Helsinki Center of Economic Research

Discussion Papers

The effect of government subsidies on private R&D: evidence from geographic variation in support program funding

Elias Einiö
University of Helsinki and HECER

Discussion Paper No. 263
May 2009

ISSN 1795-0562

HECER – Helsinki Center of Economic Research, P.O. Box 17 (Arkadiankatu 7), FI-00014 University of Helsinki, FINLAND, Tel +358-9-191-28780, Fax +358-9-191-28781,
E-mail info-hecer@helsinki.fi, Internet www.hecer.fi

HECER
Discussion Paper No. 263

The effect of government subsidies on private R&D: evidence from geographic variation in support program funding*

Abstract

Despite the prevalence of government research and development (R&D) support programs, evaluation studies based on explicit differences in support allocation are rare. In this article the identification of the causal effect of government support on private R&D effort is based on regional differences in eligibility for European Union Regional Development Funds (ERDF) determined by the population-density rule. Our data is constructed by linking a broad R&D survey to administrative data on all R&D support applications in Finland over the years 2000-2006. We find evidence that the support program has induced additional private R&D among the participants who entered it as a result of higher funding in their region. Among this group one subsidy euro induced additional R&D worth at least 1.5 euro.

JEL Classification: H25, H32, O38

Keywords: Policy evaluation, R&D, subsidies, technology policy, treatment effects

Elias Einiö

Department of Economics
University of Helsinki
P.O.Box 17 (Arkadiankatu 7)
FI-00014 University of Helsinki
FINLAND

e-mail: elias.einio@helsinki.fi

* Thanks are due to Richard Blundell, Chiara Criscuolo, Erkki Koskela, Tuomas Pekkarinen, Otto Toivanen, Matti Sarvimäki, Jouko Verho and numerous seminar participants for helpful comments and suggestions. The data used in this article are confidential but the author's access is not exclusive. Financial support from the Finnish Cultural Foundation, the Jenny and Antti Wihuri Foundation, the OP-Pohjola Group Research Foundation, and the Yrjö Jahnsson Foundation is gratefully acknowledged.

1 Introduction

Economic growth depends essentially on the application of new knowledge in order to develop improved products and production processes. Several authors have emphasized the role of research and development (R&D) as the engine of growth (Romer, 1990; Grossman and Helpman, 1991; Aghion and Howitt, 1992). A substantial amount of new knowledge is produced in innovation projects conducted by firms in the private sector. However, due to positive externalities arising from incomplete appropriability of the results and uncertainty about their success, firms may engage in less R&D than is socially optimal (Nelson, 1959; Arrow, 1962). In order to foster innovative activities and economic growth governments in numerous countries have introduced R&D support programs aimed at increasing R&D effort in the private sector.[1] Although support programs are common across industrialized economies, there is lack of evidence of their impacts on private R&D effort based on explicit differences in support policies. The main contribution of this article is to provide such evidence by exploiting geographic variation in public R&D support funding arising from the European Union population density rules governing state aid to private businesses.

Direct government subsidies can induce firms to perform R&D that without the support would be privately unprofitable.[2] The efficiency of the support program depends crucially on the qualities of the projects that are taken into it. A major concern is that program managers may be encouraged to support projects with the best technical merits and the highest potential for commercial success. As these projects typically have high private returns they will be undertaken even in the absence of the support. In this case government support may induce only a little additional R&D if any at all.

The main econometric challenge in evaluating R&D support programs arises from the fact that subsidies are typically not randomly assigned, and as a result groups of supported and unsupported firms are not directly comparable. Moreover, some of the characteristics affecting the selection, such as research productivity, are very seldom observed by the researcher. In this case OLS and other methods based on the assumption that, condi-

[1] Public R&D grants covered about 7.5 percent of private R&D in the OECD countries in 2004 (OECD, 2006). For an overview of R&D tax credits which are another commonly used fiscal incentive for R&D investment, see Bloom, Griffith, and Van Reenen (2002).

