Creative Destruction and the Innovative Core:
Is Innovation Persistent at the Firm Level?

An empirical reexamination from CIS data
comparing the propensity score and regression methods *

Emmanuel Duguet[1]

Stéphanie Monjon (YE)[2]

Abstract

At the macroeconomic level, the persistence of innovation allows sustainable growth. But does growth come from the same set of firms or originate always from different innovators? On this point, the assumptions of endogenous growth models differ and innovation persistence at the macroeconomic level can be supported by different firm-level behavioral assumptions. The aim of this article is twofold. First, we evaluate the empirical pertinence of the different views of the dynamics of the innovative process by estimating the degree of innovation persistence at the firm level. Secondly, we explore the determinants of innovation persistence by testing the empirical implications of three theoretical models. We show that the innovation persistence is essential at the firm level and that the origin of the persistence depends on the size of the firm.

Keywords: Community Innovation Surveys, Creative destruction, Innovation, Learning-by-doing, Matching, Persistence, Propensity score, Research and development.

JEL Classification: C14, O31, O32.

---

* We thank T. Brodaty, B. Crépon, D. Encaoua and C. Lebas for helpful suggestions and comments as well as the participants to the Spring Meeting of Young Economists (Copenhagen, April 2001), the CERAS-EUREQua-LEI seminar in Industrial Organisation (Paris, May 2001), the CEPR-ECARES workshop (PMFM, Brussels, May 2001), the EEA annual meeting (Lausanne, August 2001), the EARIE meeting (Dublin, August 2001), the 50th AFSE annual congress (Paris, September 2001) and the ADRES seminar (Paris, November 2001).

[1] Université de Bretagne Occidentale and EUREQua - CNRS UMR 8594 – Université de Paris I - 106-112, Boulevard de l'Hôpital, 75647 Paris Cedex 13. Tel/Fax : 01.44.07.82.20./31. e-mail : eduguet@univ-paris1.fr

[2] EUREQua - CNRS UMR 8594 - Université de Paris I - EUREQua, UMR CNRS 8594, 106-112, Boulevard de l'Hôpital, 75647 Paris Cedex 13. Tel/Fax : 01.44.07.82.17/31. e-mail : monjon@univ-paris1.fr.

1. Introduction

Innovation is a primary source of economic growth. But do innovations come from the same set of firms or, on the contrary, from a continuous renewal of innovators? On this point, the assumptions underlying the endogenous growth models are different. While Romer (1991) suggests a strong persistence of innovators, Aghion and Howitt (1992) take the opposite view and develop a neo-Schumpeterian model in which the process of creative destruction leads to an infinite renewal of innovators.[3] More recently, Aghion, Harris and Vickers (1997), Encaoua and Ulph (2000) and Hörner (2001) have shown that a rich number of individual patterns of innovation persistence are in fact possible. Hence, innovation persistence at the macroeconomic level can be supported by different firm-level behavioral assumptions. What is the empirical relevance of these different microeconomic foundations?

Answer to this question relates directly to central problems in economic theory. In the first place, the frequency with which a firm introduces innovations plays a central role in the analysis of technical progress and economic growth (Aghion, Harris and Vickers, 1997). In the second place, understanding whether innovation is persistent or not at the firm level constitutes an important piece of evidence for finding and improving current theories of industrial dynamics, where some forms of dynamic increasing returns play a major role in determining degrees of concentration, the evolution of market shares and their stability over time (Geroski, 1995).

Few empirical works have examined the issue of innovation persistence at the firm-level data (Crépon and Duguet, 1997; Geroski, Van Reenen and Walters, 1997; Malerba and Orsenigo, 1999; Cefis and Orsenigo, 2001). Globally, these works conclude either that there is a small degree of innovation persistence, or that innovation is persistent only in a small number of firms. Hence, as long as the majority of innovators would be involved in *a creative destruction process*, there would still remain *an innovative core*. However, several considerations led us to investigate the robustness of this conclusion.

The first motivation of this paper is to clarify the links between the theoretical explanations of innovation persistence and the empirical results. The previous studies do not always allow to identify the theoretical model that generates the data. Indeed, some data limitations prevent them from testing some theoretical propositions. Moreover, the previous studies differ according to the set of explanative variables, such that their results are not easily

[3] This difference of assumptions can be seen by comparing the innovation value equations of these models.

comparable. This paper is an attempt to identify the theoretical models that are compatible with the facts, in order to achieve a better understanding of the empirical regularities. We fulfilled this objective by using a larger set of explanative variables and by referring to the theoretical literature explicitly when we interpret our empirical results.

The second motivation of this paper originates from the way innovation is measured in the previous studies that use only patent data or "major" innovations data.[4] By definition, these data tend to underestimate the number of innovative firms, hence the persistence of innovation. For the patent data, many problems emerge, mostly from the fact that a patent is not an innovation (Griliches, 1990). Among the problems often quoted in the applied literature, patents are not the most effective means of knowledge protection and this explains why some industries report patent figures much below their innovative level (Levin et al., 1987; Cohen et al., 1997; Duguet and Kabla, 1998).[5] Moreover, even if patent protection was perfect, a patent would involve both to innovate and to be *the first* to innovate. This means that patent data measure the persistence of innovative leadership rather than the persistence of innovation. In the same way, the major innovations data involve some kind of leadership as well. Consequently, the evaluation of innovation persistence requires separating innovation from its commercial performance and from the intellectual protection strategies of the firms. The Community Innovation Surveys (CIS) provide information that satisfy these conditions.

Lastly we examine the robustness of the standard econometric methods. Ideally, we would like to measure the difference between, on the one hand, the innovative performance a firm makes today knowing that the firm has innovated in the past and, on the other hand, the innovative performance the same firm *would have done* if it had not innovated in the past. There is no way to observe these two quantities at the same time for any firm. Indeed, either the firm innovated and we cannot observe what it would have done if it had not innovated, or the firm did not innovate and we cannot observe what it would have done if it had innovated. The standard econometric methods assume that the non-observable outcomes can be obtained from a regression model. This implies that there should exist a firm with exactly the same characteristics but with a different past innovation status. If this past innovation status depends on the firms'

---

[4] In the applied literature, "major" innovations refer to innovations that have met a large commercial success

[5] Duguet and Kabla (1988) show that on average only a third of the innovations are patented because of the information disclosure implied by the patent documents. Moreover, after controlling for the propensity to patent, R&D investments, market share and industry of the firm, the determinants "will to avoid litigation" and "technology negotiations" remain strongly significant in the explanation of the number of patent applications. Hence, strategic aspects are omnipresent in the patent numbers, which precludes from using it as a mere innovation indicator. This

characteristics, this assumption could not be valid. In order to solve this problem, we use the propensity score approach introduced by Rubin (1974), Rosenbaum and Rubin (1983) and developed by Heckman, Ichimura and Todd (1997).

We reach three conclusions. First, we find that innovation persistence is much stronger with Community Innovation Surveys data than with patent data. Second, we find that the origin of innovation persistence depends on the size of the firms. While learning-by-doing effects in the production of innovations appear to play a major role in innovation persistence in the small firms, their importance steadily decreases with the size of the firms. In the largest firms, these effects vanish and the persistence of innovation originates from the persistence of the formal research and development investments. Third, we find that the standard econometric methods provide correct results on average but conceal an interesting composition effect.

Section 2 presents the theoretical models of innovation persistence, discusses their empirical implications and summarizes the empirical results from previous studies. Section 3 presents the data and section 4 gives the estimation methods. In section 5, we discuss the results. Section 6 concludes the paper.

2. Theoretical framework and previous empirical analysis

2.1. The theoretical framework

The persistence of innovation at the firm level can be explained by three types of models, which lead to different empirical predictions. The *linear model* of innovation establishes a simple relationship between the research and development (R&D) expenditures of a firm and its innovations. The firms that can support the sunk costs of R&D make inventions that lead to product or process innovations (Cohen and Klepper, 1996). In this model the successive innovations originate from the continuity of R&D expenditures and are not directly connected. According to this vision of the innovative process, innovation is persistent only if R&D is.

