Discussion Papers on Entrepreneurship, Growth and Public Policy

MAX-PLANCK-GESELLSCHAFT

0504

Internal and External R & D: A Sample Selection Approach

by

Claudio A. Piga

Marco Vivarelli

Number of Pages: 28

The *Papers on Entrepreneurship, Growth and Public Policy* are edited by the Group Entrepreneurship, Growth and Public Policy, MPI Jena.
For editorial correspondence,
please contact: egppapers@mpiew-jena.mpg.de

ISSN 1613-8333

Max Planck Institute for
Research into Economic Systems
Group Entrepreneurship, Growth and Public Policy
Kahlaische Str. 10
07745 Jena, Germany
Fax: ++49-3641-686710

*Internal and External R&D: A Sample Selection Approach**

Claudio A. Piga

Nottingham University Business School and University of Sassari

Marco Vivarelli

Catholic University, Piacenza; International Labour Office, Geneva; and IZA, Bonn

December 2003

Abstract

This study explicitly takes into account that the decision to enter into an external R&D relationship is related to an antecedent decision to carry out R&D. This calls for a methodological approach that, at the same time, permits the joint analysis of the determinants of the two decisions and corrects for the sample selectivity that is intrinsic in the analysis of external R&D. Based on a sample of Italian manufacturing firms, the results confirm the need to consider explicitly the selectivity issue in the empirical analysis of external R&D. Another contribution made by this study is the managerial implications that can be derived from its empirical model, which provide a better understanding of the factors determining a firm's decision to engage in R&D both independently and with external partners.

Keywords: Cooperative R&D, vertical relationships, R&D management, bivariate probit, selectivity.
JEL Class.: L2, O32

Corresponding author:
Claudio Piga
Nottingham University Business School
Wollaton Road
Nottingham NG8 1BB
England, UK
claudio.piga@nottingham.ac.uk
Tel.: 0044-(0)115-9515484, Fax: 0044-(0)115-846 6667

* We are very grateful to the editor, Jonathan Temple, and an anonymous referee of this journal for extremely helpful suggestions. An earlier version of this paper was presented at the 2002 EARIE conference in Madrid. We thank participants for their comments. The usual disclaimer applies.

1. Introduction

While innovation seems to increase a firm's performance, not all firms engage in R&D activities and – among the innovative firms – not all of them engage in external R&D. This paper addresses the question of determining which factors facilitate internal and external R&D. It deals with these two related issues by emphasising some of their peculiarities that are of interest both from a methodological and an empirical point of view.

As far as the methodological perspective is concerned, this study discusses two aspects that may play an important role in the empirical analysis of the determinants of an innovative firm's decision to engage in external R&D.

First, we argue that it may not be appropriate to analyse the decision to carry out external innovative activity using a single equation framework, since such a decision is related to the antecedent decision to engage in R&D. This observation is consistent with the important methodological result obtained by Colombo and Garrone (1996), pointing out that the determinants of internal R&D and of cooperative technological agreements should be jointly modelled, since neither of the two decisions can be treated as exogenous in the estimation of the parameters of the other. This implies that the two binary decisions need to be studied jointly using a Bivariate Probit framework, i.e. we take into consideration the possibility that the disturbances in the external R&D equation are correlated with those in the R&D equation, and therefore we assume the existence of unobservable characteristics that affect both decisions under study.[1] In this paper, we study the two decisions jointly by estimating a Bivariate Probit model using a sample of Italian manufacturing firms.

Second, the Bivariate Probit approach is particularly useful because it allows us to take the issue of sample selectivity in the external R&D decision into account. Indeed, the sample used in the analysis of the external research activity is not randomly

[1] The use of separate Probit regressions is appropriate only under the hypothesis that such disturbances are uncorrelated (Greene, 2003, Section 21.6). Early literature overlooked this methodological issue: for instance, Kleinchnecht and Rejinen (1992) study the determinants of R&D cooperation in a sample of Dutch firms by confining their analysis to less than 50% of the firms which reported any R&D activity, thereby disregarding any link between the behaviour of the firms in the two subsamples. Similarly, Cassiman and Veugelers (2002a) estimate a Probit model of the same decision on the subsample of the innovative firms that constitute 60% of the total sample of Belgian firms at their disposal. See also, *inter alia*, Audretsch *et al.* (1996), Bayona *et al.* (2001) and Fritsch and Lukas (2001) for other articles that considered a single equation setting. Other papers, instead, explicitly take into account the simultaneity issue: see Veugelers, 1997; Cassiman and Veugelers, 2002b; Kaiser, 2002.

selected, but depends on the decision to conduct R&D activities.[2] This selectivity issue calls for a further refinement of the Bivariate Probit model which we used in this paper (see Section 5).[3]

Taking into account the methodological issues discussed above, in this paper we find evidence suggesting that biases arise both as a consequence of the single equation approach, and from disregarding the selectivity issue (see results in Section 6).

With regards to empirical findings, the results of this study lend support to the new hypotheses put forward. First, as far as the decision to engage in R&D is concerned, the following determinants seem to make innovative behaviour more likely: access to long-term debt, product diversification, being part of a business group in controlling other firms, not being dependent on a small number of customers, having large firms as main competitors. Second, among the possible determinants of external R&D with other firms, a significant positive role has been found for ownership concentration, an innovation strategy focused on multiple objectives (both process and product innovation) and the presence of outsourcing agreements. Among the possible determinants facilitating external R&D with universities and research labs, public ownership emerges as a key factor. Finally, the presence of a subsidy may be important in fostering external R&D in general (both with other firms and with universities and research centres).

The paper is organised as follows: Section 2 describes the characteristics of the dataset used, while the model is developed in Sections 3 and 4, where results from previous literature are also discussed and some descriptive statistics are provided. The methodological features of the bivariate probit model with sample selection are analysed in Section 5. The main findings are reported in Section 6, which is followed by some concluding remarks.

2. Data

All the variables in this study originate from a survey conducted in 1998, where both balance sheet data and questionnaire replies were gathered. The questionnaire was prepared by an Italian investment bank, Mediocredito Centrale (see www.mcc.it), the

[2] For instance, Veugelers and Cassiman (1999) acknowledge that they never observe firms cooperating while not performing any in-house R&D.

[3] Such a methodology is used in Greene (1998) to study the probability of loan default in the credit card market and in Montmarquette *et al*. (2001) to study the determinants of university dropout rates.

unit of observation being the firm. For each firm, more than 500 variables were included, with balance sheet data for up to 9 years (1989-1997) relating to 4,495 firms with more than 10 employees. The procedures for data collection were mixed: a sampling procedure was adopted for firms hiring less than 500 employees. The stratification was in accordance with size, industry and location. The sample dimension for each stratum was determined according to Neyman's formula, so as to allow rescaling to the universe at the level of each administrative geographical region. For firms with more than 500 employees, the survey covered the entire universe. Overall, the survey generated a sample which is considered to be representative of the Italian manufacturing industry (see Mediocredito Centrale, 1999).

The survey design considered three types of data: 1) balance sheet data for the 1989-1997 period; 2) data related to measurable company characteristics for the 1995-1997 period (i.e. employment, investment and R&D outlays etc.); 3) questionnaire data regarding firm's relationship with customers and suppliers, composition of sales, competitive environment, group membership and position within the group, industry characteristics, ownership concentration, and other qualitative information.

3. Dependent Variables

As mentioned in the Introduction, the purpose of this paper is to explain both the decision to engage in R&D and the subsequent decision to engage in external R&D together with other firms or other institutions. In this section we introduce our dependent variables and we also refer to the previous empirical literature on the subject. In this and the following section the discussion is organised under subheadings introducing the single variables, giving the corresponding acronyms in parentheses.

