RAND Journal of Economics
Vol. 36, No. 4, Winter 2005
pp. 930–950

Barriers to innovation and subsidy effectiveness

Xulia González*

Jordi Jaumandreu**

and

Consuelo Pazó*

We explore the effects of subsidies by means of a model of firms' decisions about performing R&D when some government support can be expected. We estimate it with data on about 2,000 performing and nonperforming Spanish manufacturing firms. We compute the subsidies required to induce R&D spending, we detect the firms that would cease to perform R&D without subsidies, and assess the change in the privately financed effort. Results suggest that subsidies stimulate R&D and some firms would stop performing in their absence, but most actual subsidies go to firms that would have performed R&D otherwise. We find no crowding out of private funds.

1. Introduction

■ Public sectors of all industrialized countries spend considerable amounts of money to support commercial R&D in manufacturing firms. Firms apply for research grants, and agencies choose the research to be funded. The economic justification for these programs lies in the presumed 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.

* Universidade de Vigo; xgzlez@uvigo.es, cpazo@uvigo.es.

** Universidad Carlos III de Madrid; jordij@eco.uc3m.es.

We are grateful to Manuel Arellano, Isabel Busom, Raquel Carrasco, Miguel Delgado, Bronwyn Hall, Rupert Harrison, Lars-Hendrick Röller, Peter Zemsky, and the audiences at the 2nd CEPR Conference on Applied IO (Lisbon, 2000) and at Universidad Carlos III de Madrid for their suggestions. We thank the Editor, Ariel Pakes, and two anonymous referees for useful comments that have greatly improved the article. We acknowledge support from projects SEC2000-0268 and BEC2002-02527 (CICYT and FEDER), and PGIDT03PXIC300014PN (Xunta de Galicia).

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

Wallsten (2000) estimates a simultaneous model of expenditure and funding for a sample of U.S. firms and claims that, controlling for the endogeneity of grants, no effort effect is detected and that a full crowding-out effect is present. Busom (2000) estimates effort equations for a Spanish sample divided into subsidized and nonsubsidized firms, controlling for selectivity, and concludes that full crowding-out effects cannot be ruled out for 30% of the firms and partial crowding out may be important. In contrast, Lach (2002) estimates the relative increase in R&D expenditures of subsidized versus nonsubsidized firms using panel data on a sample of Israeli companies, and finds that small firms enjoy a positive (dynamic) total effect, whereas this effect fades in the larger firms. Almus and Czarnitzki (2003) also compare the average effort of subsidized East German firms with the effort of similar (in probability of subsidy) nonsubsidized firms in a matched sample, obtaining a significant difference of four percentage points.

The heterogeneity of the results mirrors the diversity of methods and approaches for dealing with the two problems that must be addressed in order to make estimates convincing, namely, the selectivity of subsidy receivers and the endogeneity of subsidies. Furthermore, available datasets often impose severe limits on addressing these problems. For example, many samples include only R&D performers and many show a reduced time dimension.

This article aims to explore the effects of commercial R&D subsidies by focusing on the modelling of firms' decisions when some government support can be expected: whether or not to carry out 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 (R&D performers, subsidized or not, and nonperformers) to identify the model parameters. From the estimates we derive profitability thresholds and gaps for expenditure on innovative activities for every firm. For nonperforming firms, we then 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 carry out R&D if subsidies were eliminated. In addition, we assess subsidy efficiency for the performing firms.

The model considers each firm as a product-differentiated competitor capable of shifting the demand for its product by enhancing product quality through R&D.[2] Demand characteristics, technological opportunities, and setup 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. Firms 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 when determining the size of planned R&D expenditures.

This framework naturally leads to a Tobit-type modelling of a censored variable, which we will call "optimal nonzero 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 are endogenous. We estimate expected subsidies and use them in explaining effort by applying methods for dealing with selectivity and endogeneity in a context that allows for autocorrelated errors.

To estimate the model, we use an unbalanced panel of more than 2,000 Spanish manufacturing firms observed during the period 1990–1999. 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 the primary source of support for innovations. Firm sample behavior is, however, heterogeneous. About 25% of the firms with more than 200 workers, and about 80% of the firms below this size, do not report

[2] Innovative investments shift the demand for the firm product instead of the production function. The model can be taken as a variant of the classical Griliches (1979) R&D "capital" framework.

carrying out formal R&D. Furthermore, only a fraction of performing firms, increasing with firm size, obtain subsidies.

The results contribute a series of interesting empirical findings. On the one hand, a significant proportion of nonperforming firms is estimated as "stimulable" by financing sensible fractions of their expenses, and some real R&D investments are estimated to depend, in fact, on the anticipated public support. But at the same time, most actual subsidies are detected going to firms that would have performed innovative activities even had they not received the subsidy. On the other hand, subsidies seem to induce only a very slight change in the level of private expenditures chosen by the firms that would, in any case, perform innovative activities, but no crowding out of private funds or inefficient use of subsidies is observed. On the whole, Spanish manufacturing subsidies, which amount to 4–5% of R&D expenditure, are estimated to increase total R&D expenditure by 8%. Half of this effect comes from the firms stimulated to perform R&D, which are mainly small firms. Thus, the results suggest that market failures[3] do matter and that subsidies can play a role, and play it effectively, in stimulating R&D activities. However, they also suggest that most subsidies in fact go to firms that would have perfomed R&D anyway and therefore actual public policy may, in part, be neglecting the inducing dimension of public support.

The article is organized as follows. In Section 2 we describe the data and the main facts on innovation activities and subsidies. Section 3 explains our modelling of the firms' R&D decisions. Section 4 presents the estimation procedure and explains how the results are used to measure subsidy effects. In Section 5 we report the results, in Section 6 the subsidy effects, and the conclusions are in Section 7. Two Appendixes detail the econometrics and the data.

2. Data and description

■ The basic dataset is an unbalanced panel of Spanish manufacturing firms surveyed during the 1990s.[4] 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 answers initially represented approximately a self-selected 60% of firms within this size.[5] Our particular sample includes a total of 2,214 firms, observed during the period 1990–1999, 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 centers, 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 1990s, 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.[6] Sample and variable details are given in Appendix B. In what follows, we summarize some facts about R&D expenditures and granted subsidies.

Table 1 and the first two columns of Table 2 report the degree to which Spanish manufacturing firms engage in formal R&D activities. Table 1 shows that the probability of undertaking R&D

[3] We refer to situations in which some R&D investment is not carried out due to its cost, but the addition of the net consumer surplus increase derived from the investment would give a positive global surplus.

[4] The survey was sponsored by the Spanish Ministry of Industry under the name "Encuesta sobre Estrategias Empresariales" (*Survey on Firm Strategies*, available at www.funep.es/esee/esee.asp).

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

[6] Namely, the European Framework program, which reached a very small number of firms; the Ministry of Industry programs, which include the subsidies granted by the specialized agency CDTI (Center for Industrial Technological Development), and the technological actions of regional governments.

