Sources and characteristics of software patents in the European Union: some empirical considerations[*]

Francesco Rentocchini[†]

September, 2009

Abstract

This work provides a deeper account of an increasing phenomenon characterising the EU in recent years, namely software patenting. Although the European patent convention expressively prohibits it, this has not been a major problem for firms and inventors that have seen their applications being granted as well. First of all, a new reliable database on European software patents is presented and its quality carefully checked. According to that, more than 30 000 software patents have been granted so far to both European and Non-European companies. Second, we investigate the relevant factors explaining firm-level software patenting at the EPO. To this respect, a large part of them has been found to belong to American and Japanese firms. Moreover, software patents are found to be characterised by an higher average length of the granting procedure and by the fact that firms belonging to the software sector 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 of the literature relative to strategic patenting.

Keywords: Software Patenting; Intellectual Property Rights; IPRs, Patent; Innovation; Empirical
JEL Classification: O34; D23; L86

1 Introduction

During the last ten years the number of filed and granted patents at the main three Patent offices - i.e. 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 center of a debate at the European level.

Since a long time, the economic literature has recognized the importance of the patent system in shaping and directing the rate of appropriation of the innovative effort of firms (Arrow, 1962). Besides '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 started to stress that, depending on appropriability conditions of industries, patents might be, or not, a useful institutional mechanism in order to promote the variety of technological solutions and the selection by market forces via competition (Merges and Nelson, 1990).

Hence, on the one side, empirical literature has shown how patents might not be suitable appropriability mechanisms in a high number of sectors (Cohen, Nelson, and Walsh, 2000), but, on the other side, we have witnessed an explosion in the number of patents filed in recent years. A set of research questions arise from these seemingly contradictory patterns of diffusion. Among the others, 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) stressed how in industries where the innovation process relies mainly on improvements made by others, namely cumulative system technologies, it is more likely to find strategic patenting behaviors such as cross-licensing, blocking rivals or extracting licensing revenues.

Another example of strategic patenting refers to submarine patents that is a patent issued after an extraordinarily long period of pre-grant review inside a patent office thus allowing the applicant to reveal it only after that the new patent covered widely-adopted technologies. In this way the owner of the original patent could pursue infringement actions, or seek injunction, against other technology adopters. Graham and Mowery (2004) have investigated the role of procedural revisions of patent applications (called 'continuations')

in software patents in the United States during 1987–1999. The authors individuate a rapid growth in the use of continuations between 1987 and 1995 in software patenting.[1]

The main contribution of the present article is to give an account of this phenomenon at the EPO in a cumulative system technology such as software. To do that, first of all a reliable data-set containing software patents is to be be framed. The main problem is that there's no one to one correspondence between software technology and IPC class or sub-class, thus calling for a tailored methodology in order to identify software patents among general ones. Before going on with this, we provide a literature background dealing with the issue of software patents for both US and EU revealing the most striking results concerning strategic patenting in this area. While for the US patent system several works have already been presented, the EU has been mainly disregarded with the exception of a couple of works (see section 2). Section 3 presents the theoretical model explaining factors affecting software patenting at the firm level which is going to be tested in the empirical part. 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 2000-03 is put forward in section 4 which links the number of software patents filed at the EPO with the 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. Econometric analysis is then performed to find out the most relevant factors affecting software patenting decisions for firms belonging to different industries (section 4). Finally, results are presented and discussed (sections 5 and 6).

2 Literature background

Studies dealing with software patents refer primarily to the US patent system where software is a patentable subject matter since 1981.[2] Allison and Lemley (2000) and Allison and Tiller (2003) are 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 in order to test the general belief that internet business method patents have not been properly searched

---

[1]Year 1995 marked a sharp change in this growth because of changes in the US patent law aimed at hitting directly this kind of practice. Prior to that, continuation applications were used more intensively by large packaged-software firm and patents subject to continuation have been found to be characterised by longer examination delays and to be more valuable.

[2]see court decision *Diamond v. Diehr*, 450 U.S. 175 (1981).

for relevant prior art, meaning that they are likely to be of poor quality.[3] The main conclusion of the study points out that there is little support for the main criticism given that internet-related patents are found to be characterised by the same amount of prior art references as more general patents.

To our knowledge, one of the first general methodology to identify software patents has been proposed by Graham and Mowery (2003) who examined all of the patents falling into identified IPC classes defining them software patents as such.[4] Main findings can be 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; (iv) a decreasing propensity to copyright software is found, thus pointing to a substitution effect.

An alternative methodology to identify software patents has been put forward by Bessen and Hunt (2007), who developed a search algorithm, based on a fixed number of keywords, in order to find 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 explain the rising propensity to patent software in the sample. They find evidence that capital intensive firms tend to patent more because of 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 occurred in the 80s when cost-effectiveness of software patents has been reduced considerably. [5]

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 which

---

[3]The rationale is that software development has been taking place for a long period before software was deemed patentable, so that when USPTO begin to issue patents of this kind they did have neither examiners with the relevant training nor adequate database with software prior art.

[4]These IPC classes were individuated by analysing overall patenting by the six largest US producers of personal computer software based on their 1997 calendar revenues.

[5]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.

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 different methodologies put forward to identify software patents have been fruitfully put together by Hall and MacGarvie (2006). First, they identified all the US patent class-subclasses combinations where 15 software firms patent. After that, in order to minimize errors of both first and second type they merge their database with another one built following the method of Graham and Mowery (2003) and intersect the resulting database with the one obtained using Bessen and Hunt (2007) method. Two main results are worth noticing: (i) the expansion of patentability affected negatively 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; (ii) software patents turn out to be valued more by the market than ordinary patents. For hardware producers this is likely to reflect the strategic value of software patents rather than their technological value. Contrary to that, software patents are found to be technologically valuable for software firms.

An interest in the study of software patenting in the European patent system has been increasing during recent years. Difficulties in data collection and the absence of a clear legislation relative to software patents[6] have constrained the analysis to rely solely on surveys rather than adopting more general approaches like in the US case.[7] Nevertheless, recent attempts have tried 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, an overall increase of 60% of software patent applications is

---

[6]Article 52 of the European Patent Convention expressively 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 as an invention and, for this reason, susceptible of patenting. To 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, EPO regularly provides a plurality of actors with patents on software; this being possible given the presence of numerous interpretations on the definition of software.

[7]Providing a literature background of studies on software patents adopting a survey approach is beyond the scope of our work. These articles collect data mainly through the use of questionnaires and interviews. The main drawbacks of such an approach are that the operation is normally highly time consuming and usually confined to a delimited geographical area (see Olsson and McQueen (2000) and Blind and Edler (2003)).

found in the last 12 years. In general, the work by McQueen (2005), despite 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 database thus constructed 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, Thoma, and Torrisi (2007) instead present a more elaborated method consisting of three main steps. First, they conduct a keyword search on the EPO data-set using the same algorithm set forward by Bessen and Hunt (2007) for the US patent system. Second, they analyse the IPC classes of the patent portfolios of the world's 15 largest software firms expanding it in order to give an account of European firms as well. Third, they accept a restrictive definition of software patent as one falling in the intersection of the two sets defined by both the keyword and IPC methods. The authors use the database thus constructed to explore whether software patents in Europe 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 while, if weighted by their intrinsic quality level, the difference disappears. Thus, the authors conclude that the value of software patents in Europe depends more on their number rather than the quality of the invention they protect.

Although the work by Hall, Thoma, and Torrisi (2007) introduce many important insights into the topic of software patenting at the EPO, it presents some limitations that our work aims at bridging. First of all, search methodology is surely good but not as accurate as it would be as reading a patents one by one. In particular, the authors do not provide any kind of comparison of the database of software patents thus constructed with a trustworthy dataset containing true software patents. In our paper, we try to overcome such a shortcoming by comparing our dataset of software patents with a sample of 78 patents containing both software and non-software patents. Second, Hall, Thoma, and Torrisi (2007) exclude from the analysis non-European firms and their sample is strongly centered 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, they are more concerned with the issue of different valuation of software patents by firms in the European and US patent systems, while we are more interested in the presence of strategic patenting in the form of 'threat' effect by other firms.

6

3 The Model

3.1 Theoretical Background

The study on the effect of research and development spending, and other factors, on the number of patents filed has relied mainly on the Knowledge Production Function (henceforth KPF) approach. The main idea is that the R&D expenditure 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).