*The choice to engage in R&D ("**R&D**").* The dependent variables under study are all binary. The first, denoted as "R&D", indicates whether a firm has declared engaging in any R&D activity during the 1995-1997 period or not. By R&D activity we mean any in-house or external (or a combination of the two) research undertaken by the firm.

*The choice to engage in external R&D ("**Tot.Ext.**").* If a firm has indeed declared engaging in R&D, then the second dependent variable, denoted as "Tot.Ext.", is observed. It takes the value of 1 if the firm declares carrying out its R&D projects using the research facilities of other external organisations, such as universities, specialised research centres or other firms, and zero if the innovative activity was carried out using exclusively internal facilities.

"Tot.Ext." can be used to shed some light on the determinants that lead a firm to outsource, at least partly, the execution of various parts of an innovative project. It is noteworthy that "Tot.Ext." mostly represents the decision between full integration, i.e. a value of zero, and tapered integration, i.e. the value of 1, whereby the firm organises its innovation both by "Making" (internal research structures) and "Buying" (external ones) (Veugelers and Cassiman, 1999). Indeed, only a minority of firms in our sample (8.6% of firms with positive R&D outlays) entirely delegated their R&D projects to an external organization, that is, adopted a pure "Buy" strategy, while the majority of innovating firms (53%) used exclusively internal facilities, that is, resorted to a pure "Make" strategy. In the remainder of the paper "Tot.Ext" defines the decision to conduct external R&D, without distinguishing whether it involves joint research efforts between the firm and its external partner or simply a form of contracting out (Veugelers and Cassiman, 1999; Audretsch *et al*., 1996). Such a limitation, which is due to the lack of data on the reasons that led the firm to use external facilities, is not particularly relevant in the present case. Indeed, the regressors used to explain the recourse to external partners are not associated to a particular form of external research activity (i.e., outsourcing vs. cooperation) but rather, explain why external R&D may be a cost-effective solution.

*The choice of R&D partner: other firms ("**Ext.Firms**") or universities and research centres ("**Ext.Unice**").* Although the emphasis of this paper is on the effects of the sample selection mechanism that the construction of "Tot.Ext" entails (see Section 5), we also verify if differences exist in the factors that explain recourse to the various contracting organisations available to the firm (Veugelers and Cassiman, 2003; Belderbos *et al.*, 2003). Thus, in the econometric model, two further different dependent variables allow the refinement of the analysis of a firm's external R&D activity. The dependent variables "Ext. Firms" and "Ext.Unice" report a value of 1 when firms have used the research facilities of – respectively - other firms, and universities or specialized research centre; and zero otherwise. A similar approach is made in Fritsch and Lukas (2001), where a distinction is made between cooperation with customers, suppliers, other firms and publicly-funded research institutions.[4]

[4] This study finds evidence supporting the notion that the propensity to cooperate with different kinds of partners is driven by similar factors.

4. The explanatory variables

This section introduces and describes the regressors used in this study and the main rationales underlying their adoption. We begin by introducing the drivers of external research activity, then we will move on to the determinants of the decision to engage in R&D, and finally to some descriptive statistics useful in the characterisation of our samples.

4.1. The determinants of external R&D

*Diversification in R&D strategy ("**R&D_Diver**").* The extent to which a firm focuses its innovative efforts on pursuing a specific objective, as opposed to being involved in reaching a number of different goals, is captured by the variable "R&D_Diver", obtained using the formula $\sqrt{\sum_{i=1}^{2} r_i^2}$, where r_i are the shares of the R&D budget used for innovating processes and products.[5] It is assumed that the complexity entailed by a more diversified innovative strategy aimed at combining both process and product R&D, as opposed to a more focused one, is likely to induce a firm to seek the services of one or more external partners (Kaiser, 2002). Furthermore, note that the maximum value of "R&D_Diver" is 100, when the firm pursues only a single strategy, either process or product R&D, while its minimum value is 70.71 when the total R&D budget is equally shared between the two strategies. Thus a negative coefficient for "R&D_Diver" is expected.

*Public ownership ("**State**").* The variable "State" represents the share of the responding firm's ownership held by the state. This variable is included to test whether the participation in the ownership structure of a State-run holding or institution facilitates the recourse to external sources such as universities, which in Italy are publicly funded, or public research centres. A positive and significant coefficient for this variable would suggest the beneficial role that state ownership plays in supporting the establishment of links between different innovative public organisations.

*Ownership concentration ("**Herf_Own**").* The data set contains three variables detailing the ownership shares of the three largest shareholders (or owners if the firm's capital is not divided in shares) that exercise direct control over the firm. The variable

[5] Note that these shares of R&D expenditures include both the introduction of new processes and products and the enhancement of existing ones.

"Herf_Own" is obtained by taking the square root of the sum of the squared values of these variables, and dividing it by 100; thus, high values of "Herf_Own" indicate a more concentrated ownership. While Holderness and Sheehan (1988) provide empirical evidence that concentrated shareholdings are associated with higher levels of R&D, the relationship between ownership concentration and the decision to do R&D jointly with other organisations is largely unexplored. A traditional *agency* argument could be usefully applied to the analysis of such a relationship: a more concentrated ownership may imply that the incentives to behave opportunistically are reduced as cheating engenders a greater loss if the relationship is terminated. Therefore, ownership concentration can work as a credible signal that induces compliance among the partners; hence, a higher ownership concentration should be positively related with the likelihood to engage in external R&D.

*Public subsidies ("**Subsidy**").* Kleinknecht and Reijnen (1992) find that the use of various types of government facilities for the promotion of innovation, such as credits and subsidies for R&D activities, seems to increase the probability that firms cooperate in R&D.[6] Accordingly, we expect a positive coefficient for the dummy "Subsidy".

*R&D intensity ("**Intensity**").* The theoretical literature has shown that when the level of exogenous spillovers is sufficiently high, cooperation in R&D is associated with higher levels of R&D expenditures than in the competitive case (D'Aspremont and Jacquemin, 1988; Kamien *et al*., 1992). Such a result also holds both when spillovers are endogenized (Poyago-Theotoky, 1999; Kamien and Zang, 2000;) and when asymmetric spillovers occur (see Amir and Wooders, 1999). To sum up, the game theory literature suggests a positive correlation between R&D cooperation and R&D intensity, measured here as the average of total R&D expenditures per employee in the period 1995-1997 ("Intensity"). While the theoretical predictions point in one specific direction, the empirical evidence is mixed. For instance, Kleinknecht and Reijnen (1992) do not find compelling evidence supporting the notion that R&D intensity enhances the propensity of Dutch firms towards cooperation, while Arora and Gambardella (1990) document that the number of agreements concluded by a sample of US chemical and pharmaceutical companies is positively correlated with R&D intensity. Both Fritsch and Lukas (2001) and Belderbos *et al.* (2003) found that firms engaged in R&D cooperation tend to have a higher share of R&D employees.

[6] Our data set provides information on whether the firm obtained any financial subsidy for applied research and technological innovation in the period 1995-1997 via the Italian National law N. 46/82. It is noteworthy that such a law does not specifically require the applicants to engage in innovative activities jointly with other partners.

Still on the subject of the effects of R&D intensity, in empirical work it is important to recognise that the decision to enter into an external research relationship and the decision regarding the R&D budget (both total and internal) are simultaneously determined. That is, exogenous changes in the economic environment may alter the propensity to conduct external research and, at the same time, affect R&D expenditures (Veugelers, 1997; Kaiser, 2002). This endogeneity problem is likely to be responsible for the observed correlation in some of the previous studies. We tackle it by using a two-stage procedure – also used in Guiso (1998), Dasgupta *et al.* (2000) and Cassiman and Veugelers (2002a) - which is further described in the methodological Section 5.

