ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 49/50 — 1998

Appropriation Strategy and the Motivations to use the Patent System: an Econometric Analysis at the Firm Level in French Manufacturing

Emmanuel DUGUET
Isabelle KABLA*

ABSTRACT. – This paper studies the determinants of both the percentage of innovations that are patented and the number of European patent applications by industrial firms, using data from the French survey on appropriation (EFAT). We build a two equations model including count and interval dependent variables and estimate it by asymptotic least squares. Controlling for the traditional determinants of innovation, like research and development expenditures, we find that patent disclosure is the main reason why firms do not patent all their innovations. Moreover, once we control for the differences in the propensity to patent, patent disclosure also reduces the number of patents applications. On the other hand, the will of firms to acquire a stronger position in technology negotiations and to avoid trials increases the number of patent applications.

Stratégie d'appropriation et motivations du recours au brevet : une analyse économétrique sur données d'entreprises

RÉSUMÉ. – Cet article étudie les déterminants du pourcentage d'innovations brevetées et du nombre de brevets européens déposés par les entreprises industrielles, à partir de l'Enquête Française sur l'Appropriation Technologique (EFAT) menée par le Sessi en 1993. Nous estimons un modèle structurel à deux équations dont les variables expliquées sont soit connues seulement par intervalle soit une variable de comptage. Suivant la méthode d'Amemiya, l'estimation est réalisée par les moindres carrés asymptotiques. Une fois pris en compte les déterminants traditionnels de l'innovation, incluant l'investissement en recherche et développement, nous trouvons que la principale raison pour laquelle les entreprises ne brevettent qu'une partie de leurs innovations est la divulgation des informations contenues dans le brevet. De plus, même lorsque l'on tient compte des différences de propension à breveter, la divulgation réduit le nombre de brevets déposés. A l'inverse, la volonté des entreprises d'améliorer leur position lors de négociations technologiques et d'éviter les procès en contrefaçon incite les entreprises à déposer plus de brevets.

---

* E. DUGUET: CEME, Université de Paris I et INSEE-DMSE; I. KABLA: INSEE. We thank A. ARUNDEL, B. CRÉPON, W. COHEN and an anonymous referee for helpful suggestions and comments. We also thank the participants at the X[th] ADRES *Conference on the Economics and Econometrics of Innovation* (Strasbourg, June 1996) as well as the participants at the *Econometric Society European Meeting* (Istanbul, August 1996).

1 Introduction

The patent system helps inventors to benefit from their research efforts, thanks to the temporary monopoly it provides on the patented inventions. In addition to the private incentives to undertake research that this legal device aims to provide, the patent system has also a positive effect from the social viewpoint: the disclosure of technological knowledge to other firms. In order to patent an invention, firms have to provide the patent office with a detailed technical description of the invention, which is published. This description must be comprehensive enough to allow a 'man of the art' to reproduce it. Thus, by examining patents, any firm working in the same field is able to learn from the technical advances that the patent applicant has made and to use this knowledge to reproduce the invention at the expiry of the patent or to improve on the invention. These features of the patent system seek to avoid the needless duplication of research expenditures and to promote technical progress. Globally, the patent system has a controversial effect on the social welfare due mostly to the trade-off between appropriation and diffusion. On the one hand, the patent system creates or increases the incentives to innovate. On the other hand, it temporarily distorts competition since it grants monopoly rights to the applicants. A vast literature that can be traced back to ARROW [1962] has explored the optimal level of protection that patent should confer and the optimal structure of the protection according to presumably available instruments: patent duration, patent scope or royalties in a compulsory license system (see for instance NORDHAUS, [1969]; KAMIEN and SCHWARTZ, [1974]; TANDON, [1982]; DEBROCK, [1985]; BECK, [1986]; GILBERT and SHAPIRO, [1990]; KLEMPERER, [1990]; GALLINI, [1992]).

Most articles make the assumption that inventions are patented. Less attention has been paid to firms' incentives to patent and to the implication of non-patenting strategies on social welfare. However, several surveys showed that the innovative firms do not always apply for patents and that when they do, they patent only a part of their inventions (LEVIN *et al.* [1987]; BUSSY *et al.* [1994]; ARUNDEL *et al.* [1995]; COHEN, NELSON and WALSH [1997]). One alleged reason is that the monopoly conferred by patent laws is limited in comparison to what an innovator could yield without patenting. In an analysis of the first Yale survey results on the issue of appropriation mechanisms in the United-States, LEVIN *et al.* [1987] found that patent was considered by firms to be neither the only way to appropriate innovation benefits nor as the most efficient one. Firms declared that the most important drawback of patenting was its failure to prevent competitors from inventing around the patented inventions. Moreover, they asserted that methods other than patents, like secret and lead time, could be used to protect against imitation, at least temporarily. The ability of patents to prevent imitation also appears to be limited. MANSFIELD *et al.* [1981] in a survey conducted in the United States, found that the average time required to imitate 48 new products was 70% of the time necessary to innovate and that the imitation cost was 65% of the innovation cost. Moreover, the imitation lag and costs were only moderately increased by patent protection.

Another possible reason to decide not to patent is disclosure. This appeared as an important deficiency of patent protection in the analysis of the Yale survey results by LEVIN *et al.* [1987]. The ability of a patent to provide incentives to innovate should be inversely related to the amount of useful information that leaks out to competitors, particularly when the information is difficult to obtain otherwise.

The failure of the patent system to prevent imitation, in comparison with other means of protection, and the role of patent disclosure are also the principal factors that are considered in the theoretical papers that explore or include the propensity to patent in their analyses (CRAMPES, [1986]; GALLINI, [1982]; SCOTCHMER and GREEN, [1990]; VAN DIJK, [1994]; KABLA, [1997]). The role of a patent as a signal has also been studied by HORSTMANN *et al.* [1985]). Only a few comprehensive welfare analyses have been conducted regarding the consequences of non-patenting strategies. SCOTCHMER and GREEN [1990] and KABLA [1997] both examine the question in respect to the diffusion of technologies [1].

While some innovations may remain unpatented, other inventions can lead to several patents. Therefore, the number of patents per patented innovation depends on the nature of each innovation and on the firms' patent strategy. On the one hand, a single innovation could be based upon a complex association of several inventions. On the other hand, when different and competing industrial applications of the same innovation are possible, the innovation results from a choice between them. In chemicals for instance, a whole family of molecules could produce similar desirable effects. Industrial usage will often be limited to only one of them but the innovator needs to obtain patents for the whole family in order to be adequately protected against imitation. Firms could even patent less efficient processes than the method that they intend to use, or less performing products than the one they will market, in order to increase their competitive advantage. In short, both the number of inventions behind a single innovation and the need to patent a cluster of similar inventions can result in a profusion of patent applications for a single innovation. "Over" patenting is also a strategy that can be used to increase bargaining power in negotiations with competitors, especially when overlapping technologies are involved. The stronger the patent portfolio, the stronger the position of a firm in convincing competitors that negotiation is preferable to litigation. This phenomenon also increases the risk of technology cartels, where incumbents could block entry trough the threat of a lawsuit.

Finally, patent applications can be made at different stages in the innovation process. Patenting at an early stage of the innovation could be required if the technological competition is fierce, for example when a competitor might patent a similar innovation first. Conversely, early patenting could be dangerous when a perfectible technology is involved.

1. The article of SCOTCHMER and GREEN is primarily devoted to the determination of the optimal patent scope in a two-stage patent race. The choice of the first inventor of an intermediate invention wether to patent or not is shown to generally diminish the advantage of a weak protection. KABLA is mostly interested in how the patent disclosure and the level of diffusion of technical knowledge by other means influence the optimal patent scope.

Since information leaks out to competitors through patent documents, they could catch-up and eventually leap-frog the initial inventor. The profitability aspects and the difficulty of the research process can also play a role in the optimal strategy. Early patenting could eventually result in more patent applications than the inverse strategy, if the firm also patents improved versions of its invention.

The paper examines which factors influence significantly the patenting behavior of firms. We are particularly interested in showing the relevance of dividing the analysis of patenting into two aspects: the percentage of innovations that involve patent applications (henceforth, the propensity to patent) and the number of patent applications per innovation. This work was possible to conduct because new data is available in a recent French survey on the use of the patent system and the appropriation of innovation benefits (EFAT survey). The survey provides the share of innovations that are patented at the firm level. The availability of both this measure and the number of patent applications at the firm level allows us, though with some limitations, to split the analysis into the two aspects mentioned above.

The paper is organized as follows. The data are presented in section 2. In section 3 we describe our model and the econometric methods. The results are presented and commented upon in section 4, while section 5 provides a few conclusions.

2 The Data

2.1. Sample Construction

The sample results from matching five data sets. Three of these data sets convey information on the innovative activity of industrial firms, while the last two provide accounting information.

The first data set is the "Enquête Appropriation" (French technological appropriation survey, known as EFAT), where all the variables are measured at the overall firm level. This survey is a part of an international project on the study of technological appropriation and was conducted in the United States, Japan and Europe. The questionnaires vary slightly depending on the country. The French survey was addressed in 1993 by the SESSI [2] to a representative sample of more than two thousand firms of at least 50 people in manufacturing. One of its purposes was to collect information on innovation and the use of the patent system. The survey was conducted at the firm level. It was sent to the R&D manager of the firm. The response rate was approximately 70%. Half of the respondents declared that their firm had conducted innovative activities, which represents 996 firms. We first

2. Statistics Department of the Ministry of Industry.

extracted from this sample all firms that had introduced product innovations between 1990 and 1992 and applied for at least one patent during this period. Patent applications could have been made in any patent system, including in France. The reason why the sample is limited to patent applicants is that only these firms were asked to answer questions on the patent system. This limitation reduced the sample to 546 firms.

The responses to questions in the Appropriation survey provide most of the exogenous variables used in the econometric analysis. They include the reasons for patenting and an evaluation of the deficiencies of the patent system. They also include a question on the postponement of patent applications. The questions are not related to any specific national or regional patent system. All qualitative questions on the evaluation of the patent system were asked for in absolute terms, and are not compared to an industry norm. This survey also provides information on the share of product innovations for which a patent application was made. Notice that the question refers to patent applications and not to patents granted.

The second data set is the "Enquête Recherche" (Research survey) which is an annual survey that collects information on the R&D inputs of firms with at least one full-time employee in research activities. We selected all firms that declared an internal R&D activity at least once over 1990-1992. For each we computed the deflated average R&D expenditures over the available years.

The third data set is the EPAT (European PATent) data set. It includes information on all patent applications since the creation of the European Patent system in 1978. We used this data set to compute the firm-level number of European patent applications during the period 1990-1992. This variable is the second endogenous variable in the model. We concentrate on patent applications – and not on patent grants – because the second endogenous variable, the percentage of innovations that are patented, is defined on patent applications.

