NBER WORKING PAPER SERIES

RESEARCH, INNOVATION, AND PRODUCTIVITY: AN ECONOMETRIC ANALYSIS AT THE FIRM LEVEL

Bruno Crépon
Emmanuel Duguet
Jacques Mairesse

Working Paper 6696
http://www.nber.org/papers/w6696

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 1998

We are thankful to B. H. Hall for her generous and valuable collaboration in helping us to revise and rewrite this paper. We have also benefited from comments by participants in the Conference on the *Effects of Technology and Innovation on Firm Performance and Employment* (Washington D.C., May 1995), and in various other conferences and seminars where successive incarnations of our work have been presented. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

Research, Innovation, and Productivity:
an Econometric Analysis at the Firm Level
Bruno Crépon, Emmanuel Duguet,
and Jacques Mairesse
NBER Working Paper No. 6696
August 1998
JEL No. C31, C36, L60, O31, O33

ABSTRACT

This paper studies the links between productivity, innovation and research at the firm level. We introduce three new features: (i) A structural model that explains productivity by innovation output, and innovation output by research investment; (ii) New data on French manufacturing firms, including the number of European patents and the percentage share of innovative sales, as well as firm-level demand pull and technology push indicators; (iii) Econometric methods which correct for selectivity and simultaneity biases and take into account the statistical features of the available data: only a small proportion of firms engage in research activities and/or apply for patents; productivity, innovation and research are endogenously determined; research investment and capital are truncated variables, patents are count data and innovative sales are interval data.

We find that using the more widespread methods, and the more usual data and model specification, may lead to sensibly different estimates. We find in particular that simultaneity tends to interact with selectivity, and that both sources of biases must be taken into account together. However our main results are consistent with many of the stylized facts of the empirical literature. The probability of engaging in research (R&D) for a firm increases with its size (number of employees), its market share and diversification, and with the demand pull and technology push indicators. The research effort (R&D capital intensity) of a firm engaged in research increases with the same variables, except for size (its research capital being strictly proportional to size). The firm innovation output, as measured by patent numbers or innovative sales, rises with its research effort and with the demand pull and technology indicators, either directly or indirectly through their effects on research. Finally, firm productivity correlates positively with an higher innovation output, even when controlling for the skill composition of labor as well as for physical capital intensity.

Bruno Crépon
INSEE-DMSE
Timbre G320
15, boulevard Gabriel Peri
B.P. 100
92244 Malakoff Cedex
bruno.crepon@dg75-g230.insee.atlas.fr

Emmanuel Duguet
Univerity of Paris 1 - CEME
106-112, boulevard de l'Hopital
75647 Paris Cedex 13
eduguet@univ-paris1.fr

Jacques Mairesse
CREST-INSEE
15, boulevard Gabriel Peri
92245 Malakoff, FRANCE
and NBER
mairesse@ensae.fr

1 Introduction

This paper proposes an original empirical approach to the problem of assessing both the innovation impacts of research and the productivity impacts of innovation and research. Building upon previous studies, it is also an attempt to confirm and summarize our main findings in these studies.[1] We consider a model summarizing the process that goes from the firm decision to engage in research activities to the use of innovations in its production activities. In specifying this model, we take advantage of new sources of data available for French manufacturing firms, and in estimating it, we try to correct for various selectivity and simultaneity biases that affect many research and development (R&D) and patent studies (for a survey of such studies, see Cohen and Levin, 1989, and Mairesse and Sassenou, 1991).

More precisely, we introduce three new features in the analysis. Firstly, we explicitly account for the fact that it is not innovation input (R&D) but innovation output that increases productivity. Firms invest in research in order to develop process and product innovations, which in turn may contribute to their productivity and other economic performances. Our model thus includes three relationships: the research relation linking research to its determinants, the innovation equation relating research to innovation output measures, and the productivity equation relating innovation output to productivity. The first relation corresponds in fact, as we shall see, to two equations, respectively accounting for the R&D investment decision and for its size.

Second, in addition to the more usual information on the firm current accounts, balance sheets and employment numbers, and in addition to the firm R&D expenditures collected by the annual Survey on Research, we use new data on innovation output in French manufacturing. These are the number of European patents applications which have been matched to the firm data, and the share of firm innovative sales (i.e., firm sales from the new products introduced in the last five year period), which we gathered from the 1990 French Survey on Innovation in manufacturing. From this survey, we also obtained two new indicators proxying for demand conditions and technological opportunities (or "demand-pull" and "technology push" indicators). Last, we have been able to construct a firm average market share variable and diversification index, based on detailed data by lines of business from the Annual Firm Survey.

Thirdly, we estimate our model using econometric methods that can deal with the many problems inherent to this model and to the nature of the data. Most studies on innovation are, for example, potentially affected by selectivity biases. Only a minority of firms are engaged in (formal) R&D activities, so that studies restricted to these firms are prone to such biases. Also only relatively few firms have patents, and thus analyses limited to them may be similarly biased. In addition, patents being count data and the percentages of innovative sales being recorded as interval data, both require

[1] See Cuneo and Mairesse (1984), Mairesse and Cuneo (1984), Crepon and Mairesse (1993), Hall and Mairesse (1995), Crepon and Duguet (1996), Crepon, Duguet and Kabla (1996), Crepon and Duguet (1997). All these studies like the present one concern French manufacturing and are based on firm level microdata.

specific econometric methods to handle them. Finally, there is the major issue of the endogeneity of innovative input and output, and more generally of the simultaneity in our model. R&D is endogenous in the innovation equation and patents or innovative sales are endogenous in the productivity equation. The disturbances in the equations of our model, reflecting in part unobserved variables and firm effects, are also likely to be correlated.

We treat all these estimation problems by relying on methods which do not seem to have been previously applied in the research and innovation literature. We take care of selection and of the specific nature of variables by using a generalized tobit specification for R&D investment, a heterogeneous count data specification for patents, and an ordered probit specification for the interval data on innovative sales. Our model thus amounts to a system of non-linear equations with limited dependent and count data variables, and we deal with simultaneity in this system by using a two-stage estimation procedure. The first stage amounts to applying the method of moments (M-estimation) to the reduced form equations, and we rely in the second stage on the method of asymptotic least squares (ALS-estimation) to retrieve consistent estimates of the structural parameters. This procedure requires relatively few assumptions on the distributions of the disturbances and is flexible enough with modest computational cost.

In order to be able to identify and estimate our model, we have of course to make some a priori assumptions on its overall structure and the specification of individual equations. As will be explained, these assumptions seem rather reasonable; however we cannot really test them. Clearly, the main drawback of our study in this respect is the cross-sectional nature of our data and estimates.

The organization of our paper is as follows. To keep its exposition simple, we focus on the main points in the text, and give further explanations and results in three appendices. The definition of variables and the econometric specification of the model are explained in section 2, the main results discussed in section 3, and some very brief concluding comments given in section 4. The construction of the sample and the definition of variables are presented at some length in Appendix A, the estimation methods are detailed in Appendix B, and our estimates are compared with those based on the potentially biased more widespread methods in Appendix C

2 The model: definition of variables and econometric specification

Figure 1 lays out a schematic diagram showing the general structure of our model.[2] It consists of four equations, two for research, one for innovation and one for productivity, which we shall present in turn, each of them requiring a different econometric treatment. We consider in fact two versions of the model, and for each a 'basic' and an 'extended' specification. In the first version of the model 'innovation output' is measured by the number of patents and in the second by the share of innovative sales. The extended specification includes the technology push and demand pull indicators

[2] See Pakes and Griliches (1984) for a similar diagram.

as explanatory variables in the research and innovation equations and two indicators of skill composition as controls for 'labor quality' in the productivity equation, while the basic specification does not.

2.1 The research equations

To depict the firm research behavior, we rely on a generalized tobit model (Heckman, 1976, 1979), with two equations, the first equation accounting for the fact that the firm is engaged in research activities (or that we observe that it is), and the second one for the magnitude or intensity of these activities.

More precisely, we assume that there exists a latent dependent variable g_i^* for the firm i given by a first equation:

$$g_i^* = x_{0i}b_0 + u_{0i} \tag{0}$$

where x_{0i} is a vector of explanatory variables, b_0 the associated coefficient vector and u_{0i} an error term, and where g_i^* expresses some decision criterion, such as the expected present value of the firm profit accruing to research investment. We observe that the firm invests in research if g_i^* is positive or larger than some constant threshold, overall or industry specific (provided x_0 contains a constant term or industry dummies, which is the case for all our equations in this analysis). Actually only a small proportion of French manufacturing firms engage in formal research activities and report R&D expenditures, though many more have some form of innovative activities.

We then assume that a latent or true intensity of research k_i^* for firm i is determined by a second equation:

$$k_i^* = x_{i1}b_1 + u_{1i} \tag{1}$$

with $k_i^* = k_i$, the actual research capital per employee of firm i when this firm does research (that is when g_i^* is larger than the industry threshold), where both k_i and k_i^* are expressed in logarithms, and where x_{1i} is the vector of explanatory variables, b_1 the corresponding coefficient vector and u_{1i} is a disturbance that summarizes omitted determinants and other sources of unobserved heterogeneity. Note that the explanatory variables in the two equations need not be the same, but that without good a priori reason to do otherwise we actually include the same set of variables ($x_0 = x_1$) in both equations.

Finally, because k_i^* is only observable when g_i^* is larger than the industry threshold, we have also to specify their joint distribution in order to get an estimable model. We thus assume the joint normality of the disturbances in the two equations (i.e., the generalized tobit model assumption):

$$\begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix} \overset{iid}{\leadsto} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \rho\sigma_0\sigma_1 \\ \rho\sigma_0\sigma_1 & \sigma_1^2 \end{pmatrix} \right)$$

where σ_0 and σ_1 are the standard errors of u_{0i} and u_{1i} and ρ is their correlation coefficient. The first equation is in fact a probit equation which is not fully identifiable, and we can only estimate the parameter vector b_0/σ_0 which is equivalent to normalizing the standard error $\sigma_0 = 1$.

In the implementation of our model we prefer to use a stock measure of research rather than a flow measure, and we have estimated our actual research capital (per employee) variable k_i for the firm based on the sum of firm i deflated and depreciated past R&D spending ('deflated in 1990 prices' and 'depreciated' at a given 15 percent yearly rate, as described in Appendix A). We also experimented with using a flow measure for research, with little change to the results.[3]

The explanatory variables we are able to measure are most of the ones usually considered in the literature on R&D determinants, (in the 'Schumpetarian tradition'): size, market share, diversification, and demand conditions and technological opportunities.[4] More precisely, they are:

$$x_{0i} = x_{1i} = \left(l_i, s_i^w, d_i, \delta_i^1, \delta_i^2, \delta_i^3, \tau_i^1, \tau_i^2, \tau_i^3, S_i^1, S_i^2, \ldots, S_i^{18}\right)$$

where l_i is employment, s_i^w the average market share and d_i the equivalent number of industry segments, these three variables being expressed in logarithms (like k_i and k_i^* the observed and latent research capital per employee). The δ_k's and the τ_k's, $k = 1, 2, 3$, are two sets of demand pull and technology push dummies. The S_i^j: are eighteen industry dummies equal to one if firm i belongs to industry j and zero otherwise.[5]

The average market share and the diversification index are computed from the firms' sales decomposition by lines-of-business, or product lines or industry segments, as given by the Firm Annual Survey. For each firm, we have the decomposition of sales into the different industry segments at the so called level 600 of the French industry classification (NAP 600). We first compute the firm's total sales for each industry segment at level 600 in which it operates, and then a market share for each of these industry segments. Thus a diversified firm has several market shares, one per industry segment, and in order to get an overall market share indicator s_i^w for the firm as a whole, we compute a weighted average, where the weight of each segment is its share in the total sales of the firm. Firm diversification is usually characterized by the Herfindahl index h_i of its lines of business, where h_i equals one when the firm is not diversified and decreases with increasing diversification. We find it more telling, though, to use as our diversification index d_i the inverse of the Herfindahl index. This index d_i can be interpreted as the equivalent number of industry segments if all the firm's segments had the same size, and it is bounded between one, if the firm is not diversified, and the real number of different segments, when they have equal weights.

[3] This is not of course too surprising, our analysis being basically a cross-sectional one (and the cross-sectional correlations between the stock and flow measure of research being very high).

[4] See in particular Cohen and Levin (1989), who discuss the importance of demand, technological opportunity and appropriability as determinants of research and innovative activity. They consider them, however, mostly at the industry level, while we try to characterize them here at the firm level and introduce dummies to control for industry differences.

In a previous paper, Crepon, Duguet and Kabla (1996) also used an indicator of imitation measured at the sub-industry level as a proxy for (the lack of) appropriability. We thought this proxy was somewhat too problematic and preferred not to include it in this new analysis.

[5] Note that these dummies replace the constant term, since for all i we have $\sum_{j=1}^{18} S_i^j = 1$. Each industry is thus allowed to have a different intercept (i.e., a different threshold and a different average research capital).

