

PRELIMINARY

**R&D AND PRODUCTIVITY: A REEXAMINATION IN LIGHT OF THE
INNOVATION SURVEYS**

Jacques Mairesse
(CREST, EHESS, NBER)

Pierre Mohnen
(MERIT, Univ. Maastricht)

May 2003

This is a revised but yet preliminary version (to be completed) of the paper presented at the EPIP workshop on New Challenges to the Patent System in Munich on April 24/25, 2003.

Part of this study was conducted when we were invited at the Institute for Fiscal Studies (IFS). We thank IFS for providing a very stimulating environment, in particular Rachel

Griffith, for helping us with the UK innovation survey data. We also thank OECD and Eurostat for giving us access to the CIS2 data.

ABSTRACT

This paper estimates a structural model linking R&D, innovation output and productivity along the lines of the Crépon, Duguet and Mairesse (1998) model. It is estimated using the microaggregated data of the innovation surveys of France, Germany, Spain and the United Kingdom from the second wave of CIS2, covering the years 1994-1996.

It reexamines the R&D-productivity relationship at the light of the information provided by the innovation surveys. It proposes an integrated accounting framework for innovation and productivity that can be used to compare innovation across sectors and countries. It also checks the robustness of the estimation results with respect to the measurement of certain variables, the location of exogenous variables in the structural model, the methods of estimation, and the treatment of simultaneity and selectivity.

1. INTRODUCTION

Many studies have estimated the rate of return to research and development (R&D) by relating some measures of output (or a dual measure of it such as cost or profit) to R&D, while controlling for other possible influences.¹ But the R&D-productivity relationship so estimated remains elusive. Another strand of empirical studies has tried to estimate a production function of knowledge relating R&D to patents², or innovations.³ Crépon, Duguet and Mairesse (1998) combine the two types of models and in so doing add some structure to the R&D-productivity relationship. They decompose the total link into a relationship between R&D (an input in the innovation process) to either patents or the share of innovative sales (two alternative measures of the output of innovation), and a relationship between either of the two output measures of innovation and labor productivity.

We follow the lines of Crépon, Duguet and Mairesse (1998), CDM for short, and reexamine the R&D-productivity relationship at the light of the information provided by the innovation surveys. These surveys contain information on R&D (doing R&D or not, continuously or not, and the R&D intensity), on the share of innovative sales (sales of products new to the firm or new to the market), on the occurrence of patent applications as well some characteristics of the reporting enterprises. We use the microaggregated data of the innovation surveys of France, Germany, Spain and the United Kingdom from the second wave of CIS (Community Innovation Surveys), CIS2, which pertains to the years 1994-1996.⁴

In particular, we decompose the rate of return to R&D (in terms of productivity) into a direct effect and an indirect effect that operates through the production of innovation output, that itself feeds onto productivity.

We set up an innovation accounting framework similar to the one presented in Mairesse and Mohnen (2002) to compare innovativeness across industries and across the four countries. Furthermore this innovation accounting framework is integrated with an accounting framework for productivity.

Finally, from a more technical side, we also examine the sensitivity of the results to various measures of innovation input and output, to a level versus a growth rate specification, to various subsamples, and to alternative estimation procedures.

The paper is structured as follows. In section 2 we review the state of existing empirical work on the CDM model. In section 3 we present our simultaneous equations model of

¹ See the surveys by Mairesse and Mohnen (1990), Mairesse and Sassenou (1991), Griliches (1994, 1995), and Hall (1995).

² See for instance the survey by Griliches (1990).

³ See for instance the studies collected in Kleinknecht (1996) and Kleinknecht and Mohnen (2002).

⁴ The Spanish innovation survey had no question on patent applications.

R&D, innovation and productivity and the way we go about estimating it. In section 4 we describe our preferred specification. The data and the measure of some variables are described in section 5. Section 6 presents preliminary results.

2. REVIEW OF THE LITERATURE ON THE CREPON-DUGUET-MAIRESSE MODEL

The basic CDM model consists of a system of three equations: a tobit model explaining R&D (actually composed of two equations: a probit for selecting R&D performers and an R&D-intensity equation for R&D performers), an equation linking innovation output to R&D and an equation linking labor productivity to innovation output. Innovation output is measured by either the number of patents or the share of innovative sales. Each of the three endogenous variables of the model (R&D, innovation output and productivity) can have its idiosyncratic as well as common determinants. The model is applied to French individual data and estimated by asymptotic least squares (minimum distance estimator). Besides offering a structural explanation to the R&D-productivity link, the CDM model also tackles two econometric problems: accounting for selection and simultaneity.

As shown in table 1, variants of this model have been estimated on data from France, Germany, the Netherlands, Chili, Sweden, Norway, Finland, China, the UK, and Italy. These studies like CDM (1998) are based on individual data. They differ from the original CDM model in the choice of endogenous and explanatory variables, or in the estimation method that they use. Some use innovation expenditures in lieu of R&D, some use profitability or revenue per employee instead of productivity, some distinguish between radical and incremental innovation or between product and process innovations, some allow for a feedback effect from productivity on innovation output, some explain productivity growth and not just productivity levels, one study adds training and product quality to the model, two of them compare innovation across countries and a few use panel data. Only two studies, besides CDM, use the ALS estimation procedure. Most use two step procedures to correct for simultaneity and use the inverse Mills' ratio to correct for a selection bias due to censored data.