The other two data sets provide accounting information. The fourth data set is the "Enquête Annuelle d'Entreprises" (EAE, Industry census) from which we extracted sales in 1992 and the primary industry of the firm. The fifth data set is the line-of-business industry census (EAE "fractions") which gives the decomposition of sales between different industry lines for diversified firms. As in a previous study, we computed the domestic average market share of the firm, the Herfindahl index of industrial concentration and a diversification index at the firm level (CRÉPON, DUGUET, KABLA, [1996]). The definitions are given in appendix 3.

The final data set includes 299 firms. Since there is a strong selection, we computed the share of total industry sales due to the firms in the sample. The sample is mostly representative of 6 industries : transportation equipment (49%), basic metals (45%), computer and electronics (36%), chemical products (24%), instruments (24%) and non electrical machinery (15%). These figures are high if we consider that our sample represents a

3. The detailed industry classification is given in appendix 1.

little more than 1% of the number of firms in manufacturing [4]. Then, the conclusions in this study apply to relatively large R&D performers.

2.2. **Sample Statistics**

Table 1 shows the distribution of the propensity to patent. About half of the sample applied for patents for less than a fifth of their product innovations. Although the degree of novelty of each innovation plays an important role in these figures, it is not the only one, essentially because the French patent system is not very restrictive. Each firm assessment of the relative advantages and disadvantages of patenting certainly plays an important role in this outcome.

Table 2 gives the distribution of the number of European patent applications. Almost 40% of the firms in the sample did not apply for a European patent during the period under study, while almost 20% of them applied for at least 10 patents. This is a standard profile among R&D performers in French manufacturing (CRÉPON and DUGUET, [1997b]). All the firms in the sample declared that they had applied for at least one patent during the period 1990-92, but some of the firms have applied in the French system only. Another reason why we observe many firms with a small number of patents is that part of them patent only a small fraction of their innovations. One important aim of our work is to disentangle the determinants of the propensity to patent from the factors that influence the number of patents.

Table 3 gives, for each decile of R&D expenditures, the average number of patents, the number of patents per million French francs (FRF) spent on R&D, and the percentage of patented innovations for each R&D decile [5]. The average number of patents strongly increases with R&D, from 0.6 patents in the first decile to 80 patents in the last one. This confirms the positive relationship between R&D and the number of patents often identified in the literature (GRILICHES, [1990]). The second column shows a decrease in the number of patents per million FRF spent on R&D, which does not mean that patent output decreases with R&D but rather that the value of a patent differs among firms with different R&D budgets [6]. The percentage of patented innovations shows less variation with R&D than the number of patents, remaining around 20% over the entire sample, except for the last two R&D deciles where about half of all innovations are patented.

4. There are about 25000 firms of at least 20 people in French manufacturing.

5. In order to compute the average propensity to patent, we set the propensity to patent of a firm equal to the center of its interval (see Table 1, only intervals are available in the Appropriation survey). We do not make this assumption in the regressions.

6. In a companion paper (DUGUET and IUNG, [1997]), based on European patent renewals, we show that firms with a bigger R&D budget support longer patent lives than other firms, once controlled for size, sector, diversification and market share differences. This suggests that the value of their patents is higher, following the model by PAKES and SCHANKERMAN [1986].

Most of the R&D is undertaken in three sectors (see Table 4): computers and electronics (38%), ships aircraft and rail (33%), and motor vehicles (14%), followed by drugs (5%) and chemicals (4%). Table 5 gives the industry averages [7] for the propensity to patent, patents per million R&D, the R&D to sales ratio and diversification. The propensity to patent is higher (between 34 and 38%) in the most innovative sectors, which possibly reflects the novelty requirements. The number of patents per million R&D is the highest in motor vehicles, which is likely to reflect an ability to patent components more easily than in other sectors. But these figures may also reflect differences in the patenting propensity across sectors. The R&D to sales ratio is the highest in ships, aircraft and rail (11%) followed by computers and electronics (10%) and drugs (7%). Finally, we also provide a diversification index. We find that the most diversified firms are in chemicals (2.33), rubber and plastics (1.98), instruments (1.65) and drugs (1.64). At the other end, motor vehicles and non metallic products (both 1.3) are the least diversified sectors.

How do these innovative firms use the patent system? What are the reasons why these firms patent? What deficiencies of the patent system do they fear the most? What are the main differences between industries in the assessment of the patent system? Tables 6 and 7 contain descriptive statistics of the reasons for patenting and patent deficiencies for each industry [8].

The most frequent motivation for patenting is to prevent imitation, which is cited by 92% of the firms. Avoiding litigation initiated by competitors and using patents in technology negotiations are rated second (62% of the respondents). The other reasons are noted by half as many firms: earning license revenue, using patents to enter foreign markets where licensing to a domestic firm is required, and rewarding or evaluating researchers. Some differences appear between industries. Preventing imitation is strong in all sectors, but it is not the case of the two other important motivations for patenting. Avoiding litigation initiated by competitors concerns mostly basic metals (87%), motor vehicles (86%), chemical products (71%), rubber and plastics (76%), instruments (67%) and computers and electronics (65%). This includes most of the innovative sectors (except aircraft, at 53%), where appropriation considerations are likely to play an important role in the patenting decisions. The result is slightly different with technology negotiation. Here the issue is partly linked to R&D cooperation. The industries that value technology negotiation are ships, aircraft and rail (88%), drugs (75%), fabricated metals (70%), and motor vehicles and basic metals (both 67%). Indeed, we found in a previous study on European patents (DUGUET, [1995]) that motor vehicles, chemicals and equipment goods

7. Notice that we do not compute the ratios on the sectoral aggregates but that we take the average of the firm level ratios.

8. The motives for patenting differ according to the country where the survey took place. Compared to European surveys, the Japanese and the American surveys propose an additional motivation to patent: preventing other firms from patenting related inventions (or "blocking"). In the American survey, this motive appeared to be second in importance, after the prevention of copying (COHEN, NELSON and WALSH, [1997]).

account for most European *joint* patents (i.e., patents applied for by several firms at the same time) in French manufacturing. In these sectors, patenting to acquire a strength in technology negotiations and jointly patenting with competitors could represent two alternative strategies for a similar purpose.

The patent deficiencies variables are based on answers to a five-point scale (0 to 4) of the importance of each deficiency. We indicate the mean score for each industry [9]. The strongest limitations are the inability of patents to prevent imitation (2.2) and the disclosure of too much information (1.96). But the costs to get the patent, maintain and defend it are also important, although slightly less than the two other limitations. The industries in which the failure to prevent imitation is of comparatively little importance include instruments (1.67), drugs (1.75) and motor vehicles (1.95), while other innovative sectors such as computers and electronics (2.19) and aircraft (2.35) find this to be a much more important limitation. Notice that the answer seems to be related to the kind of goods produced: more homogenous goods could be easier to imitate than vertically differentiated ones. This could help to explain why patents are less effective in preventing imitation in non metallic mineral products (2.85), fabricated metals (2.35) and basic metals (2.33). The second most important patent deficiency is disclosure. This is a problem when the technology is easy to circumvent from the patent documents. The highest score is for non metallic products (2.77), followed by basic metals (2.47) but high scores are also found in more innovative industries such as chemical products (2.29), drugs (2.07), aircraft (1.94) and computers and electronics (1.89). The industries where this deficiency is the lowest are instruments (1.47) and motor vehicles (1.48). This suggests that, in these two sectors, it would be more difficult to infer a full product from the patent documents, or that other ways to get access to the technology are available to competitors. The assessment of the problem of patent disclosure does not depend on the complexity of the product only (e.g., aircraft), since it also depends on the degree of competition in the final market.

3 The Model

We estimate the following two equations model:

$$\begin{cases} p_i^* = \exp(x_{1i}\pi_1 + v_{1i}) \\ E(n_i|p_i^*, x_{2i}, u_{2i}) = p_i^* \exp(x_{2i}b_2 + u_{2i}) \end{cases} \quad i = 1, \dots, m$$

9. This type of ordinal data is problematic because of possible differences in the appraisal of the scale by each correspondent. The transformation of the data into dummies using a given threshold allows one to partly solve this problem but at the cost of a loss of information. In the regressions, we compare these two approaches in order to examine the robustness of the results: the first variables used are the original variables of the survey (levels 0 to 4) and the second ones are the corresponding dummies for the highest value (equal to 1 if the answer is 4, 0 otherwise). The statistics that we present in Table 7 use the first convention: the mean of the original variables by industry.

Table 1

Propensity to Patent 1990-1992.

Patenting propensity	Number of firms	Percentage
$0 < p_i^* \leq 0.2$	151	50.5
$0.2 < p_i^* \leq 0.4$	51	17.1
$0.4 < p_i^* \leq 0.6$	36	12.0
$0.6 < p_i^* \leq 0.8$	34	11.4
$0.8 < p_i^* \leq 1$	27	9.0
Totals	299	100.0

Table 2

Number of European Patent Applications 1990-1992.

Number of patent applications	Number of firms	Percentage
0	118	39.5
1	38	12.7
2	28	9.4
3	21	7.0
4 to 10	26	8.7
11 to 20	24	8.0
21 to 50	26	8.7
51 and more	18	6.0
Totals	299	100.0

Table 3

R&D Decile Average.

R&D decile	Average number of patents	Number of patents per million R&D FRF	Average propensity to patent %
1	0.6	0.38	23.3
2	0.9	0.31	29.3
3	1.8	0.39	26.0
4	1.5	0.23	22.0
5	3.1	0.37	35.3
6	3.0	0.19	25.7
7	6.1	0.21	31.3
8	12.1	0.22	30.0
9	23.7	0.17	46.0
10	80.2	0.10	51.4
Sample average	13.1	0.26	32.3

TABLE 4

Industry Contributions.

Industry *% of the sample total*	Number of firms %	Sales %	R&D M RFR
Chemical products	5.7	6.2	4.3
Drugs	9.4	5.5	5.1
Rubber and plastics	5.7	1.7	0.5
Non metallic products	4.3	2.4	0.5
Basic metals	5.0	10.2	1.7
Fabricated metals	7.7	1.0	0.3
Non-electrical machinery	17.7	5.3	1.9
Computers and electronics	21.1	22.0	37.7
Ships, aircraft and rail	5.7	9.8	32.7
Motor vehicles	7.0	32.8	14.2
Instruments	5.0	1.4	0.6
Textile, wood, paper and others	5.7	1.7	0.5
Totals (= 100%)	299 firms	759,423 M	46,224 M

where p_i^* is the share of innovations for which patents have been applied for, n_i the number of European patent applications and m the number of firms. The first relationship explains the propensity to patent by covariates x_1 and the second relationship explains the expected number of European patent applications over 1990-92, which has two components. The first component is p_i^*, the endogenous percentage of innovations that are patented; the second component $\exp(x_{2i}b_2 + u_{2i})$ is the number of innovations times the number of patent applications per patented innovation. This model can be written under the log-linear form:

$$\begin{cases} \ln p_i^* = x_{1i}\pi_1 + v_{1i} \\ \ln E(n_i | p_i^*, x_{2i}, u_{2i}) = \ln p_i^* + x_{2i}b_2 + u_{2i} \end{cases} \quad i = 1, \ldots, m$$

We estimate this model in two steps, by asymptotic least square. This method is close to what is known in the econometric literature as Amemiya's method [10] (LEE, [1981]). First, we estimate the reduced form of the model and, secondly, we infer the structural parameters from the reduced form parameters. In the asymptotic least squares terminology the reduced form parameters are the auxiliary parameters and the structural form parameters are the parameters of interest.

