 patents					2		2
Total	**1**	**7**	**1**	**3**	**3**	**173**	**188**

Table 7: patent application count in relation to RnD expenditures

Summary statistics (Table 2) show highly positive skewness and kurtosis with which we are accustomed by now. Thus far this problem has always been tempered by log-transforming the underlying data. However, patent application counts are a discrete nonnegative variable, whereas other log-transformed variables are continuous by nature. Therefore, application counts will not be log-transformed. It is a "count" variable and in accordance with the literature negative binomial count model will be used when modelling application counts in section 8.3 rather than log-transforming it and rely upon OLS estimation.

8.1.4 R&D

In the data construction section we explained R&D expenditures were only retrieved for those firms applying for at least one patent in the observed timeframe. As a result, about 400 annual reports were retrieved and consulted for R&D expenditures. Of those 400 annual reports 64 positive entries of 17 different firms originated, hence bringing about a majority of zero entries. As explained in the literature these excess zero observation are speculated to have two underlying processes: non-perfoming and non-reporting R&D. It is obvious some (probably generic) firms will perform none or very little R&D on Belgian soil. Further, as less than 20% of the firms with patent applications effectively reported R&D, the issue of non-reporting also appears to be supported. Section 8.2 confronts the existence of two driving zero-generating processes by adopted Heckman (two-stage) estimation.

The previous two subsections outlined the level of R&D activities generally corresponds with the size of a firm as well as with the number of patent applications. Although the small number of observation implies conclusions have to be drawn with more caution, especially if non-reporting firms have systematically different R&D expenditure patterns. Summary statistics for non-transformed variables are nevertheless similar to those for other continuous variables. Contrarily log-transformed R&D capital and intensity are slightly negatively skewed. It suggests a long left-end tail, though overall not substantially different from normality. Moreover, on first sight, kernel densities of Ln(R&D) and Ln(R&D/Employee) seem to confer augmented insight (Figure 12 & Figure 13).

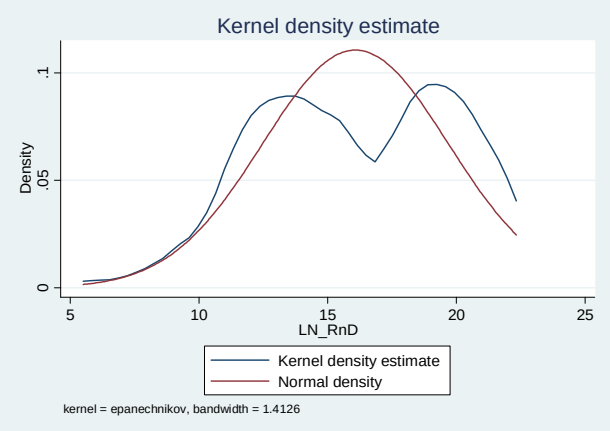


Figure 12: Kernel density of Ln(R&D).

Figure 12 indicates two distinct levels if R&D expenditures are reported: a couple of hundred thousand (≈e^{13}) and a couple of hundred million euros (≈e^{20}). It supports the big pharma-small biotech rationale. Furthermore, Figure 14 comprehensively shows more R&D reporting firms are observed towards the end of the timeframe. This is possibly explained by changing reporting regulations between 2006-2008, as it was witnessed the formulation of accounting codes (8021) changed slightly. Another explanation is changing subsidizing regimes as the Belgian governments' tax income with respect to pharmaceuticals changed from a surplus into a deficit within the timeframe, causing more R&D to be reported (Pharma.be, 2016).

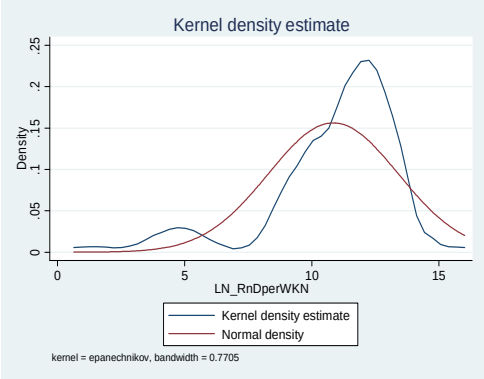


Figure 13: Kernel density of R&D intensity.

Finally, R&D capital intensity seemingly suggests two distinct levels as well (Figure 13). However, in this case the conclusion can be doubted. The density elevation to the left end is solely provoked by R&D reported by Sanofi Belgium, making it a potential outlier and thus a candidate for elimination from the dataset. Sanofi R&D expenditures drop from 500 million euros in 2006 to several ten thousands in the for all years thereafter. This firm, which is headquartered in France, might have changed its locus of research elsewhere, into another legal entity (Genzyme) or abroad. Another possibility is that a mistake happened in constructing the data. Yet the latter seems unlikely since 2006 expenditures are in line with those of other similar firms and 2007-2012 reported expenditures are consistent. Other than Sanofi R&D-intensity percentiles demonstrate at least 3000 euros (≈e^{8}) per capita are invested when R&D expenditures were reported.

Figure 14: RnD Reporting firms per observed year.

8.2 Relationship: economic variables – R&D propensity and level

The first part of the three-equation model estimates the R&D intensity. The literature identified six plausible determinants: size, firm age, capital intensity, competition, lagged innovation and open innovation. They will make up the base case formulation. Size, age and competition are Schumpeterian determinants classically included in studying the determinants of R&D (Crépon et al, 1998). Physical capital intensity is included to resemble budgetary constraints (Hottenrott & Peters, 2011). Lagged innovation will be accounted for by a dummy variable. It indicates whether the firm applied for any patents the year before thereby vocalizing the success breeds success hypothesis (innovative persistence; Vancauteren et al., 2014). Finally, the open innovation proxy constructed in the philosophy of Guellec & van Pottelsberghe de la Potterie (2001) was included to explore the effect of increased collaboration on R&D.

Because R&D observations are believed to result from two different processes, a sample selection model is used confirming to previous empirical research (for example: Crépon et al., 1998). Here, we rely upon two stage tobit estimation (Heckman, 1979). It first estimates whether a firm decides to invest in R&D. If R&D expenditures are postulated, Heckman proceeds to estimate its intensity (the fortitude of commitment). Although its validity was already established in similar studies, the findings using Heckman estimation will be checked for their robustness by using other estimation methods. Equivalently, meaningful results could also have been obtained by estimating the level of R&D activities in the second stage.

Initially, the base case regression corresponding to those specifications was augmented by including time dummies. Including these time dummies allows to take advantage of the time-series nature of panel data. Ideally, they filter out year specific heterogeneity. However, Stata was not able to reach convergence (log pseudo likelihood ratios turned out not to be concave). When this happens multicollinearity is often the problem. As a consequence, pairwise correlation was revisited. Because of a high correlation (0.66) between physical- and R&D capital intensity, physical capital intensity became a first suspect. What is more, its z-score is extremely high ($z = 15$) pointing in the same direction.

Even so, physical capital intensity is preserved in the base case scenario after more thorough investigation. Namely, the individual variance inflating factor (vif) of 1.39, obtained after OLS estimation, is reassuring physical capital intensity poses no problem. In addition, the overall variance inflating factor of 3.89 is within the acceptable region statistically. Furthermore, its sign conforms to theory and it is consistent using different specifications and regression models.

On second thought, the time dummies were excluded as overfitting with dummies is a common problem. At the same time the advantages of panel data are therefore abandoned. In spite of deleting these dummies convergence still could not be reached. Stata automatically omitted firm age. Although its vif of 1.5 does not raise suspicion, firm age was eventually deleted as a variable. This resulted in robust findings and they are elaborated further downward this section. Withal

valuable information is potentially lost concerning the maturity of Belgian pharmaceutical firms, since neither entry and exit dummies or firm age.

Instead of pointing at collinearity which can be alleviated by increasing the number of observations in the sample, it is also possible to reverse the rationale: multicollinearity problems are caused by the lack of R&D observations (annually). The explanation is twofold. First, analogue to overfitting dummies, the number of variables included in the regression relative to (annual) observations is fairly large. Building on the first argument, secondly, Stata stigmatizes variables that lack variation. In that respect, the software has a hard time to assign variation in the dependent variable to independent variables. The inability to disentangle effects is in essence the same problem encountered with multicollinearity. Because time dummies and firm age very much comply to this logic their omission is justified. The base case model is hence reduced to five variables.

Additionally, Sanofi (possible outlier) was retained in the sample to maintain as many observations as possible. In fact, the omission of Sanofi from the dataset did not prove to change results substantially. It led to a similar prediction of the observations and statistical moments of R&D-intensity but with a smaller standard error. As such, its inclusion does not threaten statistical inference beyond the significance of the results. Moreover, it is difficult to confirm Sanofi's status as an outlier due to the small number of observations.

With reference to the two-stage estimation, the consulted literature did not proclaim different vectors of independent variables for the decision and intensity, a priori. Nevertheless, the first stage estimation vector (probit part) was expanded with four variables constructed from the BvD database: tax-assets ratio, liquidity (current) ratio, solvability ratio and ln(value added per employee). More (relevant) variables generally allow for a better estimation and the pairwise correlation matrix does not hint at potential problems. Furthermore, an exclusion restriction (different vectors) is usually required to obtain trustworthy estimates. Arguments are similar to those of instrumental variable regression. Extra variables alleviate the collinearity problem and are assumed to help estimating the latent variable better. Conceptually, value added per employee potentially renders insight on specialization and cooperation (Fritsch & Lukas, 2001), tax-assets ratio is particularly interesting for pharmaceuticals because the government is heavily involved in tax cuts and subsidy schemes (pharma.be, 2016) and liquidity and solvability ratios can provide information on the influence of corporate financial health (Bodie et al., 2013). Although their effects on R&D intensity are plausible as well, they are not underpinned in thesis master thesis and likewise not included in the second stage.

