International Knowledge Flows from and into a Small Open Economy: Patent Citation Analysis

Ruslan Lukach[1] and Joseph Plasmans

Abstract

This paper presents a study of backward and forward 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 citations data provide evidence of inter- or intra-firm and inter- or intra-industry knowledge spillovers which are very industry specific. Therefore we advocate for differentiated regulation policies directed at stimulating R&D cooperation in different industries.

Keywords: Knowledge Spillovers, R&D, Patent Citation, Limited Dependent Variable Regression

1. Introduction

Presented research aims at tracking down knowledge spillovers in small open economies by following some of their "trails". 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 internalise 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] Corresponding author. Address: Department of Economics, University of Antwerp, Prinsstraat 13, 2000 Antwerp, Belgium. Tel: +32-(0)3-220-4008. Fax: +32-(0)3-220-4585. E-mail: Ruslan.lukach@ua.ac.be.

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. Specialisation is characterised 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.
Our research relies on the assumption 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 recognises the potential value of the invention. Of course, this does not mean that non-patented knowledge is worthless, but we should advocate that the patented knowledge is the one most likely to be commercialised.
The innovating firms rely on their intangible assets as a source of their market value and competitive position. Therefore, the flow of knowledge among such firms is not only a process of pure information sharing, but also contributes to the increase/decrease of their market value, competitive and economic efficiency. In the 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.
As we advocate that patents encapsulate an important part of the commercially valuable knowledge, it is rational to consider the advantages of utilising patent data in the analysis of firms’ strategic R&D behaviour. The content of a patent consists of the information verified and submitted afterwards to a controlling body. Thus, the patent citation is 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 a patent citation determines a spillover of one protected (i.e. recognised as potentially valuable) knowledge pool to another.
In this paper we consider two different types of citations: backward (patent) citations and forward (patent) citations. Backward citations are citations listed in a particular patent and represent the technological knowledge acquired by the inventor. Forward citations occur when a particular patent gets cited representing in this way the diffusion of knowledge encapsulated in this patent.
The study of Duguet and MacGarvie (2003), based on the results of the Community Innovation Survey in France, shows that backward citations are correlated with firms’ R&D and innovation activities, while forward citations are correlated with firms’ answers on questions about their actions in disseminating their knowledge. Thus, even though backward and forward knowledge citations

contribute to knowledge spillovers in a similar manner, the underlying economic rationales of these two processes differ.

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 (2003), describes knowledge spillovers as the situation in which one economic agent benefits from R&D efforts of another economic agent without any tangible remuneration. These two definitions are given on the firm level and depending on the particular setup can describe both horizontal and vertical spillovers.

Gandal and Scotchmer (1993) 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 d'Aspremont and Jacquemin (1988), the study of Lukach and Plasmans (2000) investigated the optimal R&D and production strategies of firms that have different capabilities in research and production, which is very often the case in international markets.

Arrow's (1962) work points out that the competitive behaviour 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 behaviour 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, the profit maximising 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 'nature do its job'. 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.

The study of patent citations has its own limitations. Advantages and disadvantages of using patent citations 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, such as: 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.

Though we should admit the importance of other non-patent-citation ways of knowledge exchange, it is necessary to point out that only a patent citation is to a large extent finalised as a representation of such exchange. Patent information is better protected than other forms of knowledge, 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 .

An extensive study of Verspagen (1997) analyses patent citations 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, Verspagen (1999) investigated the impact of large Dutch companies on domestic knowledge diffusion in the Netherlands by studying patent-to-patent citations data, provided by the EPO. This study employed a network analysis to analyse the place of Dutch multinationals in the domestic technology infrastructure.

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 international 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.

In this paper we conduct a comparative analysis of the data and test the methodology for qualitative response variable analysis based on the recent research of Jaffe and Trajtenberg (1998) who constructed a probit-type binary choice model of knowledge flows using only backward patent citations from the USPTO. They have built a likelihood measure for the citation probability for any given patent pair. This allows a numerical evaluation of the ‘citation frequency’ 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 1995 and 2003 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).

Secondly, by constructing separate datasets for backward and forward citations we have an opportunity to compare the industrial patterns of knowledge utilisation (represented by backward citations) and knowledge dissemination (represented by forward citations). The backward citations data yield a final time-invariant picture of knowledge flows into Belgium via patent citations made by Belgian innovating firms between 1995 and 2003. The forward citations data contain all citations received by the patents granted between 1995 and 2003 during the same time period.

Here we should point out one important assumption we made in order to analyse the forward citations data. Patents receive citations as we speak. The forward citations dataset in this study is a snap-shot picture of a dynamic process as it was by the end of 2003. Therefore we do not attempt to derive any time-related implications in this paper. Yet it is rational to assume that the industrial structure of citations remains the same over time.

Here we assume that the probability that a particular patent will become cited by a patent from a particular industry remains the same over time. Based on this assumption we can then analyse the industrial structure of knowledge dissemination using the forward citations dataset at our disposal.

The main goal of this study is to uncover different patterns of citations in and to the recent Belgian patents, which belong to different industrial classes. The studied citations data provide evidence of inter- or intra-firm and inter- or intra-industry knowledge spillovers which are very industry specific. There are also differences in the patterns of backward and forward patent citations. Hence, the environmental factors of knowledge spillovers determining firms' incentives to cooperate in innovation are also different, which asks for adopting differentiated policies by the regulator.