The estimation of the reduced form is as follows. The first equation of the model is already under its reduced form, therefore the only estimation problem is that we do not observe p_i^* but the interval it lies in. It implies

10. The difference is that we combine maximum likelihood and pseudo maximum likelihood estimators in the first step, so that the estimation of the reduced form is different from the standard one (namely, maximum likelihood).

TABLE 5

Industry Averages.

Industry *Averages on firm level data*	Propensity to patent %	Patents per M. R&D FRF	R&D/Sales %	Diversification
Chemical products	28.8	0.14	4.9	2.33
Drugs	33.6	0.09	7.0	1.64
Rubber and plastics	21.8	0.21	1.9	1.98
Non metallic products	22.3	0.16	2.2	1.29
Basic metals	28.7	0.27	1.3	1.55
Fabricated metals	32.6	0.43	3.2	1.54
Non-electrical machinery	31.5	0.34	3.3	1.39
Computers and electronics	38.6	0.19	9.7	1.50
Ships, aircraft and rail	37.0	0.10	11.0	1.57
Motor vehicles	33.8	0.59	3.9	1.30
Instruments	36.7	0.28	3.2	1.65
Textile, wood, paper and others	22.9	0.24	1.8	1.35
Sample	32.3	0.26	5.7	1.55

TABLE 6

Reasons for Patenting
(share of affirmative answers).

Industry *Averages on firm level data*	Preventing imitation	Avoiding trials	Technology negotiation	License fees	Rewarding researchers	Entering foreign markets
Chemical products	0.82	0.71	0.47	0.29	0.12	0.24
Drugs	0.89	0.46	0.75	0.43	0.14	0.36
Rubber and plastics	0.88	0.76	0.59	0.24	0.18	0.24
Non metallic products	0.92	0.46	0.54	0.31	0.23	0.23
Basic metals	1.00	0.87	0.67	0.40	0.33	0.27
Fabricated metals	0.96	0.43	0.70	0.26	0.04	0.26
Non-electrical machinery	0.96	0.42	0.45	0.21	0.09	0.23
Computers and electronics	0.92	0.65	0.65	0.25	0.35	0.21
Ships, aircraft and rail	0.94	0.53	0.88	0.29	0.24	0.12
Motor vehicles	0.86	0.86	0.67	0.33	0.05	0.33
Instruments	0.93	0.67	0.60	0.07	0.13	0.13
Textile, wood, paper and others	0.94	0.59	0.53	0.29	0.29	0.35
Manufacturing	0.92	0.62	0.62	0.28	0.18	0.25

TABLE 7

Patent Deficiencies
(mean scores).

Industry *Averages on firm level data*	Costly to get and to maintain	Costly to defend	Does not prevent imitation	Too much disclosure
Chemical products	1.76	1.94	2.29	2.29
Drugs	1.71	1.54	1.75	2.07
Rubber and plastics	2.12	1.76	2.06	1.82
Non metallic products	2.38	1.92	2.85	2.77
Basic metals	1.27	1.73	2.33	2.47
Fabricated metals	2.09	2.22	2.35	1.65
Non-electrical machinery	2.30	2.21	2.30	1.92
Computers and electronics	2.14	1.86	2.19	1.89
Ships, aircraft and rail	1.76	1.47	2.35	1.94
Motor vehicles	1.38	1.81	1.95	1.48
Instruments	2.00	1.67	1.67	1.47
Textile, wood, paper and others	1.53	1.88	2.00	1.71
Manufacturing	1.87	1.83	2.17	1.96

that we cannot use OLS. The "Appropriation" survey provides a *qualitative* variable p_i that is equal to (table 1):

$$p_i = \begin{cases} 1 & \text{if } a_0 < \ln p_i^* \leq a_1 \\ 2 & \text{if } a_1 < \ln p_i^* \leq a_2 \\ 3 & \text{if } a_2 < \ln p_i^* \leq a_3 \\ 4 & \text{if } a_3 < \ln p_i^* \leq a_4 \\ 5 & \text{if } a_4 < \ln p_i^* \leq a_5 \end{cases} \quad i = 1, \ldots, m$$

where $a_0 = \{-\infty\}$, $a_1 = \ln 0.2$, $a_2 = \ln 0.4$, $a_3 = \ln 0.6$, $a_4 = \ln 0.8$ and $a_5 = \{+\infty\}$. In order to estimate this equation, we use the ordered probit model (see MADDALA, [1992]). This is equivalent to postulate that the distribution of the disturbances v_{1i} are identically and independently distributed as normal variates $N(0, \sigma_1)$. From this assumption, we can estimate the parameters by maximum likelihood [11].

The log-likelihood of the first equation comes directly from the underlying model; it is given by:

$$\ln L_1(p|x_1; \pi_1, \sigma_1) = \sum_{i=1}^{m} \sum_{k=1}^{5} d_{ik} \ln \Pr[p_i = k]$$

11. If this assumption does not hold, the estimates may be biased. Then, an extension of this work could look at the robustness of the results with different distributions or turn to non parametric techniques. If, on the contrary, the assumption holds, the maximum likelihood estimator presented below is asymptotically efficient.

with:

$$d_{ik} = \begin{cases} 1 & \text{if } p_i = k \\ 0 & \text{otherwise} \end{cases} \quad i = 1, \ldots, m \quad k = 1, \ldots, 5$$

and [12]:

$$\begin{aligned} \Pr[p_i = k] &= \Pr[a_{k-1} < \ln p_i^* \leq a_k] = \Pr[a_{k-1} < x_{1i}\pi_1 + v_{1i} \leq a_k] \\ &= \Pr\left[\frac{a_{k-1} - x_{1i}\pi_1}{\sigma_1} < \frac{v_{1i}}{\sigma_1} \leq \frac{a_k - x_{1i}\pi_1}{\sigma_1}\right] \\ &= \Phi\left[\frac{a_k - x_{1i}\pi_1}{\sigma_1}\right] - \Phi\left[\frac{a_{k-1} - x_{1i}\pi_1}{\sigma_1}\right] \quad i = 1, \ldots, m \end{aligned}$$

where Φ is the cdf of the standard normal distribution. We have estimated this first equation with SAS-IML software [13].

The second reduced form equation explains a positive integer dependent variable: the number of patent applications. Following the econometric literature on count data we use a heterogeneous Poisson model (GOURIÉROUX, MONFORT and TROGNON, [1984a]). For this second equation, we do not need to make a specific assumption on the distribution of the residual v_{2i}. We use the specification of the conditional mean of the distribution of the dependent variable only. Therefore, our estimates are robust to the distributional assumptions on the disturbance of the patent numbers equation. We estimate this equation by pseudo maximum likelihood. Based on previous studies on similar data (CRÉPON and DUGUET, [1995a, 1997a]), we use a negative binomial pseudo distribution.

We write the reduced form of the second equation as:

$$\ln E(n_i | p_i, x_i, u_{2i}) = x_i\pi_2 + v_{2i}$$

where x is the full column rank matrix of all exogenous variables in x_1 and x_2.

---

12. Taking the limits, we obtain the specific cases:

$$\Pr[p_i = 1] = \Phi[(a_1 - x_{1i}\pi_1)/\sigma_1] \quad \text{and} \quad \Pr[p_i = 5] = 1 - \Phi[(a_4 - x_{1i}\pi_1)/\sigma_1].$$

13. With known thresholds, the standard error of the disturbance is identified. We make the parameter change $h = 1/\sigma_1$ and $\beta = \pi_1/\sigma_1$ so that the log-likelihood is concave according to h and β. We use a Newton-Raphson algorithm with analytical Hessian. The original parameters (π_1, σ_1) and their covariance are calculated from Slutsky's theorem (so-called "δ-method"). The estimation program we use is %PROBITO, a SAS-IML macro command presented in a previous working paper and available from the authors (CRÉPON and DUGUET, [1995b]).

The pseudo log-likelihood to be maximized is thus [14]:

$$L_2(n|x;\pi_2) = \sum_{i=1}^{m} n_i x_i \pi_2 - (1+n_i)\ln(1+\exp(x_i\pi_2))$$

The second step of the estimation is to infer the *structural* parameters of the second equation. We do it by asymptotic least squares also called "minimum distance" (MALINVAUD, [1970]). First, we estimate the correlation matrix of the reduced form estimates. Since we have used different estimation methods and separate regressions, we need a unifying principle. We use M-estimation (see GOURIÉROUX and MONFORT, [1996]). Consider the following function:

$$L(p,n|x;\pi_1,\sigma_1,\pi_2) = L_1(p|x_1;\pi_1,\sigma_1) + L_2(n|x;\pi_2)$$

The first order conditions on this new objective function are exactly the same as the one implied by the separate maximum likelihood and pseudo maximum likelihood estimations of our reduced form [15]. Then, the global asymptotic distribution is the one of the M-estimator defined by the objective above. Let the auxiliary parameter be $\pi = (\pi_1\,\pi_2\,\sigma_1)$', the M-estimator is defined by [16]:

$$\widehat{\pi} = \arg\max_{\pi} L(p,n|x;\pi)$$

Under the usual regularity conditions, given in GOURIÉROUX and MONFORT [1996], the asymptotic distribution of this M-estimator of the auxiliary parameter is [17]:

$$\sqrt{m}(\widehat{\pi}-\pi) \xrightarrow{A} N(0, J^{-1}IJ^{-1})$$

with

$$J = \underset{0}{E}\,\underset{X}{E}\left[-\frac{\partial^2 L(p^*,n|x;\pi)}{\partial\pi\partial\pi'}\right]$$

14. The negative binomial distribution depends on a nuisance parameter θ. Its density is equal to:

$$f_\theta(n) = \frac{\Gamma(n+1/\theta)}{\Gamma(n+1)\Gamma(1/\theta)}\,(\theta\mu)^n\,(1+\theta\mu)^{-(n+1/\theta)}$$

where μ is the mean of the distribution. In our application we set $\theta = 1$ which does not affect the consistency of our estimates. For more details on count data pseudo maximum likelihood estimation see GOURIÉROUX, MONFORT and TROGNON [1984b]. We also use SAS-IML to estimate this equation, the program is %PMVGNEG and is also available. It uses a Newton-Raphson algorithm with analytical second order derivatives.

15. If there were constraints between the parameters of different equations, this property would not hold. Nevertheless, it would not create a severe problem since it is possible to impose the constraints at the second step of the estimation method. Thus, this method is more general than is seems.

