A joint Initiative of Ludwig-Maximilians-Universität and Ifo Institute for Economic Research

CESifo Working Papers

MEASURING KNOWLEDGE SPILLOVERS USING BELGIAN EPO AND USPTO PATENT DATA*

Joseph Plasmans
Ruslan Lukatch

CESifo Working Paper No. 430

March 2001

CESifo
Center for Economic Studies & Ifo Institute for Economic Research
Poschingerstr. 5, 81679 Munich, Germany
Tel.: +49 (89) 9224-1410
Fax: +49 (89) 9224-1409
e-mail: office@CESifo.de

* The authors thank Bart Verspagen of ECIS for assistance in data acquisition, Eddy Laveren and Wim Meuseen of UFSIA-RUCA faculty, University of Antwerp for useful suggestions.

CESifo *Working Paper No. 430*
March 2001

MEASURING KNOWLEDGE SPILLOVERS USING BELGIAN EPO AND USPTO PATENT DATA

Abstract

This paper investigates two major issues of the patenting behavior of Belgian firms. Firstly, it studies the probabilistic distribution of the patent citations among several major sectors. Secondly, the firm-oriented data is studied to investigate the relationships between the Belgian firms' size and their patent citation behavior. The modeling results conclude that there is evidence that the smaller firms tend to be more active in patent citation than larger ones. Analyzing the implications from the probabilistic models of citations the paper concludes, that there are different patterns of citation behavior in different sectors. Some sectors exhibit more openness toward inter-firm or inter-industry spillovers, while others do not. Moreover, different industrial sectors exhibit different relationships between the probability of a citation to occur in this sector and the relative time lag between the citing and cited patents.

JEL Classification: O30, O34, C50

Joseph Plasmans
University of Antwerp
Department of Economics
Prinsstraat 13
2000 Antwerp
Belgium

Ruslan Lukatch
University of Antwerp
Faculty of Applied Economics
Prinsstraat 13
2000 Antwerp
Belgium
e-mail: Ruslan.Lukatch@ufsia.ac.be

1. INTRODUCTION

Knowledge spillover is a phenomenon, which is quite easy to imagine, but very difficult to see. According to the definition given by De Bondt (1996), the concept of a 'knowledge spillover' is specified like 'an involuntary leakage or voluntary exchange of useful technological information'. According to Bernstein and Nadiri (1988), knowledge spillovers can be classified as vertical or horizontal. Horizontal spillovers occur between competitors, and vertical spillovers flow between firms in different industries.

Knowledge spillovers play an important role among the factors determining the R&D behavior of firms. The model constructed by d'Aspremont and Jacquemin (A&J (1988)) shows that the spillover effect influences the strategic behavior of firms engaged in R&D in a competitive environment. Spillovers mainly determine the willingness of firms to cooperate in research. Generally, when spillovers are high enough, firms have strong incentives to cooperate with each other. In addition, under strong knowledge spillovers the cooperating firms tend to invest more in R&D than the competing ones. In several other extensions to the A&J (1988) model (for example: Suzumura (1992), De Bondt and Henriques (1995), Leahy and Neary (1997), Petit and Tolwinski (1999), Lukatch and Plasmans (2000)) the effect of spillovers on competitive (cooperative) strategies of firms remains substantial. Thus constructing an efficient measure for such spillovers can greatly contribute to the understanding of the firm's strategic R&D behavior and of the mechanisms of possible inter-firm cooperation.

The empirical study of knowledge spillovers has another important area of application. Correct assessment of 'spillover environment' contributes to the development of the most appropriate regulating policy in the area of innovation activity. For example, in industries with weak knowledge spillovers, it is better to support the competitive mode of the firms' R&D activity. On the other hand, in conditions with strong knowledge spillovers it is desirable to stimulate joint research ventures and other kinds of R&D cooperation. R&D cooperation subsidies, tax subsidies and other policy measures can be applied to achieve the desired effect.

Plasmans *et al.* (1999) advocate that the entrepreneurial innovative behavior can reasonajbly be explained by the patent 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. They applied it to the panel data for core EU countries and different industries.

Verspagen (1997) conducted an extensive study of patent citation data in application 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 (European Patent Office). 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. This study gave another empirical evidence of usefulness of citation data.

The goal of our paper is to construct an efficient knowledge spillover measure by using patent-to-patent citation information and apply it to the case of the Belgian economy using econometric methodology. We base our approach on the recent research of Jaffe and Trajtenberg (1998) who constructed a 'probit-type' model of knowledge flows using patent citations. They have built a likelihood measure for the citation probability for any given patent pair. This allows numerical evaluation of the 'citation frequency'[1] between different sectors of the economy as well as between different geographical areas. The study of Jaffe & Trajtenberg was based only on data provided by the USPTO (United States Patent and Trademark Office) and concentrated on the industry 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, thus widening our data-scope. We also augment the study by a comparative analysis of these two sources.

Our fundamental data units are provided by all patent applications submitted by Belgian firms to EPO in 1997 and all granted patent applications submitted by Belgian firms to USPTO in 1997. We study patent-to-patent citations obtained from these documents together with an additional information from other sources. We determine Belgian firms, which have applied for patents in 1997, 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. We expect that patterns of citing and cited EPO and USPTO patents have certain differences, providing additional insights about the geographical preferences of patenting agents. In general, patent citation is applicable to a wide variety of comparative analysis areas.