Concerning the results, the base case specification [1] demonstrates a negative significant influence of the intercept, firm size and competition on R&D-intensity (Table 10). Vice versa, the first lag patent dummy, capital intensity and open innovation exhibit a positive and significant sign. None of the regressors is a statistical determinant (at 10%) of the decision to perform R&D (both positive and reported). Yet, the pseudo-R2 of the probit first stage (0.695) and Chi2 (71.06; 9 df) of the log likelihood ratio test (LR test: H_0: ρ = correlation[error selection equation; error regression equation] = 0) affirm Heckman estimation is appropriate. Further, note, sigma (σ) the

standard error of the residual in the second stage regression part is 0.641 which is fairly small, indicating a good fit. It is confirmed by the Wald Chi2 statistic that is clearly rejecting the null hypothesis (H_0: all $B_i = 0$; 5 df; 509.95). However, lambda (sample selection bias: $\lambda = \rho\sigma$) is statistically insignificant. Given rho and sigma (small), this is most likely due to the small number of observations rather than non-existence of selectivity. The outcomes and statistical tests of specification [2], which replaces the competition variables based on industrial concentration (HHI) with an indicator based on profit margin (~Lerner), largely confirm the results of [1] and might even provide an improvement.

In view of the conceptual model established in section 6.2, the emphasis of analysis is on the second stage of the Heckman estimation: the regression part, diagnosing the determinants of R&D intensity. Namely, R&D intensity approximates innovative effort better than merely looking at the decision whether or not to invest in R&D. The first stage still accounts for the determinants of the decision to invest in R&D though. At the same time, using separate stages certifies unbiasedness of the coefficients in the second stage through dealing with the sample selection problem (linear OLS = biased: $\rho \neq 0$).

Revisiting the coefficients, the negative sign on firm size suggests R&D expenditures do not rise proportionally to corporate size ([1]-[2]; [3]: $H_0 : B_{size} = 1$ rejected at 5%). It hints diseconomies of scale with respect to R&D exist, consistent with findings for the Dutch pharmaceutical industry (Sanyal, 2014). Although the strength of these diseconomies cannot be pinpointed precisely ([1]-[2]: 0.381 – 0.728 compared to unity). Besides, practices such as venture capital may cause the R&D intensity of SMEs to be inflated temporarily. Moreover, various other reasons that do not pertain to risk-adjust return on the portfolio of R&D projects might cause firms to invest proportionally less as size increases. Notwithstanding diseconomies of scale, overall, the role of firm size as an indicator of the decision to invest in R&D is affirmed (for example: Artz et al., 2010).

As suggested by diseconomies of scale, smaller corporations can eventually be expected to produce more innovation since their relative effort is higher. As such, they are an important building stone of the Belgian pharmaceutical industry. On the other hand, the consistent significant positive entry of capital intensity broadens the perspective. If measured correctly, it is interpreted as R&D intensity being higher if the budgetary constraint is lower (thus the funds pool larger; Hottenrott & Peters, 2011). Accordingly, with steeply rising R&D costs, pooling funds has been an important reason for mergers and acquisitions in the past (CRA, 2004; Khanna, 2012). Seemingly, empirical and practical findings are hence contradictory.

Subsequently taking a look at competition, the negative sign on the log of the inverse of the Herfindahl index implies less competition yields greater innovative efforts. Less competition equals greater market power and price setting possibilities, inasmuch it is an incentive to propagate new inventions. The prospect of monopoly profits can therefore explain augmented R&D efforts. A logic that is reinforced by the heavy use of patents as an appropriability mechanism. Conversely, higher industrial concentration parallels less and larger firms which balances with the finding diseconomies of scale. What is more, the insignificant coefficient of price margin is reason to

question results of the Herfindahl-Hirschman index [2]. Further, recall the literature did not univocally agree on the sign of competition. Therefore, it is concluded the effect of competition on innovative effort is plausibly non-linear in accordance the recent literature (Aghion et al., 2005).

Next, the consistent and significant positive sign on lagged patent dummy underpins the success breeds success hypothesis (Vancauteren et al., 2014) or at least innovative persistence. It means firms that applied for a patent in the past (here: lag = 1 year) more heavily invest in R&D. They want to sustain previous successes and are likely to have built up a knowledge stock. It enables them to have a competitive advantage. Furthermore, several years go over the R&D process and patent grants. Using distributed lag models on past patent applications and past R&D expenditures will hence deepen the understanding with respect to timing.

The last remaining member of the base case vector of independent variables is open innovation. The results invariably designate a positive significant effect on R&D intensity. Stated differently, an increasing average percentage of foreign nationalities on patents means an applicants' innovative efforts are higher. If these different nationalities adequately represent different collaboration with other innovative participants, practices of open innovation increase the research intensity.

Concluding the discussion of the results, the first stage estimation is observed. The t-values on the coefficients of the individual variables are as good as always insignificant. Ergo, the specification can be criticized. A first point is the insignificance of the four auxiliary variables that were not (sufficiently) underpinned by the literature. Their inclusion may therefore be questioned. On that matter specifications [1] – [3] do suggest solvency ratio (financial health) and tax-assets ratio might be viable (almost significant statistical tests). The other two may be omitted. In addition, tax-assets ratio has a very large weight in the investment decision. It is a cause to rethink including it in second stage which adds up with the observation tax regimes play a major role in the industry (Pharma.be, 2016).

After determining the coefficients of the vector of independent variables, they were used to estimate the true (latent) R&D expenditures and intensities for each firm-year combination in the dataset. Two different estimations were completed using the coefficients of specification [1] and [2]. They differ in their use of competition variable, yet the resulting estimated values are in line (Figure 15 & Figure 16; Table 8 & Table 9 below). Negative R&D-intensities were replaced by 0 observations (3 cases and 2 cases respectively). The statistical moments and percentiles are tabulated in table. Fitted R&D intensity is within acceptable boundaries to label it normally distributed (figures kernel density). The predicted means are substantially lower than that of the observed values though. Estimated means are 2100 euro per capita (HHI-based) and 9800 euro per capita (PM-based) whereas actual observations had a mean R&D expenditure of 51000 euro per capita (compare Table 6 to Table 7 and Table 8). Clearly, estimated lower estimated values imply that less intensive innovators are less likely to report their efforts. Another disparity is that 147 (HHI-based) and 123 (PM-based) observations are predicted in contract to the 64 that were actually observed. The shape of all three kernel density estimated are similar.

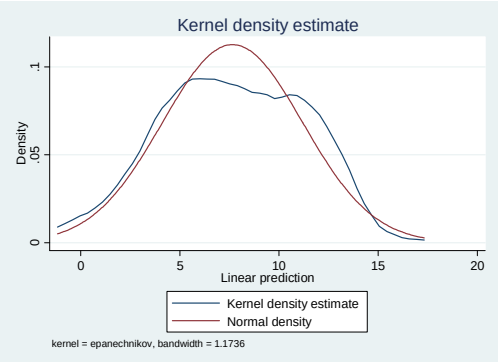


Figure 15: Predicted R&D intensity using the HHI based competition indicator.

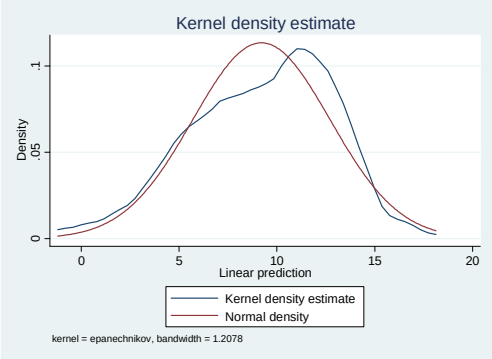


Figure 16: Predicted R&D intensity using the profit margin based competition indicator.

Linear prediction

	Percentiles	Smallest		
1%	0	0		
5%	1.921301	0		
10%	3.075664	0	Obs	147
25%	5.006764	.2685716	Sum of Wgt.	147
50%	7.530639		Mean	7.654537
		Largest	Std. Dev.	3.537978
75%	10.50754	13.54549		
90%	12.31027	13.6273	Variance	12.51729
95%	12.72125	14.16026	Skewness	-.1109246
99%	14.16026	16.15442	Kurtosis	2.282939

Table 8: Summary statistics of predicted R&D intensity using the HHI based competition indicator.

Linear prediction

	Percentiles	Smallest		
1%	0	0		
5%	3.385238	0		
10%	4.671109	.3363605	Obs	123
25%	6.594618	2.282022	Sum of Wgt.	123
50%	9.449016		Mean	9.20872
		Largest	Std. Dev.	3.513483
75%	12.18403	14.89507		
90%	13.03061	15.16333	Variance	12.34456
95%	13.28151	15.72807	Skewness	-.4316404
99%	15.72807	16.92552	Kurtosis	2.679932

Table 9: Summary statistics of predicted R&D intensity using the profit margin based competition indicator.