The demand pull and technology push dummies δ_1, δ_2 and δ_3, and τ_1, $\tau_2$and τ_3 are obtained from the 1990 Innovation Survey. They respectively express whether, in the opinion of the firm, demand and technology factors had a 'weak', 'moderate' or 'strong' influence on its innovative activities over the last five year period 1986-90, relative to an answer of 'no' influence.[6]

2.2 The innovation equations

The next equation in our model is an innovation function, whose exact formulation depends on whether we proxy innovation output by the number of patents n_i or by the share of innovative sales t_i.

2.2.1 The patent equation

Patents being observed as integer numbers, we specify the patent equation as a heterogeneous count data process with an expectation n_i^* conditional on research and other variables given by:

$$n_i^* = \mathrm{E}\left(n_i | k_i^*, x_{2i}, u_{2i}; \alpha_K, b_2\right) = \exp\left(\alpha_K k_i^* + x_{2i} b_2 + u_{2i}\right) \tag{2a}$$

where k_i^* is our latent research variable, x_{2i} is a vector of other explanatory variables (supposedly exogenous), and u_{2i} is the error or heterogeneity term, about which we do not need to make any specific distributional assumption.[7] Our patent variable, like our research capital variable, is also a stock measure; it is simply the total number of European patents applied by the firm over the five year 1986-1990 period. The coefficient α_K is the elasticity of the expected patent numbers relative to research capital, a measure of the impacts or returns to research on innovation output, while b_2 is the vector of coefficients for x_2. As exogenous variables x_2 we take:

$$x_{2i} = \left(l_i, , d_i, \delta_i^1, \delta_i^2, \delta_i^3, \tau_i^1, \tau_i^2, \tau_i^3, S_i^1, S_i^2, \ldots, S_i^{18}\right)$$

with the same notations as above. Note that we thus explicitly assume that market share and diversification do not enter directly in the innovation equation, but only indirectly through research. This imposes some a priori structure on the model, which seems reasonable enough, and which helps identification by allowing us to take the two market share and diversification variables as instruments. By contrast, it does not seem unlikely that demand pull and technology push factors could affect innovation output both directly and indirectly.

Including size in the equation also allows us to test whether the effect of firm size on innovation output passes completely through the size of the research activities, i.e., or whether there is 'constant returns to firm size'. For convenience sake, we normalize our patent equation so that n_i and n_i^* are in fact the actual and expected number of

[6] Note that we keep only three dummies to characterize demand pull or technology push, the answer 'no' dummy adding up to one with them, and thus being redundant (exactly collinear) with the constant or the set of industry dummies.

[7] Actually, we estimate the patent equation (more precisely the corresponding reduced form equation) by maximizing the pseudo log likelihood function corresponding to the assumption of a negative binomial pseudo distribution (see Appendix B, subsection B.2.2).

patents per employee (just as k_i and k_i^* are the actual and expected research capital per employee), and hence the coefficient of l_i in the equation estimates the deviation to constant returns to firm size.

2.2.2 The innovative sales equation

The French Innovation Survey, on which we rely to construct the demand and technology indicators, also provides information on the percentage share of firm innovative sales.[8] The firm is asked what is approximately the percentage share of its 1990 sales which comes from products launched in the market in the last five years 1986 to 1990, the firm having to answer on a four point scale: 0-10%, 10-30%, 30-70% and 70-100%. The underlying (unobserved) true share of sales t_i^* can be viewed as another innovation intensity variable, where innovation is measured by a number of sales weighted innovations (instead of the number of patents). Since the share is only known by intervals, we specify the innovative sales equation as an ordered probit model:

$$t_i^* = \alpha_k k_i^* + x_{2i} b_2 + u_{2i} \tag{2b}$$

where t_i^* is the underlying (unobserved) true share expressed in logarithm and the explanatory variables are the same as for the patent equation, and where we assume that the error term u_{2i} is normally distributed with mean zero and variance σ_2^2 .Note that all the parameters are identifiable (including σ_2^2), since the threshold values are known.

2.3 The productivity equation

Our last relationship is the productivity equation, for which we take, as most usually done, an augmented Cobb-Douglas production function with physical capital, employment, skill composition, and innovation output. In this equation innovation output is measured either by expected patents per employee $\ln(n_i^*)$ or by the latent share of innovative sales t_i^* , both in logs. We have

$$q_i = \alpha_I \ln(n_i^*) + x_{3i} b_3 + u_{3i} \tag{3a}$$

or

$$q_i = \alpha_I t_i^* + x_{3i} b_3 + u_{3i} \tag{3b}$$

where q_i is the logarithm of labor productivity defined as log-value added per employee, and where the vector of the factors of productivity, other than innovation output, x_{3i} is:

$$x_{3i} = \left(l_i, c_i, E_i, A_i, S_i^1, S_i^2, \ldots, S_i^{18}\right)$$

with c_i being the logarithm of physical capital per employee (and physical capital the gross book value of fixed assets adjusted for inflation), and with E_i and A_i being respectively the shares of engineers and administrators in the total number of employees.

[8] There are in fact two questions in the Survey, one for total sales and one for export sales; in this study, we only use the former.

The coefficient α_I is the elasticity of total factor productivity with respect to innovation output, while b_3 consists of the following coefficients: the elasticity of scale (more precisely its deviation from unity), that of physical capital, and the skill composition parameters, reflecting percentage differences in efficiency of skilled labor (engineers and administrators) relative to 'unskilled' labor.[9] Note that by having expected patents instead of actual patents in the productivity equation, we do not restrict estimation to the subsample of firms with at least one patent, but use all the sample observations. Note also that we do not make a specific distributional assumption for the disturbance u_{3i} in the productivity equation.

2.4 The overall model and its estimation

Taken together, the research, innovation, and productivity equations form a recursive nonlinear system, for which we have two versions, one with patent counts and the other with the share of innovative sales. The first two equations are the same in the two versions:

$$g_i^* = x_{0i}b_0 + u_{0i} \tag{0}$$

$$k_i^* = x_{1i}b_1 + u_{1i} \tag{1}$$

while the two last equations re respectively the following:

$$\ln n_i^* = \ln \mathrm{E}\left(n_i | k_i^*, x_2, u_2\right) = \alpha_K k_i^* + x_{2i}b_2 + u_{2i} \tag{2a}$$

$$q_i = \alpha_I \ln n_i^* + x_{3i}b_3 + u_{3i} \tag{3a}$$

and:

$$t_i^* = \alpha_K k_i^* + x_{2i}b_2 + u_{2i} \tag{2b}$$

$$q_i = \alpha_I t_i^* + x_{3i}b_3 + u_{3i} \tag{3b}$$

In both versions, we allow for arbitrary correlations among the disturbances u_{0i}, u_{i1}, u_{2i} and u_{3i}. In estimating this system of equations, we want to take account of the nature of the available data: research investment and hence research capital are truncated,

[9] The precise interpretation of the skill composition parameters is the following. Assume that labor corrected for quality ('skill composition') should enter in the production function as :

$$L^* = f_U L_U + f_E L_E + f_A L_A$$

instead of uncorrected labor (total employment) :

$$L = L_U + L_E + L_A$$

where subscript U stands for 'unskilled' production workers , subscript E for engineers, and subscript A for administrators. We can write:

$$\begin{aligned} L^* &= f_U (L - L_E - L_A) + f_E L_E + f_A L_A \\ L^* &= f_U L \{1 + (f_E/f_U - 1)L_E/L + (f_A/f_U - 1)L_A/L\} \\ \ln L^* &\simeq \ln f_u + \ln L + \mu_E E + \mu_A A \end{aligned}$$

and thus the coefficients of E and A in the production function are :

$$b_{3E} \simeq (\partial Q/\partial L)\mu_E \quad and \quad b_{3A} \simeq (\partial Q/\partial L)\mu_A$$

patents are count data and innovative sales are interval data. We also endeavor to take into account the fact that research capital is endogenous in the innovation equations, and innovation output endogenous in the productivity equation. Thus we hope to avoid the potentially most serious selectivity and simultaneity biases.

This of course argues for the use of some kind of simultaneous equations system estimator. Using maximum likelihood would be impossible. The joint distribution of observable variables does not have a closed form, and numerical integration seems intractable (considering the number of integrals involved and the large size of the sample). We have to rely on another (possibly less efficient) method of estimation. We could have chosen the generalized method of moments (GMM), but preferred to implement the asymptotic least squares method (henceforth, ALS).

Three arguments favor the ALS method.[10] First, the efficiency comparisons that have been made by Lee (1981) for a number of models show that in a number of situations the asymptotic least squares method is more efficient in large samples.[11] Second, the computational cost is less: while optimal GMM requires at least two steps, each of them involving all the sample observations, ALS requires two steps as well, but the second one only involves the number of auxiliary parameters estimated in the first step. Third, ALS can be easily generalized to more complicated systems, providing an unified and tractable framework for the estimation of limited dependent variables systems.[12] The application of the ALS method to our model is thoroughly explained in Appendix B.

3 The results

3.1 Simple statistics

Table 1 presents some simple statistics for the variables in our model for two samples of firms: the full sample of firms (N=6145) for which we could match the different data sources we use, and the subsample of those firms (N=4164), reporting in the Innovation Survey that they have made, during the 1986-1990 period, some kind of innovation (either minor or major, new for the market or only for the firm, either product, process, organizational, or marketing). Only these firms were asked to answer the questions on the share of innovative sales, and on the importance of demand pull and technology push. Although the subsample of these firms, which we designate as the innovation sample, is about two thirds of the full sample in size, and its average characteristics are a little different from that of the full sample, the comparable estimations we performed on both samples are quite close. We will only comment here on the main estimates we get with the innovation sample, for which we have the innovative sales, demand and technology variables. In Appendix C, however, we document all estimates for both

[10]For details on this method of estimation, see Gouriéroux, Monfort, and Trognon (1985) or Gouriéroux and Monfort (1996).

[11]Lee's study (1981) does not experiment with the generalized tobit and the count data models. However, in all the cases he considers, including probit, simple tobit, censored models, and linear regression, ALS is more efficient.

[12]Note also that the Chamberlain's PI matrix method for panel data is in the ALS class (see Chamberlain 1982 and Crepon and Mairesse 1996).

samples.[13]

Of particular interest in Table 1 are the variables relating to research and patents. Only about 11 percent of the firms in the full sample and about 15 percent in the innovation sample are R&D-doing firm, and only about 12 percent of the firms in the full sample and about 16 percent in the innovation sample have at least one patent, and half of these in fact having no more than two patents.[14] It is also noteworthy that the overlap between R&D-doing and patenting firms is about 50 percent in the two samples, i.e., half of the R&D-doing firms have patents and half of the patenting firms do R&D.

Looking at the other variables, we see that the innovating firms are slightly larger than those in the full sample, with a median number of employees of 85 versus 69, and are more productive and more capital intensive by about respectively 5 percent and 10 percent at the median. The firms in the two samples tend to have very low market shares, less than half a percent at the median, and to be not diversified, about 70 percent being in fact concentrated in one industry at the level 600 of the classification. Their shares of engineers and of administrators are in the same range, with a first quartile of about 2 percent and a third quartile of about 6 percent.

The variables we use from the Innovation Survey are summarized in Figures 2, 3 and 4. Figure 2 shows the distribution of the answers to the question about the share of innovative sales that comes from products launched on the market between 1986 and 1990. From this figure one can see that the distribution is skewed, with close to 17 percent of the firms deriving more than 30 percent of their sales from newly introduced products. Figures 3 and 4 show the breakdown of the answers to the two key questions about the influence of market demand conditions and technology developments, on which our demand pull and technology push indicators are based. It is clear that a larger fraction of firms overall believe that market demand is a stronger force than technological opportunity, but it is still true that over half the firms view technology as at least moderately important in propelling them to innovate. The important thing about these two charts is that they show that we do have sizable variability in our two 'exogenous' indicators.

The following two subsections of the paper present the results of estimating our model by the ALS method. We first present estimates for a simple specification (the basic model), comparing them, when possible, to the results found in the literature as recently summarized by Cohen and Klepper (1996, henceforth CK96). We shall then present estimates for an extended specification (the extended model), where we add the Innovation Survey demand pull and technology push indicators in the research and innovation equations, and the two shares of engineers and administrators variables roughly controlling for skill in the productivity equation.

---

[13] Note that one can assume that the share of innovative sales is nil, or at least belongs to the smaller interval from 0% to 10%, for those firms which are in the full sample but not in the innovation sample (i.e., since they did not report any kind of innovation). This is what we do for the estimations for the full sample presented in Appendix C.

[14] This is of course in the precise sense we give to these expressions here. The R&D doing firms are those which have reported R&D expenditures at least three years in the 1986-1990 five year period (and for which we have computed a R&D capital measure), and the patenting firms are those that have been granted European patents (one at least), over the same 1986-1990 five year period.

3.2 The basic model: a first look at the estimates

The preferred ALS estimates of the model in its basic form are presented in Table 2 for the innovation survey sample. The other more conventional econometric estimates of the basic model, for both the innovation sample and the full sample, as well as for the subsamples of firms reporting R&D expenditures and/or having one or more patents, are shown and discussed in Appendix C. The two first columns of Table 2 give the estimates of the research equations (common to both the patent and the innovative sales versions of our model), while the two middle columns give the estimates for the patent equation and the productivity equation with patents, and the two last columns give them for the innovative sales equation and the productivity equation with innovative sales. Columns 1 to 4 thus provide the estimates for the patent version of the model, and columns 1, 2, 5 and 6 the estimates for the innovative sales version.

