Information Economics and Policy 23 (2011) 141–157

Contents lists available at ScienceDirect

Information Economics and Policy

journal homepage: www.elsevier.com/locate/iep

Sources and characteristics of software patents in the European Union: Some empirical considerations

Francesco Rentocchini *

Department of Economics, University of Trento, via Inama 5, 38100 Trento, Italy

A R T I C L E   I N F O

Article history:
Received 1 September 2008
Received in revised form 23 December 2010
Accepted 28 December 2010
Available online 8 January 2011

JEL classification:
O34
D23
L86

Keywords:
Software patents
Intellectual property rights (IPRs)
Innovation
EPO

A B S T R A C T

Software patenting is an increasingly important phenomenon in the European Union. Using a novel database of more than 30,000 software patents granted to both European and Non-European companies, we investigate the relevant factors explaining firm-level software patenting at the European Patent Office. We find that software patents are mainly applied for by American and Japanese firms, that they are characterised by a higher than average length of the granting procedure and that firms belonging to the software sector generally do not apply for them. Finally, results from non-linear panel data estimation reveal that patents are not deemed as useful appropriability instruments by software firms and that a "threat effect" by hardware firms is growing in importance. This last result is in line with recent developments in the literature relative to strategic patenting.

1. Introduction

During the last ten years, the number of filed and granted patents at the main three patent offices – United States Patent Office (USPTO), European Patent Office (EPO) and Japanese Patent Office (JPO) – has increased spectacularly. This increase has been driven mainly by patent filings in high-tech classes (Hall, 2004). Among these, software patents attract particular interest mainly because of the nature of the technology and because software patentability has been, quite recently, at the centre of a debate in Europe (Borrás and Kahin, 2009).

For a long time, the economic literature has recognised the importance of the patent system in shaping and directing the rate of appropriation of the innovative effort of firms (Arrow, 1962). In addition to classical contributions, the literature developed to explain the recent trends in worldwide patenting has relied on Schumpeter's contributions to economic thought (Schumpeter, 1942). More recently, evolutionary economics (Nelson and Winter, 1982) has focused on the role of patents in enhancing or hindering innovation, depending on industries where firms compete. Therefore, a number of authors began to stress that, depending on appropriability conditions of industries, patents may or may not be a useful institutional mechanism for promoting the variety of technological solutions and the selection by market forces via competition (Merges and Nelson, 1990; Boldrin and Levine, 2002; Bessen and Maskin, 2009).

Hence, on the one hand, empirical literature has shown how patents may constitute a suitable appropriability mechanism in a high number of sectors (Cohen et al., 2000); however, on the other hand, we have witnessed an explosion in the number of patents filed in recent years. A set of research questions arises from these seemingly

* Tel.: +39 0461 882162; fax: +39 0461 882222.
E-mail address: francesco.rentocchini@economia.unitn.it

0167-6245/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.infoecopol.2010.12.002

contradictory patterns of diffusion. Among them, why such a trade-off exists and which factors are likely to explain it at the micro-level.

One major explanation put forward in recent contributions highlights the role played by the strategic behaviour of firms aimed at hindering competition, obtaining licensing revenues and increasing their power in negotiations. In particular, Mazzoleni and Nelson (1998) stress how, in industries where the innovation process relies mainly on improvements made by others, such as cumulative system technologies, it is more likely that strategic patenting behaviours, such as cross-licensing, blocking rivals or extracting licensing revenues, are found.

Another example of strategic patenting refers to submarine patents, that is, a patent issued after an extraordinarily long period of pre-grant review, thus allowing the applicant to reveal it only after the new patent covers widely adopted technologies. In this way, the owner of the original patent can pursue infringement actions, or seek injunction, against other technology adopters. Graham and Mowery (2004) investigated the role of procedural revisions of patent applications (called "continuations") in software patents in the United States from 1987 to 1999. They individuate a rapid growth in the use of continuations between 1987 and 1995 in software patenting.

The main contribution of the present article is to provide evidence of strategic patenting at the EPO in a cumulative system technology such as software. Although, according to article 52(2)(c) and (3) of the European patent convention, computer programs "as such" are not patentable, if the subject matter specifies words relating to computers or another conventional programmable apparatus, it is to be examined as a computer-implemented invention. As the practice of EPO indicates, computer-implemented inventions are patentable provided that such an invention has a technical character that involves a technical contribution to the prior art (Freedman, 2000). Evidently, the distinction between "pure" software patents, i.e., patents protecting inventions that can be fully considered to be software and computer-implemented inventions, is a tricky exercise that turns out to be even more difficult in practical terms. It is so because "a software patent concerns an invention about a software-based computer implementation, while a computer-implemented invention is about an invention that may be implemented in software" (Bergstra and Klint, 2007, p. 277). To disambiguate such issues and eliminate differences existing in the protection of computer-implemented inventions in different EU States, the directive on computer-implemented inventions was proposed in early 2002. The same directive was eventually rejected by the European Parliament in mid-2005. Therefore, a clear legislative distinction between "pure" software patents and computer-implemented inventions remains to be drawn, and the difference between the two has yet to be found on a case-by-case basis.

To address the issue, we put forward a novel dataset comprising software patents. Given the problem discussed in the previous paragraph, we acknowledge the fact that our dataset may contain a significant number of computer-implemented inventions. We nevertheless provide a number of reliability checks on our dataset to convince the reader that our database can be taken as a proxy of software patents accorded by the EPO in recent years. Before proceeding, we provide a literature background dealing with the issue of software patents, revealing the most striking results concerning strategic patenting in this area. For the US patent system, several works have already been presented, but the EU has been mainly disregarded with the exception of a couple of works (see Section 2). Section 3 proposes a theoretical model aimed at explaining factors affecting software patenting at the firm-level. Particular interest will be dedicated to the question of whether strategic patenting is actually an issue in the EU and whether different behaviours refer to different industries under scrutiny. An original dataset for the period 2000–2003 is put forward in Section 4, which links the number of software patents filed at the EPO with R&D spending and other relevant variables related to applicants. Consistency and representability of the database is also carefully checked compared to other methodologies implemented in the literature. Negative binomial panel data estimation is then performed to discover the most relevant factors affecting software-patenting decisions for firms belonging to different industries (Section 4). Finally, the results are presented and discussed (Sections 5 and 6).

2. Background

Studies dealing with software patents refer primarily to the US patent system, where software has been patentable subject matter since 1981.[1] Allison and Lemley (2000) and Allison and Tiller (2003) were the first to carry out a detailed analysis of more than 200 software patents, defined as such by reading the description of every single patent. Their main interest lies in the comparison of internet-related patents and general patents to test the general idea that internet business method patents have not been properly searched for relevant prior art, meaning that they are likely to be of poor quality. The main conclusion of the study points out that there is little support for the main criticism because internet-related patents are found to be characterised by the same amount of prior art references as more general patents.

Although the work carried out by Allison et al. is seminal in the field of software patenting, the proposed methodology presents the drawbacks of being highly time-consuming and not able to provide sufficiently large samples for statistical analysis. For these reasons, subsequent works have tried to develop a methodology to distinguish as correctly as possible patents that protect software from patents protecting other technologies. In this respect, one of the first methodologies of this kind was proposed by Graham and Mowery (2003). They examine all of the patents falling into identified IPC classes and define software patents as such.[2] The main findings can be

1 See court decision *Diamond v. Diehr*, 450 US 175 (1981).

2 These IPC classes are individuated by analysing overall patenting by the six largest US producers of personal computer software, based on their 1997 calendar revenues.

summarised as follows: (i) larger and older firms tend to increase their patent propensities; (ii) large electronic systems firms are more important than packaged software ones in software patenting; (iii) the ratio between the number of citations received by patents owned by the top 100 packaged software firms and the number of citations of software patents overall is increasing, except for electronic firms; and (iv) a decreasing propensity to copyright software is found, thus pointing to a substitution effect.

An alternative methodology for identifying software patents has been put forward by Bessen and Hunt (2007). The authors develop a search algorithm, based on a fixed number of keywords, aimed at identifying the number and characteristics of software patents accorded by the USPTO during the period 1976–2002. It comes out that software patents are mainly developed by US inventors and are owned by US assignees. Moreover, they are more likely to be obtained by large firms, established firms and firms in manufacturing. The authors then put forward an econometric model to test, which factors contribute to explaining the rising propensity to patent software in the sample. They find evidence that capital-intensive firms tend to patent more because of the threat of hold-up by rivals. Furthermore, industries with a high propensity to patent are also those characterised by a high patent propensity in general. Overall, the rising patent propensity is not explained by any of the controls, thus leading the authors to conclude that this is caused by legal changes that occurred in the 1980s, when cost-effectiveness of software patents was reduced considerably.[3]

In line with the previous study, Chabchoub and Niosi (2005) adopt a keyword method to identify software patents and combine this information with company data from other sources. Then, they concentrate on factors affecting the propensity to patent software by American and Canadian firms during the period 1986–2002. Results from the study show that firms that are more likely to patent software are large firms, are characterised by a higher share of revenues in products and belong to clusters of innovative firms.

The two aforementioned methodologies have been fruitfully combined by Hall and MacGarvie (2010). The authors claim that combining the two techniques is a good way to minimise type I and type II errors (see Section 4) in the process of identification of software patents. First, they identify all of the US patent class-subclass combinations for which 15 software firms patent. By referring to these classes, they generate a first dataset, following the method adopted by Graham and Mowery (2003), and intersect the resulting database with the one obtained using Bessen and Hunt (2007) method. Two main results are obtained by analysing the resulting dataset: (i) the expansion of patentability negatively affected firms without patents and firms in the downstream sectors mainly because firms had to ask for licenses to have applications to run on middle-ware and operating systems and (ii) software patents turn out to be valued by the market more than ordinary patents. For hardware producers, this is likely to reflect the strategic value of software patents rather than their technological value. On the contrary, however, software patents are found to be technologically valuable for software firms.