This model implies that there should be no innovation difference between firms once controlled for R&D expenditures. Hence, it is possible to test this model by regressing a measure

---

strategic dimension of patents is more and more present in the firms' decisions to patent their innovations (Encaoua and Hollander, 2001).

of innovation on a measure of past innovation and a measure of R&D. The issue is then whether there is an additional effect of past innovation on present innovation. If R&D and past innovation are both significant, other models should be investigated.

A second model underlines the importance of the *financial constraints* related to the R&D activities (Nelson et Winter, 1982). Giving the difficulty of funding the R&D activities, the commercial success of past innovations helps to fund the current innovative activities: successful innovative firms make profits that can be allocated to future R&D investments. This problem of funding appears when the financial markets are imperfect. A firm that reached a commercial success in the past has more chance to innovate in the future merely because it reinvests its benefits in its research projects. Hence, “success breeds success”: a past innovation that met commercial success becomes a necessary condition to finance the future research projects.

Empirically, the correlation between innovation and research is expected to be weaker than in the linear model, since past innovation is a necessary condition of present R&D investments. One way to test this model is to control the differences of financial constraints between the firms before examining the relation between innovation and research. This control can be made with a variable of size.[6] If the differences of size between firms are controlled for, research should have a significant impact on present innovation and past innovation should not be significant.

The third model relies on the idea that innovations result from an accumulation of specific competencies (Rosenberg, 1976; Malerba et al., 1997). More precisely, this model considers that the innovative abilities that a firm develops when it invests in research projects do not necessarily depreciate rapidly over time. Therefore, the same knowledge or know-how may be applied to develop several innovations at successive times. The competencies related to innovation do not only refer to the scientific knowledge available to the firm but also to a specific know-how in the production of innovations. Chandler (1990) emphasized the importance of such effects in the production of innovations. According to that view, firms benefit from *learning-by-doing* effects in the production of innovations, which leads to a strong innovation persistence at the firm level.

Empirically, a simple way to test this hypothesis is to examine whether past innovations still have an effect on present innovation, once both R&D and size have been controlled for: according to this third model, past innovation should remain significant. This conclusion differs from the ones of the two previous models. Such an effect can be supported neither by the linear

[6] On the relationship between size and liquidity constraints, see Audretsch and Elston (2002).

model because R&D differences are accounted for, nor by the Nelson-Winter model because the financial constraints are controlled for.

Finally, even if these models do not explain the persistence of innovation in the same way, they all suggest that we should observe a significant degree of persistence. Indeed, according to the first model, a weak persistence would mean that firms do not invest in R&D continuously, which contradicts the evidence from the R&D surveys. Only a part of the firms performs R&D but these firms generally invest in R&D activities continuously.

According to the second model, a weak innovation persistence implies either that the capital markets would not be able to fund the innovation projects, or that firms would not reinvest their profits from past innovations in their current research projects. These two conditions are unlikely. On the one hand, the financial markets have considerably increased the funds available for the innovation projects and many public policies supporting innovation exist in OECD countries (Hall and Van Reenen, 2000). If there remains a funding problem, it is unlikely that it is so large that innovation would stop being persistent. On the other hand, it is equally unlikely that firms do not reinvest their profits from past innovations in their current research projects, giving the importance of innovation for their competitiveness. In some industries, competition principally centers in innovation (Encaoua and Hollander, 2002). When a firm reinvests its profits in innovation projects, it has more chance to create or maintain its market power.

Lastly, according to the third model, a weak innovation persistence would mean that learning-by-doing effects in the production of innovations are not important. Some recent empirical analyses suggest the contrary. Kim (1997) showed the central role played by the development of internal innovative know-how in Samsung's technological successes in the semi-conductors industry. Henderson and Cockburn (1996) and Nightingale (2000) studied the pharmaceutical industry and showed that the pharmaceutical firms have succeeded in lowering the production costs of new drugs by adopting new experimentation methods and organizational changes.

If we combine these models, we should thus expect a significant degree of innovation persistence, at least in a part of the firms. But, until now, previous studies rather conclude that the innovation persistence at the firm level is weak. We think that this apparent paradox is due for a significant part to the data used to measure innovation.

2.2. *The previous empirical results*

The issue of innovation persistence at the firm level has recently been studied by Geroski, Van Reenen and Walters (1997). The authors use two different measures of innovation: annual patents data and annual major innovations data. The extent of innovation persistence of a firm is defined as the number of consecutive years during which it innovates. The authors do not account for innovation inputs differences (like R&D expenditures) but for size differences for major innovations. Their main result is that : "it is very hard to find any evidence at all that innovative activity can be self-sustaining over anything other than very short periods of time". Three explanations related to the measurement of innovation could explain this result.

The first point is that there is a strong difference between a patent and an innovation. The most important difference for this study is that a patent implies *leadership* in innovation. Indeed, an innovative firm needs to be the first to apply for a patent in order to be properly registered in the data set. Patent data do not only measure innovation but also the fact that a firm won an innovation race. Hence, a sample reporting firms that win the innovation race from time to time would show up a weak persistence of innovation, even if these firms innovate all the time. The second point is that leadership itself may be poorly measured by patent data. This problem is linked to the intellectual property strategy of the firms. Firms can have an obvious interest in avoiding to patent an intermediate innovation in order to conceal the knowledge that could be used by their competitors. There is a strong empirical evidence of this kind of behavior (Cohen et al., 1997, on American data; Duguet and Kabla, 1998, on French data). Lastly, the firms tend to patent more their product innovations than their process innovations (Arundel and Kabla, 1998). The patent data are thus biased in favor of product innovations. Hence it is unlikely that the persistence of innovation can be adequately measured with such data.

The second type of data that Geroski, Van Reenen and Walters (1997) used to measure persistence concerns innovations that met a commercial success. Here, the data avoid the biases associated to the patenting strategy of the firms but there remains one problem. Since this definition involves a commercial success, the firms considered as innovative are likely to be either innovation leaders or commercial leaders. In the latter case, the data measure the ranking of the firms on the market, that is their ability to adapt the available innovations to consumers' tastes.

The work of Crépon and Duguet (1997) is equally related to the issue of innovation persistence. They use a panel of R&D performers operating in France and they also use patent data to measure innovation.[7] They estimate a dynamic count data model that links the current number of patents to both the previous year number of patents and the amount invested in R&D. They also account for an individual fixed effect that represents the differences of size, of technological opportunities and of the firm's propensity to patent. Contrary to Geroski et al. (1997), they find that innovation persistence is strong among R&D performers. Moreover they find a positive effect of lagged patents on the current number of patents, which suggests that *learning-by-doing* effects play a role in the production of innovations. At a first glance, the results of Crépon and Duguet (1997) seem to contradict the results of Geroski et al. (1997). However, the study of Malerba and Orsenigo (1999) suggests that these two works are in fact complementary.

The descriptive statistics study by Malerba and Orsenigo (1999) examined the issue of persistence by using patent data of six countries over the period 1978-1991.[8] They conclude that a large fraction of innovators is casual. Nevertheless, there would still remain a stable group of innovators that apply for a large share of patenting. The results of this study are confirmed by Cefis and Orsenigo (2001) who find that both "great-innovators" and non-innovators have a strong probability to stay in the same innovative state. Then, there would be a strong persistence of innovation in a core of firms.

Globally, two interesting elements emerge from these studies. On the one hand, the weakness of persistence degree found in several empirical studies contradicts the basic theoretical prediction and suggests to use different data sources in order to avoid measuring innovation by patents. On the other hand, few studies pay attention to the determinants of innovation persistence, such that we do not know the empirical relevance of the different theoretical models.

In order to determine which of the previous theoretical models is relevant, one should care about the whole list of explanative variables included in a regression, since the significance of each coefficient depends on all the explanative variables included. Current innovation must be explained by past innovation, research and size, such that a significant effect of past innovation on current innovation can be interpreted as a learning-by-doing effect since the financial constraints and size are controlled for. Crépon and Duguet (1997) take into account these different variables but their sample is limited to the significant R&D performers only. Moreover,

---

[7] R&D performers are defined according to the Frascatti criterion (at least one research working full time on R&D).