*Outsourcing in purchases ("**OutsourceP**").* A firm that purchases a large share of inputs and services from subcontractors via an outsourcing agreement rather than relying on market transactions, should find it easier to extend such an approach to the organization of its innovative activity.[7] Thus the variable "OutsourceP", measuring the percentage of total purchases of input goods and services from outsourcing agreements, should be positively related with the likelihood of entering into external research agreements.

*Long and short term debt towards banks ("**LT_Debt**" and "**ST_Debt**").* The efficient allocation of funds to firms can be hindered by information asymmetries. Guiso (1998) argues that information problems are more severe for high-tech firms because outside investors cannot assess the prospects of truly new innovative projects accurately, as past experience can offer them little guidance. On the other hand, the entrepreneurs seeking financing have, if not better information, at least a better perception of the likelihood of success of their innovative projects. The lack of internal financial resources may limit the capacity of a firm to conduct innovation, thereby inducing it to share the cost of research with external organisations. We control for the effects that a firm's debt liabilities have on both its likelihood to undertake R&D and on its decision to rely on external research, by considering the following two variables: the ratio of long- and short-term debt with banks over total assets ("LT_Debt" and "ST_Debt", respectively).[8] Positive values for these variables' coefficients in the "R&D" equation would suggest that innovative firms were facilitated by access to bank loans. Such a result would be reinforced if innovative firms were characterised by a greater proportion of long-term debt, which is offered at a lower interest rate than short-

[7] Levin and Reiss (1988) consider the extent to which upstream materials suppliers and equipment suppliers contribute to expenditures in process R&D (see also Fritsch and Lukas, 2001).
[8] They correspond to the mean values over the years 1993 and 1994, i.e. immediately before the firms' innovative behaviour is observed.

term debt. With regards to the decision to carry out external R&D, it is not clear *a priori* if firms with a good credit rating(i.e., those firms which have access to more long-term debt) are more likely to externalise their research efforts.

*Firm's size ("**Size**").* The Schumpeterian notion that large firms are especially likely both to undertake and be successful in research activities has constituted a constant theme in the literature (Schumpeter, 1943). Such a notion has been challenged from a theoretical point of view (Arrow, 1962) and mixed empirical evidence to support it has been found (Cohen and Levin, 1989; Audretsch, 1991; Audretsch, 1995; Breschi *et al*., 2000). We control for the effect of firm's size both on the likelihood to undertake R&D and on the decision to engage in external research. "Size" is obtained by taking the natural log of the 1994-1995 mean sales.

*Pavitt's taxonomy ("**Pavitt**").* The argument that there may be technological trajectories whose characteristics are common among firms belonging to different economic sectors, led to the aggregation of manufacturing sectors into four categories (Pavitt, 1984): traditional "supplier dominated", "scale-intensive", "specialised equipment suppliers" and "science-based". We constructed a dummy variable that is equal to 1 if a firm falls into the specialised suppliers or science-based clusters, and zero otherwise. We expect its coefficient to be negative, because firms in the traditional and scale-intensive sectors should rely more on innovative strategies based both on the acquisition of innovation embodied in capital goods developed by external suppliers and on receiving information and skills from firms belonging to the other two high-tech categories.

Regional dummies. Recent research has focused on 'regional systems of innovation' (RSI) defined as "the localised network of actors and institutions in the public and private sectors whose activities and interactions generate, import, modify and diffuse new technologies" (Evangelista *et al*., 2002, p.174). To control for regional effects reflecting different technological opportunities available to firms in their geographical area, four dummies for Italian macro-regions have been included.

Sectoral dummies. In order to control for sectoral fixed effects, ten industry dummies have been included in both regressions in all the adopted models.

4.2. Factors affecting the decision to engage in R&D

In this subsection we focus attention on a number of variables that, together with "LT_ Debt", "ST_Debt", "Size" and sectoral dummies are included in the "R&D" regression.

*Control over other firms within a business group ("**Group_Head**").* Various studies have recognised that the group organisation tends to play an important role in promoting and supporting innovation (Filatotchev *et al.*, 2003). Moreover, group organisation facilitates a more rapid diffusion of process technology within SME districts, which are a peculiarity of Italian manufacturing. Thus, the dummy variable "Group_Head", which is equal to 1 if the firm is the holding or controls other firms in the group, aims at capturing the effects that being part of a network of companies engender on the likelihood to engage in innovative activities. A positive coefficient for this variable would indicate that, within the group, the tasks involved in the carrying out of innovative projects are centralised at the level of the holding firm, in line with the findings in Filatotchev *et al*. (2003).

*Product diversification ("**Prod_Diver**")* R&D and economies of scope are closely linked. The new ideas developed in one research project may be of help in another project. Thus a firm with a diversified portfolio of products may be better positioned to determine the general applicability of new ideas than a firm with a narrower portfolio of products, because it can capture the internal knowledge spillovers (Henderson and Cockburn, 1996). We constructed an index of product diversification as follows: $\text{Prod\_Diver} = \frac{1}{\sqrt{\sum_{i=1}^{3} s_i^2}}$, where s_i is the percentage of total sales from product category *i*. Such an index increases with the degree of diversification: the lowest value is obtained when the firm sells only one category of products. Its expected sign is positive.

*Customer concentration ("**Few_Cust**").* To study how being involved in a close-knit vertical relationship affects the decision to conduct R&D, we refer to the Property Rights approach to the theory of the firm (Grossman and Hart, 1986; Hart and Moore, 1990). In particular, we are interested in understanding which party is more likely to invest in the specific asset represented by an innovative project. The theory predicts that the ownership of such an asset will be held by the party that can use it more efficiently, thereby creating the greatest surplus gain. When contracts are incomplete, by holding the residual rights of control over the asset, the owner can determine the use of the asset when there are missing contractual provisions. Consider the extreme case of a supplier that sells all its output to a downstream buyer. Developing innovative equipment reduces the supplier's cost of production, but by investing in such a relationship-specific asset the supplier exposes itself to the risk of being held-up, that is, the buyer can appropriate all the rents generated by the seller's innovative efforts. We should therefore expect the coefficient of “Few_Cust”, which denotes the percentage of total sales to the firm's 3 main clients, to be negative.

*Outsourcing in sales (“**OutsourceS**”).* We also control for other forms of the hold-up problem that may arise in vertical relationships by including the variable “OutsourceS”, which measures the percentage of sales made within outsourcing agreements. It is not clear *a priori* if a firm that operates mainly as a supplier of other firms within an outsourcing agreement will tend to show a higher propensity towards innovation. On the one hand, the presence of many other potential suppliers may provide the firm with the incentive to keep abreast of the latest technological opportunities. On the other hand, if the buyer is locked into the relationship, and cannot easily find substitutes for the firm’s products, then the firm may be induced to slacken its innovative efforts.

*Competition from large firms (“**Big_Comp**”).* Finally, to take into account how competitive pressure affects the firms' innovative behaviour, we use the dummy “Big_Comp”, which is equal to 1 if the firm's main competitors are big firms, indicating more challenging technological competition and so inducing more innovative behaviour.

Table 1 briefly describes the regressors used in this study and their expected signs in the two regressions.

INSERT TABLE 1

4.3. Analysis of the regressors

This subsection is devoted to a descriptive analysis aimed at investigating: 1) the representativeness of the subsamples used in the regression analysis; 2) the complementarity between internal and external R&D; 3) possible occurrence of collinearity problems.