Looking first at the decision to engage in R&D (probit column) shows that the probability of doing R&D (and reporting it) increases significantly with the firm size (number of employees), and also with their market share and diversification. The increase with size is familiar in the literature and corresponds to Cohen and Klepper (1996) stylized fact 1.[15] The finding that market share and diversification, after controlling for size and sector, are associated with a higher probability of undertaking R&D is newer, but very plausible too, agreeing with the Schumpeterian conjectures.[16] The order of magnitude of the estimates are such that a doubling of the number of employees or the equivalent number of activities or market share, for a firm at the median in our sample would augment its probability of doing R&D by about 8, 6 and 5 percent respectively.[17]

Looking then at the R&D intensity equation shows that the R&D intensity of the firms engaging in R&D does not depend on their size (i.e., the elasticity of R&D capital to size is one). This agrees similarly with CK96 stylized facts 2 and 3.[18] Note however that this result holds only in the presence of the market share variable (i.e., if we do not include it, the estimated elasticity of R&D capital to size is of about 1.3). Contrary to size, market share and diversification appear to have a quite significant and large impact on the firm R&D effort, besides that on their decision to invest in research.[19] A doubling of market share or of the equivalent number of activities for an R&D doing

[15] Cohen and Klepper (1996) 'STYLIZED FACT 1: The likelihood of a firm reporting positive R&D effort rises with firm size and approaches one for firms in the largest size ranges' (p. 928).

[16] See Crepon, Duguet and Kabla (1996) for similar results. However, they used total sales as the firm size variable, which most likely explains why they tend to find much lower (or even negative) estimates for the market share variable, and higher estimates for size.

[17] We have: $[0.40 \ln(2)]\ 0.30 = 0.08$; $[0.29 \ln(2)]\ 0.30 = 0.06$ and $[0.22 \ln(2)]\ 0.30 = 0.05$, where 0.30 is the density of the normal distribution at the median value for the right hand side of the probit (0.76).

[18] Cohen and Klepper (1996) 'STYLIZED FACT 2: Within industries, among performers of R&D, R&D rises monotonically with firm size across all firm size range...' (p. 928) and 'STYLIZED FACT 3: Among R&D reporting firms, in most industries there is no evidence of a systematic relationship between firm size and the elasticity of R&D with respect to firm size across the full range of firm sizes. Also, in most industries it has not been possible to reject the null hypothesis that R&D varies proportionately with size across the entire firm size distribution' (p. 929).

[19] By contrast to ours, most past studies did not find clear effects of either market share or diversification (Cohen and Levin, 1989).

firm in our sample would correspond to an increase of about 20 to 25 percent of its R&D capital-labor ratio.[20]

Considering next the patent equation and the innovation intensity equation, we find that, once the differences in R&D effort are taken into account, there is no significant impact of the firm size on its innovation output, be it the average number of patents (cumulated number per employee) or the percentage of innovative sales.[21] Since all three R&D intensity, patent and innovation intensity equations exhibit constant returns to scale, we can also state that the numbers of patents and (sales weighted) innovations per dollar of R&D do not differ significantly with the firm size (or the scale of their R&D programs). This result is at variance with CK96 stylized fact 4.[22] Note however that we also find decreasing returns for the patent equation when we limit the analysis to the firms with positive numbers of patents and exclude the many firms with zero patents (see Table 5 in Appendix C).[23]

Whereas firm size has no impact, that of R&D intensity is quite strong: for patents the elasticity of the firm R&D capital intensity is about 0.9 (not significantly different from unity), and it is of about 0.4 for innovative sales. A 10 percent increase of R&D intensity will thus have an impact of the same order of magnitude on the firm total number of patents and of nearly 5 percent on its innovative sales (or total number of sales weighted innovations).

Finally, the estimates of the productivity equation confirm the results we are accustomed to find in cross-sectional regressions when knowledge capital is simply proxied by an R&D capital variable (see Table 7 in Appendix C).[24] We find constant returns to scale and a physical capital elasticity of about 0.2, while the estimated elasticity of knowledge capital is 0.13 when proxied by the number of patents and is 0.10 when proxied by the share of innovative sales or the number of sales weighted innovations.[25]

3.3 The extended model: a further look at the estimates

The ALS estimates of the model in its extended version are given in Table 3, in the same format as for the basic version, but with six additional rows in the research

[20] We have: $0.36 \ln(2) = 0.25$ and $0.30 \ln(2) = 0.21$.

[21] This is true even when the market share variable is not included in the R&D equation.

[22] Cohen and Klepper (1996) 'STYLIZED FACT 4: Among R&D performing firms, the number of patents and innovations per dollar of R&D decreases with firm size and/or the level of R&D, and among all firms, smaller firms account for a disproportionately large number of patents and innovation relative to their size' (p. 930).

[23] Bound, Cummins, Griliches, Hall and Jaffe (1984), who include non-patentees in their large sample of U.S. manufacturing firms, find as we do, constant returns to size in their patents to R&D relationship.

[24] See Table 7 in Appendix C. Examples of such results for France can also be found in Cuneo and Mairesse (1984), Mairesse and Cuneo (1985), Crepon and Mairesse (1993), Hall and Mairesse (1995), Mairesse and Hall (1996). For a general survey, see Mairesse and Sassenou (1991).

[25] As explained in Appendix A, our labor, physical capital and value added variables are corrected for R&D double counting. Our estimated elasticities of knowledge capital are thus not affected by downward biases due to such double counting and should not be given an excess return interpretation, as usually done when corrections cannot be made. These biases are mainly due to the lack of correction of the labor variable (the employees working in research activities being included in the total number of employees) and tend to be much more severe for cross-sectional estimates such as ours.

and innovation equations for the weak, moderate and strong types of answers which define our demand pull and technology push indicators (relative to the no answers), and with two other rows in the productivity equation for our two skill composition variables: the shares of engineers and administrators (relative to the other categories of employees). With the exception of the elasticity of knowledge capital (proxied by patents or innovative sales) in the productivity equation, all previous estimates of the basic specification are practically unchanged by the inclusion of these new variables. Thus we shall not comment again on these estimates, but shall only discuss the new ones.

Both market demand and technological opportunities, as measured here by our two indicators based on the firm's own assessment, appear to have positive effects on the firm R&D engagement and on its intensity (in addition to the market share, diversification and size variables). These effects tend to increase across types of answers, as should be expected, from weak to moderate and from moderate to strong. However, although all twelve effects (three types of answers of the two indicators in the two research equations) are indeed positive, only five of them are statistically different from zero (at the conventional 5 percent level of confidence): one for the strong demand pull dummy in the probit equation, and the four others for the moderate and strong technology push dummies in both the probit and tobit equations.

These significant effects are very sizable, and specially so in comparison to that of the other variables. Consider for example an interquartile difference in market share (from 0.1 percent to 1.0 percent), such difference, 'other things equal', would imply a 15 percent higher probability of doing R&D for a firm at the median of our sample, and a 80 percent higher R&D capital intensity for an R&D doing firm: these effects are about the same size than the respective effects of a 'strong' technology push. A similar interquartile difference in the equivalent number of firm lines of business (from 1 to 2) corresponds to a 6 percent higher R&D probability and a 20 percent higher R&D capital intensity: these effects are roughly two third and one third of the respective effects of a 'moderate' technology push, which are themselves about two third of the effects of a 'strong' technology push.[26]

The estimated effects of the demand pull and technology push indicators on patenting are very different from their effects on R&D. The effects of the three demand pull dummies are statistically insignificant (and small) and the effects of the technology push dummies are negative, with both the 'moderate' and 'strong' effects being large and just significant at the 5 percent confidence level. Such results are a little surprising. They are, however, conditional on a given R&D intensity, since the patents equation controls for firm differences in R&D capital intensity, and thus do not mean that the total effects of demand and technology are negative or negligible. These total effects are about equal to the sum of the estimated coefficients in columns 2 and 3 of Table 3, the R&D capital elasticity estimated in the patents equation being close to one. Adding these two columns together shows that the total effects of market demand and technological opportunities on patenting are in fact similar; they are insignificant for the moderate and weak values of both indicators, but significant and large for their 'strong' value, with an order of magnitude of one more patent for a firm with two

[26]The computation of these different effects both on the R&D doing probability and R&D capital [illegible]

patents.[27]

The fact that innovative output is not always patented or all patentable, and all product patents are not exploited, is one reason for considering the innovative sales variable besides the number of patents. Indeed, in contrast to the effects found on patenting, both demand pull and technology push indicators have positive effects on innovative sales, and significantly so for the 'strong' demand pull dummy and all three technology push dummies. These direct effects are for most of them larger than the indirect ones going through increased R&D investment.[28] The direct and indirect effects of a 'strong' demand pull thus correspond to respective increases of about 5 and 1 percent of the share of innovative sales for a firm with a (nearly) median share of 10 percent.[29] The comparable figures for a 'strong' technology push are of about 4 and 3 percent.[30]

Our two skill composition variables enter very significantly the productivity equation, with a large positive coefficient of about 1.7 for both. This implies that the productivity of engineers and that of administrators are at the margin much higher, by a factor of 2.7, than the productivity of the other categories of employees; these differences, as might be expected, are actually roughly consistent with the corresponding differences in wages and labor costs. The important point is that the estimated elasticities of knowledge capital decrease by one third when we include controls for skill.[31] The estimated elasticity of the number of patents in the last five year is about 0.09 (instead of 0.13), and that of the last five year sales weighted number of innovations about 0.06 (instead of 0.10). This means, for example, that an interquartile difference in the number of patents (from 1 to 6) corresponds to a 16 percent higher productivity, or roughly one third of the interquartile range of productivity.[32] Similarly a difference

[27] These are computed as: 2[exp(0.41+0.07)-1] = 1.2 and 2[exp(0.91-0.48)-1] = 1.1.

[28] Note that the estimated R&D capital elasticity is of about 0.30 in the innovative sales equation (0.45 in the basic version) while is it is slightly above one in the patents equation (0.90 in the basic version).

[29] These are computed as 10 [exp(0.40)-1] = 4.9 and 10 {exp[(0.41)(0.30)] -1} = 1.3 respectively.

[30] These are computed as 10 [exp(0.33)-1] = 3.9 and 10 {exp[(0.91)(0.30)] -1} = 3.1 respectively.

[31] This is also what we find when we proxy knowledge capital by an R&D capital variable. See Mairesse and Cuneo (1985) and Crepon and Mairesse (1993).

The decrease in the estimated elasticity of knowledge capital reflects its positive correlation with skill. This correlation raises a delicate problem of interpretation of whether knowledge capital and skill are substitutable factors as the Cobb-Douglas specification assumes (within the limits of an approximation) or complementary factors. If the first hypothesis is approximately true, our lower estimates (when controlling for skill composition) are the appropriate ones. If the second hypothesis is more relevant and in the extreme case where knowledge capital and skill are perfect complements, our higher estimates (when not controlling for skill composition) would be the right ones. However this would mean that increases in firm research efforts and knowledge capital do not by themselves result in increased productivity, but must be accompanied by related increases in skill (not to mention other likely conditions in reality). On the basis of the cross-sectional information available to us, it does not seem reliable enough to estimate the degree of substitutability between knowledge capital and skill using a more general production function than the Cobb-Douglas. In Crepon Mairesse (1993), we estimate a translog production function but do not view it as a better approximation of the 'true average production function'. Instead we consider that including the square and cross-product terms (in the logs of the factors) is a convenient way to account for some heterogeneity in the firm production function (by allowing between- firm differences in factor elasticities and expressing simply these differences as linear functions in the logs of the factors).

[32] The interquartile range of log-productivity for the innovation sample is of 0.47=log(314/196) or

in innovative sales from the lowest share interval to the highest share (say from 10 to 70 percent) corresponds to a 13 percent higher productivity, or roughly one fourth of the productivity interquartile range.

4 Conclusion

In this paper we have tried both to summarize and confirm the main results of previous studies, which intended to assess the innovation impacts of research and the productivity impacts of innovation and research. We have introduced three new features in the analysis concerning the model, the data and econometric methods. We consider a four equations model that relates productivity to innovation output, innovation output to research, and research to its determinants. We take advantage of new data on French manufacturing firms: the number of European patents and the percentage share of innovative sales, as well as an average market share variable, a diversification index, and two indicators proxying for the influence of 'demand pull' and 'technology push' factors on research and innovation. We use appropriate econometric methods taking into account selectivity and simultaneity, as well as the statistical nature of the data: only a small proportion of firms engage in research activities and/or apply for patents; productivity, innovation and research are endogenously determined; research investment and capital are truncated variables, patents are count data and innovative sales are interval data.

