Barriers to innovation and subsidy effectiveness*

Xulia González†, Jordi Jaumandreu‡ and Consuelo Pazó§

This version: January 2004

Abstract

This paper explores the effects of R&D commercial subsidies by means of a model of firms' decisions about performing R&D when some government support can be expected. The model is estimated with an unbalanced panel sample of more than 2,000 performing and non-performing Spanish manufacturing firms. For the non-performing firms, we compute the trigger subsidies required to induce R&D spending. Among the performing firms, we detect those that would cease to perform R&D if subsidies were eliminated. We also explore the change in the privately financed R&D effort of the performing firms. Results suggest that subsidies stimulate R&D activities, and even show that some firms would stop performing these activities in their absence, but also reveal that most actual subsidies go to firms that would have performed R&D otherwise. In these firms, however, subsidies are found to enlarge expenses with no crowding out of private funds.

Keywords: R&D, innovation, subsidies, thresholds, Tobit model

JEL Classification: O32,L11,C24

*We are grateful to Manuel Arellano, Isabel Busom, Raquel Carrasco, Miguel Delgado, Bronwyn Hall, Lars-Hendrick Röller, Peter Zemsky, and the audiences at the 2nd CEPR Conference on Applied IO (Lisbon) and at Universidad Carlos III de Madrid for their suggestions. We thank to the editor, Ariel Pakes, and two anonymous referees for useful comments which have greatly improved the current version. We acknowledge support from projects SEC2000-0268 and BEC2002-02527 (CICYT), and PGIDT00PXI30005PN.

†Dpto. Economía Aplicada. Universidade de Vigo. E-mail: xgzlez@uvigo.es

‡Dpto. de Economía. Universidad Carlos III de Madrid. E-mail: jordij@eco.uc3m.es

§Dpto. Fundamentos da Análise Económica. Universidade de Vigo. E-mail: cpazo@uvigo.es

1. Introduction

Public sectors of all industrialized countries spend considerable amounts of money on supporting commercial R&D in manufacturing firms. Firms apply for subsidies for research, and agencies choose the research to be funded. The economic justification for these programs lies in the presumable failure of the market to provide incentives to firms to allocate enough resources to innovative activities (Arrow (1962), Nelson(1959)). Positive externalities affecting other firms and consumers induce a divergence between the social and private returns of such activities.

Despite the spread of these subsidies, the evidence of their effects on firms' behavior remains relatively modest and controversial (see, for example, the survey on microeconometric evidence by Klette, Moen and Griliches (2000)[1]). Researchers are currently trying to determine whether subsidies stimulate R&D, in the sense that firms undertake projects that otherwise would not have been carried out, and also whether public funds crowd out the company-financed R&D expenditure. The most recent firm-level econometric studies still offer conflicting answers.

Wallsten (2000) estimates a simultaneous model of expenditure and funding for a sample of US firms and claims that, controlling for grants endogeneity, no effort effect is detected and full crowding out is present. Busom (2000) estimates effort equations for a Spanish sample divided into subsidized and non-subsidized firms, controlling for selectivity, and concludes that full crowding out effects cannot be ruled out for 30% of the firms, while partial crowding out may be important. On the contrary, Lach (2002) estimates the relative increase in R&D expenditures of subsidized vs non-subsidized firms, using panel data on a sample of Israeli companies, and finds that small firms enjoy a positive (dynamic) total effect which fades in the bigger firms. And Almus and Czarnitzki (2003) compare the average effort of East German subsidized firms with the effort of similar (in probability of

[1] Or the related works by Hall and Van Reenen (2000) on fiscal incentives, and David, Hall and Toole (2000) on public/private R&D. See also the interesting account of the Israeli experience by Tratjtenberg (2002).

subsidy) non-subsidized firms in a matched sample, obtaining the significant difference of four percentage points.

The heterogeneity of the results mirrors the diversity of methods and focus in dealing with the two problems which must be addressed to make estimates convincing; namely, the selectivity of subsidy receivers and the endogeneity of subsidies. Furthermore, available data sets often impose severe limitations when it comes to addressing these problems. For example, a significant number of samples include only R&D performers and many show a reduced time dimension.

This paper aims to explore the effects of commercial R&D subsidies by focussing on the modelling of firms' decisions when some government support can be expected: whether or not to perform R&D projects, and the associated level of R&D effort (R&D expenditure over sales). It tries to shed light on the questions of interest by constructing a simple but explicit structural framework to explain why and how the firms' investments can ultimately be inhibited, and by employing a sample of highly heterogeneous firms to identify the model parameters (a sample that in particular includes both R&D performers, subsidized or not, and non-performers).

>From the estimation of the model we derive profitability thresholds and gaps for the expenditure on innovative activities for every firm. For non-performing firms, we compute the trigger subsidies required to induce R&D spending. Among the performing firms, we detect those that would move back across the profitability threshold and cease to perform R&D if subsidies were eliminated. In addition, we can assess subsidy efficiency for the performing firms. The results suggest that market failures do matter and that subsidies can play a role, and play it effectively, in stimulating R&D activities, but also that most actual subsidies in fact go to firms that would have performed innovative activities had they not received the subsidy. However, subsidies are added to the private funds allocated by these firms to R&D, enlarging investments, with no evidence of funding crowding out funds or inefficiency of use.

The model considers each firm a product-differentiated competitor, which can shift the demand for its product by enhancing product quality through R&D expenditures. Demand characteristics, technological opportunities and set-up costs of R&D projects interact to determine the attainable innovative outcomes and a spending profitability threshold. Below this threshold, R&D costs are not completely recovered by means of the sales increment[2]. Firms can then find it more profitable not to undertake innovative activities, but this decision can change if expected subsidies (the fraction of expenditure that is expected to be publicly supported) reduce the cost of R&D. The same framework explains how performing firms take expected grants into account to determine the size of R&D planned expenditures.

This framework naturally leads to a Tobit type modelling of a censored variable, which we will call "optimal non-zero effort," for estimating the model parameters and, particularly, the effect of subsidies. But subsidies are presumably granted by agencies according to the effort and performance of firms, and hence are the result of selection and endogenous. We estimate expected subsidies and use them in explaining effort, applying methods to deal with selectivity and endogeneity in attempting to obtain consistent estimates.

To estimate the model, we use an unbalanced panel of more than 2,000 Spanish manufacturing firms observed during the period 1990-99. The data come from a random sample drawn by industries and size strata, and hence results can be claimed to be valid for the whole industry. During the period, several commercial R&D subsidy programs accounted for innovations' primary source of support. Firm sample behavior is, however, heterogeneous. About 25% of the firms with more than 200 workers and about 80% of the firms under this size do not report performing formal R&D. And only a fraction of performing firms, increasing with firm size, obtains subsidies.

The paper is organized as follows. Section two describes the data set and the main facts about innovation activities and subsidies. Section three develops the model for the R&D decisions of firms and relates them to subsidies. Section four presents the econometric model and explains how it can be used to measure subsidy effects. Section five reports the results and section six discusses the implied subsidy effects. Section seven concludes. Three appendices detail several aspects of the model, the econometrics and the data.

[2] These situations can be characterised as reflecting a market failure when the addition of the consumer net surplus increase derived from the investment would give a positive global surplus.

2. Data and description

The basic data set is an unbalanced panel of Spanish manufacturing firms surveyed during the nineties[3]. At the beginning of the survey, firms with fewer than 200 workers were sampled randomly by industry and size strata retaining 5%. Firms with more than 200 workers were all requested to participate, and the positive answers initially represented approximately a self-selected 60% of firms within this size[4]. Our particular sample includes a total of 2,214 firms, observed during the period 1990-99, selected according to data availability.

The data provide information on the total R&D expenditures of the firms, including intramural expenditures, R&D contracted with laboratories or research centres, and technological imports, that is, payments for licensing or technical assistance. We consider a firm to be performing technological or innovative activities when it reports some R&D expenditure. Our central interest lies in the firms' R&D expenditures and their technological effort, defined as the ratio of R&D expenditures to firm sales. To explain these variables, we use the extensive information on the firms' activities covered by the survey and the data on subsidies. During the nineties, subsidies as a whole were the main public incentive available for manufacturing firms to undertake research programs. Our subsidy measures refer to the total amount of public financing received for each firm under different program headings[5]. Sample and variable details are given in the Appendix C. In what follows, we summarize some facts about R&D expenditures and granted subsidies.

Tables 1 and 2 report some facts about the degree to which Spanish manufacturing firms engage in formal R&D activities. Table 1 shows that the probability of undertaking R&D

[3]The survey was sponsored by the Spanish Ministry of Industry under the name "Encuesta sobre Estrategias Empresariales" (Survey on Firm Strategies).

[4]To preserve representation, samples of newly created firms were added every subsequent year. Exits from the data base come both from death and attrition, but they can be distinguished and attrition was maintained under sensible limits.

[5]Commercial R&D subsidies in Spain had at the time three sources: the European Framework program, with a wide variety of subprograms, but which reached a very small number of firms; the Ministry of Industry programs, which include the subsidies granted by the specialised agency CDTI (Centre for Industrial Technological Development), and the technological actions of regional governments.

activities increases sharply with size[6] (average probability is 21% for firms with fewer than 200 workers and 73% for firms with more than 200 workers.) This probability, which shows some procyclical features, has been increasing slightly over time for the smallest firms. Table 2 adopts another perspective by distinguishing stable and occasional performers during the period. Stable R&D performers are firms that report R&D expenditures every year they remain in the sample. Occasional performers are the firms that report R&D expenditures only some of the years they remain in the sample. Stable performance of R&D activities is strongly correlated with size, while occasional performance shows an inverted u-shaped relationship with respect to size.

Expenditures among the R&D performers are highly unequal, but tend to show a lower critical value, suggesting set-up costs. Figure 1 depicts the (standarized) distributions of the logs of firms' expenditures, keeping the corresponding expenses in thousands of euros as labels[7]. Both distributions tend to fit very well the standarized normal, and hence expenses can be taken as lognormal. The vertical dashed lines point out the modes of the lognormal distributions[8], with values of about 4 and 54 thousand of euros, which we take as approximate critical expenditure values (associated probabilities of observing lower expenditures are 5.8 and 7.1% respectively). To assess their importance in relative terms, we average observed minimum industry sales over a breakdown of manufacturing in 110 industries. Absolute critical expenditures divided by average minimum sales give rough critical values for R&D effort of 1.9 and 0.8 percentage points respectively. Absolute critical expenditures for the smallest firms are smaller, but they seem to be harder in relative terms.

Sales of the R&D performers are, on the other hand, clearly associated with R&D expenditures. Figure 2 depicts the non-parametric regression curves of log of firm sales on log of firm R&D. The slope at each point can be interpreted as the local elasticity of sales with

[6] Here, as in the rest of tables, size is referred to the first year the firm is in the sample.

[7] Representation is based on the standarized values of the data after dropping 2.5% of the values at each tail. Heterogeneity is likely to influence the variance of the distribution by mixing the typical expenditure amounts of different activities (some of them very low).

[8] If $x \sim lognormal(\mu, \sigma^2)$, $\text{mode}(x) = e^{\mu-\sigma^2}$. According to their means and standard deviations we assume distributions as $lognormal(3.85, 1.57^2)$ and $(6.15, 1.47^2)$.

respect to R&D expenditure. Sales show an average elasticity of about 1/3 with respect to R&D. Simultaneous determination of sales and R&D and the firms' heterogeneity prevent us from reaching any conclusion, but the figure suggest an underlying relationship between R&D and sales of the type we assume in the next section.

Tables 3 and 4 report the main facts about grants. Table 3 shows that only a fraction of R&D performers receive subsidies and that the proportion of subsidized firms tends to increase with firm size and stable performance. Figure 3 depicts the distribution of the subsidy amounts. Many subsidies are small, but the spread is also important. Table 4 shows that the typical subsidy covers between 20% and 50% of the R&D expenditures and also that the rate of subsidized expenditure is inversely related to firm size (particularly for the stable performers).

Tables 5 and 6 provide a first look at the relationship between subsidies and effort, based on the comparison of the R&D effort of subsidized and non-subsidized performers' data. Both tables show a positive association between the granting of subsidies and R&D effort, both in the whole period and year to year. The data show more than "additionality," in the sense that subsidized efforts minus the part of this effort attributable to subsidies are higher than non-subsidized efforts. Figure 4 provides a first look at the relationship between the privately financed expenditure and the amount of the subsidy for subsidized firms[9]. According to the figure, private expenses tend to show a unit elasticity with respect to public funds.