In figure 1 we present the classical rational for the KPF approach, augmented of a set of factors which we deem as very important in our analysis[8]. The main factors contained in the figure are:

- K represents the dynamics of the stock of knowledge of firm i at time t;

- a_i 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;

- $R\&D$ represents the amount of R&D expenditure of firm i at time t;

- Pat is the number of patents filed by firm i at time t

- u and U are the error terms for the measurement of R&D expenditure and patent count respectively;

- $Trend$ is a factor controlling for the presence of specific trend patterns as time t passes.

So 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, a good index of the output of this process is needed. At this regard, the economic literature has relied on the number of patents filed by a single firm in a fixed point in time. Even if this index has relevant drawbacks, among which the fact that not all new innovations are patented and that patents differ in their economic impact, it has been widely adopted in applied studies[9] (Griliches, 1990).

---

[8]In fig. 1 triangles represent unobservable quantities, circles are observable quantities, while squares are disturbance terms.

[9]This happened because patent statistics are easily accessible, which is even more true now after that worldwide patent offices (USPTO, JPO, EPO and WIPO) have computerized their data and have granted the public access through the web.
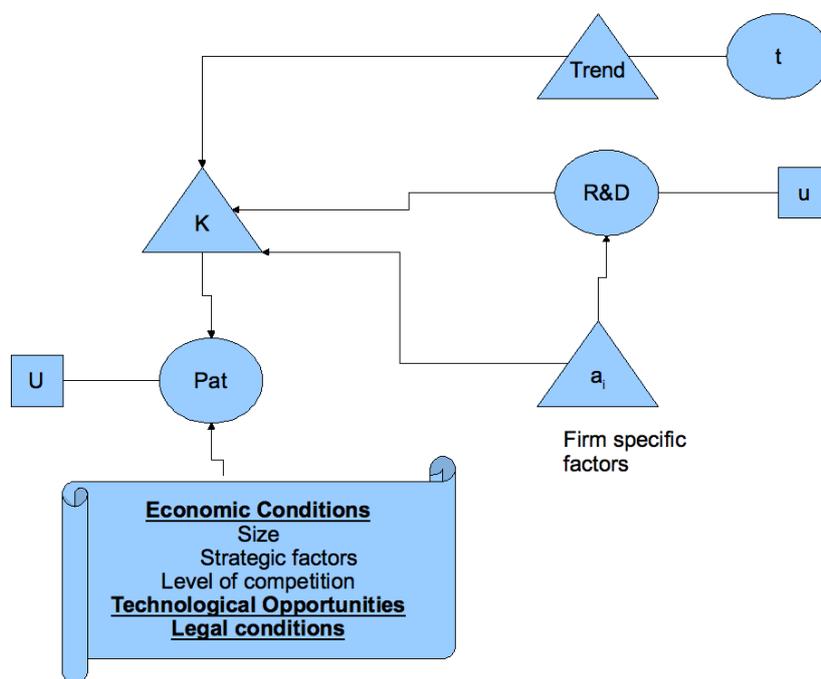
Figure 1: Knowledge Production Function (KPF) Theoretical Model

The present model is quite 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. 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 by three main factors:

- *Size.* Size influences the innovation process of the firm according 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, Nelson, and Walsh, 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 Pot-terie, 2006). Fourth, large firms are more likely to be endowed with a legal department which handles IPRs matters (Lerner, 1995).

- *Level of competition.* 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 in R&D caused by their dominant position in the market (Arrow, 1962). Second, 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).

- *Strategic factors.* These are factors explaining recent trends in patent-ing strategies at the firm level. In fact, while 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[10], recent contributions assert that the high number of patents filed by companies, in particular larger ones, are instead a strategy aimed at hindering com-petition and increasing their monopolistic position (Hall and Ziedonis, 2001). This happens mainly in 'cumulative system' technologies, that is technologies where 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 activi-ties. These factors are of different nature and content. Among them, it is worth reminding cross- licensing, threat effects, patent 'thickets', and so on.

The second set of factors, i.e. **legal conditions**, can be proxied by geographical factors. Indeed, different opportunities may arise from being located in different regions having different legislations. Among them, four main macro-areas have been identified: the European Union, the United States, Japan and other countries[11]

---

[10]An inventor, deprived of the exclusive right to exploit its invention for a definite period of time, would not had even started the inventive activity was he aware of it. This is obviously related to the public nature of knowledge (Arrow, 1962).

[11]While European and American legal regimes have been extensively discussed by the literature (Graham, Hall, Harhoff, and Mowery, 2002), Japan and other countries are worth being mentioned. In 1988 Japan has changed its patent system from a single-claim to a multiple-claim one. This reform has induced two main effects on the Japanese patent system: (i) a decrease in the number of patent applications, (ii) overlapping patent claims have been extensively used to defend strategically acquired inventions. This major change in the Japanese patent system has influenced patenting by both Japanese and American firms (Sakakibara and Branstetter, 2001). On the contrary, with other countries we indicate fast growing countries (such as, for instance, India and China) the patent system of which

The third set of factors is constituted by **technological opportunities**. These are proxied through the industrial sector of activity of the firm. 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). To our concern, technological opportunity is of particular interest. In fact, we want to investigate the different behaviour taken by firms belonging to two separated sectors, namely hardware and software producers. It has been showed that, during the last 10 years, main patenters 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 instruments account for more than the 60% of software patents accorded at the USPTO. While software publishers and firms from other software industries contribute only for the 7% to the overall share of software patents (Bessen and Hunt, 2007). Hence, if firms not belonging 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.

3.2 Method of Estimation

On the grounds of the previous section, the main focus of our analysis will be now the number of patents a firm applies for. Before continuing and going into the details of how the database was built up, we would like to spend some time on the peculiarity of the different estimation methods that have been implemented.

The main object of the analysis is to explain which factors influence the number of software patents a firm 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 Hausman, Hall, and Griliches (1984) and Wooldridge (2005) specifications. While the former is usually advocated as the seminal contribution in these kind of models, the latter is a straightforward procedure which allows to take into account dynamics without using GMM estimation of the parameters of interest[12].

The most used type of count-panel-data model is the Poisson panel data model. The estimation of the model can be done via both fixed and random effects. These are two methods common to linear panel data models but

---

is not as trustworthy as the one of more developed countries. Indeed, the enforcement of international regulations concerning IPRs is not always completely pursued in countries such as China and India (Traphagan, 1998).

[12]Models implementing Generalized Method of Moments (GMM) estimation to analyze filed patents in a Count-Panel-Data setting are reviewed by Cameron and Trivedi (1998).

the procedure differs from the former given its non-linear nature. Indeed, while in the linear case the fixed effect is cancelled through a first difference procedure, this is not the case in the non-linear case. Implementing a first difference will not allow us to cancel out the unobserved heterogeneity term. Moreover, while in the linear case the estimation proceeds via OLS or GLS, in the non-linear one maximum likelihood estimation is the correct procedure to adopt. In particular, to drop out the heterogeneity term, we rely on conditional maximum likelihood estimation (Andersen, 1970) which allows the computation of the log-likelihood. The latter, once maximised, yields both consistent and efficient estimates of the parameter of interest (Hausman, Hall, and Griliches, 1984).

Contrary to the fixed effect procedure, random effects implies a distribution of the heterogeneity term as in the linear case. The difference pertains to the kind of distribution assumed, that in this case is of gamma type rather than normal[13].

Even if very appealing, the Poisson panel data model present three relevant drawbacks:

1. Conditional mean and conditional variance cannot vary independently. Formally:

2. The main assumption behind the model is that $p_{i,t} \sim Poisson(\mu_{i,t})$. This assumption does not hold in the case zero outcome originates from a separate decision process or when non-linearities in the innovation process are present. In the first case, problems can arise because firms prefer a strategy of secrecy. The zero outcome can be due to the fact that either 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.

3. In case autocorrelation is present, then, it is impossible to introduce dynamics into the model in order to reduce it. Indeed, in case a lagged dependent variable is introduced into the model then the method of estimation, i.e. conditional maximum likelihood, does not provide consistent estimates of the parameters of interest anymore.

All of the mentioned drawbacks can be overcome through the implementation of several methodologies:

1. The adoption of a Negative Binomial panel data model, introduced by Hausman, Hall, and Griliches (1984), allows conditional variance to vary with respect to conditional mean. Nevertheless, we run regressions

---

[13]For a formal treatment of the two different method of estimation see Cameron and Trivedi (1998).