Although the US has been the main area of reference for the analysis of software patenting, interest in the study of software patenting has increased in Europe in the last decade. Difficulties in data collection and the absence of a clear legislation relative to software patents[4] have constrained the analysis to rely solely on surveys rather than adopting more general approaches, such as in the US case. Nevertheless, there have been recent attempts to overcome such a limitation and to implement more general methodologies.

For example, McQueen (2005) relies on a bibliometric technique to individuate software patents among more general ones and computes the distribution of software patents accorded at the EPO in 15 EU countries, US and Japan for the years 1987, 1990, 1993, 1996 and 1999. He finds that 49% of software patents are assigned to European countries (with Germany accounting for 50% of the total amount), 18% goes to Japan and 29% to US. Moreover, over the last 12 years, there has been a 60% overall increase in software patent applications. In general, the work by McQueen (2005), despite being path-breaking for the European patent system, presents some drawbacks that are worth mentioning. First of all, the search on the EPO database is conducted with a fairly simple keyword search, which is likely to maximise the number of false negatives. Moreover, the constructed database is used only to provide a descriptive analysis of the phenomenon without a proper test on important research questions linked to the topics put forward in the US case.

Hall et al. (2007) instead present a more elaborate method comprising three main steps. First, they conduct a keyword search on the EPO dataset, adopting the same algorithm set forward by Bessen and Hunt (2007) for the US case. Second, they analyse the IPC classes of the patent portfolios of the world's 15 largest software firms, expanding it in to give an account of European firms as well. Third, they accept a restrictive definition of "pure" software patents as one falling in the intersection of the two sets defined by both the keyword and IPC methods. Thus, the authors use the constructed database to explore whether these patents are valued more than other patents. Results show that software patents, when only the crude number is taken into consideration, are more valuable than other patents, but the difference disappears if weighted by their intrinsic quality level. Thus, the authors conclude that the value of software patents in Europe depends more on their

[3] In particular, the authors claim that the formation of the Court of Appeals for the Federal Circuit in 1982 lowered standards of patentability, allowing the rights of patent holders to be more easily enforced in court.

[4] Article 52 of the European Patent Convention expressly prohibits software patents, however, the EPO rule of practice has put forward the idea that in case a "technical contribution" to the prior art is found, then software must be considered an invention and, therefore, capable of patenting. In this regard, the situation is not completely clear, and the interpretation of the "technical contribution" is not uniform. In this vacuum of legislation, although patenting of software and business methods is not permitted by the EPC, the EPO regularly provides a plurality of actors with patents on software because of the presence of numerous interpretations on the definition of software.

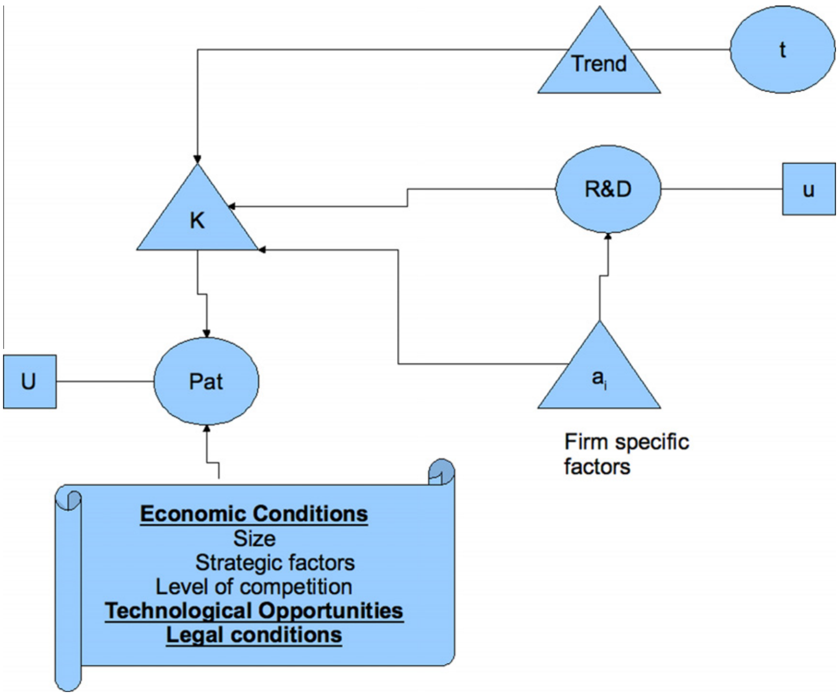

Fig. 1. KPF Theoretical model.

number rather than the quality of the invention they protect.

Although this work introduces many important insights into the topic of software patenting at the EPO, it presents some limitations that our work aims to address. First of all, the authors do not provide any kind of comparison of the constructed database of software patents with a trustworthy dataset containing true software patents. In our paper, we try to overcome this shortcoming by comparing our dataset of software patents with a sample of 78 patents containing both software and non-software patents. Second, Hall et al. (2007) exclude non-European firms from the analysis, and their sample is strongly centred on the UK due to constraints on the availability of R&D data. On the contrary, we incorporate non-European companies, knowing that, given their better experience in dealing with software patents (especially for American firms), foreign companies are likely to constitute the largest share of companies patenting software at the EPO. Finally, Hall et al. (2007) are more concerned with the issue of differing valuations of software patents by firms in the European and US patent systems, whereas we are more interested in the presence of strategic patenting in the form of a ''threat'' effect by other firms.

3. The model

The study of the effect of R&D spending and other factors on the number of patents filed has relied mainly on the Knowledge Production Function (KPF) approach. The main idea is that R&D expenditures at the firm level can be interpreted as a correct proxy for the production of knowledge. Then, if we are able to calculate the stock of knowledge for a certain firm at a fixed point in time, this value is likely to be a correct proxy for the output of the KPF (Pakes and Griliches, 1984).

Fig. 1 contains the classical rationale for the KPF approach, augmented by a set of factors that we deem to be important in our analysis.[5] The main factors therein contained are the following:

- *K*, which represents the dynamics of the stock of knowledge of firm i at time t;
- a_i, which indicates firm-specific factors constant through time. Managerial ability, opportunities and other similar factors are all examples of the mentioned variable. Indeed, managerial ability and other firm-specific conditions may have an influence on both the amount of R&D spending of the firm as well as on the output of the innovation process and hence on the stock of knowledge produced by the firm;
- *R&D*, which represents the amount of R&D expenditures of firm i at time t;
- *Pat*, which is the number of patents filed by firm i at time t;
- u and U, which are the error terms for the measurement of R&D expenditures and patent counts, respectively;
- *Trend*, which is a factor controlling for the presence of specific trend patterns as time t passes.

Thus far, the problem remains that what is produced through the R&D effort of the firm is a rather unobservable quantity, namely, technological knowledge. Hence, an in-

[5] In Fig. 1, triangles represent unobservable quantities, circles are observable quantities and squares are disturbance terms.

dex of the output of this process is needed. In this regard, the economic literature has relied on the number of patents filed by a single firm at a fixed point in time (Griliches, 1990).

The present model is simple and contributed to the understanding of the relationship between the amount of R&D spent by the firm and the output of the innovation activity. Obviously, the amount of R&D cannot be thought to be a simple amount of spending that is done once per year and whose value stays constant through time. On the contrary, the R&D diminishes its own value as time passes, that is, it depreciates. For this reason, the concept of R&D stock is implemented (Griliches and Mairesse, 1981). Moreover, R&D stock is able to provide a reason why a certain amount of R&D at time t is affected by past quantities.

Together with the amount of R&D expenditure, other factors contribute to the understanding of the output of the KPF. These factors are crucial as well. They have been classified into three main groups: economic, technological and legal conditions.

The first group, **economic conditions**, is composed of three main factors. First, *size* influences the innovation process of the firm due to four main reasons. First, large firms benefit from economies of scale and scope. In this way, they are more competitive than smaller ones (Cohen et al., 2000). Second, large firms benefit from complementarities and spillovers coming from other departments. Third, capital markets are more prone to finance risky innovation projects of larger firms other than small ones (Peeters and Van Pottelsberghe de la Potterie, 2006). Fourth, large firms are more likely to be endowed with a legal department that handles IPR matters (Lerner, 1995). Second, the *level of competition* is likely to play an important role. Two opposite effects are present in this case: first, a "replacement effect", according to which firms with a high market power are less likely to invest in R&D and, as a consequence, to innovate. The main reason resides in their lack of incentives to spend more on R&D, due to their dominant position in the market (Arrow, 1962). The second effect is the "efficiency effect", which states that firms with a high market power are more likely to innovate because they do not face any kind of competition for the exploitation of the results of their innovative activity (Gilbert and Newbery, 1982). Finally, *strategic factors* explain recent trends in patenting strategies at the firm-level. In fact, although traditional "incentive theory" advocated for a long time that the monopoly power, accorded to the patent holder, acts as an incentive to R&D expenditure, recent contributions assert that the high number of patents filed by companies, in particular larger ones, are instead a strategy aimed at hindering competition and increasing their monopolistic position (Hall and Ziedonis, 2001). This behaviour is likely to take place in "cumulative system" technologies, that is, technologies for which the innovation process is highly cumulative. Therefore, the software sector, for the essential cumulativeness of its embedded technology, is also prone to be threatened by strategic patenting activities.