Therefore, data suggest non-negative and even positive R&D effects of subsidies. But this can be the sole consequence of other non-controlled variables or because the relationship goes either way: firms with more R&D are more likely to receive subsidies, and bigger subsidies the higher the R&D expenses. Only the development of an econometric analysis can provide further insights on this relationship, by providing evidence on how these data patterns can be interpreted in terms of "causal" effects.

[9] Representation is carried out by dropping the subsidies higher than their associated yearly R&D expense values (see Section 5.1) and a 2.5% of subsidy values at each tail.

3. R&D with set-up costs

Assume firm i competes in prices in a product-differentiated industry, facing a negatively sloped demand. Demand, however, can be shifted by enhancing the quality of the product. Write demand as $q_i(p_i, s_i)$[10], where p_i stands for the own price, s_i for the level of quality, and suppose $\partial q_i/\partial s_i > 0$ and $\partial q_i^2/\partial s_i^2 \leq 0$. We drop the subscript i for simplicity.

Quality can be improved by incurring R&D expenditures, denoted henceforth by x, according to some technological rules. In particular, to surpass the current industry standard quality $s(0)$, firm i must incur some set-up costs per period that we will denote by F. Beyond F, R&D expenditures affect quality according to the function $s = s(x)$, where $\partial s/\partial x > 0$ and $\partial^2 s/\partial x^2 \leq 0$[11]. Set-up costs are in principle avoidable by producing at quality $s(0)$, and hence innovative inaction is possible, while innovative activity implies incurring F.

A firm can search for a suitable public support program and apply for having its R&D expenditures subsidized by a monetary fraction ρ which lowers the cost of its innovative activity. But the firm must make its decisions ex-ante (at the time of setting its R&D plans)[12] and we assume that they are based on expected profitability and hence on the firm's expectation about the subsidy. Moreover, public funds can be associated with either a higher or lower level of expenditure efficiency[13]. Accordingly, we will parametrize the expected cost of a unit of efficient R&D as $E[(1-\rho)^\beta]$, where β is a parameter of efficiency (if $\beta = 1$, public funds leave efficiency unchanged) and E indicates the expectation over ρ

[10] We assume that price competition can be taken as stable over time and the effect of the price of the rivals p_{-i} subsumed in the relevant own-price demand elasticity. Accordingly, we write $\widetilde{q}_i(p_i, p_{-i}, s_i) = q_i(p_i, s_i)$. We relax this assumption in the empirical exercise by considering eventual competition changes.

[11] Innovative investments shift the demand for the firm product instead of the production function. In this sense the model can be taken as a variant of the classical Griliches (1979) R&D "capital" framework. See also Pakes and Schankerman (1984) for a formalization of the gross increments of K in terms of a "knowledge production function." Our specification formalizes the effects of the indivisibility of some resources plus input complementarity in the production of innovations. For more on all of this, see Appendix A.

[12] Subsidies are granted to firms committed to R&D investments.

[13] On the one hand, public funding often gives access to other facilities or advantages (e.g., access to public laboratories and researchers). On the other, public funds can be mainly envisaged as easing liquidity constraints and allowing for less financing discipline, which implies less expenditure efficiency.

values. Firms have subjective conditional distributions of probability, which depend on their beliefs in the chance of succes in the search for a subsidy program, and in the likelihood of being granted a subsidy by the agency.

Suppose now that production marginal cost is c. To set the product price and decide the pertinence and level of R&D expenditures, the firm solves the (variable profits) problem

$$\max \ (p-c)q(p,s) - E[(1-\rho)^{\beta}]x$$
$$s.t. \quad s = \begin{cases} s(0) & \text{if } x < F \\ s(x) & \text{otherwise} \end{cases} \qquad [1]$$

which turns out to be a problem with a non-convex constraint. Equilibrium will be characterized by the pair (p^e, x^e) such that $\Pi(p^e, x^e) = \max\{\Pi(p^*, x^*), \Pi(p^{**}, 0)\}$, where $p^*$and p^{**} may diverge and (p^*, x^*) is the interior solution. Let us also define $\overline{x}$, the expenditure level which (given optimal pricing) makes the firm indifferent to performing R&D or not.

Equilibrium admits the straightforward representation of Figure 5. Net attainable revenue R (supposing optimal pricing) is represented as a function of x[14]; isoprofit curves are linear with slope equal to the (expected) effective cost of R&D, $E[(1-\rho)^{\beta}]$, and firm decision is dictated by the maximum of two ordinates: the profit Π_0 corresponding to $x = 0$ and a profit as Π_1 or Π_2, say, associated respectively with optimal non-zero solutions x_1^* or x_2^*. Firms with profit Π_1 would perform R&D and firms with profit Π_2 would not. A firm with optimal non-zero solution at $\overline{x}$ (the non-represented slope of R at this point crosses the y-axis at Π_0) would be indifferent to performing R&D or not[15].

Assume that firms can in any case obtain a non-negative profit from performing R&D, that is, $\Pi(p^*, x^*) \geq 0$. Then, there exists an effort for both performing and non-performing firms, which we will call optimal non-zero effort, which can be summarized in the unique expression

$$E^* \equiv \frac{x^*}{p^*q^*} = \left(\frac{s}{q}\frac{\partial q}{\partial s}\ \frac{x}{s}\frac{\partial s}{\partial x}\right) / \left(-\frac{p}{q}\frac{\partial q}{\partial p}\ E[(1-\rho)^{\beta}]\right) \qquad [2]$$

[14] $R(x) = (p(x)-c)q(p(x), s(x))$ where, using the envelope theorem, $\frac{\partial R}{\partial x} = (p-c)\frac{\partial q}{\partial x}$ and $\frac{\partial^2 R}{\partial x^2} = (p-c)\frac{\partial^2 q}{\partial x^2}$.

[15] Two parametric representations which may help to explore the model easily are $R(x) = x^{\alpha}$, for the relevant x range, and the (constant price elasticity) demand specification $q(p, s(x)) = q_0(p)(1 + \varepsilon \ln \frac{x}{F})$, where it is helpful to use $F/pq_0(p)$ as a measure of set-up costs.

which results from rearranging the FOC interior conditions of [1], and is a Dorfman-Steiner (1954) type of condition. Non-performing firms, however, would only choose this (local maximum) allocation if they didn't have the more profitable alternative of inaction [16].

R&D level expenditures x and effort E can be used interchangeably because the model and assumptions imply that E increases monotonically with x for a given firm[17]. In particular, let us define $\overline{E}$ as the effort which corresponds to $\overline{x}$. $\overline{E}$ is unique and has the same determinants as expenditure $\overline{x}$. Optimal non-zero effort E^* will only be observed when it surpasses this threshold effort $\overline{E}$.

Formula [2] shows that optimal non-zero effort increases with the elasticity of demand with respect to R&D expenditure, which can be conceptually decomposed in the elasticity of demand with respect to quality (demand conditions) and the elasticity of quality with respect to R&D expenditure (technological opportunities)[18]; with the degree of market power (the inverse of the price elasticity) for a given form of rivalry; and with the expected subsidy. "Lack of appropriability," as a factor which discourages R&D, can be easily discussed in this framework. For example, high knowledge spillovers mean a high likelihood of a rapid matching of product innovations by rival firms, and hence a lower (net) demand elasticity with respect to quality. For given F, this increases the likelihood of an optimal non-zero effort below the threshold effort.

An important consequence of this model is that (expected) subsidies have two potential different effects. The first is that they can induce firms to perform R&D. The second is that they can enhance the R&D expenditures of the firms that would perform innovative activities in any case.

[16] It could also be assumed that firms cover either totally or in part the fixed expenditures through (interperiod) sunk investments. Sunk investments can reduce the size of the avoidable per period set-up costs, rendering inaction less probable.

[17] To see this, simply think of a representation of gross revenue pq as a function of x analogous to that employed in Figure 5. E equals the inverse of the slope of the ray which goes from the origin to any relevant point of the revenue function.

[18] We can assume the standard account of determinants of innovative activities to be underlying these elasticities (see, for example, Cohen (1995) or Cohen and Levin (1989)).

4. Barriers to R&D and subsidy effects.

4.1 Econometric model

Let e^* and $\overline{e}$ stand for the logs of E^* and $\overline{E}$, respectively. Starting from the previous model we assume

$$e^* = -\beta ln(1-\rho^e) + z_1\beta_1 + w \quad [3]$$

$$\overline{e} = z\beta_2 + u_2 \quad [4]$$

$$\rho^e = E(\rho|z_\rho) = g(z_\rho, \lambda) \quad [5]$$

where e^* is only observed when $e^* - \overline{e} > 0$, ρ^e is the expectation about ρ, and w represents an autocorrelated error of the form $w_t = \gamma w_{t-1} + \varepsilon_{1t}$ (for simplicity, time subindexes are used only when needed to avoid confusion). We assume that (ε_1, u_2) is bivariate Normal, with zero mean, independent of z and z_ρ (z_1 is a subset of z) and serially independent, with $V(\varepsilon_1) = \sigma_1^2$, $V(u_2) = \sigma_2^2$ and $Cov(\varepsilon_1, u_2) = \sigma_{12}$.

The effort equation [3] is obtained by taking logs in [2], substituting $\beta ln[1 - E(\rho|z_\rho)]$ for $lnE[(1-\rho)^\beta|z_\rho]$[19], and letting z_1 stand for the vector of variables that determine the value of the (log of) elasticities. Expected subsidies enter the effort equation in the way they appear in the first order condition [2], but elasticities are endogeneous unobservable variables of the underlying model that we replace with a set of reduced form determinants (i.e., variables exogenous or predetermined with respect to (ε_1, u_2)[20]). Autocorrelated disturbance w takes into account that we are not likely to be able to fully specify optimal non-zero effort determinants. For example, the degree of ability acquired over time by doing R&D or the degree of involvement in sunk investments are hardly observable, and this total or partial

[19] By using a Taylor second order expansion of $(1-\rho)^\beta$ around $E(\rho)$ it can be shown that $\ln E[(1-\rho)^\beta] \simeq \beta \ln[1-E(\rho)] + \ln[1+\frac{1}{2}\beta(\beta-1)c_V^2]$, where c_V is the coefficient of variation of $(1-\rho)$, i.e. $c_V = \frac{[Var(1-\rho)]^{1/2}}{E(1-\rho)}$. The second term of this equality is likely to be small, of order $\frac{1}{2}\beta(\beta-1)E(\rho^2)$ and, under certain circumstances, constant.

[20] Some variables are taken to be predetermined in the sense that $(\varepsilon_{1t}, u_{2t})$ is assumed to be uncorrelated with their current and past values but feedback effects from lagged errors are not ruled out. Predetermined variables include lagged values of endogenous variables.

unobservability is a likely source of autocorrelated errors.

Equation [4] models thresholds. We take firms as having idiosyncratic stochastic thresholds, which can be presumed to be a function of the same variables that determine e^* and perhaps some others of the same kind (z contains at least all variables in z_1). The coefficients give the height of the "barriers" to the profitability of R&D.

Equation [5] states our assumption that the unobservable firms' expectations ρ^e can be related to observable data through the function $g(z_\rho, \lambda)$, with z_ρ such that (ε_1, u_2) is independent of z_ρ. The function gives the financial support each firm presumes it can obtain given its characteristics and the allocations observed from agencies. In particular, any agency evaluation of firm conditions is anticipated through firm attribute indicators.The function is likely to be highly non-linear and z_ρ is only partially overlapping with z (it contains at least the variables not in z needed for identification).

Equations [3]-[5] define a rather standard Tobit type model[21]. R&D performance, and hence observation of the optimal non-zero effort e^*, is determined by the sign of $e^* - \overline{e}$ (selectivity or decision equation). But the model also has some non-standard features.

Firstly, disturbances of the effort equation are assumed to be autocorrelated. This implies that predetermined variables are likely to be correlated with these disturbances. To ensure consistency, the effort equation must then be specified in the pseudo-differenced form $e_t^* = \gamma e_{t-1}^* - \beta(\ln(1-\rho_t^e) - \gamma \ln(1-\rho_{t-1}^e)) + (z_{1t} - \gamma z_{1t-1})\beta_1 + \varepsilon_{1t}$ and this raises the difficulty that the latent variable e^*, only partially observable, also becomes an explanatory variable.