[8] The six countries are : France, Germany, Italy, Japan, the U.K. and the U.S.A. Their patent data are available for different periods: 1978-1982, 1982-1985, 1986-1988 and 1988-1991.

they use patent data. Our objective is thus to analyze this latter model in order to be able to identify all the determinants of innovation persistence with a more representative sample of firms and another measurement of innovation.

3. The data

One purpose of this study is to use other data sources than patent data in order to separate the leadership, commercial or innovative, from innovation itself. The recent Community Innovation Surveys (CIS) provide information about the implementation of innovation at the firm level, without any reference to their commercial success or their patenting status.[9]

The data come from four data sets, including three about innovation. The first data set is the Innovation Survey "*L'Innovation Technologique dans l'Industrie*" conducted by the SESSI that was performed in order to prepare all the other innovation surveys in France. It was conducted in 1991 and provides information about the period 1986-1990. In order to assess innovation persistence, we also use two Community Innovation Surveys (henceforth CIS). The CIS1, conducted in 1993, provides information about the period 1990-1992, while the CIS2, conducted in 1997, gives information about the period 1994-1996. The remaining data set provides accounting information. They come from the annual industry census 1985 (in French, "*Enquête Annuelle d'Entreprises*") that provides data about the line of business and sales at the firm-level.[10] The merger of these data sets provides a sample of 808 firms operating in manufacturing and covers the period 1986-1996.[11] Year 1993 is missing since no innovation survey covers this year.

Table 1 provides the percentage of innovators by industry. These percentages are not directly comparable since the first survey spans 5 years instead of 3. Hence the first survey gives the highest innovation figures. The maximum is reached in equipment goods (including cars) and the

---

[9] Moreover, these surveys have been conducted in many European countries such that international comparisons are now possible.

[10] EAE is the « Enquête Annuelle d'Entreprises » (the industry census). It is compulsory for all firms above 20 employees.

[11] Notice that we imposed the presence of firms in each innovation survey and in the 1985 E.A.E. survey. We thus examine the innovation persistence conditional to the existence of the firms in 1985 and to their survival up to 1997. The integration of entry and survival issues of innovative firms in our analysis would require another data which we do not posses. Nevertheless, the impact of new entrants on the date at which the installed firms decide to innovate is taken into account. Moreover, the role of new innovative firms, in particular of start-ups, could be relatively minor in France. In a recent study, Arora et al. (2000) show that in Europe 90% of research projects in biotechnological sectors are due to large installed firms, whereas in United States more than 50% of the research projects are

minimum in consumer goods. The two following surveys (CIS1 and CIS2) provide lower figures as expected, but they give comparable results. On a three-year period, the percentage of innovators is 61%. These figures seem high but there are two reasons for it: they cover a three-year period or a five-year period and the definition of innovation includes both new products and new processes.[12] Since these data do not refer to the market performance of the firms or to their patenting strategy, we should be able to better identify the persistence of innovation. The advantage of this innovation definition is that it allows to measure innovation in a firm that persistently innovates but in a different manner over time, for instance by introducing a new product and then by improving on the production process. The case studies confirm the view of the innovative process according to which the persistence originates from the diversity of innovative behaviors (Kim, 1997).

-Insert Table 1-

The innovation surveys also include information about the inputs that have been used to innovate. The 1991 innovation survey distinguishes formal R&D according to the Frascatti criterion and the informal R&D identified as "technical studies". The answers are provided on a four-point scale ("none", "weak", "moderate" and "strong"). The figures are reported in Table 1. Over 1986-1990, about the two thirds of the innovative firms declare to have conducted moderate or strong informal R&D, 43% have conducted strong formal R&D and 29% have conducted no research at all. The CIS1 does not allow to separate the formal R&D from the informal one. The only available data aggregates both definitions. We have grouped the weak and moderate levels to obtain a four-point scale. It appears that a third of the firms have conducted very strong formal or informal R&D, a figure that is close to the one obtained in the previous survey for informal R&D.

Table 1 equally presents some descriptive statistics on the size of the firms measured by their sales in 1985 and on the degree of technological opportunities in their activity. The advantage of

---

accompanied by the creation of a new firm. Chesbrough (1999) finds similar results in the industry of semi-conductors.

12 The definition used in the innovation surveys includes the five following types of innovation : (1) significant improvement of an already existing product; (2) introduction of a product that is new both for the firm and for the market; (3) introduction of a product that is new for the firm but not for the market; (4) significant improvement of an already production process and (5) process breakthrough. Notice that the full decomposition is available in the first survey only.

measuring the size of the firms by sales rather than employment is that is it less sensitive to the differences in the capital-labor ratio (Cohen and Levin, 1989; Kleinknecht, 1996). The degree of technological opportunities does not come from an industry classification but is available at the firm level in the first innovation survey that provides information about the degree of innovation of a firm's line of business (four levels: "none", "weak", "moderate" and "strong"). Based on previous works, we define the "low-tech" activities as "none" or "weak", and the "high-tech" as "moderate" or "strong". This variable has proved to be useful in previous econometric studies in French manufacturing where it revealed significant differences of performances that were not attributable to the firm-level innovation.[13] We see that the innovators have generally a larger size than the non-innovators and they are situated in sectors where the technological opportunities are more important.

4. Methodology

Let y_i denote the innovative performance of firm i and T_i denote past innovation. Each firm has two different innovative performances depending on whether it innovated in the past ($T_i = 1$) or not ($T_i = 0$). We denote them $y_i(1)$ and $y_i(0)$. The effect of past innovation on the whole population, called the causal effect in the litterature, is defined as $c = \mathrm{E}(y_i(1) - y_i(0))$. If it was possible to observe the performance of each firm in both states 0 and 1, the average effect of past innovation could be consistently estimated by the difference of the corresponding sample averages. Since such data are not available, one needs to construct a *counterfactual*, that is an estimation of $y_i(1)$ for the firms that did not innovate and an estimation of $y_i(0)$ for the firms that innovated. The most widespread method is to use a parametric model $y(.)$ that explains the performance y_i by past innovation T_i and the characteristics of the firm (denoted X_i) : $y_i = y(T_i, X_i)$. The counterfactuals are simply obtained by $y_i(0) = y(0, X_i)$ and $y_i(1) = y(1, X_i)$.

When the performance is a binary variable (e.g., innovate or not), the evaluation can be obtained from a probit model, that explains the probability to innovate today by past innovation and the characteristics of the firm (size, industry, R&D expenditures). It is given by :

---

[13] This variable could indirectly measure a spillover potential available to the firm. See Barlet et al. (1998).

$$y(T_i, X_i) = \Phi[X_i b + gT_i],$$

where Φ is the cumulative distribution function of the standard normal distribution. The average effect of past innovation is estimated by :

$$\hat{c} = \frac{1}{N}\sum_{i=1}^{N}\{\hat{y}(1, X_i) - \hat{y}(0, X_i)\} = \frac{1}{N}\sum_{i=1}^{N}\left\{\Phi\left(X_i \hat{b} + \hat{g}\right) - \Phi\left(X_i \hat{b}\right)\right\},$$

where $\hat{b}$ and $\hat{g}$ denotes the maximum likelihood estimations from the probit model and N is the number of firms in our sample.[14]

The main drawback of the probit method is that it assumes that there is always a *perfect counterfactual* given by the parametric model, that is it assumes there can exist a firm that did not innovate in the past and that has *exactly* the same characteristics than the firm that innovated (i.e., the same $X_i b$). Therefore this method could be invalidated when innovative firms have significantly different characteristics X_i from the non-innovative firms. The basic reason why the firms are not comparable may be simply that innovative firms do more R&D or have a size that differs from the size of non-innovative firms.[15] In this case, it is possible that the probit method provides inconsistent estimates of the effect of past innovations.