11

using negative binomial specification and this yield results similar to the poisson specification. Hence, given the possibility of incorporating dynamics into the model only through the Poisson specification we decided to rely on the latter one.

2. The different processes yielding a zero outcome can be tested adopting a logit panel data model. In this way, we are able to check whether factors influencing software patenting are robust to different specifications of the econometric model.

3. Introducing dynamics into the model is done via a particular specification of the Poisson panel data model introduced by Wooldridge (2005). There, one lag of the dependent variable is implemented which considerably lowers the extent of autocorrelation without affecting the consistency of the estimates.

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 where software patents can be comfortably found, thus it is important to produce a reliable dataset able to minimise errors both of first and second type. Type I error refers to the error committed when many false negatives are detected, that is when a patent that should have 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 software patent.

Contrary to studies proposed so 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. The first step of the current section (4.1) is to propose a general description of the Gauss database, with some relevant statistics concerning European software patents. Its reliability has been checked by several means, among which 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 Gauss database.

After that, in section 4.2, we investigate the construction of our sample concerning both European and foreign firms patenting software at the EPO. In order to do that, 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, underlining certain

characteristics that call for the use of defined econometric techniques and providing descriptive statistics for the sample itself.

Our last step (see section 4.3) will be 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 Gauss database. This database has been created from different sources. First of all, 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 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 about 150 words 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 gone public since 1999 and in a wiki form. By doing this, not only the maintainers of the dataset are able to modify the patents contained therein[14], but also all registered users have been able to keep it up to date and helped identifying software patents and removing non-software ones.[15]

In order to check the ability of the database to minimise errors of first and second type, we relied on two different methods. First of all, we build a control dataset following the procedure contained in Hall, Thoma, and Torrisi (2007)[16](henceforth HTT database) and provide comparison statistics for the two datasets (Gauss and HTT database).[17] Second, we rely on the statement by Allison and Lemley (2000) according to which the most reliable way to individuate software patents is actually to read the description for every single item and classifying it accordingly (Allison and Lemley, 2000). Although this statement can be criticized following the argument that patent drafters sometimes obscure the use of software by making the language in the application sound like special purpose industrial equipment, we believe that a person expert in the field will be nevertheless able to identify software patent from the written description. Indeed, patent drafters have several strong

---

[14]Gauss is mainly maintained by an informal group composed of six persons whose background is rather diversified. Indeed, the group is composed by three financial analysts, two founders of start-up software companies as well as one physics researcher.

[15]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 as a successful experiment aimed at bringing this form of collaborative organisation to the patent system.

[16]Refer to section 2 for details on the procedure.

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

reasons to make clear the real nature of the invention in the description. First, if they do not do so, they risk finding their own product may not be covered. In addition, without such clear description, the patent may not satisfy the inventive step requirement.[18]

Thus, we constructed a dataset containing 59 software patents and 19 non software patents and checked how the two datasets (Gauss and HTT database) perform with respect to this benchmark case. In both cases a consistent robustness of the Gauss database has always been found. As for the first step, table 1 provide a statistical comparison of Gauss database and HTT database. In particular, we checked 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. Results are quite clear, with the exception of IPC section/class, no significant difference is found between the two datasets.

Table 1: 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

As for the second point, a set of 78 patents has been built as composition of two main 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 defined as such because they are EU equivalents of software patents granted by the USPTO (Campbell Kelly and Valduriez, 2005); (iii) 35 patents analysed by supporters of the FFII who participated to an initiative aimed at detecting software patents accorded by the EPO. In this way, we have been able to obtain with a high degree of certainty two distinct groups of patents: (i) a set of 59 patents classified as true software patents; (ii) a set of 19 patents which do not protect software technology. The main results contained in table 2 clearly point to a better performance of 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 HTT database only identifies 10% of them. As for type I error, HTT database performs better than Gauss but the difference is minimal (5% vs 10%).

---

[18]This line of reasoning has been further validated by some interviews had with Italian patent attorneys expert in the field. In particular, one of them make the argument quite clear by stating: "almost 80% of Italian patent attorneys in the field are engineers and, for this reason, they share a common technical language that can be understood by colleagues with a certain degree of randomness. Furthermore, providing a vague patent description

Table 2: 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.

The results thus obtained are likely to point out how the method originally proposed by HTT for the US (Hall and MacGarvie, 2006) and translated to the European system in order to detect software patents (Hall, Thoma, and Torrisi, 2007) may regularly fail in recognising software patents. Our conclusion is that this bias might be partially due to the method proposed by Bessen and Hunt (2007), which constitutes an important part of the methodology set forth by Hall, Thoma, and Torrisi (2007), meaning that it may not be the most efficient one for detecting software patents filed at the EPO. We provide preliminary evidence of this fact by conducting a simple qualitative data analysis on titles and abstracts of the 59 patents classified as true software patents. As expected, ranking words and ad-hoc phrases contained in the patent abstract and patent title by frequency of appearance does not return any of the words/frequencies contained in the search algorithm proposed by Bessen and Hunt (2007). Providing an explanation of why this is the case it is beyond the scope of the present work and constitute food for thought of future research avenue. Nevertheless, a tentative explanation may lie in institutional differences as well as different ways patent attorneys draft software patents in the two patent systems.

Overall, Gauss database is composed by patents filed between 1978 and 2004. Many information have been extracted from the dataset. In particular, statistics concerning designated countries, yearly evolution in the number of filed and granted software patents, country of residence for both inventors and applicants and patents' software domain. As it can be seen from fig. 2 the number of software patents filed at the EPO has increased steadily starting from 1984. During the second half of the 1990s the increase has been impressive, jumping from 4,500 patents in 1995 to almost 12,000 for the year 2001. After 2001 the amount of software patents filed dropped consistently. One of the reasons for this fall can be reconnected to the burst of the 'dotcom' bubble that took place in the period 2000 - 2001. Indeed, the crisis of many firms making business in the ICT sector could have implied diminishing patent applications.

From the figure we note that pattern of granted patents follows closely that of filed ones, nevertheless the gap between the two is increasing. This

---

may result in a rejection due to the lack of an essential requirement".
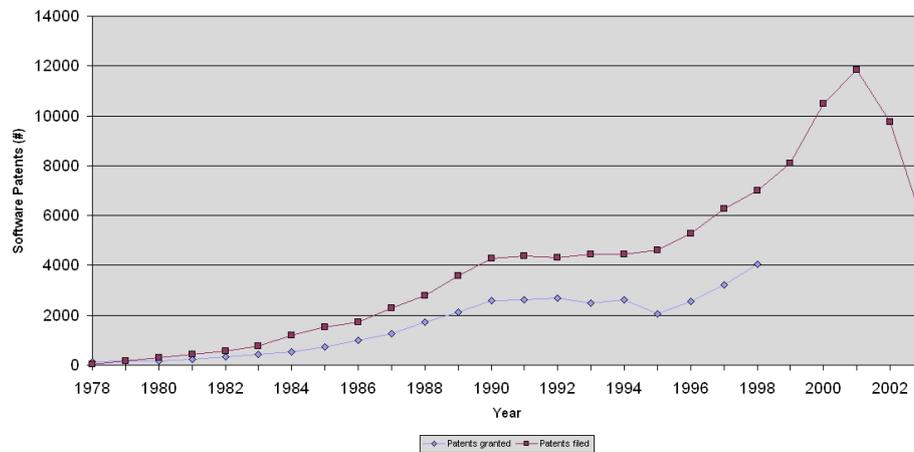
Figure 2: Yearly evolution of filed and lagged granted software patents (1978-2003)

might be an hint on the increasing strictness of the EPO concerning this patent typology. A proof of this statement can be inferred also from fig. 3 which graphs, for the top twenty applicants, the number of software patents granted as a percentage of filed patents. From this, we see that it never happened that more than the 50% of software patents filed have been granted at the EPO.