Thus, having at our disposal a numerical measure for knowledge spillovers, we can use it in an assessment of the innovation and knowledge diffusion environments. We are able to establish a direct link between the theoretical parameter β, which describes knowledge spillovers in the A&J (1988) model[2], and obtain efficient empirical measures for this parameter. Finally, we use the asymmetric duopolistic model of R&D and production behavior (Lukatch and Plasmans, (2000)) to derive political implications for the innovation activity in Belgium.

2. PATENT CITATIONS AND KNOWLEDGE SPILLOVERS: 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 'snap-shot' picture of the 'patent-driven' knowledge spillovers in Belgium. In particular, we study the set of the patent applications submitted by Belgian firms to the EPO in 1997 and the same-year applications to the USPTO, which were later granted in 1998 and 1999.

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 data currently available to us are limited to the year 1998.

The raw dataset is presented by the patent citations indicated in the patent applications submitted by Belgian corporate applicants to the EPO; and submitted (and afterwards granted during 18 months) to USPTO in 1997. 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 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 reference 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-

---

[1] 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 $\tau \leq t$.

[2] The spillover parameter β satisfying $0 < \beta < 1$ is a discount parameter, which implies that some benefits of each agjent's R&D flow without payment to other agents. The value $\beta = 0$ indicates the absence of spillovers and $\beta = 1$ indicates that knowledge freely flows between agents.

type models, technological proximity matrices, and network analysis. They both conclude the high informational value of such arrangement.

The study of patent citations has its own limitations. Advantages and disadvantages of using the patent citation data are extensively discussed by Griliches (1990) and Jaffe *et al.* (1993). Patent citations are linked to the patenting procedure itself. Thus, they capture only the knowledge flows, which only occur between patented 'pieces' of innovation. Other means of knowledge transfer are not captured by the patent citations. There is also a so-called 'noise' component in the citation list. Applicants are legally obliged to indicate any previously patented or published knowledge, which they are aware of and which was used in their invention. Moreover, the examiner has a right to add other citations he/she finds applicable in the given case, even though the inventor may not know about the inventions added. This is one of the major weaknesses of a patent citation as an indicator of knowledge spillovers.

The USPTO dataset provides data on all the applications resulted in granted patents and in this case already contains the citations indicated by the patent office investigators. On the other hand, the EPO data describe all applications submitted and thus, contains citations indicated only by the applicants themselves. Because of this, we consider these two data pools separately in the effort to figure out to which extent these procedural differences influence the main implications from the model.

All other advantages, including a vast pool of data available, explicitness of patent claims and, of course, absence of another equivalent alternative make the patent citation a good object for knowledge transfer analysis. 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 date the patent was granted, 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 classes indicated by ISIC.

The patent citation pool is used to build another interesting dataset. We aggregate the citation data and summarize it 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 EPO and USPTO), the number of citations from patents applied for with EPO, the number of citations from patents applied for with USPTO, the total number of citing patents applied for (both with EPO and USPTO), the number of citing patents applied for with EPO, the number of citing patents applied for with USPTO.

These data allow us to calculate the value of the 'propensity to cite' – the ratio of the total number of citations from patents and the total number of citing patents. It is interesting to investigate the relationship between the propensity to cite and the size of the firm, which is engaged in the patenting process. We are going to use this parameter as a crude measure for the knowledge spillovers from which the firm acquires benefits. In addition to that, we have at our disposal the 1998 annual turnover numbers to obtain a measure of the firm's 'visibility' and 'size'.

3. A 'SURFACE SCOOP'

After a preliminary processing of the data, we are able to make certain preliminary observations and conclusions and in this way to build up some basis for the further model-based analysis. These results come from basic aggregation of the data on the number of patents and citations, corresponding to different firms, countries, and industries. We present the preliminary ('surface scoop') results, which can be grouped in several sections:

- geographical 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 in ISIC.

Geographical distribution of citations. First, we consider the basic geographic distribution of citations made by Belgian applicants. In **Table I** we list ten countries, from which come the largest number of citations, together with the number of citations coming from other countries and the overall number of analyzed citations. Results are presented separately for the patents applied with EPO and the granted patents applied with USPTO (see **Table II**).

As it can be seen, the list of 'top ten performers' is almost the same in both European and American offices. The USA patents are the ones cited the most. According to data from the EPO, Belgian patents are the second most cited, and according to USPTO, Belgium is on the third position after USA and Japan. 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), p. 6-7) than the more distant ones. This seems not to be the fact in both patent offices. Patents from the United States are the most frequently cited by Belgian companies, which allows us to assume existence of a very strong 'transatlantic' knowledge flow. The 'Japanese' knowledge spillover channel is also quite strong. The rest of positions are occupied by the countries of the European Union (EU) among which four countries, Germany, Great Britain, France, and The Netherlands, occupy (in varying order) positions from 4 to 7 in both tables.