4.3.1. Subsamples' statistical representativeness

Table 2 describes the variables by providing some statistics derived from three different samples. The purpose of Table 2 is to test whether the estimation subsample used in the regressions continues to be representative of the entire sample or is instead biased in one or more variables, because of an unbalanced distribution of missing values. We perform this check by comparing the mean values and the standard deviations of all the relevant variables introduced in the previous sections.

First, we consider the full sample of 4,495 firms, for which some of the observation values may be missing (see column 7). Second, the statistics are worked out using the estimation sample which includes only the observations that have no missing values for all the regressors in both the "R&D" and "External R&D" regressions.[9] A comparison between the characteristics of these first two samples enables us to infer whether missing values may bias our analysis. Third, from the second sample we evaluate the statistics for the subsample of innovative firms with positive R&D outlays.

Table 2 includes our measure of size, to be used for evaluating the characteristics of the various samples. The reported variable was found to be highly correlated with other traditional measures of size, such as number of employees, investment expenditures and total assets; thus significant differences in this variable would cast doubts on the representativeness of the estimation samples relative to the full one. As can be seen, this is not the case in the samples used in this study.

INSERT TABLE 2

Overall, the values reported in Table 2 in the 'Full' and 'Estim' columns are very similar, especially as far as the variables `Size', `LT_Debt, `ST_Debt', `Pavitt', `OutsourceP', `OutsourceS', `Herf_Own', `Prod_Diver' and `Few_Cust' are concerned. This suggests that missing values were randomly distributed, and that therefore the

[9] The difference in the number of observations for the "R&D dummy" (2,620) and the variable for R&D intensity (2,503) is due to the fact that a firm may have declared carrying out some form of R&D, but may not have reported the amount it spent.

observations used to estimate the regressions constitute a representative subsample of all the firms that were originally included in the survey.

Table 2 indicates that the innovative firms are slightly over-represented in the estimation sample (+4.4%). As Table 3 shows, this is due to the inclusion of a larger share of innovative small firms, that is with less than Lit. 4,900 million-worth of sales in 1995.[10] Indeed, this class size includes 33% of innovative firms in the estimation sample, against only 19.1% in the full sample. However, there are at least three reasons why this does not raise any concern regarding potential bias due to missing values. First, as discussed above, the descriptive statistics for all the explanatory variables appear not to differ in the two samples, thereby suggesting that the impact of these innovative firms is negligible, mostly because non-innovating firms continue to make up the great majority even in the estimation sample (62.3%). Second, the increase in the share of small innovative firms appears to be distributed evenly over the whole country, as the third part of Table 3 indicates, thereby suggesting that there was no geographical bias in the distribution of missing values. Third, figures reported in Table 3 appear to be consistent with general empirical evidence emerging from other studies based on larger samples of Italian manufacturing firms. For instance, Evangelista *et al*. (1997) report that in the 22,787 firms constituting the Italian sample of the European Community Innovation Survey, the percentage of innovative firms in the size groups with less than 50 and with 50-100 employees equals 25.9% and 40.8%, respectively. These values are very similar to those reported for the same groups in Table 3 using our estimation sample. Furthermore, no noticeable differences appear to exist in the distribution of external R&D activities, either in terms of size or location.

Turning our attention to the R&D subsample in Table 2, differences are not a cause of concern (in contrast with the previous comparison) but rather, convey some useful descriptive information supporting hypotheses worthy of further investigation in the regression analysis. Apart from the obvious differences concerning “R&D_Diver”, and “Intensity” (these variables are observed only in the case of innovative firms), the notable differences in the R&D subsample compared with the estimation sample, regard the intensity of competition from big firms, the number of holding firms in a group, the state ownership share and the proportion of firms in the science-based and specialized sectors, all of which are greater for the innovative firms. The latter also tend to sell less via outsourcing agreements.

---

[10] The sales groups were constructed by using the values of sales at the 25, 50 and 75 percentile over the full sample.

Table 3 also indicates that R&D intensity tends to increase with size, measured both in terms of sales and employment – in accordance with the "Schumpeterian hypothesis"

INSERT TABLE 3

4.3.2. Complementarity between internal and external R&D

Tables 4 and 5, which were obtained using the subsamples of innovative firms, provide some indication as to the complementarity between internal and external innovative activities, and so lend further support to our methodological choice of studying concurrently the two decisions to engage in R&D and to engage in external R&D.

INSERT TABLE 4

Firms with external links with all types of partners spend, per employee, significantly more than pure "make" firms (i.e., firms with no external relationships, the values being 6.8 and 4.21 respectively). Generally, external research relationships are associated with greater R&D intensity. As complementarity between internal and external R&D implies that doing more of one increases the return on doing more of the other (Veugelers and Cassiman, 1999), Table 4 suggests that firms allocate more funds when they are in a position to take advantage of the higher returns that the combination of internal and external efforts engenders. In other words - as empirically found in Cassiman and Veugelers (2002b) - an adequate internal R&D base may serve as an "absorptive capacity" to taking advantage from the outcomes of external R&D activities (Cohen and Levinthal, 1989). This hypothesis is reinforced by the figures in Table 5, which reports the percentage use of internal research facilities by typology of external links.

INSERT TABLE 5

Obviously, firms with no external relationships have used exclusively internal structures. More interestingly, the 149 firms reporting links with all types of partners made a more intensive use of internal labs (60.35%) than those firms with exclusive relationships with either universities and research centres (55.73%) or other firms (54.8%). Considering that, out of the 474 with external relationships, only 72 firms used exclusively external facilities for research while 276 firms conducted the majority of their research in internal facilities, we conclude that generally a firm's own R&D is needed to lead and support the external sourcing effort.[11]

4.3.3. Correlation matrix

The linear correlation analysis among the regressors is reported in Table 6, showing that – with one exception - the value for the correlation between two regressors is not greater than 0.29 between "Pavitt" and the fitted values of "Intensity". This suggests that no collinearity exists between the regressors (the exception is the expected correlation between "Size" and "Group_Head" (0.40)).

INSERT TABLE 6

5. Methodology

In this section we briefly discuss the methodological foundations of our econometric specification. In the bivariate probit model with censoring setting, data on y_1 may be observed only when another variable, y_2, is equal to 1.

Formally, the model is as follows:

(1)
$$y^*_{i1} = \beta'_1 x_{i1} + \varepsilon_{i1},\ y_{i1} = 1 \text{ if } y^*_{i1} > 0,\ 0 \text{ otherwise}$$
$$y^*_{i2} = \beta'_2 x_{i2} + \varepsilon_{i2},\ y_{i2} = 1 \text{ if } y^*_{i2} > 0,\ 0 \text{ otherwise}$$
$$(\varepsilon_1, \varepsilon_2) \sim \text{BVN}(0,0,1,1,\rho)$$
$$(y_{i1}, x_{i1}) \text{ is observed only when } y_{i2} = 1$$

Thus, there are three types of observation in the sample with unconditional probabilities that need to be taken into account in the construction of the log-likelihood function:

$$(2)\ L_{ss} = \sum_{y_1=1, y_2=1} \ln\left(\Phi_2\left[\boldsymbol{\beta}'_1 \mathbf{x}_{i1}, \boldsymbol{\beta}'_{i2} \mathbf{x}_2, \rho\right]\right) + \sum_{y_1=0, y_2=1} \ln\left(\Phi_2\left[-\boldsymbol{\beta}'_1 \mathbf{x}_{i1}, \boldsymbol{\beta}'_2 \mathbf{x}_{i2}, -\rho\right]\right) + \sum_{y_2=0} \ln\left(1 - \Phi\left[\boldsymbol{\beta}'_2 \mathbf{x}_{i2}\right]\right)$$

where Φ_2 denotes the bivariate normal cumulative distribution function with $\rho = Cov[\varepsilon_{i1}, \varepsilon_{i2}]$. Eq. (2) has to be maximised with respect to the parameters β_1, β_2 and ρ.[12]

[11] This conclusion is consistent with empirical findings in Veugelers (1997).