We find that using the more widespread methods, and the more usual data and model specification, may lead to appreciably different estimates. We find in particular that simultaneity tends to interact with selectivity, and that both sources of biases must be taken into account together. However, our main results are consistent with many of the stylized facts of the empirical literature. The probability of engaging in research (R&D) for a firm increases with its size (number of employees), its market share and diversification, and with the demand pull and technology push indicators. The research effort (R&D capital intensity) of a firm engaged in research increases with the same variables, except for size (its research capital being strictly proportional to size). The firm innovation output, as measured by patent numbers or innovative sales, rises with its research effort and with the demand pull and technology indicators, either directly or indirectly through their effects on research. Finally, firm productivity correlates positively with an higher innovation output, even when controlling for the skill composition of labor as well as for physical capital intensity.

In order to identify and estimate our model, we have made a number of assumptions on its overall structure and the specification of the individual equations, which seem rather reasonable. We have assumed for example that the market share and diversification variables only belong to the research equations (not to the innovation and productivity equations), and are exogenous in the model.[33] The main problem of our analysis arises from cross-sectional nature of our data and estimates. We know in

47 percent (see Table 1).

[33] Likewise we have assumed that our demand pull and technology push indicators do not belong to the productivity equation and that our skill indicators only belong to this equation, all four variables being exogenous.

particular that time-series type estimates of R&D capital elasticity in the patent equation and in the productivity equation are much smaller and even not significant (see for example Crepon and Duguet 1997 for the patent equation and Cuneo and Mairesse 1984 or Hall and Mairesse 1995 for the productivity equation). It is not clear actually which type of estimates should be preferred. As a general rule cross-sectional estimates tend to be much less fragile than time-series ones, and indeed our cross-sectional estimates in this study seem quite plausible.[34] One should of course, as usual, do more investigations to go farther and deeper, and in particular one should try to gather firm panel data, which will be richer than our mainly cross-sectional data here, and to implement a dynamic model, which could give a better description of the complex relations between research, innovation and productivity.

[34] Time-series estimators control for firm correlated effects, which can take care of fixed or slowly changing omitted variables, while cross-sectional estimators may be affected by the potentially severe biases arising from such correlated effects.. Conversely, biases from measurement errors in variables or timing errors which will be negligible for cross-sectional estimators will tend to be much magnified in the time dimension of the data and to strongly affect time-series estimators. It is also more problematic to assume that a variable is exogenous or predetermined in the cross-sectional dimension of the data than in its time dimension, but difficult in the two dimensions to find instruments which will be both relevant and good. For a discussion of these issues in the context of the identification and estimation of the production function, see Mairesse (1990) and Griliches and Mairesse (1997).

References

Auzeby F. et J. P. François, 1992, L'innovation technologique dans l'industrie, *Le Quatre Pages du SESSI*, n°1.

Bound J., C. Cummins, Z. Griliches, B.H. Hall and A. Jaffe, 1984, Who Does R&D and Who Patents? In Z. Griliches ed., *R&D, Patents and Productivity*, National Bureau of Economic Research, Chicago: University of Chicago Press, pp. 21-54.

Bussy J.-C., C. Carpentier and I. Kabla, 1996, The Utilisation by French Firms of the European Patent. *Insee Studies in Economics and Statistics*, 1, pp. 12-28.

Chamberlain G., 1982, Multivariate Regression Models Applied to Panel Data, *Journal of Econometrics* 18, pp. 5-46.

Cohen W. and R. Levin, 1989, Empirical Studies of R&D and Market Structure. In R. Schmalensee and R. Willig, eds., *Handbook of Industrial Organization*, Amsterdam: North-Holland, pp.1059-1107.

Cohen W. and S. Klepper, 1996, A Reprise of Size and R&D. *The Economic Journal*, 106, pp. 925-951.

Crépon B. and J. Mairesse, 1993, Recherche et Développement, qualification et productivité des entreprises. In D. Guellec ed., *Innovation and Compétitivité*, pp. 181-222.

Crépon B. and E. Duguet, 1995, Une bibliothèque de macro-commandes pour l'économétrie des variables qualitatives et de comptage. *Document de travail CREST* 9525.

Crépon B .and J. Mairesse, 1996, Chamberlain and GMM Estimates: An Overview and Some simulation Experiments. In L. Mathias and P. Sevestre eds., Boston: Kluwer Academic Publishers, pp.321-391.

Crépon B., E. Duguet and I. Kabla, 1996, A Moderate Support to Schumpeterian Conjectures from Various Innovation Measures. In A. Kleinknecht ed., *Determinants of Innovation: the Message from New Indicators*, London: Mac Millan Press, pp.63-98.

Crépon B. and E. Duguet, 1997, 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. and E. Duguet, 1997, Estimating the Innovation Function from Patent Number: GMM on Count Panel Data. *Journal of Applied Econometrics, 12*, pp.243-263.

Cuneo P. and J. Mairesse, 1984, Productivity and Research-Development at the Firm Level in French Manufacturing. In Z. Griliches ed., *Research and Development, Patents and Productivity, Chicago:* The University Press of Chicago, pp.375-392.

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

Gouriéroux C. and C A. Monfort, 1996, *Statistics and Econometric Models* (translated by Q. Vuong), Cambridge: Cambridge University Press.

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

Gouriéroux C., A. Monfort and A.Trognon, 1985, Moindres carrés asymptotiques. *Annales de l'INSEE*, 58, pp. 91-122.

Griliches Z., and J. Mairesse, 1997, Production Functions: The Search for Identification. *Document de travail CREST* 9730. To appear in 1999, in S. Ström ed. , *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium*, Cambridge: Cambridge University Press.

Hall B. H. and J. Mairesse, 1995, Exploring the Relationship between R-D and Productivity in French Manufacturing Firms. *Journal of Econometrics*, 65(1), pp. 263-293.

Hausman J., B. H. Hall and Z. Griliches, 1984, Econometric Models for Count Data with an Application to the Patent-R&D Relationship. *Econometrica*, 52(4), pp. 909-938.

Heckman J., 1976, The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for such Models. *Annals of Economic and Social Measurement*, 5, pp. 475-492.

Heckman J., 1979, Sample Selection Bias as Specification Error. *Econometrica*, 47,pp. 153-161.

Lee L.-F., 1981, Simultaneous Equations Models with Discrete and Censored Dependent Variables. In C. Manski and D. Mc Fadden eds., *Structural Analysis of Discrete Data with Econometric Applications*, Boston: Massachusetts Institute of Technology Press, pp. 346-364.

Mairesse J. et P. Cuneo , 1985, Recherche-Développement et performances des entreprises: une étude économetrique sur données individuelles, *Revue Economique*, 36(5), pp. 1001-1041.

Mairesse J., 1990, Time -series and Cross-sectional Estimates on Panel data: Why Are They Different and Why Should They Be Equal ?. In J. Hartog, G. Ridder and J. Theeuwes, eds., *Panel Data and Labor Market Studies*, Amsterdam: North Holland, pp. 81-95.

Mairesse J., and M. Sassenou, 1991, R&D and Productivity: a Survey of Econometric Studies at the Firm Level. *Science-Technology-Industry Review*, OECD, 8 ,pp. 9-43.

Mairesse J., and B. H. Hall, 1996, Estimating the Productivity of Research and Development in French and United States Manufacturing Firms: An Exploration of Simultaneity Issues with GMM. In Wagner K., and B. van Ark, eds., *International Productivity Comparisons*, Amsterdam: North-Holland, pp.285-315.

Malinvaud E., 1970, *Statistical Methods of Econometrics,* Amsterdam: North-Holland.

Pakes A. and Z. Griliches, 1984, Patents and R&D at the Firm Level in French Manufacturing: A First Look. In Z. Griliches ed., *Research and Development, Patents and Productivity,* Chicago: The University Press of Chicago, pp.55-72.

SESSI (Service des Statistiques Industrielles), 1994, *Les chiffres clés: L'Innovation Technologique.* Paris: Dunod.

Table 1: Descriptive Statistics for the Full and Innovation Samples

Sample	Full: 6145 firms				Innovation: 4164 firms			
Statistics	Q1	Me	Q3	%	Q1	Me	Q3	%
R&D capital per employee* K_i (1990)	89.9	192.1	460.2	10.7	89.1	192.1	461.9	15.3
Number of employees L_i (1990)	39	69	183	100	43	85	242	100
Market share (%) $S_i \times 100$ (1988)	0.1	0.3	1.0	100	0.1	0.4	1.3	100
Equivalent number of activities D_i (1988)	1.2	1.6	2.0	27.4	1.2	1.6	2.0	29.8
Number of patent applications N_i (1986-1990)	1	2	5	11.6	1	2	6	16.1
Value added per employee* Q_i (1990)	184	231	322	100	196	243	314	100
Physical capital per employee* C_i (1990)	104	182	322	100	123	208	353	100
Engineers/Employment (%) $E_i \times 100$ (1990)	2.3	3.7	6.3	73.2	2.4	3.8	6.6	79.1
Administrative/Employment (%) $A_i \times 100$ (1990)	1.8	3.1	5.4	65.7	1.8	3.1	5.4	72.2

Q1 : first quartile, Me : median, Q3 : third quartile.
% : percentage of firms included in the computations.
* Thousands of 1990 FRF.

Table 2: Basic Model (Innovation sample)

Left-hand variables :
Logarithm of research capital per employee (k_i)
Number of patents per employee† (n_i)
Logarithm of innovation intensity (t_i)
Logarithm of value added per employee (q_i)
(standard errors between parentheses)

Model	R&D		Patents		Innovation	
	Probit	Tobit	n_i	q_i	t_i	q_i
R&D capital per employee k_i	–	–	0.881 (0.111)	–	0.431 (0.057)	–
Number of patents per employee n_i	–	–	–	0.130 (0.017)	–	–
Innovation intensity t_i	–	–	–	–	–	0.104 (0.016)
Market share s_i	0.221 (0.031)	0.365 (0.054)	–	–	–	–
Equivalent number of activities d_i	0.294 (0.102)	0.300 (0.134)	–	–	–	–
Number of employees l_i	0.398 (0.038)	–0.001 (0.068)	–0.021 (0.059)	–0.017 (0.012)	–0.033 (0.029)	0.011 (0.005)
Physical capital per employee c_i	–	–	–	0.206 (0.007)	–	0.212 (0.007)

Optimal asymptotic least squares, with 18 industry dummies.
Sample of 4164 manufacturing firms .
† The regression is equivalent to the number of patents per employee.

Table 3: Extended Model (Innovation sample)

Left-hand variables:
Logarithm of research capital per employee(k_i)
Number of patents per employee[†](n_i)
Logarithm of innovation intensity (t_i)
Logarithm of value added per employee (q_i)
(standard errors between parentheses)

Model	R&D		Patents		Innovation	
	Probit	Tobit	n_i	q_i	t_i	q_i
R&D capital per employee k_i	–	–	1.078 (0.166)	–	0.304 (0.064)	–
Number of patents per employee n_i	–	–	–	0.089 (0.015)	–	–
Innovation intensity t_i	–	–	–	–	–	0.065 (0.015)
Market share s_i	0.221 (0.031)	0.356 (0.053)	–	–	–	–
Equivalent number of activities d_i	0.302 (0.103)	0.333 (0.142)	–	–	–	–
Number of employees l_i	0.387 (0.040)	-0.043 (0.066)	-0.066 (0.073)	-0.014 (0.007)	-0.002 (0.028)	0.007 (0.004)
Physical capital per employee c_i	–	–	–	0.194 (0.007)	–	0.198 (0.007)
Engineers/Personnel E_i	–	–	–	1.614 (0.123)	–	1.649 (0.123)
Administrative/Personnel F_i	–	–	–	1.744 (0.143)	–	1.765 (0.142)
Demand pull :						
– weak ($\delta_{1,i}$)	0.261 (0.192)	0.377 (0.308)	-0.071 (0.332)	–	-0.040 (0.150)	–
– moderate ($\delta_{2,i}$)	0.191 (0.170)	0.342 (0.277)	-0.068 (0.315)	–	0.164 (0.137)	–
– strong ($\delta_{3,i}$)	0.343 (0.164)	0.412 (0.272)	0.074 (0.304)	–	0.399 (0.133)	–
Technology push :						
– weak ($\tau_{1,i}$)	0.173 (0.120)	0.173 (0.207)	-0.237 (0.254)	–	0.229 (0.089)	–
– moderate ($\tau_{2,i}$)	0.322 (0.113)	0.604 (0.198)	-0.540 (0.255)	–	0.268 (0.092)	–
– strong ($\tau_{3,i}$)	0.444 (0.117)	0.907 (0.202)	-0.483 (0.277)	–	0.333 (0.105)	–