Figure 3: Granted versus non-granted patents for top 20 applicants

Other facts can be desumed thank to figures from 4 to 6. First the most designated countries, that are countries inside the EU where patent applicant is asking expressively for protection, are Germany, UK, France, Italy and The Netherland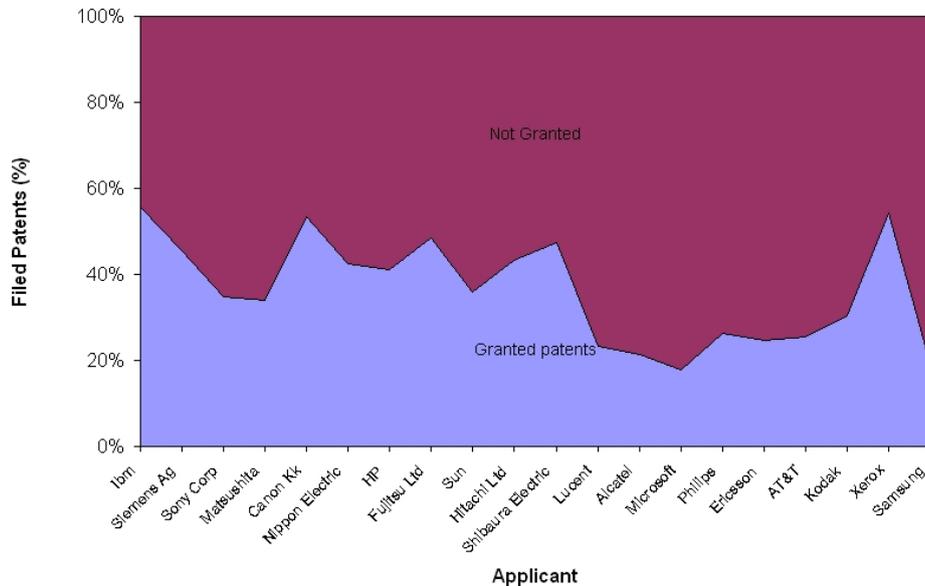s. But at the same time other countries, mainly European Union new entrants, are gaining importance, meaning that either seeking protection or behaving strategically is an active strategy by software patents applicants. Second, both applicants and inventors applying for software patents are mainly from US (respectively 39% and 40%) and Japan (respectively 25% and 26%) with a minor role played by European Union inventors and applicants. Germany, which is one of the best performing one, accounts only for 10% and 9%. This is mainly due to the leading role in ICT-related products by the US and Japan and from the fact that, at least for the US, software is susceptible of patenting since the beginning of the 1980s. This has allowed American firms to acquire expertise in both dealing with application procedures and identifying more valuable inventions to be patented.
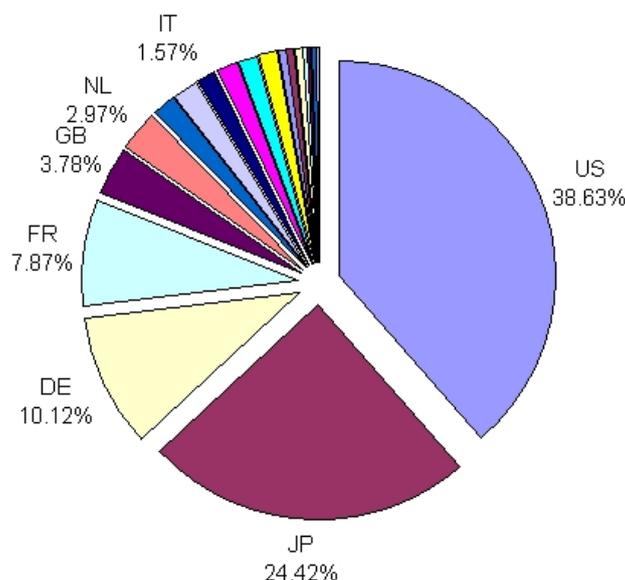

Figure 4: Country of residence for top 20 applicants

If we focus our attention on software patent concentration, then a highly concentrated pattern is discovered. According to fig. 7 the top 50 applicants account for more than the 50% of patents accorded at the EPO.

Taking a look to a subset of the database in the period between 1995 and
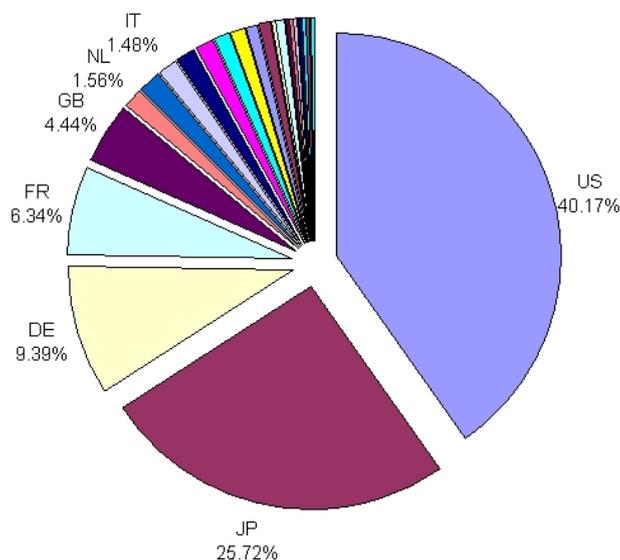
17

Figure 5: Country of residence for top 20 inventors

2003 provides useful insights on the recent dynamics in software patenting in the EU. Data shows an increasing number of filed patents which 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 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 of 3,5 years: while granted patents in 1997 had been filed about 1 year earlier, those granted in 2003 took, as an average, more than 5 years to complete the granting procedure. It must be mentioned that getting closer to 2003 the database updating procedure has a relevance in justifying a lower share of granted patents. It is possible to point out, anyway, that in the period 1995-2003 a lower number of patents had been granted against a fast 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 is mainly due to two main reasons: 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 means that the number of

18
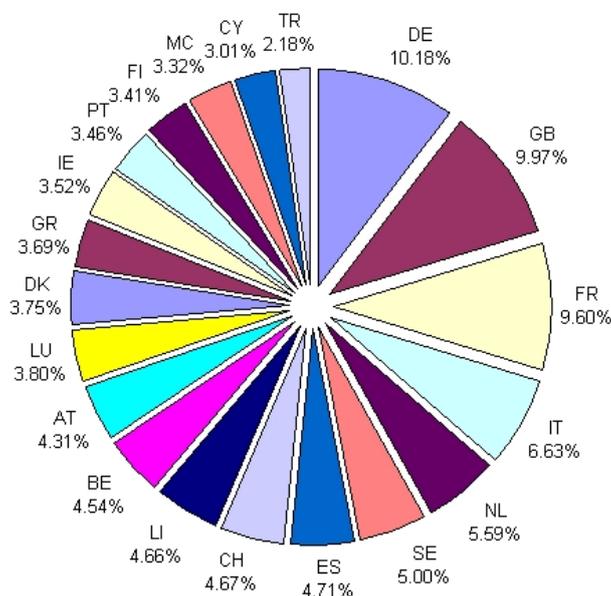
Figure 6: Top 20 designated countries

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. This has leaded international patent applications to be redirected there, given the higher quality assured in the granting procedure. This fact has additionally increased the already huge number of patent applications to be processed.

All of these reasons can not 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 against 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.

4.2 Sample Construction and Description

In order to investigate the determinants of software patenting at the firm level, a link has been established between a subset of Gauss database[19] and a set of other databases. In particular, we followed a three stage procedure. First of all, we matched firm's name from the '2004 EU Industrial Research

---

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

19

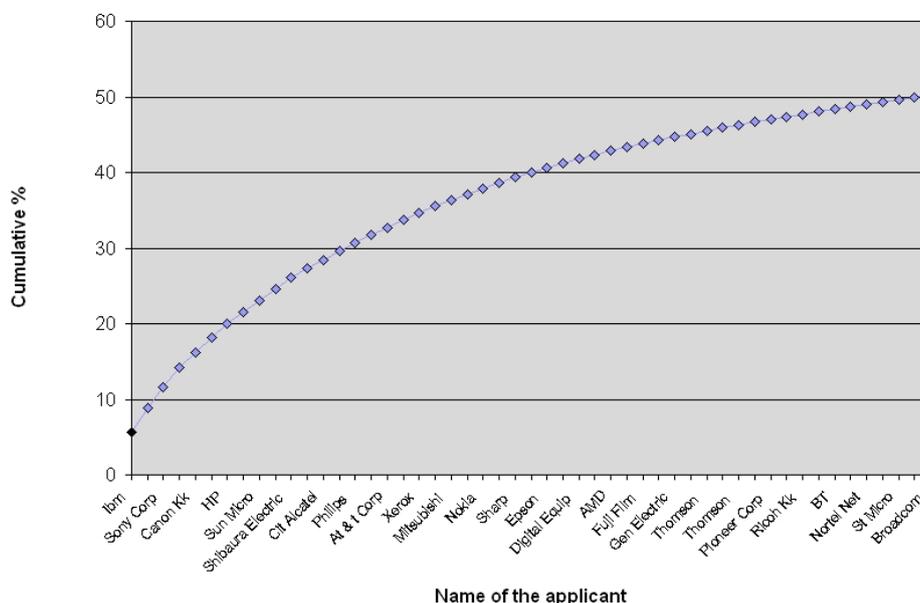Figure 7: Cumulative percentage of software patent applications by top 50 applicants

Investment Scoreboard'[20](henceforth R&D scoreboard) with patent assignee name from Gauss.[21] After that, we re-matched the two database 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 firm's name in order to retrieve additional information for our analysis (see fig. 8).[22]

---

[20]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.