With the Rubin method, the counterfactuals are not obtained from a parametric model, but from the *actual* data. This methodology was first proposed by Rubin (1974) and developed by Rosenbaum and Rubin (1983) and Heckman, Ichimura and Todd (1997) among others. The intuition is the following. If we had an experimental sample, a direct comparison of the percentage of innovation between the two sets of firms, defined by T_i, would provide a consistent estimate of c. The reason is that the performance average difference between past innovative and non-innovative firms could only come from past innovation and the empirical average would be a consistent estimator of the expected causal effect. But past innovation is not allocated at random between firms so that a generalization of this method is needed. Rubin (1974) showed that it is possible to evaluate the effect c if the following condition if fullfiled :

$$(y_i(0), y_i(1)) \perp T_i | X_i \qquad \textit{(H-1)}$$

where $\perp$ denotes statistical independence. This implies that one can evaluate the counterfactuals by :

---

[14] The same computation can be done for any binary explanative variable. It is thus possible to compare the importance of past innovation with the importance of other explanative variables, especially with R&D.

$$\mathrm{E}(y_i(1)|X_i, T_i = 0) = \mathrm{E}(y_i(1)|X_i, T_i = 1) = \mathrm{E}(y_i(1)|X_i)$$

$$\mathrm{E}(y_i(0)|X_i, T_i = 1) = \mathrm{E}(y_i(0)|X_i, T_i = 0) = \mathrm{E}(y_i(0)|X_i)$$

The only practical problem with this method is that it implies to match firms on large number of variables X_i. Fortunately, Rosenbaum and Rubin (1983) have shown that this condition can be simplified to conditional independence on the one-dimensional *propensity score* defined as $\Pr(T_i = 1|X_i)$. This selection probability plays an especially important role for the following reason. Suppose that we have a group of firms with the same probability to have innovated in the past. Inside this group, there are firms that innovated in the past and firms that did not. Hence, the difference of past innovation between these firms can be considered as random. The comparison of the average performances *inside* homogeneous probability groups is therefore relevant. More precisely, Rosenbaum and Rubin (1983) have shown that:

$$(y_i(0), y_i(1)) \perp T_i | X_i \Rightarrow (y_i(0), y_i(1)) \perp T_i | \Pr(T_i = 1|X_i).$$

In practice, the propensity score is replaced by its estimation. The propensity score can be estimated by a probit or a logit model, with X_i as explanative variables. But it may also depend on a firm-level fixed effect, which represents any time-invariant factor that influences the innovative performance of the firms.[16] For instance, this effect may represent firms that have research teams with more successful researchers. We denote this fixed effect by a_i. The identification condition becomes:

$$(y_i(0), y_i(1)) \perp T_i | (X_i, a_i) \Rightarrow (y_i(0), y_i(1)) \perp T_i | \Pr(T_i = 1|X_i, a_i).$$

The fixed effect raises an issue because it is unobservable. There are two ways to deal with this problem. The first method is to estimate a fixed-effect logit model on panel data and to use the predicted probability to match firms. But, unfortunately, this is not possible with the Community Innovation Surveys for the two following reasons:

1. In the fixed effect logit model, the estimation proceeds by the conditional maximum likelihood method, where the conditioning variable is the number of times that a firm innovates (i.e., the sum of the innovation dummies). This implies that two kinds of events must be excluded from the regression (Hsiao, 1986). First, one should exclude the firms

---

[15] For evidence, see Cohen and Levin (1989) and Kleinknecht ed. (1996) on European data.

[16] Taking into account an individual fixed effect implies that the selection equally comes from unobservable variables.

that have always innovated and, secondly, one should exclude the firms that have never innovated. The reason why the always-innovators are excluded by this method is that the conditional probability to innovate knowing that one has always innovated is equal to one, such that it does not contribute to the conditional likelihood (i.e., provides no relevant information for the estimation). The reason why never-innovators are excluded is that the conditional probability to innovate knowing one has never innovated is equal to zero. Hence *the higher the persistence of innovation, the less there will be firms available for the fixed-effect logit regression.* And we have a majority of such firms.

2. In a logit model with a fixed effect and two years of data, one needs to do a regression on the difference of the explanative variables (Hsiao, 1986). Unfortunately the definition of research inputs changes over time. In the first survey, formal and informal R&D are separated and the firms answer on a four point scale ("Not important", "Weakly important", "Moderately important", "Strongly important"). In the second survey, formal and informal R&D are grouped together and firms answer on a 5 points scale (0,1,2,3,4). Hence, there is no way to take the difference.

Therefore, the applied researcher has no other choice than to use another method. The second method is to find out observable variables that are strongly correlated to the fixed effects such that the differences between firms can be controlled for. Thus, we need variables that, for example, give information on the competencies of the research team of a firm. The most obvious set of variables is the past innovative history of the firm. If a firm has a successful research team, its innovation history should score better than the one of another firm.

The first innovation survey is especially useful since it provides information on the innovation history of the firms over five years. The long length of inquiry of this data set is an advantage when trying to control for individual effects. Firms that did not innovate over five years may not have successful research teams, while firms that innovated at least once in five years may have a more competent research teams. Another variable that can be used is the degree of technological opportunities. We denote these two variables by Z_i. Replacing the fixed effect formulation by its observable counterpart, we use the following condition:

$$(y_i(0), y_i(1)) \perp T_i | W_i \Rightarrow (y_i(0), y_i(1)) \perp T_i | \Pr(T_i = 1 | W_i)$$

where $W_i = (X_i, Z_i)$. We perform the matching on the corresponding estimated propensity score.

Whichever the method used, the first precaution to take is to check that the support of the probabilities has a sufficient overlap between the two groups of firms (i.e. treated and non-treated firms). Some firms can be excluded from the sample because there exists no relevant counterfactual. It is then not possible to evaluate a causal effect for these latter firms. It is an important difference with the probit method.

Our first evaluation of the causal effect is based on the construction of several probability classes.[17] This method allows to compute the effect of past innovation by regression and to compute heteroskedasticity-robust standard errors in a simple way[18]. However, the matching is not perfect. In order to fix this problem, we have grouped observations into strata defined on the estimated propensity score such that the covariates within each stratum are balanced (Rosenbaum and Rubin, 1983; Dehejia and Wahba, 1998).[19] Once that the probability classes are constructed, we compute the average difference of performance *inside each propensity score class* by the following regression:

$$y_i = a + c\, T_i + u_i ,$$

where u is a disturbance. The OLS estimator simply gives the difference of the means between the past innovators and the other firms: $\hat{c} = \bar{y}_1 - \bar{y}_0$. A second evaluation is based on an extension of the previous regression :[20]

$$y_i = a + bm_i + T_i \left(c + dm_i\right) + v_i ,$$

where $m_i = W_i \hat{\boldsymbol{d}}$ is the score obtained from the (selection) probit model, $\hat{\boldsymbol{d}}$ the corresponding maximum likelihood estimate and v a disturbance.[21] Here, the effect of past innovation depends on the characteristics of the firm through the score. When the score is centered, the average causal effect is still given by the coefficient of the treatment since we have :

$$\mathrm{E}\left(y_i(1) - y_i(0)\right) = \mathrm{E}\left(c + dm_i\right) = c .$$

17 An imperfect matching allows keeping more firms in the sample but implies to test for homogeneity inside the classes. On the problems associated to matching, see Heckman et al. (1998).

18 It is why we impose that at least 30 treated and 30 non-treated are included in each probability class.

19 For this reason, we perform a test for the statistical significance of differences in the distribution of covariates, focusing on the two first moments.

20 See Crépon and Iung (1999) for a justification.

21 The propensity score is strictly increasing with m_i such that it could be used to match firm as well. The reason why we take this score instead of the propensity score is that it is a linear function of the variables, such that we get a standard-looking regression.

The previous estimator provides valid inference for the whole population of firms, but we are also interested in the causal effect on the treated. The difference between the two quantities is that the causal effect can be different for each firm such that there is no reason *a priori* why the effect of the treatment should be the same among the treated and the non-treated. More precisely, we evaluate $c_1 = \mathrm{E}(c_i | T_i = 1)$. Here, the identifying assumption is less demanding :

$$y_i(0) \perp T_i | W_i \Rightarrow y_i(0) \perp T_i | \Pr(T_i = 1 | W_i)$$

The simplest way to evaluate this quantity is to use the nearest-neighbor. We match each firm that innovated in the past with the past non-innovative firm that has the closest probability to innovate (Rubin, 1977). When several neighbors have the same probability, we select one at random. The estimator is now:

$$\hat{c}_1 = \left(\sum_{i=1}^{n} T_i\right)^{-1} \left(\sum_{i / t_i = 1} y_i(1) - \tilde{y}_i(0) \right)$$

where $\tilde{y}_i(0)$ denotes the innovative performance of the firm that did not innovate in the past and whose propensity score is the closest to the one of the firm *i.* In this case, we do not need to regroup our data into probability classes since we take *the nearest neighbor* of each past innovator. The standard error of this estimator can be complicated to compute such that we have used the bootstrap method with 100 simulations (Efron and Tibshirani, 1983).[22]

Notice that this estimator can also be used to estimate the effect over the whole population. This is what we have done in order to easier the comparison with the effect on the treated.