Table 1 Variants of the Crépon-Duguet-Mairesse (1998) model

Study	Individual Data	Endogenous variables	Estimation method	Other comments
Crépon-Duguet-Mairesse (1998)	France 1986-1990	R&D, patent (or share of innovative sales), labor productivity	ALS	Censored data for R&D
Duguet (2002)	France 1986-1990	Radical innovation, incremental innovation, TFP growth	FiML logit for innov., 2SLS or GMM for TFP growth	Separate estimation for various technological opportunities
Galia and Legros (2003)	France 1994-1996	R&D, innovation output, training, quality, profitability	ALS	Censored data for R&D and training, dichotomous data for quality; allows for feedback effects
Janz, Lööf and Peters (2003)	Germany and Sweden, 1998-2000	Innovation expenditures/employee, innov. sales/employee, and sales/employee	FiML for gen. Tobit on innov. expend., other equations by 2SLS with correction for selection bias	Censored data for innovation expenditures; feedback effect from productivity on innov. output
van Leeuwen-Klomp (2001)	Netherlands 1994-1996	Innovation input (R&D or innov. expend.), innovation output, productivity (in levels or growth rates)	OLS, 3SLS limited system, or 3SLS full system (with or without correction for selectivity)	Productivity measured by revenue per employee or value added per employee; feedback effect from revenues on innov. output
van Leeuwen (2002)	Netherlands Panel data	R&D, innovation output, growth in	FiML gen. tobit for R&D or innovation	Dynamic model for 1994-96 or pooled

	from CIS2 and CIS2.5	revenue/employee	output; separate FIML for growth of revenue/employee with correction for selection bias	model for 1994-96 and 1996-98; innov. output measured by new sales or by new and improved sales.
Benavente (2002)	Chili	R&D, patent (or share of innovative sales), labor productivity	ALS	Censored data for R&D
Lööf and Heshmati (2002a)	Sweden	Innov. expend. per employee, innovative sales per employee, and value added per employee	FIML for generalized Tobit on innov. expend., other equations by 2SLS with correction for selection bias	Also estimated with only radical innovations; productivity estimated in levels and growth rates; feedback effect from productivity on innov. output
Lööf and Heshmati (2002b)	Sweden	Innov. expend. per employee, innovative sales per employee, and labor productivity	FIML for gen. Tobit for innov. input, other equations by 3SLS with correction for selection bias	Labor productivity measured as innov. sales/employee or value added/employee; feedback effect from productivity on innov. output
Lööf, Heshmati, Apslund and Nåås (2002)	Finland, Norway and Sweden 1994-1996	Innov. expend./employee, innovative sales/employee, and labor productivity	FIML for gen. Tobit for innov. input, other equations by 2SLS and 3SLS with correction for	Estimation for all innovations and for radical innovations; feedback effect from productivity on innov. output

				selection bias	
Jefferson, Huamao, Xioajing and Xiaoyun (2002)	China Panel data 1995-1999	R&D, share of innovative sales, productivity (or profitability)	Separate estimation of each equation by OLS and IV	Square term on innovative sales	
Criscuolo and Haskel (2002)	UK 1994-1996	Labor productivity growth, and innovation	Separate estimation of each	Innovation measured by innov. dummy(ies), R&D intensity, or patents	
Parisi, Schiantarelli and Sembenelli (2002)	Italy, panel data 1992-1994 and 1997-1995	Labor productivity growth, product innovation, process innovation	Product and process innovations estimated by logit or conditional logit, product. growth estimated by IV		
Hu and Jefferson (2003)	China (Beijing area) 1991-1997	R&D, output and profit	Individual and SUR estimation of 2 or 3 equations with correction for selection bias		

3. THE MODEL

We start from the CDM model and experiment with various generalizations of it. Our choice of modeling is partly determined by the structure of the innovation survey questionnaires. After a few basic questions on their present and past turnover, number of employees, and the main industry they belong to, respondents are asked three central questions which determine whether they have to respond to the rest of the questionnaire or not, namely whether they have introduced a technologically new or improved product or process or whether they are in the process of, or were unsuccessful in, doing so. If and only if they respond positively to one of those questions, they are asked to fill out the rest of the questionnaire and give information about the sources of the information relevant to their innovation, the reasons why they have innovated, possible cooperations in innovation, the amount and organization of their R&D activity, and the share in sales due to new or improved products.

Consequently, we have little information about non-innovators, in particular, we have information about R&D activities only for innovators. Instead of treating non-innovators as non-R&D performers, we have decided to concentrate on innovators. Not all innovators perform R&D (around 15% do not). In order to sharpen the selection, and also in order to concentrate on serious R&D performers, we have decided to examine only the R&D behavior of continuous R&D performers. Since the correct way of modeling R&D would be via a stock measure, in the absence of time series data the restriction to continuous R&D performers is a way to better approximate the behavior of the stock of R&D. We shall account for a possible R&D selection bias in estimating our model, but we shall not correct for an innovator selection bias (although we plan to test for its sensitivity). The innovation outputs depend on the R&D-intensity, estimated in a separate equation, and both affect labor productivity, also estimated in a separate equation.

The model has three main building blocks: one for R&D (I), one for innovation output (II) and one for productivity (III). Block I comprises a selection equation for continuous R&D performers

$$(I.1) s_i = 1 \text{ if } s_i^* = x_{0i}b_0 + u_{0i} > 0, \text{ and } s_i = 0 \text{ otherwise;}$$

and an equation explaining the intensity of their R&D

$$(I.2) r_i = r_i^* = x_{1i}b_1 + u_{1i}, \text{ if } s_i = 1, \text{ and } r_i = 0 \text{ otherwise;}$$

where u_{0i} and u_{1i} follow a bivariate normal distribution with correlation coefficient ρ , and standard errors 1 (for reasons of identification) and σ_1 respectively. Equation (I.1) is a probit equation on R&D, determining a latent variable s_i^* that defines the threshold above

which one is an R&D performer. Equation (2) determines the intensity of R&D for R&D performers. The two equations form a generalized tobit model.

