Measuring Knowledge Spillovers Using Patent Citations: Evidence from the Belgian Firm's Data

Ruslan Lukach
Joseph Plasmans

An electronic version of the paper may be downloaded
- *from the SSRN website: www.SSRN.com*
- *from the CESifo website: www.CESifo.de*

MEASURING KNOWLEDGE SPILLOVERS USING PATENT CITATIONS: EVIDENCE FROM THE BELGIAN FIRM'S DATA

Abstract

This paper conducts a comprehensive study of patent citations in patents granted to Belgian corporate applicants by the United States and the European Patent Offices using qualitative response variable analysis. The analysis uncovered different patterns of citations in patents, which belong to different industrial classes. The studied citation data provides evidence of inter- or intra-firm and inter- or intra-industry knowledge spillovers, which are quite industry specific. In the framework of R&D cooperation stimulating policies, our results advocate for considering those differences in order to balance the effect of knowledge spillovers driven market incentives and direct regulation.

JEL Classification: C25, D21, O31, O34.

Keywords: knowledge spillovers, R&D, patent citation, limited dependent variable regression.

Ruslan Lukach
Faculty of Applied Economics
UFSIA-RUCA
University of Antwerp
Prinsstraat 13
2000 Antwerp
Belgium
Ruslan.Lukatch@ufsia.ac.be

Joseph Plasmans
Faculty of Applied Economics
UFSIA-RUCA
University of Antwerp
Prinsstraat 13
2000 Antwerp
Belgium
Joseph.Plasmans@ufsia.ac.be

1. INTRODUCTION

The presented research aims at tracking down knowledge spillovers in Belgium by following some of their "trails". There are no doubts about the importance of knowledge spillovers for economic growth. In a contemporary knowledge and technology driven economy the role of knowledge exchange and dissemination is often as important as, for example, the role of direct investment. Firstly, such spillovers allow a better penetration and diffusion of innovation among economic agents increasing their competitiveness (through lower costs of new technologies). Secondly, they stimulate cooperation in R&D by creating additional incentives for innovators to try to internalize knowledge flows and pool the resources in joint research efforts. Both of these types of effects eventually result in faster technological progress and economic growth in the country.

1.1. Knowledge, Spillovers, Competition and Economic Growth: the Theory

There is one important aspect worth mentioning, which is the difference between information and knowledge. Information is free to obtain and can be freely disseminated (almost) without any cost. On the other hand, knowledge is something, which can hardly be enumerated or codified. Knowledge is exactly that intangible asset that can have a certain economic value if properly used and commercialized.

Although a knowledge spillover is a phenomenon, which is quite easy to imagine, it is much more difficult to actually come up with an effective measure of it. According to the definition given by De Bondt (1996), the concept of a 'knowledge spillover' is specified as an 'involuntary leakage or voluntary exchange' of technological knowledge. Another definition, presented in Nieuwenhuijsen and van Stel (2000), describes knowledge spillovers as the situation, in which one economic agent benefits from R&D efforts of another economic agent without any tangible remuneration.

Bernstein and Nadiri (1988) classify knowledge spillovers as vertical or horizontal. Horizontal spillovers occur between competitors, and vertical spillovers flow between firms in different industries. Both these types of spillovers are directly linked to three factors of economic growth (Glaeser *et al.* (1992)): specialization, competition and diversity. Specialization is characterized by a higher intensity of intra-industry knowledge spillovers, while diversity goes together with more extensive inter-industry knowledge exchange. Subsequently, the competition factor affects the degree of inter-firm innovation flows.

There already exists an extensive literature about the importance of knowledge spillovers as a factor determining the optimal R&D strategies of firms. The very basis was laid by the study of d'Aspremont and Jacquemin (A&J (1988)), which has shown that the spillover effect influences the firms' willingness to cooperate in R&D. Their main conclusion states that, when knowledge spillovers are relatively strong, symmetric economic agents have strong incentives to cooperate with each other.

Gandal and Scotchmer (1991) advocate that it is more efficient to delegate research efforts to the agent with the highest ability by means of a Research Joint Venture (RJV) and this will lead to better private and social results. In the framework of A&J (1988), the study of Lukach and Plasmans (2000) investigated the optimal R&D and production strategies on firms, which have different capabilities in research and production. It concludes that in RJVs the firm with a lower marginal per unit cost of R&D conducts the overwhelmingly larger part of joint R&D. This finding provides the additional evidence of delegation described by Gandal and Scotchmer. Moreover, under conditions of stronger knowledge spillovers, the creation of an RJV leads to a better social welfare position.

In his famous 'Learning by Doing' model, Arrow (1962) points out that the competitive behavior of firms in the economy yields a smaller amount of aggregate investment compared to the socially desirable one. By stimulating firms to cooperate in R&D, the social planner shifts the mode of their R&D and production behavior from a competitive to a less competitive position with a higher value of the welfare function. In order to stimulate R&D cooperation among innovative firms, the regulator has a number of tools to achieve the desired effect. Such tools can be direct and tax subsidies, government's R&D investment and expenditures policies.

For example, in the described theoretical framework, the profit maximizing firms in industries with weak knowledge spillovers tend to compete in R&D, rather than to cooperate. Thus, if the regulator wants to induce R&D cooperation, it should come up with some tangible way to stimulate these firms' cooperation. On the other hand, in conditions with strong knowledge spillovers market forces provide a certain stimulus for companies to cooperate in research and thus the regulator can save resources by letting 'the nature doing its job'. Thus, if we consider the regulator's task in stimulating the economic growth by inducing R&D cooperation, it becomes clear that the correct assessment of the knowledge spillovers' environment can be one of the important elements for the success of such regulating policy.

1.2. Knowledge Spillovers: the Patent "Trail"

Our research relies heavily on the observation that the decision to patent a certain innovation is a 'strategic decision' (Jaffe et al. (1993)). If the firm decides to apply for a patent, it recognizes the potential value of the invention. Of course, this does not mean that not patented knowledge is worthless, but we should advocate that the patented knowledge is the one most likely to be commercialized. There have been a number of historical developments that created more incentives for firms to protect their innovation by means of a patent. First, the main purpose of a patent is to protect an 'individual' act of invention and to grant the inventor temporary rights to exclusively benefit from his/her innovative idea. But as the volume of 'invention by investment' (Kingston, 2001) increased over time, patents had to evolve to accommodate also the rights of the investor, who employed the inventor and/or provided him/her with means of research. Gradually, patents became protectors not only of the individual act of creativity, but also of the result of directed and managed investment in research and development.

Previously intended as a mere legal document, the patent turned into a tool of strategic competitive behavior. Firms build their intellectual property portfolios, trade patents, sell licenses, and create patent pools with other firms. In some industries patents have a crucial strategic importance. In the pharmaceutical industry, for example, it is not enough to patent only one molecular structure for the efficient protection of the invention. A small molecular variation of the same active component must also be patented. Thus, firms in chemical and pharmaceutical industries have to apply for numerous patents to protect their innovative effort and investment. Firms in other industries are also getting more active in patenting.

Plasmans *et al.* (1999) advocate that the entrepreneurial innovative behavior can be explained reasonably well by the entrepreneur's patenting behavior. They use the average propensity to patent (the number of patents per million constant PPP dollars of R&D expenditures) as a crude measure for the absence of knowledge spillovers and apply it to panel data for core EU countries and different industries (over the sample period 1989-1995).

In their contribution to the publication of The National Innovation System of Belgium, Capron and Cincera (2000) studied the technological performance of Belgian companies using patent and scientific-publication information as output indicators of technological and innovation activity from 1980 to 1996. This study aimed to determine the areas of comparative technological advantage and the regional distribution of innovative efforts in Belgium.

As we conclude that patents encapsulate an important part of the commercially valuable knowledge, it is rational to consider the advantages of utilizing patent data in the analysis of firms' strategic R&D behavior. It was said above that, technically, a patent is a legal document. Its content consists of the information verified and submitted afterwards to a controlling body. Thus, the patent citation is a certified evidence of previous knowledge used by the inventor(s), who obtain(s) a given patent. This previous knowledge, eventually, comes from the same patented domain. Hence, we conclude that the patent citation shows the spillover of one protected (i.e. recognized as potentially valuable) knowledge pool to another.

The study of patent citations has its own limitations. Advantages and disadvantages of using patent citation data are extensively discussed by Griliches (1990) and Jaffe *et al.* (1993). Patent citations are linked to the patenting procedure itself. They capture only the knowledge flows, which occur between patented 'pieces' of innovation, thus underestimating the actual extent of knowledge spillovers. Other means of knowledge transfer are not captured by patent citations. These means are: purchase of capital goods with embodied technologies, employment of engineers and other creative staff from other firms and institutions, voluntary knowledge exchange at conferences and in scientific publications, etc (see also Dumont and Tsakanikas (2001)). Though we should admit the importance of other non-patent-citation ways of knowledge exchange, it is necessary to point out that only the patent citation is to a large extent finalized as a representation of such exchange. The knowledge acquired informally or indirectly is likely to become an object of a dispute with other economic agents. Such disputes are common in business practice and they add a substantial amount of disturbance in data, when it is used for the analysis of innovative information exchange. Patent information is better protected from such disruption, because it clearly indicates the ownership over a particular piece of knowledge, which is protected by law. Patent disputes are also possible, but these are usually resolved quickly by the authoritative institutions[1].