[12] See Greene (2003, pp. 710-712) for technical details regarding the calculation of the maximum likelihood estimates.

It is noteworthy that the methodology subsumed in (2) differs from the two-step - Heckit - procedure by Heckman (1979) for the case of a continuous dependent variable. Indeed, here the use of a Full Information Maximum Likelihood approach to maximise (2) does not imply the calculation of the Inverse Mill's Ratio (IMR), so that in practice the sample selection problem is not dealt with as one of an omitted variable.[13] As (1) indicates, sample selection arises because the observation of y_1 is not random but conditional on the observation of $y_2 = 1$. More specifically, note that when no correction for selection is needed, we have the standard bivariate probit model. In this case, the likelihood function in (2) would take into account the combination of outcomes that are not feasible in the selection model, that is, ($y_{i1} = 1$; $y_{i2} = 0$) and ($y_{i1} = 0$; $y_{i2} = 0$).

The interpretation of the correlation between the residuals from the equations in the bivariate probit model shares some similarities with that of the residuals from the two equations in Heckman's approach. In the latter, the coefficient of the IMR in the second-stage equation represents the extent to which the disturbances in the first and second step are independent, and hence, is an indication of the appropriateness of the overall approach. Similarly, when $\rho = 0$, the standard bivariate model can be estimated using independent probit equations.

Finally, as discussed in Section 4.1, one of the regressors, namely R&D intensity, is likely to be correlated with the disturbances of the external R&D equations. Such an endogeneity problem is solved in Guiso (1998), Dasgupta *et al.* (2000) and Cassiman and Veugelers (2002a) by adopting a two-stage, instrumental variable approach where the endogenous variable is replaced in the second stage by its fitted value from a first-stage regression. A further complication in the present case is that R&D intensity is itself subject to selectivity, as it is observed only for those firms that report formal R&D activity. Therefore, the fitted value of "Intensity" is worked out by using a Heckit procedure: the same regressors as specified in section 4.2 are used in the Probit selection equation for R&D, while the regression of the endogenous variable "Intensity" includes as instruments, in addition to the previous ones, the share of export sales in 1997, accumulated earnings up to 1995, the ratio of 1994 intangible assets plus three dummies identifying three Italian macro regions.[14] While the mean value for the fitted variable (4.95) does not appear to differ significantly from that of the original variable reported in the R&D sample in Table 2 (5.1), the linear correlation between the fitted and the original value equals 0.283. However, the robust χ^2 statistics ($\chi^2(34)$ =

[13] This implies that no coefficient for the IMR has to be calculated in the equation of y_2.

117.34) show that the model variables have significant explanatory power, thereby suggesting that the fitted values represent a good proxy of total R&D intensity when purged of the simultaneity effects.

6. Results

Taking into account the issues raised and the variables introduced in the previous Sections 3 and 4, Table 7 reports the estimates from the Bivariate Probit models with sample selection. The sample size, after omitting all the relevant missing values, is reduced to 2,620 for the full sample, which is used for the analysis of "R&D", and to 988 for the subsample of innovative firms used to study external R&D. The three reported models use the same regressors but differ in the dependent variable for external R&D (total external R&D; external R&D with other firms; external R&D with universities and research centres, see Section 3).[15]

INSERT TABLE 7

Before entering a discussion of the empirical findings, it is worthwhile to focus attention firstly on the methodological issue raised in Section 1 and discussed in Section 5. In this regard, it is crucial to investigate the significance of the correlation between the residuals from the equations in the three bivariate probit models. As can be seen, the coefficient ρ in Table 7 is significantly different from zero in the first and second models, thereby validating the methodological approach used in this study. In at least two cases out of three, single equation estimates might generate biased coefficients and incorrect inferences (this is certainly the case for the first model, where Hausman's test significantly rejects the similarity between the estimated coefficients and the coefficients obtained using a single probit approach).

Turning our attention to the empirical results, Table 7 reveals an interesting pattern: some of the explanatory variables that are not significant in the "Tot.Ext." regression become significant in the equation of either "Ext.Firms" or "Ext.Unice".

---

[14] Fitted values were also calculated for those observations reporting a missing value for "Intensity".

[15] Some authors have recently pointed out the possible complementarity between various forms of external R&D activity (Belderbos *et al.*, 2003; Veugelers and Cassiman, 2003). If such is the case, one can cast some doubts on the reliability of the estimates for "Ext. Firms" and "Ext. Unice". However, only a minority of firms within our estimation subsample are engaged in external research with both other firms and universities and research centres (see Tables 4 and 5). At any rate, we checked for potential complementarity between "Ext. Firms" and "Ext. Unice" by running a bivariate Probit model of these two variables, then testing whether the residuals in the two equations were correlated. The results revealed that - at least in our sample and after accounting for the same determinants of the two external activities reported in Table 7 - the two decisions about external activity appear to be unrelated .

More precisely, the regressors that are significant only in the equation of external research with other firms are: "R&D_Diver", whose negative sign indicates that the pursuit of multiple objectives creates an incentive to seek the collaboration of other firms which can contribute complementary skills and assets; "Herf_Own" - characterised by a significant positive coefficient (and a large marginal effect, see Table A1) - suggesting that a concentrated ownership structure signals the firm's willingness to avoid opportunistic behaviour; and "Pavitt", illustrating that the firms in a traditional or scale-intensive sector tend to be more involved in innovative cooperation with other firms (notice both a significant negative coefficient (see Table 7, columns 1 and 3) and a large magnitude of the marginal effect in the first two models (see Table A1, columns 1 and 2).[16] Similarly, the evidence indicates that those firms that have a public firm or institution as a stakeholder tend to be more actively involved only in external links with universities and public research centres ("State" turns out to be significant, although not characterised by a very large marginal effect, see Table A1). On the contrary, obtaining a subsidy and relying on outsourcing agreements significantly enhance the likelihood of conducting external R&D with both types of partners (in particular, the impact of "Subsidy" seems to be particularly important in terms of the magnitude of its marginal effect, see Table A1).

Total R&D expenditure per capita ("Intensity") does not seem to increase a firm's propensity to engage in any form of external R&D. In regressions not included here to save space – (available on request) – we found a positive and highly significant coefficient for the original variable for R&D intensity.[17] These combined results lend support to the notion of joint determination of R&D budgets and external research strategies (see also Section 4.3.2); hence the suspicion that this endogeneity problem may have been responsible for the observed correlation between external R&D and R&D intensity found in previous studies.

Since regional dummies are not significant in eight cases out of nine, our estimation outcomes do not give further support to the theory of the "regional systems of innovation".

Financial structure does not appear to affect the decision to conduct external research significantly, although the estimates point to a tendency for firms with high levels of short-term debt to establish external relationships. Moreover, capital structure appears to be a strong determinant of the propensity to do R&D. As expected, in the

[16] The two latter explanatory variables are also significant in the "Tot.Ext" regression.

"R&D" equation the coefficient for long-term debt is greater and more significant than that for short-term debt, thereby suggesting that long-term credit and consequently lower interest rates are more important in inducing an activity such as R&D investment, whose returns are risky and unevenly distributed over the long run. Furthermore, a change from zero to the mean values of both the short term and long term debt variables, appears to be associated (although almost always non-significantly) with high changes in the probability to engage in external activities, see Table A1.

The other results from the "R&D" regression show that the firms selling a high share of their sales to three main clients are significantly less likely to be involved in the running of innovative projects. This is consistent with the hypothesis that in such circumstances the greatest benefit from innovation would accrue to the buyers. Such a result is reinforced by the negative sign of the coefficient for the "OutsourceS" variable, which is however barely significant.