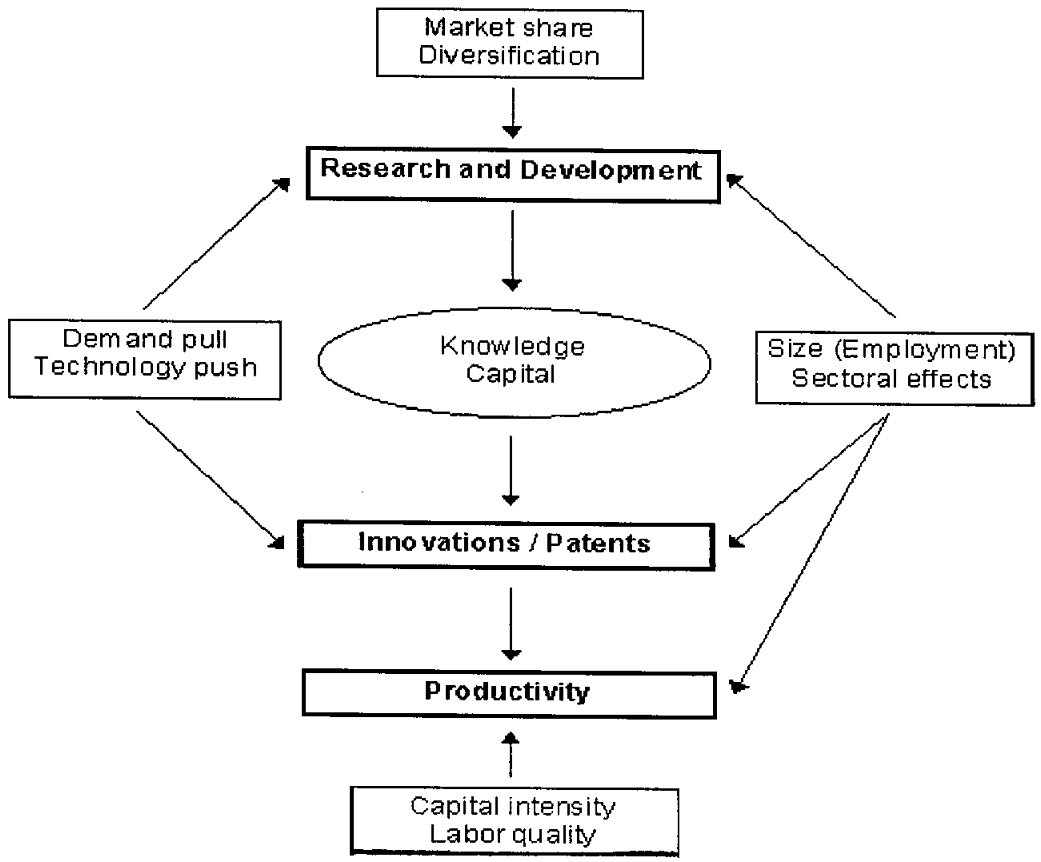

Figure 1: Diagram of the Model

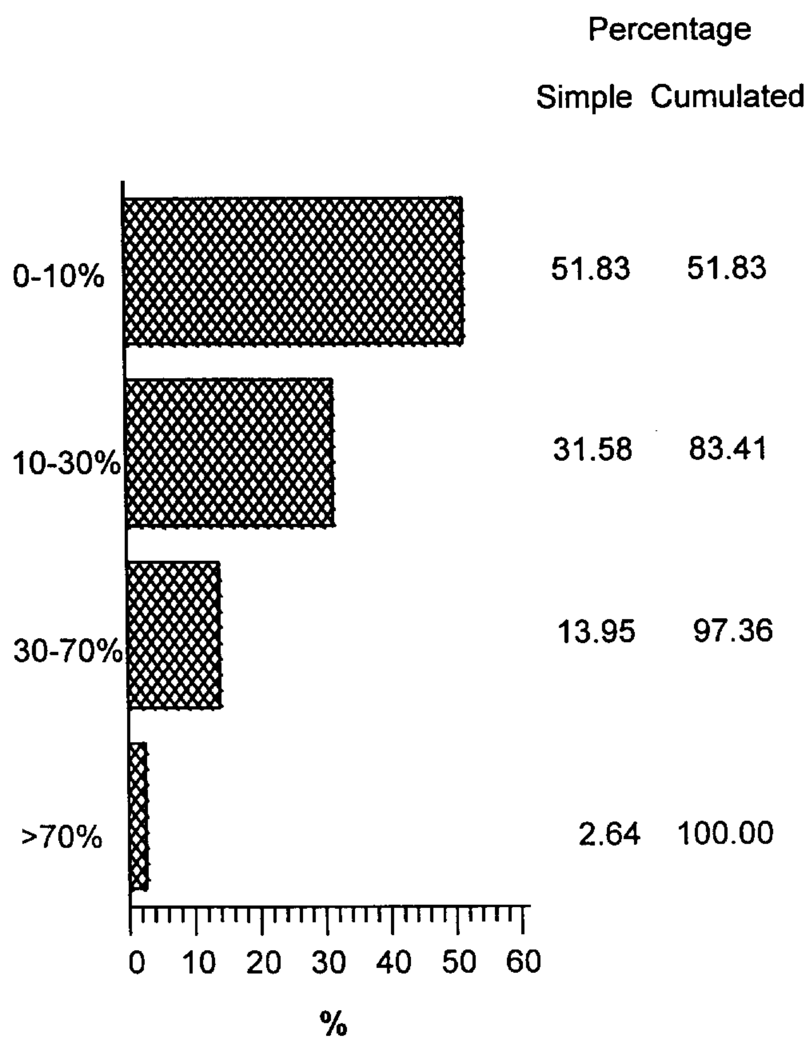

Figure 2: Share of Innovative Sales (Innovation Sample)

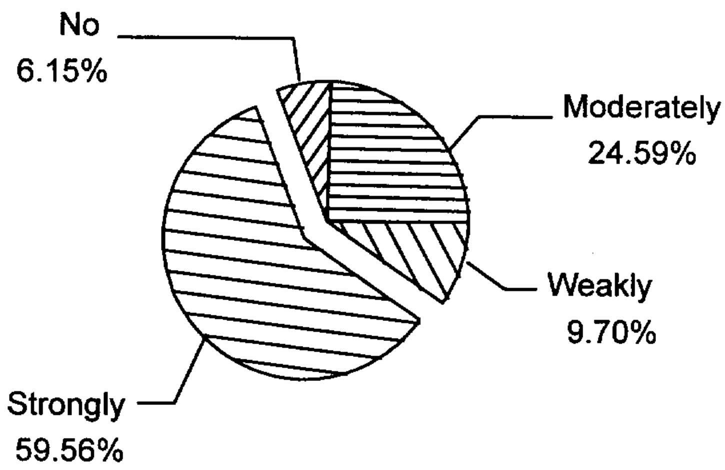

Figure 3: Demand Pull Indicator (Innovation Sample)

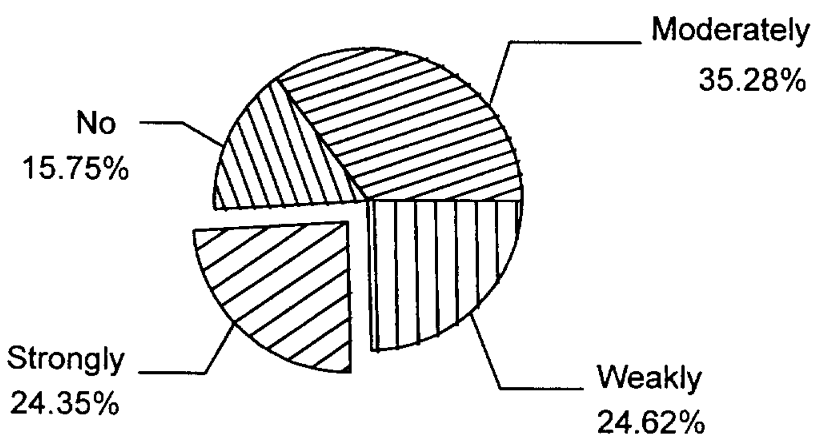

Figure 4: Technology Push Indicator (Innovation Sample)

A APPENDIX: Data construction

We have constructed our data set by merging various sources of information, using the firm national identification code SIREN (for 'Système Informatique de Répertoire des ENtreprises').

A.1 Accounting and employment data

The information on the firm current accounts and balance sheets, and on the number of employees, comes primarily from the SUSE files ('Système Unifié de Statistiques d'Entreprises'). From them, we mainly extracted the firm value added, its fixed assets gross bookvalue (at the beginning of year) and its total number of employees (average over the year), and we simply computed our labor, productivity and physical capital intensity variables as l_i = number of employees, q_i = value added per employee, and c_i = physical capital per employee. Physical capital is the fixed assets gross bookvalue, approximately adjusted for inflation on the basis of an estimated average age of fixed assets derived from the net to gross book value ratio (see for example Crepon and Mairesse, 1993). These variables are taken in logarithms in our model. Table 1 in the text shows that the 'median firm' in the innovation sample has 85 employees, a value added per employee of 243 thousands 1990 FRF (FRench Franc), and a physical capital per employee of 208 thousands 1990 FRF (a current approximation being 1 US Dollar = 5 FRF). These numbers are 5 to 15 percent lower for the median firm in the full sample.

The information on the distribution of employees by occupational categories is taken from the 'Enquête sur la Structure des Emplois' (ESE). Based on previous work and some experiments, we only kept here two main indicators of the importance of management in the firm: the shares E_i and A_i of engineers and administrators (administrative executives) in total employment (see Crepon and Mairesse, 1993). Table 1 in the text shows that the median shares of engineers and administrators are about 3 to 4 percent for the 65 to 80 percent of the firms which actually have engineers and/or administrators among their personnel.

The average firm market share variable and diversification index are computed from the detailed information provided by the 'Enquête Annuelle d'Entreprises' (EAE). This is the main annual firm survey in France, which among other information gives the decomposition of the firm sales in its different lines of business, or industry segments at the 'level 600' of the French industrial classification (NAP 600 in the 'Nomenclature des Activités et Produits'). We have in fact only considered the decomposition of domestic sales for the 227 different 'level 600' industry segments pertaining to manufacturing.[35]

Let $S_{i,k}$ be the sales of firm i for its product k in the industry segment or market k , $S_i = \sum\limits_k S_{i,k}$ and $S_k = \sum\limits_i S_{i,k}$ are respectively the overall sales of firm i (over all its

[35] The decomposition for non manufacturing sales is also available, but it seems of a lesser quality, and we did not use it.

products) and the overall sales on market k (over all firms).[36] The market share $s_{i,k}$ of firm i on market k and the share of product k in firm i total sales are thus equal to:

$$s_{i,k} = S_{i,k}/S_k \quad and \quad b_{i,k} = S_{i,k}/S_i$$

Note that we have $\sum_k b_{i,k} = 1$ for each firm i , but that we have not $s_k = \sum_i s_{i,k}$ for each market k , since we thought indeed more appropriate to compute the firm product market share $s_{i,k}$ relative to the total of sales in this market S_k for all the firm available in the EAE data files, and not only for those firms that we could keep in our samples, after merging all the data sources we used.

Then for each diversified firm i we defined the weighted average market share s_i^w and the diversification index d_i as:

$$s_i^w = \sum_k b_{i,k} \times s_{i,k} \quad and \quad \frac{1}{d_i} = h_i = \sum_k b_{i,k}^2$$

with d_i being the inverse of the Herfindahl concentration index h_i of the firm sales. For a non diversified firm (i.e., with only one k), we of course have $s_i^w = s_i$ and $d_i = h_i = 1$. For a diversified firm with n product lines of equal weights (i.e., $b_{i,k} = 1/n$ with $k = 1, ..., n$), we have: $s_i^w = s_{i,k}$ for all k ($= 1, ..., n$), and $d_i = n$ (or $h_i = 1/n$). The diversification index d_i for the firm i can be interpreted as the equivalent number of product lines with *equal* sales, such as its sales Herfindahl concentration index would the same as the one observed h_i. The higher the d_i, the higher the diversification of the firm (and the lower the h_i and the concentration of its sales).

In the model we introduce both variables s_i^w and d_i in logarithms. We can see in Table 1 that only about 30 percent of the firms are diversified at the industry level 600, and only 25 percent of these have an equivalent number of product lines of 2 or more (see third quartile Q_3). We can also see that most of firms have very small market shares, only 25 percent having a market share higher than 1 percent (see Q_3).

Note finally that we include in all the equations of the model a full set of 18 industry dummies (or 17 plus the constant) : $S_i^1, S_i^2, \ldots S_i^{18}$ with $\sum_{j=1}^{18} S_i^j = 1$. These are defined at a higher level of classification ('level 40' of the NAP) than the average market share and diversification variables, on the basis of the firm main industrial activity (with the highest market share).

A.2 Research and Development data

The Research and Development (R&D) data come from the annual firm research expenditures survey conducted by the 'Ministère de la Recherche' since 1974. For this study, we kept mainly the information on the firm total R&D which we 'convert' in 1990 prices using an overall manufacturing R&D deflator.[37] From the firm total deflated

[36] More precisely, the decomposition of sales is given for the firm total and export sales, and hence by difference for the firm domestic sales, and we thought preferable to use the decomposition of total sales to construct the diversification index, and that of domestic sales to construct the average market share variable. Hence, in fact $S_{i,k}$ denotes here the total sales of firm i for its product k in the computation of d_i , and the domestic sales of firm i for its product k in the computation of s_i^w.

[37] We computed this R&D deflator based on the decomposition of the overall Manufacturing R&D spending (for all the Manufacturing firms in the survey), in its labor, materials and physical investment components, and on the corresponding wage and prices indices.

