Innovation and productivity: a firm-level analysis for French Manufacturing and Services using CIS3 and CIS4 data (1998-2000 and 2002-2004)

Preliminary version, March 2008

Jacques Mairesse[♣], Stéphane Robin[♠]

Extended abstract:

This paper investigates the effect of innovation on labour productivity in France, using a general framework that accounts for research activities and for both product and process innovation. Using a variant of the model proposed by Griffith, Huergo, Peters and Mairesse (2006), we estimate a three stages econometric model, where the estimated output at a given stage is used as an input in the next stage. First, we control for selection into R&D activities and identify the determinants of R&D intensity using a Heckman-type procedure (Generalized Tobit model). Second, we estimate the "knowledge production function" using a bivariate Probit model, distinguishing between product and process innovation. Finally, we estimate the impact of both types of innovation on labour productivity.

This 'structural' model is estimated using two waves of the French component of the Community Innovation Survey (CIS): CIS3, which covers the 1998-2002 period, and CIS4, which covers the 2002-2004 period. A comparative analysis is conducted on the manufacturing industries using both waves of the survey, and a specific analysis is conducted on services using CIS4 only. We find a significantly positive effect of product innovation on labour productivity in the manufacturing industry over both observation periods. Process innovation also has a positive effect on labour productivity, but this effect is more significant when process innovation is conducted together with product innovation. The analysis conducted on the services industry over the 2002-2004 period reveals a similar pattern.

The rest of this extended abstract is organized as follows: in Section 1, we detail the structural econometric model. In Section 2, after giving a brief presentation of the data, we define the explanatory variables used in the various stages of the structural model. Finally, Section 3 gives a more detailed review of our results, with complete tables of estimates for each stage of the model.

1. Econometric modelling

The structural model of innovation and productivity developed by Crépon *et al.* (1998) – usually referred to as the CDM model – has enjoyed a certain amount of success in the literature over the recent years. CDM-type models are generally built as three-stage econometric models that relate productivity to the production of knowledge, which depends on firms' R&D effort, which itself is determined by a number of firm- and environment-

[♣] CREST-ENSAE, 15 Boulevard Gabriel Peri, 92245 Malakoff Cedex (France). MERIT-UNU (Maastricht University) and NBER . mairesse@ensae.fr

[♠] BETA, University of Strasbourg 1, 61 Avenue de la Foret Noire, 67085 Strasbourg Cedex (France), robin@cournot.u-strasbg.fr

specific factors. The model proposed by Griffith *et al.* (2006) is particularly interesting, as it extends the CDM framework in an attempt to take into account process as well as product innovation.

What we propose here is to estimate a variant of the latter model on two waves of the French component of the CIS, in both manufacturing and the services. Our primary objective is to observe the evolution of the innovation – productivity relationship in France over two periods of time. Our secondary objective is to compare this relationship in the manufacturing and services industries in the recent period. The model follows a now classic three-stage framework, which is detailed below.

1.1. Determinants of firms' R&D effort

The first stage of our model accounts for firms' R&D effort (or investment in R&D). For any firm *i*, we write:

$$(1) \qquad r_i^* = z'_i \beta + e_i$$

where r_i^* is an unobserved latent variable, z_i a vector of explanatory variables, β the associated vector of parameters to be estimated, and e_i a random error. Firms' R&D effort is measured by their R&D expenditures, denoted by r_i. However, firms do not systematically report these expenditures (either because they do not invest in R&D, or because they wish to keep this information secret). Thus, Equation (1) can only be estimated at the risk of selection bias. To avoid such bias, we introduce the following selection equation describing whether a firm is doing (and/or reporting) R&D or not:

$$(2) \qquad D_i = \begin{cases} 1 & \text{if } D_i^* = w'_i \alpha + \varepsilon_i > c \\ 0 & \text{if } D_i^* = w'_i \alpha + \varepsilon_i \le c \end{cases}$$

where D_i is an observed binary variable equal to 1 for firms reporting R&D expenditures, and to 0 non-R&D firms. Firms report R&D if and only if the corresponding latent variable D_i^* is above a certain threshold level *c*. The latent variable D_i^* is assumed to be a linear function of a vector of explanatory variables w_i (to which is associated the vector of unknown parameters α), and of a random error term ε_i.

The amount of resources invested in R&D is observed conditional on firms reporting R&D, which we write:

$$(3) \qquad r_i = \begin{cases} r_i^* & \text{if } D_i = 1 \\ 0 & \text{if } D_i = 0 \end{cases} \quad \Leftrightarrow \quad r_i = \begin{cases} z'_i \beta + e_i & \text{if } D_i = 1 \\ 0 & \text{if } D_i = 0 \end{cases}$$

Assuming the error terms e_i and ε_i are bivariate normal (with zero mean, variances $\sigma_e = 1$ and σ_ε, and correlation coefficient $\rho_{e\varepsilon}$), we estimate the system of Equation (2) and Equation (3) as a generalized Tobit model by maximum likelihood.