[21]In order to establish proper linking relations, a semi-automatic data process to match companies to applicants has been performed. A specific small software application has been developed performing automatic matching between firms' values and requiring explicit operator's confirmation only in cases in which applicants were not univocally identified. This procedure has been coupled with a time consuming manual processing of the data in order to strongly increase the reliability of the sample.

[22]Amadeus business directory contains account data of European companies and their subsidiaries located in the EU, together with subsidiaries of non-EU companies, while Osiris business directory contains contains account data of 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,
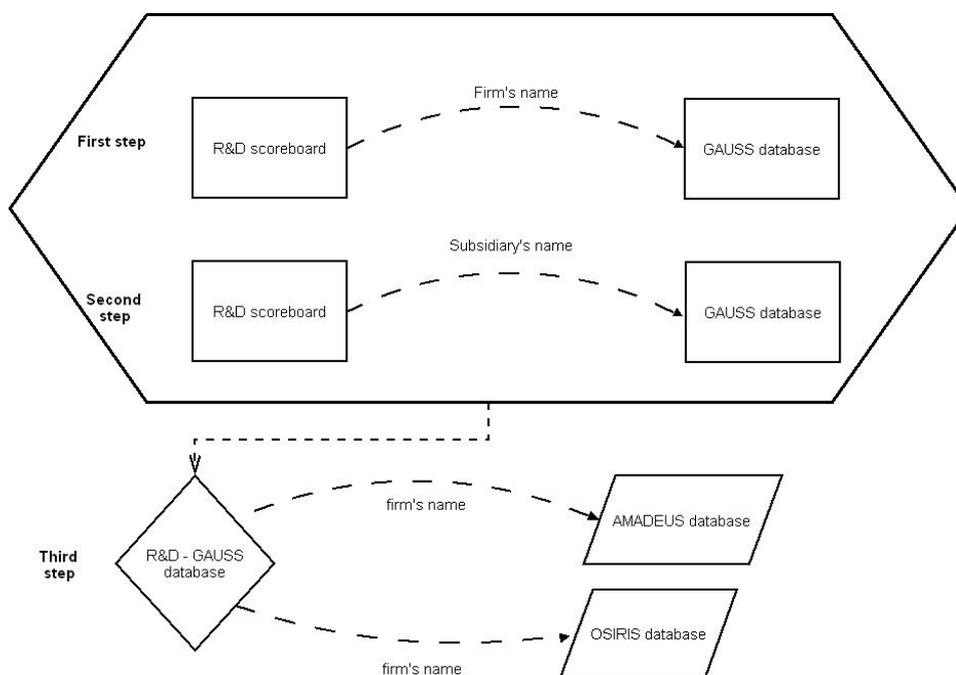
Figure 8: Database construction procedure

A resulting dataset obtained by linking the information available in the mentioned sources is composed by 979 firms whose data concerning Research and Development spending, sector and geographic classification, number of software patents filed are available for the period 2000-2003.

In order to check the representability of our sample we performed two main comparisons:

1. *ANBERD database vs R&D-Gauss.* In this first phase we compared 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 the comparison is done at the sector level, the R&D database accounts for the 71,35% of R&D performed[23].

---

useful to consolidate the data at the level of the ultimate parent company.

[23]In this second case the comparison is not as reliable as in the previous one. This is mainly due to the different sector classification characterizing the analysed datasets. Indeed, while OECD data follow NACE industrial classification R&D scoreboard are organised through Financial Times and London Stock Exchange (FTSE) one. In order to compare the two data sources we have built a NACE-FTSE table of comparison, nevertheless this is far from being totally satisfactory.

2. *EUROSTAT vs R&D-Gauss.* While from the R&D spending point of view we are pretty sure that the sample taken into consideration is representative, when the number of enterprises is analysed this might not be correct anymore. So, for the sake of comparison, we decided to rely on European statistics retrieved from EUROSTAT[24]. The comparison between our database and data on firm's 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 for the whole population of companies at the EPO, but that it gives a clear and reliable picture of the 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[25].

Once checked for the reliability of our sample, we can now shift to provide a general description of the dataset. Table 4 presents the distribution of companies by industry showing that IT hardware, electronic &electrical and software & computer services are those sectors where software patents are mainly present. Furthermore, companies patenting software at the EPO are found to be mainly US and Japanese companies (see table 5). Tables 3 and 4 also reveal a clear pattern of the sample, that is there is a high number of firms not applying for any patent. Indeed, only 406 companies out of 979 do applied for at least one software patent in the period 2000-03. Thus, the structure of the dataset calls for the implementation of a sound econometric model able to take into account data's specific pattern. At this respect, the choice made of adopting count data models is supported from both the nature of the depending variable and the structure of the data. Moreover plotting the number of filed patents against R&D spending, highlights the better fit reached by a Poisson distribution with respect to a linear one (the smoother line in figures 9 fits better to the data meaning that an exponential specification performs better).

---

[24]We also checked for the presence of statistics on firm population at the OECD level, but our attempt has been disappointing because of the presence of a high quantity of missing values.

[25]We have not been able to implement a post-stratification procedure to correct for existing biases. This is due to the unavailability of OECD data for the years of interest. In particular numerous missing values jeopardize the possibility of drawing statistics for the population of interest, whose contribution to the construction of appropriate weights for the post-stratification procedure is essential. For a more comprehensive discussion on the representability of R&D scoreboard for large companies see Frietsch (2004).

Table 3: Descriptive statistics

	N	Mean	Median	Standard Deviation	Min	Max
Stock of R&D expenditures	3560	14.85	4	35.49	0.02	333.84
Sales	3824	8225.15	2071.44	18849.06	0.7	212971.5
Employees	3849	27204.84	8316	51063.54	1	477100
Stock of filed software patents	3916	46.28	0	222.24	0	2285
Stock of granted software patents	3916	8.62	0	47.74	0	727.83
Index of sectoral concentration	3916	0.32	0.26	0.13	0.21	1
Strategic patenting	2937	4596.14	1922.25	5789.02	0	17835

Table 4: Distribution of companies by industry

	with R&D	with software patent
FTSE sector		
Aerospace & defence (21)	25	15
Automobiles & parts (31)	60	27
Banks (81)	1	0
Beverages (41)	4	0
Chemicals (11)	78	32
Construction & building (13)	21	9
Diversified industrials (24)	17	7
Electricity (72)	17	4
Electronic & electrical (25)	72	43
Engineering & machinery (26)	91	38
Food & drug retailers (63)	2	0
Food producers (43)	22	3
Forestry & paper (15)	7	0
General retailers (52)	7	1
Health (44)	38	9
Household goods & textiles (34)	29	10
IT hardware (93)	138	95
Leisure & hotels (53)	3	0
Media & entertainment (54)	17	7
Mining (04)	5	0
Oil & gas (07)	19	7
Personal care & household (47)	14	5
Pharma & biotech (48)	126	28
Software & computer services (97)	94	41
Speciality & other finance (87)	3	0
Steel & other metals (18)	15	6
Support services (58)	17	6
Telecommunication services (67)	22	12
Tobacco (49)	4	0
Transport (59)	3	0
Utilities - other (73, 78)	8	1
Total	**979**	**406**