As an answer to hypothesis 3 the findings are summarized. First, size is found a determinant of R&D, relating negatively to its intensity. R&D expenditures do not rise proportionally to firm size. A similar negative impact on R&D intensity is true for competition based on the Herfindahl index. Although the profit margin indicator and plausible non-linearities are reason for nuance. Furthermore, R&D intensity rises if more funds are available and it is higher if the firm has experienced innovative successes in the past. The results also suggest that open innovative practices are associated with more intense innovators and that tax regimes deserve more attention in subsequent studies. Finally, in the following empirical sections the estimated R&D values will be used as regressors rather than observed values in order to address the sample selection bias.

Dependant Variable **Independent Variable**	Probit RnD=0/1	Ln of RnD per Employee	Probit RnD=0/1	Ln of RnD per Employee	Probit RnD=0/1	Ln of RnD	Ln of RnD per Employee	Ln of RnD per Employee
Specification number	[1]		[2]		[3]		[4]	[5]
Ln(Employees)	1.027	-0.619**	1.836**	-0.272*	1.027	0.381	-0.389	-0.316
	[0.861]	[0.304]	[0.766]	[0.144]	[0.861]	[0.304]	[0.298]	[0.232]
First Lag Patent Dummy	0.487	2.335***	0.681	2.000***	0.487	2.335***	2.838***	2.651***
	[0.750]	[0.782]	[0.900]	[0.709]	[0.750]	[0.782]	[0.871]	[0.627]
Ln(Capital Intensity)	-0.002	1.670***	-0.105	1.671***	-0.002	1.670***	1.125***	1.175***
	[0.189]	[0.110]	[0.186]	[0.085]	[0.189]	[0.110]	[0.201]	[0.187]
Ln(Competition)	-0.168	-0.337**			-0.168	-0.337**	-0.046	-0.062
	[0.334]	[0.133]			[0.334]	[0.133]	[0.083]	[0.061]
Profit Margin (as %)			-0.049	0.001				
			[0.033]	[0.009]				
Open Innovation	-1.207	3.249***	-0.171	3.561***	-1.207	3.249***	2.321***	2.684***
	[1.331]	[0.558]	[1.545]	[0.544]	[1.331]	[0.558]	[0.778]	[0.724]
Intercept	5.103	-5.671**	-34.316*	-9.516***	5.103	-5.671**	-1.996	-3.464
	[14.791]	[2.672]	[17.637]	[1.371]	[14.791]	[2.672]	[2.607]	[2.381]
Tax-Assets Ratio	-18.068		-16.244		-18.068			
Ln(Value Added per Employee)	-0.843		1.685		-0.843			
Solvency Ratio	0.031		0.091*		0.031			
Current Ratio	-0.143		-0.196		-0.143			
Log-likelihood	-15.613		-10.402		-15.613		-91.577	
Pseudo R2	0.695		0.793		0.695		0.176	0.387
rho	1.238 -> 1		0.043		1.238 -> 1			
sigma	0.641		0.418		0.641		1.568	
							[0.158]	
lambda	0.641		0.018		0.641			
	[0.690]		[0.505]		[0.690]			
N Observation	84		84		84		49	49
Censored	59		60		59			
Uncensored	25		24		25		49	
Estimation Method	Tobit II		Tobit II		Tobit II		Tobit	Quantile Regression
Stastistical significance indicated by stars (*: 10%, **:5%, ***:1%). Standard errors between brackets.								

Table 10: Determinants of R&D

8.3 Relationship: R&D – patents

The discussion on the determinants of R&D allows to cope with the problem of R&D observations. Estimating R&D-intensity, using its determinants, was the first step in sequentially progressing through the conceptual framework. The second step relies upon the estimated R&D values as a foundation to investigate the link between innovation (efforts) and firm performance. Inasmuch, it is at the center of all CDM-type models (1998). Section 8.3 empirically tests the relationship between these innovative efforts and innovative output. In the process, it deals with hypothesis 2 which advances a positive impact. If hypothesis 2 can be supported, innovative output will be tied to firm performance (H1) thereby sketching their interlinkage for the Belgian pharmaceutical industry.

A brief recap of the literature (5.6) identifies five relevant variables on top of R&D-intensity: size, firm age, competition, lagged innovation and open innovation. They are incorporated as control variables. Still, the coefficient on open innovation is of interest regarding hypothesis 4. Moreover, the discussion on patents in sections 5.4-5.5 implies a positive coefficient on size and autoregressive patenting observations can be expected due to the time series dimension. The directions of the influences of firm age and competition were not hypothesized. Apart from that, competition appears to be a complex construct with non-linearities and different manifestations have potential to offset one another. A modest endeavor to meet these difficulties led to different specifications using HHI- and profit margin- based measures of competition interchangeably. In accordance two different R&D estimates based on them where used (see 8.2 for their estimation).

Whereas the log transformation of variables that did not satisfy normality conditions was advocated in the descriptive statistics part (8.1) another approach was used. As outlined, patents counts are a discrete and highly skewed variable. As a consequence, it is appropriate to fit a count model to the data. The base case formulation opts for Zero Inflated Negative Binomial (ZINB) estimation [1] and was tested for its fit (Alpha and Vuong test statistic). ZINB is superior to Poisson models in case overdispersion (variance > mean) characterizes the dependent variable (Table 11; tested by alpha). It is superior to standard NB models when excessive zero counts are observed (82.5% of observations; tested by Vuong). Table 11 comprises both ZINB and NB specification and is augmented FE, RE and Pooled OLS estimation of the log-transformed dependent variable.

Throughout the regressions R&D-intensity shows a consistent positive influence on patent counts that is highly significant. The size of the effect of R&D intensity can be interpreted as an increase of 0.08-0.30 units of patent counts, on average, if the intensity doubles (+100%). This is seemingly low, yet it can be explained by the skew in the data: few firms filing for a lot of patents while many file no or very few patents (see 8.1.3). Still, a positive impact is solidly established. In addition, equations [3a-c] and [6] show the elasticity (patent counts - R&D intensity) varies between 10 and 22.5%. It suggests a more comprehensive 10 to 22.5% percentage increase in patents if R&D efforts double. Moreover, the returns to scale are confirmed to hoover about 20% if fitted R&D expenditures are imputed as explanatory variable. As a consequence, the Belgian

pharmaceutical industry is labelled exhibiting decreasing returns to scale. Withal results are in line with expectations and hypothesis 2 is confirmed.

Those expectations theorized signs on firm size and lagged innovation to be positive as well. Both are supported by the data. The positive beta on the innovation dummy corresponds to stating that firms that were more innovative in the past will be more innovative today. Similarly, larger firms are found to be relying on patents more. As with fitted R&D intensity the elasticities have the more meaningful interpretation being that patents will increase by 20% of the relative increase in corporate size. Since fitted R&D intensity is assumed to represent the effect of innovative efforts on its success, the results hint at increased strategic uses.

Concerning competition, no meaningful direct impact could be observed within the data. Seemingly competition impacts innovation only indirectly as a determinant of its efforts. Nevertheless, strategic patenting and portfolio building are an expression of the competitive environment a firm experiences. Conjointly an impact of maturity on patenting cannot be established statistically. If any the results suggest a negative impact of age on patenting. It is important to note, size, which is positively correlated to maturity, was controlled for. The interpretation is more mature firms make less use of the patenting system. It could be that they have less innovative success, but they may also make more use of other appropriation mechanisms because of expertise and their established position.

The different ZINB model specifications are inconclusive on the entry of open innovation. Using the restricted version of the inflation model it enters univocally positive. In contrast the augmented inflation model (with R&D intensity) reverses the sign, though not statistically significant. NB and GLS estimation do not provide consolation either. In addition, pairwise correlation between the OI and R&D intensity is about 70% due to the high number of zero observations. Moreover, Vuong's test statistic log likelihood values indicate [1b] and [4b] specifications are superior to their a-versions. As a consequence, the null hypothesis cannot be rejected even though it enters several specifications highly significant.

Pertaining to the test statistics, they are ambiguous in identifying the best specification. The base case [1a] rejects using the inflation model and suggests NB estimation is to be preferred (Vuong). Whereas the NB specifications draw the same conclusions with only slightly different coefficients. The augmented inflation model's alpha is insignificant and suggest it reduction to a Poisson model. Nevertheless, section 8.1 on descriptive statistic clearly showed excessive zero counts and overdispersion. The inflation model is therefore incomplete. In order to resolve this issue, the auxiliary variables used in stage 1 of the Tobit II model were readdressed. Unfortunately, the specifications did not succeed in converging or only changed the results marginally (inclusion of solvency ratio). On the grounds of the descriptive statistics and the log-likelihood values (-493.814), model [4b] can thus be labelled the best count model.

To supplement the count models log-transformation was persevered normalizing patent counts (Table 4). Equations [3a-c] and [6] are the resulting specifications using pooled OLS, fixed effects and random effects. Most importantly they lead to similar conclusions on innovation input-output

relationship. Fixed effects regression [3a] allows to control for individual heterogeneity among the entities effectively assigning them different intercepts. Random effects regression [3b] superimposes a common intercept, as pooled OLS does. It assumes individual deviations are simply the result of a random process. The difference to the fixed effects model is therefore that the unobserved heterogeneity is uncorrelated with the vector of explanatory variables. The difference to pooled OLS is that generalized least squares allows transformation that controls for serial correlation and other violations of error term assumptions, a priori. Consequently, it assures unbiasedness of the pooled OLS estimators.