Block II explains the determinants of innovation output, which is split into process and product innovation. Process innovation is measured by a dichotomous variable

$$(II. 1) \text{pc}_i = 1 \text{ if } \text{pc}_i^* = x_{2i}b_2 + u_{2i} > 0, \text{ and } \text{pc}_i = 0 \text{ otherwise}$$

where u_{2i} follows a standard (for reasons of identification) normal distribution. Equation (II. 1) corresponds to a probit. Product innovation is measured either by the share of innovative sales

$$(II. 2) \text{zinno}_i = x_{3i}b_3 + u_{3i}$$

where u_{3i} follows a normal distribution with mean 0 and standard error σ_3 respectively, and where zinno_i is the logit transformation of the share of innovative sales, $\text{zinno}_i = \log[\text{inno}_i / (1 - \text{inno}_i)]^5$; or by the fact that a patent has been applied for in the last three years

$$(II. 3) \text{pat}_i = 1 \text{ if } \text{pat}_i^* = x_{4i}b_4 + u_{4i} > 0, \text{ and } \text{pat}_i = 0 \text{ otherwise}$$

where u_{4i} follows a standard (for reasons of identification) normal distribution. Equation II.2 is again a probit and equation II.3 corresponds to a simple linear regression. The logit transformation is applied so that the dependent variable in equation (4) is defined from $-\infty$ to $+\infty$, just as the assumed error term in this equation.

Block III explains the determinants of labor productivity (prod) measured by gross output divided labor:

$$(III) \ln(\text{prod}_i) = x_{5i}b_5 + u_{5i}$$

where u_{5i} follows a normal distribution with mean 0 and standard error σ_5 .

Each matrix of explanatory variables x_{ki} ($k=1, \dots, 5$) contains some common, general control variables, and some idiosyncratic variables. Identification of each equation will be obtained by using exclusion restrictions. R&D will appear as an argument in the innovation output equations, and the innovation output indicators will appear in the labor productivity equation. We shall also test for a possible direct effect of R&D on labor productivity, which would indicate that R&D influences productivity through other channels than innovation output. At first, we shall not allow for feedback effects of productivity on innovation or from innovation output on innovation input.

⁵ Because of the logit transformation on the shares of innovative sales, shares lower than 1% and or higher than 99% for innovators were replaced by 1% and 99% respectively. The same was done with the various components of innovative sales. For turnover from new products and turnover from improved products, which both sum to turnover from innovative sales, the lower bound was fixed at 0.5%.

We have a system of simultaneous equations with a recursive structure. The parameters of the structural form are estimated by the method of asymptotic least squares (ALS). In a first step, we get consistent estimates of the reduced form equations by maximum likelihood. The log-likelihood function is given by:

$$(7) \ln L = \sum_i \{ 0_{\text{RDCON},i} \ln \Phi(-z_{0i}\pi_0) + 1_{\text{RDCON},i} [-\ln(\xi_1) + \ln \phi((r_i - z_{1i}\pi_1)/\xi_1) + \ln \Phi([z_{0i}\pi_0 + \rho(r_i - z_{1i}\pi_1)/\xi_1]/(1-\rho^2)^{0.5})] + 0_{\text{PC},i} \ln \Phi(-z_{2i}\pi_2) + 1_{\text{PC},i} \ln \Phi(z_{2i}\pi_2) + \text{ipat}_i [0_{\text{PAT},i} \ln \Phi(-z_{3i}\pi_3) + 1_{\text{PAT},i} \ln \Phi(z_{3i}\pi_3)] + (1 - \text{ipat}_i) [-\ln(\xi_4) + \ln \phi((z_{\text{inno}i} - z_{4i}\pi_4)/\xi_4)] - \ln(\xi_5) + \ln \phi((\text{prod}_i - z_{5i}\pi_5)/\xi_5) \},$$

where ipat is a dichotomous variable taking the value one when the patenting equation is estimated.

The first two lines of (7) correspond to the log-likelihood function of the generalized tobit, the next two lines correspond to the log-likelihood function of a probit, and the last two lines to the log-likelihood function of an ordinary least squares estimation. The z_{ki} ($k=1, \dots, 5$) are the regressor matrices of the reduced form version of the model. If all explanatory variables appear in the first two equations $z_{0i}=z_{1i}=z_{2i}=z_{3i}=z_{4i}=z_{5i}$.

The structural form parameters are given by $b = [b_0, b_1, b_2, b_3, b_4, b_5, \sigma_1, \sigma_3, \sigma_5, \text{ and } \rho]$. The reduced form parameters are given by $\pi = [\pi_0, \pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \xi_1, \xi_3, \xi_5, \text{ and } \rho]$. There is a relationship between the reduced form and structural form parameters $g(b, \pi) = 0$, one for each reduced form parameter, hence g has the dimension of the number of reduced form parameters. The idea of ALS is to choose b so as to minimize

$$g(b, \hat{\pi})' \left[\frac{\partial g(\cdot)}{\partial \pi'} \hat{\Omega} \frac{\partial g'(\cdot)}{\partial \pi} \right]^{-1} g(b, \hat{\pi}) \text{ where } \hat{\Omega} \text{ is the estimated covariance matrix of the}$$

estimated π ($\hat{\pi}$). When there are as many reduced form as structural form parameters, it is possible to reduce the quadratic form to zero. When there are overidentifying restrictions, the distance is not necessarily equal to zero. A Sargan test of the overidentifying restrictions consists in checking whether the quadratic form is significantly different from zero. The test statistic is the quadratic form, which is distributed as a Chi-square with degrees of freedom equal to the number of overidentifying restrictions.