Table 5: Distribution of companies by country

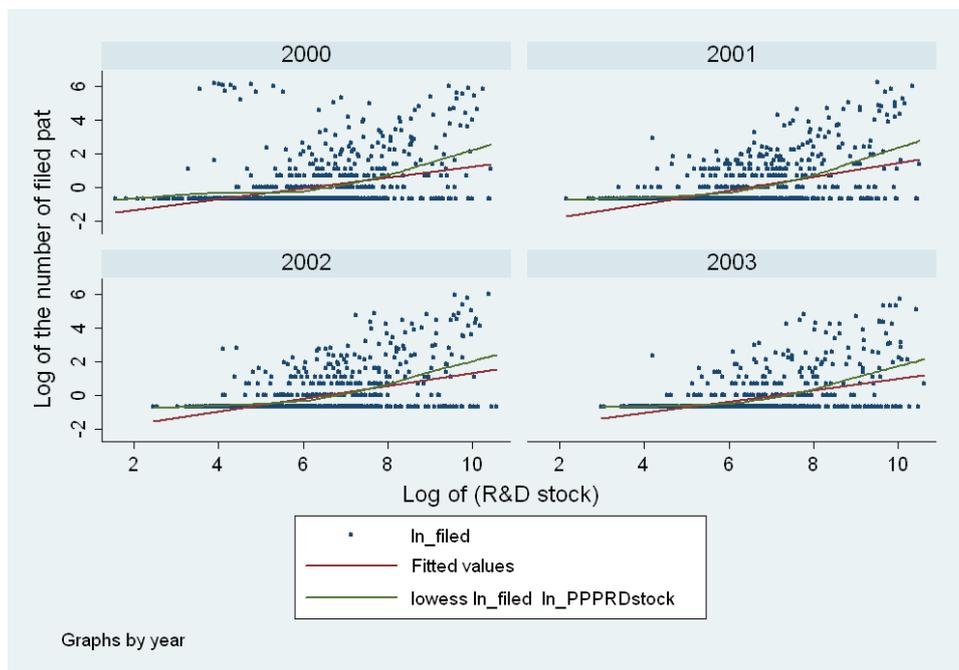|	with R&D	with software patent
Country		
Australia	2	2
Austria	10	1
Belgium	16	1
Canada	7	6
Denmark	28	4
Finland	28	7
France	66	26
Germany	100	37
Greece	2	1
Hungary	2	0
Ireland	4	1
Italy	17	6
Japan	153	97
Luxembourg	2	0
Norway	3	1
South Korea	9	3
Spain	9	1
Sweden	44	12
Switzerland	18	8
The Netherlands	22	6
UK	149	36
USA	288	150
Total	**979**	**406**

Figure 9: Plots of filed patents against R&D stock, 2000-03

4.3   Variables

In this section we will proceed to describe the variables implemented during the econometric estimation. A more systematic and clear-cut definition of variables is presented in table 6 as well. Dependent variables change according to the type of model adopted. In particular, the dependent variable is of a numeric type when either the poisson or negative binomial model are implemented[26] On the contrary, dependent variable is dichotomous when a logit panel data model is investigated. This is due to the main goal of our analysis in this case, that is whether the firm is applying for patents or not[27]

On the other side, independent variables can be divided into two main groups: structural and control variables. **Structural variables** include all those variables that are object of the analysis throughout different specifications. These variables are:

1. *R&D spending ($R\&D_{i,t}$).* This is the amount of the R&D spending performed by firm i at time t. The amount has been transformed in purchasing power parity dollars (PPP$) to allow comparability among different countries and it has been expressed as a stock (see table 6 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. At 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, Griliches, and Hausman, 1986; Montalvo, 1997).

2. *Sales ($Sales_{i,t}$).* This is the amount of sales achieved by firm i in year t. The same transformations as for R&D spending has been performed.

3. *Employees ($Empl_{i,t}$).* The number of employees for firm i at time t. This variable proxies for firm size and it influences the number of software patents filed. In fact, larger firms are likely to have more resources in order to apply for more patents. This is even more likely

---

[26]This happens because, in this last case, we are interested in the number of patents filed by the single firm in a given year. At this regard, the variable is computed as the stock of filed patents a firm files in the present year.

[27]This is why the variable is taking only two values: value one in case firm files a software patent, zero otherwise.

to happen in the EU where the average cost of a patent is higher than in other patent systems (Malerba and Montobbio, 2002).

4. *Sector concentration (Sec_j).* This has been computed as the total sales of the four largest firms in terms of sales in firm's i main sector of activity (indicated by j) divided by the overall amount of sales of the same sector[28].

5. *Strategic rivalry ($Strat_{i,t}$).* This is the stock of software patents filed by firms belonging to the same sector of firm i in 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' innovation and to decrease their average return to R&D while affecting own market value (Noel and Schankerman, 2006).

6. *Stock of software patents filed in the previous year (Pat_{t-1}).* This is the stock of software patents filed by firm i in year $t-1$. This variable takes into account the effect on software patenting decision by the number of software patents filed in the former year.

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

1. *Year dummy ($Year_t$).* These are a set of four dummy variables[29] which take into account the effect of external outcomes to the knowledge production function. In particular, it gives a hint on the institutional context where the firm is operating and on the different happenings taking place.

2. *Geographical proxy ($Country_i$).* We use these variables to disentangle the effects produced by different patent systems a firm has been dealing with[30]. A firm that is used to operate inside the US have a deep knowledge concerning 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 like a hindering mechanism.

---

[28]This has been done following the FTSE sector classification. By definition, this is based on firms operating in the same sector and it does not always reflect the impact of direct competition. Unfortunately, this is the best approximation we have been able to make according to available data.

[29]In particular, years from 2000 to 2003 are taken into consideration.

[30]In line with the theoretical considerations of section 3.1, we analyse fourth main categories: American, European, Japanese and 'remaining' patent system.

3. *Sectoral/Technological proxy ($Tech_i$).* These dummy variables proxy the technological opportunities arising from the seven sectors of activity a firm can belong to.

Table 6: Variables definition

Variable	Name in the regression table	Description
Stock of R&D spending	R&D	Stock of R&D expenditures 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%
Number of employees	Empl	Number of employees
Sales	Sales	Sales expressed in millions of PPP\$
Stock of software patents in previous year	Pat_1	Stock of software patents filed in the previous year 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%
Stock of software patents in current year	Pat	Stock of software patents filed in the present year 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%
Sector concentration ratio	Sector_conc	Sector concentration ratio relative to the FTSE sector the firm belongs to. It is computed as 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 patenting	Strategic	Stock of software patents filed in the previous year 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
Sectoral dummies	DS*	Variable assuming value 1 if the firm belongs to one particular FTSE sector and 0 otherwise
Country dummies	DC*	Variable assuming value 1 if the firm is located in one particular area (US, Japan and the EU) and 0 otherwise

5 Results

Results from the econometric exercise are presented in tables ranging from 7 to 9. All of these show Poisson panel data models in the specification as conceived by Wooldridge (2005). In all of these specifications, dependent variable is computed as the stock of filed software patents (see table 6 for details on variable construction).[31]

In table 7 we report the results for the Poisson panel data model estimation for the whole sample adopting both fixed and random effects. Random effects models are of three different typologies in our specification. The typology varies according to the set of control variables that are introduced, i.e. year dummy, geographical dummy and technological dummy. Results show that the stock of software patents applied for in the previous year is always significant but its value is negligible. The amount of sales is the only other variable always significant. In the random effects case the variable proxying for strategic factors affecting software patenting is significant and positive for all of the three specifications. Among the control variables, the most striking fact is the significance and high value for almost all the sectors taken into consideration. Among the others, being part of electronics, hardware, media and software sectors explains the likelihood to patent software at the EPO. After the estimation, we checked for the best specification through the implementation of the Hausman test (Hausman, 1978). According to the result, fixed effects must be preferred because they provide both consistent and efficient estimators of the parameters of interest. While in the former case the random effects are rejected by the Hausman test, when we proceed to an analysis at the single sector level the Hausman test is not able to give us proper information on the issue. Indeed, this points to the fact that implementing random effects, that is allowing for the presence of unobserved permanent differences across firms, yields reliable estimates of the parameters of interest (Hall and Ziedonis, 2001). There are important information to be drawn from variables which vary among firms but that are constant through time, i.e. geographical and technological proxies. Relying exclusively on the fixed effect specification will sweep them out. In this way, we will not be able to take into account their effect on the estimation procedure. For all of the above mentioned reasons, we decided to concentrate our analysis respectively on the software and hardware sector relying mainly on random effects estimates (see tables 8 and 9).

In table 8, R&D spending is not significantly related to the number of software patents a firm applies for. This result is likely to confirm the fact that patent is not considered a useful appropriability instrument for the results of

---

[31]We also controlled for patent quality by running alternative regressions with the stock of granted software patents as dependent variable. Results in this case are in line with those presented here, that is size and significance of coefficients are basically unaltered. We are grateful to a referee for pointing this out.