16. This objective is concave.

17. Notice that, since the estimation problem is separable, we have the following simplifications:

$$\frac{\partial^2 L(p^*,n|x;\pi)}{\partial\sigma_1\partial\pi_{2'}} = 0 \qquad \text{and} \qquad \frac{\partial^2 L(p^*,n|x;\pi)}{\partial\pi_1\partial\pi_{2'}} = 0.$$

and

$$\underset{0}{E}\ \underset{x}{E}\left[\frac{\partial L(p,n|x;\pi)}{\partial \pi}\,\frac{\partial L(p,n|x;\pi)}{\partial \pi'}\right]$$

We estimate the J and I matrices by their sample counterparts:

$$\widehat{J} = -\frac{1}{m}\sum_{i=1}^{m}\frac{\partial^2 L(p,n|x;\widehat{\pi})}{\partial \pi \partial \pi'}$$

and

$$\widehat{I} = \frac{1}{m}\sum_{i=1}^{m}\frac{\partial L(p,n|x;\widehat{\pi})}{\partial \pi}\,\frac{\partial L(p,n|x;\widehat{\pi})}{\partial \pi'}$$

This method can be applied to any mix of well-behaved M-estimators. In the present case, the underlying model is linear so that the relationship between the interest (i.e., structural) parameters and the auxiliary (i.e., reduced form) parameters is obtained as in standard linear models. We have three cases:

(i) Variables present in both equations. If the variable, indexed j, is present in both equations the structural parameter of this variable is simply given by $b_2^j = \pi_1^j - \pi_2^j$.

(ii) Variables present in the first equation only: the parameter is given directly by the reduced form estimate of the first equation.

(iii) Variables present in the second equation only: $b_2^j = \pi_2^j$ and π_1^j must be estimated under this constraint. But it requires a strong exclusion assumption so that we do not use it here. In fact, what we want to do is to test which variable of the "Appropriation" survey is significant in each equation. Then, we use mostly the first case out of the three.

The derivation of the covariance matrix of the structural parameters is straightforward, following Slutsky's theorem.

4 The Results

We study the propensity to patent first, then the innovation function linking the number of patents to research expenditures. We made three groups of regressions, labeled A to C. Each group of regressions includes three tables giving the estimates of, respectively, the propensity to patent equation, the reduced form of the equation for the number of patents and the structural form of the patent numbers equation. The first group of regressions (A) does not control for industry effects. This is used as a benchmark to infer the biases that are likely to arise when the industry effects are omitted.

The first model (A) explains the propensity to patent and the number of patents by R&D expenditures, sales, the average market share, the average Herfindahl concentration index and the diversification index. Two sets of additional firm-level categorical variables, based on the EFAT survey, summarize the reasons to patent and the deficiencies: six dummies for the motivations to patent (see Table 6) and four ordinal variables for the patent deficiencies (see Table 7). We also use a dummy variable that indicates whether the firm has postponed a patent application over 1990-92 in order to achieve a more important innovation. The deficiency variables are five-points categorical variables. In the first regressions (A) they are used directly, without transformation. In the second group of regressions (B), we have added 11 industry dummies in order to control for industry effects. Then, we are able to see which patent deficiencies are really significant at the firm level. In the last group of regressions (C), the deficiency variables are based on dummies that capture the highest response level ("very important" deficiency). In this last group of regressions we compare the model with and without controlling for industry effects.

4.1. The Propensity to Patent

It clearly appears from the regressions that two main factors influence the propensity to patent an innovation: patent disclosure, whose impact is negative, and R&D expenditures, which positively influences this propensity (Tables 8, 11 and 14). A third variable, although with a weaker effect, seems to play a significant negative impact on the propensity to patent: the decision to postpone a patent application. However, it is weakened by the introduction of the industry effects. Another interesting result is the lack of significance of all the reasons to patent. Among the deficiencies, only disclosure is significant.

• Patent Disclosure

The main reason that influences the decision to patent an innovation is disclosure. The effect is present and strong whatever the measure of this deficiency we use (the five-level categorical variable or the dummy). The control for industry effects does not affect this result, although it slightly lowers the coefficient of the disclosure variable. The effect of patent disclosure is always significant at the 5% level in the selected [18] regressions (column 5 in Tables 8, 11 and 14).

The reason why patent disclosure may be such an obstacle to patenting is obvious. Patent disclosure makes the circumvention of innovations by competitors easier since it grants them useful technical knowledge. Although it may demand additional research effort from competitors in order not to infringe the invention, it increases their capacity to reproduce the patented innovation by lowering both the imitation lag and its cost. Yet a patent is supposed to forbid the marketing by competitors of product imitations. However, in numerous cases imitation could be hidden and difficult to detect, moreover the ability to invent around limits the effectiveness of patents. In addition, patent disclosure could be dangerous for the firm when the innovation is a technological breakthrough that can be followed by

18. When the strategic postponement dummy is added in the regression, this coefficient falls. However, the latter variable is not significant at the 5% level in the regression, so that we have to retain the column 5 results.

improvements. Although a competitor has no right, theoretically, to market any improvement on the original patent without taking a license on it, the pioneering firm may believe that it will not be able to enforce its rights through litigation. Moreover, it would be unable to block an improvement based upon a non-infringing imitation of its patented invention.

• R&D Expenditures

The level of R&D expenditures has a positive impact on the propensity to patent. However, the effect cannot be assimilated to a pure size effect, since sales are never significant. Moreover, it does not capture sectoral differences since the addition of industry dummies only slightly decreases the R&D coefficient from around 0.19 to 0.15 (see Table 14). It is always significant at the 5% level. Several interpretations can be given to this result. First, we could attribute this effect to the relationship between the research effort of the firm and the magnitude of its innovations or fixed costs in R&D. Second, the highest R&D budgets are found among firms that already have a technical advantage over their competitors, so that they are likely to benefit from better appropriability conditions allowing them to patent a higher share of their innovations. Third, the R&D budget effect could tell us that the larger firms in our sample have internal legal departments responsible for patenting, thus diminishing its cost.

Since R&D, sales and market share logarithms are strongly correlated, we have performed some additional tests presented in appendix 4. The first point to examine is whether R&D captures a size effect or a R&D effect. The second point is simultaneity between the propensity to patent and research investments. From the regressions, it clearly appears that size is significant only when it is the only right-hand variable. Then our results imply that R&D matters, independently from size. For simultaneity, we replace the average R&D expenditures over 1990-1992 by lagged R&D capital in 1989. The results are not affected by this change as well. Overall, what we find is that all the effect of size passes through R&D, not directly.

The elasticity of the patenting propensity to R&D expenditures lies between 0.21 and 0.14, depending on the regression. In regression (3) of Table 11, for instance, where industry dummies and patent deficiency variables are included, the coefficient is around 0.16. Thus a twice bigger R&D budget is associated to a 16% higher patenting share. This effect is relatively low compared to the one of disclosure. Its coefficient in the same regression is around –0.15 and this variable is on a five points scale, from 0 to 4. Thus the difference between firms that suffer the most from disclosure and the ones that do not is $0.15 \times 4 = 60\%$ of the patenting share [19]. This explains stronger differences than the research budget. We can illustrate this point by the following comparison: in order to compensate a simple move from 0 to 1 on the disclosure categorical variable, a firm would need to multiply its R&D budget by a factor $2.5 = \exp(0.15/0.16)$.

---

19. The specification with the corresponding dummy variable gives a close figure of 50% (table 14).

This suggests that under-patenting results mainly from disclosure: the dissemination of knowledge through patents may be strongly undermined by appropriability problems.

• Postponing Patent Applications

The decision of firms to postpone their patent applications in order to achieve more advanced versions of their innovations seems to decrease the proportion of innovations that are patented. The risk that a competitor will patent a similar innovation during the period probably explains the negative sign of the coefficient. The estimated coefficient is –0.30 and significant at the 5% level when it is the only exogenous variable in the model related to the evaluation of patent (column 4 of Table 8). It is –0.24 and significant at the 10% level when industry dummies are added (column 4 of Table 11). This indicates that a part of the effect is sectoral: the low patenting propensity of some sectors could find its origin in postponement tactics.

However the effect is no longer significant when other patent variables are added (column 6 in Tables 8, 11 and 14). The effect is, in fact, captured by the disclosure variable. Simultaneously, the presence of the postponement variable tends to decrease the coefficient of the disclosure variable. Indeed both variables are strongly correlated, and one may infer that patent disclosure is precisely one of the reasons to postpone a patent. One aspect of patent disclosure is that it helps to invent around innovations that are not adequately protected by patents. The quality of the protection strongly depends on the ability of the applicant to associate precise claims with its invention. This ability increases with the maturity of the invention.

• No Effect of the other Patent Motives and Deficiencies

No alleged advantage of patenting seems to make patent applications worthwhile. The unsignificance of the first objective, to prevent imitation, is of particular interest. It is mostly linked to the very high number of firms that do mention this objective, so that it does not explain much variations between firms. Indeed, the coefficient is rather high (around 0.3) but never significant at the 10% level. Therefore, there could be a measurement problem due to the fact that the variables corresponding to the objectives are dummies and contain less information than the other variables used in the regressions. The availability of 5-levels categorical variables for the motives could perhaps alter this result.

Neither patenting costs nor the costs of legal action have a significant effect on the propensity to patent. Firms do not appear to base their decision to patent on the patent costs [20]. Our results suggest looking for other explanatory variables.

20. This result may also be linked to the low European patent renewal fees, at least during the first years. For France, it rarely goes over FRF 2000 per year and per patent (about USD 400).

TABLE 8

Propensity to Patent
Model A

Left-hand variable: percentage of innovations patented (logarithm)
With a constant term
Maximum likelihood estimation of the ordered probit model
(Asymptotic t statistic between parentheses).

Variables	(1)	(2)	(3)	(4)	(5)	(6)
R&D Expenditures	0.205	0.189	0.186	0.213	0.172	0.179
(logarithm)	(3.80)	(3.37)	(3.32)	(3.94)	(2.96)	(3.09)
Sales	–0.080	–0.096	–0.079	–0.079	–0.098	–0.101
(logarithm)	(0.89)	(1.05)	(0.84)	(0.88)	(1.03)	(1.27)
Average market share	0.100	0.101	0.095	0.104	0.097	0.101
(logarithm)	(1.35)	(1.34)	(1.20)	(1.38)	(1.23)	(1.28)
Average concentration	–0.032	–0.005	–0.009	–0.039	0.010	0.004
(logarithm)	(0.41)	(0.05)	(0.11)	(0.49)	(0.12)	(0.05)
Diversification	–0.006	–0.029	–0.031	–0.033	–0.049	–0.064
(logarithm)	(0.04)	(0.17)	(0.18)	(0.20)	(0.29)	(0.37)
Reasons for patenting:						
(dummies 0/1)						
1) Preventing imitation		0.402			0.268	0.270
		(1.80)			(1.11)	(1.09)
2) Avoiding trials		0.077			0.099	0.106
		(0.59)			(0.73)	(0.78)
3) Technology negotiation		0.10			–0.013	–0.001
		(0.07)			(0.09)	(0.01)
4) License fees		0.211			0.219	0.210
		(1.51)			(1.52)	(1.45)
5) Rewarding researchers		–0.120			–0.030	–0.038
		(0.73)			(0.17)	(0.21)
6) Entering foreign market		0.095			0.086	0.099
		(0.65)			(0.58)	(0.66)
Patent deficiencies:						
(categorical, 0/4)						
1) Costly to maintain			–0.029		–0.018	–0.019
			(0.43)		(0.26)	(0.27)
2) Costly to defend			0.008		–0.001	–0.001
			(0.12)		(0.01)	(0.01)
3) Does not prevent imitation			–0.056		–0.054	–0.061
			(0.79)		(0.73)	(0.82)
4) Too much disclosure			–0.164		–0.158	–0.134
			(2.69)		(2.51)	(2.09)
Postponing an application				–0.300		–0.207
(dummy 0/1)				(2.33)		(1.49)
$\widehat{\sigma}_1$	1.079	1.067	1.045	1.049	1.037	1.033
	(19.3)	(18.4)	(16.6)	(17.2)	(15.6)	(15.6)

TABLE 9

Reduced Form of the Patent Equation
Model A

Left-hand variable: number of patents
With a constant term
Pseudo maximum likelihood estimation of the heterogeneous Poisson model
Negative binomial pseudo distribution
(Asymptotic t statistic between parentheses).