There is one more shortcoming in patent citation data. The patent examiner has the right to add other citations he/she finds applicable in the given case, even though the inventor may not be aware of the inventions added. Addition of new citations by the patent examiners is widely practiced in the USPTO (United States Patent and Trademark Office), and the EPO (European Patent Office) examiners are able to do that as well. We have interviewed one of the USPTO examiners and she informed us that for the majority of US patents (as they are published in the published databases) it is not possible to effectively distinguish between the 'original' citation and a citation added by the examiner. Actually, we can consider those added records as an indication of knowledge spillovers, which are not officially recognized by the inventors, but from which he or she could have benefited as well. But all the other advantages, including a vast pool of data available, and,

---

[1] David and Hall (2000) point at the possibility of 'patent races', imitation activity, excess correlation between research projects of different enterprises, which can be considered as a waste of R&D resources. In that case it is advisable to study these phenomena and to determine the actual degree of R&D waste in certain industries.

most of all, explicitness of patent claims make the patent citation a good object for knowledge transfer analysis (Jaffe *et al.* (1993) and Verspagen (1997)).

An extensive study of Verspagen (1997) analyses patent citation data in relation to the productivity growth analysis for a cross-country, cross-sectional sample. He advocates that patent citations provide a measure for knowledge spillovers, which is different from other conventional measures. In addition, in 1999, Verspagen investigated the impact of large Dutch companies on domestic knowledge diffusion in the Netherlands by studying patent-to-patent citation data, provided by the EPO. This study employed a network analysis to analyze the place of Dutch multinationals in the domestic technology infrastructure.

Another Dutch study investigated the citations to Dutch-authored research papers on granted USPTO patents (Tijssen (2001)) to figure out the impact of the Dutch-authored innovations on other patented knowledge.

Our study derives itself from the previous investigation of knowledge spillovers in Belgium (see Plasmans and Lukach (2001)). This study presented a 'snap-shot' picture of knowledge flows through the mechanism of patent citations in all the Belgian firms' patent applications to the EPO, and the granted applications submitted to the USPTO in 1997. In this paper we conducted a comparative analysis of the data and tested the methodology for qualitative response variable analysis (probit, logit and Weibull modeling), which was based on the recent research of Jaffe and Trajtenberg (1998) who constructed a probit-type binary choice model of knowledge flows using patent citations. They have built a likelihood measure for the citation probability for any given patent pair. This allows a numerical evaluation of the 'citation frequency'[2] between different industries as well as between different geographical areas. The study of Jaffe & Trajtenberg was based only on data provided by the USPTO and concentrated on the industrial and national levels. We apply a similar technique to estimate the impact of knowledge spillovers (domestic and international) among different industries in Belgium, but we employ two sources, the USPTO and the EPO databases, thus widening our data's scope by building two compatible datasets.

In the current study we managed to achieve several important improvements and extensions for such analysis. First, we managed to obtain two compatible datasets from the EPO and the USPTO. Our fundamental data units are represented by all patents granted to Belgian firms by the EPO and the USPTO during the period between 1996 and 2000 inclusive. We consider not only the citations between the patents issued by the same office, but also the citations, in which one patent was issued by the EPO and another by the USPTO (cross-patent-office citation)[3]. We determine all the Belgian firms which have patents granted during the observed period and aggregate them in industries. Together with the industrial structure of spillovers, we are also able to build a geographic pattern of the Belgian patent citations.

2. OVERVIEW OF THE DATA

In this paper we analyze patenting data from two major sources: the EPO and the USPTO. The main purpose of this research is to create a picture of the 'patent-driven'

---

[2] According to the definition given by Jaffe and Trajtenberg (1998), a 'citation frequency' is a likelihood measure for the probability that any particular patent h granted in year t will cite some particular patent k granted in year $\tau \leq t$.

[3] To the best of our knowledge, we are not aware of substantial differences in European and US patent offices regulations, which could prevent us from using cross-patent office citations and from pooling these two samples.

knowledge spillovers in Belgium. In particular, we study the set of the patents obtained by Belgian firms during the five years from 1996 to 2000.

We are interested in a firm-level analysis of the data. Thus, we intend to adjust the list of considered firms using the shareholding and subsidiary relationship information collected by the National Bank of Belgium (NBB) and provided by the Bureau van Dijk (BVD's BelFirst database). The dataset, currently available to us, describes the firms' corporate governance structure as presented in their 1998 annual reports. Since 1998 is a median year in the observed period, we assume that it provides a good approximation for the typical firms' corporate governance structure for the period 1996 - 2000.

The raw dataset is presented by the patent citations indicated in the patents granted to Belgian corporate applicants by the EPO or the USPTO. Among those, we select all citations, corresponding to the applicants, which are identifiable in the BelFirst database. This allows us to adjust the ownership of patents belonging to the firms, which are involved in shareholder-subsidiary relationships. Thus, the primary object of our analysis is the patenting behavior of the Belgian firms.

Our primary source of information lies in 'patent citation pairs'. This kind of data supplies a good opportunity to study knowledge flows, indicated by the citation references in the patent application. For example, Jaffe and Trajtenberg (1998) and Verspagen (1999) conducted analyses of different patent citation datasets using different methodologies: econometric probit(logit)-type models, technological proximity matrices, and network analysis.

The dataset, which we use, provides data on all the applications that resulted in granted patents and already contains the citations indicated by the patent office investigators. We have obtained additional advantage by using the data from two different patent offices simultaneously. In the large majority of previous studies only one source was used and only one particular part of citations was studied. If the data were derived from the EPO database, then the sole citations studied were (mainly) the citations where one EPO patent cites another EPO patent (similarly with USPTO). In our case we use not only citations between patents issued by one patent office, but also the citations when the patent issued by the EPO cites the patent issued by the USPTO and vice versa. This is a very important new development, which significantly expands the sample and improves its representativeness.

In our primary dataset each line represents a single patent citation accompanied by several descriptive characteristics, which are: the patent number, the applicant's name, the applicant's country, the year in which the patent was granted, and the patent's class according to the International Patent Classification (IPC). In addition to that, we use the IPC-ISIC (ISIC – the International Standard Industrial Classification of All Economic Activities of the United Nations) concordance table compiled by Verspagen *et al.* (1994) to transform somewhat ambiguous IPC classes into more business-oriented groups indicated in the ISIC (compatible with the familiar NACE classification).

The patent citation pool is used to build another interesting dataset. We aggregate the citation data and summarize them in a firm-oriented sample, where the basic observation is the firm, which is 'identifiable' and can be linked to the NBB's information. Thus, there exists a number of variables, attributed to each firm: the total number of citations from patents applied for (both with the EPO and the USPTO), the number of citations from patents applied for with the EPO, the number of citations from patents applied for with the USPTO, the total number of citing patents applied for (both with the EPO and the USPTO), the number of citing patents applied for with the EPO, the number of citing patents applied for with the USPTO.

3. PRELIMINARY DATA ANALYSIS

After an initial processing of the data, we are able to make certain preliminary observations and conclusions. In this way we plan to build up some basis for the further model analysis. These results come from a basic aggregation of the data on the number of patents and citations, corresponding to different firms, industries, and countries. The source dataset is a pooled sample of all patents granted by the EPO and the USPTO to Belgian firms during the period between 1996 and 2000. It contains 2738 patents (1135 from the EPO and 1603 from the USPTO), which produce 11840 initial patent-to-patent citations (2513 originating from the EPO patents, and 9327 from the USPTO). Our conclusions and observations are grouped in several sections:

- geographic distribution of citations;
- firm-oriented distribution of patents and citations;
- the structure of the 'citation time lag' between citing and cited patents;
- the distribution of citations among different industries.

Geographic distribution of citations. First, we consider the basic geographic distribution of citations made by Belgian applicants. **Table 1** lists the ten countries, from which most cited patents originate. The list of 'top ten performers' consists of countries with more than 1% of total citations and covers the vast majority of these citations (96.2%).

According to the data from both patent offices, the USA patents are the ones cited the most. The second and third places are held by Japan and Belgium, although in the USPTO sample Belgium is the third, but in the EPO sample it is second. Rationally, we would have expected that Belgian patents will be the mostly cited (i.e. in the first place), driven by the argument that intra-firm and intra-country citations are more likely to occur (Jaffe & Trajtenberg (1998), pp. 6-7) than the more distant ones. Patents from the United States are the most frequently cited by Belgian companies, which allows us to assume the existence of a very strong 'transatlantic' knowledge flow. The 'Japanese' knowledge spillover channel is also quite strong. The other positions are occupied by the countries of the European Union (EU) and the Czech Republic, which is now holding the candidate status. Thus, we conclude that the 'geographic proximity' assumption is not strongly supported by the collected information: domestic patents are not the most frequently cited; although citing domestically cannot be rejected at the first site, because we observe the Belgian patents in the top three group. We can explain this finding by the fact that Belgium is a very small open economy relative to US and Japan.

In **Table 1** we also observe that the citations of American patents account for more or less comparable parts in the USPTO and the EPO samples. So, if we assume that the citations added by the examiners at USPTO do have a certain bias towards adding more citations to the American patents, this disturbance is not strong.

Table 1. Geographic distribution of patent citations of Belgian firms' 1996-2000 patents granted by the EPO and USPTO.

	Country	USPTO	EPO	Total
1	United States of America	42.15%	35.98%	40.62%
2	Japan	18.64%	17.11%	18.26%
3	Belgium	17.18%	20.42%	17.98%
4	Germany	6.07%	7.36%	6.39%
5	France	3.17%	4.15%	3.41%
6	Great Britain	3.19%	3.88%	3.36%
7	Italy	1.63%	2.15%	1.76%
8	Czech Republic	1.93%	0.85%	1.65%
9	Switzerland	1.55%	1.60%	1.56%
10	The Netherlands	0.90%	1.95%	1.16%
	Other	**3.59%**	**4.55%**	**3.84%**

Table 2. Percentage of <u>patents granted</u> to selected Belgian firms by the EPO and the USPTO and in total during the period 1996-2000.