4. SPECIFICATION OF THE MODEL

R&D is defined as intramural plus extramural R&D. As an alternative we shall test the sensitivity of our results to the use of innovation expenditures as a measure of innovation input. Those are measured as the sum of the expenditures connected to product or process innovations for intramural and extramural R&D, the acquisition of machinery and equipment, the acquisition of other external technology, industrial design and other production preparations, training, and the market introduction of technological innovations. As size is measured in terms of employment, the intensity of R&D or

innovation expenditures is also measured with respect to employment (rather than sales) to avoid a spurious correlation with labor productivity due to measurement errors in sales. Innovation output will be measured by the share of innovative sales, i.e. the share in total sales of technologically new or improved products introduced between 1994 and 1996 - where new can mean new for the firm or new to the market - when it comes to product innovations, and by the fact of being or not a process innovator, when it comes to process innovations. We also experiment with the dichotomous variable patenting (“having applied for at least one patent in the last three years”) as an alternative measure of product innovation. Labor productivity is measured by the ratio of turnover per employee in 1996 (in logarithms). The absence of capital stock data precludes the use of a total factor productivity measure.

In each equation we control for industry effects (captured by industry dummy variables), country effects (also captured by dummies), the fact that the enterprise belongs to a group, the size of the enterprise (measured by the number of employees in logarithms), and the occurrence of government support for innovation. The other variables that we introduce in the model are a demand-pull effect, a cost-push effect, the proximity to basic research (proxied by the occurrence of cooperation with universities or other higher education institutes, or with government or private non-profit research institutes), and the importance of various sources of information leading to the innovation (the enterprise itself, clients, suppliers and patent disclosures). Demand pull is measured by a dummy variable taking the value one when the four objectives of innovation related to demand (“replace products being phased out”, “improving product quality”, extend product range”, and “open up new markets or increase market share”) receive on average a score greater than 2. Cost push is measured by a dummy variable taking the value one when the four objectives of innovation related to cost (“improve production flexibility”, “reduce labour costs”, “reduce materials consumption”, “reduce energy consumption”) receive a mean score greater than 1.5. The cut-off points for demand pull and cost push were chosen so as to cut the sample roughly in two classes of equal size. The various sources of information are constructed as dummy variables taking the value one when their individual score is above the median response for all innovating firms in our sample. For suppliers, the variable can also take the value 1 when cooperation with suppliers is reported.

Our strategy is to put as much structure as possible on the model, that is to introduce these additional explanatory variables in the R&D equations only and then let them affect the innovation output equations via R&D and labor productivity via the innovation outputs. Hence they operate indirectly through the R&D-innovation propagation. To identify the innovation process from the innovation product equation we introduce an equation-specific demand-pull effect for product innovation and an equation-specific cost-push effect for process innovation. On the basis of some initial specification runs which showed that some parameter estimates were insignificant or had a dubious sign we removed demand-pull in the patenting equation, and cost-push and information from suppliers in the R&D equation. Since information from patent disclosures is a parameter of interest, we also introduced it separately in each equation. In the productivity equation, it did not turn out significant and was therefore removed from it. The preferred

specification at this stage, with its identification restrictions and assumed “causality” effects, can be visualized in table 1.

Table 1
Identification and causality structure of the model

Determinants	Continuous R&D performers	R&D intensity	Process innovation	Product innovation/ patent	Labor productivity
Industry effects	x	x	x	x	X
Country effects	x	x	x	x	X
Size	x	x	x	x	X
Group belonging	x	x	x	x	X
R&D-intensity			x	x	(x)
Process innovation					X
Product innovation					x
Government support	x	x	x	x	x
Demand pull	x	x		(x)	
Cost push			x		
Proximity. to basic R&D	x	x			
Information from enterpr.	x	x			
Information from clients	x	x			
Information from patents	x	x	x	x	
Information from suppliers			x		

As indicated in table 1,

$x_{com} = \{\text{industry, country, size, group, gov}\}$,

$x_0 = \{x_{com}, dp, basic, sent, scli, spat\}$, $x_1 = \{x_{com}, dp, basic, sent, scli, spat\}$,

$x_2 = \{x_{com}, r, cp, supp, spat\}$, $x_3 = \{x_{com}, r, dp, spat\}$, $x_4 = \{x_{com}, r, dp, spat\}$, $x_5 = \{x_{com}, pc, zinno\}$.

At the outset, it is clear that we are limited with the data at hand. Because we work with micro-aggregated data we are limited in the choice of explanatory variables by those that are contained in the innovation surveys. It would be better to have data on the number of patents, but unfortunately we only have information on the occurrence of patents in 1994-1996. We also lack data on capital stock and therefore our productivity equation is somewhat poorly specified.

5. THE DATA

We use the microaggregated data from CIS2 for France, Germany, Spain and the United Kingdom. At first, we only examine the manufacturing firm data of the scientific sectors, i.e. chemicals (NACE 23-24), machinery (NACE 29), electrical products (NACE 30-33), and transportation equipment (NACE 34-35). For the UK, pounds were converted to Euros for turnover and R&D expenditures using the annual exchange rate for 1997 from the Bank of England (1.45€/£).