The second set of factors, i.e., **legal conditions**, is proxied by geographical controls. Indeed, different opportunities may arise from being located in different regions having different legislation. Among them, four main macro-areas have been identified: the European Union, the United States, Japan and other countries.[6]

The third set of factors is constituted by **technological opportunities**. We partially control for the presence of technological opportunities using the industrial sector of activity of the firm.[7] Indeed, the effect of formal R&D spending on the innovation output, mediated by the rate of formation of the stock of knowledge capital, depends on the sector of activity of the firm (Mansfield, 1986). For our purposes, the industry is of particular interest. In fact, we want to investigate the different behaviours of firms belonging to two distinct sectors, hardware and software producers. It has been demonstrated that, during the last 10 years, main patentees at the USPTO are likely to be part of electrical, computing and instrument industries (Hall, 2004). Moreover, if only software patents are taken into account, firms belonging to electrical, machinery and instrument industries account for more than the 60% of software patents accorded at the USPTO. Software publishers and firms from other software industries contribute only 7% to the overall share of software patents (Bessen and Hunt, 2007). Hence, if firms that do not belong to the software sectors are more likely to patent software inventions, then it seems reasonable to suppose that they are doing it for reasons intrinsically different from spurring innovation spending.

4. Data

As discussed before, our main aim is to give an account of factors affecting software patenting by firms applying for patents at the EPO with a particular eye on strategic patenting. Unfortunately, there are no specific IPC classes in which software patents can be easily found, thus it is important to produce a reliable dataset able to minimise both type I and type II errors. Type I error refers to the error committed when many false negatives are detected, that is, when a patent that should have been included among software patents is actually excluded. On the contrary, type II error refers to false positives, that is, when a patent that is not related to software is instead classified as a software patent.

Contrary to studies proposed thus far that have built up static datasets according to well-defined methodologies, we rely on a database (i.e., Gauss database) made available and maintained by a group of practitioners, which is continuously updated and improved, thanks to its wiki nature. Section 4.1 proposes a general description of the Gauss database, with some relevant statistics concerning European software patents. Its reliability has been checked by

[6] While European and American legal regimes have been extensively discussed in the literature (Graham et al., 2002), Japan and other countries are worth mentioning. For example, Japan has changed its patent system from single-claim to a multiple-claim in 1988. Sakakibara and Branstetter (2001) show that, after the reform, overlapping patent claims have been extensively used to defend strategically acquired inventions.

[7] We are aware of the fact that sector dummies capture many effects overall and are not a fully reliable proxy of technological opportunities, but we use such dummies mainly as controls of specific industry-level effects, which are not captured by the other explanatory variables.

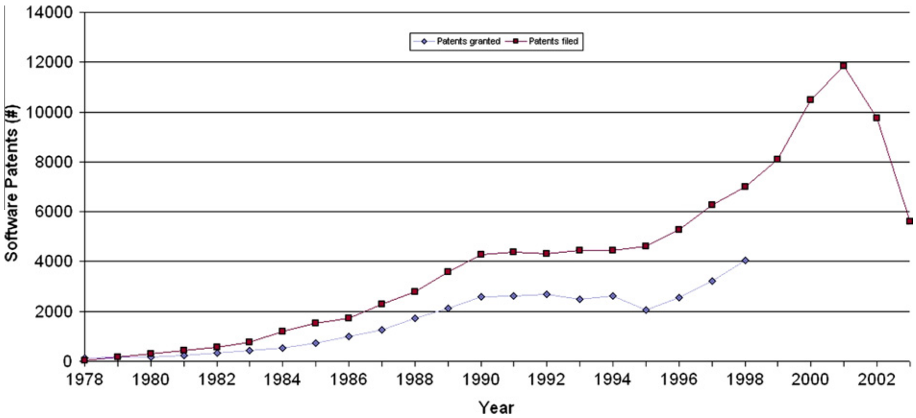

Fig. 2. Yearly evolution of filed and lagged granted software patents (1978–2003).

several means, including comparing it with another database built via a more standard methodology. Finally, a descriptive analysis of software patenting in the EU is carried out relying on the Gauss database.

Section 4.2 presents our sample comprising both European and foreign firms patenting software at the EPO. We first explain the procedure used to build the sample concerning firms' patenting strategies. Then, we check whether the sample is biased, comparing it with both ANBERD and EUROSTAT population statistics. Finally, the sample subset of data is presented, stressing the peculiarities that call for the use of defined econometric techniques and providing descriptive statistics for the sample itself.

The last step, taken in Section 4.3, is to discuss the ratio behind the adoption of particular variables in our analysis, together with the discussion of some technical issues concerning the econometric model adopted.

4.1. Gauss database and descriptive statistics

As mentioned, the present analysis of recent trends in software patenting inside the European Union relies on the Gauss database. This database has been created from different sources. First, it comprises a total of 1901 patent applications filed at EPO with existing equivalent USPTO patents falling in USPTO patent class 705, which constitute a class devoted exclusively to business methods and are thus more likely to contain software patents as well (Wagner, 2008). Second, a set of searches of patent documents has been conducted, according to the name of the applicant (mainly software companies) and approximately 150 words commonly occurring in software patents. Furthermore, focused searches in selected ECLA classes with a high probability of containing software patents have been carried out as well. Finally, the database has been public and in a wiki form since 1999. By doing this, not only are the maintainers of the dataset able to modify the patents contained therein,[8] but also all registered users have been able to keep it up-to-date and have helped to identify software patents and remove non-software ones.[9]

Appendix A provides a series of robustness checks on the reliability of the data at hand. These controls are done to show that our data are able to minimise both type I and type II errors and that the dataset performs well compared to the alternative dataset of software patents developed by Hall et al. (2007).

In this section, we will present the main features characterising the Gauss database. Overall, the Gauss database is composed of patents filed between 1978 and 2004. Much information has been extracted from the dataset. In particular, statistics concerning designated countries, the yearly evolution of the number of filed and granted software patents, country of residence for both inventors and applicants and patents' software domain. Fig. 2 depicts the steady increase in the number of software patents filed at the EPO since 1984. During the second half of the 1990s, the increase has been impressive, jumping from 4500 p patents in 1995 to almost 12,000 for the year 2001.[10] From the figure, we note how the pattern of granted patents follows closely that of filed ones; nevertheless, the gap between the two is increasing. This gap might be an indication of the increasing strictness of the EPO concerning this patent typology.

Other preliminary results can be inferred from Fig. 3. Applicants of software patents come mainly from US and Japan (39% and 25%, respectively), with a minor role played by European applicants. Germany, which is one of the best performing countries, accounts for only 10%. This pattern is mainly due to the leading role in ICT-related products by the US and Japan and the fact that, at least for the US, software is susceptible of patenting since the beginning of the

[8] Gauss is mainly maintained by an informal group composed of six people with diversified backgrounds, including financial analysts, founders of start-up software companies and physics researchers.

[9] Collaborative forms of information processing and filtering has shown to be very effective in recent years, e.g., the Wikipedia project together with other numerous Open Source programs. Gauss can be thought of as a successful experiment aimed at bringing this form of collaborative organisation to the patent system.

[10] Since 2001, the amount of software patents filed has dropped consistently. One of the reasons for this fall may be connected to the burst of the "dotcom" bubble that took place in the period 2000–2001. Indeed, the crisis of many firms in the ICT sector could have implied diminishing patent applications.

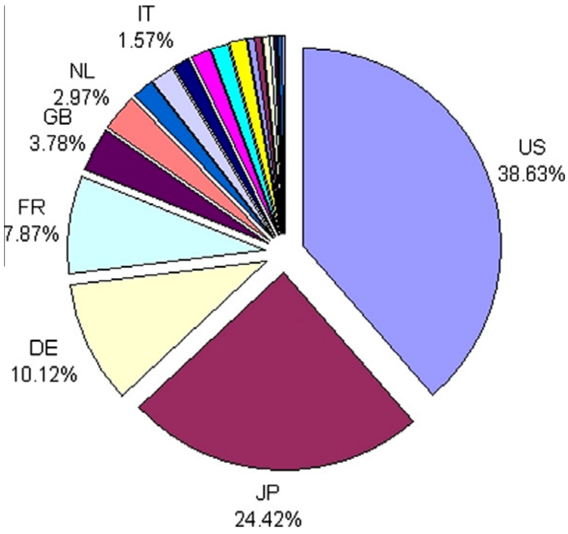

Fig. 3. Country of residence for top 20 applicants.

1980s. This scenario has allowed American firms to acquire expertise in both dealing with application procedures and identifying more valuable inventions to be patented.

If we focus our attention on software patent concentration, then a highly concentrated pattern is discovered. According to Fig. 4, the top 50 applicants account for more than the 50% of patents accorded at the EPO.

Examining the subset of the database in the period between 1995 and 2003 provides useful insights on the recent dynamics in software patenting in the EU. Data shows an increasing number of filed patents that are not granted or not yet granted: while about 60% of patents filed in 1995 switched to the granted state before the end of 2003, 83% of patents filed in 2000 have not been granted yet. This pattern justifies the low share of granted patents included in the mentioned subset, and it is connected to an increase in the time required to complete the granting process, whose average length is 3.5 years: while granted patents in 1997 were filed approximately one year earlier, those granted in 2003 took, on average, more than 5 years to complete the granting procedure. In the period 1995–2003, a lower number of patents were granted compared to an increasing number of filed requests, and in general, the granting procedure slowed. This finding can be explained by different means. First, the productivity of the EPO is decreasing. This drop in productivity is mainly due to two factors: the growing number of patents filed in general and an additional weight constituted by international patent applications. The former factor is due to the rising importance of patents among other IPRs. All patent offices around the world are facing a huge number of patent applications. These are not counterbalanced by an adequate investment in internal personnel. This fact means that the number of patents per employee is steadily increasing, leading the granting procedure to slow down. At the same time, EPO has been selected as the more efficient patent office around and, for this reason, it has attracted international patent applications, given the higher quality assured in the granting procedure. This fact has additionally increased the already huge number of patent applications to be processed.