TABLE 1 **R&D Activities and R&D Effort (Yearly Averages of Nonzero Efforts)**

	≤200 Workers			>200 Workers		
		Total R&D Effort			Total R&D Effort	
Year	Firms with R&D (%)	Without Subsidies	With Subsidies	Firms with R&D (%)	Without Subsidies	With Subsidies
1990	17.3	2.3	4.5	76.6	1.7	4.2
1991	18.8	2.2	4.8	75.0	1.7	4.3
1992	18.0	2.1	5.6	71.4	1.7	3.8
1993	18.9	2.1	4.0	70.1	1.8	3.6
1994	19.6	2.0	4.0	74.4	1.9	3.4
1995	20.2	1.6	4.2	69.3	1.5	4.1
1996	20.4	1.9	4.4	72.1	1.6	3.3
1997	22.3	1.9	3.8	71.3	1.8	3.3
1998	25.6	1.6	4.3	74.4	1.7	3.4
1999	26.0	1.6	4.2	77.0	1.4	4.1

activities increases sharply with size[7] (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. The first two columns of Table 2 adopt 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 those 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.

Expenditures among the R&D performers are unequal, with a high probability that the observed amounts exceed nonnegligible positive values, which suggests the involvement of setup costs. Figure 1 depicts the (standarized) distributions of the logs of firms' expenditures, keeping the corresponding expenses in thousands of euros as labels.[8] Both distributions tend to fit the standardized normal very well, and hence expenses can be taken as lognormal. The vertical dashed lines point out the modes of the lognormal distributions,[9] with values of about 4 and 54

TABLE 2 **R&D Activities and Subsidies During the Period 1990–1999**

	Firms with R&D (%)		Firms Granted at Least One Year (% of R&D-performers)			Subsidy/R&D Expenditures (in %, granted firms)			Total R&D Effort (averages of nonzero efforts)	
Firm Size	Stable Performers[a]	Occasional Performers[b]	Stable Performers[a]	Occasional Performers[b]	All Performers	Stable Performers[a]	Occasional Performers[b]	All Performers	Without Subsidies	With Subsidies
≤20 workers	4.1	20.3	31.0	9.9	13.5	69.9	65.3	67.5	2.2	4.9
21–50	11.2	23.6	31.7	16.7	21.5	49.5	57.0	53.1	2.0	3.8
51–100	19.1	36.3	43.3	24.6	31.0	53.9	26.0	42.4	1.7	5.0
101–200	39.1	28.2	31.6	17.5	25.7	29.5	75.8	38.1	1.6	3.9
201–500	54.1	31.7	52.7	26.6	43.1	23.0	47.1	26.6	1.7	3.7
>500	69.0	20.7	54.3	23.7	47.3	15.0	42.4	17.3	1.8	3.8

[a] Firms reporting R&D expenditures every observed year.

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

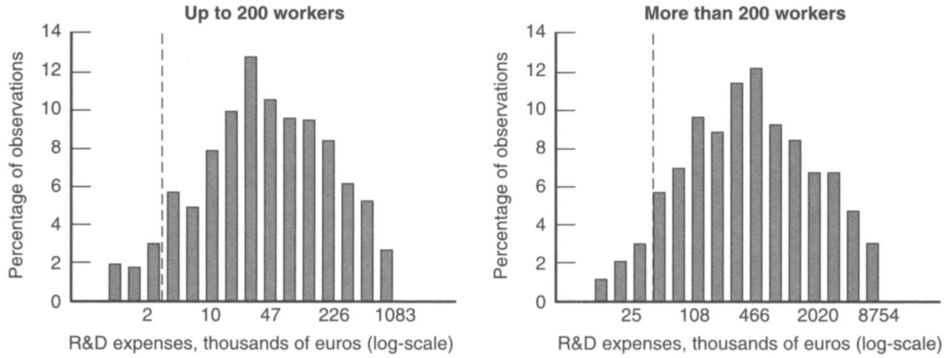
[7] In all tables, "size" reflects the first year the firm is in the sample.

[8] 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).

[9] If x ~lognormal (μ, σ^2), mode$(x) = e^{\mu - \sigma^2}$. According to their means and standard deviations, we assume distributions to be lognormal$(3.85, 1.57^2)$ and $(6.15, 1.47^2)$.

FIGURE 1

THE DISTRIBUTION OF R&D EXPENDITURES



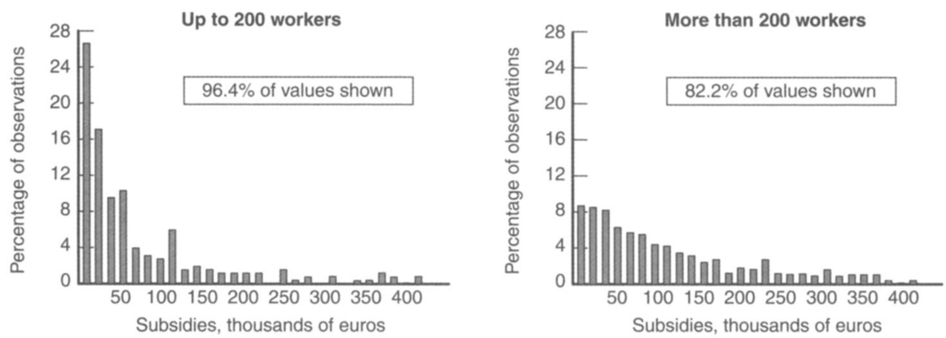
thousand euros. Take these values as a descriptive measure (among those possible) of "critical" expenditure values (associated probabilities of observing lower expenditures are small, 5.8 and 7.1%, respectively). To assess their importance in relative terms, we average observed minimum industry sales 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 .8 percentage points, respectively. Absolute critical expenditures for the smallest firms are smaller, but they appear to be higher in relative terms.

Table 2 reports the main facts about grants. Columns 3 to 5 show 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 2 depicts the distribution of the subsidy amounts. Many subsidies are small, but the spread is also important. Columns 6 to 8 of Table 2 show that the typical subsidy covers between 20% and 50% of R&D expenditures and also that the rate of subsidized expenditure is inversely related to firm size (particularly for the stable performers).

Table 1 and the two final columns of Table 2 also provide a first look at the relationship between subsidies and effort, based on the comparison of the R&D effort of subsidized and nonsubsidized performers' data. Both tables show a positive association between the granting of subsidies and R&D effort, during the period as a whole and from year to year. The data show more than "additionality," in the sense that subsidized efforts minus the part of these efforts attributable to subsidies are higher than nonsubsidized efforts. Figure 3 provides a first look at the relationship between the privately financed expenditure and the amount of the subsidy for subsidized firms.[10]

FIGURE 2

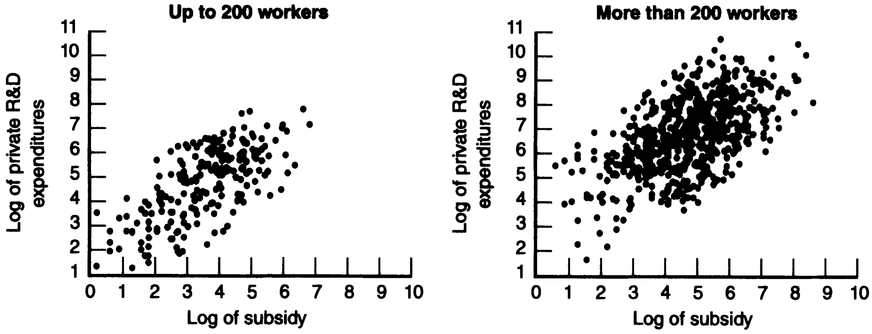
THE SIZE DISTRIBUTION OF SUBSIDIES



[10] Representation is carried out by dropping the subsidies higher than their associated yearly R&D expense values (see Section 5) and 2.5% of subsidy values at each tail.