Secondly, we have the unobservable ρ^e. Observed subsidies are granted by agencies according to, among other things, the contemporary capacity, effort and performance of firms and hence they are presumably endogenous (their values are likely to be correlated with the random term $\varepsilon_1$and hence with u_2). Our framework assumes, however, that relevant subsidies are the subsidies expected in advance by firms, ρ^e, which can be expressed in terms of a set z_ρ of exogenous or predetermined variables. But, as ρ^e is unobservable, we need to substitute the generated regressor $g(z_\rho, \widehat{\lambda})$ for the expectation.

[21] A type II Tobit model, in the words of Ameniya (1985) (see also Wooldridge (2002)). Econometric models of censored variables with stochastic thresholds date back to Gronau (1973) and Nelson (1977).

4.2 Estimation procedure

Estimation is carried out by a two-step procedure: first we estimate the conditional expectation of subsidies, and then we estimate the Tobit model, by maximum likelihood methods. Let us explain these steps in turn.

To estimate the unobserved variable $\rho^e = E(\rho|z_\rho) = g(z_\rho, \lambda)$, we decompose the expectation as follows

$$\rho^e = E(\rho|z_\rho) = P(\rho > 0|z_\rho)E(\rho \mid z_\rho, \rho > 0) \qquad [6]$$

where $P(\rho > 0|z_\rho)$ stands for the conditional expectation of getting a grant and $E(\rho \mid z_\rho, \rho > 0)$ for the expected value of the subsidy conditional on z_ρ and its granting. This allows us to use two natural "rationality" or "correctness" restrictions on the expectations to estimate the $E(\rho|z_\rho)$ function. On the one hand, we assume that firms which effectively receive a subsidy are able to forecast the amount of the subsidy up to a zero mean error. Accordingly, we use the subsample of observations in which firms are granted a subsidy to consistently estimate the parameters of the granting conditional expected subsidy function. On the other hand, we assume that firms correctly forecast the probability of getting a subsidy (which obviously is not the same as anticipating if whether are going to get a subsidy). Consequently, we use the grants observed in the whole sample to estimate the conditional probability function[22].The expected subsidy function can be computed from estimates on these two expectation functions.

We specify $P(\rho > 0|z_\rho)$ by means of a probit of parameters λ_1. And we assume $\ln \rho|(z_\rho, \rho > 0) \sim N(z_\rho\lambda_2, \sigma^2)$ to estimate of $E(\rho \mid z_\rho, \rho > 0)$. Using the estimated parameters, expected subsidies are then computed as $\widehat{\rho^e} = \Phi\left(z_\rho\widehat{\lambda}_1\right)\exp\left(z_\rho\widehat{\lambda}_2 + \frac{1}{2}\widehat{\sigma}^2\right)$ for all firms in the sample.

Substituting $\widehat{\rho^e}$ for ρ^e in the effort equation, we can face the estimation of the Tobit model by partial maximum likelihood. Amemiya (1985) discusses alternative identification conditions of the thresholds model (see also Maddala (1983) and Wooldridge (2002)). One of these conditions is the availability of at least one variable that enters the equation for the censored variable but can be excluded on theoretical grounds of the thresholds equation.

[22] A more structural approach to the probability function is unfortunately prevented by the fact that we cannot separately identify the sample of applying firms.

This condition arises naturally in our model, where expected subsidies can be safely excluded from the determinants of thresholds[23].

The likelihood of the model is based on the specification of the joint density associated with the disturbances of the effort equation, ε_1, and the decision equation, $v_2 = \varepsilon_1 - u_2$ (see Appendix B). But specification requires to provide some solution for the unobservability of e^*_{t-1}.We are going to explore the results and insights provided by four versions of the model (see Appendix B for details).

If disturbances of [3] are assumed not to be autocorrelated ($\gamma = 0$), parameters β, β_1 and β_2 can be estimated using the effort equation in levels and applying standard partial maximum likelihood methods. We call this Model I. Estimates of this model will show, as expected, evidence of simultaneity bias.

Autocorrelated errors ($\gamma \neq 0$) imply that the effort and decision equations must include the lagged-latent variable e^*_{t-1}. But the lagged-latent variable is not observed for many of the firms' data points. Estimates must then rely on the remaining sample, which constitutes an (exogenously selected) sample consisting of the firms' observations with positive effort at $t-1$[24]. Selection is here exogenous because observability of e^*_{t-1} is not related to $(\varepsilon_{1t}, u_{2t})$. This is Model II. The main problem with this estimate is the scarce proportion of observations of non-performance ("zeroes"), which in addition only correspond to firms that stop performing R&D at just that moment ("stopping zeroes"). Consistency is reached at a high price in estimation efficiency.

We assume that efficiency in estimation can be improved by using more "zeroes." One way is to reformulate the model in such a manner so that we do not need to observe the lagged-latent variable. This is accomplished by using a pseudo-differences transformation of the decision equation, which amounts to examining the sign of the pseudo-difference $(e^*_t - \overline{e}_t) - \gamma(e^*_{t-1} - \overline{e}_{t-1})$. This sign is always right (agrees with the sign of $e^*_t - \overline{e}_t$) when the sign of $e^* - \overline{e}$ changes from one period to the other, but it must be assumed to be right

[23]This happens because thresholds for profitable technological activities are defined in terms of the total expenditure needed, independently of its composition.

[24]See Arellano, Bover and Labeaga (1999) for an application of this solution in a different context.

when positive and negative differences $e^* - \overline{e}$ tend to remain unchanged. The assumption is more likely if γ is not very big. This is Model III. The decision equation of this model shows a composite disturbance including u_{2t-1}. This implies that any endogenous variables included among the predetermined should be lagged twice to avoid correlation with this disturbance, and that we induce some autocorrelation in the likelihood score which must be taken into account.

Finally, we can trace the differences in observable behavior (never performers, occasional performers, firms always performing) back to an order of magnitude of the differences $e^* - \overline{e}$ in addition to their sign. This seems particularly sensible if we take the size of these differences to be closely related to the accomplishment of sunk investments (or to their absence). This insight suggests the use of an enlarged Tobit model in which the observability of the latent variable depends on the value reached by the "cut parameter" of an ordered probit. This is model IV.

Models I to IV are estimated using partial maximum likelihood estimators with a generated regressor; these estimators solve $\max_\theta \sum_i \sum_t \log L_{it}(\theta, \widehat{\lambda})$. Asymptotic standard errors are computed taking into account the variance of $\widehat{\lambda}$ and, eventually, possible correlations between the scores at different periods of time (see, for example, Wooldridge (2002)). Maximum likelihood estimation is carried out through a grid over the values of the disturbances correlation coefficient r, beginning at $r = 0$ (see Nawata and Nagase (1996)). Models in pseudo-differences are estimated performing a combined grid over the r and γ values.

4.3 Measuring profitability gaps and subsidy effects

Given parameter estimates of the model, one is ready to compute individual optimal non-zero effort and threshold estimates and use them to assess the effects of subsidies. We will do this relying on the non-stochastic components of the equations, that is, evaluating the relationships at the (zero) expected value of the disturbances[25].

[25] Let us distinguish two possible measures of (efforts, thresholds and) gaps: $E[\exp(e^*) - \exp(\overline{e})|z, w = v = 0] = \exp[E(e^* - \overline{e})]$, which gives the level values corresponding to the (zero) expected value of the disturbances, and $E[\exp(e^*) - \exp(\overline{e})|z]$, which also averages the unobserved heterogeneity. The model predicts R&D performance using the first gaps, and we choose to base our

Let us first define profitability gaps. We define profitability gaps as the difference between the optimal non-zero effort in the absence of subsidy and the threshold effort. If negative, they give the R&D expenditure (in terms of % of sales) by which the firm falls short of undertaking profitable innovative activities. If positive, they give the R&D expenditure (in terms of % of sales) that the firm would make, in the absence of subsidies, in addition to the minimum profitable amount. We compute them as $\exp(z\widehat{\beta}_1) - \exp(z\widehat{\beta}_2)$.

Given estimated profitability gaps, we can evaluate the (actual and potential) roles of subsidies in the performance of innovative activities. Let us first focus on trigger subsidies. We define them as the value of the ρ^e's that would induce non-performing firms to undertake innovative activities (by filling their negative profitability gaps). They can be estimated as the values of ρ^e that solve the equations $-\widehat{\beta}ln(1-\rho^e) + z(\widehat{\beta}_1 - \widehat{\beta}_2) = 0$ for observations at which this expression, evaluated at the estimated expected subsidy, is negative.

Let us then evaluate the role of a subsidy withdrawal. Some firms are likely to be performing innovative activities because the support effect of the expected subsidy fills in the negative profitability gap that would exist in its absence. We identify the observations at which $-\widehat{\beta}ln(1-\widehat{\rho}^e) + z(\widehat{\beta}_1 - \widehat{\beta}_2) > 0$ but with $z(\widehat{\beta}_1 - \widehat{\beta}_2) < 0$ (negative profitability gap).

All this refers to the ability of subsidies to induce firms (potentially or effectively) to invest in R&D. But, according to the model, how do subsidies change the expenditure of firms that perform innovative activities? Firstly, notice that R&D expenditures are expanded in the model to increment sales and, therefore, the rate of change in effort constitutes a lower bound for the rate of change in expenditure[26]. Secondly, changes in effort depend on subsidies in a complex way, because all the elasticities in [2] may change with the firm equilibrium. We will use an approximate measure of the change in effort which becomes exact in the simplest case in which elasticities remain constant.

Call $E^*(\rho^e)$ total effort with subsidy and $E^*(0)$ total effort in its absence. Write $(1-\rho^e)E^*(\rho^e)$

measures on these gaps. We also report values for the second gap measure.

[26]The change in expenditure may be conceptually decomposed in the sum of two changes: the change due to sales and the change in effort. An assessment of the sales effect of subsidies would only be possible with a more complete specification of the demand.

for private effort when expenditures are subsidized. It is easy to check that

$$\frac{(1-\rho^e)E^*(\rho^e)-E^*(0)}{E^*(0)}=[(1-\rho^e)^{-(\beta-1)}-1]\lesseqgtr 0 \text{ if } \beta \lesseqgtr 1$$

Therefore, if subsidy efficiency β is unity, private effort will remain the same, which means that privately-financed expenditures will increase by the same amount as sales. On the contrary, if β exceeds unity, the subsidy will increase private effort, and total effort will become higher than the sum of the public fraction and private effort without subsidy. If β were less than unity, private effort would be reduced[27].

5. Empirical specification and results

5.1 Expected subsidies

We estimate the unobservable firms' expectations ρ^e using the probit and OLS specification of [6]. Recall that we want to predict the expected outcome by means of a set of variables which can be considered exogenous or, at least, predetermined. In what follows, we detail them.

First of all, subsidies and their amount tend to persist over time. This persistence can be based either on projects spread over several years or the renewal of grants by particular firms. To pick up persistence, we specify both equations as dynamic, including the dependent variable (the subsidy dummy and the log of the subsidy) suitably lagged. We consider two alternative specifications of the equations: we will use in turn the dependent variables lagged one and two periods. On the other hand, the subsidy amount can be zero for the (one or two periods) lagged values. Hence, this variable is included in OLS regressions split in two: a variable taking the value of the log of the subsidy when positive and zero when

[27] Other studies take the value of the derivative of private expenses with respect to subsidy (see Wallsten (2000) or Lach (2003)). With sales controlled for, this derivative amounts to a linear partial effect (independent of the subsidy value and without demand-induced effects). We can compute an average subsidy effect of this type by evaluating at some point the first term of the right hand of the identity

$$\frac{(1-\rho^e)x(\rho^e)-x(0)}{\rho^e x(\rho^e)-0}=\frac{(1-\rho^e)E^*(\rho^e)-E^*(0)}{\rho^e E^*(\rho^e)}+\frac{E^*(0)}{\rho^e E^*(\rho^e)}\frac{S(\rho^e)-S(0)}{S(\rho^e)}$$

where S is a shorthand for sales.

the subsidy is zero, and a dummy which takes the value one when this is the case[28].

We use the same set of additional variables to estimate both equations. We first include a series of firm characteristics that may enhance the willingness to apply and/or the eligibility of firms: their size, age, an indicator of the degree of technological sophistication and firms' capital (in equipment goods and machinery) growth. We then include three indicators of firm situations that can turn out to be significant to granting agencies for politico-economic reasons: a dummy characterizing whether the firm is a domestic exporter, a dummy denoting whether the firm has foreign capital, and another indicating whether the firm is likely to have significant market power. A number of these variables are considered predetermined and always included as lagged one period (Size, Domestic exporter, Firm with market power); others are assumed strictly exogenous or predetermined longer in advance (Age, Technological sophistication, Foreign capital)[29]. Finally, we add three sets of dummy variables to account for sectorial heterogeneity (industry dummies), differences in regional support policies (region dummies), and changes over time (time dummies).