Vonortas (1997) in his study on cooperation in research and development states that the current understanding of different environments in which the innovating

firms are functioning does not allow a constructive discussion on policy differentiation.
By this study we are setting a step forward in preparing the set of indicators that helps to assess such an environment and determine the factors and incentives for firms to cooperate in R&D.

2. Overview of the data

In this paper we analyse 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' knowledge spillovers in Belgium. The raw dataset is presented by the patent citations indicated in the patents granted to Belgian institutional applicants by the EPO or the USPTO. Among those, we select all citations, corresponding to the applicants, which are identifiable in the BelFirst database (data collected by the National Bank of Belgium (NBB) and provided by the Bureau van Dijk). 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 behaviour 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.
We run our model in two main data collections, the sets of backward citations made in Belgian firms' patents during 1995-2003 and forward patent citations, which cite these Belgian patents. The backward citations data present us with the picture of knowledge utilisation flows and the forward citations describe the picture of knowledge dissemination.
We gained an 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 a patent issued by the EPO cites a patent issued by the USPTO and vice versa.
In the 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).

3. Preliminary data analysis

The source dataset is a pooled sample of all patents granted by the EPO and the USPTO to Belgian firms during the period between 1995 and 2003. It contains 6663 patents (2709 from the EPO and 3594 from the USPTO), which produce 29797 initial backward patent-to-patent citations (6413 originating from the EPO patents, and 23384 from the USPTO) and 5424 forward citations (360 originating from the EPO patents, and 5064 from the USPTO). Unlike backward citations, the number of forward citations changes every day as new patents become granted by both USPTO and EPO. Our forward citations dataset represents the state of things by the end of the year 2003.

Table 14.1. Geographic distribution of backward patent citations of Belgian firms' 1995-2003 patents granted by the EPO and USPTO.

Country	USPTO	EPO	Total
United States	44.58	37.36	43.02
Japan	21.53	21.41	21.86
Belgium	13.93	17.70	14.74
Germany	6.61	7.93	7.15
France	2.85	3.92	3.08
Great Britain	2.45	2.65	2.49
Switzerland	1.30	1.66	1.37
Italy	1.17	1.69	1.28
Netherlands	0.98	1.80	1.16
Canada	0.65	0.70	0.66
Other	3.95	3.18	3.19

First, we consider the basic geographic distribution of backward citations made by Belgian applicants. Table 14.1 lists ten countries, from which the most cited

patents originate (96.8% of the total sample). 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. Rationally, we would have expected that Belgian patents will be the mostly cited, 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 and Japan are the most frequently cited by Belgian companies, which clearly indicates the importance of these two countries. The other positions are occupied by the countries of the European Union (EU), Switzerland and Canada.
In Table 14.1 we also observe that the citations of American patents account for quite comparable shares in the USPTO and the EPO samples. Therefore, 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 14.2. Geographic distribution of forward patent citations by the end of

Country	USPTO	EPO	Total
United States	50.25	11.50	47.96
Belgium	18.70	49.77	20.54
Japan	15.41	10.33	15.11
Germany	3.79	14.08	4.40
Sweden	1.96	0.47	1.87
France	1.93	3.29	2.01
Canada	1.52	0.23	1.44
Switzerland	1.14	1.41	1.15
Netherlands	0.90	2.58	1.00
Great Britain	0.86	0.94	0.86
Other	3.54	5.40	3.66

2003 of Belgian firms' 1995-2003 patents granted by the EPO and USPTO.

In Table 14.2 we present ten countries from which the most forward citations to Belgian patents are originating. As in the previous case, more than 96% of the forward citations come from ten countries. The list of the top-ten countries producing forward citations is almost the same as the list of countries receiving the most backward citations with one exception. Italy occupies a place in the list of the top-ten countries by backward citations, but in the list of the top-ten by forward citations its place is taken by Sweden.

We see that both backward and forward citations datasets give use almost the same composition of countries in the list, but different relative shares of citations. In the backward citations dataset the share of Great Britain is 2.49%, but in the forward citations it is only 0.86%.

The first place in both lists is occupied by the United States making it the most important source and destination of knowledge flows initiated by the Belgian patents in our dataset. Belgium is in third position according to the backward citations data. Observing the forward citations, we see that Belgium holds the second position for citations in patents contained in the USPTO and the total datasets, but holds the first place with the most citations originating in the patents issued by the EPO. Thus, we observe that patents granted to Belgian firms in the years 1995-2003 are a quite important source of knowledge spillovers for other Belgian patents.

In general, we conclude that the ‘geographic proximity’ assumption for knowledge spillovers is not strongly supported by the collected information: domestic patents are not the most frequently cited and are not those most frequently citing; 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.

Analysing percentages of citations made in the patents granted by the EPO and the USPTO in years 1995-2003 to individual Belgian companies, we observe that that the top 20 companies with the highest percentages account for more than four-fifths of the patent citations. This indicates quite a substantial concentration of patenting activities in a small number of Belgian private enterprises.

These results are closely related to the findings already presented by Plasmans et al. (1999), which are based on the study of the patenting behaviour 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 account 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 55% of all patents issued between 1995 and 2003 (inclusive) by the USPTO and the EPO.

When it comes to the ‘size’ characteristics of patenting companies, we observe some enterprises which are quite big and known (Agfa-Gevaert, Solvay, Janssen Pharmaceutica, Glaverbel, Bekaert), but also some much smaller firms which are also active in patenting (Esselte, Xeikon, Sofitech, Owens - Corning). Thus, the large size of a company does not necessarily indicate that it will be more active in patenting than its smaller companions.