Nevertheless, all of these reasons cannot fully explain the high difference in the average grant of the granting procedure between patents in general and software patents, i.e., 3.5 years compared to 5 years. The cause for this difference must be found in other factors, such as the complexity of the patenting matter and the absence of clarity concerning decision procedures. Moreover, the lack of a well-defined prior art contributes to the uncertainty surrounding the granting procedure.

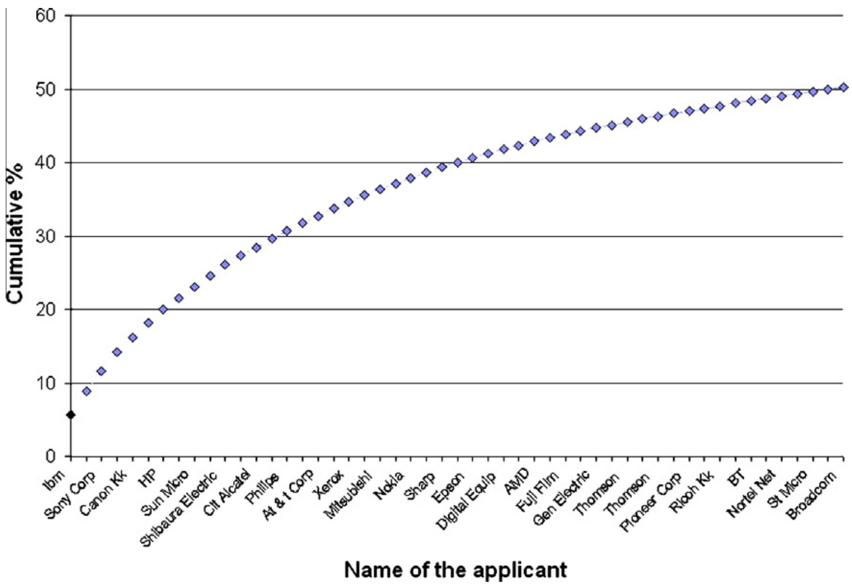

Fig. 4. Cumulative percentage of software patent applications by top 50 applicants.

Table 1
Descriptive statistics for the pooled sample.

	N	Mean	Standard Deviation	Min	Max
Number of software patents	3916	3.19382	19.2438	0	399
R&D intensity	3741	0.0023699	0.0141569	0.00000145	0.4911295
Number of employees	3849	27204.84	51063.54	1	477100
Industry concentration	3916	0.3198684	0.1285628	0.2120746	1
Strategic rivalry	2937	310.5781	555.2262	0	1852

4.2. Sample construction and description

To investigate the determinants of software patenting at the firm-level, a link has been established between a subset of the Gauss database and a set of other databases.[11] In particular, we followed a three-stage procedure. First, we matched the firm's name from the "2004 EU Industrial Research Investment Scoreboard"[12] (henceforth, R&D scoreboard) with the patent assignee name from Gauss.[13] Next, we re-matched the two databases through the name of subsidiaries, and we assigned the number of patents filed to the relative parent company. From these first two steps, we obtained a new sample comprising firms contained in the R&D scoreboard with the relative number of patents they have applied for in the period 2000–2003, defining it R&D-Gauss. Finally, we matched R&D-Gauss with both Amadeus and Osiris consolidated data by the firm's name to retrieve additional information for our analysis.[14]

A resulting dataset obtained by linking the information, available in the mentioned sources, is composed of 979 firms, whose data concerning R&D spending, sector and geographic classification and number of software patents filed are available for the period 2000–2003. To check the representability of our sample, we performed two main comparisons:

1. *ANBERD database vs. R&D-Gauss*. In this first phase, we compare the representability of our dataset with data from ANBERD. The latter is a comprehensive database containing information on the R&D spending in 21 OECD countries. Our database is found to perform very well with respect to this. Indeed, it accounts for the 73% of R&D conducted by countries contained in it. Moreover, whether or not the comparison is done at the sector level, the R&D database accounts for the 71.35% of R&D performed.
2. *EUROSTAT vs. R&D-Gauss*. From the R&D spending point of view, we are fairly confident that the sample taken into consideration is representative, but when the number of enterprises is analysed, this may not be correct anymore. Therefore, for the sake of comparison, we decide to rely on European statistics retrieved from EUROSTAT. The comparison between our database and data on firm population displays the low representability of our sample. Nevertheless, it has been possible to conduct the comparison only with respect to few countries, given the mismatch between our sample (comprising firms from several countries around the world) and population statistics from EUROSTAT (comprising only a limited number of European countries).

From the previous analysis we can conclude that our sample is not representative of the whole population of companies at the EPO, but that it gives a clear and reliable picture of R&D spending and of other relevant variables. We interpreted this fact as the ability of our sample to describe correctly the behaviour and characteristics of large firms applying for software patents at the EPO (Frietsch, 2004).

We turn now to provide a general description of the dataset. Table 1 provides descriptive statistics for the pooled sample. Table 2 presents the distribution of companies by industry, showing that IT hardware, electronic, electrical, software and computer services are those sectors where software patents are mainly present. Furthermore, companies patenting software at the EPO are mainly US and Japanese companies (see Table 3). Tables 2 and 3 also reveal a clear pattern of the sample, that is, there are a high number of firms not applying for any patent. Indeed, 693 companies out of 979 did apply for at least one software patent with priority year 2000–2003. Thus, the structure of the dataset calls for the implementation of a sound econometric model able to take into account the data's specific pattern. In this respect, the choice made of adopting count data models is supported by both the nature of the depending variable and the structure of the data.

4.3. Method of estimation and variables

The main objective of our analysis is to explain which factors influence the number of software patents a firm

[11] The subset of data refers to the information collected for the period 2000–2003.

[12] This is produced as a part of the "Investing in research: an Action Plan for Europe COM(2003)226 - EC DG Joint Research Centre" and lists the R&D spending together with other relevant information, of the top 500 EU and top 500 Non-EU corporate R&D investors for the period 2000–2003.

[13] To establish proper linking relations, a specific small software application has been developed, performing automatic matching between firm values and requiring explicit operator confirmation only in cases in which applicants were not univocally identified. This procedure was coupled with time-consuming manual processing of the data to strongly increase the reliability of the sample.

[14] Amadeus business directory contains account data of European companies and their subsidiaries located in the EU, together with subsidiaries of non-EU companies, and Osiris business directory contains account data for non-EU companies with their subsidiaries together with subsidiaries of EU companies located in non-EU countries. These sources also provide information concerning the ownership status, affiliates and subsidiaries, useful for consolidating the data at the level of the ultimate parent company.

Table 2
Distribution of companies by industry for the pooled sample.

FTSE code	With R&D	With software patents
Aerospace & defence (21)	100	31
Automobiles & parts (31)	240	75
Banks (81)	4	0
Beverages (41)	16	3
Chemicals (11)	312	68
Construction & building (13)	84	20
Diversified industrials (24)	68	15
Electricity (72)	68	15
Electronic & electrical (25)	288	132
Engineering & machinery (26)	364	87
Food & drug retailers (63)	8	0
Food producers (43)	88	10
Forestry & paper (15)	28	3
General retailers (52)	28	6
Health (44)	152	37
Household goods & textiles (34)	116	23
IT hardware (93)	552	255
Leisure & hotels (53)	12	4
Media & entertainment (54)	68	29
Mining (04)	20	1
Oil & gas (07)	76	12
Personal care & household (47)	56	15
Pharma & biotech (48)	504	90
Software & computer services (97)	376	140
Speciality & other finance (87)	12	5
Steel & other metals (18)	60	9
Support services (58)	68	13
Telecommunication services (67)	88	31
Tobacco (49)	16	3
Transport (59)	12	1
Utilities – other (73, 78)	32	6
Total	3916	1139

Table 3
Distribution of companies by country for the pooled sample.

Country	With R&D	With software patents
Australia	8	4
Austria	40	7
Belgium	64	9
Canada	28	8
Denmark	112	15
Finland	112	22
France	264	70
Germany	400	102
Greece	8	1
Hungary	8	2
Ireland	16	5
Italy	68	12
Japan	612	255
Luxembourg	8	4
Norway	12	2
South Korea	36	13
Spain	36	8
Sweden	176	39
Switzerland	72	8
Netherlands	88	15
UK	596	106
USA	1152	432
Total	3916	1139

applies for at the EPO. Hence, our dependent variable is of a count data type, that is, it can assume only positive integer values. Given this particular feature, together with the fact that we are facing micro-level data repeating through time, we rely on count panel data models. In particular, we adopt specifications contained in Hausman et al. (1984) and Wooldridge (2005). Whereas the former is usually advocated as the seminal contribution in models of this kind, the latter is a straightforward procedure that allows us to take into account dynamics, without relying on GMM estimation of the parameters of interest (Cameron and Trivedi, 1998).