Table 7 reports the results of the estimation. Results are sensible and turn out to be similar in the two specifications (dependent variables lagged once and twice). The goodness of fit of probit models is checked using the explained percentage of ones and zeroes when the critical value is suitably selected (samples have only about 8% of ones). The OLS model explains approximately 50% of the variance of the observed subsidies' values.

Persistence turns out to be significant. Industry dummies tend to reveal heterogeneity across manufacturing. Region dummies show a significantly greater probability of subsidies for two particular regions. Although the characterization of the granting process is not the main target of these estimations, the estimated equations seem good enough to provide a stylized summary of it: the big, mature, technologically sophisticated and expanding firms are more likely to obtain grants for their innovative activities, as well as the domestic

[28] In addition, a small number of sample subsidy values (33) are higher than their associated yearly R&D expenditures. We assume that this reflects simple accounting imperfections in the time allocation of subsidies.

[29] Exceptionally, the capital growth variable, already in differences, will be alternatively used contemporaneously and lagged once to avoid losing extra data points.

exporters, but agencies seem to apply some criteria in expenditure coverage favoring the relatively small, new, domestic and competitive firms.

Computed expected subsidies are sensible. Average probability is near 8%, average expected subsidy conditional on its granting 28% and average expected subsidy is about 2%, with a standard deviation of 4%[30]. Only a negligible number of predictions for expected conditional values slightly surpasses 100%, and no prediction of expected subsidy lies outside the relevant interval (with a maximum value of 59%).

5.2 Tobit Model

Let us now detail the specification of equations [3] and [4]. According to the model, there are three main types of variables to be considered: indicators of market power/competition conditions, variables used to reflect the sensitivity of demand with respect to product quality and product quality with respect to R&D expenditure, and variables employed to approximate set-up costs and the heterogeneity of thresholds among firms. Obviously, no variable can claim to exclusively pick up the effects of one of these headings, but it seems useful to classify them in order to summarize the empirical effects.

With the important exception of expected subsidies, it must be admitted that the same variables can play a role in explaining the optimal non-zero efforts and the thresholds. This happens partly because we have to rely on indirect indicators, but also because thresholds tend to depend on the same factors as effort. However, we will find it both statistically acceptable and useful to impose some exclusion constraints on the effort equation.

Let us briefly detail the main variables included in both equations and their expected roles. Two variables are intended to perform as indicators of market power/competition conditions: the firms' market share and a dummy variable representing concentrated markets. Two variables are included to perform respectively as indicators of a high sensitivity of demand with respect to product quality and/or product quality with respect to R&D expenditure: the advertising/sales ratio and the average industry patents excluding the patents obtained by the firm (a classic formal technological opportunities measure). Finally, we include a dummy variable which takes the value one for the firms with negative

[30] These are the values obtained using the two last columns of Table 7.

cash flow, to represent serious financial difficulties in carrying out innovation activities.

Six variables are included exclusively in the decision equation to account for different aspects of set-up costs. The list consists of the following indicators: presence of foreign capital, location in a region with high spillovers (geographical opportunities), capital growth, a market which has been in recession, a product highly sensitive to quality controls, and employment of highly skilled workers. All these variables are likely to be associated with lower set-up expenses, and some of them also with a high sensitivity of the demand with respect to quality.

A number of these variables are considered predetermined and always included as lagged one period (Market share, Concentrated market, Advertising/sales ratio, Negative cash flow and Recessive market), while others are assumed to be strictly exogenous or predetermined longer in advance (Average industry patents, Foreign capital, Geographical opportunities, Quality controls, Skilled labor). Capital growth enters contemporaneously and lagged once.

In addition, in both equations we include a set of dummy variables of size, measured according to the number of employees, to control for any remaining threshold size effect. Moreover we include a set of 18 sector dummies, to control for permanent differences arising from activities. Details on the employed variables can be found in Appendix C.

Table 8a reports the results of carrying out the estimation of the different versions of the model. Samples change according to the estimated version for two reasons: the "usable" time observations[31] and the exogenous selections performed in each case. Variables, instead, are always kept the same (although lags used to predict expected subsidies change from estimates a to estimates b).

Expected subsidy is included in the form $-ln(1-\widehat{\rho}^e)$ and it would be surprising to obtain a β estimate very far from unity when estimating consistently. In fact, the sequence of estimates of Table 8a strongly confirms what we expect from theory. Estimates in levels (Model I) show clear signs of bias, both when they are carried out with the unselected

[31] Levels estimation including lagged variables requires dropping the first observation of each firm from regression, pseudo-differences require dropping the first two observations, and pseudo-differenced equations using a regressor generated employing variables lagged twice require dropping the first three observations.

sample and the selected sample used next to obtain a consistent estimate. The extremely large β coefficient can be attributed to the correlation between the generated regressor and an autocorrelated disturbance. The estimate of Model II supports the presence of autocorrelated disturbances ($\gamma = 0.69$) and shows a dramatic change in the coefficient value, which falls to unity with autocorrelated residuals controlled for. But, as discussed above, Model II provides a consistent estimate at the price of constraining the sample to observations for which the latent variable past value is observed. This induces a big efficiency loss, which is in fact very apparent in the σ estimate and the variances-covariances of the remaining parameter estimates (not shown in the table). Model II uses hardly a fourth of the available observations and includes a scant 13% of zero effort observations.

Model III provides an interesting alternative for consistent estimation of parameter β. The parameter estimate is both sensible when the subsidy regressor is generated using both the one-lag and the two-lags alternatives, but Model IIIb implies a more judicious choice from the point of view of the model assumptions (subsidies lagged twice are expected to be orthogonal to the first lag of u_2). In addition, the preserving sign assumption, on which the model transformation is based, holds ex-post in 96.5% of cases. Moreover, coefficients are sensible (see Table 8b and comments below) and fit is good. We take this model as our preferred estimate, and we will base our economic discussion on its parameter estimates.

Does the modelling of uncertainty really make a difference in estimations? To check this we alternatively estimate models II and IIIb using the simple prediction of subsidy amounts for the firms obtaining subsidies and zeroes for the rest. This can be interpreted as the relevant variable in case firms are certain about the subsidy and the only problem is endogeneity. The β parameter drops to 0.60 and 0.69, respectively. Uncertainty about subsidies is probably a key question outside of the biggest firms.

Serious difficulties appear to obtain meaningful estimates of Model IV without imposing restrictions. The table estimates impose $r = 0$, an acceptable outcome given the small r values obtained in all the (differenced) models. The β estimate turns out to be too high in variant a and sensible although imprecise in variant b (where lags should ensure no correlation). In any case, Model IVb estimates sunk costs to consist on average, for otherwise similar firms, of investments which reduce current thresholds by about 45%[32].

Table 8a (bottom) reports the results of comparing the models' predictions with the actual observations in the sample. All models but Model II behave sensibly, even keeping the standard 0.5 critical value for prediction. For Model II, highly unbalanced in terms of ones and zeroes, it is better to compute prediction with an adjusted critical value which equals the prediction outcomes. The rest of the models can also be compared in terms of adjusted critical values (see table's footnote 6).

Table 8b shows all the results of model estimation. The interpretation of coefficient estimates can be made as follows. Market power clearly influences effort, although the effect of the firm market share is somewhat imprecisely estimated. In any case, the impact of market share must be balanced against the degree of rivalry. For a given market share, R&D effort is bigger the more competitive the environment is. This is consistent with the evidence of inverted U-shaped relationships between product market competition and innovative activities (see, e.g., Aghion et al. (2002))[33]. Market power also seems to have the same type of impact on thresholds. On the other hand, spread patent protection emerges as a good indicator of technological opportunities (which perhaps also picks up part of demand sensitivity to R&D) which show a positive impact on effort. But it also performs as an indicator of the corresponding set-up costs of innovative activities, increasing thresholds. Although less precisely estimated, there appear to be two additional effects. The advertising/sales ratio seems to perform weakly as an indicator of demand sensitivity, increasing effort, and tight firm financial constraints increase thresholds.

Finally, the list of firm characteristics included to pick up threshold effects show that the presence of foreign capital, the benefits stemming from geographical spillovers, a high product sensitivity to quality and the presence of highly skilled labor reduce thresholds. The similar effect of the recessive market dummy can be interpreted as controlling for the impact

[32] The log-linear difference between the two extreme relevant thresholds is 0.61 and hence the high threshold is about 1.8 times the low one.

[33] We additionally experimented with the introduction of the variable representing competition changes, which was never fully significant and did not change the main estimation results.

on the (sales relative) threshold of an abnormally low value of sales. In addition, there seems to remain some effect of scale (biggest sizes tend to experience smaller thresholds).

6. Profitability gaps and subsidy effects

Table 9 reports the distribution of the estimated profitability gaps, which are depicted in Figure 6. They show a somewhat skewed distribution, with more density at negative values, and two modes around the zero value (one positive and the other negative). Positive gaps represent about 30% and their mean is about 0.4%, while the average of negative gaps has an absolute value of about 0.8%. Positive gaps show less heterogeneity (90% lie in the (0,1) interval), with an important mass of values concentrated at relatively uniform departures. Negative gaps show a greater heterogeneity (less than 73% lie in the (-1,0) interval), which includes, however, a significant number of firms presenting relatively small gaps[34].

Table 10 further details gap heterogeneity by reporting the distribution of trigger subsidies for the non-performing firms. Subsidies required to induce firms to engage in R&D are smaller for the biggest firms and bigger for the smallest ones. With an expected funding of less than 10% of R&D expenditures, almost 50% of the non-performing big firms will switch to performing innovative activities. On the contrary, inducing 30% of the small firms to perform R&D implies expected support accounting for up to 40% of the expenses, and inducing one firm out of two would require financing up to 50% of the expenses.

Table 11 reports the impact of subsidy withdrawal on performing firms and the expected subsidies which characterize the presumably R&D-abandoning firms. Interestingly enough, subsidy withdrawal would induce stopping innovative activities in a significant number of performing observations (93 observations, about 6% of all positive gap observations), particularly among the smallest firms (almost 14%). More than half of the deterred firms show expected subsidies lower than 10%, but some small firms show more important expected funding. These results suggest that not all funding is allocated to firms with positive prof-

[34] In the distribution $\exp(z\widehat{\beta}_1 + \frac{1}{2}\widehat{Var}(w)) - \exp(z\widehat{\beta}_2 + \frac{1}{2}\widehat{Var}(u_2)) = 1.83\exp(z\widehat{\beta}_1) - 1.61\exp(z\widehat{\beta}_2)$, positive gaps represent about 35%, with a mean of 0.8%, and the average of negative gaps gives an absolute value of about 1.1%.

itability gaps, which would also perform R&D activities in the absence of public financing, and hence that some part of the public financing stimulates R&D activities.

Finally, our preferred point estimate for parameter β (1.07) implies that subsidies only induce modest increases in privately-financed effort. This impact increases with the size of the subsidy, but the increase in private effort for subsidies running from 20 to 60% is by about 2 to 7%[35]. Recall, however, that this is only a lower bound for the increase in private expenses; this increase does not try to disentangle the sales growth effect of the innovative activities. In any case, there is no evidence of funding crowding out, displacement or slackness.

7. Summary and conclusions

Despite the spread of R&D commercial subsidies, the evidence on their impact on firms' decisions about R&D expenditures remains relatively modest and controversial. This paper tries to contribute a series of findings on the potential and actual roles of subsidies, based on the estimation of an explicit and theoretically based model of firms' decisions. The firm's decision on whether or not to spend on R&D emerges from the comparison of optimal non-zero effort with the effort needed to reach some profitability (threshold effort). The focus of the paper is on the impact of the expected subsidy (or fraction of the effort that is expected to be publicly supported) on this comparison, and on the firms' decisions about the level of expenditure. The model is estimated using a censored variable regression method, attempting to obtain inferences robust to selectivity and endogeneity.

The main findings, based on a representative panel sample of more than 2,000 Spanish manufacturing firms, are the following. Non-performance of innovative activities can effectively be traced back to the presence of optimal efforts below the profitability thresholds (that is, negative profitability gaps). Small firms experience the greatest negative profitability gaps, but negative gaps also affect a proportion of big firms. Data also suggest that

[35] These numbers imply a low value of the derivative of private expenses with respect to subsidies. For $\rho^e = 0.3$, our estimate gives a value of 0.06, positive (no crowding out) but small.

thresholds can differ according to the level of sunk investments.