We proceed to some data cleaning to eliminate outliers related to size, labor productivity, R&D intensity and growth in employment, production or labor productivity (see appendix 1). Missing values for the three questions defining an innovator and for the explanatory variables regarding innovators were considered as zero responses.

Table 2 gives some descriptive statistics of our sample of R&D intensive firms. In Germany and the UK, where the innovation survey was done on a voluntary basis, we have much fewer observations than in France and Spain.⁶ The average size of French enterprises is twice the average size of Spanish enterprises. On average 62 % of innovative firms are clean R&D performers; that percentage does not differ much across countries. The R&D intensity is much higher in France, be it expressed in terms of R&D/sales or in percentages of R&D personnel over total personnel. These R&D intensity figures are somewhat higher than what we are used to, but it should be kept in mind that it concerns innovative R&D performing enterprises. The innovation output indicators clearly show that R&D performing firms are more innovative than innovating firms in general. German enterprises stand out as the most innovative in patenting and product innovations, whereas Spanish firms dominate in terms of process innovations. The Spanish innovation survey did not carry the question on patenting. Average labor productivity is also higher for R&D innovating firms than innovating firms in general. French firms are slightly more productive. Of course, all these comparisons only make sense if we can consider that the variables are measured the same way in all countries. There is some variation in the structural composition of output between the four countries and hence from the technological opportunity argument we would expect some differences in innovation across the four countries. French enterprises are more often part of a group than the others, especially the German firms. More Spanish and German firms report benefiting from government support for innovation than French and UK firms. Demand pull seems to be stronger for the Spanish firms and cost push stronger for the German firms. The propensities of firms measured as close to basic research is not very different across the four countries. Finally there are some international differences in the perceived sources of information for innovation. For instance, information from patent disclosures is more frequently reported by German firms than elsewhere.

In table 3, we compare the number of employees in our samples with the number of employees from the national accounts figures. In France and Spain our samples cover roughly half of the respective populations, but in the UK and Germany the coverage is

⁶ With some possible differences due to size coverage, the response rates communicated at the 8th EEA working party meeting in Luxembourg on November 1999 for manufacturing were 29% for Germany, 75% for Spain, 85% for France, and 41% for the UK.

rather poor. Only 4% of German firms are represented in our sample of innovative firms and 12.5% in the UK, reflecting their low response rates. Differences in the averages should not affect the estimation results, but well the innovation and productivity decompositions. The average labor productivity from our sample is in line with the national accounts figure in France, and upward biased in the other three countries, as much as by 30% in Spain. We expect a somewhat higher productivity performance since we are dealing with innovating firms. Likewise we expect a higher R&D intensity in our sample of innovating firms. It is therefore surprising to have a lower figure for the UK. Labor productivity shows a lower growth rate in our sample than in the corresponding population for France, but a substantially higher growth rate in the UK. In Spain and Germany the sample based productivity growth is in line with the population based figure.

6. FIRST RESULTS

In table 4 we present the estimation results of our preferred specification of the model so far. Industry dummies, country dummies, size, and belonging to a group are introduced in each equation as control variables. We shall not report on their sign and estimates. Government support, as expected, has a positive and significant influence on the R&D decision. However, apart from its effect of R&D, it has no other direct significant effect of both types of innovation output. Its negative sign on labor productivity suggests that ailing firms get government support. Demand pull has a significant positive effect on product innovations and cost push plays positively on process innovation. Both increase the incentive to do R&D and its intensity. The information from basic research institutions, from other parts of the enterprise, from clients and patent disclosures increase significantly R&D. Information from suppliers does not affect R&D but directly process innovation. Patent disclosures reveal information that significantly boosts product innovation, but not process innovation, which also makes sense since patents are most effective and hence mostly used for product innovations. R&D affects positively and significantly both types of innovation and therefore all variables that bear on R&D also indirectly bear on innovation output. Process innovations do not significantly boost labor productivity but product innovations do.

The parameter estimates are not very meaningful as such. To give them more meaning it is necessary to compute the slope estimates of the variables explanatory variables. But before doing that, we contrast in table 5 the parameter estimates of the endogenous variables and of patent disclosures in three specifications of the model. Model I comprises the R&D decision, one innovation output measure, namely patenting, and labor productivity. Model II adds innovation process to it. Model III replaces patenting by the share of innovative sales. This is our preferred specification reported in table 4. In each case we report estimates with and without a direct effect of R&D on labor productivity. When we compare models I and II we first notice that the estimates are pretty robust to the introduction of an additional equation. There is room for a process innovation equation, that is significant in labor productivity, and that gets significantly

influenced by R&D. However, if we also let R&D play directly on productivity, we notice that process innovation ceases to be significant. Actually we notice in each model that once we introduce R&D in labor productivity, the R&D variable dominates and replaces the influence of the other innovation indicators on productivity. In model III once R&D is introduced in productivity, no innovation indicator is anymore significant. Patent disclosures have a significant impact on patenting but not on product innovation. When product innovation replaces patenting, process innovation ceases to affect significantly productivity. Of the three innovation output indicators, the share of innovative sales seems to be the strongest one.