FIGURE 3

PRIVATE R&D EXPENDITURES AND SUBSIDIES



According to the figure, private expenses tend to show a unit elasticity with respect to public funds.

Therefore, the data suggest nonnegative and even positive R&D effects of subsidies. However, this could be the consequence solely of other omitted variables or because of the two-way nature of the relationship: firms with more R&D are more likely to receive subsidies, and the larger the subsidies, the higher the R&D expenses. Only the development of an econometric analysis can provide further insight into this relationship, by providing evidence as to how these data patterns can be interpreted in terms of "causal" effects.

3. A model with barriers to R&D

■ **R&D with setup costs.** Let $R(x)$ be firm net revenue as a function of R&D expenditure (subindexes are dropped for simplicity).[11] R&D affects revenue positively at a nonincreasing pace, i.e., $\partial R/\partial x > 0$ and $\partial^2 R/\partial x^2 \leq 0$, but only if x surpasses setup costs F. To decide the pertinence and level of R&D expenditures, the firm maximizes the expected profits $E[R(x) - (1-\rho)^\beta x]$, where ρ is the fraction of R&D expenditure that is subsidized, and E indicates the expectation over ρ values.[12] We allow for public funds to be associated with either a higher or lower level of expenditure efficiency[13] by means of parameter β.

Equilibrium admits the straightforward representation of Figure 4. Isoprofit curves are linear with a slope equal to the (expected) effective cost of R&D, $E[(1-\rho)^\beta]$, and the firm's 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 x_1^* or x_2^*. Define $\bar{x}$ as the expenditure level that makes the firm indifferent to performing R&D or not (the tangent of R at this point, not shown in the figure, crosses the y-axis at Π_0).

Under fairly general conditions there is an effort by both performing and nonperforming firms, which we will call optimal nonzero effort, that can be summarized in the (Dorfman and Steiner (1954)-type) unique expression

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

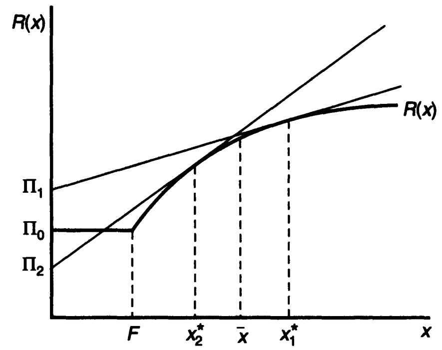
[11] Net revenue can be, for example, $R(x) = \max(p-c)q(p,x)$, where p stands for output price, c for (constant) marginal cost, and q the demand for the firm's output.

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

[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 hand, public funds can be mainly viewed as easing liquidity constraints and allowing for less financing discipline, which implies less expenditure efficiency.

FIGURE 4

THE DETERMINATION OF EQUILIBRIUM AND PROFITS $\Pi(x^{eq}) = \max\{\Pi(x^*), \Pi(0)\}$



and that will be observed if it surpasses the threshold effort which corresponds to $\overline{x}$.[14] Formula (1) shows that optimal nonzero effort increases with the elasticity of demand with respect to R&D expenditure and with the degree of market power (the inverse of the price elasticity). The numerator can be decomposed into the elasticity of demand with respect to quality (demand conditions) and the elasticity of quality with respect to R&D (technological opportunities). "Lack of appropriability," as a factor that discourages R&D, can be easily discussed in this framework.[15] (Expected) subsidies have two different potential effects: they can induce firms to perform R&D and they can enhance the R&D of the firms that would perform innovative activities in any case.

□ **Econometric model.** Let e^* and $\overline{e}$ stand for the logs of optimal nonzero effort and threshold effort, respectively. Starting from (1) we assume

$$e^* = -\beta \ln(1 - \rho^e) + z_1\beta_1 + w \qquad (2)$$

$$\overline{e} = z\beta_2 + u_2 \qquad (3)$$

$$\rho^e = E(\rho|z_\rho) = g(z_\rho, \lambda), \qquad (4)$$

where e^* is observed only when $e^* - \overline{e} > 0$, ρ^e is the expectation for 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 $\mathrm{Cov}(\varepsilon_1, u_2) = \sigma_{12}$.

The effort equation (2) is obtained by taking logs in (1), substituting $\beta \ln[1 - E(\rho|z_\rho)]$ for $\ln E[(1-\rho)^\beta \mid z_\rho]$,[16] 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 (1), but elasticities are endogeneous unobservable variables that we replace with a set of reduced-form determinants,[17] (i.e., exogenous or predetermined variables

[14] R&D level expenditures x and R&D effort x/pq can be used interchangeably because the model and assumptions imply that effort increases monotonically with x for a given firm.

[15] 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 nonzero effort below the threshold effort.

[16] An expansion of $(1-\rho)^\beta$ around $E(\rho)$ shows that $\ln E[(1-\rho)^\beta] \simeq \beta \ln[1 - E(\rho)] + \ln[1 + (1/2)\beta(\beta - 1)c_V^2]$, where c_V is the coefficient of variation of $(1-\rho)$, i.e., $c_V = [\mathrm{var}(1-\rho)]^{1/2}/E(1-\rho)$. The second term of this expression is likely to be small, of order $(1/2)\beta(\beta-1)E(\rho^2)$ and, under certain circumstances, constant.

[17] We assume the standard account of determinants of innovative activities to be underlying these elasticities. See, for example, Pakes and Schankerman (1984), Cohen and Levin (1989), or Cohen (1995).

with respect to (ε_1, u_2)).[18] The autocorrelated disturbance w takes into account that we are not likely to be able to fully specify optimal nonzero effort determinants.

Equation (3) 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 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. Here we are assuming that the error term u_2 is independent and identically distributed over time.

Equation (4) 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. Notice that we make the strong assumption that we observe all variables relevant for the expectation. In particular, any agency evaluation of firm conditions is anticipated through firm attribute indicators. The function is likely to be highly nonlinear, and z_ρ is (possibly) only partially overlapping with z.

Equations (2)–(4) define a rather standard (type-II or thresholds) Tobit model.[19] R&D performance, and hence observation of e^*, is determined by the sign of $e^* - \bar{e}$ (selectivity or decision equation). Amemiya (1985) discusses alternative identification conditions of this 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. This condition arises naturally in our model, where expected subsidies can be safely excluded from the determinants of thresholds.[20] But the model also has some nonstandard features.

First, 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.