Subsidies are potentially effective in inducing firms to invest. We estimate that almost 50% of non-performing big firms could be induced to perform innovative activities financing less than 10% of their R&D expenditures, and one out of three non-performing small firms by financing up to 40% of their expenses. We obtain evidence that actual subsidies, in fact, play this role, even if modestly. Almost 14% of the small firms' R&D performing observations (and 3% of the big firms') are estimated to depend on the (expected) subsidy, in the sense that no R&D would be observed in the absence of it. But it must be realized that subsidies go mainly to firms that would otherwise perform innovative activities. This fact, which can be seen as the result of a proper selection of applicants and risk-averse practices of agencies, suggests that public policy tends to neglect the inducing dimension of public support.

On the other hand, subsidies seem to induce only a very slight change in the level of private expenditures chosen by the firms that in any case would perform innovative activities (2-7% plus). Our parameter estimate implies that if projects were not subsidized, they would basically be carried out at the smaller size implied by the absence of public funds. However, this also implies that no crowding out of private funds nor inefficient use of subsidies is observed.

In summary, the availability of public funds seems to make an important difference in the undertaking of projects, and hence in the allocation of private funds to these projects. They also seem to expand firms' expenditures without crowding out or innefficient use, although only by a little more than the amount added to the privately allocated funds.

The employed framework, despite its simplicity, has turned out to be sensible in describing profitability gaps and exploring the impact of subsidies in a context which attepts to be robust to selectivity and endogeneity bias. Among others, two main questions call for further research: 1) the developing of dynamics, improving the ability to describe the different behavior of stable and occasional performers, and in particular the incurring of sunk investments, and 2) the modelling of the ex-post adjustments of firms.

Appendix A

The model can be specified in terms of a demand $q(p, s(K_{t-1}))$ depending on a lag of the stock of "knowledge capital" K, which accumulates according to the rule $K_t = (1-\delta)K_{t-1}+x_t$, where δ stands for depreciation. Dynamic optimization of the model in terms of "knowledge capital," with a discount factor r, gives the rule $\frac{K^*}{p^*q^*} = (\frac{s}{q}\frac{\partial q}{\partial s}\ \frac{K}{s}\frac{\partial s}{\partial K})\ /(-\frac{p}{q}\frac{\partial q}{\partial p}\ E[(1-\rho)^{\beta}](r+\delta))$. In a stationary environment $x \simeq \delta K$, and the static equation in terms of the current R&D expenditure x can be considered an approximation of the equations produced by the "knowledge capital" version.

To see how the model formalizes the effects of resource indivisibility plus input complementarity, think of the following stylized framework. Quality s may be produced with two complementary inputs: C, cumulated knowledge plus laboratories, say, and L, engineers. Resources indivisibility implies, for example, a minimum size for the first input equal to $\overline{C}$. The production function of quality is then

$$s = \begin{cases} s(0) & \text{if } \ C \le \overline{C} \\ \min\{C, L\} & \text{if } \ C > \overline{C} \end{cases}$$

For some production of quality $s > s(0)$, cost can be written $w\overline{C} + w(s - s(0))$, where $w = w_C + w_L$ is the sum of input prices. R&D expenditures x are hence given by the cost function

$$x = \begin{cases} 0 & \text{if } \ s = s(0) \\ F + C_v(s) & \text{if } \ s > s(0) \end{cases}$$

with $C_v(s(0)) = 0$. This is a standard example of cost function when there are non-sunk fixed set-up costs. Inverting this cost function for the likely case of a strictly convex $C_v(s)$ one obtains the $s(x)$ function of the text.

Appendix B.

Model I (Levels's model). Let us write x for $-\ln(1-\rho^e)$, $\mathbf{z}$ for the union of variable sets z and z_ρ, and assume that β_1 is written including the exclusion restrictions. If $\gamma = 0$, equations $e^* = \beta x + z_1\beta_1 + \varepsilon_1$ and $e^* - \overline{e} = \beta x + z(\beta_1 - \beta_2) + v_2$, where $v_2 = \varepsilon_1 - u_2$, are the structural and selectivity equations. We observe $y_1 = e^*$ if $y_2 = 1\left[e^* - \overline{e} > 0\right] = 1$. The partial conditional likelihood for one observation may be written

$$\begin{aligned} L(\theta) &= \left[P(y_2 = 0 \mid \mathbf{z})\right]^{1-y_2}\left[f(y_1 \mid y_2 = 1, \mathbf{z})P(y_2 = 1 \mid \mathbf{z})\right]^{y_2} \\ &= \left[P(y_2 = 0 \mid \mathbf{z})\right]^{1-y_2}\left[P(y_2 = 1 \mid y_1, \mathbf{z})f(y_1 \mid \mathbf{z})\right]^{y_2} \end{aligned}$$

Normality implies $y_1 \mid \mathbf{z} \sim N(\beta x + z_1\beta_1, \sigma_1^2)$ and $y_2 = 1\left[\beta x + z(\beta_1 - \beta_2) + v_2 > 0\right]$ with $v_2 \sim N(0, \sigma^2)$. Conditioning on y_1 and writing $(\delta, \delta_2) = \frac{1}{\sigma}(\beta, \beta_1 - \beta_2)$, $y_2 = 1\left[\delta x + z\delta_2 + r(y_1 - \beta x - z_1\beta_1)/\sigma_1 + \varepsilon_2 > 0\right]$ with $\varepsilon_2 \sim N(0, 1 - r^2)$ and $r = \frac{\sigma_{1v}}{\sigma_1\sigma}$. Notice that σ is identified through the relationship between δ and β. The partial conditional log likelihood for an observation is

$$\begin{aligned} l(\theta) &= (1-y_2)\log(1 - \Phi(\delta x + z\delta_2)) \\ &\quad + y_2\left[\log\Phi\left(\frac{\delta x + z\delta_2 + r(y_1 - \beta x - z_1\beta_1)/\sigma_1}{(1-r^2)^{1/2}}\right) + \log\phi\left(\frac{y_1 - \beta x - z_1\beta_1}{\sigma_1}\right) - \log\sigma_1\right] \end{aligned}$$

Model II (Pseudo differences with latent lag observed). If $\gamma \neq 0$, $e^* = \gamma e^*_{t-1} + \beta\widetilde{x} + \widetilde{z}_1\beta_1 + \varepsilon_1$, where $\widetilde{x}_t = x_t - \gamma x_{t-1}$. This equation now includes a lag of the latent variable, and this is also the case for the decision equation, which becomes $e^* - \overline{e} = \gamma e^*_{t-1} + \beta\widetilde{x} + \widetilde{z}_1\beta_1 - z\beta_2 + \varepsilon_1 - u_2$ or $e^* - \overline{e} = \beta(\widetilde{x} + \frac{\gamma}{\beta}e^*_{t-1} - \gamma z_{1t-1}\frac{\beta_1}{\beta}) + z(\beta_1 - \beta_2) + v_2 = \beta\widetilde{x}_c + z(\beta_1 - \beta_2) + v_2$.

Under our serial independence assumption, e^*_{t-1} constitutes a variable uncorrelated with $(\varepsilon_{1t}, u_{2t})$, and hence a sample selection based on a fixed rule involving e^*_{t-1} does not affect the consistency of the estimation. Consequently, we use two-year subsequences in which the lagged-latent variable is observed, i.e., all the two-year subsequences for which the indicator of performance takes the sequence of values $(1,1)$ or $(1,0)$. The partial likelihood for one observation has the same general form as before, and our assumptions now imply that $y_1 \mid \mathbf{z} \sim N(\gamma e^*_{t-1} + \beta\widetilde{x} + \widetilde{z}_1\beta_1, \sigma_1^2)$, $y_2 = 1\left[\beta\widetilde{x}_c + z(\beta_1 - \beta_2) + v_2 > 0\right]$ with $v_2 \sim N(0, \sigma^2)$, and $y_2 = 1\left[\delta\widetilde{x}_c + z\delta_2 + r(y_1 - \gamma e^*_{t-1} - \beta\widetilde{x} - \widetilde{z}_1\beta_1)/\sigma_1 + \varepsilon_2 > 0\right]$ with $\varepsilon_2 \sim N(0, 1-r^2)$ and $r = \frac{\sigma_{1v}}{\sigma_1\sigma}$. Notice that y_2 is now given by a non-linear model in the parameters, but σ is again identified.

Model III (Differenced differences). Assume that $sign\left[(e^*_t - \overline{e}_t) - \gamma(e^*_{t-1} - \overline{e}_{t-1})\right] = sign(e^*_t - \overline{e}_t)$. This is always the case for subsequences $(1,0)$ and $(0,1)$ and, if γ is not too large, it is a sensible stationarity assumption for differences $e^* - \overline{e}$ which remain positive or negative. Take the set of subsequences with a sequence of values $(0,0)$ or $(1,1)$ or $(1,0)$. The selectivity equation can be rewritten as $(e^*_t - \overline{e}_t) - \gamma(e^*_{t-1} - \overline{e}_{t-1}) = \beta\widetilde{x} + \widetilde{z}(\beta_1 - \beta_2) + \varepsilon_1 - \widetilde{u}_2$, where the lagged-latent is now not necessary. This gives an estimable model (conditional on γ) where $y_1 = e^*_t - \gamma e^*_{t-1}$ is observed when $y_2 = 1\left[(e^*_t - \overline{e}_t) - \gamma(e^*_{t-1} - \overline{e}_{t-1}) > 0\right] = 1$ (we have excluded the subsequences $(0,1)$ because $y_2 = 1$ but y_1 is not observable). The partial likelihood, conditional on γ, may be written similarly to the other models. But notice that $v_{2t} = \varepsilon_{1t} - u_{2t} + \gamma u_{2t-1}$ (a lagged disturbance enters the composite error term) and hence endogenous variables must be lagged twice.

Model IV (Differenced-ordered differences). Suppose only the smaller differences $e^* - \overline{e}$ tend to change occasionally. Write $\beta x + z(\beta_1 - \beta_2) = \mu_t$. Then $(e^*_t - \overline{e}_t) - \gamma(e^*_{t-1} - \overline{e}_{t-1}) = (1 - \gamma L)\mu_t + v_2$, where L is the lag operator. Assume that μ_t tends to persist ($\mu_t \simeq \mu$) and that it is a large positive (negative) value, with respect to the disturbances v_2, presumably linked to an initial investment (to its absence). For subsequences taken from sequences in which all values are one or zeroes it seems reasonable to assume lower and upper bounds as in $(1-\gamma L)\mu_t + v_2 > (1-\gamma)\frac{K}{2} \equiv \alpha$ and $(1-\gamma L)\mu_t + v_2 < -(1-\gamma)\frac{K}{2} \equiv -\alpha$. K is the minimum expected distance which separates the stable performers from the non-performers. Differences of the occasional performers are comprised in the intermediate area, eventually changing their sign. Consider $y^*_2 = \beta\widetilde{x} + \widetilde{z}(\beta_1 - \beta_2) + v_2$, and let y_1 take the self-explaining values {permanent zero, occasional zero, occasional effort, permanent effort} according to the (ordered) values of y^*_2. The partial likelihood can be built in the same way as in the simplest model. The model provides an estimate of the sunk costs $\widehat{K} = \frac{2\widehat{\alpha}}{(1-\widehat{\gamma})}$.

Appendix C: Variable definition and descriptive statistics.

After deleting the firms' data points for which some variable needed in the econometric exercise is missing, we retain a panel with 9,455 observations (and the lagged observations needed for some variables). In what follows, we briefly define the variables employed. Table A1 describes the sample and Table A2 gives some descriptive statistics.

Advertising/sales ratio: advertising and promotional expenditures over sales.

Age: firms' average constituent year (1975) minus the constituent year of the firm (in tens of years).

Average industry patents: yearly average number of patents registered by the firms in the same industry (excluding the patents registered by the firm), for a breakdown of manufacturing in 110 industries.

Capital growth: Real rate of growth of an estimate of the firm's capital in equipment goods and machinery.

Competition changes: dummy variable which takes the value one if the firm reports that a price variation has occurred due to market changes.

Concentrated market: dummy variable which takes the value one if the firm reports that its main market consists of fewer than 10 competitors.

Domestic exporter: Dummy variable which takes the value one if the firm is domestic (less than 50% of foreign capital) and has exported during the year.

Expected subsidy: computed as the product of the predicted probability times the predicted value.

Firm with market power: dummy variable which takes the value one if the firm reports significant market share and the market has fewer than 10 competitors.

Foreign capital: dummy variable which takes the value one if the firm has foreign capital.