The highly significant results concerning group organization reveal that innovative activity seems to be centralised within the holding company, rather than decentralised among the subsidiaries.

Product diversification and innovative activity are found to be significantly associated. Indeed, the firms selling only one category of product are less likely to engage in R&D than firms selling a broader range of products.

The competitive environment faced by a firm is found to be an incentive for innovation. Indeed, firms competing with larger firms tend to be more involved in innovative activities than those firms whose main competitors are represented by small and medium enterprises.

Finally, large firms are more likely to engage in formal innovative activities, in accordance with the "Schumpeterian hypothesis".

7. Conclusions

As mentioned in the Introduction, the purpose of this paper was twofold. From the methodological point of view, the study shows the importance of correcting for the selectivity bias arising when the observation of a binary dependent variable depends on the value of another binary dependent variable. To this end, we jointly considered both a firm's decision to conduct R&D with an external partner and its antecedent, i.e. the decision as to whether the firm engages in R&D at all. Finding a significant correlation

[17] The other regressors exhibited qualitatively similar behaviour both when the original variable was used and when it was dropped altogether.

between the residuals from the equations in two out of three estimated models, we validated this methodological approach.

Another contribution made by this study consists in the managerial implications that can be derived from its empirical model, providing a better understanding of the factors driving a firm's approach to innovation. Indeed, while some of the empirical findings support previously obtained results - e.g., that firms operating in the specialised suppliers- and science-based sectors are more likely to conduct their R&D internally, that outsourcing relationships with suppliers enhances a firm's propensity to engage in external R&D or that a large, multiproduct firm which competes with big firms has a stronger incentive to engage in R&D - others shed new light on some relatively unexplored determinants of a firm's innovative behaviour. For instance, our evidence suggests different impacts of certain explanatory variables depending on the type of external partners. Indeed, a firm partly or wholly owned by the state is more likely to engage in external R&D with universities and research centres, but not with other firms. R&D with other firms is more favoured by firms having objectives in the areas of both process and product innovation. Moreover, within a business group, the estimates indicate a tendency to centralise R&D activity at the holding firm level. The impact on the propensity to conduct R&D depending on the extent to which a firm is involved as a seller in a vertical relationship has also been taken into account. It turns out that the firms selling most of their production to a small number of influential buyers are less likely to engage in R&D. Furthermore, the estimates suggest that a firm with a concentrated ownership structure exhibits a greater tendency to seek other firms as partners.

From a policy viewpoint, firms that have received a subsidy for applied research and innovation appear more likely to have external partners, thereby suggesting that such a policy may have beneficial effects both in terms of permitting the implementation of valuable projects and in enhancing a firm's willingness to share its know-how. Finally, it was found that access to long-term credit positively affects a firm's decision to engage in R&D but has no significant bearing on a firm's decision to externalise its research activity.

References

Amir, R. and Wooders, J. (1999). Effects of One-Way Spillovers on Market Shares, Industry Price, Welfare, and R&D Cooperation. *Journal of Economics and Management Strategy,* 8(2):223-249.

Arora, A. and Gambardella, A. (1990). Complementarity and External Linkages: The Strategies of the Large Firms in Biotechnology. *Journal of Industrial Economics,* 38:361-379.

Arrow, K. (1962). Economic Welfare and the Allocation of Resources for Inventions. In Nelson, R., editor, *The Rate and Direction of Inventive Activity.* Princeton University Press, Princeton, NJ.

Audretsch, D. B. (1991). New Firm Survival and the Technological Regime. *Review of Economics and Statistics,* 73:441-450.

Audretsch, D. B. (1995). *Innovation and Industry Evolution.* Mit Press, Cambridge, MA.

Audretsch, D., Menkveld, A., and Thurik, R. (1996). The Decision between Internal and External R&D. *Journal of Institutional and Theoretical Economics,* 152:519-530.

Bayona, C., García-Marco, T., and Huerta, E. (2001), Firms' Motivations for Cooperative R&D: An Empirical Analysis of Spanish Firms, *Research Policy*, 30, 1289-1307.

Belderbos, R., Carree, M. A., Diederen, B., Lokshin, B. and Veugelers, R. (2003). Heterogeneity in R&D Cooperation Strategies, *CEPR Discussion Paper 4021*.

Breschi, S., Malerba, F., and Orsenigo, L. (2000). Technological Regimes and Schumpeterian Patterns of Innovation. *Economic Journal,* 110:388-4 10.

Cassiman, B. and Veugelers, R. (2002a). R&D Cooperation and Spillovers: Some Empirical Evidence from Belgium. *American Economic Review,* 92: 1169-1184.

Cassiman, B. and Veugelers, R. (2002b). Complementarity in the Innovation Strategy: Internal R&D, External Technology Acquisistion and Cooperation, *CEPR Discussion Paper 3284*.

Cohen, W. M. and Levin, R. C. (1989). Innovation and Market Structure. In Schmalensee, R. and Willig, R., editors, *Handbook of Industrial Organization,* pages 1059- 1107. Elsevier Science Publisher, Amsterdam.

Cohen, W. M. and Levinthal, D. A. (1989). Innovation and Learning: The Two Faces of R&D. *Economic Journal*, 99: 569-596.

Colombo, M.G. and Garrone, P. (1996), Technological Cooperative Agreements and Firm's R&D Intensity. A Note on Causality Relations. *Research Policy*, 25: 923-932.

Dasgupta, S., Hettige, H., and Wheeler, D. (2000). What Improves Environmental Compliance? Evidence from Mexican Industry. *Journal of Environmental Economics and Management,* 39:39-66.

D'Aspremont, C. and Jacquemin, A. (1988). Cooperative and Noncooperative R&D in Duopoly with Spillovers. *American Economic Review,* 78:1133-1137.

Evangelista, R., Iammarino, S., Mastrostefano, V., and Silvani, A. (2002). Looking for Regional Systems of Innovation: Evidence from the Italian Innovation Survey. *Regional Studies,* 36:173-186.

Evangelista, H., Perani, G., Rapiti, F., and Archibugi, D. (1997). Nature and Impact of Innovation in Manufacturing Industry: Some Evidence from the Italian Innovation Survey. *Research Policy,* 26:521-536.

Filatotchev, I., Piga, C., and Dyomina, N. (2003). Network Positioning and R&D Activity: A Study of Italian Groups. *R&D Management,* 33:37-48.

Fritsch, M. and Lukas, R. (2001). Who Cooperates on R&D. *Research Policy,* 30:297-312.

Greene, W. H. (1998). Sample Selection in Credit-Scoring Models. *Japan and the World Economy*, 10:299-316

Greene, W. H. (2003). *Econometric Analysis.* Prentice Hall, Upper Saddle River, NJ, 5th edition.

Grossman, S. and Hart, 0. (1986). The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration. *Journal of Political Economy,* 94:691-719.

Guiso, L. (1998). High-tech Firms and Credit Rationing. *Journal of Economic Behavior and Organization,* 35:39-59.

Hart, 0. and Moore, J. (1990). Property Rights and the Nature of the Firm. *Journal of Political Economy,* 98:1119-1158.

Heckman, J. (1979). Sample Selection Bias as a Specification Error. *Econometrica,* 47:153-161.

Henderson, R. M. and Cockburn, I. M. (1996). Scale, Scope and Spillovers: Determinants of Research Productivity in the Pharmaceutical Industry. *Rand Journal of Economics,*27:32-59.

Holderness, C. and Sheehan, D. P. (1988). The Role of Majority Shareholders in Publicly Held Corporations. *Journal of Financial Economics,* 20:317--346.