[2] For an extensive discussion on the micro- and macro-economic effects of public R&D support, see David, Hall, and Toole (2000). For an R&D-driven growth model incorporating R&D subsidies, see Davidson and Segerstrom (1998).

tional on observed factors, the support is randomly assigned are likely to yield biased estimates of the causal effect of the program because it seems unlikely that conditioning on the observable attributes is sufficient to avoid differences in the expected performance between the supported and unsupported groups in the absence of the treatment (Jaffe, 2002).[3] In order to assess the selection on unobservable attributes Wallsten (2000) uses an instrumental variables (IV) approach based on the idea that differences in government R&D-support funding across industries induce variation in the likelihood of receiving the support and finds that the grants allocated by the SBIR program in the US crowded out private R&D expenditures dollar for dollar.[4] However, there is a concern that the government may allocate its support partly in line with technological opportunities. Such opportunities may differ across industries and also affect R&D investment decisions, and variation in R&D support funding across industries is likely to be endogenous as a result (David, Hall, and Toole, 2000).

In this article, the identification of the effect of program participation is based on geographic variation in potentially available R&D-support funding arising from the allocation of European Union Regional Development Funds (ERDF) in Finland.[5] These differences in allocation induce variation in the probability of program participation, which facilitates the identification of the causal effect of program participation on R&D effort. The advantage of our approach is that it is based on explicit differences in public policies with well-defined publicly stated allocation criteria. The regional provision of ERDF in Finland is especially suitable for program evaluation purposes because there are regions receiving the highest levels of European Union regional development aid because of low population density (rather than because of low levels of R&D investment or poor economic performance, for example). Furthermore, we address the potential endogeneity of this eligibility rule by controlling for the effect of population density on the outcome of interest, firm-level R&D expenditure.

We believe that results concerning the Finnish program may in a significant way im-

[3]For studies in which identification is based on the assumption of random assignment after conditioning on observable characteristics, see Howe and McFetridge (1976), Irwin and Klenow (1996), Lerner (1999), and Almus and Czarnitzki (2003). Lach (2002), and Görg and Strobl (2007) also control for time-invariant unobserved firm characteristics affecting R&D effort.

[4]For studies using a structural approach see González, Jamandreu, and Pazó (2005) and Takalo, Tanayama, and Toivanen (2008).

[5]Criscuolo, Martin, Overman, and Van Reenen (2007) also use regional variation in the European Commission rules on regional-support eligibility in a study evaluating an investment subsidy program in the UK.

prove our understanding about the potential of R&D subsidies to enhance private innovative efforts in technologically advanced economies because Finland has a large export-oriented high-tech sector and there are no other R&D policy instruments, such as tax exemptions, to complicate the evaluation analysis. Using data constructed by linking a broad R&D survey to administrative information on all recipients of government R&D subsidies over the years 2000-2006 we find evidence of the positive effect of the program on R&D expenditure among firms that entered it as a result of higher ERDF funding in their region. A conservative estimate is that one subsidy euro induces 1.5 euro of additional company R&D. Our findings indicate that government assistance may induce additional R&D, and that the impact may be substantial.

The rest of the article is organized as follows. Section 2 describes the institutional setting of the R&D support policy in Finland and Section 3 presents the data. The empirical strategy for identifying the causal effect of the participation in the R&D support program on private R&D effort is explained in Section 4, Section 5 presents the results and several robustness tests verifying the identification strategy and Section 6 concludes the article.

2 The Institutional Setting

The implementation of the Finnish technology policy is centralized in the Funding Agency for Technology and Innovation (Tekes), which is the only authority responsible for allocating government subsidies to R&D projects conducted by private businesses. During the period 2000-2005 the agency granted direct R&D subsidies worth 970 million euro. The Tekes budget comes mainly from the Finnish Government. Some areas receive funds from the ERDF in addition to national funding.