Geographical opportunities: dummy variable which takes the value one if the firm has its main plant in the autonomous communities of Madrid, Catalonia or Valencian Country.

Industry dummies: set of 18 industry dummies.

Market share: market share reported by the firm in its main market. Firms are asked to split their total sales according to markets and report their market shares. If a firm reports that its share is not significant, market share is set to zero.

Negative cash flow: dummy variable which takes the value one if sales minus production cost is negative.

Quality controls: dummy variable which takes the value one if the firm reports that it carries out quality controls on a systematic basis.

Recessive market: dummy variable which takes the value one if the firm reports that its main market is in recession.

Region dummies: set of 17 autonomous community (regions) dummies.

R&D effort: ratio of total R&D expenditures to sales. Total R&D expenditures include the cost of intramural R&D activities, payments for outside R&D contracts, and expenditures on imported technology (patent licenses and technical assistance).

R&D effort dummy: dummy which takes the value one if effort is positive.

Skilled labor: dummy which takes the value one if the firm possesses highly qualified workers (engineers and graduates).

Size: number of employees (in hundreds).

Size dummies: set of 6 dummy variables.

Subsidy: ratio of total public subsidies to total R&D expenditures.

Subsidy dummy: dummy which takes the value one if the subsidy is positive.

Technological sophistication: dummy variable which takes the value one if the firm uses automatic machines, or robots, or CAD/CAM, or some combination of these procedures, multiplied by the ratio of engineers and graduates to total personnel.

Time dummies: set of yearly dummy variables.

References

Aghion, P., N. Bloom, R. Blundell, R. Griffith and P. Howitt (2002), "Competition and innovation: An inverted U relationship," WP02/04, Institute for Fiscal Studies.

Almus, M. and D. Czarnitzki (2003), "The effects of public R&D subsidies on firms' innovation activities: The case of Eastern Germany," *Journal of Business and Economic Statistics, 21, 226-236.*

Amemiya, T. (1985), *Advanced Econometrics*, Basil Blackwell.

Arellano, M., O. Bover and J.M. Labeaga (1999), "Autoregressive models with sample selection for panel data," in C. Hsiao et al. (eds.), *Analysis of Panel and Limited Variable models*, Cambridge University Press.

Arrow, K. (1962), "Economic welfare and the allocation of resources for inventions," in R. Nelson (ed.), *The Rate and Direction of Inventive Activity*, Princeton U.Press.

Busom, I. (2000), "An empirical evaluation of the effects of R&D subsidies," *Economics of Innovation and New Technology*, 9, 111-148.

Cohen, W. (1995), "Empirical studies of innovative activity," in P. Stoneman (ed.), *Handbook of the Economics of Innovation and Technological Change*, Blackwell.

Cohen, W. and R. Levin (1989), "Empirical studies of innovation and market structure," in R. Shmalensee and R. Willing (eds.), *H. of Industrial Organization*, North Holland.

David, P., Hall, and Toole, A. (2000), "Is public R&D a complement or substitute for private R&D? A review of the Econometric Evidence," *Research Policy*, 29, 497-529

Dorfman, R., and P. Steiner (1954), "Optimal advertising and optimal quality," *American Economic Review*, 44, 826-836.

Griliches, Z. (1979), "Issues in assessing the contribution of R&D to productivity growth," *Bell Journal of Economics*, pp.92-116.

Gronau, R. (1973), "The effect of children on the housewife's value of time," *Journal of Political Economy*, 81, 168-199.

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

Klette, J., Moen, J. and Z. Griliches (2000), "Do subsidies to commercial R&D reduce market failures? Microeconometric evaluation studies," *Research Policy*, 29, 471-495.

Lach, S. (2002), "Do R&D subsidies stimulate or displace private R&D? Evidence from Israel," *Journal of Industrial Economics*, 50, 369-390.

Maddala, G.S. (1983), *Limited-dependent and qualitative variables in econometrics*, Cambridge University Press.

Nawata, K. and N. Nagase (1996), "Estimation of sample selection bias models," *Econometric Reviews*, 15, 387-400.

Nelson, R. (1959), "The simple economics of basic scientific research," *Journal of Political Economy*, 76, 297-306.

Nelson, F. (1977), "Censored regression models with unobserved stochastic censored thresholds," *Journal of Econometrics*, 6, 309-327.

Pakes, A. and M. Schankerman (1984), "An exploration into the determinants of research intensity," in Z. Griliches (ed.), *R&D, Patents and Productivity*, NBER.

Trajtenberg, M. (2002), "Government support for commercial R&D: Lessons from the Israeli experience," *Innovation Policy and the Economy*, 2, 79-134.

Wallsten, S. (2000), "The effect of government-industry R&D programs on private R&D: the case of the small business innovation research program ," *Rand Journal of Economics*, 1, 82-100.

Wooldridge, J.M. (2002), *Econometric Analysis of cross section and panel data*, MIT Press.

Table 1.- Firms with R&D activities
(percentages of firms)

Year	≤ 200 workers	> 200 workers
1990	17.3	76.6
1991	18.8	75.0
1992	18.0	71.4
1993	18.9	70.1
1994	19.6	74.4
1995	20.2	69.3
1996	20.4	72.1
1997	22.3	71.3
1998	25.6	74.4
1999	26.0	77.0

Table 2.- Firms with R&D activities during the period 1990-1999
(percentages of firms)

Firm size	Stable performers[1]	Occasional performers[2]	All firms
≤ 20 workers	4.1	20.3	24.5
21-50	11.2	23.6	34.8
51-100	19.1	36.3	55.4
101-200	39.1	28.2	67.3
201-500	54.1	31.7	85.8
>500	69.0	20.7	89.7

[1] Firms reporting R&D expenditures every observed year

[2] Firms reporting R&D expenditures some of the observed years

Table 3.- R&D performers granted at least one year during the period 1990-1999
(percentages of firms)

Firm size	Stable performers	Occasional performers	All performers
≤ 20 workers	31.0	9.9	13.5
21 y 50	31.7	16.7	21.5
51-100	43.3	24.6	31.0
101-200	31.6	17.5	25.7
201-500	52.7	26.6	43.1
>500	54.3	23.7	47.3

Table 4.- Average ratios of public funding to R&D expenditures
(subsidy/R&D expenditure, in %, granted firms)

Firm size	Stable performers	Occasional performers	All performers
≤ 20 workers	69.9	65.3	67.5
21 - 50	49.5	57.0	53.1
51-100	53.9	26.0	42.4
101-200	29.5	75.8	38.1
201-500	23.0	47.1	26.6
>500	15.0	42.4	17.3

Table 5.- Total R&D e®ort with and without subsidies
(period averages of non-zero e®orts)

Firm size	Without subsidies	With subsidies
· 20 workers	2.2	4.9
21-50	2.0	3.8
51-100	1.7	5.0
101-200	1.6	3.9
201-500	1.7	3.7
>500	1.8	3.8

Table 6.- Total R&D e®ort with and without subsidies (by year)
(averages of non-zero e®orts)

	· 200 workers		>200 workers	
Year	Without subsidies	With subsidies	Without subsidies	With subsidies
1990	2.3	4.5	1.7	4.2
1991	2.2	4.8	1.7	4.3
1992	2.1	5.6	1.7	3.8
1993	2.1	4.0	1.8	3.6
1994	2.0	4.0	1.9	3.4
1995	1.6	4.2	1.5	4.1
1996	1.9	4.4	1.6	3.3
1997	1.9	3.8	1.8	3.3
1998	1.6	4.3	1.7	3.4
1999	1.6	4.2	1.4	4.1

Table 7.- Estimates of the equations P (½ > 0 j y) and E(ln ½ j ½ > 0; y)
Dependent variable: (indicator function and log of) ½

	Equations with endogenous vars. lagged once (¿ = t ¡ 1)		Equations with endogenous vars. lagged twice (¿ = t ¡ 2)	
	Probability equation[1]	Subsidy equation[1]	Probability equation[1]	Subsidy equation[1]
Constant	-2.83 (-12.7)	-0.40 (-1.3)	-2.62 (-11.4)	-0.67 (1.7)
Abnormal subsidy dummies[2]	-0.79 (-3.8)	2.12 (14.5)	-0.45 (-1.8)	2.33 (14.1)
1(½$_¿$ > 0)[3]	1.89 (23.9)		1.47 (15.4)	
ln[1(½$_¿$ > 0)½$_¿$ + 1(½$_¿$ = 0)][3]		0.38 (8.3)		0.28 (5.2)
1(½$_¿$ = 0)[3]		-0.58 (-5.1)		-0.41 (-3.2)
Size$_{t ¡ 1}$	0.04 (4.3)	-0.02 (-2.7)	0.05 (3.4)	-0.02 (-1.8)
Age	0.04 (2.6)	-0.08 (-3.3)	0.05 (2.5)	-0.12 (-3.3)
Technological sophistication	2.48 (5.7)	-0.48 (-0.8)	2.94 (6.0)	-0.50 (-0.6)
Capital growth$_{¿ +1}$	0.18 (3.3)	0.16 (1.1)	0.09 (1.2)	0.32 (1.5)
Domestic exporter dummy$_{t ¡ 1}$	0.47 (7.8)	0.14 (1.3)	0.50 (7.3)	0.26 (1.7)
Foreign capital dummy	0.17 (2.3)	-0.37 (-3.1)	0.17 (2.0)	-0.40 (-2.5)
Firm with market power dummy$_{t ¡ 1}$	0.03 (0.5)	-0.10 (-1.2)	-0.01 (-0.2)	-0.06 (-0.6)
Industry, region and time dummies[4]	included	included	included	included
¾		0.96		1.02
Estimation method:	Probit	OLS	Probit	OLS
N^o of ¯rms:	2,214	321	1,916	270
N^o of observations:	9,455	727	7,241	571
Correctly predicted observations[5]:				
zeroes	0.84		0.81	
ones	0.83		0.81	
R^2		0.51		0.49

[1] Coe±cients and t-ratios (standard errors robust to heteroskedasticity and serial autocorrelation).
[2] Dummies to account for a total of 33 subsidy coverages higher than yearly expenditure. Included in Probit estimations dated at t ¡ 1 and t ¡ 2, respectively, and in OLS at t.
[3] 1(.) stands for the indicator function.
[4] 17 industry dummies, two particular region dummies (Navarre and Basque Country), and yearly dummies for periods 1992-99 and 1993-99 respectively.
[5] Using 0.055 and 0.065 as critical values respectively.

Table 8a: The e®ect of public funding on R&D decisions:
Alternative estimates of the thresholds model

Dependent variable: (log of and indicator of) R&D e®ort. Estimation method: Maximum Likelihood

Variables	Parameters	Model I Levels (Total sample)	Model I Levels (Latent lag observed)	Model II Pseudo-di®s. (Latent lag observed)	Model IIIa Pseudo-di®s. (Di®erenced di®erences)	Model IIIb[1] Pseudo-di®s. (Di®erenced di®erences)	Model IVa Pseudo-di®s. (Di®.-ordered di®erences)	Model IVb[1] Pseudo-di®s. (Di®.-ordered di®erences)
R&D e®ort equation[2]								
Constant[3]		-4.74(-14.1)	-4.27(-10.8)	-5.18(-6.0)	-4.72(-9.2)	-5.11(-13.9)	-4.20(-8.4)	-4.67(-11.4)
Expected subsidy[4]	¯	2.38(7.1)	2.00(4.6)	1.00(2.0)	1.18(3.9)	1.07(2.0)	1.71(4.5)	1.05(1.6)
Other variables; Size and ind. dummies[5]						(see Table 8b)		
R&D decision equation[2]								
Constant[3]		-2.14(-8.8)	-0.12(-0.4)	-0.33(-0.9)	-4.36(-7.9)	-4.86(-8.0)	-3.57(-7.8)	-3.75(-7.6)
Expected subsidy[4]	± = ¯=¾	6.05(5.4)	1.17(1.3)	0.25(2.0)	4.69(5.9)	5.11(5.0)	3.45(5.9)	4.27(4.2)
Sunk costs	K						1.22 (3.8)	0.61 (1.5)
Other variables; Size and ind. dummies[5]						(see Table 8b)		
	$¾_1$	1.36	1.39	0.91	0.95	0.94	0.89	0.91
	¾	0.39	1.71	3.91	0.25	0.21	0.49	0.25
	$¾_{1v}$	-0.07	-2.15	0.14	-0.03	-0.01	0.00	0.00
	°			0.69	0.50	0.52	0.41	0.38
	r	-0.14	-0.90	0.04	-0.11	-0.05	0.00	0.00
Nº of ¯rms		2214	849	849	1891	1396	1916	1413
Nº of observations		9455	2532	2532	6891	5076	7241	5325
Log-likelihood		-0.989340	-1.731081	-1.454862	-0.780667	-0.773197	-1.194843	-1.186829
Correctly predicted obs.[6]:								
0's				0.74	0.90	0.90	0.87	0.87
1's				0.74	0.75	0.76	0.74	0.75

[1] Endogenous variables used to predict subsidies have been lagged twice.
[2] Coe±cients and t-ratios (standard errors corrected for two-stage estimation and correlation in the score).
[3] Firm with less than 20 workers, eighteenth industry.
[4] Generated regressor ¡ ln(1 ¡ ½b)
[5] Additional set of variables common to all versions of the model. Includes 17 industry dummies and 5 size dummies (see Table 8b).
[6] For Model II, predictions for the critical value which equals the predicted percentages. Modi¯ed critical values predictions give 0.83 in models III and 0.81-0.82 in models IV.