		USPTO	EPO	Total
1	Agfa-Gevaert	38.86%	34.71%	37.14%
2	Solvay	9.30%	10.93%	9.97%
3	Janssen Pharmaceutica	7.86%	3.17%	5.92%
4	Esselte	2.62%	2.64%	2.63%
5	Raychem	2.74%	2.82%	2.78%
6	Dow Corning	1.87%	1.06%	1.53%
7	Xeikon	1.75%	1.06%	1.46%
8	Fina Research	2.43%	3.00%	2.67%
9	Glaverbel	1.62%	0.26%	1.06%
10	Heraeus Electro-Nite International	2.06%	1.85%	1.97%
11	Bekaert	2.18%	2.03%	2.12%
12	Plant Genetic Systems	2.74%	0.35%	1.75%
13	Innogenetics	1.75%	0.79%	1.35%
14	Smithkline Beecham Biologicals	0.94%	0.97%	0.95%
15	U.C.B.	1.00%	1.32%	1.13%
16	Michel Van De Wiele	1.62%	0.97%	1.35%
17	Picanol	1.06%	1.67%	1.31%
18	Owens – Corning	0.50%	0.09%	0.33%
19	Bayer Antwerpen	0.56%	0.88%	0.69%
20	Lernout & Hauspie Speech Products	0.87%	0.00%	0.51%
	Other	**15.66%**	**29.43%**	**21.37%**

Table 3. Percentage of <u>patent citations</u> generated in the patents granted to selected Belgian firms by the EPO and the USPTO and in total in 1996-2000.

		USPTO	EPO	Total
1	Agfa-Gevaert	34.19%	36.01%	34.58%
2	Solvay	9.37%	9.99%	9.50%
3	Janssen Pharmaceutica	5.46%	3.66%	5.08%
4	Esselte	4.71%	3.50%	4.45%
5	Raychem	4.17%	3.26%	3.98%
6	Dow Corning	3.20%	1.35%	2.80%
7	Xeikon	2.89%	1.51%	2.60%
8	Fina Research	2.04%	3.10%	2.26%
9	Glaverbel	2.52%	0.24%	2.04%
10	Heraeus Electro-Nite International	2.10%	1.79%	2.04%
11	Bekaert	2.14%	1.59%	2.03%
12	Plant Genetic Systems	2.29%	0.36%	1.88%
13	Innogenetics	1.66%	1.11%	1.55%
14	Smithkline Beecham Biologicals	1.27%	1.19%	1.25%
15	U.C.B.	1.23%	1.11%	1.21%
16	Michel Van De Wiele	1.30%	0.76%	1.18%
17	Picanol	0.89%	1.35%	0.99%
18	Owens – Corning	1.21%	0.04%	0.96%
19	Bayer Antwerpen	1.00%	0.72%	0.94%
20	Lernout & Hauspie Speech Products	1.14%	0.00%	0.90%
	Other	**15.22%**	**27.34%**	**17.80%**

Firm-oriented distribution of patents and citations. The second block of preliminary results deals with the 'top 20 performers' among the investigated firms. **Table 2** contains percentages of patents granted to these companies. **Table 3** presents the list of firms with the highest number of patent citations indicated in patents granted by the EPO and the USPTO during the period from 1996 to 2000. In this table we see that the top 20 companies (or 9.6% of all firms in our dataset) account for more than four fifths of the patent citations.

These results are closely related to the findings already presented by Plasmans *et al.* (1999), which are based on the study of the patenting behavior in 22 major industrial sectors of EU core countries during the period 1989 – 1995. This study indicates that a very limited number of companies actually accounts for the significantly larger part of patents granted by the EPO. In our data we observe a similar picture: the three companies at the top of the list own 56.02% of all patents issued between 1996 and 2000 (inclusive) by the USPTO and the 48.81% of patents issued by the EPO during the same period. Results of Capron and Cincera ((2000), p. 178) indicate that there was a tendency of growing concentration of patenting among the limited number of bigger players. Our findings show that during the recent half decade such concentration got even stronger. For example, Capron and Cincera ((2000), p.179) indicate that during the period 1980-1996 the top 20 Belgian patenting firms covered 49.8% of the EPO patents and 65.7 of the USPTO ones. But since 1996 we see that the top 20 players are responsible for 70.57% and 84.34% of patents respectively. Therefore, there is strong evidence of progressing concentration of patenting efforts in Belgium.

Table 4 shows the 'aggregated size' characteristics of companies mentioned above. We have obtained weighted consolidated turnover figures for each firm as the sum of the firms' own turnover and the turnovers of their subsidiaries weighted by the total participation share. A similar procedure was applied to the average annual employment as well. These variables serve as proxy measures for the firms' relative size characteristic.

Among these companies, some are quite big and known (Agfa-Gevaert, Solvay, Janssen Pharmaceutica, Glaverbel, Bekaert), but also some are much smaller companies (Esselte, Xeikon, Sofitech, Owens - Corning). This indicates that, although the biggest firms occupy the top three positions, there are also small companies engaging in the active patenting process. Thus, the large size of a company does not necessarily indicate that it will be more active in patenting than its smaller companions.

Table 4. Profiles of selected Belgian firms (based on 1998 annual financial reports). Source: Bureau van Dijk

	Name	Weighted Consolidated[4] Turnover (billion BEF)	Weighted Consolidated Average Employment (employees)
1	Agfa-Gevaert	66.1	5701.62
2	Solvay	82.9	3629.04
3	Janssen Pharmaceutica	48.2	3864.98
4	Esselte	5.4	571.98
5	Raychem	11.8	849.07
6	Dow Corning	2.0	394.00
7	Xeikon	4.8	274.00
8	Fina Research	2.6	474.00
9	Glaverbel	36.1	4278.78
10	Heraeus Electro-Nite International	3.1	471.00
11	Bekaert	33.61	4965.00
12	Plant Genetic Systems	1.1	167.00
13	Innogenetics	0.7	379.80
14	Smithkline Beecham Biologicals	26.4	1442.00
15	U.C.B.	36.5	3692.74
16	Michel Van De Wiele	6.9	583.89
17	Picanol	13.8	1764.98
18	Owens - Corning	16.0	906.00
19	Bayer Antwerpen	39.1	2575.00
20	Lernout & Hauspie Speech Products	4.3	297.61

The structure of the 'citation time lag' between citing and cited patents. Based on the data about the time lag between citing and cited patents we can derive the implications about the time structure of knowledge spillovers. **Figure 1** illustrates the

---

[4] We obtained weighted consolidated turnover figures for each firm as the sum of the firms' own turnover and the turnovers of their Belgian subsidiaries weighted by the total participation share. A similar procedure was applied to the average annual employment as well. These variables serve as proxy measures for the firms' relative size characteristic.

distribution of cited patents among the different years. The basic shape of the distribution is very much alike to the shape of the estimated citation frequency functions obtained by Jaffe and Trajtenberg (1998). The figure shows that recent patents (relative to the date of the citing patent) are more likely to be cited than older ones. The specifics of the patent examination process actually allows for the (small) negative citation lag values to occur as one patent can cite a published application for another patent, which is granted later than the citing patent itself, or when the cited patent is reissued.

We should point out another fact worth noting. The time structure of the citation lag is pretty much the same in both the USPTO and the EPO samples. This can serve as additional evidence of compatibility of the data in these two samples and that pooling of these two samples is feasible.
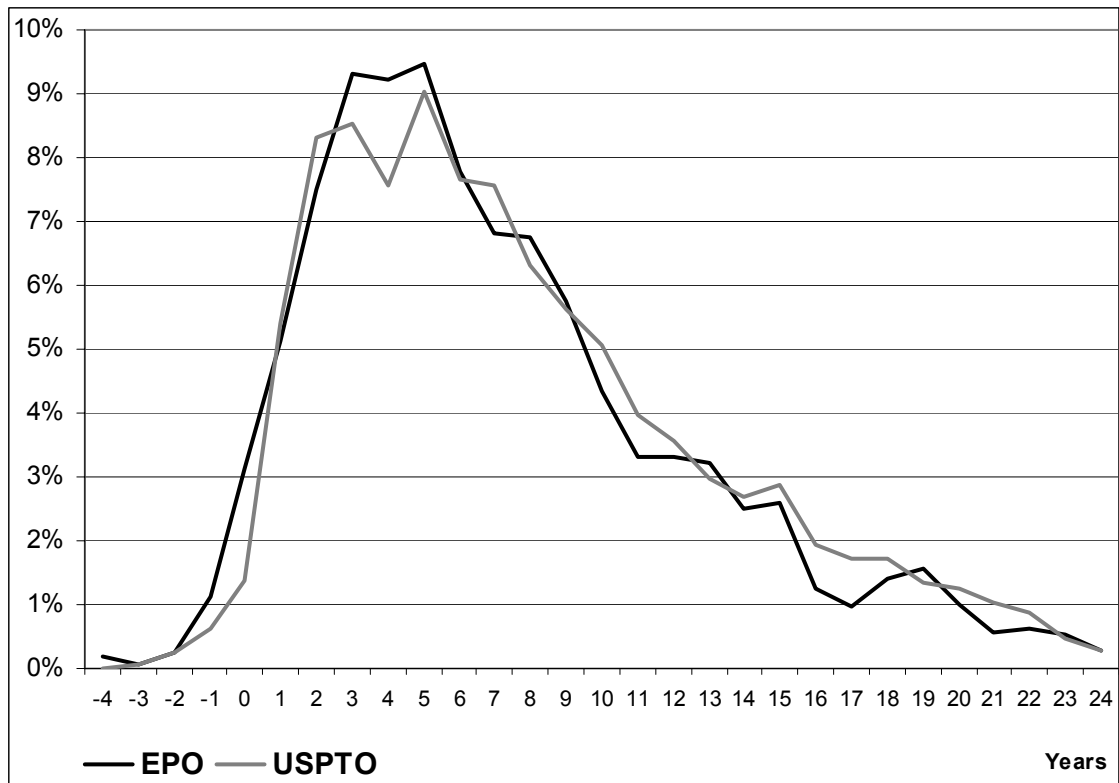
Figure 1. Time Lag Structure based on the Belgian patents granted by two different patent offices during 1996-2000.