1.2. The knowledge production function

The second stage of the model deals with the knowledge production function (or innovation production function), which relates measures of knowledge to innovation inputs (represented here by R&D expenditures) and to characteristics of firms and their environment. Our model accounts for both product and process innovation. Our proxy variables for the knowledge output are simply two binary variables indicating (i) whether a firm does product innovation or not, and (ii) whether a firm does process innovation or not.

Let g_1 and g_2 respectively be these indicators, we can assume they are related to two latent variables g_1^* and g_2^*, so that for any firm *i*:

(4) $$g_{1i} = \begin{cases} 1 & \text{if } g_{1i}^* > 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad g_{2i} = \begin{cases} 1 & \text{if } g_{2i}^* > 0 \\ 0 & \text{otherwise} \end{cases}.$$

We then relate g_1^* and g_2^* to firms' R&D expenditures and to other explanatory variables, assuming in each case a linear relationship:

(5) $$\begin{cases} g_{1i}^* = \hat{r}_i \gamma_1 + x'_{1i}\,\delta_1 + u_{1i} \\ g_{2i}^* = \hat{r}_i \gamma_2 + x'_{2i}\,\delta_2 + u_{2i} \end{cases}$$

where $\hat{r}_i$ is the predicted value of R&D expenditures (obtained from the generalized Tobit model estimated in the previous stage), x_{ji} (j = 1, 2) a vector of other explanatory variables, and γ_j and δ_j (j = 1, 2) their respective associated vectors of parameters.

Taken together, Equation (4) and Equation (5) specify our knowledge production function as a bivariate Probit model, if we assume the error terms u_{ji} (j = 1, 2) to be bivariate normal with correlation coefficient ρ. The bivariate specification takes into account the fact that product and process innovation can be jointly determined. Taking the predicted value of R&D expenditures allows us to estimate the bivariate Probit model for all firms, and not only for those reporting R&D expenditures. By using the predicted value, we also instrument the R&D effort and take into account the fact that it could be endogenous to the knowledge production function. In other words, we account for the fact that unobservable characteristics could increase both firms' R&D effort and firms' "innovativeness" or "innovativity" (i.e. their "productivity" in producing innovations, as defined in: Mairesse and Mohnen, 2002; Mohnen *et al.*, 2006). If that were the case, r_i^* and u_{ji} would be positively correlated, and the γ_j parameters in Equation (5) would be biased upward. Using the predicted value from the selection and R&D effort equations correct for this, as long as w_i and z_i are independent of u_{ji}.

1.3. Labour productivity

The third and final stage of the model consists in estimating the impact of product and process innovation on firms' labour productivity[1] y_i (defined as the natural logarithm of turnover per worker). From the bivariate Probit estimated in the previous stage, we can infer three different predictions: the probability to do product innovation only, the probability to do process innovation only, and the probability to do both product and process innovation. We specify firms' productivity as a linear function of these predicted probabilities:

(6) $$y_i = \pi_1\,\hat{\Pr}_i(g_1 = 1,\ g_2 = 0) + \pi_2\,\hat{\Pr}_i(g_1 = 0,\ g_2 = 1) + \pi_3\,\hat{\Pr}_i(g_1 = 1,\ g_2 = 1) + \pi_0 x_{0i} + v_i$$

where x_{0i} is a vector of control variables, π_0 its associated vector of parameters, and v_i a normally distributed random error term. Using the predicted probabilities from our knowledge production function allows us to control for the possible endogeneity of knowledge outputs in Equation (6).

[1] Our data does not allow us to compute Total Factor Productivity.

In a nutshell, our structural model consists of the five equations numbered (2), (3), (4), (5) and (6). Equation (2) and Equation (3) are estimated simultaneously; so are Equation (4) and Equation (5). Equations (4) and (5) use the prediction from Equations (2) and (3) as an input, while Equation (6) uses the predictions of Equation (4) and (5) as inputs. Assuming a recursive model structure that does not allow for feedback effects, we follow a three-step estimation procedure: in the first step, we estimate the generalized Tobit model defined by Equations (2) and (3), using Maximum Likelihood. In the second step, we jointly estimate the knowledge production functions for product and process innovations, using Maximum Likelihood to estimate the bivariate Probit model defined by Equations (4) and (5). In the final step, we estimate the productivity equation, Equation (6), by OLS.

2. Data and variables

2.1. The French CIS3 and CIS4 databases

The present study uses data from the third and fourth waves of the Community Innovation Surveys (CIS3 and CIS4). The CIS is a harmonised survey that is carried out by national statistical agencies in all 25 EU Member States under the co-ordination of Eurostat. CIS3 was conducted in 2001 and provides information for the period 1998-2000, whereas CIS4 was conducted in 2005 and provides information for the period 2002-2004. Both surveys cover R&D activities, product and process innovation, as well as organizational changes and, to some extent, innovations in marketing.