Variables	(1)	(2)	(3)	(4)	(5)	(6)
R&D Expenditures	0.613	0.566	0.601	0.637	0.564	0.577
(logarithm)	(7.96)	(7.45)	(7.91)	(8.61)	(7.52)	(7.80)
Sales	0.288	0.282	0.334	0.288	0.316	0.309
(logarithm)	(2.12)	(2.14)	(2.40)	(2.17)	(2.41)	(2.41)
Average market share	0.024	0.054	0.049	0.021	0.076	0.072
(logarithm)	(0.18)	(0.50)	(0.42)	(0.16)	(0.78)	(2.41)
Average concentration	–0.095	–0.075	–0.107	–0.109	–0.077	–0.085
(logarithm)	(0.86)	(0.79)	(1.04)	(1.00)	(0.86)	(0.94)
Diversification	0.422	0.410	0.425	0.366	0.415	0.376
(logarithm)	(1.83)	(1.90)	(1.96)	(1.61)	(2.03)	(1.87)
Reasons for patenting:						
(dummies 0/1)						
1) Preventing imitation		0.259			0.195	0.161
		(0.80)			(0.60)	(0.48)
2) Avoiding trials		0.487			0.495	0.532
		(2.85)			(2.89)	(3.15)
3) Technology negotiation		0.724			0.687	0.697
		(3.60)			(3.45)	(3.56)
4) License fees		0.040			–0.059	–0.046
		(0.21)			(0.30)	(0.24)
5) Rewarding researchers		–0.097			–0.080	–0.060
		(0.43)			(0.36)	(0.26)
6) Entering foreign markets		–0.041			0.021	0.035
		(0.21)			(0.11)	(0.19)
Patent deficiencies:						
(categorical, 0/4)						
1) Costly to maintain			–0.026		–0.012	–0.005
			(0.32)		(0.14)	(0.06)
2) Costly to defend			0.143		0.118	0.100
			(1.72)		(1.42)	(1.23)
3) Does not prevent imitation			0.042		0.053	0.048
			(0.49)		(0.64)	(0.58)
4) Too much disclosure			–0.262		–0.245	–0.204
			(3.12)		(3.14)	(2.55)
Postponing an application				–0.416		–0.321
(dummy 0/1)				(2.26)		(1.76)

TABLE 10

Structural Form of the Patent Equation
Model A

Left-hand variable: number of patents
With a constant term
Asymptotic least squares
(Asymptotic t statistic between parentheses).

Variables	(1)	(2)	(3)	(4)	(5)	(6)
R&D Expenditures	0.409	0.376	0.415	0.424	0.392	0.339
(logarithm)	(4.70)	(4.64)	(4.83)	(5.11)	(4.78)	(4.24)
Sales	0.369	0.378	0.413	0.366	0.392	0.409
(logarithm)	(2.56)	(2.76)	(2.72)	(2.58)	(2.76)	(2.94)
Average market share	–0.077	–0.046	–0.046	–0.083	–0.022	–0.029
(logarithm)	(0.53)	(0.37)	(0.35)	(0.59)	(0.19)	(0.25)
Average concentration	–0.063	–0.069	–0.098	–0.069	–0.087	–0.089
(logarithm)	(0.53)	(0.67)	(0.85)	(0.59)	(0.85)	(0.89)
Diversification	0.429	0.439	0.456	0.399	0.464	0.440
(logarithm)	(1.74)	(1.90)	(1.89)	(1.63)	(2.06)	(1.97)
Reasons for patenting:						
(dummies 0/1)						
1) Preventing imitation		–0.143			–0.074	–0.109
		(0.47)			(0.23)	(0.33)
2) Avoiding trials		0.410			0.396	0.426
		(2.24)			(2.11)	(2.30)
3) Technology negotiation		0.714			0.700	0.699
		(3.47)			(3.41)	(3.46)
4) License fees		–0.171			–0.277	–0.256
		(0.92)			(1.40)	(1.31)
5) Rewarding researchers		0.023			–0.049	–0.022
		(0.10)			(0.20)	(0.09)
6) Entering foreign markets		–0.136			–0.066	–0.116
		(0.64)			(0.31)	(0.53)
Patent deficiencies:						
(categorical, 0/4)						
1) Costly to maintain			0.004		0.006	–0.064
			(0.05)		(0.07)	(0.30)
2) Costly to defend			0.135		0.119	0.101
			(1.52)		(1.42)	(1.25)
3) Does not prevent imitation			0.098		0.108	0.109
			(1.08)		(1.19)	(1.20)
4) Too much disclosure			–0.097		–0.086	–0.071
			(1.04)		(0.97)	(0.82)
Postponing an application				–0.116		–0.114
(dummy 0/1)				(0.63)		(0.63)

TABLE 11

Propensity to Patent
Model B

Left-hand variable: percentage of innovations patented (logarithm)
With 12 industry dummies
Maximum likelihood estimation of the ordered probit model
(Asymptotic t statistic between parentheses).

Variables	(1)	(2)	(3)	(4)	(5)	(6)
R&D Expenditures	0.175	0.144	0.163	0.181	0.134	0.139
(logarithm)	(2.73)	(2.18)	(2.47)	(2.83)	(1.97)	(2.04)
Sales	0.005	–0.004	–0.007	0.003	–0.020	–0.023
(logarithm)	(0.05)	(0.04)	(0.06)	(0.03)	(0.19)	(0.21)
Average market share	0.053	0.057	0.047	0.060	0.051	0.056
(logarithm)	(0.61)	(0.64)	(0.51)	(0.69)	(0.55)	(0.60)
Average concentration	0.053	0.076	0.062	0.040	0.083	0.075
(logarithm)	(0.58)	(0.81)	(0.64)	(0.43)	(0.84)	(0.77)
Diversification	–0.008	–0.024	–0.033	–0.019	–0.046	–0.051
(logarithm)	(0.04)	(0.13)	(0.18)	(0.10)	(0.24)	(0.27)
Reasons for patenting:						
(dummies 0/1)						
1) Preventing imitation		0.366			0.223	0.233
		(1.54)			(0.87)	(0.90)
2) Avoiding trials		0.128			0.163	0.170
		(0.92)			(1.15)	(1.19)
3) Technology negotiation		0.010			–0.008	0.002
		(0.07)			(0.05)	(0.01)
4) License fees		0.242			0.242	0.232
		(1.61)			(1.58)	(1.51)
5) Rewarding researchers		–0.118			–0.040	–0.038
		(0.69)			(0.22)	(0.21)
6) Entering foreign markets		0.098			0.094	0.103
		(0.55)			(0.61)	(0.67)
Patent deficiencies:						
(categorical, 0/4)						
1) Costly to maintain			–0.045		–0.037	–0.035
			(0.63)		(0.49)	(0.47)
2) Costly to defend			–0.082		–0.021	–0.001
			(1.17)		(0.29)	(0.01)
3) Does not prevent imitation			–0.062		–0.064	–0.069
			(0.86)		(0.84)	(0.91)
4) Too much disclosure			–0.146		–0.140	–0.122
			(2.28)		(2.12)	(1.82)
Postponing an application				–0.241		–0.165
(dummy 0/1)				(1.72)		(1.10)
$\widehat{\sigma}_1$	1.055	1.041	1.025	1.049	1.014	1.012
	(17.3)	(16.5)	(15.3)	(17.2)	(14.7)	(14.7)

4.2. **The Number of Patent Applications**

The regressions point to several factors that influence the number of patent applications. We distinguish the reduced form of the patent equation (Tables 9, 12 and 15) from its structural form (Tables 10, 13 and 16). The difference is that the latter equation gives the number of patent applications controlled for the propensity to patent: this measure is closer to the number of innovations. However, we cannot disentangle the number of patent applications from the number of patents applications per patented innovation.

Some variables have an effect on both the reduced and structural form of the patent equation. These include R&D expenditures, the use of patents to avoid litigation and the use of patents to strengthen technology negotiations with other firms.

We also find that the significance of firm size, market share and diversification depends on the inclusion of industry dummies. Without industry effects, sales and diversification are significant, while market share is the only significant variable when industry differences are controlled for.

It is important to notice that the postponement variable affects the number of patents through its impact on the patenting propensity only and not directly. In other words, it has a significant coefficient in the reduced form of the model but it is not significant in the structural form. The patent disclosure variable remains significant in the structural form when it is taken as a dummy and not as a five-level variable (Table 16).

The following comments are restricted to the results with industry dummies in the regressions (Tables 12, 13, 15 and 16).

• Size-Linked Effects

An important issue is the R&D elasticity in the patent equation. That is, do we find constant returns in the "innovation production function"? Most cross-section studies exhibit unit elasticities (COHEN and LEVIN, [1989]), contrary to panel data studies. The latter find decreasing returns with coefficients varying from 0.3 to 0.6, depending on the estimation method used (HAUSMAN, HALL and GRILICHES, [1984]; BLUNDELL, GRIFFITH and WINDMEIJER, [1995]; CRÉPON and DUGUET, [1997b]). An explanation for this discrepancy is the propensity to patent: if it remains constant over time while differing between firms, it ends up in a fixed effect and the cross- section estimates that do not account for the differences in patenting propensities are biased. The original features of our data allow to correct for differences in the patenting propensity in a cross-section. The question is then: do we find research elasticities that are closer to the panel data estimates? The answer is positive.

The R&D elasticity (Tables 12-15) is around 0.6-0.7 in the reduced form. Once accounting for the differences in patenting propensities, it lowers to around 0.4-0.5 (Tables 13 and 16). Thus, our regressions suggest that the share of innovations that are patented would indeed intervene in the fixed effect, and would allow for a reduction of the coefficient to around 0.4-0.5. This result is consistent with the panel data estimates mentioned above. The regressions show two types of size effects. One is captured by the R&D

TABLE 12

Reduced Form of the Patent Equation
Model B

Left-hand variable: number of patents
With 12 industry dummies
Pseudo maximum likelihood estimation of heterogeneous Poisson model
Negative binomial pseudo distribution
(Asymptotic t statistic between parentheses).