Intra-industry citations in different industries. Let us consider the industrial structure of patent citations indicated in a pooled sample (the USPTO and EPO samples together). **Figure 2** presents the 'surface' of intra- and inter-industry citations. Each point on the surface represents the percentage of the citations between two industry codes in the overall sample. The industries presented in the figure were determined from the patent's main IPC, transformed using the IPC-ISIC concordance table (Verspagen *et al.* (1994)). In determining the category of a patent, which indicates several categories in application, we used the first category listed. **Table 5** lists all the industries indicated in the ISIC, accompanied by the corresponding percentages of citations calculated in the pooled sample.
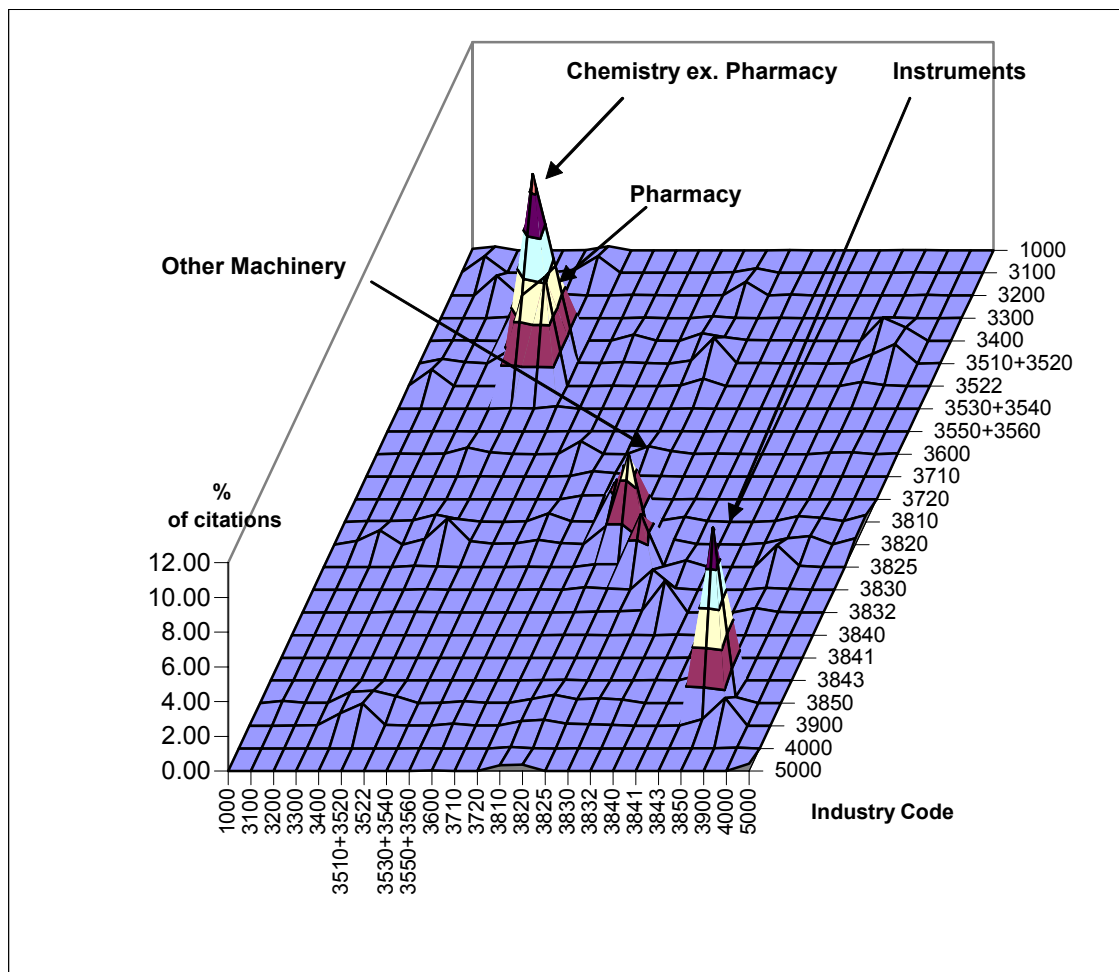
Figure 2. Relative Frequencies of 1996-2000 Citations Among Industries Surface (for industry codes see Table 5).

Actually, the figure is a graphical representation of the cross-industry citation matrix, calculated over the whole citation sample. This matrix closely resembles the widely used 'Yale matrix' (see e.g. Verspagen (1997)). As we expected, these diagonal elements are quite 'high' (see the 'Main Diagonal Ridge' on **Figure 2**), i.e. there is evidence that intra-industry citations are more numerous than the citations between different industries.

The highest peaks correspond to intra-industry citations in 'Chemistry excluding Pharmacy' (10.9% of all citations), 'Instruments' (10.1%), 'Pharmacy' (6.28%), and 'Other Machinery' (5.17%) industries. There is also a number of peaks outside the main diagonal, which point at active streams of knowledge flow between certain industries. These flows are primarily symmetric (relatively strong in both directions between two industries), but there are several asymmetric peaks corresponding to one-directional spillovers, such as between 'Paper, Printing and Publishing' and 'Instruments' (1.34%). Among the symmetric cross-industry knowledge flows, the strongest ones occur between 'Chemistry excluding Pharmacy' and 'Pharmacy' industries (5.6% of citations one way and 5.25% in the opposite direction), between 'Chemistry exd. Pharmacy' and 'Other Machinery' (1.48% one way and 1.51% in the opposite direction).

12

Table 5. Citation Percentages in Different Industries (as a fraction of all citations 1996-2000).

ISIC code	Industry	% of citations
3510+3520	**Chemistry, except pharmacy**	**21.29%**
3850	**Instruments**	**14.04%**
3522	**Pharmacy**	**13.41%**
3820	**Other machinery**	**11.55%**
3400	**Paper, printing and publishing**	**6.94%**
3810	**Metal products, exd. machines**	**6.38%**
3825	**Computers & office machines**	**5.38%**
3900	**Other industrial products**	**5.22%**
3100	Food, beverages, tobacco	2.56%
3832	Electronics	2.50%
3600	Stone, clay and glass products	2.23%
3200	Textiles, clothes, etc.	2.22%
3830	Electric mach., exd. electronics	1.79%
5000	Building and construction	1.23%
3710	Ferrous basic metals	0.70%
3720	Non ferrous basic metals	0.58%
1000	Agriculture	0.56%
3843	Motor vehicles	0.45%
3300	Wood and furniture	0.33%
3530+3540	Oil refining	0.22%
4000	Utilities	0.19%
3550+3560	Rubber and plastic products	0.15%
3840	Other transport	0.09%
3841	Shipbuilding	0.01%

There are eight major industries which account for the largest part (84%) of all citations considered: 3510+3520 (Chemistry excluding Pharmacy), 3850 (Instruments), 3522 (Pharmacy), 3820 (Other Machinery), 3400 (Paper, Printing and Publishing), 3810 (Metal Products, excluding Machinery), 3825 (Computers and Office Machines), and 3900 (Other Industrial Products). Individual shares of these industrial sectors are presented in **Figure 3**.
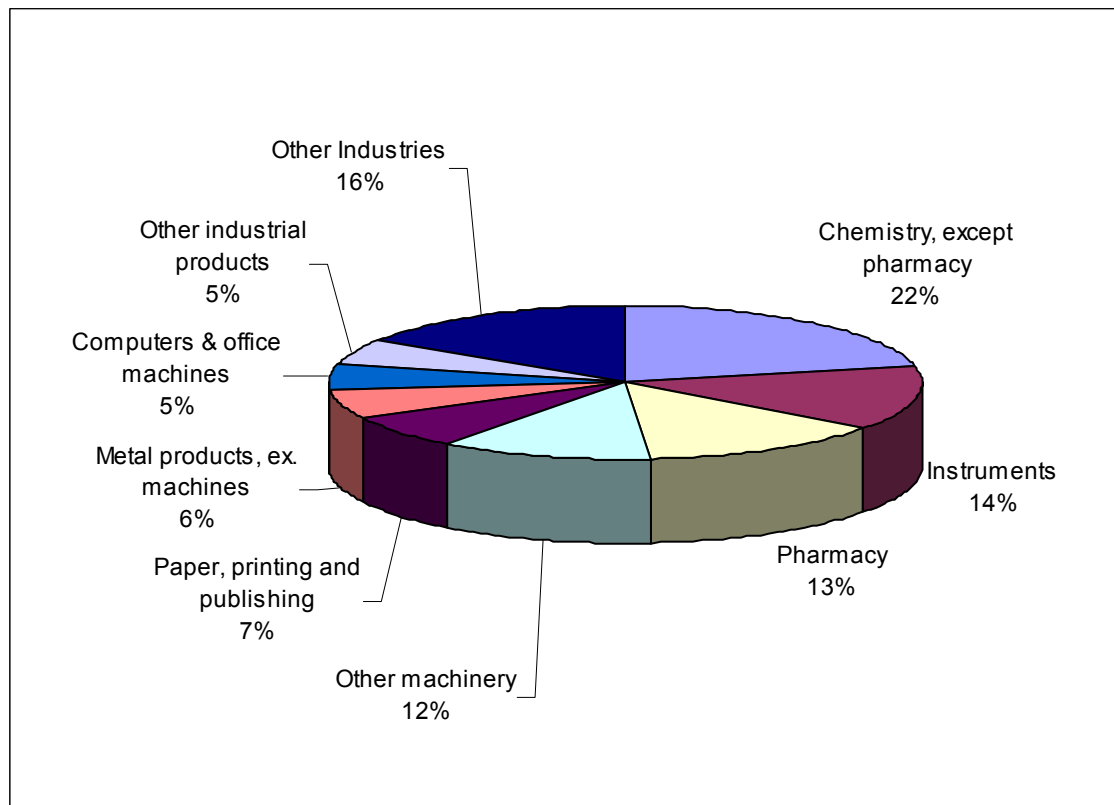
Figure 3. Percentages of 1996-2000 Citations by Industries

4. MODELS AND ESTIMATION

4.1. Citation Pairs Modeling

Now we intend to employ an econometric methodology to try to get a deeper insight into the knowledge spillovers pattern, 'encoded' into patent citation data. Previous researchers' experience (Jaffe and Trajtenberg (1998)) shows that patent citation data are best to be analyzed using a binary choice qualitative response model. The occurrence of a citation with particular attributes represents a binary event (occurrence or not), of which it is possible to estimate the probability of occurrence.