R&D expenditures series, R_{it} for firm i in year t , we simply compute the R&D capital K_{it} (at the beginning of year t), by the so-called 'permanent inventory method', i.e., by the following iterative formula:

$$K_{it} = (1 - \delta)\, K_{i(t-1)} + R_{i(t-1)}$$

which assumes a geometric depreciation of R&D capital at the constant rate δ . Consistently with previous studies on French manufacturing (see for example Cuneo and Mairesse, 1984, or Crepon and Duguet, 1997), we adopt a depreciation rate δ of 15%. For the firms which have regularly invested in R&D since 1974, we take $K_{i(74)} = 5 \times R_{i(74)}$ as starting value. This assumes an R&D constant pre-sample growth rate γ of 5% (i.e., $K_{i(74)} = R_{i(74)}/(\gamma + \delta)$). For the firms which have engaged in R&D in a latter year τ, the starting value at the beginning of this year is zero ($K_{i\tau} = 0$). Note however that we consider here that a firm has engaged consistently in R&D activities if it has reported R&D expenditures for a least three years in the period 1986-1990, which is the relevant period for our innovative sales variable and our demand pull and technology push indicators (see below); we have only computed the firm R&D capital in this case.[38] Actually the choice of starting value and that even of the depreciation rate in the computation of R&D capital has little consequence on the estimated elasticities of interest in our model, for such cross-sectional estimates as ours. It affects however the absolute value of R&D capital (such as the sample median or average) and hence the estimated impact of a given difference in R&D investment or capital (in dollar terms).

We have also used the information on research employees, and on the physical investment and materials components of total R&D expenditures, also available in the annual survey, to correct our labor, physical capital and value added variables for R&D double counting. The number of employees working in research activities is deducted from the total number of employees; the gross book value of fixed assets used in research activities (estimated on the basis of the physical investment component) is similarly subtracted from the total gross book value of fixed assets. The accounting value added measure is also corrected by adding back R&D expenditures (materials) which are expensed out in the firm standard current accounts. As discussed in particular in Cuneo and Mairesse (1984) and Hall and Mairesse (1995), the estimated elasticities of knowledge capital (either patents, innovative sales or R&D capital) in the productivity equation are thus not affected by R&D double counting biases, and thus should not be given an excess return interpretation, as usually done when the corrections cannot be made.

To summarize, g_i is a research dummy equal to 1 if firm i has reported R&D for at least three years in the period 1986-1990, 0 otherwise; and k_i is the log of R&D capital at the beginning of year 1990 for an R&D doing firm i (if $g_i = 1$). We see in Table 1 that relatively few firms are consistently engaged in R&D, even in the innovation sample. Only about 15% of the firms in this sample are R&D doing firms with a median R&D capital of about 190 thousand '1990 FRF'.

[38] Note also that for the few R&D doing firms which had a one or two years of interruption in reporting to the research survey, we simply interpolate their R&D expenditures for these years.

A.3 Patent data

The patent numbers come from the European PATent (EPAT) data base. Since the firm ID codes SIREN were not available in this data base, it has been necessary to carefully match SIREN and firm names.[39] We have taken as our patent variable n_i the total numbers of patents applied by the firm i during the five year period 1986-1990. We have considered the number of patents applied rather than the number of patents granted, which is often viewed as a more appropriate measure of innovation output. However these two measures should be practically equivalent for our cross-sectional type estimates.

As in the case of R&D, Table 1 shows that relatively few firms have applied for an European patents in five years, even in the innovation sample. About 16% of the firms in this sample have applied for at least a patent in the 1986-1990 period, with a median number of patents of two and a third quartile of six. Also only about half of them (50%) are R&D performers, and conversely half of the R&D firms have applied for at least a patent in five years. For similar numbers in U.S. manufacturing, see Bound et al. (1984).

A.4 Innovation data

The information on the share of innovative sales and on the demand conditions and technological opportunities is provided by the 1990 Innovation survey performed by SESSI (Service des Statistiques Industrielles).[40] Actually this information is based on questions that are only asked in the survey to the 'innovating firms', which make up our innovation sample. The innovating firms are those which answered 'yes' to at least one of the following eight questions: did you performed in the five last years (between January 1st 1986 and December 31 1990) an innovation of the following type: (i) product improvement; (ii) new product for the market; (iii) product imitation (i.e., new for the firm but not for the market); (iv) technological breakthrough; (v) process improvement; (vi) packaging innovation (explicitly excluded from i, ii and iii in the questionnaire); (vii) organizational innovation linked to the introduction of technological change and (viii) marketing innovation. About 60% of the French manufacturing firms have innovated according to this definition (a little more in our own sample which consists of slightly larger firms in average).

The Innovation survey which was conducted in 1991 for all manufacturing firms with more than 20 employees, achieved an excellent coverage (all firms of more than 500 employees and more than 80% of the sales of the smaller firms). Our own full sample (N=6145) corresponds to roughly one third of the firms actually surveyed; this is reasonable considering we had to match several different data sources.[41] Detailed

[39] This work has been performed at INSEE by J.-C. Bussy, C. Carpentier, P. Corbel and I. Kabla (1996), with the collaboration of INPI (Institut National de la Propriété Industrielle). See *INSEE studies in Economics and Statistics no1.* (1996), which also present a number of studies using the matched EPAT data set.

[40] For a presentation of the survey and its questionnaire, see François (1991).

[41] We also had to do a very modest amount of cleaning to eliminate of few 'outliers', which is necessary when analyzing firm level micro data (see Hall and Mairesse, 1995, for some details).

information on the 1990 Innovation survey can be found in François (1991) and Auzeby and François (1992), and studies using it in SESSI (1994).

The share of innovative sales is based on a question asking innovating firms what percentage of their 1990 sales percentage of sales is imputable to new products launched between 1986 and 1990. Firms answered on an interval scale: 0-10%, 10-30%, 30-70% and more than 70%. We can see from Figure 2 that more than the half of innovating firms have less than 10% of innovative sales, while less than 3% have more than 70% of innovative sales.

The firm level demand pull and technology push indicators are respectively based on the two following questions: 'Do you consider that in your firm innovation is determined through the impetus given by the market (relationships with customers, competitors)?' and ' Do you consider that in your firm innovation is determined by technology specific dynamics?'. Firms answered on a four point scale: 'no', 'weakly', 'moderately' and 'strongly'. Figures 3 and 4 show that the market demand plays a moderate or strong role for about 85% of innovating firms while technology plays such a role for about 60% of them.

B APPENDIX: Method of estimation

The key idea of the method of estimation is to estimate reduced form coefficients in each of the equation of the model separately, and infer from these auxiliary parameters the structural form parameters of the model using a minimum distance estimator or so called Asymptotic Least Squares method, henceforth ALS (see Malinvaud 1970, Gourieroux, Monfort and Trognon 1985, Gourieroux and Monfort 1996, chapter 9). The intuition is the same as that of indirect least squares, although ALS is more general. We have thus two estimation problems: first, estimating the reduced form parameters and their joint covariance matrix; then, estimating the structural form parameters in a consistent (and efficient) way. The first problem is solved by interpreting the maximum likelihood and pseudo maximum likelihood estimators of the reduced form equations as specific M-estimators (i.e., Maximization-estimators, see Gourieroux and Monfort 1996, chapter 8). The second problem comes down to writing explicitly the relationships between the structural and reduced form parameters (or the parameters of interest and auxiliary parameters).

B.1 Reduced form estimation

Once we have written the reduced form of the model, the estimation problem becomes that of a series of single equation, and we can apply the estimation method which is most appropriate for each equation, i.e., a maximum likelihood or pseudo maximum likelihood method. We can thus write:

$$\widehat{\gamma}_k \in \arg\max_{\gamma_k} L_k(\gamma_k), \quad k = 1, 2, 3$$

where γ_k is the reduced form parameter in equation k and L_k denotes the likelihood or pseudo likelihood function. The problem remains to estimate the joint covariance matrix of the $\widehat{\gamma}_k$'s. We can solve it by considering that our estimators can also be defined globally by:

$$\widehat{\gamma} \in \arg\max_{\gamma} L(\gamma) \text{ with } \gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix} \text{ and } L(\gamma) = \sum_{k=1}^{3} L_k(\gamma_k)$$

Maximizing L according to γ gives exactly the maximum likelihood and pseudo maximum likelihood estimators. This property arises from the separability of $L(\gamma)$ with respect to the $\widehat{\gamma}_k$'s.[42] These estimators can be interpreted as M-estimators. Under the usual regularity conditions for M-estimators:

$$\sqrt{N}(\widehat{\gamma} - \gamma) \underset{N \to +\infty}{\xrightarrow{d}} N(0, \Omega)$$

with asymptotic covariance matrix:

$$\Omega = J^{-1} I J^{-1}$$

[42] If such separability does not hold, we can always assume that it does by reparametrizing in the first step of the estimation (changing the names of the reduced form parameters in each equation so that they be different), and imposing the cross-equations constraints only in the second step. The method is thus quite general.

and

$$I = \underset{Z\,0}{\mathrm{EE}}\left[\frac{\partial L}{\partial \gamma}(\gamma_0)\frac{\partial L}{\partial \gamma'}(\gamma_0)\right], \quad J = \underset{Z\,0}{\mathrm{EE}}\left[-\frac{\partial^2 L}{\partial \gamma \partial \gamma'}(\gamma_0)\right]$$

where γ_0 denotes the true value of parameter γ , and the expectations are taken with respect to the distributions of the exogenous variables (index Z) and of the endogenous variables (index 0).

The I matrix and the J matrix can be estimated by their sample counterparts:

$$\widehat{I} = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial L_i}{\partial \gamma}(\widehat{\gamma})\frac{\partial L_i}{\partial \gamma'}(\widehat{\gamma}), \quad \widehat{J} = -\frac{1}{N}\sum_{i=1}^{N}\frac{\partial^2 L_i}{\partial \gamma \partial \gamma'}(\widehat{\gamma})$$

where N is the sample size. Note that since the objective function is separable, the off-diagonal terms (the cross derivatives) in $\widehat{J}$ (and J) are equal to zero.

Finally, we can use as the estimator of the covariance matrix of the estimated reduced form parameters:

$$\widehat{\Omega} = \widehat{J}^{-1}\widehat{I}\widehat{J}^{-1}$$

Through $\widehat{\Omega}$ (and $\widehat{I}$), the correlations between the perturbations of the different equations of the model are taken into account, without making specific distributional assumptions on their *joint* distribution.

B.2 Actual Implementation

B.2.1 The research equations

The research equations are already reduced form equations that we directly estimate by maximum likelihood.[43] If we denote the reduced form first order parameters: $\pi_0 = b_0$ and $\pi_1 = b_1$, we have $\gamma_1 = (\pi_0, \pi_1, \sigma_1, \rho)'$, with σ_0 normalized to 1.

B.2.2 The patent equation

The reduced form of the patent equation is given by:

$$\mathrm{E}\left(n_i | k_i^*, x_{2i}, u_{2i}; \alpha_k, b_2\right) = \exp\left\{x_{1i}\pi_1\alpha_K + x_{2i}b_2 + \alpha_K u_{1i} + u_{2i}\right\}$$

which can also be written as:

$$\mathrm{E}\left(n_i | k_i^*, z_{2i}, v_{2i}; \gamma_2\right) = \exp\left\{z_{2i}\pi_2 + v_{2i}\right\}$$

where z_{2i} is the vector of exogenous variables, grouping x_{1i} and x_{2i} variables without replication so that z_{2i} is full column rank (of dimension r_2), π_2 is the corresponding reduced form parameter vector, and $v_{2i} = \alpha_K u_{1i} + u_{2i}$ the reduced form error term. This is a count data model with heterogeneity term $\exp(v_{2i})$. Note that even if the structural equation is not heterogeneous, the reduced form equation will be (which is a general argument in favor of heterogeneous count data models). We do not need

[43] This is done with a SAS-IML program (%TOBITGEN),based on a Newton-Raphson algorithm with Levenberg-Marquardt modifications and analytical second order derivatives (see Crépon- Duguet 1995, and Crépon- Duguet-Kabla, 1996).

any specific distributional assumption on u_{2i} to estimate this model, only that the expectation of $\exp(v_{2i})$ is constant (overall or by industry) and its variance finite, and can apply the pseudo maximum likelihood method (see Gourieroux, Monfort and Trognon 1984). Based on our previous experience, we used a pseudo log likelihood function based on the negative binomial pseudo distribution.[44]

B.2.3 The innovative sales equation

The latent reduced form of the innovative sales equation is given by:

$$t_i^* = x_{1i}\pi_1\alpha_K + x_{2i}b_2 + \alpha_K u_{1i} + u_{2i}$$

which we can write as:

$$t_i^* = z_{2i}\pi_2 + v_{2i}$$

Since u_1 and u_2 are both assumed to be normal $v_{2i} = \alpha_K u_{1i} + u_{2i}$ is also normal and we can estimate this equation as a ordered probit model. Since our observed indicator corresponds to known intervals of the underlying share of innovative sales t_i^*, $\gamma_2 = (\pi_2', \sigma_2)'$ is fully identifiable.