Variables	(1)	(2)	(3)	(4)	(5)	(6)
R&D Expenditures	0.665	0.581	0.640	0.669	0.570	0.581
(logarithm)	(7.31)	(6.60)	(7.03)	(7.60)	(6.40)	(6.60)
Sales	0.029	0.064	0.078	0.034	0.095	0.085
(logarithm)	(0.20)	(0.48)	(0.52)	(0.23)	(0.69)	(0.63)
Average market share	0.265	0.283	0.281	0.271	0.290	0.296
(logarithm)	(2.23)	(2.60)	(2.36)	(3.11)	(2.64)	(2.69)
Average concentration	–0.062	–0.058	–0.093	–0.091	–0.067	–0.087
(logarithm)	(0.55)	(0.53)	(0.82)	(0.79)	(0.62)	(0.81)
Diversification	–0.001	–0.007	0.032	0.009	0.026	0.026
(logarithm)	(0.01)	(0.03)	(0.15)	(0.04)	(0.12)	(0.12)
Reasons for patenting:						
(dummies 0/1)						
1) Preventing imitation		0.257			0.220	0.219
		(0.83)			(0.75)	(0.74)
2) Avoiding trials		0.333			0.356	0.393
		(0.20)			(2.17)	(2.75)
3) Technology negotiation		0.790			0.764	0.784
		(4.46)			(4.29)	(4.38)
4) License fees		0.177			0.105	0.083
		(0.92)			(0.53)	(0.42)
5) Rewarding researchers		0.016			0.016	0.037
		(0.08)			(0.08)	(0.17)
6) Entering foreign markets		–0.091			–0.025	–0.013
		(0.49)			(0.13)	(0.07)
Patent deficiencies:						
(categorical, 0/4)						
1) Costly to maintain			–0.011		–0.014	–0.008
			(0.13)		(0.16)	(0.09)
2) Costly to defend			0.099		0.072	0.056
			(1.25)		(0.89)	(0.70)
3) Does not prevent imitation			0.019		0.022	0.020
			(0.23)		(0.27)	(0.24)
4) Too much disclosure			–0.183		–0.164	–0.137
			(2.86)		(2.22)	(1.80)
Postponing an application				–0.296		–0.276
(dummy 0/1)				(1.70)		(1.53)

TABLE 13

Structural Form of the Patent Equation
Model B

Left-hand variable: number of patents
With 12 industry dummies
Asymptotic least squares
(Asymptotic t statistic between parentheses).

Variables	(1)	(2)	(3)	(4)	(5)	(6)
R&D Expenditures	0.490	0.431	0.477	0.488	0.436	0.442
(logarithm)	(3.95)	(4.31)	(4.42)	(4.69)	(4.27)	(4.38)
Sales	0.024	0.068	0.085	0.032	0.116	0.108
(logarithm)	(0.15)	(0.47)	(0.51)	(0.20)	(0.75)	(0.71)
Average market share	0.211	0.225	0.234	0.211	0.239	0.240
(logarithm)	(1.77)	(2.01)	(1.93)	(2.43)	(2.08)	(2.11)
Average concentration	–0.116	–0.134	–0.154	–0.131	–0.149	–0.162
(logarithm)	(0.98)	(1.17)	(1.24)	(1.09)	(1.25)	(1.37)
Diversification	0.008	0.017	0.066	0.028	0.073	0.077
(logarithm)	(0.03)	(0.07)	(0.27)	(0.11)	(0.30)	(0.31)
Reasons for patenting:						
(dummies 0/1)						
1) Preventing imitation		–0.109			–0.03	–0.014
		(0.37)			(0.01)	(0.05)
2) Avoiding trials		0.205			0.192	0.223
		(1.13)			(1.04)	(1.23)
3) Technology negotiation		0.780			0.772	0.782
		(4.15)			(4.04)	(4.09)
4) License fees		–0.065			–0.136	–0.149
		(0.33)			(0.66)	(0.72)
5) Rewarding researchers		0.135			0.056	0.075
		(0.58)			(0.23)	(0.30)
6) Entering foreign markets		–0.189			–0.119	–0.116
		(0.89)			(0.055)	(0.53)
Patent deficiencies:						
(categorical, 0/4)						
1) Costly to maintain			0.034		0.023	0.027
			(0.36)		(0.24)	(0.28)
2) Costly to defend			0.091		0.070	0.089
			(1.01)		(0.80)	(0.96)
3) Does not prevent imitation			0.081		0.087	–0.015
			(0.91)		(0.94)	(0.16)
4) Too much disclosure			–0.037		–0.024	–0.111
			(0.41)		(0.27)	(0.61)
Postponing an application				–0.054		–0.112
(dummy 0/1)				(0.31)		(0.62)

expenditures variable while the other is captured by direct measures of firm size (sales, diversification, market share). Interestingly, when we control for industry effects, the significance of sales and diversification vanishes while the market share remains significant.

Given the impossibility to distinguish the number of innovations from the number of patents per patented innovation, there are two ways to interpret this market share effect: an effect on the number of innovations or an effect on the number of patent applications per innovation. On the one hand, it could be that the number of innovations increases with market power. On the other hand the multiplication of the number of patents may be a way to reinforce the monopoly power provided by the patent system.

• Technology Negotiations and Trials

Some of the reasons to patent have an influence on the corrected (i.e. structural form) patent count: a firm applies for more patents if it uses patents as a tool for negotiations or in order to prevent infringement suits from competitors.

The coefficients apply to dummy variables so that they directly give the differences in percentage between the firms that answered "Yes" and the others. Acquiring a strength in technology negotiations explains an average difference of 75% in the corrected patent number (Table 15, column 6). It is stronger than the effect of R&D (doubling R&D would increase the patent count by 44%). The second effect is also strong: the will to avoid trials by competitors entices firms to patent about 42% more [21].

This result probably means that using patents in negotiations is a motivation to increase the number of applications per innovation. Negotiations may take place with suppliers, customers, and even competitors. The reasons for negotiation may be to fix the license price or other relevant aspects in negotiations with a licensee, to exchange information on technologies under development with a customer or a supplier, to exchange rights with competitors for technologies, or to avoid litigation. Indeed a large portfolio of patents may be an advantage in these kind of negotiations. In the case of a lawsuit, the length and the probable cost of the litigation is correlated with the quantity of technical arguments that each party can provide. Alternatively, when two competing firms develop overlapping technologies, they may be interested in exchanging free licenses rather than entering a conflict.

---

21. The structural form coefficients given in Table 16 are misleading in this case, since when a coefficient is not significant in the patenting propensity equation, its reduced form estimate, given in Table 15, is the right one. This is equivalent to say that the best estimate is obtained by setting the insignificant coefficients to zero in the first (patenting propensity) equation.

TABLE 14

Propensity to Patent
Model C

Left-hand variable: percentage of innovations patented (logarithm)
Maximum likelihood estimation of the ordered probit model
(Asymptotic t statistic between parentheses).

	With constant term			With 12 industry dummies		
Variables	(1)	(2)	(3)	(4)	(5)	(6)
R&D Expenditures	0.190	0.176	0.183	0.168	0.141	0.145
(logarithm)	(3.45)	(3.03)	(3.10)	(2.51)	(2.04)	(2.10)
Sales	–0.051	–0.068	–0.067	0.026	0.013	0.011
(logarithm)	(0.57)	(0.73)	(0.71)	(0.25)	(0.12)	(0.10)
Average market share	0.094	0.094	0.098	0.044	0.050	0.056
(logarithm)	(1.21)	(1.21)	(1.26)	(0.48)	(0.54)	(0.61)
Average concentration	–0.015	0.011	0.003	0.063	0.084	0.072
(logarithm)	(0.19)	(0.13)	(0.04)	(0.68)	(0.89)	(0.77)
Diversification	0.006	–0.013	–0.036	0.003	–0.046	–0.017
(logarithm)	(0.04)	(0.08)	(0.20)	(0.02)	(0.24)	(0.09)
Reasons for patenting:						
(dummies 0/1)						
1) Preventing imitation		0.359	0.361		0.326	0.338
		(1.42)	(1.38)		(1.25)	(1.27)
2) Avoiding trials		0.063	0.075		0.117	0.129
		(0.47)	(0.55)		(0.83)	(0.91)
3) Technology negotiation		–0.011	0.008		–0.010	0.008
		(0.08)	(0.06)		(0.07)	(0.05)
4) License fees		0.223	0.210		0.240	0.224
		(1.55)	(1.45)		(1.57)	(1.45)
5) Rewarding researchers		–0.109	–0.108		–0.116	–0.108
		(0.65)	(0.64)		(0.67)	(0.62)
6) Entering foreign markets		0.146	0.155		0.137	0.147
		(0.99)	(1.03)		(0.90)	(0.96)
Patent deficiencies:						
(dummies on the highest level, 0/1)						
1) Costly to maintain	0.075	0.124	0.132	0.038	0.082	0.100
	(0.29)	(0.48)	(0.52)	(0.14)	(0.30)	(0.37)
2) Costly to defend	0.234	0.200	0.175	0.225	0.188	0.170
	(0.97)	(0.82)	(0.72)	(0.87)	(0.72)	(0.65)
3) Does not prevent imitation	0.104	0.192	0.186	0.095	0.169	0.159
	(0.43)	(0.77)	(0.74)	(0.38)	(0.66)	(0.61)
4) Too much disclosure	–0.592	–0.592	–0.457	–0.518	–0.500	–0.471
	(2.61)	(2.52)	(1.90)	(2.13)	(1.98)	(1.83)
Postponing an application			–0.241			–0.221
(dummy 0/1)			(1.80)			(1.50)
$\hat{\sigma}_1$	1.059	1.047	1.039	1.039	1.026	1.021
	(17.6)	(16.9)	(16.5)	(16.2)	(15.5)	(15.5)

TABLE 15

Reduced Form of the Patent Equation
Model C

Left-hand variable: number of patents
Pseudo maximum likelihood estimation of the heterogeneous Poisson model
Negative binomial pseudo distribution
(Asymptotic t statistic between parentheses).