Second, we have the unobservable ρ^e. Observed subsidies ρ are granted by agencies according to, among other things, the contemporary effort and performance of firms and hence are presumably endogenous (their values are likely to be correlated with the random term ε_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 because ρ^e is unobservable, we need to substitute the generated regressor $g(z_\rho, \widehat{\lambda})$ for the expectation.

4. Estimating the model and measuring subsidy effects

■ **Estimation procedure.** Estimation is carried out through 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 $\rho^e = E(\rho \mid z_\rho) = g(z_\rho, \lambda)$, we decompose the expectation as follows:

$$\rho^e = E(\rho \mid z_\rho) = P(\rho > 0 \mid z_\rho) E(\rho \mid z_\rho, \rho > 0), \tag{5}$$

where $P(\rho > 0 \mid z_\rho)$ stands for the conditional expectation of receiving 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

[18] 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.

[19] Econometric models of censored variables with stochastic thresholds date back to Gronau (1973) and Nelson (1977).

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

natural "rationality" or "correctness" restrictions on the expectations to estimate the $E(\rho \mid z_\rho)$ function. On the one hand, we assume that firms that 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 that they are going to receive a subsidy). Consequently, we use the grants observed in the whole sample to estimate the conditional probability function.[21] The expected subsidy function can be computed from estimates on these two expectation functions.

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

Substituting $\widehat{\rho^e}$ for ρ^e in the effort equation, we can estimate the Tobit model by partial maximum likelihood. The likelihood is based on the specification of the joint density associated with ε_1 and $v_2 = \varepsilon_1 - u_2$ (which are the disturbances of the pseudo-differences of the effort equation and the disturbances of the decision equation; see Appendix A). The allowance for serial correlation in the disturbances of the effort equation has come at the price of the presence of the partially unobservable variable e^*_{t-1} among the explanatory variables of both equations. We are going to explore the results and insights provided by approaching this problem in three ways (see Appendix A for details).

If disturbances of (2) are assumed not to be autocorrelated ($\gamma = 0$), the terms in e^*_{t-1} disappear, and parameters β, β_1, and β_2 can be estimated by 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 we must include the lagged-latent variable e^*_{t-1}. But the value of e^*_{t-1} is not observed for many firms' data points (when observed effort at $t-1$ is zero). Estimates must then rely on the remaining data points, which constitute an (exogenously selected) sample consisting of the firms' observations with positive effort at $t-1$.[22] Selection here is 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 small proportion of observations of current nonperformance ("zeros"), which in addition correspond only to firms that discontinue R&D at precisely that moment ("stopping zeros"). Consistency is reached at a high price in estimation efficiency.

We assume that efficiency in estimation can be improved by using more "zeros." One way is to reformulate the model in such a manner that we do not need to observe the lagged-latent variable. This can be 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 (it 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 when positive and negative differences $e^* - \overline{e}$ tend to remain unchanged. The assumption is more likely if γ is not very large, but if it does not hold, its violation will be a source of bias in estimation. This is model III. We expect it to contribute a large reduction in variance with a negligible bias. On the estimation side, 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.

Models I to III 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 possible correlations between the scores at different periods of time (see, for example, Wooldridge, 2002). Maximum-likelihood estimation

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

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

is carried out through a grid covering 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.

□ **Measuring profitability gaps and subsidy effects.** After the estimation of the model, we are ready to compute individual optimal nonzero effort and threshold estimates, then use them to assess the effects of subsidies. We will do this relying on the nonstochastic components of the equations, that is, evaluating the relationships at the (zero) expected value of the disturbances. Let us distinguish between these gaps and the gaps that average the unobserved heterogeneity.[23] The model predicts R&D performance using the first gaps, and we choose to base our measures on these gaps. We also report values for the second gap measure.

Let us define profitability gaps. These are the difference between the optimal nonzero effort in the absence of subsidy and threshold effort. If negative, they provide the R&D effort wherein the firm falls short of undertaking profitable innovative activities. If positive, they provide the R&D effort 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 nonperforming 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 for which this expression, evaluated at the estimated expected subsidy, is negative.

Let us now evaluate the role of a subsidy withdrawal. Some firms are likely to be carrying out 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).

The above refers to the ability of subsidies to induce firms (potentially or effectively) to invest in R&D. But how, according to the model, do subsidies change the expenditure of firms that perform innovative activities? First, 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.[24] Second, changes in effort depend on subsidies in a complex way, because all the elasticities in (1) may change with the firm's equilibrium. We will use an approximate measure of the change in effort that becomes exact in the simplest case in which elasticities remain constant.

Call $\varepsilon(\rho^e)$ total effort with subsidy and $\varepsilon(0)$ total effort in its absence. Write $(1-\rho^e)\varepsilon(\rho^e)$ for private effort when R&D is subsidized. It is easy to check that

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

Therefore, if subsidy efficiency β is unity, private effort will remain the same, which means that privately financed expenditures will increase at the same pace as sales. In contrast, if β exceeds unity, the subsidy will increase private effort, and total effort will be higher than the sum of the public fraction and the private effort without subsidy. If β were less than unity, private effort would be reduced. 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).[25]

[23] $E[\exp(e^*) - \exp(\overline{e}) \mid z, w = u_2 = 0] = \exp[E(e^* - \overline{e})]$, which gives the level values corresponding to the (zero) expected value of the disturbances, and $E[\exp(e^*) - \exp(\overline{e}) \mid z]$, which also averages the unobserved heterogeneity.

[24] 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 be possible only with a more complete specification of the demand.

[25] We can compute an average subsidy effect of this type by evaluating at some point the first term of the right-hand side of the identity below (where S is a shorthand for sales): $\frac{(1-\rho^e)x(\rho^e)-x(0)}{\rho^e x(\rho^e)-0} = \frac{(1-\rho^e)\varepsilon(\rho^e)-\varepsilon(0)}{\rho^e\varepsilon(\rho^e)} + \frac{\varepsilon(0)}{\rho^e\varepsilon(\rho^e)}\frac{S(\rho^e)-S(0)}{S(\rho^e)}$.

5. Empirical specification and results

■ **Expected subsidies.** We estimate the unobservable firms' expectations ρ^e using the probit and OLS specification of (5). Recall that we want to predict the expected outcome by means of a set of variables that can be considered exogenous or, at least, predetermined. This will be explained in the following section. Details on all the employed variables can be found in Appendix B.

First of all, subsidies 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 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 the subsidy is zero, and a dummy that takes the value one when this is the case.[26]

We use the same set of additional variables to estimate both equations. We first include a series of the firm's 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 capital (in equipment goods and machinery) growth. We then include three indicators of situations for the firm 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 lagged one period; others are assumed to be strictly exogenous or predetermined longer in advance.[27] 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 3 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 zeros 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 large, mature, technologically sophisticated and expanding firms, as well as the domestic exporters, are more likely to obtain grants for their innovative activities, 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 about 2%, with a standard deviation of 4%.[28] 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%).

□ **Tobit model.** Let us now detail the specifications of equations (2) and (3). According to the model, there are three main types of variables to be considered: indicators of market

---

[26] 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.

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

[28] These are the values obtained using the last two columns of Table 3.