Further data analysis yields results illustrating the small open economy feature of Belgium. We constructed two variables for indicating the events where: i) the patent citation occurred between the patents owned by the same assignee; ii) the patent citation occurred between the patents owned by the assignees in the same country. In the backward citations dataset the “same firm” citations account for 37% of all citations and the “same country” citations account for 34%, respectively. In the forward citations dataset these percentages are 21% and 21%, correspondingly.

The correlation coefficient between the same firm and the same country indicators is 0.96 in the backward citations and 0.99 in the forward citations datasets.

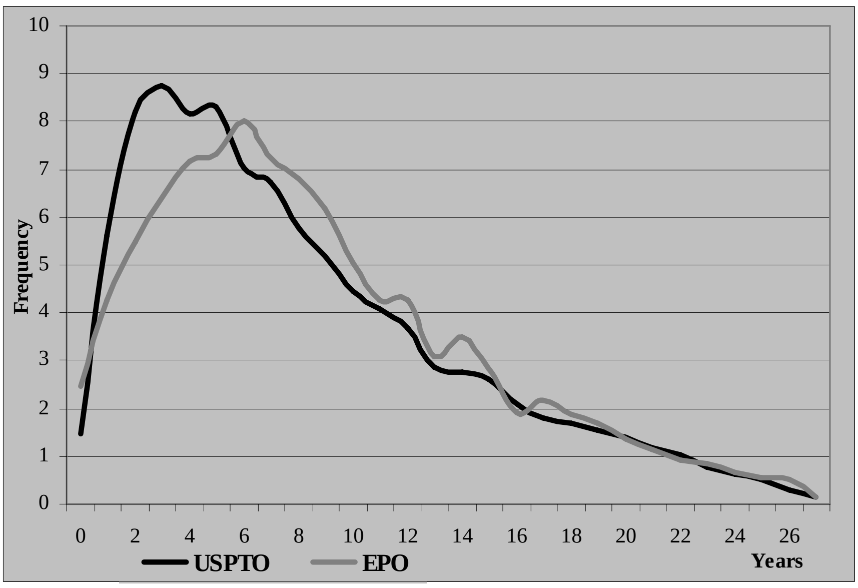
This allows us to argue that in Belgium the most inter-firm citations are also the international citations and vice versa.

We can mention two main factors contributing to this situation. First, the most R&D in Belgium in each particular industrial sector is conducted by a very small number of firms, which makes it more likely that if some other firm’s knowledge is cited, that firm will be from the other country. The second factor, in our opinion, is centralisation of the intellectual property protection in large multinational enterprises (such as AGFA-Gevaert, Monstanto, Solvay, and Janssen Pharmaceutica) in one legal entity in one country.

We conclude that the patent citations data are appropriate for analysing the international knowledge flows between different enterprises, but are not able to catch the international knowledge exchange occurring inside one large multinational corporation.

What concerns the time-related features of the data, based on the time lag between citing and cited patents we can derive the implications about the time structure of knowledge spillovers. Figure 14.1 illustrates the distribution of cited patents across different years. The basic shape of the distribution is very much like 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 the older ones.

Figure 14.1. Backward Citations Time Lag Structure based on the Belgian



patents granted by two different patent offices during 1995-2003.

We also observe that the time structure of the citation lag is very similar in both the USPTO and the EPO samples. This provides an additional argument for compatibility of the data in these two samples and that pooling of these two samples is feasible. As it was already mentioned before, the similar time lag analysis of forward citations data is not feasible due to their dynamic nature, variability of which is especially strong if we analyse relatively recent data pools.

Figure 14.2. Relative frequencies of backward citations between industries.

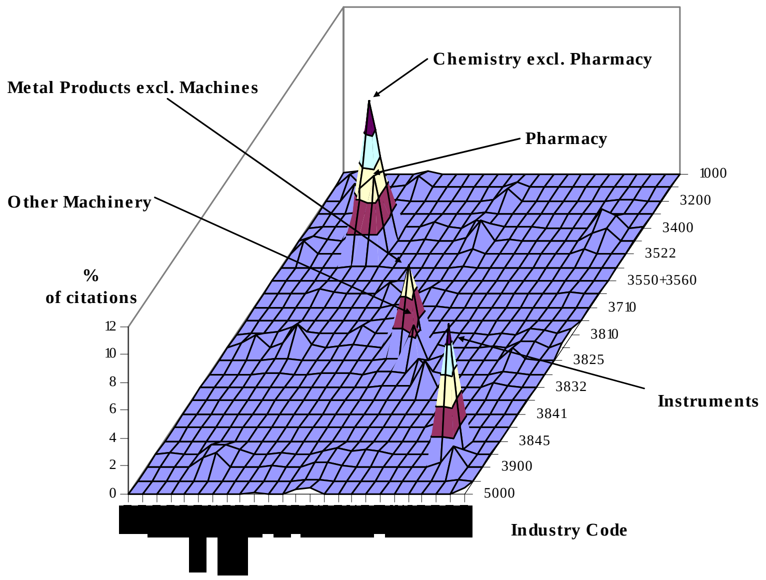
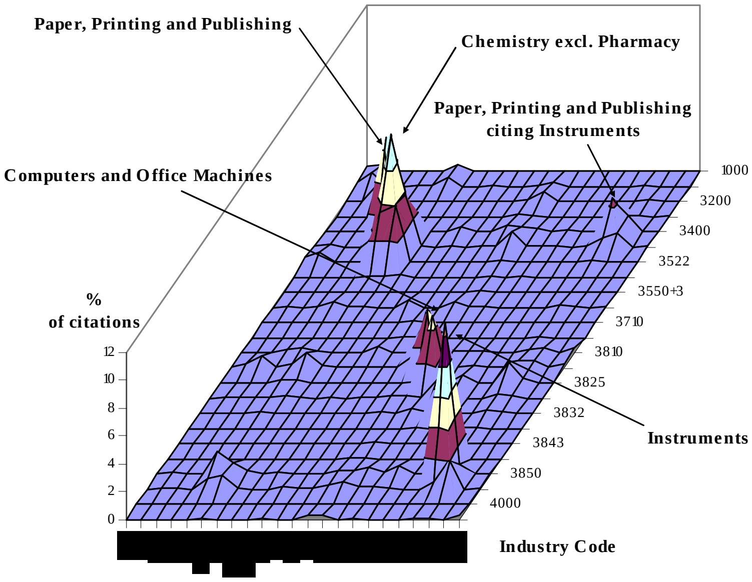