In table 6, we present the labor productivity decomposition from models II and III. Each cell is expressed in deviation with respect to the average European country, which is a fictitious construct defined as having as values of each variable the average of the 4 European countries (three in the case of model II, where Spain has no observations on patenting). As a result, the sum of the average deviations with respect to average Europe is equal to zero. As we noticed already when reading the descriptive statistics, French R&D-intensive manufacturing firms are more productive than their German, Spanish and English counterparts. If we take the first column in model II, our decomposition states that close to one third of it can be explained by the industrial structure, a quarter by the combined effects of size and group, innovation output explains close to nothing, and more than one third of labor productivity (the residual) cannot be explained. In model III the unexplained residual is even higher as French innovation output is lower the average innovation output in Europe. Germany is more innovative than the average European country, but less productive. Thus there again innovation does not go the way we expected. Spain is close to the average European country regarding labor productivity level. According to its industry structure and innovation it should be more productive. Again a great deal of the productivity difference remains unexplained. Only for the UK, does innovation contribute to half of the explanation of the difference with respect to average Europe.

In table 7, we report the marginal effects of patent disclosures on each of the endogenous variables of the model. The same computations could be done for each other explanatory variable. Firms that find patent disclosures an important source of information have a 30 percentage points higher R&D intensity than the other firms. Patent disclosures have no significant total effect on process innovation, but they affect patenting or the share of innovative sales, in the case of patenting through a direct and indirect effect, in the case of the share of innovative sales only indirectly, i.e. via through its effect on R&D and the effect of R&D on product innovation. In model III firms which find patent disclosures important have in total a 12.8 percentage points higher share of innovative sales, all other things being equal. In model II, patent disclosures tend to be associated with a 4.2 percentage points higher labor productivity because of the effect of patent disclosures on patenting and its marginal effect on productivity, and if we add the indirect effects via R&D on patenting and on process innovation and the effects of those two innovation output indicators on productivity we attribute a 5.3 percentage points higher labor productivity to patent disclosures.

APPENDIX 1: DATA CLEANING

Were deleted observations with:

- Number of employees missing, less than 20, or more than 100000 (225 observations in the UK)
- Logarithm of labor productivity in 1996 roughly more than four times the sample standard error away from the sample mean (10 firms in France, Germany and Spain combined, 7 in the UK).
- Non-scientific sector NACE 37 (15 firms in the UK, 19 in France, Germany and Spain combined)
- R&D/sales or R&D personnel/total personnel equal to 0
- R&D/sales greater than 50% (25 enterprises in France, Germany and Spain combined, none in the UK) or R&D personnel over the number of employees greater than 50% (17 enterprises in France, Germany and Spain combined, 2 in the UK)
- Growth in employment lower than -75% and greater than 150%, growth of production or labor productivity lower than -100% or greater than 200%, provided 1994 and 1996 employment and output figures were available, and the level of labor productivity in 1994 outside the same boundaries as in 1996 (635 enterprises).

INDUSTRY NAME	NACE CODE (rev. 1)	INDUSTRY DEFINITION
High-R&D Sectors		
Vehicles (D10)	34-35	Manufacture of motor vehicles, trailers, semi-trailers, and other transport equipment
Chemicals (D4)	23-24	Manufacture of coke, refined petroleum products and nuclear fuel, manufacture of chemicals and chemical products
Machinery (D8)	29	Manufacture of machinery and equipment NEC
Electrical (D9)	30-33	Manufacture of office machinery and computers, electrical machinery And apparatus, radio, television and communication equipment and Apparatus, medical, precision and optical instruments, watches and clocks
Low-R&D Sectors		
Food (D)	15-16	Manufacture of food, beverages and tobacco
Textiles (D)	17-19	Manufacture of textiles, wearing apparel, dressing and dyeing of fur, tannings, and dressing of leather, luggage, handbags, saddlery, harness and footwear
Wood (D)	20-22	Manuf of wood and products of wood and cork, except furniture, manufacture of straw and plaiting materials, pulp, paper, and paper products, publishing, printing, and reproduction of recorded media
Rubber, Plastic (D)	25	Manufacture of rubber and plastic products
Non-metallic (D)	26	Manufacture of other non-metallic mineral products
Metals (D)	27-28	Manufacture of basic metals, fabricated metal products, except machinery and equipment
NEC (D)	36	Manufacture of furniture, manufacturing NEC

APPENDIX 2: INDUSTRY DEFINITIONS

REFERENCES

Benavente, J. M. (2002), "The role of research and innovation in promoting productivity in Chile", mimeo.

Crépon, B., E. Duguet and J. Mairesse (1998), "Research and Development, Innovation and Productivity: An Econometric Analysis at the Firm Level", *Economics of Innovation and New Technology*, 7(2), 115-158.

Criscuolo, C. and J. Haskel (2002), "Innovations and productivity growth in the UK", CeRiBa discussion paper.

Duguet, E. (2002), "Innovation height, spillovers and TFP growth at the firm level: Evidence from French manufacturing", EUREQua 2002.73.

Galia, F. and D. Legros (2003), "Research and development, innovation, training, quality and profitability: Econometric evidence from France", mimeo.

Griliches, Z. (1990), "Patent statistics as economic indicators: a survey", *Journal of Economic Literature*, 28(4), 1661-1707.

Griliches, Z. (1994), "Productivity, R&D and the data constraint", *American Economic Review*, 84(1), 1-23.

Griliches, Z. (1995), "R&D and Productivity: Econometric Results and Measurement Issues", in P. Stoneman (ed.), *Handbook of the Economics of Innovation and Technical Change*, Blackwell Handbooks in Economics,

Hall, B. (1995), "The private and social returns to research and development", in *Technology, R&D, and the Economy*, Bruce L.R. Smith and Claude E. Barfield (eds.), The Brookings Institution, Washington D.C.

Hu, A.G. and G.H. Jefferson (2003), "Returns to research and development in Chinese industry: Evidence from enterprises in Beijing", mimeo.