TABLE 3 **Estimates of the Equations $P(\rho > 0 \mid y)$ and $E(\ln \rho \mid \rho > 0, y)$ Dependent Variable: (indicator function and log of) ρ**

	Equations with Endogenous Variables							
	Lagged Once ($\tau = t - 1$)				Lagged Twice ($\tau = t - 2$)			
	Probability Equation[a]		Subsidy Equation[a]		Probability Equation[a]		Subsidy Equation[a]	
Constant	−2.83	(−12.7)	−.40	(−1.3)	−2.62	(−11.4)	−.67	(1.7)
Abnormal subsidy dummy[b]	−.79	(−3.8)	2.12	(14.5)	−.45	(−1.8)	2.33	(14.1)
$1(\rho_\tau > 0)$[c]	1.89	(23.9)			1.47	(15.4)		
$\ln[1(\rho_\tau > 0)\rho_\tau + 1(\rho_\tau = 0)]$[c]			.38	(8.3)			.28	(5.2)
$1(\rho_\tau = 0)$[c]			−.58	(−5.1)			−.41	(−3.2)
Size_{t-1}	.04	(4.3)	−.02	(−2.7)	.05	(3.4)	−.02	(−1.8)
Age	.04	(2.6)	−.08	(−3.3)	.05	(2.5)	−.12	(−3.3)
Technological sophistication	2.48	(5.7)	−.48	(−.8)	2.94	(6.0)	−.50	(−.6)
Capital $\text{growth}_{\tau+1}$	.18	(3.3)	.16	(1.1)	.09	(1.2)	.32	(1.5)
Domestic exporter dummy_{t-1}	.47	(7.8)	.14	(1.3)	.50	(7.3)	.26	(1.7)
Foreign capital dummy	.17	(2.3)	−.37	(−3.1)	.17	(2.0)	−.40	(−2.5)
Firm with market power dummy_{t-1}	.03	(.5)	−.10	(−1.2)	−.01	(−.2)	−.06	(−.6)
Industry, region and time dummies[d]	included		included		included		included	
σ			.96				1.02	
Estimation method	Probit		OLS		Probit		OLS	
Number of firms	2,214		321		1,916		270	
Number of observations	9,455		727		7,241		571	
Correctly predicted observations[e]								
0's		.84				.81		
1's		.83				.81		
R^2			.51				.49	

[a] Coefficients and t-ratios (standard errors robust to heteroskedasticity and serial autocorrelation).
[b] Dummy 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.
[c] $1(\cdot)$ stands for the indicator function.
[d] 17 industry dummies, 2 particular region dummies (Navarre and Basque Country), and yearly dummies for periods 1992–1999 and 1993–1999 respectively.
[e] Using .055 and .065 as critical values respectively.

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 setup costs and the heterogeneity of thresholds among firms. Obviously, no variable can claim to pick up exclusively 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 nonzero efforts and the thresholds. This happens partly because we have to rely on proxies, 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.

The main variables included in both equations are: the firms' market share and a dummy variable representing concentrated markets (both lagged one period) as indicators of market power/competition conditions; the advertising/sales ratio (lagged) and average industry patents (excluding the patents obtained by the firm) as indicators of a high sensitivity of demand with respect to product quality and/or product quality with respect to R&D; and a dummy variable that takes the value one for the firms with (lagged) negative cash flow, to represent serious financial difficulties in carrying out innovation activities.

Six variables are included exclusively in the decision equation to account for setup costs. The list consists of the following indicators: presence of foreign capital, location in regions in which the

TABLE 4 The Effect of Public Funding on R&D Decisions: Alternative Estimates of the Thresholds Model
Dependent Variable: (log of and indicator of) R&D effort

Variables	Parameters	Model I Levels (Total sample)	Model I Levels (Latent lag observed)	Model II Pseudo-diffs. (Latent lag observed)	Model IIIa Pseudo-diffs. (Differenced differences)	Model IIIb[a] Pseudo-diffs. (Differenced differences)
		Estimation Method: Maximum Likelihood				
R&D decision equation[b]						
Constant[c]		−4.74 (−14.1)	−4.27 (−10.8)	−5.18 (−6.0)	−4.72 (−9.2)	−5.11 (−13.9)
Expected subsidy[d]	β	2.38 (7.1)	2.00 (4.6)	1.00 (2.0)	1.18 (3.9)	1.07 (2.0)
Other variables; size and industry dummies[e]						(see Table 5)
R&D decision equation[b]						
Constant[c]		−2.14 (−8.8)	−.12 (−.4)	−.33 (−.9)	−4.36 (−7.9)	−4.86 (−8.0)
Expected subsidy[d]	$\delta = \beta/\sigma$	6.05 (5.4)	1.17 (1.3)	.25 (2.0)	4.69 (5.9)	5.11 (5.0)
Other variables; size and industry dummies[e]						(see Table 5)
	σ_1	1.36	1.39	.91	.95	.94
	σ	.39	1.71	3.91	.25	.21
	σ_{1v}	−.07	−2.15	.14	−.03	−.01
	γ			.69	.50	.52
	r	−.14	−.90	.04	−.11	−.05
Number of firms		2,214	849	849	1,891	1,396
Number of observations		9,455	2,532	2,532	6,891	5,076
Log-likelihood		−.989340	−1.731081	−1.454862	−.780667	−.773197
Correctly predicted observations[f]						
0's				.74	.90	.90
1's				.74	.75	.76

[a] Endogenous variables used to predict subsidies have been lagged twice.
[b] Coefficients and t-ratios (standard errors corrected for two-stage estimation and correlation in the score).
[c] Firm with fewer than 20 workers, 18th industry.
[d] Generated regressor—$\ln(1 - \widehat{\rho^e})$.
[e] Additional set of variables common to all versions of the model. Includes 17 industry dummies and 5 size dummies (see Table 5).
[f] For model II, predictions for the critical value which equals the predicted percentages. Modified critical values predictions give .83 in models IIIa and IIIb.

accumulation of private and public technological assets ensures higher spillovers (geographical opportunities), capital growth, a market that 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 setup expenses, and some of them also with a high sensitivity of the demand with respect to quality.

In addition, in both equations we include a set of dummy variables of size (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.

Table 4 reports the results of carrying out the estimation of the different versions of the model. Samples change for two reasons, according to the estimated version: the "usable" time observations[29] and the exogenous selections performed in each case. Variables, on the other hand, are always kept the same (although lags used to predict expected subsidies are different for model IIIb).

[29] 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.

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 in Table 4 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 sample and when the selected sample used next to obtain a consistent estimate is employed. 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 ($\gamma = .69$) and shows a dramatic change in the coefficient value, which falls to unity with autocorrelated residuals controlled for. However, as discussed above, model II provides a consistent estimate but at the price of constraining the sample to observations for which the latent variable past value is observed. This induces a considerable loss of efficiency, which is in fact apparent in the σ estimate and the variances-covariances of the remaining parameter estimates (not shown in the table). Model II uses scarcely a fourth of the available observations and includes a scant 13% of zero-effort observations.