The output gives several hints FE regression is best used. They will not be inconsistent even if the true underlying model is that of random effects. FE is adequate because the correlation between the unobserved heterogeneity and the explanatory variables is very large (60%) considering RE and Pooled OLS models assume it to be zero. More formally the Hausman test also favors FE. It implies revisiting the effect of the lagged patent dummy and corporate size, since their impact is no longer significantly positive. However, count models and theory support the findings by RE and Pooled OLS. Therefore, the correlation may be interpreted as omitted variable bias. Omitted variables enter in the unobserved heterogeneity, are not time constant and by definition correlate with the depend variable. Improvement should thus be sought in that area. Nevertheless, the measures of model fit indicate a lot of work has already been done.

Summarizing the results, the most important finding is the confirmation innovative efforts affect outcomes positively. Thereby hypothesis 2 is authenticated, suggesting innovation is relevant to the success of a firm which will be tested in section 8.4. Furthermore, expected positive associations between firm size, previous patenting activity and patent counts could also be established. In the margin more mature firm could be argued to use the patenting system to a lesser extent. Impacts of other variables are contested. Finally, combining the findings within the theory, the descriptive statistics and the regression results ZINB can be concluded an adequate modelling patent counts for the Belgian pharmaceutical industry. But the regression results revealed potential for omitted variable bias and unobserved heterogeneity to be explained.

Dependant Variable **Independent Variable**	Simple Patent Application Count	Simple Patent Application Count	Simple Patent Application Count	Ln (Patent Count)	Ln (Patent Count)	Ln (Patent Count)	Simple Patent Application Count	Simple Patent Application Count	Simple Patent Application Count	Ln (Patent Count)
Specification number	[1a]	[1b]	[2]	[3a]	[3b]	[3c]	[4a]	[4b]	[5]	[6]
Fitted Ln(R&D Intensity)	0.298***	0.103***	0.316***	0.109***	.150***	0.223***				
based on HHI	[0.040]	[0.027]	[0.034]	[0.032]	[0.035]	[0.025]				
Fitted Ln(R&D Intensity)							0.303***	0.084***	0.303***	0.199***
based on PM							[0.033]	[0.029]	[0.033]	[0.026]
Ln(Employees)	0.360***	0.478***	0.343***	0.004	0.190***	0.219***	0.301***	0.564***	0.301***	0.223***
	[0.111]	[0.068]	[0.107]	[0.101]	[0.071]	[0.036]	[0.073]	[0.044]	[0.073]	[0.034]
Firm Age	-0.103	-0.109	-0.095			-0.043**	-0.037	-0.082	-0.037	-0.040*
	[0.112]	[0.114]	[0.105]			[0.017]	[0.119]	[0.117]	[0.119]	[0.021]
First Lag Patent Dummy	1.721***	0.955***	1.877***	-0.068	0.161	0.402***	1.964***	1.103***	1.964***	0.486***
	[0.296]	[0.258]	[0.228]	[0.031]	[0.135]	[0.152]	[0.249]	[0.284]	[0.249]	[0.169]
Ln(Competition)	0.063	-0.029	0.067*	-0.025	0.004	0.011				
1/HHI	[0.386]	[0.030]	[0.039]	[0.031]	[0.016]	[0.011]				
Profit Margin (as %)							0.001	-0.002	0.001	0.003
PM							[0.007]	0.005	[0.007]	[0.003]
Open Innovation	1.932***	-0.416	2.191***	-0.009	-0.062	-0.366	1.952***	-0.326	1.952***	-0.496
	[0.492]	[0.316]	[0.402]	[0.319]	[0.340]	0.305	[0.428]	[0.342]	[0.428]	[0.342]
Intercept	-3.239**	0.664	-3.623**	0.624	-0.567	-0.511*	-3.460***	-0.402	-3.460***	-0.447**
	[1.500]	[1.411]	[1.400]	[0.622]	[0.379]	[0.300]	[1.121]	[1.141]	[1.121]	[0.189]
Ln (Alpha)	0.748***	-0.008	0.822***				0.838***	0.012	0.838***	
Log Likelihood (Full Model)	-691.374	-587.213	-691.813				-586.143	-493.814	-586.143	
Vuong's z test statistic	0.18	5.47					-0.16	4.92		
Pseudo R2			0.262			0.668			0.275	0.661
rho				0.799	0.610					
Inflation model	Logit	Logit					Logit	Logit		
Fitted Ln(R&D Intensity)		-74.78						-82.166		
Ln(Employees)	-0.252	0.440					-0.077	0.775		
First Lag Patent Dummy	-22.624	-3.146***					-22.605	-4.289***		
Intercept	0.026	3.059**					-14.417	3.054*		
N Observations	601	601	601	601	601	601	556	556	556	556
Nonzero Observations	151	151					125	125		
Zero Observation	450	450					431	431		
Estimation Method	ZINB	ZINB	Negative Binomial	Fixed Effects GLS	Random Effects GLS	Pooled OLS	ZINB	ZINB	Negative Binomial	Pooled OLS
Stastistical significance indicated by stars (*: 10%, **:5%, ***:1%). Standard errors between brackets.										

Table 11: Innovative input-output

8.4 Relationship: patents – productivity

The third and last equation in the CDM framework sketches the relationship between innovative output and firm performance (testing hypothesis 1). In section 8.3, the positive relationship between innovative efforts and innovative success, as measured by patent counts was already established. Here, the natural logarithm of actual observed patent counts is consequently considered as the main variable of interest. Observing its coefficient enables formulating a blueprint of the dynamics of innovation and firm performance in the Belgian pharmaceutical industry. It indirectly models the effect of innovative efforts on firm performance.

As outlined in the literature review, firm performance can be pronounced in many different ways, such as social, environmental and economic impacts of its activities. Traditionally economic and monetary measurement instrument are used since they are readily available and reasonably objective. Likewise the natural logarithm of value added per employee is used as a dependent variable here. As alternative, turnover per employee or more sophisticated multifactor productivity (growth) measures are also frequented.

Other than innovative output, innovative input is also directly linked to firm performance in order to identify a possible direct channel. Again, two different fitted R&D intensities are used. As in the previous section, serial correlation is dealt with by including the first lag of the dependent variable. Moreover, firm size, competition and open innovation are selected as control variables. To exploit the panel data nature underlying the dataset, random effects and fixed effects regression are used as an alternative to pooled OLS.

Inspecting the results reveals two invariable findings (**Error! Reference source not found.**). One is a negative influence of competition, with less competition boosting labor productivity. Namely, profit margin and more industrial concentration are associated with a better firm performance when it measured by value added per employee. In concrete terms, it translates to an increase of one percentage point in profit margin being equivalent to an increase of 1.5% in the dependent variable. Even so, labor productivity most likely did not increase in real terms. Rather increased market concentration and price setting possibilities enable corporation to appropriate a larger portion of the consumer surplus. Moreover, the HHI-based competition indicator was abandoned from equation [2] onwards due to its variance inflating character (high absolute correlation with size). The other finding is that labor productivity is clearly autoregressive. Its first lag turns out to be a dominant predictor of the dependent variable.

Concerning the effect of innovation on labor productivity pooled OLS and RE regression confirm hypothesis one: firms that innovate more are more productive. Although FE regression does not succeed in statistically rejecting the hypothesis innovative output has no effect on firm performance. What is more, FE has the Hausman test backing it. On the other hand, RE is advocated by the fact that the fixed effects model is incapable of proving correlation between unobserved heterogeneity and the regressors. The fixed effects specifications do not appear strong. Likewise, their respective F-statistics testing whether all beta coefficients are zero

simultaneously is unconvincingly low (yet significant). Besides, quantile regression also supports the Pooled OLS/RE findings.

Adding up with the statistics and theory the results of the regression we therefore formulate a strong suspicion patent counts are positively associated with labor productivity. Yet, the results of the FE model refrain us from concluding on it. Inasmuch hypothesis one is only partially supported with an assumed affect that lies between a 4% to 7% increase in value added per employee if the number of patent counts doubles. Using citation counts, other estimation methods or resolving potential omitted variables are all sources that could lead to improvement.

Progressing to the other variables, none of the results suggests an impact consistently. Thus firm size, R&D intensity and open innovation do not appear to affect firm performance as measured here. A direct linkage between R&D and performance is not observed, as expected considering the sequential logic underlying CDM. In addition, patent counts are more likely to adequately represent the impact of the innovative process on productivity, effectively controlling the correct formulation of hypothesis 1. Readdressing the conceptual framework, it may be concluded CDM adequately represent the innovation-performance dynamics underlying the Belgian pharmaceutical industry. The next section elaborates on open innovation.