Janz, N., H. Lööf, and B. Peters (2003), "Innovation and productivity: a cross-country comparison between Germany and Sweden", mimeo.

Jefferson, G., B. Huamao, G. Xioajing and Y. Xiaoyun (2002), "R and D Performance in Chinese Industry", mimeo.

Kleinknecht, A. (ed.) (1996), *Determinants of Innovation. The Message from New Indicators*. Macmillan Press, London.

Kleinknecht, A. and P. Mohnen (2002) (eds.), *Innovation and Firm Performance. Econometric Explorations of Survey Data*. Palgrave, Hampshire and New York.

Licht, G. und H. Stahl, "Ergebnisse der Innovationsforschung 1996", ZEW Dokumentation Nr. 97-07.

Lööf, H. and A. Heshmati (2002), "Knowledge capital and performance heterogeneity: A firm-level innovation study", *International Journal of Production Economics*, 76(1), 61-85

Lööf, H. and A. Heshmati (2002), "On the relationship between innovation and performance: A sensitivity analysis", mimeo.

Lööf, H., A. Heshmati, R. Apslund and S.O. Nås (2002), "Innovation and performance in manufacturing firms: A comparison of the Nordic countries", mimeo.

Mairesse, J. and P. Mohnen (1990), "Recherche-développement et productivité: un survol de la littérature économétrique", *Economie et Statistique*, 237-38, 99-108.

Mairesse, J. and P. Mohnen (2001), "To be or not to be innovative : An exercise in measurement", *STI Review. Special Issue on New Science and Technology Indicators*, OECD, 27, 103-129.

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

Mairesse, J. and M. Sassenou (1991), "R&D and productivity: a survey of econometric studies at the firm level", *STI Review*, OECD, 8, 9-46.

Parisi, M.L., F. Schiantarelli and A. Sembenelli (2002), "Productivity creation and absorption, and R&D: Micro evidence for Italy", mimeo.

Van Leeuwen G. (2002), "Linking innovation to productivity growth using two waves of CIS", STI working paper 2002/8, OECD.

Van Leeuwen, G. and L. Klomp (2001), "On the contribution of innovation to multi-factor productivity growth", mimeo

Table 2
 Summary statistics: CIS 2, micro-aggregated data, 1996,
 "R&D-intensive" sectors
 for the sample of innovative firms (INNO)

Variable	Sample	France	Germany	Spain	United Kingdom	Four countries
Number of firms	INNO	1078	330	935	367	2710
Average number of employees	INNO	712.3	464.9	330.5	568.6	531.0
Percentage of R&D firms	INNO	62.3	62.4	63.6	58.6	62.3
R&D/person in 10 ³ euros per person	R&D	9.5	5.2	4.7	5.4	6.8
R&D/sales in %	R&D	6.0	4.1	3.5	3.9	4.6
Patenting in %	INNO	52.6	58.2	n.a	33.5	--
	R&D	67.6	74.8	n.a	47.9	--
Process innovation in %	INNO	72.5	76.4	83.9	63.8	75.7
	R&D	76.6	82.5	85.7	68.8	79.6
Share of innovative sales in %	INNO	28.2	51.0	42.0	31.6	36.2
	R&D	31.1	53.0	45.5	33.3	39.1
Log Average Labor Productivity	INNO	4.96	4.78	4.83	4.77	4.87
	R&D	5.02	4.82	4.91	4.85	4.94

“INNO” indicates the sample of Innovative firms, and “R&D” the nested subsample of R&D expenditures reporting firms only.

Table 2: continued

Variable	Sample	France	Germany	Spain	United Kingdom	Four countries
% of firms by industry	D4 Chem.	24.1	12.7	33.3	16.6	24.9
	D8 Mach.	29.9	45.2	19.0	18.3	26.4
	D9 Elec.	32.7	35.5	26.1	44.1	32.3
	D10 Veh.	13.3	6.6	21.6	21.0	16.4
% of firms belonging to a Group	INNO	75.2	42.7	50.1	63.2	61.0
Government support	INNO	26.3	33.9	35.9	18.8	29.6
Demand pull	INNO	64.1	65.8	70.4	64.6	66.5
Cost push	INNO	46.9	70.6	60.5	58.6	56.1
Proximity to basic research	INNO	27.9	25.8	26.7	26.7	27.1
DSCLI	INNO	68.0	86.7	79.4	60.8	73.2
DSPAT	INNO	50.0	78.8	55.6	51.5	55.6
DSENT	INNO	60.9	71.8	81.2	58.3	68.9
SUPP	INNO	58.7	66.7	52.7	71.7	59.4

Table 3
 Representativeness of CIS 2, micro-aggregated data sample,
 "R&D-intensive" Sectors, 1996

Variable	Sample	France	Germany	Spain	United Kingdom
Total number of employees (in 10 ³)	INNO	767.9 (55.8)	153.4 (4.0)	309.0 (44.5)	208.7 (12.5)
	NACC	1377 (100.0)	3815 (100.0)	695 (100.0)	1667 (100.0)
Average Labor Productivity	INNO	216.0 (-0.3)	156.9 (9.6)	196.4 (32.2)	183.6 (15.0)
	NACC	216.6	143.1	148.6	159.6
R&D/sales in %	INNO	6.1 (35.6)	4.9 (11.4)	2.2 (83.3)	2.5 (-32.4)
	NACC	4.5	4.4	1.2	3.7
Log rate of growth (96/94) of labor (in %)	INNO	1.6	-7.0	0.7	-1.9
	NACC	-0.7	-5.6	2.8	7.2
Log rate of growth (96/94) of productivity (in %) in current prices (gross production)	INNO	7.2	15.8	19.8	17.9
	NACC	10.8	15.8	20.7	6.9

“NACC” corresponds to the four country National Accounts figures (Source: OECD).