Model III provides an interesting alternative for the estimation of parameter β. The parameter estimate is 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 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 and consistency are based, holds *ex post* in 96.5% of the cases. Moreover, coefficients are sensible (see Table 5 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 values for the firms obtaining subsidies and zeros 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 .60 and .69, respectively. Uncertainty about subsidies is probably a key question outside of the largest firms.

Table 4 (bottom) reports the results of comparing the models' predictions with the actual observations in the sample. All models except for model II behave sensibly, even when keeping the standard .5 critical value for prediction.[30]

Table 5 shows all the results of model estimation. Let us interpret the estimates. Market power clearly influences effort and thresholds, although the effect of the firm's market share is somewhat imprecisely estimated. In any case, the impact of market share must be balanced against the degree of rivalry. For any given market share, R&D effort is greater when the environment is more competitive. This is consistent with the evidence of inverted U-shaped relationships between product market competition and innovative activities (see, e.g., Aghion et al., 2005).[31] 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 that show a positive impact on effort. Nevertheless, it also performs as an indicator of the corresponding setup 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 inclusion of the list of firm characteristics to pick up threshold effects shows 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 on

---

[30] For model II, highly unbalanced in terms of ones and zeros, it is better to compute prediction with an adjusted critical value that equals the prediction outcomes. The rest of the models can also be compared in terms of adjusted critical values (see the table's footnote f).

[31] 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.

TABLE 5 **The Effect of Public Funding on R&D Decisions: Estimate of the Thresholds Model (Pseudo-Differences, Endogenous Lagged Twice)**
Dependent Variable: (log of and indicator function of) R&D effort

		Estimation Method: Maximum Likelihood		
Variables	Parameters[a]	R&D Effort[b]	R&D Decision[b]	Threshold[b]
Constant[c]		−5.11 (−13.9)	−4.86 (−8.0)	−4.10 (−5.9)
Expected subsidy[d]	$\beta, \delta = \beta/\sigma$	1.07 (2.0)	5.11 (5.0)	
Market share$_{t-1}$		.27 (1.4)	.22 (1.0)	.22 (1.2)
Concentrated market dummy$_{t-1}$		−.17 (−2.1)	.20 (2.7)	−.21 (−2.5)
Advertising/sales ratio$_{t-1}$		1.12 (1.1)	2.81 (1.7)	.53 (.5)
Average industry patents		.12 (3.9)	.12 (2.5)	.09 (2.5)
Negative cash flow dummy$_{t-1}$		.08 (1.0)	−.19 (−2.7)	.12 (1.4)
Foreign capital dummy			.45 (2.6)	−.09 (−1.4)
Geographical opportunity dummy			.73 (4.0)	−.15 (−1.6)
Capital growth$_{t-1}$			.02 (.2)	−.00 (−.2)
Recessive market dummy$_{t-1}$			.12 (2.2)	−.02 (−1.4)
Quality controls dummy			.81 (8.3)	−.17 (−1.7)
Skilled labor dummy			.89 (6.6)	−.19 (−1.7)
Size dummies: 21–50 workers		.19 (.8)	.76 (4.8)	.03 (.1)
51–100 workers		.22 (.6)	1.20 (4.9)	−.04 (−.1)
101–200 workers		.23 (.8)	2.48 (10.2)	−.28 (−.7)
201–500 workers		−.05 (−.2)	3.11 (12.6)	−.70 (−1.5)
>500 workers		.22 (.8)	4.19 (12.2)	−.66 (−1.1)
Industry dummies		included	included	included
	$\sigma_1, \sigma, \sigma_2$	.94	.21	.97
	σ_{1v}, σ_{12}		−.01	.89
	$\gamma = .52$			
	$r = -.05$			

[a] 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.
[b] Coefficients and t-ratios (standard errors corrected for two-stage estimation and correlation in the score). Blank spaces stand for exclusion restrictions.
[c] Firm with fewer than 20 workers, 18th industry.
[d] Generated regressor $-\ln(1-\widehat{\rho^e})$.

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

6. Profitability gaps and subsidy effects

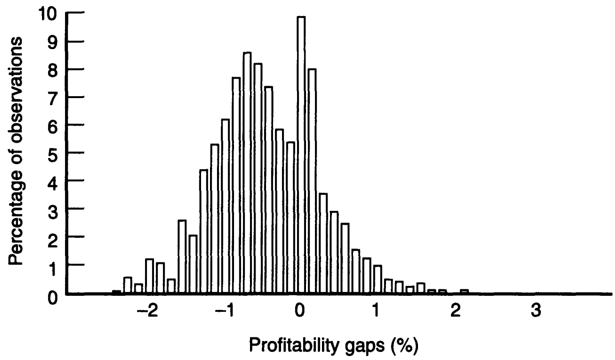
■ Figure 5 depicts the distribution of the estimated profitability gaps (see also the numbers in the web Appendix). Positive gaps represent about 30% and their mean is around .4%, while the absolute value of the negative gaps mean is about .8%. Positive gaps show less heterogeneity (90% lie in the (0,1) 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.[32]

Table 6 further details gap heterogeneity by reporting the distribution of trigger subsidies for the nonperforming firms. Subsidies required to induce firms to engage in R&D are smaller for the largest firms and bigger for the smallest ones. With an expected funding of less than 10% of R&D expenditures, almost 50% of the big nonperforming firms will switch to performing innovative activities. In contrast, inducing 30% of the small firms to carry out R&D implies

[32] In the distribution $\exp(z\widehat{\beta_1} + (1/2)\widehat{\text{var}}(w)) - \exp(z\widehat{\beta_2} + (1/2)\widehat{\text{var}}(u_2)) = 1.83\exp(z\widehat{\beta_1}) - 1.61\exp(z\widehat{\beta_2})$, positive gaps represent about 35%, with a mean of .8%, and the average of negative gaps gives an absolute value of about 1.1%.

FIGURE 5

THE DISTRIBUTION OF PROFITABILITY GAPS



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 6 also reports the impact of subsidy withdrawal on performing firms and the expected subsidies that characterize the firms that presumably abandon R&D. Interestingly enough, subsidy withdrawal would induce the cessation of 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 expected funding to be more important. These results suggest that not all funding is allocated to firms that have positive profitability gaps and would carry out R&D activities even in the absence of public financing, thus indicating that some part of public financing does, in fact, stimulate R&D activities.

Finally, our preferred point estimate for parameter β (1.07) implies that subsidies induce only modest increases in privately financed effort. This impact grows 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%.[33] This is, however, only a lower bound for the increment in private expenses, which 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.

What is implied, then, by our model with regard to the overall effect of subsidies on Spanish manufacturing? Because our sample has a known representativeness, this can be roughly computed from the following exercise. Take (predicted) R&D expenditures in the presence of subsidies and in the absence of subsidies. We will distinguish between firms whose R&D performance decision is not affected by subsidies and firms that begin carrying out R&D thanks to subsidies. We build manufacturing aggregate numbers (for the whole period).[34] The numbers say that aggregate R&D expenditure increases by about 8% as the result of subsidies.[35] Interestingly enough, total expected subsidies (observed subsidies) amount to 4.4% (5%) of total R&D expenditure. Hence subsidies are helping to increase total expenditure by slightly more than their amount.