Thus, we conclude that the 'geographic proximity' assumption is not strongly supported by the collected data: domestic patents are not the most frequently cited. However, we should remember that the presented research gives only a snapshot picture of the Belgian applicants' patent citation behavior in 1997. This fact definitely must be tested over time to derive implications that are more time-consistent.

Firm-oriented distribution of patents and citations. The second block of preliminary results deals with the 'top-ten performers' among the investigated firms. **Table II** presents the list of firms with the highest number of patent applications submitted to the EPO and the successful applications with the USPTO. The total number of analyzed applications with the two patent offices is different: 601 submitted to the EPO; 290 submitted, and afterwards granted, to the USPTO.

Table III shows the top ten firms ranked by the number of citations listed in their patent applications with the European and American patent offices. The interesting fact is that although the number of patent applications submitted to the EPO was much larger, than of those submitted to the USPTO, the total number of citations 'generated' exhibits quite an opposite relationship. The smaller number of applications with the USPTO (48% of those with EPO) yields roughly twice the number of citations. Our main explanation of this fact is based on the knowledge that due to the procedure requirements the patent office investigators assign additional patent citations as they find appropriate.

The results, presented in **Table III** are closely related to the findings, already presented by Plasmans *et al.* (1999). A very high percentage of all patent applications in 1997 was submitted by the Flemish firms led by Agfa-Gevaert, which 'is responsible' for a quarter of all Belgian EPO patent applications and more than one third of all Belgian grants with the USPTO. We should note that in both offices, a small group of firms (17 overall, see **Table IV**) generates a much larger fraction of patent applications/grants and patent citations respectively.

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

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 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 average citation lag computed using EPO's data equals 5.79 years and the average citation lag computed using USPTO's data is 8.66 years. Thus, it appears that patent applications to the EPO tend to cite patents that are 'fresher' than those, which are cited by the applications submitted to the USPTO. Such fact can be explained by a number of older citations added by the USPTO officers, who, in the thorough investigation, discover more (sometimes older) patents relevant to the particular invention.

Intra-industry citations in different industries. **Table V** presents the percentages of intra-industry citations among all the citations occurred in a given industry. The industries in the table were determined from the patent's main IPC, transformed using the IPC-ISIC concordance table. In determining the category of a patent, which indicates several categories in application, we used the first category listed.

Actually, the data in the table represent an extraction from 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)). We do not show the complete matrix in this paper due to its bulkiness and present only the 'main diagonal' elements corresponding to the citations, which occurred in the same industry. As we expected, these diagonal elements are quite 'heavy', i.e. there is evidence that intra-industry citations are more numerous than the citations between different industries. In **Table VI** we can see the industries, which account for the largest part of all citations considered. For example, according to both the EPO and USPTO data, patents from six classes: 3510+3520 (Chemistry, excluding Pharmacy), 3522 (Pharmacy), 3810 (Metal Products, excluding Machinery), 3850 (Instruments), 3820 (Other Machinery), and 3825 (Computers and Office Machines); account for a majority (74.1% in the USPTO and 74.4% in the EPO dataset) of the overall number of citations.

4. MODELS AND ESTIMATION

4.1. CITATION PAIRS MODELING

The Model

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

We pay our attention to one particular kind of event, 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 structure of knowledge spillovers, indicated by the patent citations of Belgian applicants in 1997. Our dependent variable is a weighted indicator, which has value 1 if the citation occurs in the patent of a given particular class, and equals 0 otherwise. On the basis of preliminary analysis we have choosen patents from six major industrial classes (mentioned above) 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 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 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 accessory 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 given ISIC class, thus their product in the citation pair gives certain likelihood measure of the patent citation itself to fall into this ISIC class. Moreover, 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 our equations using two different specifications of the binary choice model: probit model and logistic regression. In the preliminary estimation we compared two specifications for this model. One was a probit and another was a logit regression. Both these specifications showed almost identical goodness of fit, thus we have chosen the standard probit model, due to its better interpretability.

The estimation was conducted in two different datasets: the sample consisting of citations indicated in EPO data, another sample consisting of citations indicated in USPTO data. Every dataset contains a list of citation pairs, which already have occurred. Thus, if we consider the probability of a citation to occur in patent pairs from our dataset, it is equal to 1. In this population we select several other sub-events: 'the citation has occurred in the citing patent, which comes from class X'.

$\Pr ob(Y = 1) = F(\beta' x_i) = \int_{-\infty}^{\beta' x_i} \phi(t)dt$, with $\phi(t) := \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}$, $i = 1,2,..,n$, where n is the number of observations. In our case have:

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

The dependent variable Y is an indicator that the patent citation occurred in the particular patent class (see above).

Estimation results. There are several notes to be made about the interpretation of the results. Among the explanatory variables in our model we have two binary variables and one integer. It is known that the estimated coefficients of a probit model do not necessarily yield the value of the marginal effect of the independent variable. For the probit model, the marginal effect for an independent variable is calculated as the product of the corresponding equation coefficient and the value of the standard normal density function calculated at the means of regressors:

$\frac{\partial F(x_i'\beta)}{\partial x_{ij}} = f(x_i'\beta)\beta_j$, $i = 1,2,..,n$, $j = 1,2,.,k$, where $f(x_i'\beta)$ is the standard normal density of the structural part of the model[3]. Corresponding slopes are presented in the output tables (**Tables VII – XII**) together with the regression coefficients.