They however differ on some key aspects: first, CIS3 gives some information about firms' investment in physical capital, and distinguishes among different type of human capital (i.e., low-skill and high-skilled, the latter being measured by the proportion of employees with higher education in the workforce). This information has disappeared in CIS4, which has been extended into other directions: first, CIS4 now samples firms with 10 employees or more, where CIS3 only included firms with 20 employees or more. Moreover, CIS3 was focused mostly on manufacturing firms, whereas CIS4 covers the services industry quite extensively.

TABLE 1 ABOUT HERE

We will therefore use CIS3 and CIS4 to draw a comparison in two directions: first, we will examine how the relationship between R&D activities, innovation, and productivity in the French manufacturing industry has evolved across time. To do so, we will compare the sub-sample of manufacturing firms provided by CIS3, and covering 1998-2000, to the one provided by CIS4, and covering 2002-2004. Then, we will use CIS4 to examine whether this relationship differs in the manufacturing and services industries in the recent period.

Table 1 gives of breakdown by industry of the three sub-samples that we use for this analysis (manufacturing firms observed in CIS3, manufacturing firms observed in CIS4, and services firms observed in CIS4). We use industry categories similar to those used in Griffith *et al.* (2006) and based upon the 2-digit NACE classification. It is interesting to notice that the distribution of manufacturing firms across industry categories is quite similar in the third and fourth waves of the survey. We cannot, unfortunately, make the same comparison for services firms, since only CIS4 provides a complete sampling of the services industry.

2.2. Choice of variables

In this sub-section, we detail the specification of the three econometric models that compose our structural model, in terms of choice of variables. The first component of our structural model is the generalized Tobit model defined in Section 1.1. In this model, Equation (2) is the selection equation (specified as a Probit function) and Equation (3) is the intensity equation (specified as a linear function augmented by the inverse Mills ratio).

In our empirical application, the dependent variable D_i from Equation (2) is defined as a binary indicator of **continuous R&D engagement**: it is equal to 1 if firm *i* reports continuous engagement in R&D activities during the observation period (i.e. the two years that precede the year of the survey), and to 0 otherwise. The vector of explanatory variables w_i used in Equation (2) include:

- A binary indicator of **international competition,** which is equal to 1 if firm *i*'s most significant market is international, and to 0 otherwise.
- Two dummy variables that characterize **appropriability conditions**, i.e. that describe how firms protected their inventions or innovations during the observation period:
  - ***Formal protection***, which is equal to 1 if firm *i* used patents, design patterns, trademarks or copyrights, and to 0 otherwise.
  - ***Strategic protection*,** which is equal to 1 if firm *i* used complexity of design, secrecy or lead-time advantage on competitors, and to 0 otherwise.
- **Firm size**, which is measured by the number of employees *two years before* the year of the survey, and is represented by a categorical variable. The 5 categories are: less than 50 employees, 50 to 99 employees, 100 to 249 employees, 250 to 999 employees, and 1000 or more employees.
- An **industry fixed-effect** using the 2-digit industry dummy variables presented in Table 1.
- Variables indicating to which extent innovation was **demand pulled** or **technology pushed** in the 3-digit industry where firm *i* operated during the observation period. These variables are built using a question specific to French CIS surveys, which asks firms about the respective importance of market and technological conditions. In each case, three variables give the share of firms where innovation was weakly / relatively / strongly influenced by market (or technological) conditions, while a fourth variable indicating no influence at all is taken as the reference.

The dependent variable we use in Equation (3) is the natural logarithm of R&D intensity, defined as the ratio of R&D expenditures per employee at the time of the survey. To identify the generalized Tobit model, some explanatory variables used in Equation (2) must be excluded from Equation (3) (i.e. from the R&D intensity equation). We chose to exclude firm size: indeed, previous studies have shown that, in several European countries (including France), firm size influences the probability to do R&D, but not R&D investments. The other explanatory variables from Equation (2) are also included in Equation (3). In addition, we include some explanatory variables that are only observed when firms report R&D expenditures. These additional explanatory variables are:

- **Cooperation**, a dummy variable equal to 1 if the firm had some cooperative agreements on innovation activities during the observation period, and to 0 otherwise.
- A 4-category variable indicating whether a firm received some amount of **public support** in the form of funding for innovation projects during the observation period. The four categories are:

- *Local funding*, which indicates local or regional funding.
    - ***National funding***, which indicates funding from the national government.
    - ***EU funding***, which indicates funding from the European Unions, such as funding received through participation in a Framework Programme.
- A set of dummy variables indicating whether a firm relied on various **sources of information** for its innovation activities. This dummy variables include:
    - ***Internal sources within the firm***, a dummy variable equal to 1 if information from internal sources within the firm was of high importance during the observation period, and to 0 otherwise.
    - ***Internal sources within the group***, which indicates whether or not information from internal sources within the group was of high importance.
    - ***Suppliers***, which indicates whether suppliers were a highly important source of information during the observation period.
    - ***Customers***, which indicates whether customers were a highly important source of information during the observation period.
    - ***Competitors*****,** which indicates whether competitors were a highly important source of information during the observation period.
    - ***Universities*****,** which indicates whether universities (or other higher education institutions), government labs, or non-profit research institutes were a highly important source of information during the observation period