Figure 14.3. Relative frequencies of forward citations between industries.



Let us consider the industrial structure of patent citations indicated in a pooled sample (the USPTO and EPO samples together). Figures 14.2 and 14.3 present 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 14.3 lists all the industries indicated in the ISIC, accompanied by the corresponding percentages of citations calculated in the pooled sample.

Table 14.3. Citation Percentages in Different Industries (as a fraction of all citations 1995-2003).

ISIC code	**Industry**	**% of backward citations**	**% of forward citations**
3510+3520	Chemistry, except pharmacy	19.52	15.98
3850	Instruments	13.34	16.28
3820	Other machinery	13.02	9.75
3522	Pharmacy	12.14	8.97
3400	Paper, printing and publishing	7.81	15.98
3810	Metal products, ex. machines	6.81	5.58
3900	Other industrial products	4.99	5.55
3825	Computers & office machines	4.59	8.00
3832	Electronics	2.88	3.82
3100	Food, beverages, tobacco	2.62	2.36
3600	Stone, clay and glass products	2.58	1.83
3200	Textiles, clothes, etc.	2.51	2.11
3830	Electric mach., ex. electronics	2.21	2.81
5000	Building and construction	1.37	0.92
1000	Agriculture	0.6	1.25
3843	Motor vehicles	0.57	0.42
3710	Ferrous basic metals	0.55	0.24
3720	Non ferrous basic metals	0.48	0.15
3300	Wood and furniture	0.45	0.45
3550+3560	Rubber and plastic products	0.31	0.23
3840	Other transport	0.25	0.07
3530+3540	Oil refining	0.19	0.22
4000	Utilities	0.17	0.01
3841	Shipbuilding	0.01	0.00
3845	Airspace	0.00	0.02

There are eight major industries which account for the largest part (82%) 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).

Figures 14.2 and 14.3 are graphical representations 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, 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’ (10.6% of all backward and 8.1% of all forward citations), ‘Instruments’ (9.38% and 10.91% respectively), ‘Pharmacy’ (5.61% and 4.03%), and ‘Other Machinery’ (5.97% and 4.24%) industries. There are 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% of backward and 2.36% of forward citations). Among the symmetric cross-industry knowledge flows, the strongest ones occur between ‘Chemistry excluding Pharmacy’ and ‘Pharmacy’ industries (4.97% backward and 3.7% forward citations one way and 4.67% of backward and 3.61% of forward citations in the opposite direction).

4. Model and estimations

4.1. Citation Pairs Modelling

Now we intend to employ an econometric methodology to try to get a deeper insight into the knowledge spillovers pattern, ‘encoded’ into patent citations data. Previous researchers’ experience (Jaffe and Trajtenberg (1998)) shows that patent citations data are best to be analysed 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 on one particular kind of event, which takes place as a 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 (occupying the first eight places in Table 14.3) to be analysed 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 modelled 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 log-log and complementary log-log (Long 1997). After comparing the forecasting performance of these specifications (see Appendix) we have chosen the complementary log-log distribution as the basis for our model. The complementary log-log distribution is asymmetric. The distribution of our dependent variable is also likely to be 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).

Distributions of the independent variables are asymmetric too. As we return to the graph (Figure 14.1) 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.
There are several notes to be made about 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. We immediately substitute the estimated coefficients by the corresponding slopes or marginal effects (see Appendix), which are presented in Table 14.4 for backward citations and Table 14.5 for forward citations.

Table 14.4. Estimated complementary log-log marginal effects in the backward citations dataset.

	SameFirm	Same Industry	TimeLag	Year
3510+3520 'Chemistry, excluding Pharmacy'	0.0730***	-0.1195***	0.0007**	-0.0039***
3850 'Instruments'	-0.0017	0.1789***	-0.0010***	-0.0021***
3522 'Pharmacy'	0.0688***	-0.0413***	-0.0002	0.0018***
3820 'Other Machinery'	-0.0437***	-0.0169***	0.0037***	0.0018***
3400 'Paper, Printing and Publishing'	0.0272***	-0.0286***	-0.0033***	0.0025***
3810 'Metal Products, excluding Machines'	-0.0651***	-0.0409***	0.0023***	-0.0026***
3825 'Computers and Office Machines'	-0.0196***	0.0386***	-0.0024***	0.0010***
3900 'Other Industrial Products'	-0.0063***	-0.0410***	-0.0012***	-0.0036***

Table 14.5. Estimated complementary log-log marginal effects in the forward citations dataset.