5. The results

5.1. The probit models

The results of the probit models are reported in Table 2. Two different sets of explanative variables are examined. The aim of the model 1 is to evaluate the degree of innovation persistence, such that the current innovation (1994-96) is explained only by lagged innovations, industry dummies and a firm-level dummy indicating whether the firm belongs to a high-tech

[22] The propensity score is re-estimated for each simulation.

activity. We controlled for industries differences because we want to evaluate the firm-level persistence of innovation. The model 2 includes past innovations, past research activities and size as explanative variables. It is only this last model that allows us to identify precisely the determinants of the innovation persistence.

We find a strong persistence of innovation at the firm level. In the model 1, all the lagged innovation variables are significant. A firm that innovated in the past has thus a stronger probability to innovate today. Moreover, the coefficient of past innovations decreases with the lag length. These findings are robust to the introduction of industry and to the introduction of the firm-level high-tech dummy, such that the past innovation variables reflect neither differences between industries, nor differences between low-tech and high-tech activities. The controls are also significant: the fact to be engaged in an activity where the technological opportunities are high favors current innovation and the equipment goods industry produces more innovations than the other lines of business. These two results are consistent with previous evidence. This first model allows us to conclude to the likely existence of an innovative core of firms that innovate persistently and whose the advantage decays relatively slowly over time since innovations that have been made 10 years before still have a significant effect.

- Insert Table 2-

In the model 2, we find that innovation over 1986-1990 is no more significant once controlled for research and development over the same period. Hence a good part of the persistence found in the model 1 comes from research. *The linear model is thus partly validated.* However, the short-run lagged innovation (1990-1992) continues to be significant. We conclude that firms that innovated in the past benefit from advantages that do not come only from formal research activities but equally from more informal activities. The learning-by-doing effects in the production of innovations thus play an important role to explain the persistence of innovation. *The relevance of the model with learning-by-doing is thus confirmed.*

Another interesting result is that while formal R&D over 1986-1990 is significant on innovation over 1994-1996, informal R&D is not. *This suggests that informal knowledge would have a higher depreciation rate than formal knowledge.* One explanation is that formal R&D can be more easily codified and transmitted to the new researchers or engineers or simply that the knowledge generated by formal R&D is relevant for a larger number of innovations than the knowledge generated by informal research. Lastly, the size variable is significant, which suggests that firms have not the same access to finance. It is thus necessary to control for these differences.

The simplest way to compare the relative importance of research and past innovation is to compute the average innovation probability difference that they imply (Table 2, columns 4 and 5). We find that a firm that has innovated over 1990-1992 has on average a 19% higher probability to innovate in 1994-1996, once controlled for all the other determinants of innovation. This figure is comparable to the difference associated to the strongest level of formal research and development (18%). According to these first estimates, the achievement of a past innovation would be as important as formal research itself.

Finally, the conclusions of the different theoretical models appear as complementary. But we will see that in fact the relevance of the theoretical models depends on the size of the firms.

5.2. *The Rubin model and the evaluation of the causal effect*

The estimation of the causal effect is based on the first-step estimate of the probability to have innovated in the past (1990-92). Then firms can be matched in different ways. The first one consists in constructing probability classes in which there are firms that innovated in the past and firms that did not innovate. In each probability class, it is possible to evaluate the causal effect, i.e. the specific effect of past innovation on current innovation. The second matching method is the nearest-neighbor method that consists in matching each firm that has innovated in the past with the past non-innovative firm that had the closest probability to innovate.

The estimate of the probability to have innovated over 1990-92, i.e. to have been treated, is equivalent to choose a set of conditioning variables. We retained size, industry dummies, high tech dummy, past research activities (1986-90) and innovation 1986-90 as explanative variable.[23] For each firm, we then get a predicted probability to have been treated and construct probability classes in function of this *propensity score.*

The construction of probability classes requires several steps. First, we exclude from our sample firms that have a propensity score superior to 85% because among these firms, there are nearly only treated firms, such that for these latter firms, it is not possible to evaluate correctly the causal effect (see figure 1). Therefore, the comparison between past innovative and past non-

---

[23] The innovation 1986-90 is used in order to correct a possible individual fixed effect. This estimate relies on a probit model.

innovative firms can only be made for a part of the sample. The probit model does not allow to correct this potential source of bias. Second, we imposed that each class includes at least 30 treated and 30 non-treated in order to be able to evaluate the causal effect. This convention increases the heterogeneity inside the classes, that we control for by adding the propensity score and the product of the treatment and of the propensity score in our regressions. Lastly, inside each class, we check that the conditioning variables are well-balanced between the treated and the non-treated firms. We performed a test for the statistical significance of differences in the distribution of observable variables, focusing on first moment.

These three steps lead to define three probability classes: 0-35%, 35-57% and 57-85%. The first class starts at strictly positive values because, even for low levels of the probability to be treated, there are enough treated and non-treated firms.

We distinguish two types of evaluation: first, we compute the difference of the average probabilities to innovate inside each probability class (Table 3 – regression 1). Secondly, we run a regression of the current innovation dummy on the score, past innovation and the cross product of the two latter variables (Table 3 – regression 2).[24] The purpose of the latter regression is to allow for firm-level variation of the causal effect.

Important differences appear between probability classes (Table 3 – regressions 1 and 2). We find that the causal effect is the strongest for the lowest probability class, around 40%, decreases in the second probability class, around 20 %, and vanishes in the last class.

- *Insert Table 3*-

The class 1 corresponds to small firms that do not invest in R&D activities. In these firms, the causal effect is the strongest. It does not mean that the persistence of innovation is the strongest in this class but that the learning-by doing effects play an essential role in these firms. Moreover, the existence of such effects leads to a strong innovation persistence in the small firms that are not engaged in R&D activities since firms that innovated in the past increase their probability to innovate once again of around 40%.

The causal effect in the class 2 is not far from the half of the causal effect in the class 1, around 20%. The persistence of innovation in these firms is explained not only by their larger size or their R&D activities but equally by the existence of learning-by-doing effects. Indeed there remains a strong effect of the past innovation on current innovation.

The causal effect of past innovation is not significant in the class 3. This result does not mean that the innovation is not persistent within these firms. This class includes firms with a larger size and more R&D. Here, the persistence of innovation comes from the persistence of R&D.

In Table 3, the nearest neighbor estimates are equally presented for each class and for each model. These last estimates are consistent with the results obtained with the probability classes method.

The origin of innovation persistence thus depends on the size of the firm. Consequently, the relevance of the different theoretical models depends on the characteristics of the firms. In the largest firms, the linear model applies whereas in the smallest firms, the relevant model includes learning-by-doing effects. This last conclusion is close to the results of Kleinknecht (1987) and emphasizes the inadequacy of R&D data to evaluate the innovative competencies in the small firms.

This suggests the following functioning of innovation: the importance of learning-by-doing decreases with the formalization of research and development activities. Clearly, the last class includes mostly firms with the highest formal R&D budgets, such that the persistence of innovation for these firms comes from the persistence of research. In order to evaluate the degree of innovation persistence at the firm level, both effects must be accounted for. The fact to omit the *learning-by-doing effect* leads to underestimate the innovation persistence, in particular in small firms.

All the methods used lead to estimate an average causal effect of past innovation on the whole sample around 20%.[25] Consequently, the learning-by-doing effects play an essential role in the persistence of innovation. This last result is very close to the result of the probit model.