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

The coefficient for *SameIndustry* variable points at the higher likelihood of 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 industry.

Concerning the time difference between the citing and cited patents, only the USPTO sample has produced statistically significant result, which indicated the negative relationship between the time lag and the probability of the citation. On the other hand, the EPO sample 'states' that such relation ship is negative, but the coefficient is not significant.

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 the time dependence of the probability of citation in this industry requires additional inquiry.

Pharmacy (Table VIII). The estimation results for the Pharmacy industry exhibits similar results as for the Chemistry industry. Pharmacy industry shows a lower likelihood of the intra-firm citation and higher probability for knowledge spillovers in the same industry. Thus, we expect knowledge exchange to be more intensive among different firms, but in the limits of the same industry.

We are as well indecisive about the influence of the time lag on the estimated probability of the citation. The only difference is that the coefficient for the *TimeLag* variable is significant in the EPO's sample and insignificant in the USPTO's, which mirrors the results observed in the Chemistry industry.

Metal Products (Table IX). In this industry we have the case, when two samples yield the same corresponding results. The probably of the patent citation's occurrence is higher when two patents belong to the same firm and come from the same industry. Thus, the knowledge spillovers in Metal Products industry are weak and some R&D cooperation inducing measures may be advisable.

There is evidence of the negative relationship between the time lag between patents and the likelihood of the citation (with reservations about lower significance of the coefficient in EPO data). Both samples show approximately 1 percent decrease in probability for each additional year of the time lag.

Instruments (Table X). There are mixed signals coming from two samples concerning the estimated coefficient for a *SameFirm* variable. The coefficient, estimated in the EPO's sample, is strong, negative, and significant. On the other hand, the coefficient, produced by the American data, is weak, positive, and not significant at all. This prevents us of coming up with the definite conclusion about this relationship. However, there is strong evidence that in the

---

[3] If the logit model is chosen, then $\phi(t) := \frac{1}{1+e^{\hat{\beta}' x_i}}$, $i = 1,2,..,n$, and the slope is $\frac{\partial F(x_i'\beta)}{\partial x_{ij}} = \frac{\beta_j e^{-x'_i \beta}}{(1+e^{-x'_i \beta})^2}$, $j = 1,2,.,k$.

Instruments industry the probability of the citation is much lower when citing and cited patents belong to the same industry class.

The final 'verdict' for the Instruments industry states that it is likely to favor intra-industry knowledge spillovers, but has an undetermined attitude towards the fact of intra-firm citation. Knowledge in this industry is not likely to 'depreciate' fast, which is supported by the evidence of a significant positive relationship between the time lag between the patents and the probability of citation.

Other Machinery (Table XI). The name of this industry is quite ambiguous and makes it difficult to extract particular policy implications, although there is significant number of patent citations falling covered by it. The results show that in this industry the time difference between two patents negatively affects the probability of the citation. Knowledge seems to depreciate with time according to the USPTO.

Regarding the existence intra-firm spillovers, we find strong support for that in the USPTO data and weaker evidence coming from the EPO. The model leaves us inconclusive about the likelihood of intra- or inter-industry citation. In the USPTO sample we see the support for the inter-industry knowledge exchange and the opposite conclusion in the EPO sample. If we take into account the diversity of patents falling into any other-type category, the conflict of conclusions does not seem to be surprising.

Computers and Office Machines (Table XII). This industry deserves special attention due to its importance in current technology driven times. The model was not able to produce definite conclusion about the knowledge flows on the firm level. The USPTO data strongly advocates for the more intensive internal knowledge usage rather than external. On the other hand the EPO data points (with a weakly significant coefficient though) at higher probability of inter-firm citations. Concerning the inter-industry knowledge spillovers, there is strong support for such fact coming from both samples. The Computer industry has higher likelihood of the external knowledge assimilation.

Despite the same sign of the coefficients for the *TimeLag* variable, we can not sufficiently rely on the evidence from the USPTO data due to its extremely low statistical significance. The EPO provides acceptable support for the positive dependence of the probability of citation on the time difference between patents, thus indicating the higher rate of older knowledge utilization.

4.2. A FIRMS' ORIENTED ANALYSIS

The Model

The second part of this research is devoted to the analysis of the firms' data and the study of their citation intensities. We had created a sample of the firms' citation intensities (the ratio between the number of citations in the patents, containing the citations, and the total number of patents, the particular firm has applied for) and some descriptive statistics: the firms' weighted consolidated annual turnover and the weighted consolidated average employment[4].