B.2.4 The productivity equation

In the first version of the model with patents, the reduced form of the productivity equation can be written as:

$$q_i = \alpha_I \ln \mathrm{E}\left(n_i | k_i^*, z_{2i}, v_{2i}; \pi_2\right) + x_{3i}b_3 + u_{3i}$$

and in the second version with innovative sales as:

$$q_i = \alpha_I t_i^* + x_{3i}b_3 + u_{3i}$$

Both educed form equations can thus written as:

$$q_i = z_{2i}\pi_2\alpha_I + x_{3i}b_3 + v_{2i}\alpha_I + u_{3i}$$

or:

$$q_i = z_{3i}\pi_3 + v_{3i}$$

We simply estimate these equations by ordinary least squares (OLS), with a robust covariance matrix.[45]

---

[44] The negative binomial pseudo log likelihood is concave with respect to the parameter π_2 so that the maximum is unique. We carried on the estimation using a SAS-IML program (%PMVBNEG), based on a Newton-Raphson algorithm with analytical second order derivatives (see Crepon and Duguet 1995 and 1997).

[45] The OLS estimator can be viewed as a pseudo-maximum likelihood estimator with a normal pseudo-distribution.

B.3 Structural form estimation

B.3.1 The relation between the reduced form and structural form parameters

The patent equation can be written under the two equivalent forms:

$$\mathrm{E}\left(n_i|k_i^*, x_{2i}, u_{2i}; \alpha_k, b_2\right) = \exp\left\{x_{1i}\pi_1\alpha_K + x_{2i}b_2 + \alpha_K u_{1i} + u_{2i}\right\} = \exp\left\{z_{2i}\pi_2 + v_{2i}\right\}$$

$\forall i = 1, \ldots, N$ where $v_{2i} = \alpha_K u_{1i} + u_{2i}$.

Taking the logarithms of both expressions gives:

$$z_2\pi_2 = x_1\pi_1\alpha_K + x_2 b_2$$

where

$$\underset{(N,r_1)}{x_1} = \begin{pmatrix} x_{11} \\ \vdots \\ x_{1N} \end{pmatrix}, \quad \underset{(N,k_2)}{x_2} = \begin{pmatrix} x_{21} \\ \vdots \\ x_{2N} \end{pmatrix}, \quad \underset{(N,r_2)}{z_2} = \begin{pmatrix} z_{21} \\ \vdots \\ z_{2N} \end{pmatrix}$$

We can define the two exclusion matrices J_{21} and J_{22} (made of zeroes and ones) simply indicating if the variables in z_2 are in x_1 or x_2:

$$\begin{array}{lll} \underset{(r_2,r_1)}{J_{21}} & \text{such as} & z_2 J_{21} = x_1 \\ \underset{(r_2,k_2)}{J_{22}} & \text{such as} & z_2 J_{22} = x_2 \end{array}$$

and write:

$$z_2\pi_2 = z_2\left(J_{21}\pi_1\alpha_K + J_{22}b_2\right)$$

Since z_2 is full column rank, this implies:

$$\pi_2 = J_{21}\pi_1\alpha_K + J_{22}b_2 = H_2\beta_2 \tag{1}$$

with

$$H_2 = \left(J_{21}\pi_1 | J_{22}\right), \quad \beta_2 = \begin{pmatrix} \alpha_K \\ b_2 \end{pmatrix}$$

The same relation applies for the innovative sales equation, since:

$$t_i^* = x_{1i}\pi_1\alpha_K + x_{2i}b_2 + \alpha_K u_{1i} + u_{2i} = z_{2i}\pi_2 + v_{2i}$$

The productivity equation can now be written as:

$$q_i = z_{2i}\pi_2\alpha_I + x_{3i}b_3 + v_{2i}\alpha_I + u_{3i} = z_{3i}\pi_3 + v_{3i}$$

or:

$$\pi_3 = J_{32}\pi_2\alpha_I + J_{33}b_3 = H_3\beta_3$$

with:

$$H_3 = \left(J_{32}\pi_2 | J_{33}\right), \quad \beta_3 = \begin{pmatrix} \alpha_I \\ b_3 \end{pmatrix}, \quad z_3 \underset{(r_3,r_2)}{J_{32}} = z_2, \quad z_3 \underset{(r_3,k_3)}{J_{33}} = x_3$$

B.3.2 Asymptotic least squares

The estimates of the reduced form parameters π being consistent and asymptotically normal, we can apply ALS to estimate our structural parameters β ,where:

$$\beta = \begin{pmatrix} \beta_2 \\ \beta_3 \end{pmatrix} \quad \text{and} \quad \pi = \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{pmatrix}$$

The true values of the β and π are assumed to be such that we have the exact relation: $g(\beta, \pi) = 0$.

In the traditional ALS terminology the structural parameters β are also called the interest parameters and the reduced form parameters π the auxiliary parameters. The basic idea of ALS is to use the estimate $\widehat{\pi}$ of π obtained in a first step and compute in a second step an estimate $\widehat{\beta}$ of β such that $g\left(\widehat{\beta}, \widehat{\pi}\right)$ is as 'close to zero' as possible. That is, we solve the program:

$$\widehat{\beta} = \arg\min_{\beta} g(\beta, \widehat{\pi})' \, \Psi^{-1} g(\beta, \widehat{\pi})$$

where Ψ is a metric.Whatever Ψ , $\widehat{\beta}$ is consistent if $\widehat{\pi}$ is consistent, and it is asymptotically efficient for Ψ given by:

$$\Psi^* = \mathrm{V}_{as}\left[g(\beta, \widehat{\pi})\right] = \frac{\partial g}{\partial \pi'}(\beta, \widehat{\pi}) \, \Omega \frac{\partial g'}{\partial \pi}(\beta, \widehat{\pi})$$

where $\Omega = \mathrm{V}_{as}(\widehat{\pi})$ is the covariance matrix of $\widehat{\pi}$. Since we need an estimate of β to estimate the optimal metric Ψ^*, we can estimate β in two steps.

In the first step, we can use for Ψ the Euclidean metric ($\Psi = Id$), and estimate β as:

$$\widehat{\beta} = \arg\min_{\beta} g(\beta, \widehat{\pi})' \, g(\beta, \widehat{\pi})$$

In our case, this is simply OLS applied to:

$$\begin{pmatrix} \widehat{\pi}_2 \\ \widehat{\pi}_3 \end{pmatrix} = \begin{pmatrix} \widehat{H}_2 & 0 \\ 0 & \widehat{H}_3 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + w = \widehat{H}\beta + w \tag{4}$$

with $\widehat{H}_2 = (J_{21}\widehat{\pi}_1 | J_{22})$, $\widehat{H}_3 = (J_{32}\widehat{\pi}_2 | J_{33})$, $\widehat{H} = \mathrm{diag}\left(\widehat{H}_k\right)$and residual $w = g(\beta, \widehat{\pi})$. We thus have:

$$\widehat{\beta} = \left(\widehat{H}'\widehat{H}\right)^{-1} \widehat{H}' \begin{pmatrix} \widehat{\pi}_2 \\ \widehat{\pi}_3 \end{pmatrix}$$

and the asymptotic distribution of $\widehat{\beta}$ is:

$$\sqrt{N}\left(\widehat{\beta} - \beta\right) \underset{N \to +\infty}{\xrightarrow{d}} N(0, \Sigma)$$

with:

$$\Sigma = \left(\frac{\partial g'}{\partial \beta}\frac{\partial g}{\partial \beta'}\right)^{-1} \frac{\partial g'}{\partial \beta}\frac{\partial g}{\partial \pi'} \Omega \frac{\partial g'}{\partial \pi}\frac{\partial g}{\partial \beta'} \left(\frac{\partial g'}{\partial \beta}\frac{\partial g}{\partial \beta'}\right)^{-1}$$

Replacing β, π and Ω by their consistent estimates $\widehat{\beta}$, $\widehat{\pi}$ and $\widehat{\Omega}$,we are now able to compute an estimated $\widehat{\Psi}^*$ such that:

$$\widehat{\Psi}^* = \frac{\partial g}{\partial \pi'}\left(\widehat{\beta},\widehat{\pi}\right)\widehat{\Omega}\frac{\partial g'}{\partial \pi}\left(\widehat{\beta},\widehat{\pi}\right)$$

with

$$\frac{\partial g}{\partial \pi'}\left(\beta,\pi\right) = \begin{pmatrix} -J_{21}\alpha_K & \mathrm{Id}_{r_2} & 0_{r_2,r_3} \\ 0_{r_3,r_1} & -J_{32}\alpha_I & \mathrm{Id}_{r_3} \end{pmatrix}$$

In the second step, we can then compute the corresponding optimal ALS estimator β^* such that:

$$\beta^* = \arg\min_{\beta} g\left(\beta,\widehat{\pi}\right)' \widehat{\Psi}^{*-1} g\left(\beta,\widehat{\pi}\right)$$

which in our case reduces to GLS applied to equation 2. We thus have:

$$\beta^* = \left(\widehat{H}'\widehat{\Psi}^{*-1}\widehat{H}\right)^{-1}\widehat{H}'\widehat{\Psi}^{*-1}\begin{pmatrix} \widehat{\pi}_2 \\ \widehat{\pi}_3 \end{pmatrix}$$

and the asymptotic distribution of β^* is:

$$\sqrt{N}\left(\beta^* - \beta\right) \underset{N\to+\infty}{\xrightarrow{d}} N\left(0,\Sigma^*\right)$$

where Σ^* can be estimated by:

$$\widehat{\Sigma}^* = \left[\frac{\partial g'}{\partial \beta}\left(\widehat{\beta},\widehat{\pi}\right)\widehat{\Psi}^{*-1}\frac{\partial g}{\partial \beta'}\left(\widehat{\beta},\widehat{\pi}\right)\right]^{-1}.$$

C APPENDIX: An assessment of biases likely to arise in innovation and productivity studies

We have tried in this study to correct for selectivity and simultaneity biases and to take into account the specific features of the data, by using appropriate econometric methods. However, it is of interest to assess the magnitude of these biases and the effects of neglecting to consider the nature of the data, when we use more widespread methods. The following subsections thus present a series of estimates, which we obtained using ordinary least squares(OLS), two stages least squares (2SLS), maximum likelihood (ML) or pseudo maximum likelihood (PML), and first and second steps asymptotic least squares (ALS) estimators. We only present these estimates for the basic specification of the equations in our model, but for both the full and innovation samples, and when appropriate for the subsamples of R&D doing firms or patenting firms or both. Depending on the equation and samples or subsamples, these estimators fail to take into account selectivity, and/or simultaneity, and/or the specific nature of R&D, patents or innovative sales variables.

C.1 The research equations

Table 4 compares the OLS estimates of the research intensity equation for the subsamples restricted to the R&D doing firms (with positive R&D expenditures) to the corresponding ML estimates for our generalized tobit specification (the tobit part) for the full and innovation samples (with both R&D and non R&D doing firms). While the ML and OLS estimates are about the same for both samples or subsamples (for the right and left panels of the Table), they do significantly differ between themselves. If we had considered only the evidence conveyed by OLS for the R&D firms we would have concluded that the firm R&D intensity decreases significantly with its size and is independent of diversification. Actually we found quite the opposite: firm R&D intensity does not depend on size and increases significantly with diversification.

C.2 The patent equation

Table 5 compares the different types of estimates of the patent equation we can consider on the complete samples (full and innovation samples), positive R&D subsamples, positive patents subsamples and both positive R&D and patents subsamples. The PML estimators (assuming a negative binomial pseudo distribution of the disturbances) can be computed in all cases, while the OLS and 2SLS estimators can be only for the positive patents subsamples (i.e., of firms with at least one patent). Note that we include an R&D dummy in the equation when we consider these estimators for a sample or subsample with non R&D firms. We also present both the first step and second steps ALS estimates for the complete samples. Note that the second step ALS estimates are estimated independently of the productivity equation and thus can differ from our preferred ones. Comparing them in Table 2 and Table 5 for the innovation sample, we can verify that they are in fact quite close.

As in the case of the research intensity equation we see that the same estimates vary very little across the full and innovation samples or subsamples. By contrast the different types of estimates can differ significantly between themselves. Both our findings that patents do not depend on size, but are strongly related to R&D intensity with an elasticity of about one, are at variance with the other estimates (other than ALS). The OLS and PML estimates indicate a much lower R&D elasticity of about 0.4 in all cases and a negative dependence on size for all the subsamples with positive patents, while the 2SLS are too imprecise to let us say much about them. Ignoring both selectivity (the fact that most of the firms do not patent) and simultaneity thus result in large downward biases. The fact that OLS, 2SLS and PML are about the same for the positive patents subsamples tend to show that selectivity is a more severe problem.[46] The fact the PML and ALS estimates differ for the complete samples (coupled with the inaccuracy of the 2SLS estimates for the subsamples) seem also to indicate that simultaneity (the endogeneity of R&D) interacts with selectivity. Simultaneity matters and can be dealt with, only if selectivity is taken into account.