The 8% increment can be decomposed in two parts: 4.4% comes from the increase in expenditures of firms that would perform R&D in any case, and 3.6% comes from the R&D contributed by firms that the model predicts to be nonperformers in the absence of subsidies. It is interesting to further decompose these numbers according to firm size. The percentage increase

[33] These numbers imply a low value for the derivative of private expenses with respect to subsidies. For $\rho^e = .3$, our estimate gives a value of .06, positive (no crowding out) but small.

[34] We add up the values for firms with 200 and fewer workers multiplied by 20 (the sample of these firms amounts to roughly 5% of the population) and for firms with more than 200 workers multiplied by 2 (the sample includes on average more or less half of the population).

[35] We compute a rough standard error of .2 associated with the aggregrate number 8% by applying the delta method, taking the weighting scheme as fixed.

TABLE 6 Subsidies Required to Engage in R&D and the Impact of Subsidy Withdrawal

Subsidies Required to Engage in R&D[a] (Percentages of observations by subsidy values)					Impact of Subsidy Withdrawal[b] (Number of observations and percentage from performing observations)				
Trigger Subsidy Values	≤200 Workers		>200 Workers		Subsidy Values	≤200 Workers		>200 Workers	
(%)	%	Cumulated %	%	Cumulated %	(%)	Stop Doing R&D	%	Stop Doing R&D	%
0–10	3.3	3.3	48.7	48.7	0–10	29	6.8	24	2.0
10–20	6.0	9.3	41.3	90.0	10–20	5	1.2	10	.9
20–30	8.1	17.4	6.9	96.9	20–30	14	3.3		
30–40	13.3	30.7	2.5	99.4	30–40	7	1.6		
40–50	22.4	53.1	0.6	100.0	40–50	2	.5		
50–60	29.5	82.6			50–60	2	.5		
60–70	17.4	100.0							
Number of observations	3,321		160			59 (13.8%)		34 (2.9%)	
Median subsidy	48.9		10.1			11.0		4.0	

[a]Firms with negative gaps even with currently expected subsidy.
[b]Firms that run into negative gaps when expected subsidy is not accounted for.

in the R&D of the smallest firms ($\leq$ 200 workers) is higher, 10.8%, with a contribution of the firms stimulated to perform R&D as high as 6.9%. The percentage increase in the R&D of the largest firms ($>$ 200 workers) is 5.9%, with a component due to the switching firms of only .9%.[36] Subsidies during the period are thus estimated to increase total R&D expenditure by more than their amount, with almost half of the effect coming from the firms stimulated to perform R&D, which are mainly small firms.

7. Summary and conclusions

■ The evidence of the impact of subsidies on firms' decisions regarding R&D remains relatively modest and controversial. This article tries to contribute a series of findings, based on a model of firms' decisions estimated with a representative panel sample of more than 2,000 Spanish manufacturing firms. The decision of whether or not to spend on R&D emerges from the comparison of optimal nonzero effort with the effort needed to reach some profitability (threshold effort). We focus on the impact of the expected subsidy (or fraction of effort that is expected to be publicly supported) on this comparison, and on the level of expenditure chosen. The model is estimated using a censored variable regression, with methods that attempt to avoid selectivity and endogeneity biases, taking into account autocorrelated errors.

We find that nonperformance 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 large firms.

Subsidies are potentially effective in inducing firms to invest. We estimate that almost half of large nonperforming firms could be induced to perform innovative activities by financing less than 10% of their R&D, and one out of three small nonperforming firms by financing up to 40% of their expenses. We obtain evidence that actual subsidies do, in fact, play a part, even if a modest one. Some small firms' R&D performing observations are estimated to depend on the (expected) subsidy, in the sense that no R&D would be observed in its absence. But it must be realized that subsidies go mainly to firms that would have performed innovative activities anyway. This fact, which can be seen as the result of a proper selection of applicants and risk-aversion practices of agencies, suggests that public policy tends to neglect the inducing dimension of public support.

[36] Standard errors associated with 10.8% and 5.9% are .35 and .23 respectively.

On the other hand, subsidies seem to induce only a very slight change in the level of private expenditures chosen by the firms that would, in any case, perform innovative activities. Our 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 or inefficient use of subsidies is observed.

On the whole, for Spanish manufacturing, subsidies are estimated to increase total R&D expenditure by slightly more than their amount, with almost half of the effect coming from the firms stimulated to carry out R&D, which are mainly small firms.

The employed framework, despite its simplicity, has turned out to be sensible in describing profitability gaps and exploring the impact of subsidies. Among others, two main questions call for further research: (1) the developing of dynamics (the different behavior of stable and occasional performers, the incurring of sunk investments, etc.), and (2) the modelling of the *ex post* adjustments of firms.

Appendix A

■ Econometric details follows.

□ **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 $e^* - \bar{e} = \beta x + z(\beta_1 - \beta_2) + v_2$, where $v_2 = \varepsilon_1 - u_2$, are the structural and selectivity equations. We observe $y_1 = e^*$ if $y_2 = 1\,[e^* - \bar{e} > 0] = 1$. 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 $y_2 = 1[\beta x + z(\beta_1 - \beta_2) + v_2 > 0]$, with $v_2 \sim N(0, \sigma^2)$. Conditioning on y_1 and writing $(\delta, \delta_2) = (1/\sigma)(\beta, \beta_1 - \beta_2)$, $y_2 = 1[\delta x + z\delta_2 + r(y_1 - \beta x - z_1\beta_1)/\sigma_1 + \varepsilon_2 > 0]$, with $\varepsilon_2 \sim N(0, 1 - r^2)$ and $r = \sigma_{1v}/(\sigma_1\sigma)$. Notice that σ is identified through the relationship between δ and β. The partial conditional log-likelihood for an observation is

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

□ **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 $e^* - \bar{e} = \gamma e^*_{t-1} + \beta\widetilde{x} + \widetilde{z}_1\beta_1 - z\beta_2 + \varepsilon_1 - u_2$ or $e^* - \bar{e} = \beta[\widetilde{x} + (\gamma/\beta)e^*_{t-1} - \gamma z_{1t-1}(\beta_1/\beta)] + z(\beta_1 - \beta_2) + v_2 = \beta\widetilde{x}_c + z(\beta_1 - \beta_2) + v_2$.

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 $y_1 \mid \mathbf{z} \sim N(\gamma e^*_{t-1} + \beta\widetilde{x} + \widetilde{z}_1\beta_1, \sigma_1^2)$, $y_2 = 1[\beta\widetilde{x}_c + z(\beta_1 - \beta_2) + v_2 > 0]$, with $v_2 \sim N(0, \sigma^2)$, and $y_2 = 1[\delta\widetilde{x}_c + z\delta_2 + r(y_1 - \gamma e^*_{t-1} - \beta\widetilde{x} - \widetilde{z}_1\beta_1)/\sigma_1 + \varepsilon_2 > 0]$, with $\varepsilon_2 \sim N(0, 1 - r^2)$ and $r = \sigma_{1v}/(\sigma_1\sigma)$. Notice that y_2 is now given by a nonlinear model in the parameters, but σ is again identified.