Dependant Variable **Independent Variable**	Ln of Value Added per Employee	Ln of Value Added per Employee	Ln of Value Added per Employee	Ln of Value Added per Employee	Ln of Value Added per Employee	Ln of Value Added per Employee
Specification number	[1]	[2]	[3a]	[3b]	[4a]	[4b]
Ln(Patent Counts)	0.071***	0.066**	0.066**	0.066*	-0.021	-0.035
	[0.024]	[0.026]	[0.034]	[0.034]	[0.071]	[0.070]
Fitted Ln(R&D Intensity)	-0.011		-0.001		-0.025	
based on HHI	[0.014]		[0.015]		[0.027]	
Fitted Ln(R&D Intensity)		-0.004		-0.004		-0.020
based on PM		[0.010]		[0.015]		[0.027]
Ln(Employees)	-0.203***	0.008	0.009	0.008	0.170	0.161
	[0.055]	[0.026]	[0.032]	[0.032]	[0.179]	[0.180]
Ln(Lagged Value Added per Employee)	0.414***	0.474***	0.472***	0.474***	0.202**	0.199*
	[0.076]	[0.075]	[0.065]	[0.066]	[0.100]	[0.100]
Ln(Competition)	-0.129***					
1/ HHI	[0.025]					
Profit Margin (as %)		0.015***	0.015***	0.015***	0.020***	0.020***
PM		[0.004]	[0.003]	[0.003]	[0.004]	[0.004]
Open Innovation	0.133	0.164	0.177	0.164	0.099	0.077
	[0.121]	[0.132]	[0.154]	[0.154]	[0.256]	[0.263]
Intercept	9.271***	6.131***	6.107***	6.131***	8.994***	9.102***
	[1.086]	[0.864]	[0.749]	[0.757]	[1.597]	[1.600]
F Value H0: all βi = 0					F(6,73) = 4.92***	F(6,74) = 4.82***
F Value H0: corr(Error,X) = 0					F(30,73) = 1.25	F(30,74) = 1.27
Wald Chi2 (df)			157.62	154.25		
Pseudo R2	0.546	0.574				
Rho			0	0	0.46	0.458
Mean VIF	3.93	1.48				
N Observations	113	111	110	111	110	111
Estimation Method	Pooled OLS	Pooled OLS	Random Effects GLS	Random Effects GLS	Fixed Effects GLS	Fixed Effects GLS
Stastistical significance indicated by stars (*: 10%, **:5%, ***:1%). Standard errors between brackets.						

Table 12: Innovative output - Labor productivity.

8.5 Open innovation

To our knowledge this study is unique in its attempt to include open innovation and collaboration in a quantitative manner while studying innovation and firm performance in the Belgian pharmaceutical industry. In other words, we did not come across such a study while reviewing the literature. OI can be very broadly interpreted and as such it is infeasible to represent all the facets quantitatively. It is also a new concept. Or at least, the term has only been around since the early 2000s. As a consequence, there is a scarcity of satisfactory indicators. Therefore, inspiration was sought in related fields bearing in mind the data availability.

Eventually, the indicator was constructed in similar fashion to the SHII indicator as proposed by Guellec & van Pottelsberghe de la Potterie (2001). They used the different nationalities of the inventors (applicants) cited on a patent to study the internationalization of technology on a country level. Among others, they found the extent of collaboration with foreign researchers was very well explained by such an indicator. Analogously, we used the percentage of foreign nationalities on patent applications for each year-entity combination as an indicator for open innovation, or collaboration in a narrower sense. Its kernel density is described by Figure 17 below.

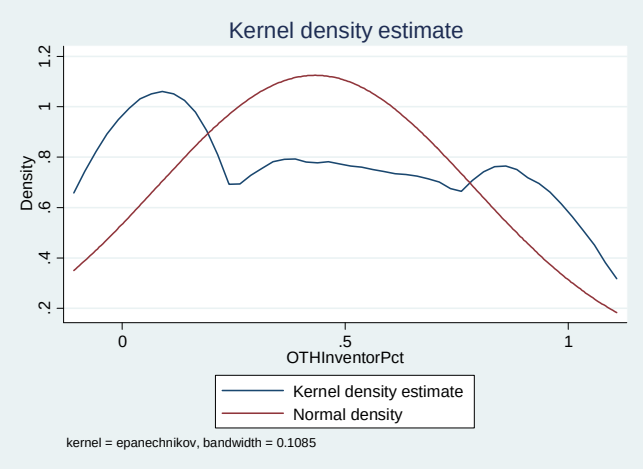


Figure 17: Kernel density on international collaboration.

An important difference is that, here, the study is conducted on firm level. Using nationalities rather than the different companies that inventors work for, is therefore an important limitation (could not feasibly be distilled). On that grounds, the constructed indicator misrepresents collaborative efforts to the extent people working for different innovative entities have the same nationalities. It is especially relevant for SMEs, whom often have a geographically smaller operation area than big pharmaceuticals.

The findings in section 8.2 - 8.4 suggest open innovation enters the CDM framework only as an explanatory variable for innovative efforts. Firms that collaborate more are found to be more research intensive. Several explanations that are not mutually exclusive of exhaustive can be formulated. As first, rising costs of the R&D process and dwindling new molecular or biological

entities might cause firm to look beyond the borders of the own organization. Secondly, a complementary set of skills or fields of expertise might also be reasons to combine resources, especially as the market environment because increasingly interweaved technologically. Finally, it might also be the case the indicator does not adequately represent collaboration on a firm level because it seems biased towards patenting firms and larger firms. The latter is supported by their significant positive pairwise correlations ranging from 0.25 to 0.60.

In equation two, the impact of OI on patent applications is ambiguous. At times, it enters the regression highly significant and positive whereas other specifications show negative though insignificant beta coefficient. Although the inflation models can be criticized, no conclusion can be drawn on a direct effect between collaboration and patent application. The regressions did however for firm size and lagged patent applications which can therefore not be introducing bias.

Concerning its beta coefficient on labor productivity, similarly, no statistical inferences can be made. Nevertheless, the beta coefficient was consistently positive throughout the regressions which might hint at an effect actually existing. Either it is very small, the specification of the indicator or the dependent variable was not ideal or simply no effect exist. Disentangling the indicator into low, moderate and high OI subsamples did not hold enhanced insights. Withal, answering hypothesis 4 only an initial effect on R&D intensity could be established.

9 Conclusions

Setting out to sketch the impact of innovation on firm performance in the pharmaceutical industry we could not built on prior econometric studies for Belgium. This master thesis is therefore largely built upon the work of Sanyal (2014) who studies innovation and firm performance in the Dutch pharmaceutical industry. Quantitative methods and measurement of the variables were frequently checked against those used by Sanyal (2014) and sometimes inspired by them (for example: ZINB estimation). Other than that study, literature on innovation in different manufacturing industries and the pharmaceutical industries of different countries provided the foundation to build a conceptual model upon. In that respect, the pioneering framework of Crépon et al. (1998) was put to work. It postulates sequential estimation of innovative efforts, success and firm performance. In the process, several hurdles such as missing data, count variables, heavy skewness and kurtosis had to be taken.

With respect to the first equation (section 8.2), which investigates the determinants of R&D, we used a sample selection model to cope with differences between observed, reported R&D and true R&D intensities. We find size and competition influence the capital intensity negatively. In other words, less innovate efforts are undertaken if competition intensifies and R&D expenditures do not rise proportionally to size. R&D expenditures are found to increase between 38% and 73% if the corporate size doubles. On the other hand, open innovation and fewer budgetary constraints are found to affect the R&D intensity positively. In fact equation one, as a determinant of R&D intensity, is where open innovation seems to enter the series of structural equations (section 8.5). Whether open innovation causes the R&D intensity the be boosted because of expanded technological opportunities or whether rising R&D expenditures as an important cost caused firms to look beyond the borders of the own organization could not be verified. It is probably not the one or the other explanation, but a combination of both. Additionally, we found that including information on tax regimes and the financial health of a firm have potential to augment the results for similar studies in the future.

In the second part (section 8.3) of the structural equations model the fitted R&D intensities were used as an explanatory variable for the number of patent application counts. It translates to concepts of innovative input and output into practice. Both count models as GLS estimation were fitted to the data. Under these specifications a positive impact could univocally be confirmed. Still, results suggest pronounced decreasing returns to scale (10-22.5%) are present. Which is in line with soaring R&D costs and the difficulties of sustaining the product line encountered in modern-day pharmaceuticals. Other than R&D, firm size and previous patenting activities entered the regressions with a positive beta. Larger firms are known to rely more heavily on the patenting system although frequently not because of desirable economic motives, but rather strategically. In addition, maturity, as measured by firm age, was associated negatively with patent counts. It suggests more mature firms make more use of alternative routes to appropriate return, or that they are less innovative.

In section 8.4 the link between innovative output and firm performance was investigated. Firm performance which is a multidimensional construct was measured in monetary terms. It was represented by labor productivity – the log value added per employee. Here, a positive relationship between innovation and firm performance is suggested by the results but not unambiguously supported. Using fixed effects regression, patent counts do not seem to have an effect on labor productivity. Otherwise the elasticity is found to lie between 4% and 7%. We suspect other another operationalization of the variables and the usage of citation weighted patents could build a stronger case considering other findings in the literature. Further a direct link of innovative efforts to labor productivity did not seems present. Fitted R&D intensity has no explanatory power. Profit margin did have a positive impact on labor productivity suggesting market power increases financial results. Moreover, labor productivity was also clearly autoregressive.

To conclude we revisit the conceptual framework. In general, the structural equations were found to be adequately representing the way innovation is translated in financial prosperity. Combining equations two (H2) and three (H1) results in concluding that firms making a higher R&D effort will perform better, ceteris paribus. Although the interlinkage is rather weak and the relationship between innovative success and labor productivity was not always established. The effect of competition is ambiguous, but overall the results hint less competition might harbor more innovation. Finally, open innovation did not seem to influence innovative output or firm performance directly. It enters as a determinants of R&D intensity.

10 Limitations

Throughout the literature review we distilled a conceptual framework and identified different ways to operationalize relevant constructs. Frequently, they do not fully represent the concepts that are studied. The results we obtained are therefore limited to the extent the specifications correctly resemble the concepts and underlying relationships in the data. Yet, this is a problem with every study and we suspect it does not threaten inferences about our results, because we mainly on validated measurement instruments and an established framework. We point out four limitations that have the potential for a more definite impact on our results

The empirical part –especially section 8.2- already elaborated on the lack of observations. It is for a fact the dataset is highly imbalanced. It is primarily a result of combining three different datasets. In addition, the ownership structure very incompletely identified consolidated structures. All threaten to introduce bias if the missing values are systematically different from those that are reported. It became very relevant with respect to R&D observations. In order to cope with the non-reporting issue, we embraced a Tobit II model and replaced actual observed R&D entries with the fitted version in the subsequent analysis.