R&D process, mostly in technologies characterised by cumulativeness in the innovation process. This is a result obtained for the EU at the general level — i.e. concerning patents in general (Arundel, 2001)— but no evidence was provided for software patents. To us, the absence of a significant relationship between the two variables supports our belief that R&D contributes to the creation of knowledge capital and innovation but that software patents do not proxy well the innovation output, meaning that it is not deemed as a suitable appropriability measure. Obviously, 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 of 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).

On the other side, the number of employees seems to play an important role for software patenting at the firm level, underlying 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 IPRs managing departments.

Finally, the variable proxying the role of strategic factors is never significant for firms belonging to the software sector. This fact sheds light on both the economic and technological nature of the inventive process typical of high–tech industries. In particular, the software sector has been found to be characterised by a technology which 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 does not deem patent protection as a useful mechanism spurring future inventive streams. On the contrary, it is common belief among practitioners that software patents are likely to hinder the rate of innovation ant to lock-in the market in favor of few monopolistic firms (Blind and Edler, 2003; Ghosh, 2007)

Results for the IT hardware sector are presented in table 9. R&D spending is not significant as in the software case while firm size is significant, even if the coefficient is lower than in the software case. On the contrary, the variable proxying for strategic factors contributes consistently to explain the number of software patents applied for. We interpret this result as a clue on the pressure put on firm patenting strategies by rivals belonging to the same sector. This is likely to point to the presence of a 'threat effect', that is a strategic interest of hardware firms to patent software for reasons other than the increase in its own inventive capacity. Indeed, the coefficient proxying for the number of software patents filed in the previous year by firms other than the firm under consideration (variable *strategic*) 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; but, at the same time, they are eager to patenting 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 neighbor firms can hinder future development of both hardware and embedded software, in this way leading the company to apply for patents as a defensive strategy.

Table 7: Poisson panel data estimation: Fixed and Random Effects

Variables	FE	RE (1)	RE (2)	RE (3)
Pat_1	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
R&D	-0.115*	0.109**	0.027	-0.056
	(0.046)	(0.040)	(0.039)	(0.039)
Sales	0.074***	0.068***	0.077***	0.080***
	(0.017)	(0.017)	(0.017)	(0.017)
Empl	-0.052	0.167***	0.093**	0.015
	(0.043)	(0.036)	(0.035)	(0.036)
Strategic	0.072	0.182***	0.186***	0.081*
	(0.039)	(0.032)	(0.032)	(0.036)
Sector_conc	0.009	0.204**	0.076	0.035
	(0.068)	(0.076)	(0.066)	(0.066)
Dyear02		-0.007		
		(0.010)		
Dyear03		-0.058***		
		(0.011)		
DCeu			0.606	
			(0.608)	
DCusa			0.348	
			(0.605)	
DCjapan			0.820	
			(0.636)	
DSelectr				1.945***
				(0.459)
DSeng				-0.446
				(0.394)
DShard				1.695***
				(0.365)
DSmedia				1.605*
				(0.813)
DSsoft				1.197**
				(0.394)
DStel				1.058
				(0.794)
cons		0.885*	0.580	1.586***
		(0.370)	(0.737)	(0.367)
χ^2	970.382	1102.615	1069.587	1116.493
N	1102.000	2504.000	2504.000	2504.000
Log-likelihood	-1705.271	-4284.131	-4305.951	-4286.582

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors are in parentheses. The specification of the model relies on Hausman, Hall, and Griliches (1984), Hall and Ziedonis (2001) and Wooldridge (2005).

Table 8: Poisson panel data estimation for the software sector: Fixed and Random Effects.

Variables	FE	RE (1)	RE (2)	RE (3)
Pat_1	0.001***	0.001***	0.000	0.001***
	(0.000)	(0.000)	(0.001)	(0.000)
R&D	0.558	0.201	0.379	0.669
	(0.435)	(0.224)	(0.261)	(0.293)
Sales	0.109	0.161*	0.102	0.095
	(0.084)	(0.065)	(0.078)	(0.070)
Empl	0.677*	0.393*	0.483*	0.785**
	(0.364)	(0.183)	(0.196)	(0.241)
Sector_conc	0.353	-1.591		-1.208
	(1.888)	(1.877)		(1.881)
Strategic	-0.111	-2.057	-6.818	-1.490
	(1.685)	(1.658)	(3.888)	(1.675)
Dyear01			1.242	
			(0.783)	
Dyear02			0.577	
			(0.368)	
DCeu				1.201
				(2.317)
DCusa				-1.290
				(2.268)
cons		16.562	58.887	9.588
		(11.901)	(33.041)	(12.591)
χ^2	264.908	279.787	281.625	292.000
N	117.000	265.000	265.000	265.000
Log-likelihood	-234.368	-519.572	-518.665	-516.397

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors are in parentheses. The specification of the model relies on Hausman, Hall, and Griliches (1984), Hall and Ziedonis (2001) and Wooldridge (2005).

Table 9: Poisson panel data estimation for the hardware sector: Fixed and Random Effects.

Variables	FE	RE (1)	RE (2)	RE (3)
Pat_1	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
R&D	0.001	0.073	0.068	0.085
	(0.066)	(0.058)	(0.058)	(0.060)
Sales	0.088**	0.091***	0.085**	0.090**
	(0.027)	(0.027)	(0.028)	(0.027)
Empl	0.082	0.159**	0.148**	0.169**
	(0.060)	(0.051)	(0.052)	(0.054)
Sector_conc	0.542	0.478		0.474
	(0.295)	(0.295)		(0.295)
Strategic	1.314**	1.209**	6.966*	1.217**
	(0.422)	(0.422)	(4.224)	(0.422)
Dyear02			0.311	
			(0.192)	
Dyear03			0.776	
			(0.507)	
DCeu				0.024
				(1.270)
DCusa				-0.639
				(1.225)
DCjapan				-0.102
				(1.364)
cons		-8.281*	-65.164	-8.094*
		(3.769)	(41.256)	(3.969)
χ^2	323.424	347.134	348.777	349.945
N	264.000	395.000	395.000	395.000
Log-likelihood	-474.096	-1112.417	-1111.478	-1111.286

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors are in parentheses. The specification of the model relies on Hausman, Hall, and Griliches (1984), Hall and Ziedonis (2001) and Wooldridge (2005).

6 Discussion and Conclusion

The main goal of this article has been to provide a deep 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 some time. Nevertheless, in the last decade the number of software patents filed at the EPO has grown rapidly despite article 52 of the European Patent Convention expressively prohibits software patenting. To investigate the topic, we present a new database containing information on software patents, i.e. Gauss database. The database has undergone an extensive check in order to prove its reliability, not only it performs 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 thanks to the contribution of several experts. These data show how software patents are an important phenomenon in the EU as well given that more than 30,000 software patents have been granted to both European and foreign firms till nowadays. In this respect, a large part of them has been accorded to American and Japanese firms. The fact that nearly the majority of granted patents belongs to foreign companies must be due to the higher experience that these firms have acquired dealing with their own patent system. For example, software is patentable since 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 the others, 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 both inventors and applicants come respectively from US and Japan. Then, the knowledge production function approach has been implemented in order 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, technological and geographical opportunities, and to deal with our specific interests, i.e. the idiosyncrasies of both software and hardware sector. Both the way in which the dataset has been built and a robustness check for the database itself have been presented. Moreover a set of different methods of estimation have been put forward and the most suitable one has been chosen. After that, the results of the chosen econometric model have been presented. The main outcome of the analysis deals with the strategic behaviour by firms belonging to the hardware sector, that file software patents

37

mainly as a 'threat effect' against competitors.