We employ the instrumental variable approach and the Generalized Method of Moments (GMM) we to obtain a significant estimator for the relationship between the firm's annual turnover and its propensity to cite. We use the firm's employment figure as an instrumental variable (the employment is strongly correlated with the turnover, but is not correlated with the propensity to patent). An appropriate kind of the GMM estimator could be found by solving the problem:

[4] 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.

$$\min_{\beta}\left\{\frac{1}{n}[Z'(y-X\beta]'V^{-1}\frac{1}{n}[Z'(y-X\beta))]\right\}.$$

with $V^{-1}=\left[Var\left[\frac{1}{n}(Z'\varepsilon)\right]\right]^{-1}$ being the inverse covariance matrix, where it is assumed that the restrictions with higher variance are given less weight than those with a lower one. If a consistent estimator $\hat{V}^{-1}$ of this inverse covariance matrix is available, we have:

$$\hat{V}^{-1}:=\frac{n^2}{\hat{\sigma}_{\varepsilon}^2}(Z'Z)^{-1}.$$

The desired estimator can be derived as:

$$\tilde{\beta}=[(X'Z)(Z'Z)^{-1}(Z'X)]^{-1}(X'Z)(Z'Z)^{-1}(Z'y),$$

where X is an explanatory variable matrix (intercept and the log of *Turnover*), Z is an instrumental variable (the log of *AvEmpl*), and y is a dependent variable (*PropCite*). This estimator is precisely the 2SLS estimator of β.

Estimation results. The estimated GMM coefficients are presented in **Table XIII**. As we can see, this model shows that there is a significant relationship between the company's turnover and its citation intensity. Surprisingly, the coefficient came out to be negative. It seems that relatively smaller enterprises with the smaller turnovers and average employment figures tend to have higher citation intensities. This result also calls for a parallel with the research of Plasmans *et al.* (1999), which explored the evidence that smaller enterprises with smaller R&D budgets have the incentive to apply for more patents per R&D dollar invested, than large firms with large R&D budgets, i.e. smaller enterprises seem to be more active in patenting. Our result tells us, that smaller enterprises are also more active in citation and thus more active in knowledge diffusion.

5. CONCLUSIONS

The objective of this study was to investigate the patenting and patent citation behavior of the Belgian private firms using the 1997 patent citation data from the USPTO and the EPO. The attention of this study was concentrated on two major aspects of patent citation behavior: a firm-oriented study and the study of a patent citation using the probabilistic approach. We have conducted an extensive preliminary analysis of the data and built two 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 small list of firms of different size and visibility.
2. The data from two sources, EPO and USPTO, allows us to conclude that the size of the firm does influence the firm's patent citation behavior. Moreover, there is evidence that smaller firms with a smaller annual turnover and a smaller average employment appear to have higher patent citation intensities, than their larger partners/competitors.
3. The occurrences of patent citations in Belgian patents in six major industries were studied.
   - The model shows that among six industries, three industries exhibit similar behavior: the Chemistry (except Pharmacy) industry, Pharmacy, and the Metal Products industry. In these industries inter-firm knowledge spillovers are more likely to occur, but intra-industry patent citations are more likely than inter-industry ones. Thus, we can label the Chemistry, Pharmacy, and the Metal Products as industries with the higher intensity of inter-firm knowledge exchange in the limits of the same industrial class.

     Another pair of industries, Instruments and Computers (and office machines), fall into the category of the industries with higher likelihood of inter-industry knowledge spillovers, but without clear evidence (i.e. support from the both EPO and USPTO) of stronger or weaker firm-to-firm knowledge exchange.

- Regarding the time aspect of patent citation in different industries, the Chemistry and Pharmacy industries do not show particular tendency towards faster or slower knowledge depreciation in the industry. Patents in the Instruments and Computers industries are likely to cite older patents, utilizing in this way the older knowledge. The Metal Products and Other Machinery industries provide evidence for more intensive usage of recent knowledge in comparison to older one.

4. The estimated probability of a patent citation, calculated given a particular set of factors, can be used as a crude but efficient measure of the knowledge spillovers environment in a certain industry, and can be applied for various competitive behavioral models. Once the special feature of the industry is determined (such as likelihood of inter- or intra-firm spillovers and the likelihood of inter-industry knowledge exchange), we receive an understanding of the intensity of knowledge spillovers. For example, now we can state that in the investigated time frame, the Chemistry industry has higher intensity of knowledge spillovers among firms, that the Instruments industry. Thus, probably, the Instruments industry could use some measures to stimulate R&D cooperation.

We should note that this research provides only a static snap shot picture of the patent-citation-induced knowledge spillovers. The development of models towards incorporation of the time factor by expanding the dataset to a number of years is considered a next most desirable step.

Table I. Geographic distribution of patent citations in 1997 applications of Belgian firms submitted to the EPO and those granted by the USPTO.

EPO					USPTO			
1	USA	254	27.55%		**1**	USA	762	46.49%
2	Belgium	204	22.13%		**2**	Japan	397	24.22%
3	Japan	182	19.74%		**3**	Belgium	219	13.36%
4	Germany	71	7.70%		**4**	Germany	109	6.65%
5	Great Britain	38	4.12%		**5**	France	44	2.68%
6	France	35	3.80%		**6**	Great Britain	31	1.89%
7	Netherlands	29	3.15%		**7**	Netherlands	16	0.98%
8	Italy	21	2.28%		**8**	Canada	11	0.67%
9	Switzerland	19	2.06%		**9**	Switzerland	11	0.67%
10	Sweden	10	1.08%		**10**	Italy	9	0.55%
	Other	59	6.40%			Other	60	3.66%
	Total	**922**				**Total**	**1639**	

Table II. Number of patent 1997 applications submitted by Belgian firms to the EPO and those granted by the USPTO.