	SameFirm	Same Industr	TimeLag	Year
3510+3520 'Chemistry, excluding Pharmacy'	0.00791	-0.1196***	0.0062***	-0.0064***
3850 'Instruments'	0.0334***	0.1846***	0.0077***	-0.0142***
3522 'Pharmacy'	0.0139**	-0.0611***	0.0043***	-0.0029*
3820 'Other Machinery'	-0.0097	-0.0534***	-0.0057***	0.0071***
3400 'Paper, Printing and Publishing'	0.0870***	0.0189***	-0.0155***	0.0215***
3810 'Metal Products, excluding Machines'	-0.0450***	-0.0608***	-0.0030***	0.0019*
3825 'Computers and Office Machines'	-0.0152**	0.0736***	0.0040***	-0.0046***
3900 'Other Industrial Products'	-0.0149***	-0.0327***	0.0047***	-0.0037***

As we can see the regressions provide the majority of slopes with a high degree of statistical significance. The slopes of variables SameFirm and SameIndustry are discussed in detail in the next section, therefore here we will concentrate our attention on the time-related independent variables.
The estimated slopes in the backward citations dataset show that older patents are more likely to be cited in the ‘Chemistry, excl. Pharmacy’, ‘Other Machinery’ and the ‘Metal Products, excl. Machines’ industry. In the

‘Instruments’, ‘Paper, Printing and Publishing’, ‘Computers and Office Machines’, and ‘Other industrial Products’ it is more likely that a more recent patent receives a citation.

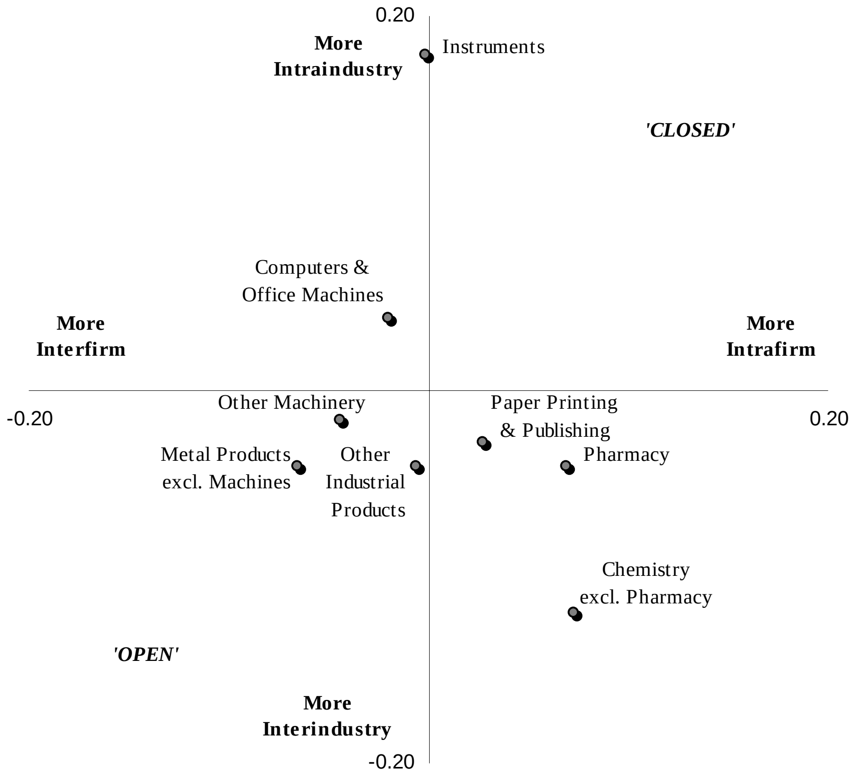
Half of the studied industries show a tendency towards making more citations in the patents granted in later years. Indeed, we expect that as more patent information resources become available to inventors and applicants, the number of citations made in a new patent application will increase. Our estimation does not provide full support for this assumption. In the ‘Chemistry, excl. Pharmacy’, ‘Instruments’, ‘Metal Products excl. Machines’, and the ‘Other Industrial Products’ industries we observe negative slope coefficients indicating that newer patents are not likely to make more citations than older ones.

Here we should point out that so far more applicable time-related implications can be made only from the backward citations pool, because it describes complete time patterns inside the fixed time interval (years 1995-2003). The time-related variables (citation lag and the patent’s publication year) are present in the forward citations estimations in Table 14.5 to make them compatible with those based on backward citations. This allows us to compare the “same firm” and the “same industry” coefficients. The time-related coefficients from the forward citations estimations seem not to be applicable for comparison, because our datasets are very recent and the time patterns of the recent forward citations dataset can change in the future. Looking at Figure 14.1 we see it takes approximately 12 years for 80% of citations to occur.

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

To obtain a better view on general results of modelling 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. Figures 14.4 and 14.5 are constructed in two dimensions, 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 1993, p. 879).