The nearest neighbor estimator equally allows to compute the causal effect on the treated (Table 3 – column 5). This effect (14%) is lower than the causal effect on the whole sample (20%). The reason is that the sample of treated includes many large firms. Moreover, this last

[24] We have centered the score before to take the cross products such that the average causal effect is given directly by the coefficient of past innovation.

[25] We recall that the sample does not contain the firms that have a propensity score superior to 85%.

result slightly differs from the average effect of past innovation computed on the sample of the treated with the probit model (18%).

6. Conclusion

The econometric results that we obtain with the Innovation Surveys contrast with the results of previous studies on patent data and seem more in accordance with the theoretical literature.

In a first step, we find that the innovation persistence is strong since, *ceteris paribus,* a firm that already innovated in the past has a stronger probability to innovate today (around +20%). This persistence has several origins. Indeed we find that the origin of the persistence depends on the size of the firm. Whereas the *learning-by-doing* model seems to play a major role in the small firms, its validity is decreasing when the size of the firm is increasing. In the largest firms, it is the linear model that is relevant. In these latter firms we do not find any significant effect of past innovation on current innovation. The innovation persistence is due to the formal research in these firms.

Consequently, the pertinence of the innovation models depends on the size of the firm, such that the *learning-by-doing* model and the linear model are not conflicting but complementary.

References

Aghion P. and Howitt P. (1998). *Endogeneous Growth Theory*, MIT Press.

Aghion P., Harris C. and Vickers J. (1997). Competition and Growth with Step-by-step Innovation: An Example. *European Economic Review*, vol. 41, pp.771-782.

Arora A., Gambardella A., Pammolli F. and Riccaboni M. (2000). The Nature and the Extent of the Market for Technology in Biopharmaceuticals, *CNRS-CEPR International conference on « Technology Policy and Innovation : Economical and Historical Perspectives »*, Paris, November.

Arundel A. and Kabla I. (1998). What Percentage of Innovations are Patented? Empirical Estimates for European Firms, *Research Policy*, 27, pp.127-141.

Audretsch D. and Elston J. (2002). Does firm size matter? Evidence on the impact of liquidity constraints on firm investment behavior in Germany. *International Journal of Industrial Organization*, 20(1), pp.1-17.

Barlet C., Duguet E., Encaoua D. and Pradel J. (1998). The Commercial Sucess of Innovations : An Econometric Analysis at the Firm Level, *Annales d'Economie et de Statistique*, 49-50, pp.457-478.

Cefis E. and Orsenigo L. (2001). The Persistence of Innovative Activities: A Cross-countries and Cross-sectors Comparative Analysis. *Research Policy*, 30, pp.1139-1158.

Chandler A.D. (1990). *Scale and Scope : the Dynamics of Industrial Capitalism*, The Belknap Press of Harvard University Press : Cambridge, MA.

Chesbrough H. (1999). Arrested Development : The Experience of European Hard Disk Drive Firms in comparison with US and Japanese Firms, *Journal of Evolutionary Economics*, 9, pp.287-329.

Cohen W. and Klepper S. (1996). A reprise of size and R&D. *Economic Journal*, vol. 106, pp.925-951.

Cohen W. and Levin R. (1989). Empirical Studies of Innovation and Market Structure. In Schmalensee et al. (eds), *The Handbook of Industrial Organization*, volume II.

Cohen W., Nelson R. and Walsh J. (1997). Appropriability conditions and why firm patent and why they do not in the American manufacturing sector. Mimeo, June.

Crépon B. and Duguet E. (1997). Estimating the knowlege production function from patent numbers: GMM on count panel data with multiplicative errors . *Journal of Applied Econometrics*, vol. 12(3).

Crépon B. and Iung N. (1999). Innovation, emploi et performance. INSEE working paper, DESE, ref. G9904.

Dehejia R.H. and Wahba S. (1998). Causal Effects in Non-experimental Studies : Re-evaluating the Evaluation of Training Programs. *NBER Working Paper Series*, 6586.

Duguet E. and Kabla I. (1998). Appropriation strategy and the motivations to use the patent system in France : an econometric analysis at the firm level. *Annales d'Economie et de Statistique*, 49-50 (Special issue on the economics and econometrics of innovation).

Efron B. and Tibshirani R.J. (1993). *An Introduction to the Bootstrap*, Monographs on Statistics and Applied Probability 57, Chapman e& Hall, New York.

Encaoua D. and Hollander A. (2002). Competition Policy and Innovation, *Oxford Economic Review*, forthcoming.

Encaoua D. and Ulph D. (2000). Catching-up and Leapfrogging? The Effects of Competition on Innovation and Growth. *Working Paper, EUREQua, University Paris 1*.

François J.-P. (1991). Une enquête sur l'innovation. *Courrier des Statistiques*, n°57.

François J.-P and Favre F. (1998). L'innovation technologique progresse dans l'industrie. *Le 4 Pages du SESSI*, n°89.

Geroski P., Van Reenen J. and Walters C. (1997). How Persistently do firms innovate? *Research Policy*, 26, pp.33-48.

Geroski P. (1995). What do we Know about Entry?, *International Journal of Industrial Organization*, 13, pp.421-440.

Griliches Z. (1990). Patents Statistics as Economic Indicators : a Survey, *Journal of Economic Litterature*, 28, pp.1661-1707.

Hall B. and Van Reenen J. (2000). How effective are fiscal incentives for R&D ? A review of the evidence. *Research Policy*, vol. 29, pp.449-469.

Heckman J., LaLonde R. and Smith J. (1998). The Economics and Econometrics of Active Labor Market Programms. *Working Paper prepared for the Handbook of Labor Economics*, III, Ashenfelter O. and Card D. editors.

Heckman J., Ichimura H. and Todd P. (1997). Matching as an Econometric Evaluation Estimator : Evidence from Evaluating a Job Training Programme. *Review of Economic Studies*, 39 (1), pp.33-38.

Henderson R. and Cockburn I. (1996). Scale, Scope and Spillovers: The Determinants of Research Productivity in Drug Discovery. *Rand Journal of Economics*, 27(1), pp.32-59.

Hörner J. (2001). A Perpetual Race to Stay Ahead, *Working Paper*.

Hsiao C. (1986). *Analysis of Patent Data*, New York: Cambridge University Press.

Kim L. (1997). The Dynamics of Samsung's Technological Learning in Semiconductors. *California Management review*, 39(3), spring, pp.86-100.

Kleinknecht A. (1987). Measuring R&D in Small Firms : How Much are we Missing?, *The Journal of Industrial Economics*, 36 (2), december, pp.253-256.

Kleinknecht A. (ed.) (1996). *Determinants of Innovation: The Message from new Indicators*, Mc Million.

Levin R., Klevorick A., Nelson R. and Winter S. (1987). Appropriating the returns from research and development. *Brookings Papers on Economic Activity*, 3, pp.783-831.

Lhuillery S. (1995). L'innovation technologique dans l'industrie. *Le 4 Pages du SESSI*, n°46.

Malerba F. and Orsenigo L. (1999). Technological Entry, Exit and Survival : An Empirical Analysis of Patent Data. *Research Policy*, 28, pp.643-660.

Malerba F., Orsenigo L. and Peretto P. (1997). Persistence of Innovative Activities, Sectoral Patterns of Innovation and International Technological Specialization. *International Journal of Industrial Organization*, 15(6), pp.801-827.

Nelson R. and Winter S. (1982). *An Evolutionary Theory of Economic Change*, The Bellknapp Press of Harvard University Press, Cambridge, MA.

Nightingale P. (2000). Economies of Scale in Experimentation: Knowledge and Technology in Pharmaceutical R&D. *Industrial and Corporate Change*, 9(2), pp.315-359.

Romer P. (1991). Progrès Technique Endogène, *Annales d'Economie et de Statistique*, 22, pp.1-32.

Rosenberg N. (1976). *Perspectives on Technology*, Cambridge UP.

Rosenbaum P. and Rubin D. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70, pp.41-55.

Rubin D. (1977). Assignment to a Treatment Group on the Basis of a Covariate. *Journal of Educational Statistics*, 2, pp.1-26.

Rubin D. (1974). Estimating Causal Effects of Treatments in Randomized and Non-Randomized Studies. *Journal of Educational Psychology*, 66, pp.688-701.