We focus our attention to one particular kind of events, which takes place as the patent citation occurs. The event is 'the citation occurs in the citing patent belonging to the particular industry class'. We study the estimated probability of this event and its relationship with a set of independent variables in order to derive analytical implications about the inter- and intra-industry/firm structure of knowledge spillovers. Our dependent variable is an indicator, which has value 1 if the citation occurs in the patent of a given particular industry, and equals 0 otherwise. We have chosen patents from the eight major industries (mentioned above and representing 84% of all citations during the period 1996-2000) to be analyzed by the model. We consider the following list of explanatory variables:

- an indicator that the patent citation has occurred between patents, owned by the same firm or institution (equals 1 if both citing and cited patents belong to the same firm, and equals 0 otherwise), it is represented by the dummy variable *SameFirm*;
- a 'concordance weighted' indicator that the citation has occurred between patents, belonging to the same ISIC-industry class (real number between 0 and 1 inclusive); it is represented by the variable *SameIndustry*;

- the year when the citing patent was issued represented by the variable *Year*;
- the value of a citation lag (i.e. the time difference between citing and cited patents, expressed in years); it is represented by the variable *CitationLag*.

We use the concordance percentage from the MERIT Concordance Table (the share of the patents in each IPC-class assigned to the corresponding ISIC category, see Verspagen *et al.* (1994)) to weigh the indicator variable for the citation occurred. For example, if two patents belong to the same industry, we calculate the product of their concordance percentages, obtaining in this way the measure of the 'citation occurrence' in this particular industry. The concordance percentage is the relative frequency of patents in the particular IPC class falling into a given ISIC class, thus their product in the citation pair represents a certain likelihood measure of the patent citation itself to fall into this ISIC class. Moreover, the usage of concordance percentages leads to the expansion of the modeled sample due to the fact that one IPC-class may fall into several industries with different weights.

It is possible to estimate several different specifications of the binary choice model: probit, logit or an extreme value distribution, such as a Weibull distribution (see Appendix and also Greene (2000), Chapter 19). In a preliminary estimation we compared probit, logit and Weibull specifications of the model. The goodness of fit criteria showed that the Weibull model predicted the probability of patent citations in our sample (in the majority of industries) best of all three[5]. The distribution of our dependent variable is most likely asymmetric, because the number of citations occurring in a certain industry (corresponding to non-zero elements in the sample) is certainly expected to be much smaller than the number of citations in other industries together (zero elements). Of course, the distributions of various independent variables are asymmetric too. As we return to the frequency graph (**Figure 1**) for the time lag variable, we see that it is quite asymmetric with more weight falling on the more recently granted cited patents. In our binary variables (such as the event indicator and variable *SameFirm*) too, we see that zero values are more numerous than non-zero ones. This is also true for the non-binary variable *SameIndustry*.

We have run another series of preliminary computations to determine the best way in obtaining the general implications about the knowledge spillovers. In this step we conducted the estimation of a Weibull model in two different sub-samples: the sample consisting of citations indicated in EPO data, and the sample consisting of citations indicated in USPTO data. We came to the conclusion that it is reasonable to use the pooled sample to study the knowledge spillovers generated by the patent citation behavior of Belgian companies. There are three arguments in support of this decision. Firstly, after estimation of the model in two different samples we observed that 25 out of 32 (8 industry equations with 4 slopes each) estimated slope coefficients[6] have the same sign with mostly sufficient statistical significance. Secondly, because the object of our interest is the overall picture of patent citation induced knowledge spillovers generated by the Belgian companies, it is preferable to consider these patent citations in a pooled sample. And thirdly, the close similarity in the time lag structure, firms' patent and citation percentages, provides additional arguments for the compatibility of these two datasets.

Estimation results. There are several notes to be made about the interpretation of the results. Among the explanatory variables in our model we have one binary variable, two integer variables, and one coming from the real numbers set. The corresponding

---

[5] The goodness of fit score is best in 5 out of 8 industries for Weibull: 'Chemistry excl. Pharmacy', 'Instruments', 'Pharmacy', 'Other Machinery', 'Paper, Printing and Publishing'.

6 In this model the slope coefficient is the product of the corresponding equation coefficient and the value of the distribution density function calculated at the means of the regressors (see Appendix).

slopes or marginal effects (see Appendix) are presented in the output tables (**Tables 6 – 13**). In these tables we marked the variables, which yielded different signs in the separate USPTO and EPO samples with a star superscript. Below we are going to refer to some results of the estimations conducted in different samples. Because of lack of space, only results from the pooled sample Weibull estimation are presented in tables.

Table 6. Weibull regression results in the 'Chemistry, excluding Pharmacy' industry.

3510+3520	Coefficient	Slope	Std. Err.	Chi-Square	Prob.
Intercept	-33.4727		11.60921	8.3133	0.0039
SameFirm	-0.1747	-0.05717	0.02176	64.4765	<.0001
SameIndustry	0.41329	0.135242	0.01794	530.7173	<.0001
Year	0.01693	0.005539	0.005811	8.4871	0.0036
TimeLag[*]	-0.00195	-0.00064	0.001507	1.682	0.1947

Chemistry, excluding Pharmacy (Table 6). The results in the 'Chemistry excl. Pharmacy' industry indicate that there is evidence of a negative relationship between the *SameFirm* dummy and the probability of the citation. This fact was also indicated in both the individual USPTO and EPO samples. It allows us to conclude that a 'chemical' patent is more likely to cite a patent belonging to a different firm, rather than its own, i.e. this industry is more oriented towards the usage of the other firms' patented knowledge.

The coefficient for the *SameIndustry* variable points at a higher likelihood of a citation to occur in the same industrial class. This is quite reasonable because of the special nature of the chemical industry. Chemical patents usually protect either molecular structures or technological sequences for their synthesis; thus this knowledge does not go far beyond the scope of the own industry.

The positive coefficient for the variable *Year* indicates that the citation is more likely to occur in the relatively newer chemical patents. Concerning the time difference between the citing and cited patents, it is difficult to make a definitive conclusion about the relationship between the time lag and the likelihood of a citation due to the (completely) insufficient statistical significance of the estimator. Moreover, different sub-samples yielded different conclusions for this coefficient (positive in the EPO and negative in the USPTO).

To summarize the results, we may state that in the 'Chemistry, excluding Pharmacy' industry the 'citation-induced' knowledge spillovers tend to be inter-firm, but intra-industry. The question about how the 'age' of a cited patent affects the probability of citation in this industry requires additional inquiry. As for citing patents, there is a clear indication that the newer patents tend to generate more citations.

Table 7. Weibull regression results in the 'Instruments' industry.

3850	Coefficient	Slope	Std. Err.	Chi-Square	Prob.
Intercept	49.01282		12.46039	15.4723	<.0001
SameFirm	0.05167	0.013128	0.02386	4.6888	0.0304
SameIndustry	-0.92024	-0.2338	0.01864	2436.454	<.0001
Year[*]	-0.02402	-0.0061	0.006236	14.8376	0.0001
TimeLag[*]	0.005356	0.001361	0.001596	11.2568	0.0008

Instruments (Table 7). Our data for this industry provides evidence of the positive relationship between the likelihood of patent citation and the fact that the citing and cited patents both belong to the same firm. There is a very strong evidence that in the 'Instruments' industry the probability of the citation is much lower when citing and cited patents belong to the same industry class.

There are mixed signals coming from two samples concerning the estimated slope coefficient for the *Year* and *TimeLag* variables, but the pooled sample delivers these coefficients with a quite high statistical significance. It corresponds to the sign, obtained from the largest (USPTO) sample, which dominates the smaller (EPO) one. It shows that the more recent citing patents indicate a smaller number of citations, and that the older patents are more likely to be cited by the patents in this industry.

The final 'verdict' for the Instruments industry states that it strongly favors inter-industry knowledge spillovers, and leans towards intra-firm patent citation. Knowledge in this industry does not seem to 'depreciate' fast, which fact is supported by the evidence of a significant positive relationship between the time lag between the patents and the probability of citation.

Table 8. Weibull regression results in the 'Pharmacy' industry.

3522	Coefficient	Slope	Std. Err.	Chi-Square	Prob.
Intercept	21.69069		11.76131	3.4012	0.0651
SameFirm	-0.24386	-0.06506	0.02128	131.3063	<.0001
SameIndustry*	0.21525	0.057429	0.0203	112.4091	<.0001
Year*	-0.01054	-0.00281	0.005887	3.2048	0.0734
TimeLag	0.005544	0.001479	0.001588	12.1814	0.0005

Pharmacy (Table 8). The 'Pharmacy' industry shows a lower likelihood of the intra-firm citation and a higher probability for knowledge spillovers in the same industry (but we get different signs in the sub-samples). Thus, in general we expect a knowledge exchange that is more intensive among different firms, but in the limits of the same industry.

It appears that the more recent pharmaceutical patents indicate fewer citations, although the coefficient is moderately significant and varies in sign in the two sub-samples. The coefficient for the *TimeLag* variable is positive and significant, and it points at a slower knowledge 'depreciation' in this industry.

Table 9. Weibull regression results in the 'Other Machinery' industry.