EPO					USPTO			
1	Agfa-Gevaert	149	24.79%		**1**	Agfa-Gevaert	106	36.55%
2	Solvay	40	6.66%		**2**	Janssen Pharmaceutica	22	7.59%
3	Fina Research	31	5.16%		**3**	Solvay	22	7.59%
4	Janssen Pharmaceutica	28	4.66%		**4**	De Ster	14	4.83%
5	New Holland Belgium	22	3.66%		**5**	Esselte	10	3.45%
6	Esselte	17	2.83%		**6**	Dow Corning	7	2.41%
7	Bekaert	15	2.50%		**7**	Raychem	6	2.07%
8	Barco	10	1.66%		**8**	Bekaert	6	2.07%
9	Raychem	9	1.50%		**9**	International Sanitary Ware – Manufacturing	6	2.07%
10	Sofitech	8	1.33%		**10**	Heraeus Electro-Nite International	6	2.07%
	Other	272	45.26%			Other	85	29.31%
	Total	**601**				**Total**	**290**	

Table III. Number of patent citations listed in 1997 patent applications submitted by Belgian firms to the EPO and those granted by the USPTO.

EPO					USPTO			
1	Agfa-Gevaert	246	26.68%		**1**	Agfa-Gevaert	532	32.46%
2	Solvay	82	8.89%		**2**	Janssen Pharmaceutica	131	7.99%
3	Fina Research	71	7.70%		**3**	Esselte	99	6.04%
4	Janssen Pharmaceutica	65	7.05%		**4**	Dow Corning	86	5.25%
5	Esselte	31	3.36%		**5**	De Ster	68	4.15%
6	Raychem	20	2.17%		**6**	Bekaert	60	3.66%
7	Barco	20	2.17%		**7**	Solvay	60	3.66%
8	New Holland Belgium	20	2.17%		**8**	International Sanitary Ware – Manufacturing	55	3.36%
9	VCST	18	1.95%		**9**	Raychem	55	3.36%
10	Raffinerie Tirlemontoise Services	14	1.52%		**10**	Glaverbel	46	2.81%
	Other	335	36.33%			Other	447	27.27%
	Total	922				**Total**	1639	

Table IV. Profiles of selected Belgian firms (as by the end of 1998), mentioned in tables and discussion. Source: Bureau van Dijk

	Name	Weighted Consolidated Turnover (billion BEF)	Weighted Consolidated Average Employment (employees)
1	Solvay	1959.64	105900.20
2	Glaverbel	1613.99	196019.67
3	Janssen Pharmaceutica	1021.43	37523.02
4	Bekaert	802.30	102277.09
5	Vcst	674.17	96617.00
6	Barco	611.42	69727.03
7	Agfa-Gevaert	594.88	34671.06
8	Esselte	142.80	24428.95
9	New Holland Belgium	90.14	3683.44
10	Raychem	35.46	1945.47
11	Heraeus Electro-Nite International	19.82	3045.00
12	De Ster	5.33	504.00
13	Fina Research	2.60	474.00
14	Dow Corning	2.00	394.00
15	Sofitech	0.62	92.00
16	International Sanitary Ware – Mfg.	-	-
17	Raffinerie Tirlemontoise Services	-	-

Table V. Intra-industry Citations Percentages (as a fraction of all citations occurred in the industry).

Industry		EPO	USPTO
1000	Agriculture	-	33.33
3100	Food, beverages, tobacco	45.95	40.00
3200	Textiles, clothes, etc.	63.64	48.53
3300	Wood and furniture	44.44	26.32
3400	Paper, printing and publishing	68.42	22.22
3510+3520	Chemistry, except pharmacy	58.42	49.19
3522	Pharmacy	49.56	43.24
3530+3540	Oil refining	28.00	100.00
3550+3560	Rubber and plastic products	50.00	5.88
3600	Stone, clay and glass products	64.58	26.76
3710	Ferrous basic metals	100.00	33.33
3720	Non ferrous basic metals	100.00	62.50
3810	Metal products, ex. machines	47.37	43.91
3820	Other machinery	48.65	50.46
3825	Computers & office machines	70.49	50.96
3830	Electric mach., ex. electronics	76.47	62.30
3832	Electronics	67.57	71.56
3840	Other transport	-	-
3841	Shipbuilding	-	-
3843	Motor vehicles	96.00	90.91
3845	Aerospace	-	-
3850	Instruments	72.62	72.30
3900	Other industrial products	29.73	26.92
4000	Utilities	-	38.89
5000	Building and construction	33.33	33.33

Table VI. Overall Citation Percentages (out of the whole sample)