White H. (1980). A Heteroskedasticity Consistent Covariance Matrix Estimator and a Direct Test of Heteroskedasticity, *Econometrica*, 48, pp.817-838.

Table 1 -Sample Statistics

	Innovation 1991 5 years (86-90)	**CIS1 (1993)** 3 years (90-92)	**CIS2 (1997)** 3 years (94-96)
% of innovators			
Consumer goods (210 firms)	63	48	50
Car industry (33 firms)	91	76	79
Other equipment goods (180 firms)	85	79	77
Intermediate goods (385 firms)	79	58	56
Sample total (808 firms)	*77*	*61*	*61*
Innovation inputs (% of firms) :			
Formal research and development			
- none / weak / moderate / strong	29/ 11 / 17 / 43	×	×
Informal research			
- none / weak / moderate / strong	15 / 17 / 32 / 36	×	×
Formal or informal R&D			
- none / weak or moderate / strong / very strong	×	9 / 17 / 40 / 34	×
Other variables :			
Sales 1985 – Millions of euros: Means*			
- Innovative firms / Non-innovative firms	20.4 / 3.0	24.0 / 4.3	24.8 / 3.3
High technological opportunities (% of firms)			
- Innovative firms / Non-innovative firms	71.4 / 10.2	71.1 / 35.7	68.1 / 40.0

Note: Sample of 808 French manufacturing firms of 20 employees or more resulting from the fusion of the three consecutive innovation surveys and of the 1985 E.A.E. survey. * Official conversion rate 1 euro = 6.55957 FRF.

[1]: Information on research activities is available only for firms that innovated during the period considered. In regressions, we set the research variables of the firms that have not innovated to 0. The 1997 innovation survey equally questions firms on their research activities but only for the year 1996 such that this information can not be used in our model.

Table 2 – The probit models and the average effects

Variables	**Model 1**	**Model 2**	**Average effect[a]** Whole sample	**Average effect[a]** Sample of treated
- Intercept	-0.79* (0.12)	-4.45* (0.62)	×	×
- Innovation 86-90	0.42* (0.13)	0.11 (0.18)	×	×
- Innovation 90-92	0.84* (0.10)	0.56* (0.21)	0.19	0.18
- Industry[b]				
- Car industry	0.46 (0.28)	0.16 (0.29)	×	×
- Other equip. goods	0.43* (0.15)	0.38* (0.16)	0.11	0.09
- Intermediate goods	0.62 (0.11)	0.04 (0.12)	×	×
-- High tech dummy	0.22 (0.11)	0.06 (.12)	×	×
- Formal R&D 86-90[c]				
- weak	×	0.31 (0.20)	0.11	0.09
- moderate	×	0.29 (0.18)	0.11	0.09
- strong	×	0.45* (0.16)	0.18	0.16
- Informal R&D 86-90				
- weak	×	0.16 (0.20)	×	×
- moderate	×	-0.06 (0.18)	×	×
- strong	×	0.03 (0.18)	×	×
-Formal or informal R&D 90-92				
- weak	×	-0.21 (0.25)	×	×
- moderate	×	0.07 (0.23)	×	×
- strong	×	0.26 (0.24)	×	×
- ln(Sales)	×	0.21* (0.03)	×	×
Log-likelihood	-452.88	-408.92	×	×
% correct predictions	73.6%	81.9%	×	×

Note: Sample of 808 French manufacturing firms of 20 employees or more resulting from the fusion of the three consecutive innovation surveys and of the 1985 E.A.E. survey. Left-hand variable : implementation of a product or process innovation between 1994 et 1996. Maximum likelihood estimates of the probit model (standard errors between parentheses). * : significant at 1%. ** : significant at 5%.

[a]: The average effects are computed only for significant binary variables of the model 3.[b]: The consumer goods industry is the industry of reference. [c]: For research activities, the reference is the modality none.

Figure 1 - Distributions of the propensity score among treated and non-treated firms

(Kernel density estimator with a gaussian kernel and a Silverman window)

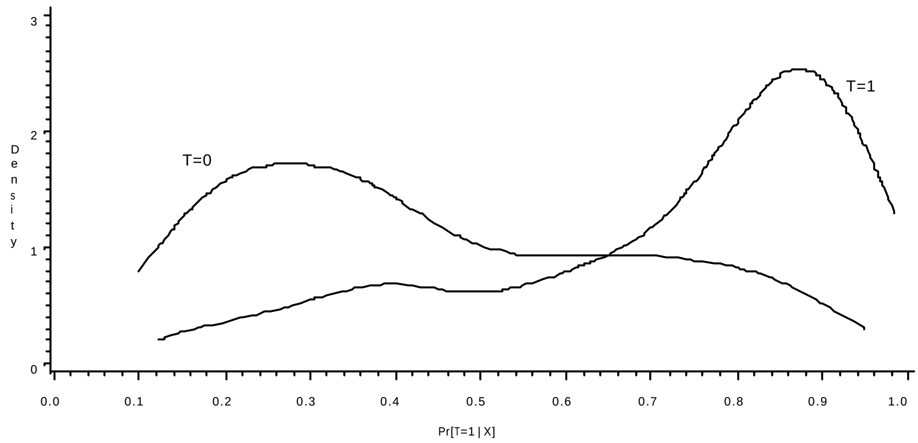

Table 3 – Average causal effect of innovation 1990-92 on innovation 1994-96 for the whole sample

Causal effect	**Causal effect on the whole sample**			**Causal effect on the treated**
Estimator	**The probability classes estimator: regression 1**	**The probability classes estimator: regression 2**	**The nearest neighbor estimator**	**The nearest neighbor estimator**
Class 1 $0\%<\Pr(T=1 \mid X)\leq 35\%$	0.38* (0.08)	0.40* (0.08)	0.37* (0.09)	×
Class 2 $35\%<\Pr(T=1 \mid X)\leq 57\%$	0.23* (0.08)	0.23* (0.08)	0.19* (0.09)	×
Class 3 $57\%<\Pr(T=1 \mid X)\leq 85\%$	0.13** (0.06)	0.07 (0.06)	0.07 (0.07)	×
Weighted Estimator	0.23* (0.04)	0.21* (0.04)	0.19* (0.05)	×
Global Estimators	×	×	0.20* (0.05)	0.14* (0.07)

* : significant at 1%. ** : significant at 5%.

Note: The probability classes estimator- regression 1: The regression comprises a constant term and the treatment variable. The heteroskedasticity robust asymptotic standard errors are between parentheses. The probability classes estimator- regression 2: The regression comprises a constant term, the treatment variable, the propensity score and the product of the treatment and of the propensity score. The heteroskedasticity robust asymptotic standard errors are between parentheses. The nearest neighbor estimator: The asymptotic standard errors obtained by the bootstrap method are between parentheses (100 simulations).

SUPPLEMENTARY SECTION

DATA APPENDIX

<u>The innovation surveys</u>

The first innovation survey in France, namely *"l'innovation technologique dans l'industrie"*, was conducted in 1991. The firms were asked to report retrospectively over the 1985-1990 period. Hence, the choice of the respondent was an important issue. Here intervenes the SESSI (Industrial Statistics Bureau of the Ministry of Industry) which is responsible of the Industry Census and of all innovation surveys in France (and more surveys). The basic organization is as follows: inside SESSI the same person always works with the same set of firms. A part of his (or her) job is to find the right interlocutor inside the firm. On each questionnaire appear the name and the phone number of the SESSI correspondent inside the firm. Here the correspondent (which is an employee of the firm) has to send the questionnaire to "a person responsible of innovation, development, strategy issues or to the boss himself" (literal translation). The name of the respondent, that can be different from the name of the correspondent, and its phone number, have to be systematically reported on the questionnaire. The respondent has a SESSI phone number he (or she) can use to have explanations on how to reply to the survey. The census file is used for the mailing that prints automatically the name of the correspondents on the questionnaire itself etc. In other words, this survey has been conducted by an administration that has for main purpose to collect data among industrial firms.