3820	Coefficient	Slope	Std. Err.	Chi-Square	Prob.
Intercept	-49.9749		11.91326	17.5972	<.0001
SameFirm	0.14916	0.036699	0.02311	41.6769	<.0001
SameIndustry	0.09567	0.023538	0.01898	25.4147	<.0001
Year*	0.02545	0.006261	0.005963	18.2167	<.0001
TimeLag	-0.02034	-0.005	0.001562	169.579	<.0001

Other Machinery (Table 9). The title for this industry is quite ambiguous and makes it difficult to extract particular policy implications, although there is a significant number of patent citations covered by it. The results show that in this industry the time difference between two patents negatively affects the probability of the citation, and that newer patents employ a larger number of external citations. Regarding the existence of intra-firm spillovers, we find strong support for this in the pooled data and in both individual sub-samples as well. It also provides strong evidence for a more intra-industry knowledge exchange.

Table 10. Weibull regression results in the 'Paper, Printing and Publishing' industry.

3400	Coefficient	Slope	Std. Err.	Chi-Square	Prob.
Intercept	49.1891		12.57772	15.2945	<.0001
SameFirm	-0.16522	-0.02977	0.02291	51.9971	<.0001
SameIndustry	0.22454	0.040453	0.02094	114.9527	<.0001
Year	-0.02419	-0.00436	0.006296	14.7621	0.0001
TimeLag	0.01228	0.002212	0.001662	54.5651	<.0001

Paper, Printing and Publishing (Table 10). As the estimations in both sub-samples 'agree' on signs of the coefficients, and as we already provided a number of explanations for coefficients above, we will try to keep short. This industry exhibits a more inter-firm, but intra-industry pattern of patent citations. Newer patents cite less and the older patents are more likely to be cited.

Table 11. Weibull regression results in the 'Metal Products, excluding Machines' industry.

3810	Coefficient	Slope	Std. Err.	Chi-Square	Prob.
Intercept	2.91598		13.13631	0.0493	0.8243
SameFirm	0.23553	0.03895	0.02751	73.2795	<.0001
SameIndustry	0.27673	0.045762	0.02169	162.8431	<.0001
Year[*]	-0.00093	-0.00015	0.006575	0.0198	0.888
TimeLag	-0.01997	-0.0033	0.001693	139.1178	<.0001

Metal Products, excluding Machines (Table 11). The industry is more 'turned towards itself'. The probability of the patent citation's occurrence is higher when two patents belong to the same firm and come from the same industry. Thus, the external knowledge spillovers in the Metal Products industry are weak and some R&D cooperation inducing measures may be advisable.

There is strong evidence for a negative relationship between the time lag and the likelihood of the citation: recent patents are cited more, indicating a faster knowledge depreciation in this industry. The effect of the patent's issue year on citation is left undetermined for the reason of no statistical significance (plus the disagreement of the two sub-samples).

Table 12. Weibull regression results in the 'Computers and Office Machines' industry.

3825	Coefficient	Slope	Std. Err.	Chi-Square	Prob.
Intercept	96.65918		14.03608	47.4236	<.0001
SameFirm	0.23978	0.034592	0.02833	71.6273	<.0001
SameIndustry	-0.33882	-0.04888	0.02011	283.9024	<.0001
Year	-0.04786	-0.0069	0.007025	46.4103	<.0001
TimeLag	0.02061	0.002973	0.001944	112.3771	<.0001

Computers and Office Machines (Table 12). This industry deserves special attention due to its importance in the current technology driven times. The model was able to produce statistically very significant coefficients with full concordance of the two sub-samples. The data strongly advocate for more intra-firm knowledge usage rather than inter-firm. Concerning the inter-industry knowledge spillovers, there is a strong support for it, meaning a higher likelihood that the knowledge from other industries will be used.

The model provides strong support for the positive dependence of the probability of citation on the time difference between patents, thus indicating the relatively higher rate of older knowledge utilization. We also see that newer patents are less likely to cite the knowledge from other patent documents.

Table 13. Weibull regression results in the 'Other Industrial Products' industry.

3900	Coefficient	Slope	Std. Err.	Chi-Square	Prob.
Intercept	-102.44		13.32543	59.0987	<.0001
SameFirm	0.02239	0.003269	0.02513	0.7935	0.373
SameIndustry	0.31166	0.045508	0.02264	189.575	<.0001
Year	0.05173	0.007553	0.00667	60.1418	<.0001
TimeLag	0.01037	0.001515	0.001792	33.5008	<.0001

Other Industrial Products (Table 13). The coefficient for the *SameFirm* dummy is not significant at all for a conclusion. The indication for mainly intra-industry patent citing behavior is very strong. The more recent citing patents cite more, and older patents have higher chances to be cited.

4.2. The Intra-Firm/Intra-Industry Positioning Of Industries

To obtain a better view on general results of modeling the knowledge spillovers, we present a map of relative positions for particular industries with relation to the likelihood of intra-firm and intra-industry citation. **Figure 4** is a two-dimensional graph, where on the horizontal axis we plot the slope coefficient for the *SameFirm* dummy and on the vertical axis is the slope coefficient for the *SameIndustry* variable. Such an arrangement is based on the interpretation of the obtained slope coefficients. A slope coefficient in our model describes the change in the probability of a patent citation at the means of the regressors (Greene (2000), p. 879).
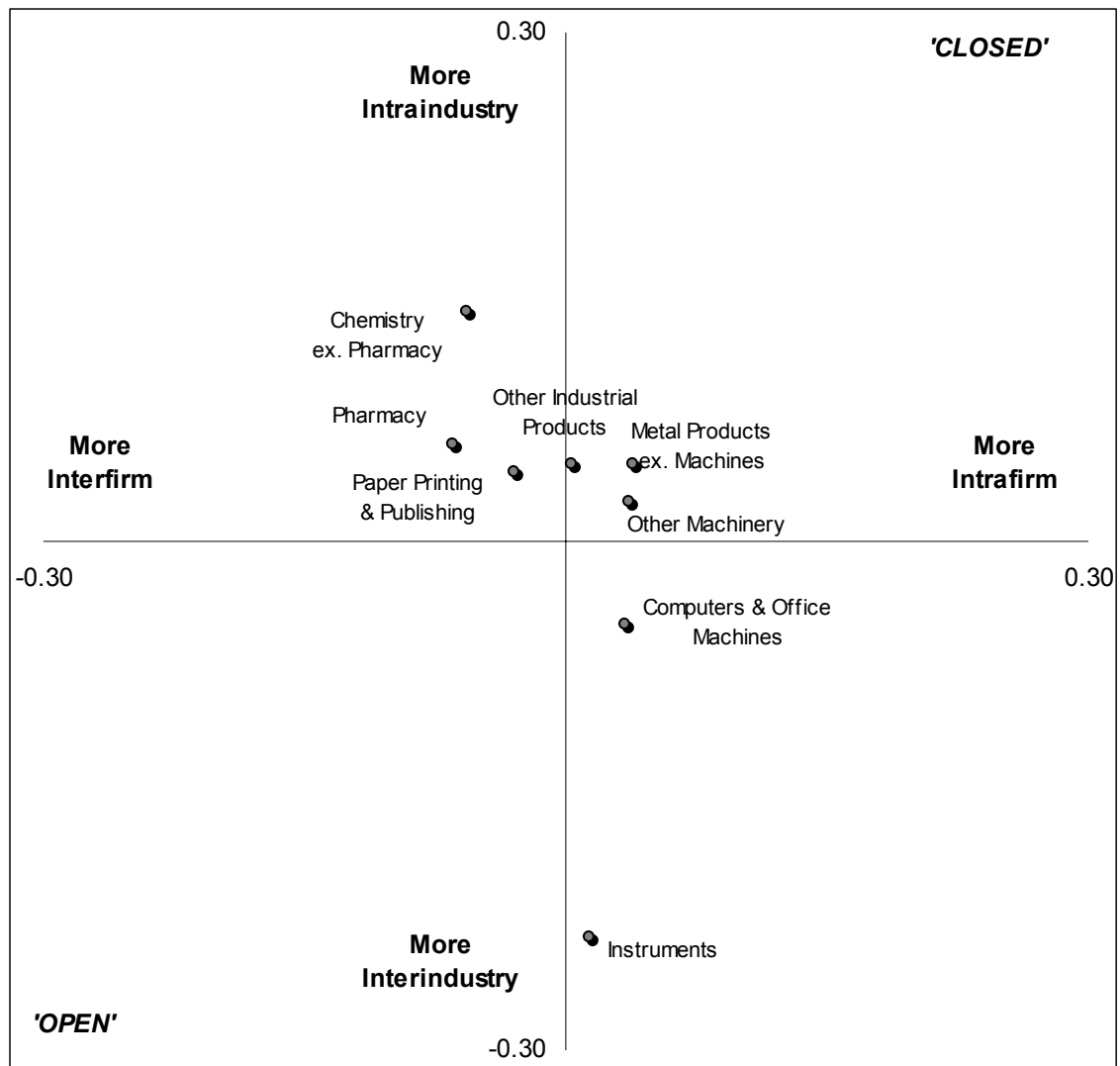
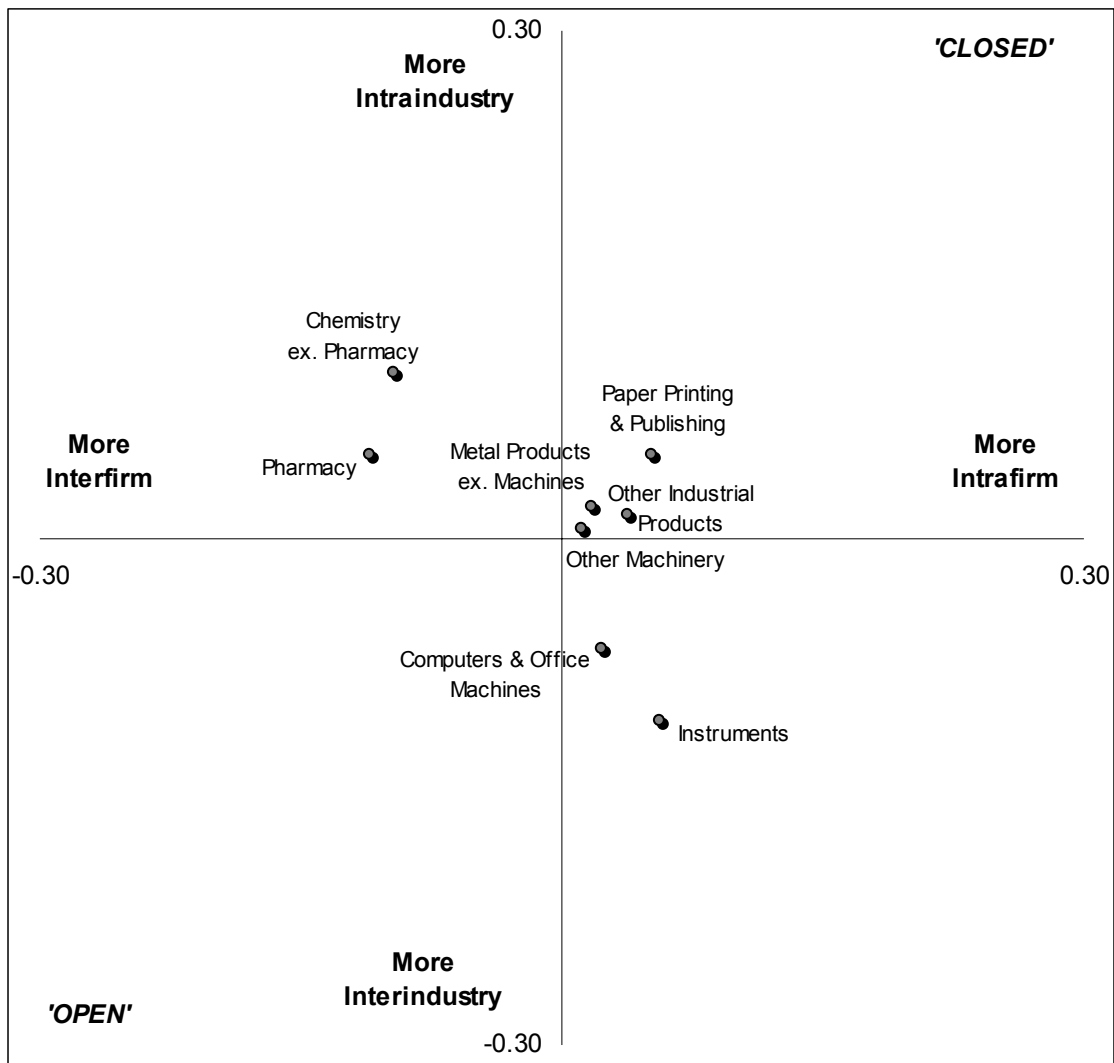**Figure 4. Positioning of Industries with Relation to Intra-firm and Intra-industry Knowledge Spillovers (based on the Weibull model).**