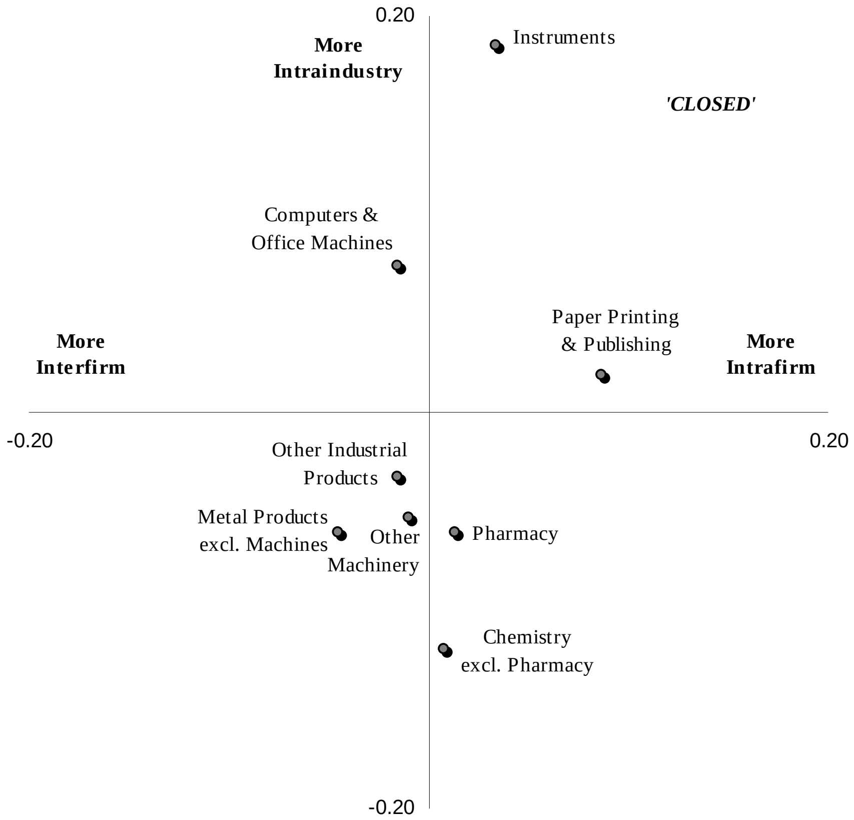
Figure 14.4. Positioning of industries with relation to Intra-firm and Intra-industry knowledge spillovers using backward citations.



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 favour 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. And the top-left quadrant combines intra-industry and inter-firm spillovers correspondingly.

Following the discussion in Section 3 on specifics of Belgian patent citations data, we can also interpret the industry's openness towards inter-firm spillovers as openness towards more international knowledge flows.

Figure 14.5. Positioning of industries with relation to Intra-firm and Intra-industry knowledge spillovers using forward citations.



In Figures 14.4 and 14.5 we see that different industries occupy positions in different quadrants. According to both backward and forward citations data a group of ‘open’ industries consists of ‘Metal Products, excl. Machines’, ‘Other Machinery’, and ‘Other Industrial Products’. These industries are more likely to use inter-firm and therefore international knowledge flows.

The backward citations model shows that there are no ‘closed’ industries among the analysed sectors. Yet if we observe results of the forward citations model, we see that the ‘Instruments’ and ‘Paper, Printing and Publishing’ industries do exhibit the closed industry’s patent citation patterns. This implies that the knowledge utilisation processes in these industries are different from their knowledge dissemination processes. For example, the Belgian patents in the 'Paper, Printing and Publishing' are more likely to use knowledge (cite backwards) from other industries, but their knowledge in its turn becomes used (cited) by the patents in the same industry. The 'Instruments' industry is in an interesting position, where there is almost no difference between the intra- or inter-firm backward citation, but the forward citations in it are more likely to be intra-firm.

Another way to interpret these results is to say that in the 'Instruments' and 'Paper, Printing and Publishing' sectors utilisation of knowledge is more internationalised than knowledge dissemination.

The 'Computers and Office Machines' industry is open for inter-firm or international knowledge spillovers, and is less inclined towards using the knowledge from other industries according to the results from both backward and forward citations datasets. The 'Chemistry, excluding Pharmacy' and the 'Pharmacy' industry exhibit greater openness for inter-industry knowledge spillovers, but are less inclined to cite the knowledge of other firms (and from other countries), which is also supported by the data from both datasets.

Augmenting these results by the data presented in Chapter 10 (Table 10.2), we see that among the industries, which are less likely to use and produce inter-firm and, therefore, international knowledge flows, there are two industries ('Chemistry, excluding Pharmacy' and 'Pharmacy') with a very high degree of foreign ownership, and one (Paper, Printing and Publishing) with a medium degree of foreign participation.

Yet the 'Computers and Office Machines' industry, which also has a very high degree of foreign ownership, tends to cite more inter-firm knowledge, thus being more open towards international spillovers. From this we conclude that the patterns of inter(intra)-firm and international(domestic) knowledge flows cannot be determined by only the degree of foreign ownership in innovating enterprises, but must take the special feature of a particular industry into account.