As explained in Section 1.2, the predicted value of (the log) R&D intensity is used as an explanatory variable in the second stage of our structural model, which consists in estimating the bivariate Probit defined by Equations (4) and (5). The dependent variables in the bivariate Probit model are two dummy variables indicating product and process innovation respectively during the observation period. The **product innovation** dummy variable is equal to 1 if the firm reports having introduced new or significantly improved products (new to the market or only new to the firm) during the observation period. The **process innovation** dummy variable is equal to 1 if the firm reports having implemented new or significantly improved production processes during the observation period. The same vector of explanatory variables is used for both product and process innovation, which can be noted: $x_{1i} = x_{2i} = x_i$. Besides the predicted value of R&D intensity, these explanatory variables include our indicators of **appropriability conditions** (as defined above), **firm size** (as defined above), and the **industry fixed-effect** defined at the 2-digit level.

Using the bivariate Probit model, we are able to predict three probabilities, as explained in Section 1.3: the probability to do product innovation only, the probability to do process innovation only, and the probability to do both product and process innovation. These three predicted probabilities are used as explanatory variables in the final stage of our model, which consists in estimating the labour productivity equation, i.e. Equation (6). Our empirical measure of **labour productivity** is the natural logarithm of sales per employee, observed in the year of the survey. Besides the aforementioned predicted probabilities, Equation (6) includes control variables regrouped in the x_{0i} vector. These other explanatory variables are simply **firm size** (as defined above) and the **industry fixed-effect** defined at the 2-digit level.

2.3. Comparison of the samples

Table 2 gives summary statistics for the variables defined in Section 2.2 across our three sub-samples: the CIS3 manufacturing sample (which includes 4222 firms), the CIS4 manufacturing sample (which includes 5969 firms), and the CIS4 services sample (which includes 8743 firms). Overall, the CIS3 and CIS4 manufacturing samples look very similar, although labour productivity seems slightly higher in CIS4 than in CIS3. Another important difference is that innovation seems to have become more demand-pulled between the late 1990s and the early 2000s: the proportion of firms (at the 3-digit industry level) where innovation was weakly demand-pulled has decreased between CIS3 and CIS4, whereas the proportion of firms where innovation was strongly demand-pulled has increased. International competition also seems to have intensified from one period to the next.

TABLE 2 ABOUT HERE

Focusing only on CIS4 emphasizes the differences between manufacturing and services in the recent period: first, the proportion of firms doing R&D continuously is much lower (about 1/3) in the services. Second, although labour productivity seems higher in the services than in manufacturing, it also presents a much larger standard deviation. The proportion of firms receiving public support (especially in the form of government funding) is also much lower in the services, whereas the respective impulsions of market conditions and technology seem quite similar. Services firms seem less involved in knowledge sourcing, and are also less concerned with appropriability conditions, which is consistent with the fact that they invest less in R&D. Finally, they seem more oriented towards local or national markets, and are facing less pressure from international competition.

3. Results

3.1. R&D effort and the R&D equations

We first present the estimates of the generalized Tobit that constitutes the first stage of our structural model. Tables 3.a and 3.b present the results of the selection and of the R&D intensity equations respectively. Firm size appears as a major determinant of the propensity to do R&D, which is consistent with the empirical I.O. literature. Moreover, and quite logically, firms that are able to better protect their inventions or innovations are also more likely to invest in R&D. These results are observed in all three sub-samples.

TABLE 3.a ABOUT HERE
TABLE 3.a ABOUT HERE

International competition and cooperation are both positively associated with a higher R&D intensity, in all three sub-samples, although the marginal effect is higher and more significant in manufacturing than in the services. Appropriability conditions, however, seem to have acquired more importance in the recent period: their marginal effect (controlling for selection into R&D) was not significant in the manufacturing industry in 1998-2000, but has become significant (and positive) in 2002-2004. In the recent period, services firms also seem to condition their investment in R&D on their ability to protect their inventions; however, they seem to rely exclusively on strategic protection, and not on formal protection.

Another interesting contrast appears between manufacturing and services firms: whether in 1998-2000 or in 2002-2004, manufacturing firms tend to invest more in R&D if they receive a funding from the EU. The other types of public support do not seem to matter. Services firms display a different pattern: they tend to invest more in R&D if they receive funding from government agencies or national authorities, whereas the other type of public support do not seem to matter. Finally, knowledge sourcing does not seem to influence R&D intensity in manufacturing, whereas it has a significant influence in the services.