We also provide a series of robustness checks on the specifications presented here. First, we control for the presence of autocorrelation by relying on a particular specification introduced by Wooldridge (2005). The dependent variable lagged by 1 year is introduced, which considerably lowers the extent of autocorrelation without affecting the consistency of the estimates. Second, we control for departures from the specification herein presented by running alternative estimations. In particular, we account for cases where zero outcomes in our dependent variable originate from a separate decision process or when non-linearities in the innovation process are present. The zero outcome can arise because firms do not patent or firms prefer to keep innovation secret. In the second case, problems arise because the first innovation is likely to be more difficult to achieve than the following ones and, for this reason, the innovation process in non-linear in nature. We run estimates based on logit, tobit and Poisson panel data specifications. Thus, we are able to check whether factors influencing software patenting are robust to different specifications of the econometric model. In all of the cases, we obtain results consistent with those of the negative binomial panel data model presented in Section 5.[15]

We now turn to describing the variables used in the empirical analysis. A more systematic presentation of these variables is provided in Table 4. Our dependent variable measures the number of software patents with priority dates between 2000 and 2003.[16] To control for the non-negligible variance of the quality of different patents, we rely on forward citations as a measure of patent quality (Hall et al., 2001). In particular, we weight the stock of software patents by the amount of forward citations received by each single patent.[17]

Independent variables can be divided into two main groups: structural and control variables. **Structural variables** include all those variables that are objects of the analysis throughout different specifications. These variables are:

1. *R&D spending* ($R\&D_{i,t}$). The amount of R&D spending performed by firm i at time t. The amount has been transformed in purchasing power parity dollars (PPP$)

[15] Results of the robustness checks are available from the authors upon request.

[16] We rely on priority date rather than filing date to control for problems of reverse causality. In particular, by relying on priority date, we are able to shorten the lag between the time when R&D expenditures are made and the time of patent application. Similar results are nevertheless obtained when the filing date is used instead of priority.

[17] Data on forward citations have been drawn from the PATSTAT database.

Table 4
Variables definition.

Variable	Name	Description
R&D spending	R&D	Natural logarithm of the stock of R&D expenditures per employee expressed in millions of PPP$. The Stock has been computed following Griliches and Mairesse (1981) and assuming a pre-sample growth rate of 1% and a depreciation rate of 15%
Employees	Empl	Natural logarithm of the number of employees
Software patents in previous year	Pat_1	Stock of software patents with priority year equal to $t-1$ rounded to the closest integer number. The Stock has been computed following Griliches and Mairesse (1981) and assuming a pre-sample growth rate of 5% and a depreciation rate of 15%. We control for patent quality by multiplying the quantity for the number of forward citations received by the patents filed by the firm in a given year
Software patents in current year	Pat	Stock of software patents with priority year equal to t rounded to the closest integer number. The Stock has been computed following Griliches and Mairesse (1981) and assuming a pre-sample growth rate of 5% and a depreciation rate of 15%. We control for patent quality by multiplying the quantity for the number of forward citations received by the patents filed by the firm in a given year
Sector concentration	Sector_conc	Sector concentration ratio relative to the FTSE sector the firm belongs to. It has been computed by taking the natural logarithm of the ratio between the sales of the four largest firms by FTSE economic group over the total sales in the same FTSE economic group
Strategic rivalry	Strategic	Natural logarithm of the stock of software patents with priority date $t-1$filed by other firms in the same FTSE sector as the observed company
Year dummies	DYear	Variable assuming value 1 in one particular year (ranging between 2000 and 2003) and 0 otherwise
Industry dummies	DInd	Variable assuming value 1 if the firm belongs to one particular FTSE sector and 0 otherwise
Country dummies	DCountry	Variable assuming value 1 if the firm is located in one particular area (US, Japan and the EU) and 0 otherwise
Firm-level patent propensity	TotPat	Stock of total patents with priority year equal to t
Industry-level patent propensity	TotPatInd	Number of total patents applied for based on the SIC code at the 4-digit level.

to allow comparability among different countries and has been expressed as a stock (see Table 4 and Griliches and Mairesse (1981)). According to our theoretical model, this should be a very important variable directly related to the stock of software patents filed at the EPO. In this regard, it can be reasonably asserted that not only the contemporaneous level of R&D spending should be used in the analysis, but also its whole lag structure should be taken into account. Indeed, the nature of R&D as a long-term investment in knowledge capital, whose results are likely to be achieved at any time and not only in the year of the investment, seems to be a reasonable assumption. Nevertheless, the empirical literature on this topic has shown that the estimated coefficient for the sum of past R&D spending is roughly equal to the estimated coefficient for the level of contemporaneous R&D (Hall et al., 1986; Montalvo, 1997).

2. *Employees* ($Empl_{i,t}$). The number of employees for firm i at time t. This variable is a proxy for firm size and influences the number of software patents filed. In fact, larger firms are likely to have more resources to apply for more patents. This is even more likely to happen in the EU, where the average cost of a patent is higher than in other patent systems (Malerba and Montobbio, 2002).
3. *Sector concentration* (Sec_j). Sector concentration has been computed as the total sales of the four largest firms in terms of sales in firm i's main sector of activity (indicated by j) divided by the overall amount of sales of the same sector.
4. *Strategic rivalry* ($Strat_{i,t}$). Strategic rivalry is the stock of software patents filed by firms belonging to the same sector of firm i in priority year $t-1$. This variable proxies for the influence of strategic factors on the software patenting of firms in the sample. Indeed, most of the time, firms apply for software patents only because this is a way to strategically hinder their competitors. Patenting inventions is a way to reduce the value of other firms' innovations and to decrease their average return to R&D, while affecting the firm's own market value (Noel and Schankerman, 2006).
5. *Patent propensity at firm and industry levels ($TotPatInd_{j,t}$ and $TotPat_{i,t}$)*. This measure of strategic rivalry is nevertheless subject to one major problem. In particular, software patents in a sector may be related to specific sector patent propensity rather than to the extent of strategic rivalry. To control for this bias, we include two additional variables in our analysis: (i) the number of total patents applied for based on the SIC code at the 4-digit level ($TotPatInd_{j,t}$) and (ii) the stock of total patents firm i applied for in year t ($TotPat_{i,t}$).
6. *Stock of software patents filed in the previous year (Pat_{t-1})*. Stock of software patents filed by firm i in year $t-1$. This variable takes into account the effect of the number of software patents filed in the prior year on software patenting decisions.

Control variables are all those variables that are implemented to control for factors that are essentially specific to the particular context in which analysis is conducted. These variables are:

Table 5
Negative binomial panel data estimation: fixed (FE) and random effects (RE).

Variables	Overall		Software		Hardware	
	(1) FE	(2) RE	(3) FE	(4) RE	(5) FE	(6) RE
Pat_1	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
R&D	0.206** (0.067)	0.268*** (0.047)	0.465* (0.232)	0.510** (0.198)	0.104 (0.135)	0.101 (0.092)
Empl	0.168** (0.055)	0.251*** (0.037)	0.510* (0.204)	0.481** (0.168)	0.067 (0.124)	0.063 (0.089)
Sector_conc	0.799** (0.293)	0.624*** (0.177)	−2.234 (1.895)		−9.632 (9.109)	
Strategic	0.005 (0.033)	0.247*** (0.025)	−0.442 (1.198)	−0.447 (1.147)	5.951* (2.345)	11.668*** (2.711)
TotPat	0.000*** (0.000)	0.001*** (0.000)	0.002 (0.002)	0.001 (0.001)	0.002*** (0.000)	0.002*** (0.000)
TotPatInd	1.375* (0.697)	0.390 (0.712)	−0.588 (1.841)	0.895 (1.596)	−0.677 (1.512)	4.695* (2.309)
DYear	No	Yes	No	Yes	No	Yes
DCountry	No	Yes	No	Yes	No	Yes
DInd	No	Yes	No	Yes	No	Yes
Cons	−2.298*** (0.603)	−3.670*** (0.431)	−5.018 (3.328)	−2.165* (1.090)	−22.202 (30.713)	92.920*** (22.530)
χ^2	83.963	579.895	33.889	42.751	113.935	424.707
N	1199	2169	152	251	273	353
Log-likelihood	−2492.220	−5440.633	−361.372	−791.530	−746.548	−1485.648

Standard errors are in parentheses.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

1. *Year dummies* ($DYear_t$). These are a set of four dummy variables that take into account the effect of external outcomes on the knowledge production function.[18] In particular, it controls for the institutional context in which the firm operates and for the presence of unexpected shifts, i.e., a structural break in the time series.
2. *Geographical dummies* ($DCountry_i$). We use these variables to disentangle the effects produced by the different patent systems a firm has been dealing with.[19] A firm that is used to operating inside the US has a deep knowledge of both the intrinsic and strategic value an invention is likely to produce once patented. On the contrary, the blurred situation characterising the European patent system should be interpreted as a hindering mechanism.
3. *Industry dummies* ($DInd_i$). To control for residual effects not captured by other explanatory variables and controls, we include in the analysis seven industry dummies aimed at controlling for specific industry-level idiosyncrasies.

[18] In particular, the years from 2000 to 2003 are taken into consideration.
[19] In line with the theoretical considerations of Section 4, we analyse four main categories: American, European, Japanese and the "remaining" patent system.

5. Results

The empirical analysis is based on a negative binomial panel data model that takes into consideration as the main dependent variable the stock of software patents by priority year (see Table 4 for details on variable construction). The estimation is conducted on three related samples. The results are reported in Table 5. In particular, we consider firms contained in the whole sample (columns (1) and (2)), as well as two subsamples comprising firms operating in software (columns (3) and (4)) and hardware industries (columns (5) and (6)). For all three specifications, we provide estimations controlling for both fixed and random effects.[20]

Results show a positive and significant coefficient for R&D spending and firm size. In particular, other things being equal, increasing the amount of R&D spending and

[20] The comparison between fixed effects and random effects is performed for the three different samples: the overall sample, software companies and hardware companies. In all cases, we check for the best specification by running the Hausman test. While for the overall sample, the test results favour fixed effects, in the other two cases, random effects should be preferred. This result is in line with the discussion contained in Hall and Ziedonis (2001), wherein the authors rely on models with random effects by assuming that the impact of permanent differences across firms within a single industry is quite modest.

being a larger firm increases the extent of software patenting at the EPO. It is worth stressing that this result is similar for the overall sample and the sample comprising software firms. We interpret the important role played by firm size in propensity to patent software at the firm-level in terms of the importance of the presence of a legal department handling IPRs (Lerner, 1995). This points to the presence of economies of scale in generating software patents. Indeed, larger firms can exploit patents better, thanks to the rich endowment of financial resources devoted to IPR management departments.