Industry		EPO	USPTO
1000	Agriculture	-	0.2
3100	Food, beverages, tobacco	2.3	0.5
3200	Textiles, clothes, etc.	2.1	2.4
3300	Wood and furniture	0.6	1.4
3400	Paper, printing and publishing	5.9	4.5
3510+3520	**Chemistry, except pharmacy**	**24.4**	**17.6**
3522	**Pharmacy**	**14.2**	**11.9**
3530+3540	Oil refining	1.6	0.0004
3550+3560	Rubber and plastic products	0.1	0.6
3600	Stone, clay and glass products	3.0	2.5
3710	Ferrous basic metals	0.1	0.2
3720	Non ferrous basic metals	0.5	0.3
3810	**Metal products, ex. machines**	**4.7**	**8.2**
3820	**Other machinery**	**11.5**	**15.5**
3825	**Computers & office machines**	**3.8**	**5.6**
3830	Electric mach., ex. electronics	2.1	2.2
3832	Electronics	2.3	3.9
3840	Other transport	-	-
3841	Shipbuilding	-	-
3843	Motor vehicles	1.6	0.4
3845	Aerospace	-	-
3850	**Instruments**	**15.7**	**15.2**
3900	Other industrial products	2.3	4.7
4000	Utilities	0.2	0.6
5000	Building and construction	0.9	1.4

Table VII. Probit regression results. Dependent variable corresponds to citations occurred in the Chemistry (excluding Pharmacy) industry.

3510+3520	Coefficient	Slope	std. err.	Chi-square		
USPTO						
Intercept	0.9901		0.0563	308.7973		
SameFirm	-0.4803	-0.3837	0.0741	41.9879		
SameIndustry	0.5307	0.4240	0.0690	59.1328		
TimeLag	-0.0184	-0.0147	0.0051	13.1062	$f(\hat{\beta}'\bar{x})$	0.837514
EPO						
Intercept	0.4994		0.0721	47.9540		
SameFirm	-0.1255	-0.0955	0.0834	2.2670		
SameIndustry	0.4647	0.3537	0.0770	36.4433		
TimeLag	0.0076	0.0058	0.0086	0.7881	$f(\hat{\beta}'\bar{x})$	0.761059

Table VIII. Probit regression results. Dependent variable corresponds to citations occurred in the Pharmacy industry.

3522	Coefficient	Slope	std. err.	Chi-square		
USPTO						
Intercept	1.1919		0.0630	358.2607		
SameFirm	-0.6436	-0.5753	0.0769	70.1087		
SameIndustry	0.5121	0.4577	0.0826	38.4269		
TimeLag	-0.0050	-0.0044	0.0058	0.7367	$f(\hat{\beta}'\bar{x})$	0.893862
EPO						
Intercept	0.7264		0.0831	76.3440		
SameFirm	-0.1803	-0.1564	0.0902	3.9972		
SameIndustry	0.3626	0.3145	0.0927	15.2941		
TimeLag	0.0481	0.0417	0.0110	19.1795	$f(\hat{\beta}'\bar{x})$	0.867341

Table IX. Probit regression results. Dependent variable corresponds to citations occurred in the Metal products (excluding machines) industry.

3810	**Coefficient**	**Slope**	**std. err.**	**Chi-square**		
USPTO						
Intercept	1.3675		0.0688	395.0380		
SameFirm	0.3340	0.3212	0.1145	8.5054		
SameIndustry	0.2162	0.2079	0.0796	7.3698		
TimeLag	-0.0119	-0.0114	0.0062	3.7016	$f(\widehat{\beta}'\overline{x})$	0.921415
EPO						
Intercept	1.5283		0.1132	182.3407		
SameFirm	0.8146	0.7833	0.2206	13.6421		
SameIndustry	0.3019	0.2903	0.1254	5.7912		
TimeLag	-0.0122	-0.0117	0.0133	0.8329	$f(\widehat{\beta}'\overline{x})$	0.96157

Table X. Probit regression results. Dependent variable corresponds to citations occurred in the Instruments industry.

3850	**Coefficient**	**Slope**	**std. err.**	**Chi-square**		
USPTO						
Intercept	1.3772		0.0650	448.7875		
SameFirm	0.0302	0.0264	0.0904	0.1115		
SameIndustry	-1.0538	-0.9210	0.0659	256.0782		
TimeLag	0.0213	0.0186	0.0057	13.9416	$f(\widehat{\beta}'\overline{x})$	0.873914
EPO						
Intercept	1.3361		0.0902	219.1672		
SameFirm	-0.2419	-0.2089	0.0939	6.6331		
SameIndustry	-0.8566	-0.7396	0.0856	100.0365		
TimeLag	0.0304	0.0262	0.0105	8.4376	$f(\widehat{\beta}'\overline{x})$	0.863475

Table XI. Probit regression results. Dependent variable corresponds to citations occurred in the Other machinery industry.