	With constant term			With 12 industry dummies		
Variables	(1)	(2)	(3)	(4)	(5)	(6)
R&D Expenditures	0.593	0.541	0.564	0.643	0.568	0.581
(logarithm)	(7.80)	(7.21)	(7.73)	(7.31)	(6.53)	(6.76)
Sales	0.284	0.287	0.273	0.036	0.064	0.055
(logarithm)	(2.14)	(2.22)	(2.20)	(0.26)	(0.49)	(0.42)
Average market share	0.028	0.048	0.049	0.261	0.269	0.276
(logarithm)	(0.21)	(0.45)	(0.48)	(2.23)	(2.49)	(2.53)
Average concentration	–0.053	–0.040	–0.051	–0.071	–0.058	–0.084
(logarithm)	(0.48)	(0.42)	(0.55)	(0.63)	(0.54)	(0.78)
Diversification	0.405	0.401	0.346	0.021	0.002	0.007
(logarithm)	(1.82)	(1.93)	(1.72)	(0.10)	(0.01)	(0.03)
Reasons for patenting:						
(dummies 0/1)						
1) Preventing imitation		0.164	0.106		0.159	0.143
		(0.54)	(0.31)		(0.54)	(0.49)
2) Avoiding trials		0.501	0.551		0.366	0.416
		(3.04)	(3.36)		(2.27)	(2.60)
3) Technology negotiation		0.663	0.659		0.738	0.749
		(3.44)	(3.45)		(4.27)	(4.28)
4) License fees		0.077	0.073		0.175	0.135
		(0.42)	(0.41)		(0.94)	(0.72)
5) Rewarding researchers		–0.073	–0.054		0.020	0.036
		(0.34)	(0.24)		(0.10)	(0.17)
6) Entering foreign markets		–0.034	–0.001		–0.079	–0.048
		(0.19)	(0.01)		(0.44)	(0.27)
Patent deficiencies:						
(dummies on the highest level, 0/1)						
1) Costly to maintain	–0.320	–0.195	–0.198	–0.078	–0.064	–0.072
	(0.99)	(0.57)	(0.60)	(0.26)	(0.21)	(0.24)
2) Costly to defend	0.208	0.067	0.023	–0.011	–0.114	–0.132
	(0.64)	(0.23)	(0.08)	(0.04)	(0.45)	(0.53)
3) Does not prevent imitation	–0.039	0.097	0.087	–0.037	0.081	0.072
	(0.15)	(0.35)	(0.32)	(0.14)	(0.30)	(0.27)
4) Too much disclosure	–1.288	–1.272	–1.208	–1.154	–1.095	–1.064
	(5.07)	(5.34)	(5.14)	(4.87)	(5.14)	(5.07)
Postponing an application			–0.420			–0.332
(dummy 0/1)			(2.40)			(1.92)

TABLE 16

Structural Form of the Patent Equation
Model C

Left-hand variable: number of patents
Asymptotic least squares
(Asymptotic t statistic between parentheses).

	With constant term			With 12 industry dummies		
Variables	(1)	(2)	(3)	(4)	(5)	(6)
R&D Expenditures	0.403	0.365	0.380	0.475	0.428	0.436
(logarithm)	(4.63)	(4.45)	(4.75)	(4.52)	(4.32)	(4.45)
Sales	0.335	0.355	0.340	0.011	0.050	0.043
(logarithm)	(2.28)	(2.54)	(2.50)	(0.07)	(0.33)	(0.29)
Average market share	–0.066	–0.046	–0.049	0.217	0.219	0.220
(logarithm)	(0.46)	(0.38)	(0.42)	(1.84)	(1.97)	(1.98)
Average concentration	–0.038	–0.051	–0.054	–0.134	–0.143	–0.156
(logarithm)	(0.32)	(0.48)	(0.52)	(1.13)	(1.23)	(1.34)
Diversification	0.399	0.414	0.382	0.018	0.012	0.023
(logarithm)	(1.62)	(1.78)	(1.68)	(0.07)	(0.05)	(0.09)
Reasons for patenting:						
(dummies 0/1)						
1) Preventing imitation		–0.194	–0.255		–0.167	–0.196
		(0.58)	(0.73)		(0.52)	(0.60)
2) Avoiding trials		0.438	0.476		0.249	0.287
		(2.37)	(2.59)		(1.37)	(1.59)
3) Technology negotiation		0.675	0.651		0.748	0.740
		(3.34)	(3.29)		(4.04)	(3.98)
4) License fees		–0.146	–0.137		–0.065	–0.090
		(0.80)	(0.76)		(0.33)	(0.45)
5) Rewarding researchers		–0.177	0.054		0.136	0.144
		(0.84)	(0.23)		(0.59)	(0.62)
6) Entering foreign markets		–0.189	–0.156		–0.216	–0.195
		(0.90)	(0.75)		(1.02)	(0.91)
Patent deficiencies:						
(dummies on the highest level, 0/1)						
1) Costly to maintain	–0.395	–0.320	–0.330	–0.115	–0.147	–0.171
	(1.38)	(1.09)	(1.13)	(0.41)	(0.53)	(0.61)
2) Costly to defend	–0.027	–0.133	–0.153	–0.236	–0.302	–0.302
	(0.08)	(0.47)	(0.54)	(0.80)	(1.14)	(1.14)
3) Does not prevent imitation	–0.143	–0.05	–0.100	–1.131	–0.088	–0.086
	(0.49)	(0.30)	(0.32)	(3.90)	(0.28)	(0.28)
4) Too much disclosure	–0.695	–0.680	–0.661	–0.636	–0.595	–0.593
	(2.26)	(2.28)	(2.23)	(2.02)	(2.00)	(2.03)
Postponing an application			–0.155			–0.112
(dummy 0/1)			(0.87)			(0.62)

5 Conclusion

The patent numbers embed various components, including the share of innovations that are patented. The "Appropriation" (EFAT) survey allows researchers to correct the patent numbers for this effect and to study the patenting decision at the same time. But the patent numbers are counts and the share of patented innovations are interval data so that one has to use count data and limited dependent variables econometrics.

Using appropriate econometric methods we find that, on the one hand, R&D expenditures and patent disclosure are the main determinants of the share of innovations that are patented. Higher research budgets are associated to higher propensities to patent. On the other hand, the compulsory disclosure implied by the publication of patent documents strongly undermines the incentive to patent. It is not certain however that this implies a welfare loss, since if the disclosure requirement was weakened, less diffusion would occur as well.

Finally, along with research expenditures and market share, two variables newly available in the survey affect the corrected patent numbers: the willingness of firms to avoid trials and to reach a stronger position in technology negotiations with other firms. This clearly confirms that the patent numbers do not rely solely on an innovation function relating innovation inputs to outputs, but also depend on strategic aspects linked to patenting. Securing knowledge benefits seems to be the main problem industrial firms face.

APPENDIX 1

Industry Classification and Representatives of the Sample

The classification is close to OECD's. We give its definition for manufacturing, from the French NAP 100 classification. We also compare the total sales in our sample with that of all manufacturing in 1992 published in the *Tableaux de l'Economie Française* (INSEE, 1994, page 133). The most important selections are the presence in the R&D survey (Frascatti criteria) and to answer to the Appropriation survey.

Classification	NAP 100	Sales (M)		Sample
		Sample	Industry	%
Chemical products	17. Industrie chimique de base 43. Industrie fil et fibres artif. et synthétiques	47112	194788	24.2
Drugs	18. Parachimie 19. Industrie pharmaceutique	41780	273646	15.3
Rubber and plastics	52. Industrie du caoutchouc 53. Transformation des matières plastiques	12862	152024	8.5
Non-metallic products	14. Production de minéraux divers 15. Matériaux de construction, céramique 16. Industrie du verre	18084	158374	11.4
Basic metals	09. Extraction, préparation de minerai de fer 10. Sidérurgie 11. Première transformation de l'acier 12. Extraction et prépar. des minerais non ferreux 13. Métallurgie, 1re transf. des métaux non ferreux	77131	170292	45.3
Fabricated metals	20. Fonderie 21. Travail des métaux	7662	232810	3.3
Non-electrical machinery	22. Fabrication de machines agricoles 23. Fabrication de machines-outils 24. Production d'équipement industriel 25. Fabrication matériel de manutention	40639	270649	15.0
Computers and electronics	27. Fabrication de matériel informatique 28. Fabrication de matériel électrique 29. Matériel électronique ménager et professionnel 30. Fabrication d'équipement ménager	167180	462177	36.2

Classification	NAP 100	Sales (M)		Sample
		Sample	Industry	%
Shipping, aircraft and rail* Motor vehicles*	31B. Autre matériel de transport terrestre 32. Construction navale 33. Construction aéronautique 31A. Construction automobile	323153	652452	49.5
Instruments	34. Fabrication d'instruments de précision	10951	46226	23.7
Miscellaneous	44. Industrie textile 45. Industrie du cuir 46. Industrie de la chaussure 47. Industrie de l'habillement 48. Travail mécanique du bois 49. Industrie de l'ameublement 50. Industrie du papier et du carton 51. Imprimerie, presse, édition 54. Industries diverses	12869	683248	1.9

** No detailed information between industries 31A and 31B is available in the* Tableaux de l'Economie Française, *so that we have aggregated them in this table.*

The Patent Numbers Components

The number of patent applications per firm results from three components: the number of innovations, the share of innovations that is patented, and the number of patent applications for each innovation applied for.

In the following, let the subscript i denotes the firm. Let n_i be the number of patent applications of firm i, m_i its total number of innovations, δ_{ik} a dummy variable that equals one when the k-th innovation of firm i is applied for (0 otherwise), and μ_{ik} the number of patent applications for the corresponding innovation ($\delta_{ik} = 1$). Then, we have:

$$n_i = \sum_{k=1}^{m_i} \mu_{ik} = \sum_{k=1}^{m_i} \delta_{ik} \times \mu_{ik} = m_i \times \frac{1}{m_i} \sum_{k=1}^{m_i} \delta_{ik} \times \frac{\sum\limits_{k=1}^{m_i} \delta_{ik} \times \mu_{ik}}{\sum\limits_{k=1}^{m_i} \delta_{ik}} k = 1, \ldots, m_i$$

The first quantity on the right-hand side is the number of innovations of firm i, then comes the share of innovations that are patented $1/m_i \sum_{k=1}^{m_i} \delta_{ik}$ and, finally, the average number of patents per patented innovation.

The share of innovations that are patented is available in the Appropriation survey. Therefore, our econometric model allows to separate the effects of this variable only. The number of patent applications divided by the share of innovations that are patented, will be referred as *the corrected patent number.*

APPENDIX 3

Measurement of Diversification and Domestic Concentration

Accounting for diversification implies to change the market share and concentration measures, since a firm can operate in several industries at the same time. In this study we use weighted averages of the traditional indicators.

The line-of-business *Enquête Annuelle d'Entreprise* (so-called "Fractions EAE") allows us to compute concentration and diversification indicators. Let Q_{ik} be the domestic sales of firm i in industry k, its market shares are equal to:

$$s_{ik} = \frac{Q_{ik}}{\sum_{i=1}^{I} Q_{ik}}, \quad i = 1, \ldots, I \quad k = 1, \ldots, K$$

Notice that in the previous formula, the number of firm is $I > m$, that is we have been computing the market share from the whole data set (size I) and not only from our sample (size m). Here, a diversified firm has a different market share in each industry [22]. To perform our study we need an overall indicator of market power, the average (or weighted) market share, as well as a direct indicator of diversification.