3.2. The knowledge production function

The estimates of the bivariate Probit model that represent our knowledge production function are given in Table 4. The most important result in this table concerns the impact of the predicted value of R&D intensity, which is obtained in the previous stage of the structural model by estimating the generalized Tobit. Table 4 shows that R&D intensity always has a significantly positive impact on the probability to do product innovation. However, the marginal effect on the propensity to do product innovation is much higher in manufacturing (it is about 0.5 in both the CIS3 and CIS4 sub-samples) than in the services (where it is equal to 0.13 only).

TABLE 4 ABOUT HERE

Predicted R&D intensity also have a significantly positive effect on the probability to do process innovation, a result which is again common to all three sub-samples. In the manufacturing industry, the marginal effect of R&D intensity on the propensity to do process innovation has increased from 0.30 to 0.42 between 1998-2000 and 2002-2004. In the services industry, the marginal effect is equal to 0.09, which is again much lower than in the manufacturing industry. Finally, we note that the correlation coefficient of the errors terms, denoted by ρ in Equation (5), is significantly different from zero in all sub-samples, which justifies using a bivariate Probit instead of two unrelated Probit models as was done in Griffith *et al.* (2006).

3.3. The labour productivity equation

Finally, we come to the impact of innovation on firms' labour productivity, using the predicted probability from the previous-stage regression. The results of the estimation of the labour productivity equation are given in Table 5. Qualitatively, all sub-samples display similar results: product innovation, process innovation, and process combined with product innovation are all associated with a higher labour productivity. However, quantitative differences appear across sub-samples.

TABLE 5 ABOUT HERE

First, in the manufacturing industry, the effect of product innovation has become more important: it has doubled between the late nineties and 2002-2004. On the contrary, between these two periods, the effect of process innovation on labour productivity has been divided by three, whereas the effect of process innovation combined with product innovation has remained roughly the same. Comparing services and manufacturing firms in the recent period reveals another trend: the respective impacts of product innovation alone and process innovation alone on labour productivity are approximately 4 times higher in the services. The impact of process innovation combined with product innovation is slightly more important in the services as well, but the difference is not quite as large.

In nutshell, we can conclude that, in spite of the significant differences that exist between the manufacturing and services industries, all types of innovation (product, process, and the combination of both) have a positive impact on French firms' labour productivity.

References

Crépon B., Duguet E. and J. Mairesse (1998) "Research, Innovation and Productivity: An Econometric Analysis at the Firm Level", *Economics of Innovation and New Technology*, 7, 115-158.

Griffith R., Huergo E., Mairesse J. and B. Peters (2006) "Innovation and Productivity across Four European Countries", *Oxford Review of Economic Policy*, 22(4), 483-498.

Mairesse J. and P. Mohnen (2002) "Accounting for Innovation and Measuring Innovativeness: An Illustrative Framework and an Application", *American Economic Review*, 92(2), 226-230.

Mohnen P., Mairesse J., and M. Dagenais (2006) "Innovativity: a Comparison Across Seven European Countries", *NBER Working Papers*, n° 12280.

Table 1: Number and proportion of firms by industry in manufacturing and services for CIS3 and CIS4

		CIS3		CIS4	
Industry	*Nace*		%		%
Food/Tobacco	15-16	698	16.5	1014	17.0
Textile	17-19	475	11.3	648	10.9
Wood/Paper	20-22	386	9.1	706	11.8
Chemicals	23-24	374	8.9	451	7.6
Plastic/Rubber	25	276	6.5	313	5.2
Non-metallic min.	26	168	4.0	273	4.6
Basic metals	27-28	607	14.4	699	11.7
Machinery	29	418	9.9	439	7.4
Electrical	30-33	452	10.7	657	11.0
Vehicles	34-35	192	4.6	483	8.1
Misc.	36-37	176	4.2	286	4.8
All Manufacturing		4222	100	5969	100
Trade	50-52			3223	36.9
Horeca	55			505	5.8
Transports	60-63			1268	14.5
Communication	64			78	0.9
Finance/Insurance	65-67			523	6.0
Housing/Real Estate	70			269	3.1
Rental	71			179	2.1
ICT services	72			406	4.6
R&D	73			130	1.5
Services to firms	74			2162	24.7
All Services				8743	100

Note: The industry definition is based on the classification system NACE (*Nomenclature générale des activités économiques dans les Communautés Européennes*) as published by Eurostat (1992), using 2-digit levels.