Thus, a pair of such coefficients for a particular industry points at its unique position on the map relative to other industries and the origin, which can be interpreted in the following manner. The bottom-left quadrant of the map contains industries, which are more inclined towards inter-firm and inter-industry knowledge spillovers (the probability of citation decreases for patents belonging to the same firm and industry class). We can call such industries 'open'. On the opposite, the top-right quadrant of the map contains more 'closed' industries, which favor intra-firm and intra-industry citation (the citation is more likely if the patent pair comes from the same industry and is owned by the same owner). The bottom-right quadrant combines a higher likelihood of inter-industry, but intra-firm spillovers, which, for example, can be the case in complex technologies (see Kingston (2001)). And the top-left quadrant combines intra-industry and inter-firm spillovers correspondingly.

On **Figure 4** we see that there are no truly 'open' industrial sectors considered in our pooled sample. A group of 'closed' industries consists of 'Metal Products, exd. Machines', 'Other Machinery', and 'Other Industrial Products' classes. The 'Metal Products, exd. Machines' industry is the most 'internally-oriented' among them all. The 'Instruments'

industry is in an interesting position, where it is almost indifferent towards the intra- or inter-firm citation, but it is strongly on the side of inter-industry knowledge utilization. The 'Computers and Office Machines' industry is open for inter-industry knowledge spillovers, and is less inclined towards using the knowledge of other firms. The 'Chemistry, excluding Pharmacy' and the 'Pharmacy' industry itself, exhibit greater openness for inter-firm knowledge spillovers, which preferably do not go far beyond the scope of the same industry. A similar behavior can be observed in the case of 'Paper, Printing and Publishing'.

As we think about the political implications of such analysis, it is recommended to turn to the main conclusions of d'Aspremont and Jacquemin (1988) and Lukach and Plasmans (2000). They state that under conditions of stronger knowledge spillovers, symmetric and asymmetric innovative firms have more incentives to engage in R&D cooperation, which results in a larger R&D investment and innovative product output. For a regulator whose goal is to induce R&D cooperation, it is important to balance the market incentives, created by stronger knowledge spillovers, and the regulative incentives.

The general guidelines for the regulator, derived from our study, can be summarized by observing the relative positioning map along the horizontal axis. The industries in the right quadrants appear to be more oriented towards intra-firm knowledge spillovers, thus there are rationales for stimulating the R&D cooperation among the firms in these industries. On the other hand, the industries, situated in the left quadrants, operate under conditions of stronger knowledge spillovers, and there are market incentives, which drive the companies towards more cooperation. The regulator in this case can stand on less intrusive positions, observing the 'natural' tendencies towards cooperation and maybe stimulating only the most interesting joint R&D projects and/or alliances.

4.3. Agfa-Gevaert: An Outlier Problem

Looking back at the data in **Table 3**, it does not take much attention to notice that one firm towers far high above the others. Patent citations coming from the patents owned by Agfa-Gevaert account for 34.58% of the pooled sample. The runner-up Solvay has a considerably lower share (9.5%) of citations. Thus, it is reasonable to assume that Agfa-Gevaert generates a strong outlier effect in our data. As an experiment we delete all the citations coming from the Agfa-Gevaert's patents from our sample, re-estimate the model and built another intra-firm/intra-industry positioning map on the basis of the new results. This new map is presented on **Figure 5.** If we compare the newly obtained map with the previous one, we can make the following observations:

- only one industry ('Paper, Printing and Publishing') out of eight changed its quadrant as a result of the outlier's deletion. The 'Paper, Printing and Publishing' industry 'took off' from the moderately inter-firm and intra-industry inclined position and 'landed' in the point with relatively strong inclination towards intra-firm and intra-industry knowledge spillovers;
- 'Chemistry excl. Pharmacy' and 'Pharmacy' industries slightly moved in the limits of the same quadrant in the direction of more inter-firm knowledge spillovers;
- 'Metal Products excl. Machines', 'Other Industrial Products' and 'Other Machinery' industries moved to the neighboring positions (although closer to the origin) in the same intra-industry intra-firm quadrant;
- 'Computers & Office Machines' moved slightly to the position with weaker intra-firm and a bit stronger inter-industry spillovers;
- the 'Instruments' industry now is more in favor of intra-firm and less, but still evident, of inter-industry knowledge utilization.

Figure 5. Positioning of Industries with Relation to Intra-firm and Intra-industry Knowledge Spillovers (excl. Agfa-Gevaert and based on the Weibull model)

Thus, comparing **Figure 4** and **Figure 5** it becomes visible that deletion of the Agfa-Gevaert's citations from the sample lead to certain changes in the estimation results, although not extremely dramatic. The majority of industries shifted around, but stayed in the same quadrant as before. From the size of the industry's shift we can judge the influence the outlier had in the sample. The jump of the 'Paper, Printing & Publishing' industry from one quadrant to another illustrates the especially active position of Agfa-Gevaert in this industry. Without Agfa-Gevaert this industry seems to be more closed, while in the full sample it looks more open. We also conclude that Agfa-Gevaert has a critical mass in the 'Instruments' industry, as we observe its quite a long jump from the very strong to moderate inter-industry spillovers' position, but into the relatively strong intra-firm position. Other industries did not exhibit such strong qualitative change which is understandable. The conducted experiment showed that Agfa-Gevaert indeed creates a quite noticeable disturbance as an outlier in the observed sample, but only in the industries of its own specialization.

We realize with a certain degree of surprise that 'Chemistry excl. Pharmacy' industry did not exhibit a considerable shift, although chemistry is traditionally considered the main area of Agfa-Gevaert's specialization. On our opinion, this fact has to do with a time frame of the analyzed data. Our data spans only over last 5 years during which Agfa-Gevaert was in the process of major shift of its production specialization from the traditional photographic products to the area of digital visual equipment, x-ray devices, and computer periphery. Thus, it becomes understandable why positioning of the chemicals industry did not change much after Agfa-Gevaert was excluded from consideration.

5. CONCLUSIONS

The objective of this study was to investigate the patenting and patent citation behavior of the Belgian private firms using the 1996-2000 patent citation data from the EPO and the USPTO. The attention of this study was concentrated on the patent citation behavior of Belgian firms using the Weibull binary response variable model (preferred after comparing it with probit and logit specifications). We have conducted an extensive preliminary analysis of the data and built empirical models. The results can be summarized in the following statements:

1. First, the study of the patent citation data proved to be useful in the analysis of innovation behavior of Belgian firms. A preliminary analysis has indicated that the majority of the patenting is conducted by a (very) small number of firms being different in size (represented by the consolidated weighted turnover and consolidated weighted average annual employment).