However, the significance of both coefficients disappears when the sample of hardware firms is taken into account. The fact that R&D spending does not significantly contribute to the propensity to patent software of hardware firms looks puzzling. Nevertheless, we interpret the result in terms of the small portion of overall R&D spending specifically devoted to developing new software. Indeed, given the nature of the data at hand, we are not able to differentiate between the share of R&D devoted to software development and the part devoted to other activities. It is also reasonable to assume that the development of hardware infrastructure should play the most important role, given that it constitutes the core business of firms operating in the hardware industry. For this reason, the impact of R&D spending on the propensity to patent software is found to me minimal.

Sector concentration also seems to be an important explaining factor of software patenting only in the overall sample. Industry affiliation also has some bearing on the firms' propensity to patent software in the overall sample. Specifically, being part of the electronics, hardware, media and software sectors explains the likelihood to patent software at the EPO.

Notably, the variable proxying for strategic rivalry is found to be significant and positively related to the number of software patents filed by firms in the overall sample. Nevertheless, the significance completely disappears when firm-level idiosyncratic factors are accounted for via fixed-effects estimation (column (1)). Given that the Hausman test, performed to choose between fixed and random effects, provided results that favour fixed effects, we cannot conclude that strategic patenting is a phenomenon characterising the whole sample of firms. As expected, strategic rivalry is not significant when the sample of software firms is analysed. On the contrary, hardware firms are strongly affected by the amount of software patents filed in the same sector of activity. This latter result is important because it shows that, even when controlling for firm-level heterogeneity via fixed effects estimation, hardware firms experience a hold-up problem by other firms in the same industry. We interpret this result as an indication of the pressure exerted on firm-patenting strategies by rivals belonging to the same sector. This finding likely points to the presence of strategic patenting, that is, specific interest of hardware firms in patenting software for reasons other than the increase of their own inventive capacities. Indeed, the coefficient proxying for the number of software patents filed in the previous year by firms other than the firm under consideration is significant and positive. We interpret this result as a sign of the presence of strong strategic factors inside the hardware sector. Firms in this sector are not likely to patent software to appropriate results of the R&D process; however, at the same time, they are eager to patent if they fear intra-industry competition. This "threat effect" is due to the nature of the software technology that is of a cumulative type. An increase in the amount of software patents accorded to neighbouring firms can hinder future development of both hardware and embedded software, thus leading the company to apply for patents as a defensive strategy. Evidently, this result is in contrast with the common belief that patents are useful appropriability measures for the result of inventive activity (Arrow, 1962; Scotchmer, 1991), but in line with recent progress in the economic literature pointing to the existence of a mix of reasons explaining the upsurge in the number of patent applications, i.e., strategic factors (Shapiro, 2001; Hall and Ziedonis, 2001; Hall, 2004). In particular, the software sector is characterised by a technology that is cumulative, sequential, path dependent and where strong interdependencies among pieces of knowledge are present (Marengo and Pasquali, 2006). For all of these reasons, firms producing software do not deem patent protection as a useful mechanism spurring future inventive streams.

6. Discussion and conclusion

The main goal of this article has been to provide a deeper account of software patenting in the EU. Although this has been a relevant and well-regulated phenomenon since the 1980s in the US, the EU lagged behind for a long time. Nevertheless, in the last decade, the number of software patents filed at the EPO has grown rapidly, despite the fact that Article 52 of the European Patent Convention expressly prohibits software patenting.

To investigate the topic, we present a new database containing information on software patents, the Gauss database. The database has undergone extensive checks to prove its reliability, not only in performing well compared to alternative datasets built according to alternative procedures (e.g., the HTT dataset and a benchmark dataset containing 78 patents) but also its wiki nature suggests that it will improve even more in the near future because of the contribution of several experts.

These data show how software patents are an important phenomenon in the EU as well, given that to date, more than 30,000 software patents have been granted to both European and foreign firms.

In this respect, a large share of patents has been accorded to American and Japanese firms. The fact that the majority of granted patents belong to foreign companies must be due to the higher experience that these firms have acquired dealing with their own patent systems. For example, software has been patentable for a long time in the US, meaning that firms have more expertise in dealing with application procedures and in identifying more valuable inventions to be patented. Together with this finding, other interesting statistics have been presented. Among patents, we have found that for software patents the average length of the granting procedure is larger than for more general patents. Moreover, we have discovered that particular industries apply for the majority of software patents, i.e., electronics and IT hardware. Despite the increasing number of applications, the granting procedure of software patents has been characterised by an increasing strictness in recent years. Finally, the ownership of software patents is highly concentrated, and applicants are mainly American and Japanese firms.

Therefore, the knowledge production function approach has been implemented to identify the major factors affecting the output of the innovation process at the firm-level. The model has been extended to incorporate factors deemed as very important to explain recent patenting strategies, i.e., strategic factors, firm size, industry and geographical controls, and to deal with our specific interests, i.e., the idiosyncrasies of both software and hardware sector. Both the way in which the dataset was built and a robustness check for the database itself ere presented. Moreover, a set of different methods of estimation was put forward and the most suitable one was chosen, after which, the results of the chosen econometric model were presented.

The results of the analysis highlight differences between software and hardware firms in the determinants of patenting software. In particular, strategic behaviour by firms belonging to the hardware sector seems to be very important. This result is in line with empirical evidence highlighted in recent studies. For example, Fosfuri et al. (2008) find that, at the sector level, hardware companies in the US have a higher propensity to patent software than to file software trademarks, whereas software firms show a higher propensity to file software trademarks than to patent. Similarly, Graham et al. (2009) find that the role of patents in helping US technology entrepreneurs compete in the market tends to be much more pronounced among hardware companies than among software firms (with an average of 27 patents compared to 6, respectively). Likewise, Bessen and Hunt (2007) argue that legal changes that occurred in the US are likely to have increased the cost-effectiveness of software patents and, consequently, have encouraged firms in the hardware sector to pursue more aggressive patenting strategies. Several reasons can be put forward to explain the differences between software and hardware firms in the determinants of patenting software.

First, hardware companies face an increasing need to protect software "embedded" in their hardware manufactures. The segment of "embedded" software has risen spectacularly in the last decade, and most of the increase has been taking place in the hardware sector (Gal and Genuchten, 2009), where the increasing complexity of hardware products incorporates software parts aimed at managing the different tasks. Thus, hardware companies have a high propensity to patent software just because they want to protect a core part of their product range (McQueen, 2005).

Coupled with this, hardware firms are characterised by a higher experience in patenting (see Hall and Ziedonis (2001) for a similar result in the context of semiconductor industry). Indeed, hardware companies have been patenting their inventions for over three decades. Therefore, they are more acquainted with the patenting process. In line with this argument, Chabchoub and Niosi (2005) show that, among software firms, those that have a strong component of hardware products tend to obtain more patents overall.

Third, although both hardware and software companies can be included under the definition of cumulative system technology (Merges and Nelson, 1990), software heavily impinges on cumulativeness more than hardware. In such a context, developing a new software often requires the ability to use existing software code, hence an environment where litigation is not a serious threat is usually preferred (Mazzoleni and Nelson, 1998). This fact contributes to explaining the different propensities to patent software between hardware and software companies.

Finally, the rise of open source software as a new market segment in the software industry contributes to the differential in the propensity to patent software between software and hardware companies. In particular, the idea of copyleft, which is at the core of the development of open source software, is at odds with the use of software patents (Lerner and Tirole, 2005). Therefore, the software industry, which is likely to be characterised by an increasing share of companies working in the open source segment, may present a lower propensity to patent software compared to hardware.

We are aware of limitations to the present work. First, we acknowledge the fact that our dataset is likely to include a non-negligible share of computer–implemented inventions. In this sense, it differs significantly from the one proposed by Hall et al. (2007). Therein the authors implement a restrictive identification strategy aimed at detecting only "pure" software patents. We nevertheless think that this distinction is problematic, and thus we decide to rely on a broader definition of software patents, including both "pure" software patents as well as computer-implemented inventions. However, we find support for our choice in the results of the content analysis carried out in Appendix A. Indeed, no significant difference between a set of patents accorded by the USPTO (where no distinction is drawn between "pure" software patents and computer-implemented inventions) and a benchmark dataset containing software patents accorded by the EPO has been found, thus highlighting that the distinction between computer-implemented inventions and "pure" software patents, which has been put forward in the European patent system, is likely to be an artificial one. This result sheds light on the need for the software industry to give careful consideration to the rules of the game, which is necessary to grasp the effects of the introduction of several computer-implemented inventions that can be hardly distinguished from "pure" software patents.

Second, the EU Scoreboard, which we rely on to link firms' characteristics with the number of software patents filed, allows us to take into consideration the behaviour of large firms only. Small and Medium Enterprises (SMEs) together with their software-patenting strategies are totally disregarded. Obviously, providing a proper representation of SMEs at the European level is a very difficult task due to the shortage of reliable data concerning both innovation strategies and patents at the firm-level. Such a limitation will be overcome once more trustworthy datasets on these issues will be available.

Third, there has not been the possibility to separate R&D expenditures used in software production from those used for other purposes. To overcome such a limitation, we refined our analysis in two particular sectors of activity, i.e., software and hardware. Indeed, firms belonging to these sectors devote a large share of their R&D spending to the production of software. Therefore, almost all of the investment in R&D is driven towards software patenting and not towards patenting other types of technology.