Another limitation is that we did not succeed in extracting patent citation data. The literature has shown that this would be an improvement on using simple patent count because it weights patents approximating their market value. Still, results are usually similar using simple patent counts and citation weighted patent counts (for example: Sanyal, 2014). On top the literature also diagnosed patent counts being used increasingly for strategic reasons, which damages its reputation as a proxy for innovative success.

Relating to the other variables the most important concerns are about our open innovation indicator. It is not validated and we can therefore not be as certain about its inferences. Furthermore, the indicator is constructed based on inventors nationalities which is a rather distant proxy for the wide perspective of open innovation activities. Surveys, for example, could provide enhanced insights. A similarly narrow interpretation should be given to the eventual measure of firm performance –labor productivity, although it is a validated measurement tool. Two critiques are that it only focusses on monetary aspects of firm performance and that it is only one aspect of financial performance therein. Multifactor productivity (growth) for example, covers a wider perspective.

Ultimately, we also come back to the estimation methods. The characteristics of the data prompted more sophisticated regression analysis. Namely, Tobit II and ZINB are not covered in our curriculum and are rather complex mathematically. It is therefore possible that we did not exploit their possibilities to the maximum. For example, the observation of the majority of the beta coefficients being insignificant in the first stage of the Tobit II model and in the inflation part of ZINB are possibly a result of lack of understanding because both models are validated and backed by the literature.

11 Bibliography

Abrahamsen, M., Acar, O., Brinded, B., & Rainisch, V. (2011). The Belgian Pharmaceutical Cluster. Harvard Business School: Institute for Strategy and Competitiveness, Cambridge, Massachusetts.

Aghion, P., Bloom, N., Blundell, R., Griffith, R. & Howitt, P. (2002) Competition and innovation: an inverted U relationship.

Aghion, P., Bloom, N., Blundell, R., Griffith, R. & Howitt, P. (2005). Competition and innovation: an inverted U relationship.

Arrow, K. (1962). Economic welfare and the allocation of resources for invention. Princeton University Press.

Artz, K. W., Norman, P. M., Hatfield, D. E., & Cardinal, L. B. (2010). A longitudinal study of the impact of R&D, patents, and product innovation on firm performance. Journal of Product Innovation Management, 27(5), 725-740.

Bodie, Z., Kane, A., & Marcus, A. J. (2013). Essentials of Investments. McGraw-Hill.

Charles River Associates (2004). Innovation in the Pharmaceutical sector. A study undertaken for the European commission. London

Chesbrough, H., Vanhaverbeke, W. & West, J. (2006). Open Innovation: Researching a New Paradigm [e-version]. New York: Oxford University Press.

Ciftci, M., & Cready, W. M. (2011). Scale effects of R&D as reflected in earnings and returns. Journal of Accounting and Economics, 52(1), 62-80.

Cohen, W. M., Nelson, R. R., & Walsh, J. P. (2000). Protecting their intellectual assets: Appropriability conditions and why US manufacturing firms patent (or not) (No. w7552). National Bureau of Economic Research.

Crépon, B., Duguet, E., & Mairesse, J. (1998). Research, Innovation and Productivity : an Econometric Analysis at the Firm Level.

EFPIA (2014). facts & figures [e-version]. obtained 8 Dec 2015 via Pharma.be

Enkel, E. Gassman, O., & Chesbrough, H. (2009) Open R&D and open innovation: exploring the phenomenon. R&D Management, 39(4), 311-316.

EPO (2013). GPI introductory guide. Consulted via www.epo.org/searching-for-patents/technical/espacenet/gpi.html, February 21th 2016

European Commission (2008). Pharmaceutical sector inquiry: preliminary report. DG competition staff working paper. Retrieved via ec.europa.eu/competition/sectors/pharmaceuticals/inquiry, May 21th 2016

European Commission (2014). Patents and Standards: a framework for IPR-based standardization, final report. Retrieved via ec.europa.eu/growth/industry/intellectual-property/patents/standards/index_en.htm, May 19th 2016

European Commission (2016). The European Single Market. Retrieved via ec.europa.eu/growth/single-market/, July 15th 2016

FOD Economie, K.M.O. Middenstand en Energie. Retrieved via economie.fgov.be/nl/ondernemingen/Intellectuele_Eigendom/octrooien May 19th 2016

Foxon, T. J. (2007). 9 Technological lock-in and the role of innovation. Handbook of sustainable development, 140.

Fritsch, M., & Lukas, R. (2001). Who cooperates on R&D? Research policy, 30(2), 297-312.

García-Fontes, W. (2011). Incentives to innovate: a survey. In Incentives for Research, Development, and Innovation in Pharmaceuticals (pp. 5-14). Springer Healthcare Iberica.

Gärtner, M. (2013). Macroeconomics

Gilbert, R. (2006). Competition and Innovation. Issues in Competition Law and Policy. Ed. Wayne Dale Collins. American Bar Association Antitrust Section.

Gilbert, R., & Shapiro, C. (1990). Optimal patent length and breadth. The RAND Journal of Economics, 21(1), 106-112.

Graves, S. B., & Langowitz, N. S. (1993). Innovative productivity and returns to scale in the pharmaceutical industry. Strategic Management Journal, 14(8), 593-605.

Griliches, Z., (1990). Patent Statistics as Economic Indicators: A Survey. Journal of Economic Literature (28), pp. 1661-1707.

Guellec, D. & van Pottelsberghe de la Potterie, B. (2001). The internationalisation of technology analysed with patent data. Research Policy, 30, 1253-1266.

Gujarati, D. N. (2009). Basic econometrics. Tata McGraw-Hill Education.

Hall, B. H., Jaffe, A., & Trajtenberg, M. (2005). Market value and patent citations. RAND Journal of economics, 16-38.

Hanel, P. (2005). Current Intellectual Property Protection Practices of Manufacturing Firms in Canada. Intellectual Property and Innovation in the Knowledge–Based Economy.

Hannigan, T., Mudambi, R., & Sfekas, A. (2013). A longitudinal analysis of blockbuster pharmaceutical drugs. DRUID Academy.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error, Econometrica, 153-162.

Henderson, R., & Cockburn, I. (1996). Scale, scope, and spillovers: the determinants of research productivity in drug discovery. The Rand journal of economics, 32-59.

Hsiao, C. (2007). Panel data analysis—advantages and challenges.

Hottenrott, H., & Peters, B. (2012). Innovative capability and financing constraints for innovation: More money, more innovation?. Review of Economics and Statistics, 94(4), 1126-1142.

Khanna, I. (2012). Drug discovery in pharmaceutical industry: productivity challenges and trends. Drug discovery today, 17(19), 1088-1102.

Lipsey, R. & Chrystal, A. (2011). Economics. 12th edition.

Lopez, A. (n.d.). Innovation and appropriability, empirical evidence and research agenda.

Mairesse, J. & Mohnen, P. (2010). Using innovation surveys for economic analysis.

Mohnen, P. & Hall, B.H. (2013). Innovation and Productivity: An Update. Eurasian Business Review, 3(1), 47-65.

Munos, B. (2009). Lessons from 60 years of pharmaceutical innovation. Nature Reviews Drug Discovery, 8(12), 959-968.

Munos, B. (2010). Can open source drug R&D repower pharmaceutical innovation?

Nagaoka, S., Motohashi, K. & Goto, A. (2010) Patent Statistics as an Innovation Indicator. Handbook of the Economics of Innovation, 2, 1083-1127.

Nickell, S.J. (1996), Competition and Corporate Performance. Journal of Political Economy, 724-746.

OECD (2009), OECD patent statistic manual, OECD Publishing, Paris.

Petrova, E. (2014). Innovation in the Pharmaceutical Industry: The Process of Drug Discovery and Development.

Rogers, M. (1998). The definition and measurement of innovation. Parkville, VIC: Melbourne Institute of Applied Economic and Social Research.

Sanyal, S. (2014). Innovation, competition and firm performance: An econometric analysis on the Dutch pharmaceutical sector.

Scherer, F.M. (1967). Market structure and the employment of scientists and engineers. American Economic Review, 524-531

Scherer, F. M. (2010). Pharmaceutical innovation. Handbook of the Economics of Innovation, 1, 539-574.

Schumpeter, J. (1942). Capitalism, Socialism and Democracy. New York: Harper and Row.

Tangen, S. (2005) Demystifying productivity and performance. International Journal of Productivity and performance management, 54.1, 34-46.

Vancauteren, M. (2014). Course: Time-series analysis, Part II, lecture notes. UHasselt.

Vancauteren, M. (2014). Course: Energy, the environment and the economics of innovation, lecture notes. UHasselt.