3820	Coefficient	Slope	std. err.	Chi-square		
USPTO						
Intercept	1.1305		0.0584	374.6987		
SameFirm	0.3228	0.2743	0.0923	12.2442		
SameIndustry	-0.1164	-0.0989	0.0638	3.3272		
TimeLag	-0.0147	-0.0125	0.0051	8.1507	$f(\hat{\beta}'\bar{x})$	0.849821
EPO						
Intercept	1.4158		0.0906	244.2295		
SameFirm	0.0667	0.0599	0.1101	0.3671		
SameIndustry	0.4726	0.4242	0.0987	22.9330		
TimeLag	-0.0629	-0.0565	0.0098	41.0171	$f(\hat{\beta}'\bar{x})$	0.897725

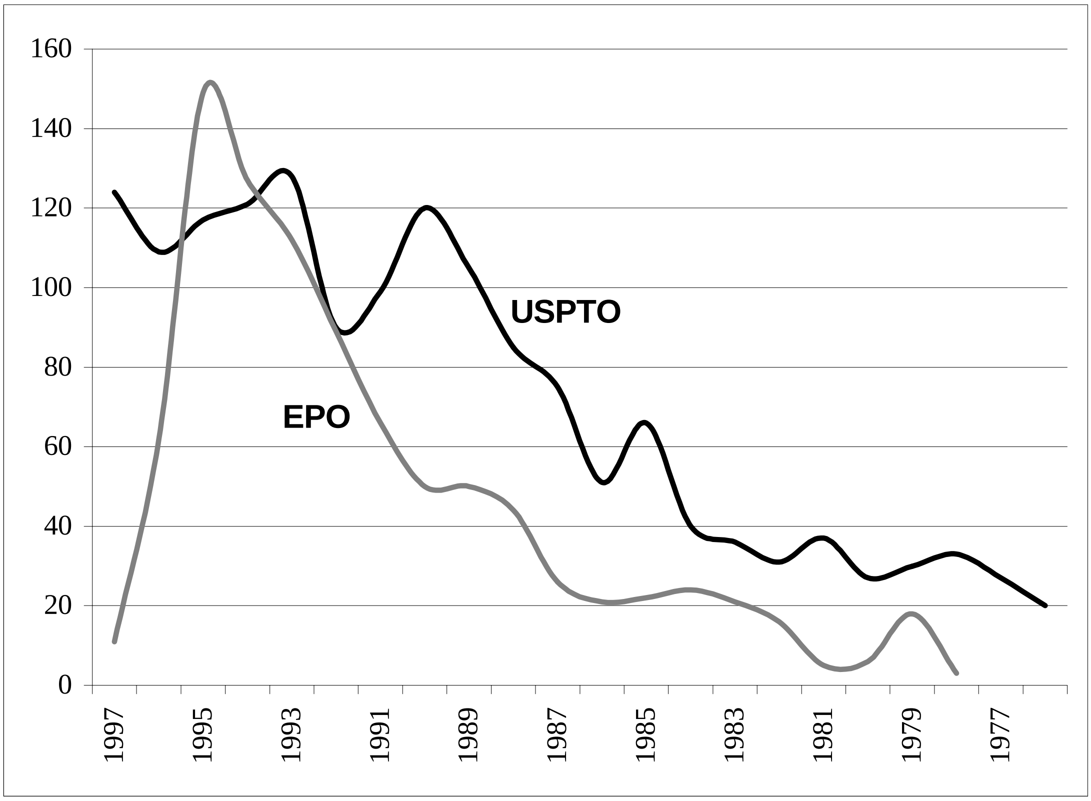
Table XII. Probit regression results. Dependent variable corresponds to citations occurred in the Computers and office machines industry.

3825	Coefficient	Slope	std. err.	Chi-square		
USPTO						
Intercept	1.6605		0.0563	869.1847		
SameFirm	0.5875	0.5575	0.1586	13.7155		
SameIndustry	-0.3148	-0.2988	0.0828	14.4543		
TimeLag	0.0003	0.0003	0.0013	0.0400	$f(\hat{\beta}'\bar{x})$	0.949088
EPO						
Intercept	2.0109		0.147325	186.2991		
SameFirm	-0.1607	-0.1558	0.141741	1.2857		
SameIndustry	-0.6576	-0.6375	0.134057	24.0628		
TimeLag	0.0305	0.0296	0.017168	3.1603	$f(\hat{\beta}'\bar{x})$	0.9695

Table XIII. GMM estimators for the Propensity to Cite Model. Dependent variable: *Prop2Cite*, Instrumental variable: log(*AvEmpl*).

	Coef.	Std. Err.	t-stat.		
C	1.835553	0.27989	6.56		
log(Turnover)	-0.077816	0.01630	-4.78	**R^2**	0.1617

Figure 1. The distribution of citations to patents issued in different years (EPO and USPTO 1997 data)



References:

Bersnstein, 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.

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

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

De Bondt, R. and I. Henriques (1995), Strategic Investment with Asymmetric Spillovers, *Canadian Journal of Economics*, XXVIII, No. 3, pp. 656-674.

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., R. Henderson and M. Trajtenberg (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.

Leahy, D. and J. P. Neary (1997), R&D Spillovers and the Case for Industrial Policy in an Open Economy, Discussion Paper No. 345, *Centre for Economic Performance*, London School of Economics and Political Science, London.

Lukatch, R. and J. Plasmans (2000), R&D and Production Behavior of Asymmetric Duopoly Subject to Knowledge Spillovers, *TEW Research Paper*, UFSIA – University of Antwerp.

Petit, M. L. and B. Tolwinski (1999), R&D Cooperation or Competition?, *European Economic Review*, No. 43, pp. 85-208.

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.

Suzumura, Kotaro (1992), Cooperative and Noncooperative R&D in an Oligopoly with Spillovers, *American Economic Review*, Vol. 82, No. 5, pp. 1307-1320.

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*.