Table 4. ALS parameter estimates of model with process innovation and the share of innovative sales as innovation output measures

Explanatory variables	Estimating equations				
	Selection	R&D intensity	Process innov.	Innov. sales	Productivity
Chemicals	-2.19***	0.05	-0.90***	-2.10***	4.82***
Machinery	-2.45***	-0.50**	-0.90***	-0.96***	4.29***
Electrical	-2.26***	0.01	-0.62***	-1.27***	4.22***
transp. Equipm.	-2.69***	-0.57**	-0.55***	-0.99***	4.25***
France	0.14*	0.58***	-0.15	-1.69***	0.14***
Spain	0.18*	-0.20**	0.42***	-0.32	-0.01
UK	0.18*	0.01	-0.40***	-1.11***	0.04
ln(employment)	0.29***	0.01	0.23***	0.12***	0.07***
Group	0.04	0.04	-0.19***	-0.21*	0.20***
government support	0.43***	0.49***	0.00	-0.25	-0.10***
demand pull	0.24***	0.19***		0.45***	
cost push			0.37***		
proximity to basic research	0.54***	0.32***			
sources of inform: within enterp.	0.47***	0.37***			
sources of inform: clients	0.18***	0.06			
sources of inform: patent discl.	0.38***	0.29***	-0.09	0.04***	
sources of inform: suppliers			0.17***		
ln(R&D/employee)			0.19*	0.76***	
process innovation					0.05
share of innovative sales					0.07***
Standard error of error term		1.06***		2.27***	0.51***
corr. coefficient error terms		0.56***			

* significant at 10%, ** significant at 5%, *** significant at 1%

Table 5. Parameter estimates of patent disclosures and endogenous variables for alternative specifications

Explanatory variable	Estimating equation				Productivity
	Selection	R&D intensity	-	Patenting	
Model I					
sources of inform: patent discl.	0.46***	0.33**		0.71***	
ln(R&D/employee)				0.84***	
Patenting					0.08***
sources of inform: patent discl.	0.46***	0.33***		0.77***	
ln(R&D/employee)				0.73***	0.25***
Patenting					-0.05
Model II					
sources of inform: patent discl.	0.46***	0.33***	-0.16	0.73***	
ln(R&D/employee)			0.39**	0.82***	
process innovation					0.16**
Patenting					0.06**
sources of inform: patent discl.	0.46***	0.33***	-0.07	0.77***	
ln(R&D/employee)			0.23	0.75***	0.23**
process innovation					0.09
Patenting					-0.05
Model III					
sources of inform: patent discl.	0.38***	0.29***	-0.09	0.04	
ln(R&D/employee)			0.19*	0.76***	
process innovation					0.05
share of innovative sales					0.07***
sources of inform: patent discl.	0.38***	0.29***	-0.09	0.04	
ln(R&D/employee)			0.19*	0.75***	0.00
process innovation					0.05
share of innovative sales					0.07

* significant at 10%, ** significant at 5%, *** significant at 1%

Table 6. Labor productivity decomposition

	France	Germany	Spain	UK	Average
Difference w/t average European country	0.121	-0.058	-	-0.064	0
Model II					
Industry composition	0.039	-0.018	-	-0.021	0
Group and size effects	0.030	-0.032	-	0.002	0
Government support for innovation	0.000	-0.015	-	0.015	0
Patenting	0.003	0.006	-	-0.008	0
Process innovation	0.003	0.009	-	-0.012	0
Productivity residual	0.047	-0.007	-	-0.040	0
Model III					
Difference w/t average European country	0.122	-0.057	0.002	-0.063	0
Industry composition	0.015	-0.043	0.064	-0.036	0
Group and size effects	0.054	-0.031	-0.035	0.012	0
Government support for innovation	0.002	-0.005	-0.007	0.010	0
Share of innovative sales	-0.046	0.065	0.013	-0.032	0
Process innovation	-0.001	0.001	0.004	-0.005	0
Productivity residual	0.097	-0.044	-0.041	-0.012	0

Table 7. Marginal direct and indirect effects of information from patent disclosures, evaluated at European average

	Model II			Model III		
	estimate	S.E.	t-stat	Estimate	S.E.	t-stat
On labor productivity						
Indirect effect through process innovation	-0.026	0.017	-1.49	-0.004	0.005	-0.72
Indirect effect through product innov/patent.	0.042	0.020	2.12	0.003	0.009	0.30
Indirect effect through R&D on process inno.	0.021	0.012	1.76	0.003	0.003	0.81
“ through R&D on product inno./patent.	0.016	0.008	1.80	0.016	0.008	2.02
Total effect	0.053	0.025	2.15	0.018	0.010	1.74
On product innovation/patenting						
Direct effect	0.291	0.048	6.06	0.019	0.063	0.30
Indirect effect through R&D	0.109	0.036	3.02	0.109	0.039	2.83
Total effect	0.401	0.030	13.50	0.128	0.048	2.64
On process innovation						
Direct effect	0.038	0.022	1.66	-0.025	0.022	-1.14
Indirect effect through R&D	-0.031	0.014	-2.21	0.017	0.010	1.58
Total effect	0.007	0.017	0.40	-0.09	0.018	-0.49
On R&D intensity						
Direct effect	0.333	0.080	4.14	0.289	0.061	4.76