5. Conclusions and policy discussion

The objective of this study was to investigate the patenting and patent citation behaviour of private firms in a small open economy. We based the study on patent behaviour of Belgian private firms using the 1995-2003 patent citations data from the EPO and the USPTO. The attention of this study was concentrated on the patent citation behaviour of Belgian firms using binary response variable models. The results of the data analysis and estimations can be summarised in the following statements:

A preliminary analysis has shown that the majority of the patenting is conducted by a (very) small number of firms different in size. The majority of patent citations occur in a limited number of main industries.
The geographical structure of citations derived from the backward and forward citations datasets are very similar in terms of the list countries, but differs in their relative weights. The most important countries in both knowledge utilisation and knowledge dissemination are United States, Belgium, and Japan, followed by the list of other European countries and Canada. Therefore, we can neither fully support nor fully reject the hypothesis of geographical localisation of knowledge flows.
The data on Belgian backward and forward patent citations show a very strong correlation between occurrences of the inter-firm and the international patent citations.
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 strength of knowledge spillovers in a certain industry, and can be applied for various competitive behaviour models.
The analysed industrial sectors exhibit different patterns of patent citation and the knowledge spillovers associated with the above. These patterns are very industry-specific and are not correlated with the degree of foreign participation and/or ownership.
Analysing the relative positioning of different industries depending on their attitude towards inter-firm knowledge spillovers allows us to make certain implications about 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.
We consider knowledge spillovers as a source of the positive externalities determining the firms' incentives to cooperate in research and development. From the social planner's point of view, it is desirable to promote R&D cooperation, since it increases the efficiency of R&D, output and social welfare (d'Aspremont and Jacquemin 1988). Under conditions of stronger knowledge spillovers, innovative firms have more incentives to engage in R&D cooperation. For a policymaker 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.

Once the special feature of the industry is determined, such as the likelihood of inter- or intra-firm spillovers (which also describes the degree of internationalisation of knowledge exchange) and the likelihood of inter-industry knowledge exchange, we obtain an understanding of the general knowledge spillovers intensity.

The general guidelines for the regulator, derived from our study, can be summarised 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. Steurs (1995) points out that inter-industry cooperation is more favourable for increasing the R&D investment and welfare than intra-industry cooperation. Hence, stimulating the inter-industry R&D cooperation among the firms gives a better positive effect than stimulating the intra-industry alliances. Such regulating measures will bring their best results if applied in the industrial sectors located in the upper quadrants of our map, because knowledge spillovers and the corresponding natural incentives to cooperate in those industries are weaker.

The main tools of such regulating policy can be measures directed at facilitating creation of the R&D consortia through subsidies or tax breaks (direct stimuli) or measures for facilitating knowledge spillovers (indirect stimuli) to create an environment in which the firms are more inclined to engage in cooperative R&D efforts.

In the ideal scenario it is desirable to have a balanced picture of knowledge utilisation and knowledge dissemination. The regulatory measures, which stimulate R&D cooperation, should take into account the type of knowledge flows prevailing in a particular industry. In industries with weak outward knowledge flows it is preferable to favour the joint R&D efforts, which are directed at better dissemination of knowledge produced by Belgian firms. In the opposite situation (weak inward knowledge spillovers) attention should be paid to stimulating better knowledge utilisation.

The understanding of the knowledge spillovers’ environment allows the regulator to develop balanced differentiated policies in each industry where the natural incentives for R&D cooperation created by knowledge spillovers are complemented by the policy measures in order to achieve the best results.

Concluding this discussion, we bring up an argument that public authorities should use a differentiated approach to the regulation of R&D activities by firms in different industries. There are market-driven incentives which induce firms' cooperation; thus it is possible for a regulator to use these incentives in combination with particular regulatory measures to achieve desired effects whether it is the higher R&D investment or better knowledge diffusion in the economy. The major outcome of such a successful policy will eventually surface in faster economic growth.
Acknowledgements. The authors thank Bart Verspagen of ECIS, Technical University of Eindhoven, for assistance in data acquisition, Michel Cincera of Université Libre de Bruxelles, Bart Clarysse of University of Ghent, Daniela Cristina and Wim De Preter of University of Antwerp, and the editors for their valuable comments.

6. Appendix

6.1. Complementary Log-Log Model for Patent Citations

The pooled dataset contains a list of citation pairs, which were made in the granted patents. 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 is made in the citing patent coming from industry A'. The complementary log-log model is specified as:

$$\mathrm{P}(y_i = 1) = F(\beta' x_i) = 1 - \exp(-\exp(\beta' x_i)),\ i = 1, 2, .., n,$$

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 is made in the patent belonging to a particular industry. It is also known that the estimated coefficients of this type of model do not give the value of the marginal effect of the independent variable. 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:

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

where $f(\overline{x}_i^{'}\hat{\beta}) = \exp(\overline{x}_i^{'}\hat{\beta} - \exp(\overline{x}_i^{'}\hat{\beta}))$ is the complementary log-log density function calculated in the mean of the estimated structural part of the model.

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 (1993, p. 878) 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.

6.2. Forecasting Performance of the Binary Choice Models

We use two forecast error statistics to compare the forecasting performance of different model specifications. The first statistic is the Root Mean Squared Error:

$$RMSE = \sqrt{\frac{1}{N+1}\sum_{i=1}^{N}\left(\hat{p}_i - p_i\right)^2}\ .$$

The second statistic is the Theil Inequality Coefficient:

$$TIC = \frac{\sqrt{\frac{1}{N+1}\sum_{i=1}^{N}\left(\hat{p}_i - p_i\right)^2}}{\sqrt{\frac{1}{N+1}\sum_{i=1}^{N} p_i^{\,2}} + \sqrt{\frac{1}{N+1}\sum_{i=1}^{N} \hat{p}_i^{\,2}}}$$

The smaller the value of the statistic, the better the forecasting ability of the model according to this criterion.