2. The estimated probability of a patent citation, calculated given a particular set of factors (*SameFirm* dummy and *SameIndustry* variable, time lag between the citing and the cited patents, the year in which the citing patent was issued), can be used as an efficient measure of the size of knowledge spillovers in a certain industry, and can be applied for various competitive behavioral models. Once the special feature of the industry is determined (such as the likelihood of inter- or intra-firm spillovers and the likelihood of inter-industry knowledge exchange), we obtain an understanding of the knowledge spillovers intensity. Additionally, the relationship between the likelihood of citation and the size of a time lag between the citing and cited patents indicates the speed on 'citable' knowledge depreciation.

3. In particular, analyzing the relative positioning of different industries depending on their attitude towards inter-firm knowledge spillovers allows us to infer implications concerning the necessity of measures to stimulate R&D cooperation. For example, it is preferred that the regulator proposes more R&D cooperation stimulating policy towards the industries with less intensive knowledge spillovers, and employs less regulation in the industries where such spillovers are stronger and create more natural incentives for firms to cooperate in R&D.

4. The occurrences of patent citations in Belgian patents in eight major industries were studied. As a result of this study, it is possible to determine the 'level of openness' of different industries toward inter-industry and inter-firm knowledge exchanges through patent citation. Industries with more complex technologies (such as 'Computers & Office Machines' and 'Instruments) are more open towards inter-industry knowledge flows. On the other hand the industries with 'uniform' technological orientation (such as 'Chemistry', 'Pharmacy', 'Metal Products', and 'Paper, Printing & Publishing') stay more oriented at intra-industry knowledge utilization. In 'Chemistry', 'Pharmacy' we conclude higher intensity of inter-firm knowledge exchange, which would indicate a

better environment for R&D cooperation. Firms in the other industries favor more internal knowledge flows and have fewer incentives to cooperate in R&D.

5. We conclude that Agfa-Gevaert introduces substantial outlier's disturbances in the pooled sample. Deletion of this firm's patents from the dataset leads to a noticeable change in the positions of the industries, where Agfa-Gevaert exhibits stronger specialization: 'Paper, Printing & Publishing', and 'Instruments' industries. Such disturbances were much weaker in other industries, although, in the case of Chemistry excl. Pharmacy (AGFA is considered to be a chemical firm) industry such a conclusion is a bit surprising.

Summarising these findings, we up with an argument that public authorities should use differentiated measures to regulate R&D activities (and especially R&D cooperation) by firms in different industries. The existing knowledge spillovers create certain market-driven incentives inducing firms to cooperate. It is possible for a regulator to use these incentives in combination with particular regulatory measures to achieve the desired effects, whether it is higher R&D investment or improved diffusion of knowledge in the economy.

For the firms operating in the industries with conditions of stronger knowledge spillovers, the regulator can adopt a less intrusive policy (which is usually a 'cheaper' one as well), observing the 'natural' tendencies towards cooperation and possibly stimulating only the most interesting joint R&D projects and/or alliances. More regulators' attention must be paid to the firms in the industries with weaker knowledge spillovers, because these firms tend to invest in R&D in a more competitive way. It will require a bigger effort from the regulator to stimulate R&D cooperation by direct subsidies and/or advantageous tax measures in such cases. The major outcome of such a successful policy will eventually surface in faster economic growth.

6. APPENDIX

The Weibull Binary Choice Model for Patent Citations

The pooled dataset contains a list of citation pairs, which have already occurred. Thus, if we consider the probability of a citation to occur in patent pairs from our dataset, it is equal to 1. Within this population, we select several other sub-events, for example 'the citation has occurred in the citing patent coming from industry A'. The basic Weibull model can be specified:

$P(y_i = 1) = F(\beta' x_i) = 1 - \exp(-\exp(\beta' x_i))$, $i = 1, 2, .., n$, where n is the number of observations. In our case we have:

$\beta' x_i = Const_i + \beta_1 SameFirm_i + \beta_2 SameIndustry_i + \beta_3 Year + \beta_4 CitationLag_i + \varepsilon_i$.

The dependent variable Y_i is an indicator that the patent citation occurred in the particular industry (see above). It is also known that the estimated coefficients of a Weibull model (probit and logit as well) do not yield the value of the marginal effect of the independent variable. For the Weibull model, the marginal effect for an independent variable is calculated as the product of the corresponding equation coefficient and the value of the density function calculated at the means of regressors:

$$\frac{\partial F(x_i'\hat{\beta})}{\partial x_{ij}}\bigg|_{x_i = \bar{x}_i} = f(\bar{x}_i'\hat{\beta})\hat{\beta}_j, \ i = 1, 2, .., n, \ j = 1, 2, .., k,$$

where $f(\overline{x}_i^{'}\hat{\beta}) = \exp(\overline{x}_i^{'}\hat{\beta} - \exp(\overline{x}_i^{'}\hat{\beta}))$ is the Weibull density function calculated at the mean of the estimated structural part of the model[7].

Since we have one binary variable in the model, another method for calculating the marginal effects should be mentioned. For a binary independent variable *b*, the marginal effect (also called *slope*) is calculated as $P\{Y=1|\overline{x}_*, b=1\} - P\{Y=1|\overline{x}_*, b=0\}$. However, Greene ((2000), p. 817) indicates that 'simply taking the derivative with respect to the binary variable as if it were continuous provides an approximation that is often surprisingly accurate'. Thus, we calculate the slopes for the binary independent variables in our model in the same way as we do this for non-binary variables.

7. REFERENCES:

d'Aspremont, C. and A. Jacquemin (1988), Cooperative and Noncooperative R&D in Duopoly with Spillovers, The American Economic Review, 78, pp. 1133-1137.

Bernstein, J. and M. Nadiri (1988), Inter-industry R&D Spillovers, Rates of Return, and Production in High-Tech Industries, The American Economic Review, 78, pp. 429-434.

Capron H. and M.Cincera (2000), Technological Performance, Chapter 8 of The National Innovation System of Belgium, edt. by H. Capron and W. Meeusen, Physica-Verlag.

De Bondt, R. (1996), Spillovers and Innovative Activities, International Journal of Industrial Organization, 15, pp. 1-28.

Devid, P. and B. Hall (2000), Heart of Darkness: Public-Private Interactions Inside the R&D Black Box, Research Policy, 29(9), pp. 1165-1184.

Dumont, M. and A. Tsakanikas (2001), Knowledge Spillovers through R&D Networking, CESIT Discussion Paper No. 2001/02.

Gandal N. and S. Scotchmer (1993), Coordinating Research Through Research Joint Ventures, Journal of Public Economics, 51, pp. 173-193.

Glaeser, E., H. Kallal, J. Scheinkman, and A. Shleifer (1992), Growth in Cities, Journal of Political Economy, 100, pp. 1126-1152.

Greene, W. (2000), Econometric Analysis, Prentice Hall International, Inc., New York.

Griliches, Z. (1990), Patent Statistics as Economic Indicators: a Survey, Journal of Economic Literature, 92, pp. 630-635.

Jaffe, A., M. Trajtenberg and R. Henderson (1993), Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations, Quarterly Journal of Economics, 108, pp. 577-598.

Jaffe, A.B. and M. Trajtenberg (1998), International Knowledge Flows: Evidence from Patent Citations, Working Paper 6507, National Bureau of Economic Research, Cambridge, Massachusetts.

Kingston, W. (2001), Innovation Needs Patents Reform, Research Policy, 30, pp. 403-423.

Krugman, P. (1991), Geography and Trade, Cambrige, MIT Press.

Lukach, R. and J. Plasmans (2000), R&D and Production Behavior of Asymmetric Duopoly Subject to Knowledge Spillovers, CESifo Research Paper 287, CES, Munich.

---

[7] For the probit model this marginal effect of a certain independent variable satisfies:

$$\left.\frac{\partial F(x_i^{'}\hat{\beta})}{\partial x_{ij}}\right|_{x_i=\overline{x}_i} = \overline{x}_i^{'}\hat{\beta} \cdot \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}\overline{x}_i^{'}\hat{\beta}^2), \quad i=1,2,..,n, \quad j=1,2,..,k,$$ while for the logit model with

$$\Phi(t) := \frac{\exp(t)}{1+\exp(t)} \text{ it satisfies } \left.\frac{\partial F(x_i^{'}\hat{\beta})}{\partial x_{ij}}\right|_{x_i=\overline{x}_i} = \frac{\hat{\beta}_j \exp(-\overline{x}^{'}_i\hat{\beta})}{(1+\exp(-\overline{x}^{'}_i\hat{\beta}))^2}.$$

Neieuwenhuijsen H. en Andre van Stel (2000), Kennis-spillovers en economische groei, Research Report 0007/N, EIM Onderzoek voor Bedrijf & Beleid.

Plasmans, J., W. Pauwels and T. Thewys (1999), Strategic R&D and Patent Behavior in some EU-countries, TEW Research Paper 99/042, UFSIA – University of Antwerp.

Plasmans, J. and R. Lukach (2001), Measuring Knowledge Spillovers Using Belgian EPO and USPTO Patent Data, CESifo Research Paper 430, CES, Munich.

Tijssen R.J.W. (2001), Global and Domestic Utilization of Industrial Relevant Science: Patent Citation Analysis of Science-Technology Interactions and Knowledge Flows, Research Policy, No. 30, pp. 35-54.

Verspagen, B. (1997), Measuring Intersectoral Technology Spillovers: Estimates from the European and US Patent Office Databases, Economic Systems Research, Vol. 9, No. 1, pp. 47- 65.

Verspagen, B. (1999), The Role of Larger Multinationals in the Dutch Technology Infrastructure. A Patent Citation Analysis., Working Paper, Eindhoven Center for Innovation Studies (ECIS), Maastricht Economic Research Institute on Innovation and Technology (MERIT).

Verspagen, B., T. van Moergastel and M. Slabbers (1994), MERIT concordance table: IPC - ISIC (rev.2), MERIT Research Memorandum 94-004.

CESifo Working Paper Series