Fourth, our definition of strategic rivalry relies on software-patenting activities by rivals of the focal firm. As suggested by McGahan and Silverman (2006), the effect of patenting activities by rivals may incorporate two ambiguous effects. On the one hand, ''market stealing'' effects can be at work: patenting by rivals may actually generate negative effects on a firm's financial-market value. On the other hand, ''spillover'' effects may dominate: inventions by rivals actually trigger greater technological opportunities for the focal firm.[21] We do not have information on firm's financial-market values, and thus we are not able to control for the two effects just described. Nevertheless, we have information on the opposition to software patents contained in our sample. Several authors claim that the likelihood of opposition increases with patent value (Harhoff et al., 2003; Harhoff and Reitzig, 2004). In our sample, we find a low share of opposed patents compared to other technologies (0.3% against 7% for biotech and 9% for financial patents) (Harhoff and Reitzig, 2004; Hall et al., 2009). We interpret this low rate of patent opposition compared to other fields as a sign of the low value of software patents compared to other patents. As a result, we are more inclined to conclude that a ''market stealing'' effect is at work in this case. Unfortunately, the low degree of opposition to patents contained in our sample did not allow us to provide a more meaningful check on the robustness of our variable on strategic rivalry. This should be considered in future research.

Finally, we provide the estimation within a narrow time frame only, 2000–2003. This is mainly due to the lack of data with respect to long-term R&D expenditures for companies located in the EU, which is a common problem in studies of this kind.[22] A second point refers to the difficult task of matching company data with names of patent applicants. Although automatic matching techniques exist, they are still far from being fully reliable and, for this reason, a manual check is often preferred. In our case, the manual check comprised three datasets (Amadeus and Osiris for account data, EPO access and EPO bulletin databases for filed and granted patents at the EPO and the EU R&D investment scoreboard for company-level data) and, therefore, extensive work processing the data manually. We are aware of the fact that a longer time period would allow us to better investigate the robustness of our results, but we faced a trade-off between the construction of a reliable dataset for a four-year time period and setting up a dataset covering a longer time frame at the expense of trustworthiness.

Future work should try to address all the points mentioned above to extend our results. We believe that the insights gained from this study will serve as a guide and foundation for future work aimed at investigating the phenomenon of software patenting in the EU.

Acknowledgements

The author gratefully acknowledges comments from Riccardo Leoncini and Giuditta De Prato to the first draft of the present paper. The same was presented for discussion at the 11th International Schumpeter Society Conference and UNU-MERIT Seminar, held respectively at Nice/Sophia Antipolis, 21–24 June 2006 and UNU- MERIT, Maastricht, 30 June 2006 where I received many helpful comments and suggestions from a number of colleagues. Among the others, I would like to mention Rishab Ghosh, Robin Cowan, Maureen McKelvey and Enrico Perotti. Moreover, I would like to thank Gauss project for providing me with part of the data necessary for the analysis and Gianluca Tarasconi for the extractions made on PATSTAT database. I am responsible for any other mistakes and omissions.

Appendix A. Robustness checks on Gauss database

To check for the ability of the database to minimise errors of the first and second types, we rely on two methods. First, we build a control dataset, following the procedure contained in Hall et al. (2007) (henceforth HTT database),[23] and provide comparison statistics for the two datasets (Gauss and HTT database).[24] Table 6 provides a statistical comparison of the Gauss and HTT databases. In particular, we check whether the two datasets differ with respect to any of the following characteristics: distribution by country of priority, distribution by designated country, distribution by IPC section/class, annual growth rate of patents published and annual growth rate of patents filed. The results are quite clear, and with the exception of IPC section/class, no significant difference is found between the two datasets.

Second, we propose a comparison between the Gauss and HTT databases on the grounds of their ability to spot type I and type II errors. In particular, we compare the accuracy by which the two datasets are able to: (i) detect true software patents in a dataset containing only software patents (type II error) and (ii) not spot patents in a dataset containing false software patents (type I error). To do so, we construct a dataset containing 59 software patents and 19 non-software patents and checked how the two

[21] We are grateful to one referee for pointing this out.

[22] Unlike in the US, where data on R&D are disclosed by companies due mainly to fiscal reasons, European companies are less likely to do that in the absence of such a provision. Some European companies located in the EU are nevertheless disclosing this information, and Hall et al. (2007) make use of these data and couple them with other data sources to fill in the blanks. Unfortunately, Amadeus only contains data on non-EU companies with subsidiaries located in Europe, but it lacks accounting data for corporations without such subsidiaries and for subsidiaries, belonging to EU or non-EU companies, located elsewhere. Given our interest in understanding different strategies put forward by the EU and foreign companies relative to software patents, we had to rely on other data sources to collect R&D data, thus constraining the possibility of extending the time frame.

[23] Refer to Section 2 for details on the procedure.

[24] We are very grateful to an anonymous referee for suggesting to us such a strategy.

Table 6
Comparison of Gauss dataset and controlling sample.

	Typology of difference test	P-value
Country of priority	Sign test	0.1
Designated countries	Sign test	1
Unique IPC section/class	Sign test	0
Annual growth rate by publication year	Paired *t*-test	0.66
Annual growth rate by filing year	Paired *t*-test	0.6

datasets (Gauss and HTT database) perform with respect to this benchmark case (henceforth benchmark dataset). In this way, the two distinct groups that constitute our benchmark dataset, i.e., 59 patents classified as true software patents and 19 patents that do not protect software technology, result from the composition of different sources: (i) 23 patents analysed in a study by Bergstra and Klint (2007), of which 14 are classified as software patents; (ii) 20 software patents that are EU equivalents of software patents granted by the USPTO; and (iii) 35 patents analysed by supporters of the Foundation for a Free Information Infrastructure (FFII), who participated in an initiative aimed at detecting software patents accorded by the EPO.[25]

Evidently, the credibility of the exercise depends on how reliable the benchmark dataset is in providing a clear distinction between software and non-software patents. To this end, we propose a methodology to inspect the robustness of our benchmark dataset, which is based upon content analysis. Our starting point is a dataset comprising 50 software patents granted by the USPTO and analysed by Campbell-Kelly and Valduriez (2005) (CAMPBELL-KELLY). The authors conduct a detailed technical examination of such patents after having read the description for every single item. Our approach consists of applying content analysis on the text of the patent title and abstract to extract relevant concepts capable of statistical validation. In this way, we have been able to extract not only the presence and frequency of relevant concepts but also measures of how the identified concepts are related to each other within the documents.[26] The same procedure has been applied on two other datasets, which constitute the sources of our benchmarking dataset, in particular, to the 14 software patents analysed by Bergstra and Klint (2007) (BERGSTRA) and the 35 patents identified by the supporters of the FFII (FFII). Finally, the extraction of relevant concepts via content analysis has been carried out for our databases, containing 59 software patents (TRUE) and 19 non-software patents (FALSE).

A comparison among the datasets is then performed to detect significant differences. In particular, we compute two relevant measures based on the identified concepts. The first measure gives an account of whether there is a significant difference between two datasets in terms of the number of the most relevant concepts that have been extracted. This measure is computed as the difference in the fraction of concepts over the total that are in common within the 50th percentile for each dataset (we briefly refer to it as concept number). The second measure provides an account of the extent of the difference between the concepts that are in common between the two datasets. This measure is provided by the Wilcoxon signed rank sum test for matched-pairs of concepts (concept intensity). Table 7 contains the results of the exercise. By comparing the CAMPBELL dataset with two of the sources of our benchmark dataset (BERGSTRA and FFII) and with the benchmark dataset itself (TRUE), we obtain no relevant differences in terms of both concept number and concept intensity. Two additional comparisons are then provided: (i) CAMPBELL vs. FALSE and (ii) TRUE vs. FALSE. In both cases, as expected, sensible differences are found in terms of both concept number and concept intensity.[27]

The results of the content analysis not only point out the reliability of our benchmark dataset but also highlight that the distinction between computer-implemented inventions and "pure" software patents, which has been put forward in the European patent system, is likely to be an artificial one. Our results show that no significant differences are present between a set of 50 software patents granted by the USPTO (where patents on software are legitimately granted) and a set of patents that may contain what the European patent system is likely to define as computer-implemented inventions.

As a final step, the comparison between the Gauss and HTT databases is performed taking our TRUE and FALSE software as a reliable benchmark dataset. In both cases, a consistent robustness of the Gauss database has always been found. The main results contained in Table 8 clearly point to a better performance of the Gauss database compared to the HTT one. Indeed, Gauss is able to detect more than the 70% of software patents contained in the benchmark dataset, while the HTT database only identifies 10% of them. As for type I error, the HTT database performs better than the Gauss database, but the difference is minimal (5% vs. 10%).

The results thus obtained point out how the method originally proposed by HTT for the US (Hall and MacGarvie, 2010) and translated to the European system to detect software patents (Hall et al., 2007) is mainly aimed at identifying "pure" software patents. In this sense, our database differs significantly from the one they proposed. In our case, we rely on a broader definition of software patents,

[25] For a list and thorough description of the patents analysed by the FFII refer to http://eupat.ffii.org/patents/samples/index.en.html.

[26] The extraction of the concepts as well as the generation of statistical measures has been carried out through the Leximancer software program. In content analysis, concepts are collections of words that travel together throughout the text. For further details about the procedure of concept identification and other relevant issues, please refer to Smith and Humphreys (2006).

[27] We also carried out additional comparisons between the mentioned datasets. In particular, we were interested in controlling for relevant differences that may arise due to the different institutional environments of the US and EU patent systems. It is the case that both time trends and differences in IPC classifications in the two patent systems may induce differences in the language structure in which the patents are drafted. We control for these differences by conditioning on the year of filing and unique IPC section class. No relevant differences are found either within or between datasets.