Vancauteren, M., Melenberg, B. & Plasmans, J. (2014) Innovation and Productivity of Dutch Firms: A Panel Data Analysis. (obtained via Prof. dr. Vancauteren)

Websites

EPO (2016) : https://www.epo.org/learning-events/materials/inventors-handbook/protection/patents.html on May 21th 2016

Thomson Reuters (2016) : http://ip-science.thomsonreuters.com/support/patents/patinf/patentfaqs/invent/ on May 21th 2016

USA Today (2010) : http://usatoday30.usatoday.com/money/companies/management/advice/2010-07-19-advice19_ST_N.htm -Eli Lilly CEO John Lechleiter (2010)

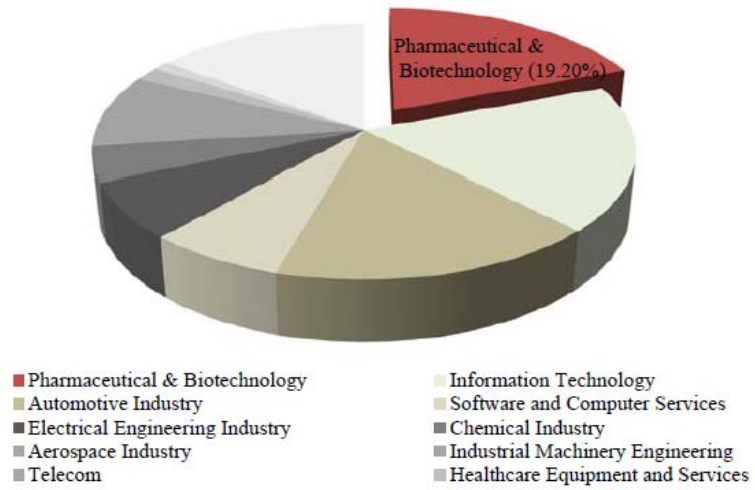
Pharma.be : Information on developments in the Belgian pharmaceutical industry.

World Intellectual Property Organization. (2015). What is intellectual property? retrieved May 23th,2015, via http://www.wipo.int/about-ip/en/).

12 Appendix

	1995	2000	2005	2006	2007	2008	2009	2010	2011	2012	CAGR last year available/2000
Production											
Pharmaceutical industry (m of EUR)	2.275,91	2.988,63	4.825,53	5.265,37	5.221,66	5.518,22	5.419,50	6.295,92	7.061,36	6.696,37	7,0
Share of chemical industry (as %)	13,7	11,9	15,3	15,2	14,7	15,2	18,7	17,8	18,3	17,2	
Share of manufacturing industry (as %)	2,4	2,5	3,5	3,5	3,3	3,6	4,4	4,5	4,7	4,5	
Exports											
Pharmaceutical industry (m of EUR)	3.011,33	7.419,77	28.046,33	30.447,62	34.574,17	34.343,00	37.459,55	38.500,70	35.909,41	36.533,95	14,2
Share of total Belgian exports	2,3	3,6	10,4	10,6	11,0	10,6	14,1	12,4	10,5	10,5	
Investments											
Pharmaceutical industry (m of EUR)	254,66	388,10	518,10	555,00	693,20	767,40	725,30	696,60	862,90	951,10	7,8
Share of chemical industry (as %)	17,0	32,1	30,2	30,7	29,7	36,1	44,7	44,3	43,3	48,9	
Share of manufacturing industry (as %)	3,8	4,6	6,9	7,0	7,6	8,7	10,1	10,5	12,0	13,0	
Expenditure on R&D											
3 companies conducting fundamental research (m of EUR)	339,58	762,18	1.550,59	1.558,71	1.884,32	1.884,14	1.809,35	1.803,70	2.057,04	2.342,98	9,8
Pharma industry expenditure on in-house R&D (m of EUR)	360,16	580,90	921,41	1.124,66	1.249,23	1.273,36	1.299,64	1.443,46	1.707,99	n.d.	6,2
Share of in-house expenditure in private sector (as %)	14,6	16,2	24,4	27,4	28,3	27,4	28,4	28,70	30,40	n.d.	
Total employment											2012/2006
Pharmaceutical industry (people employed)				30.843	31.817	32.366	32.125	31.903	32.740	32.718	1,0
Share of manufacturing industry (as %)	2,9	3,7	4,8	5,2	5,6	6,0	6,3	6,3	6,5	6,7	
Share of private sector (as %)	0,9	1,0	1,2	1,2	1,2	1,2	1,2	1,2	1,2	1,2	
Total employment in R&D											
Pharmaceutical industry (m of EUR)	n.d.	4.014	4.829	5.478	5.665,0	4.827,00	4.840,00	5.274,00	5.762,00	n.d.	3,3
Share of total R&D employment in manufacturing industry (as %)	n.d.	13,6	17,7	22,8	22,7	20,6	20,6	23,5	24,3	n.d.	
Share of total private sector R&D employment (as %)	n.d.	10,3	12,6	14,0	13,9	11,7	11,6	12,7	12,5	n.d.	
Total pharmaceutical market (ex-factory price)											
Pharmaceutical industry (m of EUR)	1.914,67	2.625,38	3.661,72	3.671,94	3.900,90	4.188,03	4.319,57	4.428,29	4.537,11	4.493,79	4,6
Expenditure on R&D as % of total pharmaceutical market	17,7	29,0	42,3	42,4	48,3	45,0	41,9	40,7	45,3	52,1	

Table 14: Figures on the Belgian pharmaceutical industry.



Source: European Commission (2008), EU Industrial R&D Investment Scoreboard

Figure 18: Share of worldwide R&D investment in different sectors. Source: European Commision (2008). Industrial R&D Investment Scoreboard.

	Revenue €mm	Employees	Region	Corporate Parent	Country of Parent	Business Description
Janssen Pharmaceutica NV	6,708	3,842	Flanders	Johnson & Johnson	US	▪ Janssen Pharmaceutica NV engages in producing and marketing pharmaceutical products in the US and internationally.
Glaxosmithkline Biologicals S.A.	5,797	1,000	Wallonia	GSK	UK	▪ Glaxosmithkline Biologicals S.A. engages in the research, development, manufacture, and supply of vaccines.
UCB SA (ENXTBR:UCB)	3,218	8,898	Brussels	N/A	Belgium	▪ UCB SA, a biopharmaceutical company, engages in the research, development, and commercialization of medicines with a focus on the fields of central nervous system and immunology disorders.
Corden Pharmachem P/A Controle B Nv	935	96	Flanders	Cambrex Corporation	US	▪ Pharmaceutical Preparations
Omega Pharma	856	1,945	Flanders	N/A	N/A	▪ Omega Pharma NV engages in the development, marketing, and sale of health and personal care products
AstraZeneca NV	759	500	Brussels	AstraZeneca	UK	▪ AstraZeneca NV operates as a pharmaceutical company.
Pharma Belgium Sa	650	200	Brussels	N/A	Belgium	▪ Pharmaceutical Preparations
Sanofi-Aventis Belgium S.A.	427	200	Flanders	Sanofi-Aventis	France	▪ Sanofi-Aventis Belgium S.A. also known as Aventis Pharma SA, engages in the development, manufacture, and sale of pharmaceutical products and vaccines in Belgium.
Eumedica SA	415	20	Brussels	N/A	Belgium	▪ Eumedica SA, a pharmaceutical company, produces, stores, and distributes medicinal products to treat serious and rare diseases in Europe.
Pfizer Manufacturing Belgium NV	391	1,000	Flanders	Pfizer	US	▪ Pfizer Manufacturing Belgium NV engages in developing and manufacturing innovative medicines for life of humans and animals.

Figure 19: Overview of Top10 Belgian Pharma Companies.

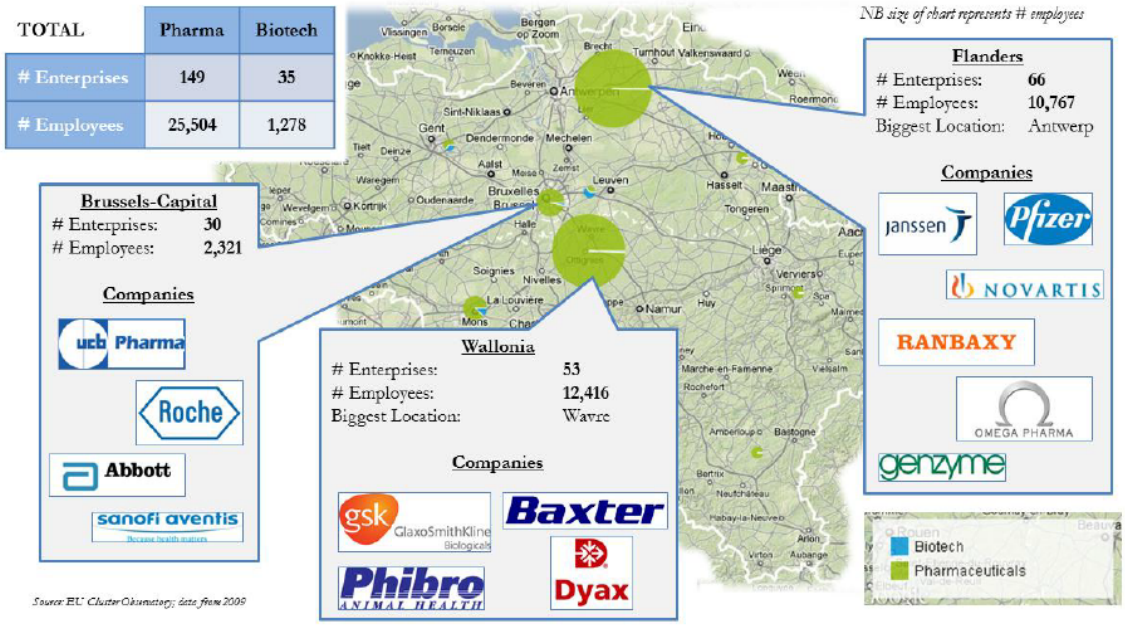


Figure 20: Geographic distribution of Companies and Employees in Belgium.