Table 8b.- The effect of public funding on R&D decisions:
Estimate of the thresholds model (pseudo-differences, endogenous lagged twice)

Dep. variable: (log of and indicator function of) R&D effort. Estim. method: Maximum Likelihood

Variables	Parameters[1]	R&D effort[2]	R&D decision[3]	Threshold[2]
Constant[4]		-5.11 (-13.9)	-4.86 (-8.0)	-4.10 (-5.9)
Expected subsidy[5]	$\beta; \delta = \beta/\sigma$	1.07 (2.0)	5.11 (5.0)	
Market share$_{t-1}$		0.27 (1.4)	0.22 (1.0)	0.22 (1.2)
Concentrated market dummy $_{t-1}$		-0.17 (-2.1)	0.20 (2.7)	-0.21 (-2.5)
Advertising/sales ratio$_{t-1}$		1.12 (1.1)	2.81 (1.7)	0.53 (0.5)
Average industry patents		0.12 (3.9)	0.12 (2.5)	0.09 (2.5)
Negative cash flow dummy$_{t-1}$		0.08 (1.0)	-0.19 (-2.7)	0.12 (1.4)
Foreign capital dummy			0.45 (2.6)	-0.09 (-1.4)
Geographical opp.dummy			0.73 (4.0)	-0.15 (-1.6)
Capital growth$_{t-1}$			0.02 (0.2)	-0.00 (-0.2)
Recessive market dummy$_{t-1}$			0.12 (2.2)	-0.02 (-1.4)
Quality controls dummy			0.81 (8.3)	-0.17 (-1.7)
Skilled labour dummy			0.89 (6.6)	-0.19 (-1.7)
Size dummies: 21-50 workers		0.19 (0.8)	0.76 (4.8)	0.03 (0.1)
51-100 workers		0.22 (0.6)	1.20 (4.9)	-0.04 (-0.1)
101-200 workers		0.23(0.8)	2.48(10.2)	-0.28(-0.7)
201-500 workers		-0.05(-0.2)	3.11(12.6)	-0.70(-1.5)
>500 workers		0.22(0.8)	4.19(12.2)	-0.66(-1.1)
Industry dummies		included	included	included
	$\sigma_1; \sigma; \sigma_2$	0.94	0.21	0.97
	$\sigma_{1v}; \sigma_{12}$		-0.01	0.89
	$\gamma = 0.52$			
	$r = -0.05$			

[1]Unless otherwise stated the first column estimates refer to parameters β_1; the second to parameters δ_2 and the third to parameters β_2. Third column estimates are based on $\beta_2 = \beta_1 - \sigma\delta_2$, and standard errors are computed from the delta method.

[2]Coefficients and t-ratios (standard errors corrected for two-stage estimation and correlation in the score). Blank spaces stand for exclusion restrictions.

[4]Firm with less than 20 workers, eighteenth industry.

[5]Generated regressor $-\ln(1-\hat{\rho})$.

Table 9.- The distribution of profitability gaps[1]
(Number and percentage of observations by gap values)

Gaps in %	N^o observations	%
<-2.5	0	0.0
-2.5 to -2	53	1.1
-2 to -1.5	240	4.7
-1.5 to -1	696	13.7
-1 to -0.5	1422	28.0
-0.5 to 0	1163	22.9
0 to 0.5	1069	21.1
0.5 to 1	301	5.9
1 to 1.5	87	1.7
1.5 to 2	34	0.7
2 to 2.5	9	0.2
>2.5	2	0.1
Total observations: 5,076		
Negative gaps: 3,574	Mean of negative gaps: -0.76%	
Positive gaps: 1,502	Mean of positive gaps: 0.39%	

[1]Estimated optimal efforts without subsidy minus estimated threshold efforts.

Table 10.- Subsidies required to engage in R&D[1]
(Percentages of observations by subsidy values)

	≤ 200 workers		> 200 workers	
Trigger subsidy values in %	%	Cumulated %	%	Cumulated %
0-10	3.3	3.3	48.7	48.7
10-20	6.0	9.3	41.3	90.0
20-30	8.1	17.4	6.9	96.9
30-40	13.3	30.7	2.5	99.4
40-50	22.4	53.1	0.6	100.0
50-60	29.5	82.6		
60-70	17.4	100.0		
Total observations: 3,481	Observations : 3,321		Observations : 160	
	Median subsidy: 48.9		Median subsidy: 10.1	

[1]Firms with negative gaps even with currently expected subsidy.

Table 11.- The impact of subsidy withdrawal[1]
(Number of observations and percentages from performing obs.)

	≤ 200 workers		> 200 workers	
Subsidy values in %	Stop doing R&D	%	Stop doing R&D	%
0-10	29	6.8	24	2.0
10-20	5	1.2	10	0.9
20-30	14	3.3		
30-40	7	1.6		
40-50	2	0.5		
50-60	2	0.5		
Total obs.: 93	59	13.8	34	2.9
	Median subsidy: 11.0		Median subsidy: 4.0	

[1]Firms which run into negative gaps when expected subsidy is not accounted for.

Table A1 : Number of firms by time spells and type of R&D performers.

			Non-performers[1]	Stable performers[2]			Occasional performers[3]		
					Mean effort			Mean effort	
N^o of years in sample	N^o firms	Observations	N^o firms	N^o firms	≤200	>200	N^o firms	≤200	>200
1	298	298	145	112	3.1	2.5	41	0.7	0.3
2	503	1006	287	129	2.6	3.1	87	0.9	0.6
3	319	957	159	74	2.0	2.1	86	0.5	0.3
4	186	744	84	56	2.7	2.2	46	0.5	0.4
5	193	965	81	54	2.4	3.3	58	0.5	0.5
6	170	1020	83	39	2.3	2.4	48	0.8	0.6
7	136	952	67	27	2.8	2.7	42	1.0	0.7
8	168	1344	85	18	4.5	3.3	65	0.6	0.7
9	241	2169	102	53	2.3	2.6	86	0.6	0.4
Total	2214	9455	1093	562	2.6	2.7	559	0.7	0.5

[1] Firms reporting zero R&D expenditures every observed year

[2] Firms reporting positive R&D expenditures every observed year

[3] Firms reporting positive R&D expenditures some of the observed years

Table A2.- Variable descriptive statistics

	All observations				Observations with positive R&D			
	Mean	St. dev	Min	Max	Mean	St. dev	Min	Max
Dependent Variables								
R&D effort (×100)	0.78	2.1	0.0	27.5	2.2	3.1	0.0	27.5
R&D effort dummy	0.36	–	0	1				
Subsidy (×100)	2.24	14.7	0.0	440.0	6.31	24.12	0.0	440.0
Subsidy dummy	0.08	–	0	1	0.22	–	0	1
Explanatory Variables								
Adv./sales ratio$_{t-1}$ (×100)	1.37	3.4	0.0	102.4	2.31	4.5	0.0	96.5
Age (×10)	0.79	16.0	-23	35	7.13	16.8	-23	35
Avge. industry patents	0.36	1.1	0.0	21.4	0.59	1.4	0.0	21.4
Capital growth	0.09	0.3	-3.5	7.3	0.10	0.3	-1.7	6.3
Concentrated market dummy$_{t-1}$	0.54		0	1	0.69		0	1
Domestic exporter dummy$_{t-1}$	0.40		0	1	0.53		0	1
Expected subsidy (×100;sub. lagged once)	1.90	4.90	0.01	56.35	4.01	7.39	0.02	56.35
Expected subsidy (×100;sub. lagged twice)	1.97	4.37	0.02	59.23	3.92	6.46	0.02	59.23
Firm with market power dummy$_{t-1}$	0.38		0	1	0.57		0	1
Foreign capital dummy	0.19		0	1	0.39		0	1
Geographical opp. dummy	0.54		0	1	0.60		0	1
Market share$_{t-1}$	0.13	0.2	0	1	0.19	0.2	0	1
Negative cash flow dummy$_{t-1}$	0.15		0	1	0.12		0	1
Quality controls dummy	0.41		0	1	0.68		0	1
Recesive market dummy$_{t-1}$	0.25		0	1	0.23		0	1
Size$_{t-1}$(×100)	168.4	336.2	1	6731	334.5	443.0	1	6731
Skilled labor dummy	0.55		0	1	0.87		0	1
Technological sophistication	0.02	0.05	0	0.52	0.04	0.06	0	0.49
Industry dummies								
Ferrous and non-ferrous metals	0.02		0	1	0.04		0	0
Non-metallic mineral products	0.07		0	1	0.06		0	0
Chemical products	0.07		0	1	0.14		0	0
Metal products	0.11		0	1	0.08		0	0
Agricultural and ind. machinery	0.05		0	1	0.08		0	0
Office and data processing machin.	0.01		0	1	0.01		0	0
Electrical goods	0.07		0	1	0.15		0	0
Motor vehicles	0.04		0	1	0.07		0	0
Other transport equipment	0.02		0	1	0.03		0	0
Meats, meat preparation	0.03		0	1	0.02		0	0
Food products and tobacco	0.11		0	1	0.07		0	0
Beverages	0.02		0	1	0.02		0	0
Textiles and clothing	0.12		0	1	0.07		0	0
Leather, leather and skin goods	0.04		0	1	0.02		0	0
Timber, wooden products	0.07		0	1	0.02		0	0
Paper and printing products	0.08		0	1	0.04		0	0
Rubber and plastic products	0.06		0	1	0.06		0	0
Other manufacturing products	0.01		0	1	0.02		0	0
Region dummies:								
Navarre	0.02		0	1	0.03		0	1
Basque Country	0.07		0	1	0.10		0	1
Size dummies:								
<20 workers	0.33		0	1	0.08		0	1
21-50 workers	0.24		0	1	0.14		0	1
51-100 workers	0.08		0	1	0.08		0	1
101-200 workers	0.09		0	1	0.13		0	1
201-500 workers	0.19		0	1	0.38		0	1
>500 workers	0.08		0	1	0.18		0	1
Time dummies								
1991	0.08		0	1	0.08		0	1
1992	0.11		0	1	0.10		0	1
1993	0.11		0	1	0.11		0	1
1994	0.11		0	1	0.11		0	1
1995	0.12		0	1	0.11		0	1
1996	0.11		0	1	0.11		0	1
1997	0.12		0	1	0.12		0	1
1998	0.13		0	1	0.14		0	1
1999	0.12		0	1	0.13		0	1