Kaiser, U. (2002), An Empirical Test of Models Explaining Research Expenditures and Research Cooperation: Evidence for the German Service Sector. *International Journal of Industrial Organization*, 20: 747-774.

Kamien, M. I., Muller, E., and Zang, I. (1992). Research Joint Ventures and R&D Cartels. *American Economic Review,* 82:1293-1306.

Kamien, M. I. and Zang, I. (2000). Meet Me Halfway: Research Joint Ventures and Absorptive Capacity. *International Journal of Industrial Organization,* 18:995-1012.

Kleinknecht, A. and Reijnen, J. (1992). Why Do Firms Cooperate on R&D? An Empirical Study. *Research Policy,* 21:347 360.

Levin, R. C. and Reiss, P. C. (1988). Cost-reducing and Demand-creating R&D with Spillovers. *Rand Journal of Economics,* 19:538-556.

Mediocredito Centrale, Ministero dell'Industria (1999), *Indagine sulle imprese manifatturiere: Settimo rapporto sull'industria italiana e sulla politica industriale*. Mediocredito Centrale, Roma.

Montmarquette, C., Maliseredjian, S., and Houle, R. (2001). The Determinants of University Dropouts: a Bivariate Probability Model with Sample Selection. *Economics of Education Review,* 20:475-484.

Pavitt, K. (1984). Sectoral Patterns of Technical Change: Towards a Taxonomy and a Theory. *Research Policy,* 3:343-373.

Poyago-Theotoky, J. (1999). A Note on Endogenous Spillovers in a Non-Tournament R&D Duopoly. *Review of Industrial Organization,* 15:253-262.

Schumpeter, J. (1943). *Capitalism, Socialism and Democracy.* Unwin University Press, London.

Veugelers, R. (1997). Internal R&D Expenditures and External Technology Sourcing. *Research Policy*, 26: 303-315.

Veugelers, R. and Cassiman, B. (1999). Make and Buy in Innovation Strategy: Evidence from Belgian Manufacturing Firms. *Research Policy,* 28:63-80.

Veugelers, R. and Cassiman, B. (2003). R&D Cooperation between Firms and Universities: Some Empirical Evidence from Belgian Manufacturing, *CEPR Discussion Paper 3951.*

Table 1: The variables and their expected signs

Variables	Description	Ext. R&D	R&D
State	Ownership share held by a state-owned company	+	
Herf_Own	Index of the three largest ownership shares	+	
Subsidy	Dummy =1 if the firm has received a subsidy for applied research and technological innovation	+	
R&D_Diver	Inverse index of diversification in product and process R&D	-	
OutsourceP	% of total purchases of goods and services from outsourcing agreements	+	
Intensity	Index of R&D intensity measured as the average of R&D expenditure per employee in the 95-97 period	+	
Size	Log of mean of sales in 1994 and 1995.		
LT_Debt	Ratio "long-term debt with banks over total assets" in the years 1993 and 1994 - Mean value	+/-	+
ST_Debt	Ratio "short-term debt with banks over total assets" in the years 1993 and 1994 - Mean value	+/-	+
Pavitt	Dummy =1 if firm is in a science-based or specialized suppliers sector, zero if in a traditional or scale-intensive sector	-	+
Prod_Diver	Index of product diversification		+
Group_Head	Dummy =1 if a firm is the holding or controls other firms within a group organization		+
Few_Cust.	% of total sales to the three main clients		-
OutsourceS	% of sales made within outsourcing agreements		+/-
Big_Comp	Dummy =1 if the main competitors are big firms		+

Table 2: Descriptive Statistics for the dependent and independent variables

Variables	Mean			Std. Dev			N		
	Full	Estim	R&D	Full	Estim	R&D	Full	Estim	R&D
Tot. Ext	.16	.18	.48	.37	.38	.50	4495	2620	988
Ext. Firms	.11	.12	.33	.31	.33	.47	4495	2620	988
Ext. Unice	.10	.11	.30	.30	.31	.46	4495	2620	988
R&D	.333	.377	1	.47	.48	0	4488	2620	988
Herf_Own	.67	.68	.71	.21	.22	.23	4256	2620	988
State	.78	1.07	1.67	8.4	9.8	12.0	4493	2620	988
Subsidy	.04	.06	.13	.20	.23	.33	4490	2620	988
R&D_Diver	28.4	32.1	85.2	41.5	42.7	17.5	4488	2620	988
Prod_Diver	1.21	1.23	1.3	.48	.52	.63	4353	2620	988
Intensity	1.49	1.8	5.1	5.0	5.6	8.5	4223	2503	879
LT_Debt	.041	.043	.048	.05	.07	.07	3403	2620	988
ST_Debt	.155	.155	.162	.14	.14	.14	3403	2620	988
Size	9.54	9.55	10.0	1.26	1.24	1.4	3302	2620	988
OutsourceP	15.6	15.4	16.5	28.7	28.0	27.2	4360	2620	988
OutsourceS	28.1	26.6	23.1	42.2	41.4	38.8	4468	2620	988
Few_Cust	35.6	34.6	32.4	25.5	25.0	24.2	3991	2620	988
Group_Head	.11	.14	.22	.31	.34	.42	4495	2620	988
Pavitt	.30	.30	.39	.46	.46	.49	4495	2620	988
Big_Comp	.32	.34	.46	.47	.47	.50	4495	2620	988

Table 3: Total R&D intensity, percentage of innovative firms (over total) and percentage of firms with external R&D (over innovative firms) by size classes (1995 sales in million lire and number of employees) and Italian macro-regions.

Cluster	“Intensity”		”R&D” (%)		”Tot.Ext.” (%	
	Full	Estim	Full	Estim	Full	Estim
Sales<=4900	0.75	0.95	19.1	33.0	43.8	45.3
4901 <Sales<=9568	1.19	1.32	25.5	26.5	50.2	49.7
9569 <Sales<=24411	1.36	1.42	32.3	32.9	48.0	49.3
Sales> 24411	2.7	2.88	54.3	56.8	51.1	47.6
Emp<=50	1.11	1.34	23.1	24.9	49.2	49.9
51 <Emp<= 100	1.53	1.67	37.7	39.4	46.1	46.8
101 <Emp<=250	1.63	1.77	51.2	52.5	48.5	46.3
250 <EMP<=500	2.96	3.3	66.5	67.3	46.9	44.1
Emp>500	4.95	5.1	78.1	79.4	57.8	54.5
North West	1.54	1.82	37.9	43.0	48.3	47.6
North East	1.53	1.99	34.1	38.6	46.6	45.5
Centre	1.98	2.28	30.8	34.7	54.0	52.7
South	.56	.76	20.3	23.5	53.8	54.2
Total	1.49	1.8	33.3	37.7	49.2	48.3

Table 4: Average R& D intensity by typology of external R&D activity for firms declaring R&D (cluster sizes in parentheses)

	Ext. Unice		
Ext. Firms	No	Yes	Total
No	4.21*,† (444)	5.32 (139)	4.48 (583)
Yes	6.07† (159)	6.8* (137)	6.4 (296)
Total	4.7 (603)	6.05 (276)	5.12 (879)

Note:
*** and † indicate couples of cluster means significantly different at 1% and 10% levels respectively (t-test).**

Table 5: Average percentage of use of internal research facilities by typology of external R&D activity for firms declaring R&D (cluster sizes in parentheses)

	Ext. Unice		
Ext. Firms	No	Yes	Total
No	100 (511)	55.73 (147)	90.11 (658)
Yes	54.8 (178)	60.35 (149)	57.3 (327)
Total	88.3 (689)	58.05 (296)	79.22 (985)