Table 2: summary statistics across samples

	CIS3	CIS4	
	Manufacturing	**Manufacturing**	**Services**
Knowledge/Innovation:			
Continuous R&D engagement	0.34 (0.47)	0.32 (0.47)	0.12 (0.33)
R&D per employee (for firms w/ continuous R&D engagement)	7.2 (19.6)	7.2 (19.2)	6.56 (14.7)
Innovator (product and/or process innovation)	0.52 (0.50)	0.54 (0.50)	0.29 (0.45)
Process innovation	0.32 (0.46)	0.38 (0.49)	0.21 (0.41)
Product innovation	0.46 (0.50)	0.42 (0.49)	0.20 (0.40)
Share of sales with new products (for firms with product innovation)	0.22 (0.26)	0.22 (0.24)	0.20 (0.26)
Labour productivity	198.5 (638.5)	224.2 (598.2)	275.8 (1552.1)
Public Support:			
Local funding	0.06 (0.23)	0.05 (0.21)	0.02 (0.13)
National funding	0.15 (0.35)	0.11 (0.31)	0.03 (0.18)
EU funding	0.05 (0.22)	0.04 (0.19)	0.02 (0.12)
Proportion of firms where innovation was:			
Not demand pulled	0.07 (0.05)	0.06 (0.04)	0.10 (0.05)
Weakly demand pulled	0.15 (0.19)	0.04 (0.04)	0.06 (0.03)
Relatively demand pulled	0.28 (0.09)	0.21 (0.06)	0.23 (0.05)
Strongly demand pulled	0.50 (0.22)	0.69 (0.09)	0.61 (0.09)
Not technology pushed	0.17 (0.08)	0.16 (0.08)	0.23 (0.09)
Weakly technology pushed	0.20 (0.10)	0.20 (0.08)	0.15 (0.05)
Relatively technology pushed	0.41 (0.10)	0.36 (0.08)	0.32 (0.07)
Strongly technology pushed	0.22 (0.12)	0.27 (0.11)	0.30 (0.12)
Sources of information:			
Internal sources within the enterprise	0.47 (0.50)	0.49 (0.50)	0.28 (0.45)
Internal sources within the group	0.22 (0.41)	0.28 (0.45)	0.19 (0.39)
Suppliers as source of information	0.28 (0.45)	0.32 (0.47)	0.19 (0.39)
Customers as source of information	0.42 (0.49)	0.36 (0.48)	0.18 (0.38)
Competitors as source of information	0.33 (0.47)	0.23 (0.42)	0.12 (0.33)
Universities/ Government as source of information	0.11 (0.31)	0.11 (0.32)	0.05 (0.21)
Appropriability conditions:			
Formal protection dummy variable	0.45 (0.50)	0.50 (0.50)	0.29 (0.45)
Strategic protection dummy variable	0.27 (0.45)	0.37 (0.48)	0.16 (0.36)
Cooperation dummy variable	0.26 (0.44)	0.28 (0.45)	0.16 (0.36)
Other:			
International competition	0.39 (0.49)	0.52 (0.50)	0.19 (0.39)
Size: <50	0.30 (0.46)	0.35 (0.48)	0.41 (0.49)
Size: 50-99	0.18 (0.38)	0.19 (0.40)	0.20 (0.40)
Size: 100-250	0.20 (0.40)	0.17 (0.38)	0.15 (0.36)
Size: 250-999	0.24 (0.43)	0.23 (0.42)	0.18 (0.39)
Size: >1000	0.08 (0.27)	0.05 (0.23)	0.05 (0.23)
Observations	4222	5969	8743

Notes: standard deviations in parentheses.
Data are from the French CIS3 and CIS4.
CIS3 variables cover the years 1998-2000, with the exception of R&D per employee, labour productivity and investment per employee (related to 2000) and size (related to the number of employees in 1998).
CIS4 variables cover the years 2002-2004, with the exception of R&D per employee, labour productivity and investment per employee (related to 2004) and size (related to the number of employees in 2000). All values are in thousands of Euros.

Table 3.a: R&D equations – selection equation

Variables	(1) CIS3 Manufacturing		(2) CIS4 Manufacturing		(3) CIS4 Services	
	Coeff.	*Marg Eff*	*Coeff.*	*Marg Eff*	*Coeff.*	*Marg Eff*
Constant	-3.28		-3.46		-2.23	
	(0.71)***		(0.57)***		(0.58)***	
International competition	0.39	0.14	0.44	0.15	0.43	0.07
	(0.05)***	(0.02)***	(0.04)***	(0.01)***	(0.05)***	(0.01)***
Appropriability conditions						
Formal protection	0.89	0.30	0.57	0.19	0.57	0.10
	(0.05)***	(0.02)***	(0.04)***	(0.01)***	(0.04)***	(0.01)***
Strategic protection	0.75	0.27	0.70	0.24	0.78	0.16
	(0.06)***	(0.02)***	(0.04)***	(0.01)***	(0.05)***	(0.01)***
Firm size (*ref.: < 50 employees*)						
50 to 99 employees	0.28	0.10	0.19	0.06	0.16	0.02
	(0.07)***	(0.03)***	(0.06)***	(0.02)***	(0.05)***	(0.01)**
100 to 249 employees	0.41	0.14	0.40	0.13	0.24	0.03
	(0.07)***	(0.03)***	(0.06)***	(0.02)***	(0.06)***	(0.01)***
250 to 999 employees	0.75	0.26	0.68	0.24	0.37	0.04
	(0.07)***	(0.03)***	(0.05)***	(0.02)***	(0.05)***	(0.01)***
≥ 1000 employees	0.95	0.35	1.08	0.41	0.71	0.11
	(0.10)***	(0.04)***	(0.09)***	(0.03)***	(0.08)***	(0.02)***
Test for 2-digit industry	0.000		0.000		0.000	
Test for "Demand pulled / Technology pushed"	0.079		0.001		0.000	

* Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level

Notes:
Robust Standard Errors in parentheses

All models include 2-digit industry dummy variables and "demand pull and technology pulled" variables in the selection equation. For the sake of concision, we do not show coefficients, but report the p-values of a test of the joint significance of each group of variables.

Since, in all models, both the selection and intensity equations are estimated simultaneously, goodness-of-fit statistics are reported for the whole model at the bottom of Table 3.b.

Table 3.b: R&D equations – intensity equation

Variables	(1) CIS3 Manufacturing		(2) CIS4 Manufacturing		(3) CIS4 Services	
	Marginal Effect		Marginal Effect		Marginal Effect	
	Uncondit.	*Conditional*	*Uncondit.*	*Conditional*	*Uncondit.*	*Conditional*
Constant	-4.42	—	-3.57	—	-11.59	—
	(0.96)***		(1.05)***		(1.93)***	
International competition	0.46	0.33	0.42	0.26	0.78	0.22
	(0.08)***	(0.07)***	(0.09)***	(0.08)***	(0.14)***	(0.13)*
Cooperation	0.31	0.31	0.35	0.35	0.21	0.21
	(0.07)***	(0.07)***	(0.07)***	(0.07)***	(0.12)*	(0.12)*
Appropriability conditions						
Formal protection	0.29	-0.01	0.44	0.23	0.79	0.05
	(0.09)***	(0.08)	(0.09)***	(0.08)***	(0.15***	(0.12)
Strategic protection	0.29	0.04	0.60	0.36	1.38	0.39
	(0.08)***	(0.07)	(0.09)***	(0.08)***	(0.16)***	(0.12)***
Funding *(ref.: no funding)*						
Local funding	0.17	0.17	-0.04	-0.04	-0.35	-0.35
	(0.11)	(0.11)	(0.13)	(0.13)	(0.24)	(0.24)
National funding	-0.09	-0.09	0.12	0.12	0.68	0.68
	(0.08)	(0.08)	(0.08)	(0.08)	(0.16)***	(0.16)***
EU funding	0.41	0.41	0.37	0.37	-0.14	-0.14
	(0.13)***	(0.13)***	(0.14)***	(0.14)***	(0.24)	(0.24)
Test for 2-digit industry	0.000		0.000		0.000	
Test for "Demand pulled / Technology pushed"	0.000		0.001		0.000	
Test for "Sources of Information"	0.165		0.133		0.025	
Rho (correlation coefficient)	0.35 (0.05)***		0.32 (0.06)***		0.69 (0.04)***	
Sigma (Inv. Mills ratio coeff.)	1.29 (0.04)***		1.49 (0.04)***		2.24 (0.10)***	
Log-likelihood	-4088.56		-6151.346		-4571.96	
Wald test of H_0: "$\beta = 0$"	604.75***		498.78***		459.49***	

* Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level

Notes:

Robust Standard Errors in parentheses

Unconditional marginal effects measure the deviation of the expected value of R&D intensity with respect to the explanatory variables, whereas conditional marginal effects measure the deviation of the expected value of R&D intensity conditional on doing R&D with respect to the explanatory variables.

All models include 2-digit industry dummy variables in the intensity equation; "demand pull and technology pulled" variables, as well as "sources of information" variables, are also included in that equation. For the sake of concision, we do not show coefficients or marginal effects, but report the p-values of a test of the joint significance of each group of variables.