We are aware of limitations to the present work. First of all, the EU Scoreboard, on which we rely 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. Second, there has not been the possibility to separate R&D expenditure used in software production from the one used for other purposes. In order 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. This means that almost all of the investment in R&D drives to software patenting and not to patent another type of technology. Third, the method of estimation that has been adopted relies heavily on a set of general assumptions. Among the others, the assumption of strict exogeneity is the most controversial one[32]. In our specific case, this condition implies that patents cannot depend on additional R&D expenditures to be fully appropriate or improved. Obviously, this statement is contrary to the common practice. In the reality, most of the firms make additional investments either to improve or to fully commercialize an invention protected by patent. Nevertheless, the fact that R&D expenditure cannot be considered completely as strictly exogenous does not invalidate the analysis carried out along this work. Indeed, most of the empirical contributions so far has mainly relied on these method of estimations. Only recent works have tried to overcome above mentioned limitations and they have not reached viable solutions yet (Montalvo, 1997; Blundell, Griffith, and Windmeijer, 2002). In particular, the adoption of GMM procedures is rather computationally intensive, meaning that a large amount of data is needed both at the individual level and through time. Thus, in our case the trade-off between data constraints and estimation efficiency has been resolved relying on more trustworthy estimation procedures. Finally, we provide the estimation on a narrow time window only, that is 2000-2003. This is mainly due to the lack of data concerning R&D for a long time frame for companies located in the EU, which is a common problem in studies of this kind.[33]

---

[32]According to this, the error term is not correlated at all with future, past and present values of the explanatory variables. In case the assumption does not hold, then the consistency of the estimators fails to be achieved.

[33]Unlikely in the US, where data on R&D are revealed by companies due mainly to fiscal reasons, European companies are less likely to do that given 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 this data and couple them with other data sources to

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, manual check comprised 3 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) thus meaning an extensive and long work in 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 4-years long time period and setting-up a dataset covering a longer time frame at the expense of trustworthiness.

References

ALLISON, J., AND M. LEMLEY (2000): "Who's Patenting What? An Empirical Exploration of Patent Prosecution," *Vanderbilt Law Review*, 53, 2099–2148.

ALLISON, J., AND E. TILLER (2003): "Internet Business Method Patents," in *Patents in the Knowledge-Based Economy*, ed. by W. Cohen, and S. Merrill, pp. 259–284. National Academies Press.

ANDERSEN, E. B. (1970): "Asymptotic Properties of Conditional Maximum-Likelihood Estimators," *Journal of the Royal Statistical Society. Series B (Methodological)*, 32(2), 283–301.

ARROW, K. (1962): "Economic Welfare and the Allocation of Resources for Invention," in *The Rate and Direction of Inventive Activity: Economic and Social Factors*, ed. by R. Nelson. NBER, Princeton University Press.

ARUNDEL, A. (2001): "The relative effectiveness of patents and secrecy for appropriation," *Research Policy*, 30(4), 611–624, available at http://ideas.repec.org/a/eee/respol/v30y2001i4p611-624.html.

BERGSTRA, J., AND P. KLINT (2007): "How to find a software patent?," Discussion paper, Informatics Institute, University of Amsterdam.

BESSEN, J., AND R. HUNT (2007): "An Empirical Look at Software Patents," *Journal of Economics & Management Strategy*, 16, 157–189.

---

fill 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 them to EU or non-EU companies, located elsewhere. Given our interest in understanding different strategies put forward by EU and foreign companies relative to software patents, we had to rely on other data sources to collect R&D data, thus constraining our possibility to extend the time.

BLIND, K., AND J. EDLER (2003): "Idiosyncrasies of the Software Development Process and Their Relation to Software Patents: Theoretical Considerations and Empirical Evidence," *Netnomics*, 5(1), 71–96, 1385-9587.

BLUNDELL, R., R. GRIFFITH, AND F. WINDMEIJER (2002): "Individual Effects and Dynamics in Count Data Models," *Journal of Econometrics*, 108, 113–131.

CAMERON, C., AND P. TRIVEDI (1998): *Regression Analysis of Count Data*pp. 275–300. Cambridge University Press.

CAMPBELL KELLY, M., AND P. VALDURIEZ (2005): "Technical Critique of Fifty Software Patents," *Marquette Intellectual Property Law Review*, 9, 249.

CHABCHOUB, N., AND J. NIOSI (2005): "Explaining the propensity to patent computer software," *Technovation*, 25(9), 971–978.

COHEN, W. M., R. R. NELSON, AND J. P. WALSH (2000): "Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)," NBER Working Papers 7552, National Bureau of Economic Research, Inc, available at http://ideas.repec.org/p/nbr/nberwo/7552.html.

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.

GHOSH, R. (2007): "Study into Effects of Allowing Patent Claims on Computer-Implemented Inventions," Discussion paper, UNU-MERIT, Study for the European Commission, DG INFSO.

GILBERT, R., AND D. NEWBERY (1982): "Preemptive Patenting and the Persistence of Monopoly," *The American Economic Review*, 72(3), 514–526.

GRAHAM, S., AND D. MOWERY (2003): "Intellectual Property Protection in the U.S. Software Industry," in *Patents in the Knowledge-Based Economy*, ed. by W. Cohen, and S. Merrill, pp. 219–258. National Academies Press.

——— (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., B. H. HALL, D. HARHOFF, AND D. C. MOWERY (2002): "Post-Issue Patent "Quality Control": A Comparative Study of US Patent Re-examinations and European Patent Oppositions," Discussion

Paper 8807, National Bureau of Economic Research, Inc, available at http://ideas.repec.org/p/nbr/nberwo/8807.html.

GRILICHES, Z. (1990): "Patent Statistics as Economic Indicators: a Survey," *Journal of Economic Literature*, 28(4), 1661–1707.

GRILICHES, Z., AND J. MAIRESSE (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., Z. GRILICHES, AND J. HAUSMAN (1986): "Patents and R&D: Is There a Lag?," *International Economic Review*, 27(2), 265–283.

HALL, B., AND M. MACGARVIE (2006): "The private value of software patents," NBER working paper 12195, NBER.

HALL, B., AND R. H. ZIEDONIS (2001): "The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry, 1979-1995," *RAND Journal of Economics*, 32(1), 101–28.

HALL, B. H., G. THOMA, AND S. TORRISI (2007): "The Market Value of Patents and R&D: Evidence from European Firms," Discussion paper, SSRN e-library.

HAUSMAN, J. (1978): "Specification Tests in Econometrics," *Econometrica*, 46(6), 1251–1271.

HAUSMAN, J., B. HALL, AND Z. GRILICHES (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), 463–95, Available at http://ideas.repec.org/a/ucp/jlawec/v38y1995i2p463-95.html.

MALERBA, F., AND F. MONTOBBIO (2002): "L'Europa dell'Alta Tecnologia: lo Scenario Internazionale in Alcuni Settori Chiave," in *Tendenze dell'Industria Italiana*, ed. by Centro Studi Confindustria, pp. 63–78. Centro Studi Confindustria.

MANSFIELD, E. (1986): "Patents and Innovation: An Empirical Study," *Management Science*, 32(2), 173–181.

MARENGO, L., AND C. PASQUALI (2006): "Non Rivalry and Complementarity in Computer Software," DIME Working Papers 11, DIME Working Papers.

MAZZOLENI, R., AND R. NELSON (1998): "The Benefits and Costs of Strong Patent Protection: a Contribution to the Current Debate," *Research Policy*, 27, 273–284.

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

MERGES, R., AND R. NELSON (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., AND S. WINTER (1982): *An Evolutionary Theory of Economic Change.* Belknap press.

NOEL, M. D., AND M. SCHANKERMAN (2006): "Strategic Patenting and Software Innovation," CEPR Discussion Papers 5701, C.E.P.R. Discussion Papers, Available at http://ideas.repec.org/p/cpr/ceprdp/5701.html.

OLSSON, H., AND D. H. MCQUEEN (2000): "Factors Influencing Patenting in Small Computer Software Producing Companies," *Technovation*, 20(10), 563–576.

PAKES, A., AND Z. GRILICHES (1984): "Patents and R&D at the Firm Level: a First Look," in *R&D, Patents and Productivity*, ed. by Z. Griliches, pp. 55–72. University of Chicago Press.

PEETERS, C., AND B. VAN POTTELSBERGHE DE LA POTTERIE (2006): "Innovation Strategy and the Patenting Behavior of Firms," *Journal of Evolutionary Economics*, V16(1), 109–135.

SAKAKIBARA, M., AND L. BRANSTETTER (2001): "Do Stronger Patents Induce More Innovation? Evidence from the 1988 Japanese Patent Law Reforms," *RAND Journal of Economics*, 32(1), 77–100, Available at http://ideas.repec.org/a/rje/randje/v32y2001i1p77-100.html.

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 *Innovation Policy and the Economy*, ed. by A. Jaffe, J. Lerner, and S. Stern, pp. 119–50. MIT Press, Cambridge.

TRAPHAGAN, M. (1998): "Software Piracy and Global Competitiveness: Report on Global Software Piracy," *International Review of Law, Computers & Technology*, 12(3), 431–451.

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.