The share of industry k in firm i total sales[23] is:

$$b_{ik} = \frac{S_{ik}}{\sum_{k=1}^{K} S_{ik}}, \quad i = 1, \ldots, I \quad k = 1, \ldots, K$$

This coefficient equals one in the main activity if firm i is not diversified, and 0 in the other activities. Following SCHERER [1983] we compute the equivalent number of activities taken from the Herfindahl index on sales decomposition (one per firm). Our *diversification indicator* d_i thus equals:

$$\frac{1}{d_i} = \sum_{k=1}^{K} b_{ik}^2, \quad i = 1, \ldots, I$$

This equivalent number of activities equals one when the firm is not diversified, and is bounded by the actual number of activities (reached when the activities have equal sales). It is a better measure than the actual number of activities since it eliminates the effect of the industries that contribute weakly to a firm's sales.

We also define the *average market share* as:

$$\bar{s}_i = \sum_{k=1}^{K} b_{ik} \times s_{ik}, \quad i = 1, \ldots, I$$

22. We use the NAP 600 French industrial classification that breaks manufacturing in 255 industries. Moreover, we kept industrial activities only to compute the diversification index presented below (services are also available).

23. It includes exports.

From the market share we also compute the average Herfindahl index. First, the Herfindahl index of concentration in market k is defined as usual by:

$$H_k = \sum_{i=1}^{I} s_{ik}^2, \quad k = 1, \ldots, K$$

Since diversified firms operate in different industries at the same time, we use the *average Herfindahl concentration index*:

$$\overline{H}_i = \sum_{k=1}^{K} b_{ik} \times H_k, \quad i = 1, \ldots, I$$

APPENDIX 4

Tests of Robustness on the Propensity to Patent Equation

Correlation Matrix
p-value between parentheses

	R&D Expenditures (logarithm)	Average Market Share (logarithm)	Diversification (logarithm)
Sales	.749	.314	.723
(logarithm)	(.001)	(.001)	(.001)
R&D Expenditures		.457	.136
(logarithm)		(.001)	(.018)
Average Market Share			.129
(logarithm)			(.026)

In all the regressions that follow, we have checked that nor concentration neither diversification is significant at the 5% level. Therefore, we dropped them out of the model in order to reduce possible multicolinearity. More generally, we kept the only variables that are significant at the 5% level in at least one regression.

The selection of firms that have R&D in 1989 reduces the sample size to 242 firms (instead of 299). Therefore, we ran the regression with the R&D expenditures again. No change occurred. For more details on the construction of R&D capital, by the permanent inventory method, see CRÉPON and DUGUET [1997a].

Propensity to Patent
Additional evidence

Left-hand variable: percentage of innovations patented (logarithm)
With 12 industry dummies (242 firms)
Maximum likelihood estimation of the ordered probit model
(Asymptotic t statistic between parentheses)

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
R&D Expenditures	–	–	.175	.174	.169	–	–	–
(logarithm)			(2.53)	(2.50)	(3.04)			
R&D Capital in 1989	–	–	–	–	–	.176	.176	.160
(logarithm)						(2.96)	(2.96)	(3.39)
Sales	.203	.140	–.016	.052	–	–.044	.018	–
(logarithm)	(3.75)	(1.70)	(0.16)	(0.66)		(1.01)	(0.23)	
Average market share	–	.084	.089	–	.081	.083	–	.061
(logarithm)		(1.00)	(1.08)		(1.25)	(1.01)		(0.94)

Patent deficiency:								
4) Too much disclosure	–.154	–.148	–.137	–.143	–.137	–.146	–.152	–.146
	(2.35)	(2.25)	(2.11)	(2.21)	(2.12)	(2.27)	(2.37)	(2.28)
$\widehat{\sigma}_1$	1.031	1.029	1.011	1.013	1.011	1.003	1.005	1.003
	(12.03)	(12.03)	(12.04)	(12.71)	(12.05)	(12.05)	(12.04)	(12.05)

• References

ARROW, K. (1962). – "Economic Welfare and the Allocation of Resources for Invention". In R. Nelson ed., *The rate and direction of economic activity*, New York: Princeton University Press.

ARUNDEL, A., VAN DE PAAL, SOETE, L. (1995). – *Innovation Strategies of Europe's Largest Industrial Firms*. MERIT, Maastricht, June.

BECK, R. (1986). – "Does Competitive Dissipation Require a Short Patent Life?" *Research in Law and Economics*, vol. 8, pp. 121-129.

BLUNDELL, R., GRIFFITH, R., WINDMEIER, F. (1995). – "Individual Effects and Dynamics in Count Data Models". Institute for Fiscal Studies working paper.

BUSSY, J.-C., KABLA, I., LEHOUCQ, T. (1994). – "La protection technologique dans l'industrie". *Le 4 Pages du SESSI*, n° 34.

COHEN, W., LEVIN, R. (1989). – "Empirical Studies of Innovation and Market Structure". In Schmalensee R. and R. Willig eds, *Handbook of Industrial Organization*, vol. 2, ch. 18, North-Holland.

COHEN, W., NELSON, R., WALSH, J. (1997). – "Appropriability Conditions and why Firm Patent and why they do not in the American Manufacturing Sector". *Mimeo*, June.

CRAMPES, C. (1986). – "Les inconvénients d'un dépôt de brevet pour une entreprise innovatrice". *L'Actualité Economique*, vol. 62, pp. 521-534.

CRÉPON, B., DUGUET, E. (1995a). – "Innovation: Measurement, Returns and Competition". *Insee studies in Economics and Statistics*, 1995, 1, pp. 82-95.

CRÉPON, B., DUGUET, E. (1995b). – "Une bibliothèque de macro commandes pour l'économétrie des données de comptage et des variables qualitatives". *CREST working paper 9525*.

CRÉPON, B., DUGUET, E. (1997a). – "Research and Development, Competition and Innovation: Pseudo Maximum Likelihood and Simulated Maximum Likelihood Methods Applied to Count Data Models with Heterogeneity". *Journal of Econometrics*, 79, pp. 355-378.

CRÉPON, B., DUGUET, E. (1997b). – "Estimating the Innovation Function from the Patent Numbers: GMM on Count Panel Data". *Journal of Applied Econometrics*, vol. 12, pp. 243-263.

CRÉPON, B., DUGUET, E., KABLA, I. (1996). – "A Moderate Support to Schumpeterian Conjecture from Various Innovation Measures". In A. Kleinknecht éd., *Innovation: The Message From New Indicators*. Mac Millan, London.

DE BROCK, L. (1985). – "Market Structure, Innovation and Optimal Patent Life". *Journal of Law and Economics*, vol. XXVIII, April, pp. 223-244.

DUGUET, E. (1995). – "Technical Cooperation through European Joint Patents". *Insee studies in Economics and Statistics*, n° 1, pp. 96-111.

DUGUET, E., IUNG, N. (1997). – "R&D Investment, Patent Life and Patent Value: an Econometric Analysis at the Firm Level". *INSEE Working paper G9705*.

Gallini, N. (1992). – “Patent Policy and Costly Imitation”. *Rand Journal of Economics*, vol. 23, n° 1, Spring, pp. 52-63.

Gilbert, R., Shapiro, C. (1990). – “Optimal Patent Length and Breadth”. *Rand Journal of Economics*, vol. 21, n° 1, Spring, pp. 106-112.

Gouriéroux, C., Monfort, A. (1996). – *Statistics and Econometric Models*. Cambridge University Press.

Gouriéroux, C., Monfort, A., Trognon, A. (1984a). – “Pseudo Maximum Likelihood Methods: Application to Poisson Models”. *Econometrica* 52 (3), pp. 701-720.

Gouriéroux, C., Monfort, A., Trognon, A. (1984b). – “Pseudo Maximum Likelihood Methods: Theory”. *Econometrica* 52 (3), pp. 681-700.

Griliches, Z. (1990). – “Patent Statistics as Economic Indicators: a Survey”. *Journal of Economic Litterature*, vol. XXVIII, pp. 1661-1707.

Hausman, J., Hall, B., Griliches, Z. (1984). – “Econometric Models for Count Data with an Application to the Patent-R&D Relationship”. *Econometrica* 42 (3), pp. 909-938.

Horstmann, I., Mc Donald, G., Slivinski, A. (1985). – “Patents as Information Transfer Mechanisms: to Patent or (maybe) not to Patent”. *Journal of Political Economy*, 93, pp. 837-858.

Kabla, I. (1997). – “Easiness of Imitation, Patent Disclosure and the Optimal Patent Scope”. *Mimeo.*

Kamien, M., Schwartz, N. (1982). – *Market structure and innovation*. Cambridge University Press.

Klemperer, P. (1990). – “How Broad Should the Scope of Patent Protection be?” *Rand Journal of Economics*, vol. 21, n° 1, Spring, pp. 113-130.

Lee, L.-F. (1981). – “Simultaneous Equations Models with Discrete and Censored Dependent Variables”. In C. Manski and Mc Fadden D. eds, *Structural analysis of discrete data with econometric applications*. MIT Press, pp. 346-364.

Lee, T., Wilde, L. (1981). – “Market Structure and Innovation: a Reformulation”. *Quaterly Journal of Economics*, 94, pp. 429-436

Levin, R., Klevorick, A., Nelson, R., Winter, S. (1987). – “Appropriating the Returns from Industrial Research and Development”. *Brookings Papers on Economic Activity*, 3, pp. 783-831.

Loury, G. (1979). – “Market Structure and Innovation”. *Quaterly Journal of Economics*, 93, pp. 395-410.

Maddala, G. S. (1992). – *Limited Dependent and Qualitative Variables in Econometrics*. Econometric Society Monograph, n° 3, Cambridge University Press.

Malinvaud, E. (1970). – *Statistical Methods of Econometrics*. North-Holland.

Mansfield, E., Schwartz, M., Wagner, S. (1981). – “Imitation Costs and Patents: an Empirical Study”. *Economic Journal*, 91, pp. 907-918.

Mansfield, E. (1985). – “How Rapidly does New Industrial Technology Leak Out”. *Journal of Industrial Economics*, 34 (2) pp. 217-227.

Nordhaus, W. (1969). – *Invention, growth and welfare: a theoretical treatment of technological change*. Cambridge, Massachussets: MIT Press.

Pakes, A., Shankerman, M. (1986). – “Estimates of the Value of Patent Rights in European Countries during the post 1950 Period”. *The Economic Journal*, December, pp. 1052-1076.

Scherer, F. (1983). – “The Propensity to Patent”. *International Journal of Industrial Organization*, 1, pp. 107-128.

Scotchmer, S. (1991). – “Standing on the Shoulders of Giants: Cumulative Research and the Patent Law”. *Journal of Economic Perspectives*, 5 (1), pp. 29-41.

Scotchmer, S., Green, J. (1990). – “Novelty and Disclosure in Patent Law”. *Rand Journal of Economics*, 21, n° 1, Spring, pp. 131-146.

TANDON, P. (1982). – “Optimal Patents with Compulsory Licensing”. *Journal of Political Economy*, 90, n° 3, pp. 470-486.

TAYLOR, SYLBERSTON (1973). – *The Economic Impact of the Patent System*. Cambridge University Press.

VAN DIJK, T. (1994). – *The limits of patent protection*. Maastricht: Universitaire Pers Maastricht.