Table 4: knowledge production function

Product Innovation Equation						
Variables	**(1) CIS3 Manufacturing**		**(2) CIS4 Manufacturing**		**(3) CIS4 Services**	
	Coefficient	*Marg. Eff.*	*Coefficient*	*Marg. Eff.*	*Coefficient*	*Marg. Eff.*
Constant	0.73		0.79		0.61	
	(0.12)***		(0.10)***		(0.12)***	
Pred log(R&D intensity)	1.34	0.53	1.44	0.55	0.59	0.13
	(0.07)***	(0.03)***	(0.06)***	(0.02)***	(0.03)***	(0.01)***
Appropriability conditions						
Formal protection	0.47	0.18	-0.04	-0.01	0.01	0.00
	(0.06)***	(0.02)***	(0.05)	(0.02)	(0.05)	(0.01)
Strategic protection	0.30	0.12	-0.23	-0.09	-0.04	-0.01
	(0.07)***	(0.03)***	(0.06)***	(0.02)***	(0.07)	(0.01)
Firm size (*ref.: < 50 employees*)						
50 to 99 employees	0.26	0.10	0.05	0.02	0.09	0.02
	(0.07)***	(0.03)***	(0.05)	(0.02)	(0.05)*	(0.01)*
100 to 249 employees	0.14	0.05	0.09	0.04	0.21	0.05
	(0.07)*	(0.03)*	(0.06)	(0.02)	(0.05)***	(0.01)***
250 to 999 employees	0.37	0.15	0.24	0.09	0.32	0.08
	(0.07)***	(0.03)***	(0.06)***	(0.02)***	(0.05)***	(0.01)***
≥ 1000 employees	0.41	0.16	0.35	0.14	0.56	0.16
	(0.12)***	(0.05)***	(0.11)***	(0.04)***	(0.08)***	(0.03)***
Test for 2-digit industry	0.000		0.000		0.000	
Process Innovation Equation						
Variables	**(1) CIS3 Manufacturing**		**(2) CIS4 Manufacturing**		**(3) CIS4 Services**	
	Coefficient	*Marg. Eff.*	*Coefficient*	*Marg. Eff.*	*Coefficient*	*Marg. Eff.*
Constant	-0.02		0.59		0.09	
	(0.11)		(0.09)***		(0.11)	
Pred log(R&D intensity)	0.89	0.30	1.11	0.42	0.35	0.09
	(0.06)***	(0.02)***	(0.05)***	(0.02)***	(0.03)***	(0.01)***
Appropriability conditions						
Formal protection	0.07	0.02	-0.37	-0.14	0.03	0.01
	(0.06)	(0.02)	(0.05)***	(0.02)***	(0.05)	(0.01)
Strategic protection	0.28	0.09	-0.11	-0.04	0.17	0.05
	(0.06)***	(0.02)***	(0.06)**	(0.02)**	(0.06)***	(0.02)**
Firm size (*ref.: < 50 employees*)						
50 to 99 employees	0.12	0.04	0.06	0.02	0.10	0.03
	(0.07)*	(0.02)*	(0.05)	(0.02)	(0.05)**	(0.01)**
100 to 249 employees	0.25	0.08	0.08	0.03	0.22	0.06
	(0.07)***	(0.02)***	(0.05)	(0.02)	(0.05)***	(0.01)***
250 to 999 employees	0.33	0.11	0.12	0.04	0.34	0.10
	(0.07)***	(0.02)***	(0.05)**	(0.02)**	(0.05)***	(0.01)***
≥ 1000 employees	0.52	0.19	0.36	0.14	0.58	0.18
	(0.10)***	(0.04)***	(0.09)***	(0.04)***	(0.07)***	(0.03)***
Test for 2-digit industry	0.000		0.000		0.000	
Rho (correlation coeff.)	0.47 (0.03)***		0.42 (0.02)***		0.56 (0.02)***	
Log Likelihood	-3609.21		-5796.54		-6625.10	
Wald test of H_0: "β_j=0"	2037.61***		2669.12***		2407.60***	

* Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level

Notes: Robust Standard Errors in parentheses.
The marginal effects reported above measure the deviation of the probability of (product or process) innovation with respect to explanatory variables. All models include 2-digit industry dummy variables in both equations. For the sake of concision, we do not show coefficients or marginal effects, but report the p-values of a test of the joint significance of these variables.

Table 5: Output production function (labour productivity equation)

Variables	(1) CIS3 Manufacturing	(2) CIS4 Manufacturing	(3) CIS4 Services
	Coefficient	*Coefficient*	*Coefficient*
Constant	5.00	5.03	5.25
	(0.05)***	(0.04)***	(0.04)***
Product innovation only	0.57	1.09	3.75
	(0.10)***	(0.13)***	(0.41)***
Process innovation only	1.12	0.31	1.44
	(0.48)**	(0.22)	(0.48)***
Product and process innovation	0.52	0.35	0.59
	(0.07)***	(0.05)***	(0.14)***
Firm size (*ref.: < 50 employees*)			
50 to 99 employees	-0.08	0.03	0.06
	(0.03)***	(0.03)	(0.02)***
100 to 249 employees	-0.07	0.09	-0.05
	(0.04)*	(0.03)***	(0.03)*
250 to 999 employees	0.04	0.15	-0.18
	(0.03)	(0.03)***	(0.03)***
≥ 1000 employees	0.27	0.37	-0.45
	(0.05)***	(0.04)***	(0.06)***
Test for 2-digit industry	0.000	0.000	0.000
R^2	0.24	0.21	0.40
***F* test of H_0: "$\beta = 0$"**	78.46***	94.34***	399.89***

* Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level

Notes: Robust Standard Errors in parentheses.

All models include 2-digit industry dummy variables. For the sake of concision, we do not show coefficients or marginal effects, but report the p-values of a test of the joint significance of these variables.