Table 6: Pairwise correlation matrix

Variables		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
State	(1)	1.00														
Size	(2)	0.08	1.00													
Herf_Own	(3)	0.10	0.22	1.00												
Subsidy	(4)	0.01	0.21	0.07	1.00											
R&D_Diver	(5)	0.05	0.28	0.09	0.22	1.00										
Intensity*	(6)	0.04	0.12	0.07	0.14	0.18	1.00									
LT_Debt	(7)	0.03	0.12	0.02	0.04	0.06	-0.12	1.00								
ST_Debt	(8)	-0.02	0.16	-0.01	0.03	0.04	-0.03	0.13	1.00							
OutsourceP	(9)	-0.00	-0.01	0.02	0.03	0.02	0.05	-0.00	0.02	1.00						
Pavitt	(10)	-0.02	-0.01	0.04	0.14	0.15	0.29	-0.07	-0.07	0.06	1.00					
Group_Head	(11)	0.05	0.40	0.16	0.22	0.19	0.23	0.05	0.04	0.03	0.04	1.00				
Few_Cust	(12)	0.05	-0.04	0.07	-0.02	-0.05	-0.01	-0.00	-0.07	0.09	0.00	-0.02	1.00			
OutsourceS	(13)	-0.00	-0.09	0.00	-0.02	-0.06	-0.11	0.02	-0.07	0.28	-0.01	-0.02	0.26	1.00		
Prod_Diver	(14)	-0.00	0.10	0.02	0.08	0.09	0.10	0.04	0.03	0.04	0.04	0.07	-0.08	-0.02	1.00	
Big_Comp.	(15)	0.05	0.27	0.09	0.09	0.20	0.12	0.03	-0.02	0.02	0.06	0.10	0.07	-0.01	0.06	1.00

Notes:
*** The variable corresponds to the predicted value from the first stage.**
The dummies for the three macro-regions and the ten sectoral dummies were not reported to save on space. The largest correlation coefficient for one of these dummies and the variables reported in the table equals 0.42 and corresponds to the correlation between the dummy for the chemicals sector and Intensity*.

Table 7 – Bivariate Probit estimates with Sample Selection

Variables	Model 1		Model 2		Model 3	
	Tot. Ext.	**R&D**	**Ext. Firms**	**R&D**	**Ext. Unice**	**R&D**
R&D_Diver (10^{-2})	-.1 (.8)		-.4 (1.9)[c]		-.01 (.1)	
Herf_Own	.39 (2.26)[b]		.47 (2.49)[b]		.08 (.4)	
State (10^{-2})	.3 (.9)		.1 (.3)		.77 (2.28)[b]	
Subsidy	.338 (2.76)[a]		.31 (2.51)[b]		.29 (2.3)[b]	
OutsourceP (10^{-2})	.5 (3.42)[a]		.5 (3.58)[a]		.3 (2.1)[b]	
Intensity* (10^{-2})	-.5 (.27)		-1.36 (.66)		1.8 (.8)	
Pavitt	-.277 (2.4)[b]		-.24 (2.07)[b]		-.12 (1.0)	
North East	-.08 (.56)		-.18 (1.27)		-.21 (1.43)	
Centre	-.15 (1.0)		-.17 (1.06)		-.33 (2.0)[b]	
South	.09 (.51)		-.04 (.21)		-.27 (1.4)	
Size	-.12 (2.4)[b]	.22 (7.9)[a]	-.05 (.82)	.22 (8.0)[a]	.057 (.9)	.22 (7.8)[a]
LT_Debt	-.7 (1.18)	.97 (2.51)[a]	-.72 (1.13)	.97 (2.5)[b]	-.1 (.1)	.94 (2.44)[b]
ST_Debt	.65 (2.28)[b]	.1 (.5)	.38 (1.29)	.09 (.5)	.18 (.6)	.1 (.5)
Prod_Diver		.175 (3.0)[a]		.16 (2.89)[a]		.18 (2.97)[a]
Group_Head		.31 (3.84)[a]		.31 (3.82)[a]		.28 (3.45)[a]
OutsourceS (10^{-2})		-.12 (1.78)[c]		-.11 (1.69)[c]		-.11 (1.72)[c]
Few_Cust (10^{-2})		-.36 (3.14)[a]		-.35 (3.0)[a]		-.38 (3.4)[a]
Big_Comp		.34 (6.0)[a]		.33 (5.8)[a]		.34 (6.1)[a]
Constant	1.41 (2.0)[b]	-2.25 (8.4)[a]	.65 (.8)	-2.2 (8.5)[a]	.32 (.38)	-2.2 (8.4)[a]
Wald Test Industry Dummies	$\chi^2(20)=166.2$[a]		$\chi^2(20)=158.1$[a]		$\chi^2(20)=166.2$[a]	
N	988	2620	988	2620	988	2620
Regressors' Wald Test†	$\chi^2(23)=86.2$[a]		$\chi^2(23)=54.1$[a]		$\chi^2(23)=43.6$[a]	
ρ	-.566 ($\chi^2(1)=3.81$[b]		-.535 ($\chi^2(1)=3.58$[b]		-.423 ($\chi^2(1)=2.1$	
Comparison Test‡	$\chi^2(24)=43.5$[a]		$\chi^2(24)=29.9$		$\chi^2(24)=14.7$	

Notes:

Robust *t*-statistics in parentheses; [a, b, c] Significant at the 1%, 5% and 10 % level respectively.

*** Variable corresponds to predicted values from the first stage.**

° 10 industry dummies were included in both regressions.

† Regressors' joint significance test.

‡ Hausman's Wald test comparing the joint significance of the difference of the regressors' coefficients estimates with those obtained using a single probit approach.

APPENDIX

Table A1: Marginal effects (dY/dX) of "external R&D" models from the Bivariate probit with sample selection regressions, calculated at the regressors' mean values.

Variables (X)	Z=Tot.Ext	Z=Ext. Firms	Z=Ext. Unice
	Y=Pr(Z=1|R&D=1)	Y=Pr(Z=1|R&D=1)	Y=Pr(Z=1|R&D=1)
R&D_Diver(10^{-3})	-.8 (.86)	-1.78 (1.81)[c]	-.06 (.1)
Herf_Own	.18 (2.33)[b]	.202 (2.6)[a]	.029 (.43)
State (10^{-3})	1.35 (.88)	.4 (.3)	2.81 (2.25)[b]
Subsidy	.149 (2.94)[a]	.138 (2.5)[b]	.114 (2.25)[b]
OutsourceP (10^{-2})	.234 (3.68)[a]	.23 (3.68)[a]	.114 (2.14)[b]
Intensity* (10^{-2})	-.24 (.26)	-.59 (.64)	.658 (.86)
Pavitt	-.125 (2.44)	-.104 (2.1)[b]	-.043 (.95)
Size	-.056 (1.94)[c]	-.017 (1.26)	.004 (.32)
LT_Debt	-.14 (.54)	-.15 (.56)	.068 (.3)
ST_Debt	.314 (2.56)[a]	.183 (1.50)	.079 (.72)
Prod_Diver	.032 (1.51)	.027 (1.63)	.02 (1.14)
Group_Head	.055 (1.62)	.051 (1.51)	.032 (1.34)
OutsourceS (10^{-3})	-.214 (1.28)	-.19 (1.28)	-.013 (1.06)
Few_Cust (10^{-3})	-.66 (1.96)[b]	-.6 (1.93)[c]	-.43 (1.39)
Big_Comp	.06 (1.94)[c]	.056 (1.87)[c]	.038 (1.38)

Notes:
***t*-statistics in parentheses.**
[a, b, c] Significant at the 1%, 5% and 10 % level respectively.
*** Variable corresponds to predicted values from the first stage.**