□ **Model III (differenced differences).** Assume that $\text{sign}[(e^*_t - \bar{e}_t) - \gamma(e^*_{t-1} - \bar{e}_{t-1})] = \text{sign}(e^*_t - \bar{e}_t)$. 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 $e^* - \bar{e}$ 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 $(e^*_t - \bar{e}_t) - \gamma(e^*_{t-1} - \bar{e}_{t-1}) = \beta\widetilde{x} + \widetilde{z}(\beta_1 - \beta_2) + \varepsilon_1 - \widetilde{u}_2$, 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 $y_2 = 1[(e^*_t - \bar{e}_t) - \gamma(e^*_{t-1} - \bar{e}_{t-1}) > 0] = 1$ (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.

Appendix B

■ **Variable definition and sample description.** 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 B1 describes the sample. Descriptive statistics are available in the web Appendix at www.rje.org/main/sup-mat.html.

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

Age: firms' average founding year (1975) minus the founding 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 growth rate of an estimate of the firm's capital in equipment goods and machinery.

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

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

Domestic exporter: dummy variable that 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 that takes the value one if the firm reports significant market share and the market has fewer than 10 competitors.

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

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

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 that takes the value one if sales minus production cost is negative.

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

Recessive market: dummy variable that 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 that takes the value one if effort is positive.

Skilled labor: dummy that 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 that takes the value one if the subsidy is positive.

Technological sophistication: dummy variable that 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 university graduates to total personnel.

Time dummies: set of yearly dummy variables.

TABLE B1 **Number of Firms by Time Spells and Type of R&D Performers**

Years in Sample	Firms	Observations	Nonperformers[a] Firms	Stable Performers[b] Firms	Stable Performers[b] Mean Effort ≤200	Stable Performers[b] Mean Effort >200	Occasional Performers[c] Firms	Occasional Performers[c] Mean Effort ≤200	Occasional Performers[c] Mean Effort >200
1	298	298	145	112	3.1	2.5	41	.7	.3
2	503	1,006	287	129	2.6	3.1	87	.9	.6
3	319	957	159	74	2.0	2.1	86	.5	.3
4	186	744	84	56	2.7	2.2	46	.5	.4
5	193	965	81	54	2.4	3.3	58	.5	.5
6	170	1,020	83	39	2.3	2.4	48	.8	.6
7	136	952	67	27	2.8	2.7	42	1.0	.7
8	168	1,344	85	18	4.5	3.3	65	.6	.7
9	241	2,169	102	53	2.3	2.6	86	.6	.4
Total	2,214	9,455	1,093	562	2.6	2.7	559	.7	.5

[a] Firms reporting zero R&D expenditures each observed year.
[b] Firms reporting positive R&D expenditures each observed year.
[c] Firms reporting positive R&D expenditures some of the observed years.

References

AGHION, P., BLOOM, N., BLUNDELL, R., GRIFFITH, R., AND HOWITT, P. "Competition and Innovation: An Inverted U Relationship." *Quarterly Journal of Economics*, Vol. 120 (2005), pp. 701–728.

ALMUS, M. AND CZARNITZKI, D. "The Effects of Public R&D Subsidies on Firms' Innovation Activities: The Case of Eastern Germany." *Journal of Business and Economic Statistics*, Vol. 21 (2003), pp. 226–236.

AMEMIYA, T. *Advanced Econometrics*. Oxford: Basil Blackwell, 1985.

ARELLANO, M., BOVER, O., AND LABEAGA, J.M. "Autoregressive Models with Sample Selection for Panel Data." In C. Hsiao et al., eds., *Analysis of Panel and Limited Dependent Variable Models*. New York: Cambridge University Press, 1999.

ARROW, K. "Economic Welfare and the Allocation of Resources for Inventions." In R. Nelson, ed., *The Rate and Direction of Inventive Activity*. Princeton, N.J.: Princeton University Press, 1962.

BUSOM, I. "An Empirical Evaluation of the Effects of R&D Subsidies." *Economics of Innovation and New Technology*, Vol. 9 (2000), pp. 111–148.

COHEN, W. "Empirical Studies of Innovative Activity." In P. Stoneman, ed., *Handbook of the Economics of Innovation and Technological Change*. Oxford: Blackwell, 1995.

——— AND LEVIN, R. "Empirical Studies of Innovation and Market Structure." In R. Schmalensee and R.D. Willig, eds., *Handbook of Industrial Organization*. New York: North-Holland, 1989.

DAVID, P., HALL, B., AND TOOLE, A. "Is Public R&D a Complement or Substitute for Private R&D? A Review of the Econometric Evidence." *Research Policy*, Vol. 29 (2000), pp. 497–529.

DORFMAN, R. AND STEINER, P. "Optimal Advertising and Optimal Quality." *American Economic Review*, Vol. 44 (1954), pp. 826–836.

GRILICHES, Z. "Issues in Assessing the Contribution of R&D to Productivity Growth." *Bell Journal of Economics*, Vol. 10 (1979), pp. 92–116.

GRONAU, R. "The Effect of Children on the Housewife's Value of Time." *Journal of Political Economy*, Vol. 81 (1973), pp. 168–199.

HALL, B. AND VAN REENEN, J. "How Effective Are Fiscal Incentives for R&D? A Review of the Evidence." *Research Policy*, Vol. 29 (2000), pp. 449–469.

KLETTE, J., MOEN, J., AND GRILICHES, Z. "Do Subsidies to Commercial R&D Reduce Market Failures? Microeconometric Evaluation Studies." *Research Policy*, Vol. 29 (2000), pp. 471–495.

LACH, S. "Do R&D Subsidies Stimulate or Displace Private R&D? Evidence from Israel." *Journal of Industrial Economics*, Vol. 50 (2002), pp. 369–390.

MADDALA, G.S. *Limited-Dependent and Qualitative Variables in Econometrics*. New York: Cambridge University Press, 1983.

NAWATA, K. AND NAGASE, N. "Estimation of Sample Selection Bias Models." *Econometric Reviews*, Vol. 15 (1996), pp. 387–400.

NELSON, F. "Censored Regression Models with Unobserved Stochastic Censored Thresholds." *Journal of Econometrics*, Vol. 6 (1977), pp. 309–327.

NELSON, R. "The Simple Economics of Basic Scientific Research." *Journal of Political Economy*, Vol. 76 (1959), pp. 297–306.

PAKES, A. AND SCHANKERMAN, M. "An Exploration into the Determinants of Research Intensity." In Z. Griliches, ed., *R&D, Patents and Productivity*. Chicago: University of Chicago Press, 1984.

TRAJTENBERG, M. "Government Support for Commercial R&D: Lessons from the Israeli Experience." *Innovation Policy and the Economy*, Vol. 2 (2002), pp. 79–134.

WALLSTEN, S. "The Effects of Government-Industry R&D Programs on Private R&D: The Case of the Small Business Innovation Research Program." *RAND Journal of Economics*, Vol. 31 (2000), pp. 82–100.

WOOLDRIDGE, J.M. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Mass.: MIT Press, 2002.