Table 7
Reliability check for the benchmark dataset based on content analysis.

	Concept number	Concept intensity
Campbell vs. FFII	0.04	−0.05
Campbell vs. Bergstra	0.14	−1.99
Campbell vs. True	0.09	−1.36
Campbell vs. False	0.4	−2.44**
True vs. False	0.34	−2.2**

Concept number is computed as the difference in the fraction of concepts over the total that are in common within the 50th percentile. Concept intensity is computed as the Wilcoxon signed rank sum test for matched-pairs of concepts.

Table 8
A comparison of control sample and Gauss in minimising errors of first and second type.

	Type II error*	Type I error**
HTT	10%	5%
Gauss	73%	10%
N	59	19

* Percentage of true software patents detected by the method.
** Percentage of patents detected as software that are not.

including both "pure" software patents and computer–implemented inventions.

References

Allison, J., Lemley, M., 2000. Who's patenting what? An empirical exploration of patent prosecution. Vanderbilt Law Review 53, 2099–2148.

Allison, J., Tiller, E., 2003. Internet business method patents. In: Cohen, W., Merrill, S. (Eds.), Patents in the Knowledge-based Economy. National Academies Press, pp. 259–284.

Arrow, K., 1962. Economic welfare and the allocation of resources for invention. In: Nelson, R. (Ed.), The Rate and Direction of Inventive Activity: Economic and Social Factors. NBER. Princeton University Press.

Bergstra, J.A., Klint, P., 2007. About trivial software patents: the IsNot case. Science of Computer Programming 64 (3), 264–285.

Bessen, J., Hunt, R., 2007. An empirical look at software patents. Journal of Economics & Management Strategy 16, 157–189.

Bessen, J., Maskin, E., 2009. Sequential innovation, patents, and imitation. RAND Journal of Economics 40 (4), 611–635.

Boldrin, M., Levine, D.K., 2002. The case against intellectual property. American Economic Review 92 (2), 209–212.

Borrás, S., Kahin, B., 2009. Patent reform in Europe and the US. Science and Public Policy 36 (8), 631–640.

Cameron, C., Trivedi, P., 1998. Regression Analysis of Count Data. Cambridge University Press.

Campbell-Kelly, M., Valduriez, P., 2005. A technical critique of fifty software patents. Marquette Intellectual Property Law Review 9 (2), 249–305.

Chabchoub, N., Niosi, J., 2005. Explaining the propensity to patent computer software. Technovation 25 (9), 971–978.

Cohen, W.M., Nelson, R.R., Walsh, J.P. 2000. Protecting their Intellectual Assets: Appropriability Conditions and Why US Manufacturing Firms Patent (or not). NBER Working Papers 7552. National Bureau of Economic Research, Inc.

Fosfuri, A., Giarratana, M., Luzzi, A., 2008. The penguin has entered the building: the commercialization of open source software products. Organization Science 19 (2), 292–305.

Freedman, C., 2000. Software and computer-related business-method inventions: must Europe adopt American patent culture? International Journal of Law and Information Technology 8 (3), 285–309.

Frietsch, R. 2004. Combining databases for forecasting purposes. patents and further economic data. Paper presented at the WIPO-OECD Workshop on the Use of Patent Statistics, 11th/12th October 2004, Geneva.

Gal, R., Genuchten, M., 2009. Release the embedded software: the electronics industry in transition. International Journal of Technology Management 12 (1), 33–44.

Gilbert, R., Newbery, D., 1982. Preemptive patenting and the persistence of monopoly. The American Economic Review 72 (3), 514–526.

Graham, S., Merges, R., Samuelson, P., Sichelman, T., 2009. High technology entrepreneurs and the patent system: results of the 2008 Berkeley Patent Survey. Berkeley Technology Law Journal 24, 1255–1328.

Graham, S., Mowery, D., 2003. Intellectual property protection in the US software industry. In: Cohen, W., Merrill, S. (Eds.), Patents in the Knowledge-based Economy. National Academies Press, pp. 219–258.

Graham, S., Mowery, D., 2004. Submarines in software? Continuations in US software patenting in the 1980s and 1990s. Economics of Innovation and New Technology 13 (5), 443–456.

Graham, S.J.H., Hall, B.H., Harhoff, D., Mowery, D.C. 2002. Post-issue Patent "Quality Control": A Comparative Study of US Patent Re-examinations and European Patent Pppositions. Technical Report 8807. National Bureau of Economic Research, Inc.

Griliches, Z., 1990. Patent statistics as economic indicators: a survey. Journal of Economic Literature 28 (4), 1661–1707.

Griliches, Z., Mairesse, J. 1981. Productivity and R&D at the Firm Level. NBER Working Papers 0826. National Bureau of Economic Research, Inc.

Hall, B., 2004. Exploring the patent explosion. Journal of Technology Transfer 30 (1-2), 35–48.

Hall, B., Griliches, Z., Hausman, J., 1986. Patents and R&D: is there a lag? International Economic Review 27 (2), 265–283.

Hall, B., MacGarvie, M., 2010. The private value of software patents. Research Policy 39 (7), 994–1009.

Hall, B., Thoma, G., Torrisi, S., 2009. Financial patenting in Europe. European Management Review 6 (1), 45–63.

Hall, B., Ziedonis, R.H., 2001. The patent paradox revisited: an empirical study of patenting in the US semiconductor industry, 1979–1995. RAND Journal of Economics 32 (1), 101–128.

Hall, B.H., Jaffe, A.B., Trajtenberg, M., 2001. The NEBR Patent Citation Data File: Lessons, Insights and Methodological Tools. NBER Working Papers 8498. National Bureau of Economic Research, Inc.

Hall, B.H., Thoma, G., Torrisi, S., 2007. The Market Value of Patents and R&D: Evidence from European Firms. NBER Working Papers 13426. National Bureau of Economic Research, Inc.

Harhoff, D., Reitzig, M., 2004. Determinants of opposition against EPO patent grants – the case of biotechnology and pharmaceuticals. International Journal of Industrial Organization 22 (4), 443–480.

Harhoff, D., Scherer, F., Vopel, K., 2003. Citations, family size, opposition and the value of patent rights. Research Policy 32 (8), 1343–1363.

Hausman, J., Hall, B., Griliches, Z., 1984. Econometric models for count data with an application to patents – R&D relationship. Econometrica 52 (4), 909–938.

Lerner, J., 1995. Patenting in the shadow of competitors. Journal of Law & Economics 38 (2), 463–495.

Lerner, J., Tirole, J., 2005. The economics of technology sharing: open source and beyond. The Journal of Economic Perspectives 19 (2), 99–120.

Malerba, F., Montobbio, F., 2002. L'europa dell'alta tecnologia: lo scenario internazionale in alcuni settori chiave. In: Tendenze dell'Industria Italiana. Centro Studi Confindustria, pp. 63–78.

Mansfield, E., 1986. Patents and innovation: an empirical study. Management Science 32 (2), 173–181.

Marengo, L., Pasquali, C., 2006. Non Rivalry and Complementarity in Computer Software. DIME Working Papers 11. DIME Working Papers.

Mazzoleni, R., Nelson, R., 1998. The benefits and costs of strong patent protection: a contribution to the current debate. Research Policy 27, 273–284.

McGahan, A., Silverman, B., 2006. Profiting from technological innovation by others: the effect of competitor patenting on firm value. Research Policy 35 (8), 1222–1242.

McQueen, D.H., 2005. Growth of software related patents in different countries. Technovation 25 (6), 657–671.

Merges, R., Nelson, R., 1990. On the complex economics of patent scope. Columbia Law Review 90, 839–916.

Montalvo, J., 1997. Gmm estimation of count-panel-data models with fixed effects and predetermined instruments. Journal of Businness & Economic Statistics 15 (1), 82–89.

Nelson, R., Winter, S., 1982. An Evolutionary Theory of Economic Change. Belknap Press.

Noel, M.D., Schankerman, M., 2006. Strategic Patenting and Software Innovation. CEPR Discussion Papers 5701. CEPR Discussion Papers.

Pakes, A., Griliches, Z., 1984. Patents and R&D at the firm level: a first look. In: Griliches, Z. (Ed.), R&D, Patents and Productivity. University of Chicago Press, pp. 55–72.

Peeters, C., Van Pottelsberghe de la Potterie, B., 2006. Innovation strategy and the patenting behavior of firms. Journal of Evolutionary Economics 16 (1), 109–135.

Sakakibara, M., Branstetter, L., 2001. Do stronger patents induce more innovation? Evidence from the 1988 Japanese patent law reforms. RAND Journal of Economics 32 (1), 77–100. Spring.

Schumpeter, J., 1942. Capitalism, Socialism and Democracy. McGraw Hill.

Scotchmer, S., 1991. Standing on the shoulders of giants: cumulative research and the patent law. The Journal of Economic Perspectives 5 (1), 29–41.

Shapiro, C., 2001. Navigating the patent thicket: cross licenses, patent pools and standard setting. In: Jaffe, A., Lerner, J., Stern, S. (Eds.), Innovation Policy and the Economy. MIT Press, Cambridge, pp. 119–150.

Smith, A., Humphreys, M., 2006. Evaluation of unsupervised semantic mapping of natural language with Leximancer concept mapping. Behavior Research Methods 38 (2), 262–279.

Wagner, S., 2008. Business method patents in Europe and their strategic useevidence from franking device manufacturers. Economics of Innovation and New Technology 17 (3), 173–194.

Wooldridge, J., 2005. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. Journal of Applied Econometrics 20, 39–54.