We compare the performance of four model specifications:

Probit: $\mathrm{P}(y_i = 1 \mid x_i) = \int_{-\infty}^{\beta' x_i} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt$, $i = 1, 2, .., n$;

Logit: $\mathrm{P}(y_i = 1 \mid x_i) = \frac{\exp(\beta' x_i)}{1 + \exp(\beta' x_i)}$, $i = 1, 2, .., n$;

Complementary log-log (Clog-log): $\mathrm{P}(y_i = 1 \mid x_i) = 1 - \exp(-\exp(\beta' x_i))$, $i = 1, 2, .., n$;

Log-log: $\mathrm{P}(y_i = 1 \mid x_i) = \exp(-\exp(-\beta' x_i))$, $i = 1, 2, .., n$.

The calculated values of these forecast error statistics are presented in Tables 14.6- 14.9.

Table 14.6. Root Mean Squared Errors for the backward citations dataset.

Industry	Probit	Logit	Clog-log	Log-log
Chemistry ex. Pharmacy	0.391712	0.391585	0.391477	0.391896
Instruments	0.319760	0.319279	0.319002	0.320366
Pharmacy	0.324893	0.324898	0.324880	0.324884
Other Machinery	0.335031	0.335042	0.335045	0.335018
Paper, Prinitng and Publishing	0.266946	0.266892	0.266872	0.266990
Metal Products ex. Machines	0.249868	0.249876	0.249878	0.249867
Computers and Office Machines	0.206924	0.206855	0.206841	0.206991
Other Industrial Products	0.216499	0.216449	0.216438	0.216543

Table 14.7. Theil Inequality Coefficients the backward citations dataset.

Industry	Probit	Logit	Clog-log	Log-log
Chemistry ex. Pharmacy	0.607218	0.606741	0.606108	0.608100
Instruments	0.594484	0.592118	0.590145	0.598102
Pharmacy	0.683292	0.683635	0.682680	0.682680
Other Machinery	0.676843	0.676920	0.677011	0.676725
Paper, Prinitng and Publishing	0.737411	0.736998	0.736809	0.737741
Metal Products ex. Machines	0.742593	0.742269	0.742299	0.743112
Computers and Office Machines	0.772280	0.770463	0.770051	0.774064
Other Industrial Products	0.780121	0.779014	0.778754	0.781119

Table 14.8. Root Mean Squared Errors for the forward citations dataset.

Industry	Probit	Logit	Clog-log	Log-log
Chemistry ex. Pharmacy	0.362590	0.362588	0.362450	0.362592
Instruments	0.353737	0.353359	0.350764	0.354273
Pharmacy	0.284839	0.284849	0.284825	0.284836
Other Machinery	0.295676	0.295664	0.295399	0.295685
Paper, Prinitng and Publishing	0.331743	0.331735	0.328185	0.331758
Metal Products ex. Machines	0.226905	0.226917	0.226803	0.226909
Computers and Office Machines	0.268105	0.268037	0.267877	0.268174
Other Industrial Products	0.228148	0.228142	0.228087	0.228155

Table 14.9. Theil Inequality Coefficients for the forward citations dataset.

Industry	Probit	Logit	Clog-log	Log-log
Chemistry ex. Pharmacy	0.638095	0.638100	0.637510	0.638055
Instruments	0.595021	0.593401	0.584650	0.597750
Pharmacy	0.722135	0.723209	0.723333	0.720687
Other Machinery	0.716078	0.716019	0.714120	0.716125
Paper, Prinitng and Publishing	0.663133	0.662900	0.647445	0.663525
Metal Products ex. Machines	0.753921	0.753044	0.752253	0.755185
Computers and Office Machines	0.721586	0.720697	0.719109	0.722547
Other Industrial Products	0.776079	0.775952	0.774926	0.776200

As we can see in the overwhelming majority of industries the complementary log-log specification of the model has better forecasting performance than other settings.

References

Arrow, K. The Economic Implications of Learning by Doing. Review of Economic Studies, 29, 1962, 155-173.

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

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

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

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

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

Duguet, E. & MacGarvie, M. (2003). How Well Do Patent Citations Measure Flows of Technology? Evidence from French Innovation Surveys, Working Paper, http://ssrn.com/abstract=452000.

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

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

Greene, W. (1993). Econometric Analysis. Prentice Hall International, Inc., New York.

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

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

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

Johnson, D. (2002). The OECD Technology Concordance (OTC): Patents by Industry of Manufacture and Sector of Use. STI Working Papers 2002/5, OECD.

Long, J. (1997). Regression Models for Categorical and Limited Dependent Variables. SAGE Publications, London.

Lukach, R. & Plasmans, J. (2000). R&D and Production Behavior of Asymmetric Duopoly Subject to Knowledge Spillovers. CESifo Working Paper No. 287, CESifo, Munich.

Nieuwenhuijsen, H.R. and A.J. van Stel (2003). Knowledge Spillovers and Economic Growth. EIM Report, Rotterdam.

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

Schmoch, U., Laville, F., Patel, P. & Frietsch, R. (2003). Linking Technology Areas to Industrial Sectors. Final Report to the European Commission, DG Research.

Steurs, G. (1995). Inter-industry R&D Spillovers: What Difference Do They Make? International Journal of Industrial Organization, 13, 249-276.

Verspagen, B. (1997). Measuring Intersectoral Technology Spillovers: Estimates from the European and US Patent Office Databases. Economic Systems Research, 9, 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., van Moergastel, T. & Slabbers M. (1994). MERIT concordance table: IPC - ISIC (rev.2). MERIT Research Memorandum 94-004, MERIT.

Vonortas, N. (1997). Cooperation in Research and Development. Kluwer Academic Publishers, London.