The survey was presented as an appendix to the Industry Census, which is compulsory. While the Census was compulsory, the appendix was not, but it *was not indicated* on the questionnaire such that the firms could have believed that it was compulsory. This is likely to be the case since the response rate to the innovation survey is the same as the one of the industry census (85% for compulsory surveys in France). The possibility of a response bias is systematically studied by the specialists of SESSI, for all the surveys. They compute the response rate after the "first wave" of the survey for each size class and each industry in order to detect abnormal response rates (e.g., below 85%). When the questionnaire does not come back, they can launch a second wave.

Last but not least, all French firms have a compulsory national identification number that is called the SIREN code. The use of this code is compulsory for all the relationship that a firm has with the administration (including taxes). Its main advantage is that it allows for matching all the surveys without loss for identification reasons.

The first innovation survey is linked to the Community Innovation Survey (CIS) since it was designed to prepare the future CIS. The information collected is made up of answers over the 1985-1990 period, such that no annual information is available. This survey provides information on the type of innovation that industrial firms have implemented, as well as which knowledge sources they have used to achieve these innovations. We have qualitative information (yes / no) about eight types of innovation including the five following types of innovation:

- Significant improvement on an existing product.
- Introduction of a product that is new for the firm and for the market;
- Introduction of a product that is new for the firm but not for the market;
- Significant improvement on an existing process;
- Process technological breakthrough ("Première de procédé technologique").

A firm that has performed at least one of these five types of innovation is considered as innovative. We take this definition because it corresponds to the one used in the CIS surveys (product or process innovation).[26] The questionnaire then turns to the sources of these innovations. The questionnaire design clearly indicates the causality in the following way (literal translation) "*Sources* of innovations: in your firm, does the introduction of innovation *result from*:" and then comes the list of the innovation inputs. The importance of innovation sources is available on a four-point scale. The scale is: unimportant, weakly important, moderately important and very important. The two inputs used in this paper are formal and informal R&D:

- Formal R&D is defined as internal research and development with at least one full-time employee. This is the definition that is used for the R&D survey and is the closest to the traditional R&D studies (i.e., from the Frascatti criterion).
- Informal R&D is defined as "internal method and technical studies". The interest of this measure is to avoid the undercounting of research in small firms, a common feature of most databases (Kleinknecht, 1987; Kleinknecht and Reijnen, 1991).

[26] The remaining three types are: organizational innovation, marketing innovation and packaging innovation.

CIS 1 and CIS2

CIS 1 is the first international survey on innovation. It was also conducted by SESSI and reached the same response rate as the first survey. Notice that we do not use the micro-aggregated version of the survey but the original French survey in which information is available at the firm level. Fewer firms were surveyed than in the original innovation survey of 1991 (see below). It provides information about the implementation of product and process innovation and on the innovation inputs. In order to keep comparable specifications across time, we have kept the formal and informal research inputs that are grouped together in only one question. The answer is on a five points scale: “not important”, “weakly important”, “moderately important”, “important” and “very important”. CIS 2 is the second international survey on innovation. It was conducted in the same conditions as the two other surveys and reaches the same response rate. It includes many informations but we just keep the implementation of a product or a process innovation. For more information about the data sources, see François (1991), Lhuillery (1995) and Favre and François (1998).

The sample

Our sample results from the merger of these three innovation surveys and of the industry census in 1985. We impose the presence in 1985 because the first survey questionnaire refers to the period 1986-1990. The sample includes 808 firms. Even though merging is easily done with the SIREN code, we lose two types of firms by performing this operation.

The first kind of firms that we lose are the firms that were not included in all the surveys. This allocation is random, such that is should not be a source of bias. The second kind of firms that we lose are the firms that did not survive the whole period. Here, we should check that our data are representative of the exit rates of the industry. Since there are both innovative and non-innovative firms in our sample, we should not expect that our firms survive longer.

The appendix tables gives the detail of the constitution of the sample and compares the entry and exit rates of the innovation survey with the ones of the whole industry. It clearly shows the innovation surveys are representative of the cohort of firms that were present in manufacturing in 1985. It is especially true of the first survey since all the firms in the CIS1 survey were respondents to the first innovation survey of 1991. The CIS2 survey, on the contrary, includes a large number of firms newly included in the survey. Globally, we can say that our sample is representative of the exit rate of the industry but that it exhibits a smaller entry rate. Therefore, we should interpret our results as valid for the cohort of manufacturing firms in 1985.

Table A.1 - Details on the Merger of the Innovation surveys (2 by 2)

File 1	**Innovation 1991**	**Innovation 1991**	**CIS1**
Number of firms in file 1 **(A)**	15498	15498	3342
File 2	**CIS1**	**CIS2**	**CIS2**
Remaining number of firms after the merger of file 1 and file 2	3329	3882	3881
Explanation of the variation of the number of firms after the merger of file 1 and file 2 :			
Entry in the survey (from file 1 to file 2)			
- old business included in file 2	0	821	2840
- new business included in file 2	2	558	284
- total	2	1379	3124
Exit from the survey (from file 1 to file 2)			
- by bankruptcy between the two surveys	2626	5330	693
- by exclusion from the second survey (these firms are still in the census)	9545	7665	1892
- total	12171	12995	2585
Entry in the survey – Exit from the survey **(B)**	-12169	-11616	+539
Remaining firms after the merger of file 1 and file 2 **(A) + (B)**	**3329**	**3882**	**3881**
Annualized Entry and exit rates (corrected for pure sampling movements):			
Exit rate from the census	10.3%	5.8%	6.1%
Exit rate from the CIS	8.1%	5.2%	4.8%
Entry rate in the census	6.6%	4.9%	5.9%
Entry rate in the CIS	<0.1%	2.3%	1.8%

Note about our sample: the merger of the three files gives 808 firms. This small figure comes mostly from the variation of the sampling of the two last innovation surveys (CIS1 and CIS2) since the exit rates are about the same in the census and in the CIS. Therefore, our sample is representative of the firms that belong to the cohort of 1985.

We equally give the results of the test for the statistical significance of differences in the distribution of observable variables and the mean characteristics of each class.

The t-tests

Classes	Class 1 0-35%	Class 2 35-57%	Class 3 57-85%
- Innovation 1986-90	-0.42 (.67)	-0.97 (.33)	-0.61 (.54)
- Sales 1985	-1.96** (.05)	-0.72 (.47)	-0.40 (.69)
- Industry			
- Car industry	-0.35 (.73)	-0.23 (.82)	-0.91 (.36)
- Other equipment goods	1.33(.18)	0.98 (.33)	-0.72 (.47)
- Intermediate goods	-0.43 (.67)	-0.61 (.54)	0.30 (.76)
-High tech dummy	0.21 (.83)	0.07 (.94)	-0.30 (.77)
-Formal R&D 1986-90			
- weak	×	-1.93** (.05)	1.26 (.21)
- moderate	PM	1.32 (.19)	-0.32 (.75)
- strong	×	-0.23 (.82)	-1.31 (.19)
-Informal R&D 1986-90			
- weak	-0.11 (.91)	-0.84 (.40)	0.43 (.66)
- moderate	-1.17 (.25)	-0.39 (.70)	0.14 (.89)
- strong	×	-0.55 (.58)	-1.31 (.19)
Number of treated	42	66	204
Number of non-treated	135	91	72

Note : Between parentheses, p-value for testing the hypothesis that the true means of two groups of observations are equal. "PM" indicates perfect matching. ×: no point in the class.

Characteristics of firms in the different classes

Means of the different samples

Classes	**1**	**2**	**3**
-Innovation 1986-90	21.5	73.2	98.2
-Sales[a]	42	223	504
-Industry			
- Car industry	1.7	1.3	6.1
- Other equipment goods	5.6	12.1	18.1
- Intermediate goods	54.2	47.1	55.4
-High tech dummy	7.9	38.2	73.5
-Formal R&D 1986-90			
- weak	0	8.3	17.4
- moderate	0.6	5.7	26.4
- strong	0	12.7	35.5
-Informal R&D 1986-90			
- weak	6.8	21.0	17.7
- moderate	3.3	19.7	34.0
- strong	0	13.4	35.5
Number of treated		401	
Number of non-treated		309	

Note: The different samples used are defined in the table 7.