C.3 The innovation sales equations

Table 6 shows the different types of estimates of the innovative sales equation that we have been able to compute on the complete samples and positive R&D subsamples. Since only the 'innovating' firms had to declare their share of innovative sales (see Appendix A, subsection A.4), in order to compute these estimates for the full sample and the corresponding positive R&D subsample (in the left panel of the Table), we assumed that in fact the percentage share of innovative sales of the 'non innovating' firms lies in the first interval from 0 to 10%. Note also that in order to compute the OLS and 2SLS estimates, we have simply assumed that the value of the innovative sales variable was equal to the interval center, that is 5%, 20%, 50% and 85%. Despite such invention of the missing or exact values of our equation right hand side variable, the corresponding estimates for the full and innovation samples and subsamples remain close as in the case of the other equations, and the OLS estimates do not differ significantly from the ML ordered probit estimates, nor the 2SLS, which again are very imprecise.[47]

As for the patent equation, the OLS and ML estimates indicate a much lower R&D elasticity than our ALS estimates, of about 0.1 instead of 0.3 or 0.4, and the 2SLS estimates are so inaccurate that they do not statistically differ from zero. Likewise the OLS, 2SLS and ML estimates are about equivalent (both for the complete samples and positive R&D subsamples), but differ markedly from our preferred ALS estimates (on the complete samples). Again this suggests that simultaneity interacts with selectivity and that the two must be taken care of together.

[46] Clearly what matters is to take into account the zero patents, not the fact that the number of patents are count data.

[47] Comparing our preferred estimates in Table 2 to the corresponding second step ALS estimates (i.e., estimated independently of the productivity equation) on the innovation sample in Table 6, shows that they are a little different, although not statistically so considering their standard errors: the estimated R&D capital elasticity is about 0.4 in Table 2 as against about 0.3 in Table 6

C.4 Productivity equation

The last three Tables 7, 8 and 9 present the different types of estimates for the productivity equation, respectively using as proxy for knowledge capital, the more usual R&D capital variable and our two newer innovation output variables: the number of patents and innovative sales (number of sales weighted innovations). What we see in these Tables largely confirm the remarks we just made as concerns the research and innovation equations, though with some exceptions.

The estimates of the same type are about the same in the full and innovation samples and in the positive R&D or positive patents subsamples. The second step ALS estimates, which are estimated independently of the research equations, are close enough to our preferred estimates in Table 2. There is, however, the surprising exception of the estimated innovative sales elasticity: about 0.5 in Table 9 (for the innovation sample), which is too large, as compared to the more reasonable 0.1 in Table 2. This is also one of the only two cases where we find significant differences between our first step and second step ALS estimates: a smaller 0.2 first step estimates of innovative sales elasticity as against the 0.5 too large second step estimate.[48] The 2SLS estimates tend to be very imprecise and to give implausible values of all parameters, and in particular much too high values of the elasticity of knowledge capital. As they do for R&D elasticity in the patent and innovative sales equations, the OLS estimates tend to indicate a smaller elasticity of knowledge capital than the ALS estimates. This is the case when knowledge capital is measured in terms of number of patents and innovative sales (number of sales weighted innovations), but not surprisingly when we use an R&D capital measure. As before this points to the need to account explicitly for both selectivity and simultaneity.

[48]The second case is that of the ALS estimates of R&D capital elasticity in Table 7: the first step estimates are about 0.05 in the full and innovation samples, as compared to about 0.10 in the full sample or about 0.15 in the innovation sample for the second step estimates.

Table 4: Research and Development Equation

Dependent variable: logarithm of research capital per employee (k_i)
(standard errors between parentheses)

Sample	**Full**			**Innovation**		
Variable	l_i	s_i	d_i	l_i	s_i	d_i
Generalized tobit:	6145 firms[a]			4164 firms[b]		
(maximum likelihood)						
– probit part	0.412 (0.038)	0.244 (0.030)	0.323 (0.097)	0.398 (0.038)	0.221 (0.031)	0.294 (0.102)
– tobit part	-0.005 (0.067)	0.378 (0.053)	0.321 (0.142)	-0.001 (0.068)	0.365 (0.054)	0.300 (0.146)
	657 firms			636 firms		
OLS	-0.198 (0.056)	0.232 (0.046)	0.136 (0.131)	-0.188 (0.056)	0.231 (0.047)	0.130 (0.134)

All regressions include 18 industry dummies.
a. Correlation among the residuals, $\widehat{\rho}$ =0.698 (0.074).
b. Correlation among the residuals, $\widehat{\rho}$ =0.709 (0.074).

Table 5: Patent Equation

Dependent variable: number of patents per employee (n_i)
(standard errors between parentheses)

Sample	Full			Innovation		
Variable	k_i	$l_i^\dagger$	g_i	k_i	$l_i^\dagger$	g_i
All observations	6145 firms			4164 firms		
Pseudo maximum likelihood *(negative binomial)*	0.410 (0.075)	0.118 (0.055)	1.334 (0.169)	0.421 (0.074)	0.076 (0.057)	1.203 (0.164)
Asymptotic least squares *(first step: OLS)*	0.854 (0.240)	-0.042 (0.069)	⋆	0.745 (0.251)	-0.049 (0.072)	⋆
Asymptotic least squares *(second step: GLS)*	1.055 (0.149)	-0.051 (0.069)	⋆	0.979 (0.155)	-0.065 (0.071)	⋆
Positive patents	712 firms			669 firms		
Ordinary least squares	0.426 (0.042)	-0.493 (0.029)	0.310 (0.078)	0.427 (0.043)	-0.481 (0.031)	0.298 (0.082)
Two stages least squares	1.154 (0.735)	-0.313 (0.205)	-0.806 (1.256)	1.158 (0.799)	-0.255 (0.269)	-1.139 (1.691)
Pseudo maximum likelihood *(negative binomial)*	0.385 (0.052)	-0.325 (0.042)	0.291 (0.115)	0.385 (0.052)	-0.319 (0.043)	0.281 (0.116)
Positive R&D	657 firms			636 firms		
Pseudo maximum likelihood *(negative binomial)*	0.485 (0.073)	0.005 (0.070)	⋆	0.487 (0.073)	0.005 (0.070)	⋆
Positive Patents and R&D	341 firms			332 firms		
Ordinary least squares	0.442 (0.051)	-0.357 (0.047)	⋆	0.443 (0.052)	-0.361 (0.047)	⋆
Two stages least squares	0.466 (0.212)	-0.358 (0.047)	⋆	0.467 (0.215)	-0.362 (0.048)	⋆
Pseudo maximum likelihood *(negative binomial)*	0.412 (0.051)	-0.248 (0.046)	⋆	0.415 (0.052)	-0.255 (0.051)	⋆

All regressions include 18 industry dummies
Instruments for 2SLS: l_i, s_i, d_i and the 18 industry dummies.
† The coefficient reported is the coefficient of the logarithm of the number of employees minus unity which measures departure from constant returns.

Table 6: Innovation Intensity Equation

Dependent variable: logarithm of innovations percentage in sales (t_i)
(standard errors between parentheses)

Sample	Full			Innovation		
Variable	k_i	l_i	g_i	k_i	l_i	g_i
All observations	6145 firms			4164 firms		
Ordinary least squares*	0.111 (0.025)	0.096 (0.010)	0.430 (0.039)	0.114 (0.029)	0.060 (0.013)	0.306 (0.046)
Two stages least squares*	4.615 (4.328)	0.053 (0.290)	1.630 (1.891)	2.803 (3.359)	0.064 (0.327)	0.780 (2.009)
Maximum likelihood†	0.127 (0.041)	0.164 (0.019)	0.549 (0.065)	0.136 (0.037)	0.079 (0.018)	0.361 (0.059)
Asymptotic least squares *(first step: OLS)*	0.434 (0.162)	0.048 (0.044)	⋆	0.260 (0.119)	0.016 (0.031)	⋆
Asymptotic least squares *(second step: GLS)*	0.461 (0.090)	0.047 (0.044)	⋆	0.280 (0.066)	0.015 (0.031)	⋆
Positive R&D	657 firms			636 firms		
Ordinary least squares*	0.107 (0.033)	0.049 (0.030)	⋆	0.121 (0.033)	0.043 (0.030)	⋆
Two stages least squares*	0.108 (0.168)	0.049 (0.030)	⋆	0.094 (0.169)	0.044 (0.030)	⋆
Maximum likelihood†	0.128 (0.036)	0.053 (0.032)	⋆	0.142 (0.036)	0.046 (0.032)	⋆

All regressions include 18 industry dummies.
Instruments for 2SLS: l_i, s_i, d_i and the 18 industry dummies.
* The logarithm of the interval center was taken as the dependent variable.
† Ordered probit estimates (with known thresholds).

Table 7: Productivity Regression with Research Capital

Dependent variable: logarithm of value added per employee (q_i)
(standard errors between parentheses)

Sample	**Full**				**Innovation**			
Variable	k_i	c_i	l_i	g_i	k_i	c_i	l_i	g_i
All observations	6145 firms				4164 firms			
Ordinary least squares	0.119 (0.010)	0.213 (0.005)	0.009 (0.004)	0.159 (0.010)	0.119 (0.010)	0.207 (0.006)	0.007 (0.005)	0.158 (0.016)
Two stages least squares	9.679 (11.052)	-0.073 (0.284)	0.306 (0.593)	-0.343 (3.356)	7.219 (10.475)	-0.174 (0.485)	0.360 (0.812)	-0.801 (4.141)
Asymptotic least squares *(first step: OLS)*	0.065 (0.014)	0.204 (0.006)	-0.027 (0.007)	⋆	0.048 (0.028)	0.207 (0.007)	-0.030 (0.008)	⋆
Asymptotic least squares *(second step: GLS)*	0.119 (0.014)	0.206 (0.006)	-0.019 (0.007)	⋆	0.149 (0.019)	0.207 (0.007)	-0.034 (0.007)	⋆
Positive R&D	657 firms				636 firms			
Ordinary least squares	0.154 (0.012)	0.096 (0.021)	0.009 (0.011)	⋆	0.155 (0.012)	0.098 (0.022)	0.007 (0.011)	⋆
Two stages least squares	0.407 (0.098)	-0.046 (0.061)	0.025 (0.015)	⋆	0.419 (0.103)	-0.062 (0.068)	0.023 (0.016)	⋆

All regressions include 18 industry dummies.
Instruments for 2SLS: l_i, c_i, s_i, d_i and the 18 industry dummies.

Table 8: Productivity Regression with Patents

Dependent variable: logarithm of value added per employee (q_i)
(standard errors between parentheses)

Sample	Full				Innovation			
Variable	n_i	c_i	l_i	g_i	n_i	c_i	l_i	g_i
All observations	6145 firms				4164 firms			
Ordinary least squares	0.040 (0.011)	0.219 (0.005)	0.016 (0.004)	0.089 (0.015)	0.044 (0.011)	0.217 (0.007)	0.014 (0.005)	0.094 (0.015)
Two stages least squares	3.197 (1.132)	0.171 (0.028)	0.208 (0.136)	0.417 (0.688)	1.596 (0.556)	0.207 (0.017)	0.102 (0.092)	0.457 (0.426)
Asymptotic least squares *(first step: OLS)*	0.066 (0.021)	0.208 (0.006)	−0.026 (0.007)	⋆	0.113 (0.029)	0.207 (0.007)	−0.025 (0.009)	⋆
Asymptotic least squares *(second step: GLS)*	0.142 (0.020)	0.208 (0.006)	−0.026 (0.010)	⋆	0.166 (0.027)	0.207 (0.007)	−0.030 (0.012)	⋆
Positive Patents	712 firms				669 firms			
Ordinary least squares	0.055 (0.012)	0.213 (0.020)	0.045 (0.012)	⋆	0.056 (0.013)	0.214 (0.020)	0.038 (0.012)	⋆
Two stages least squares	0.866 (0.492)	0.047 (0.114)	0.436 (0.239)	⋆	0.769 (0.448)	0.079 (0.010)	0.373 (0.212)	⋆

All regressions include 18 industry dummies.
Instruments for 2SLS: l_i, c_i, s_i, d_i and the 18 industry dummies.

Table 9: Productivity Regression with Innovation Intensity

Dependent variable: logarithm of value added per employee (q_i)
(standard errors between parentheses)

Sample	Full			Innovation		
Variable	t_i	c_i	l_i	t_i	c_i	l_i
All observations	6145 firms			4164 firms		
Ordinary least squares*	0.035 (0.005)	0.219 (0.005)	0.018 (0.004)	0.035 (0.005)	0.220 (0.007)	0.019 (0.004)
Two stages least squares*	0.643 (0.094)	0.165 (0.013)	−0.059 (0.014)	0.773 (0.154)	0.196 (0.016)	−0.044 (0.017)
Asymptotic least squares *(first step: OLS)*	0.057 (0.050)	0.208 (0.006)	−0.031 (0.008)	0.228 (0.121)	0.207 (0.007)	−0.034 (0.010)
Asymptotic least squares *(second step: GLS)*	0.308 (0.051)	0.209 (0.006)	−0.047 (0.014)	0.542 (0.116)	0.207 (0.007)	−0.043 (0.017)

All regressions include 18 industry dummies.
Instruments for 2SLS: l_i, c_i, s_i, d_i and the 18 industry dummies.
* The logarithm of the interval center was taken as the right-hand variable.