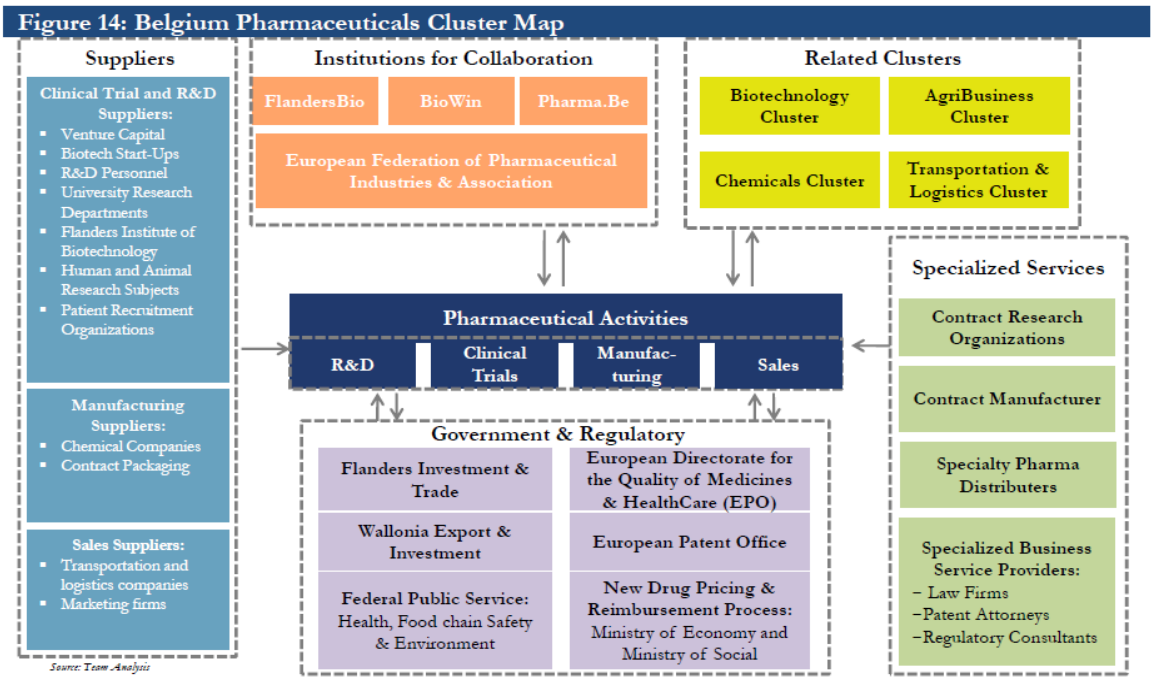


Figure 21: Belgium Pharmaceutical Cluster map

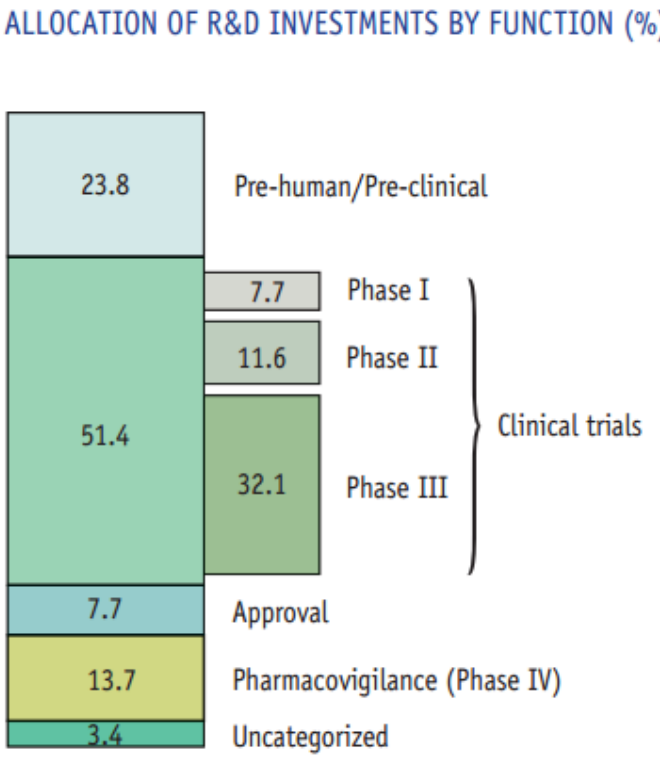


Figure 22: Allocation of R&D investments by function (%), source PhRMA, annual Membership Survey 2014)

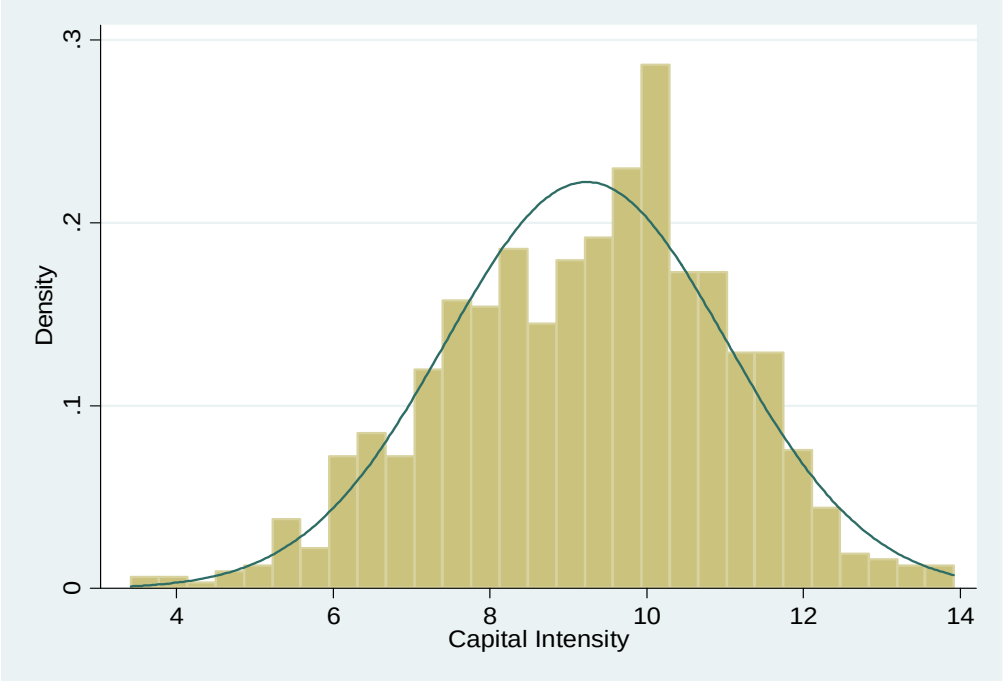


Figure 23: Capital Intensity

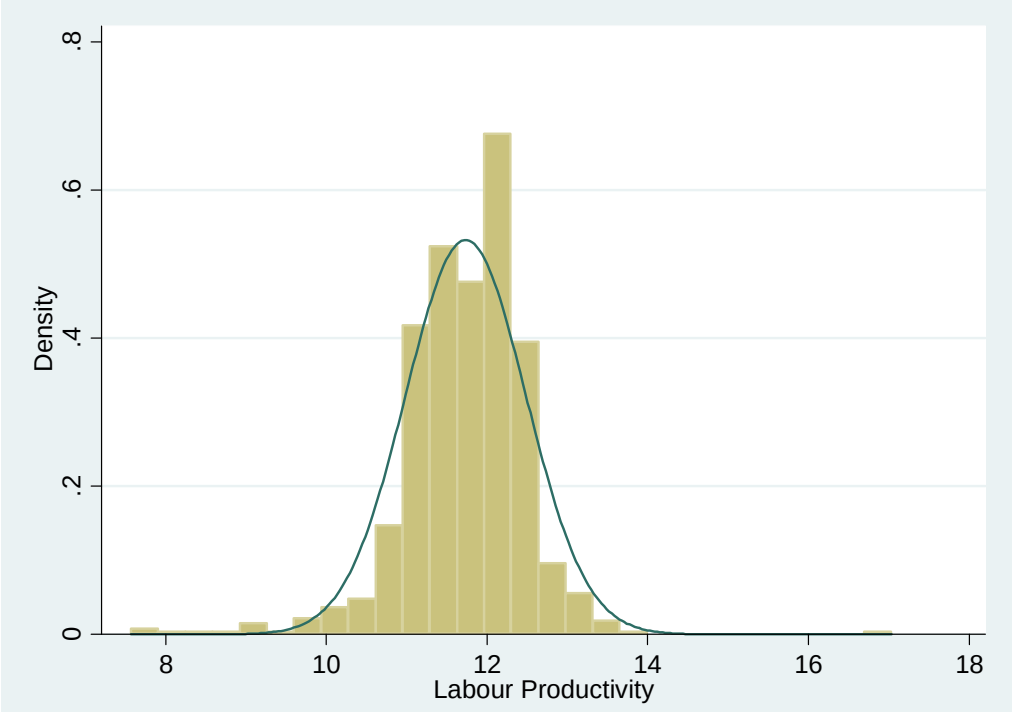


Figure 24: Labor Productivity

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Innovation and firm performance: An econometric analysis on the Belgian pharmaceutical sector

Richting: **master in de toegepaste economische wetenschappen: handelsingenieur-technologie-, innovatie- en milieumanagement**
Jaar: **2016**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Vandevenne, Sander

Datum: **22/08/2016**