Figure 1: The distribution of R&D expenditures

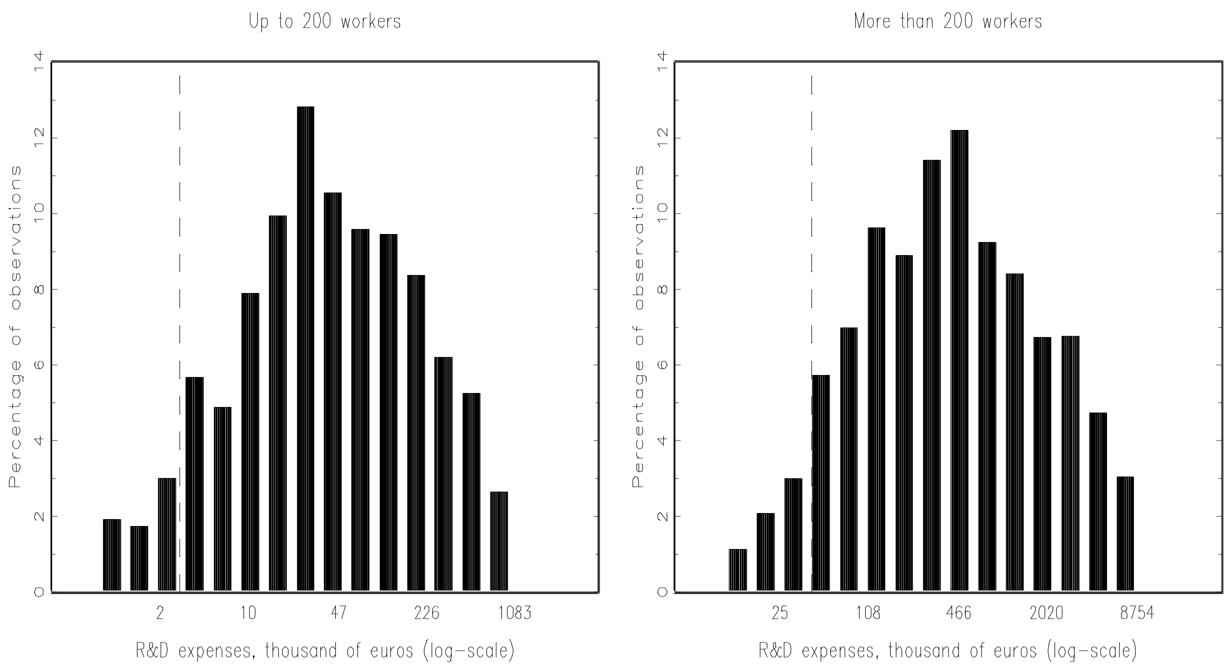
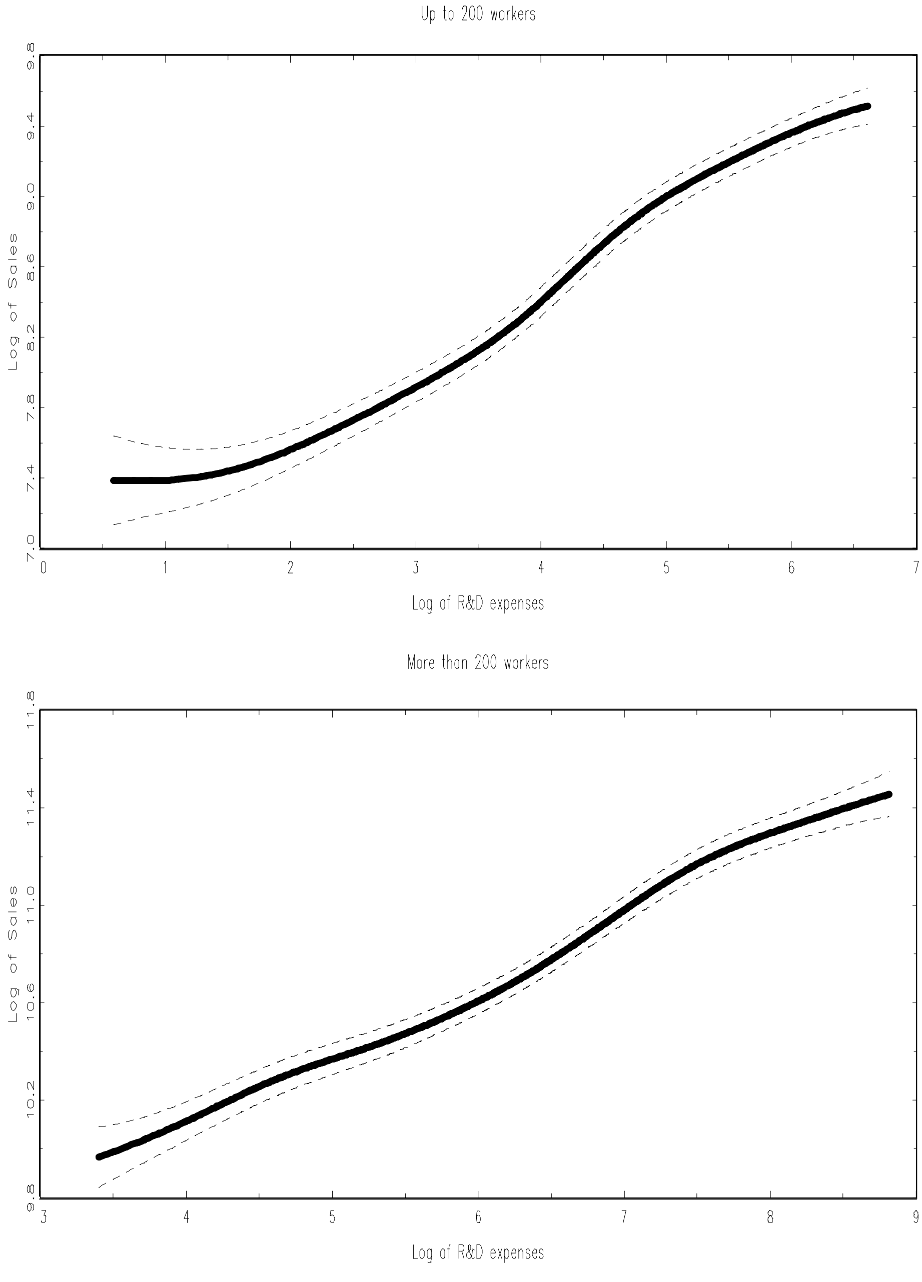
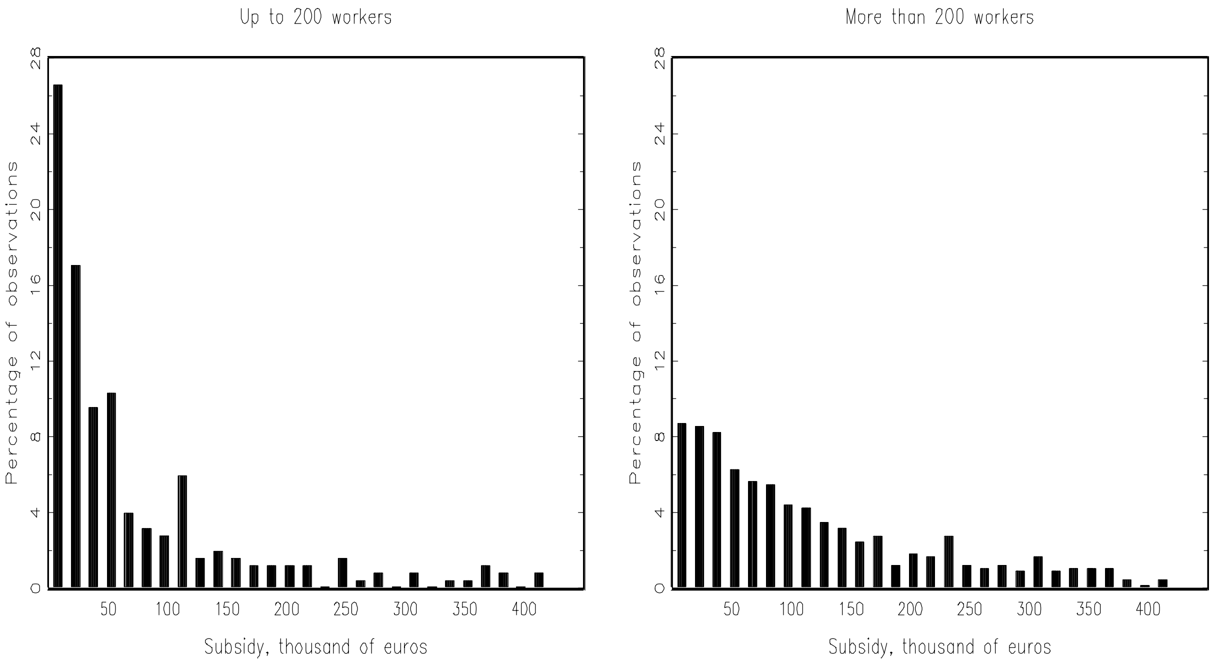


Figure 2: Sales and R&D expenditure[1]



[1] Non-parametric regression using the Nadaraya-Watson estimator, after dropping 5% of values at each tail, with a smoothing parameter $h = 1.7n^{-0.2}$. Dashed bands represent the 95% confidence intervals.

Figure 3: The size distribution of subsidies[1]



[1] Up to 200 workers: depicted 96.4% of the values.
More than 200 workers: depicted 82.2% of the values.

Figure 4: Private R&D expenditures and subsidies

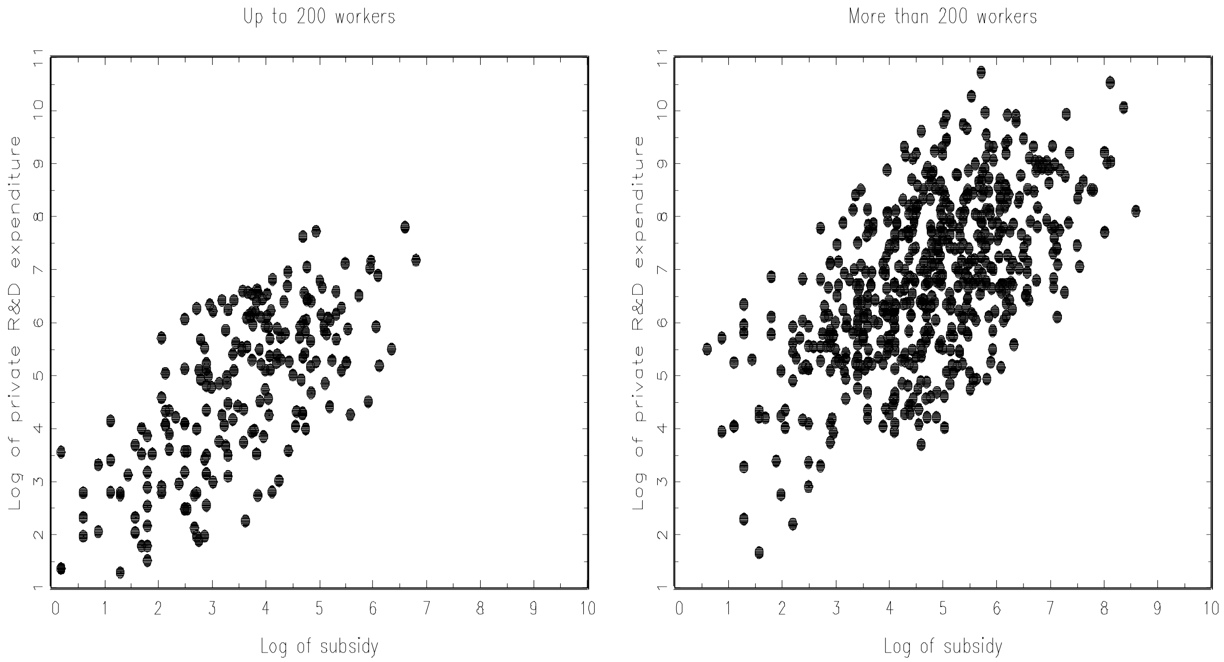


Figure 5: The determination of equilibrium and profits
$\Pi(p^e; x^e) = \max\{\Pi(p^*; x^*); \Pi(p^{**}; 0)\}$

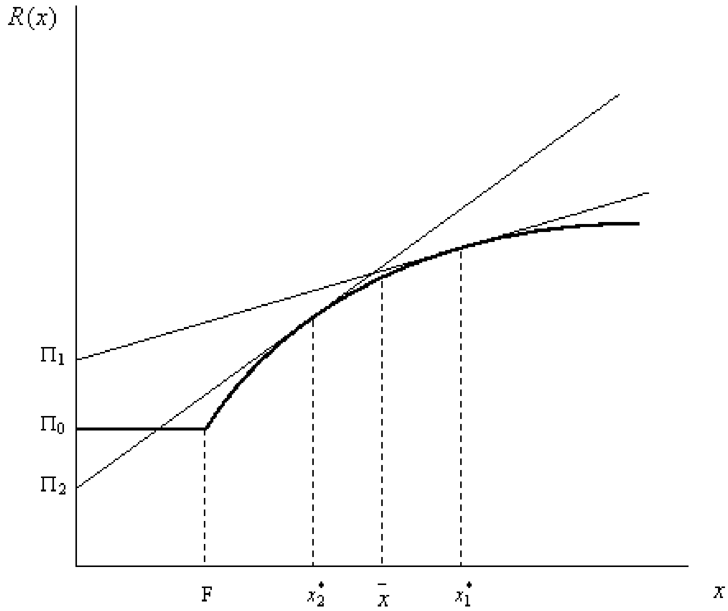


Figure 6: The distribution of profitability gaps