The Finnish R&D support program

The publicly expressed objective of the Finnish R&D support program is to encourage firms to start up new and accelerate the completion of ongoing R&D projects. All firms operating in Finland may apply for funding. The main criteria for being selected into the program include commercial potential (in terms of expected future sales), technological challenge, available resources (e.g., R&D staff qualified to conduct the project), and the importance of the agency's support to the success of the project. The agency's Internet pages emphasize that projects involving technological and commercial risks and which would not be fully implemented without agency funding are specifically supported.[6]

Firms receive funding in the form of direct subsidies and low-interest loans. Direct subsidies are granted to research projects that provide the basis for future product and process development, the results of which are not immediately commercializable. High technological challenge, which reflects the riskiness of the project, increases the probability of receiving a direct subsidy. In the case of precompetitive projects involving the development of new products or processes that have direct commercial value, loans are the main form of funding. Funding consisting partly of direct subsidies and partly of loans is also available. Direct subsidies cover 25-50 percent of the realized costs of the project up to an amount that the agency has approved as the maximum cost. In order to receive a subsidy payment firms must present an account of the realized costs.

The program has several characteristics that could be expected to induce efficiency. The most important is the criterion that agency funding is necessary for the completion of the project. The program also has to satisfy European Union regulations for state support of R&D, adherence to which is monitored by the European Commission. The European Commission states in its community framework covering R&D aid that it will make special efforts to verify that the planned support will induce firms to pursue research they would not otherwise have pursued.[7]

There are still several reasons why the program may be inefficient, however. The agency may be unable to identify the projects for which its funding is of the essence. On the other hand, it is not obvious whether or not the program is actually abiding by its stated selection rules and European Union regulations. In its own effectiveness analysis it relies on figures on the average success of the projects and does not mention the selection problem which may confound the causal interpretations. There is a need for a careful econometric evaluation addressing the selection problem in order to demonstrate the efficiency of the program. Such an analysis requires a credible source of exogenous variation in the program-participation status. In order to identify the effects of govern-

[6] Http://www.tekes.fi (Accessed April 18, 2007).
[7] Official Journal of the European Union C 45, 17.2.1996.

ment assistance on private R&D we use regional variation in the available government R&D funding.

Regional differences in government R&D funding

The Finnish regional policy is based on districts delineated in accordance with European Union criteria. Finland joined the European Union in 1995, bargaining in the accession negotiations for a large share of the country to be eligible for the highest level of regional development aid over the program period 1995-1999. The criterion for eligibility was "no more than 8 persons per km^2".[8] The region receiving the highest level of aid (Objective 1) was largely maintained for the program period 2000-2006.[9] Morover, a substantial part of the country was considered eligible for the second highest level of aid (Objective 2) and transitional support (Objective 4). Figure 1 depicts these regional provisions.

The regional variation in available government R&D funding arises because Tekes is entitled to withdraw larger amounts from the ERDF in areas eligible for higher levels of aid. The total government R&D funding to the private sector relative to GDP in 2000-2006 was 17 percent higher in regions under Objective 1 than in those under Objective 2.[10] This suggests that ERDF funding may produce substantial variation in the probability of being accepted into the program. This is verified in the empirical analysis, which indicates that the probability of receiving the subsidy is substantially higher in the regions under Objective 1 compared to those eligible for a lower level of aid.

3 Data

We constructed our data by linking several administrative datasets to a broad R&D survey, all maintained by Statistics Finland.

[8]See Protocol 6 of the EU Act of Accession of 1994.

[9]See Article 3 in the European Council regulation No 1260/1999 on laying down the general provisions on the Structural Funds (The Official Journal of the European Union L 161, 26.6.1999). Part of Central Finland became eligible for highest level of aid due to changes in the regional units within which the population densities were calculated. The only change to the previous provision was that the whole region of Pohjois-Savo was included in Objective 1, as part of this region was under Objectives 2 and 5b in the previous period.

[10]The calculation is based on data on the regional R&D funding provided by Tekes and regional GDP figures from Statistics Finland. It includes the districts of Etelä-Savo, Pohjois-Savo and Pohjois-Karjala from the Objective 1 area, and Häme and Etelä-Karjala from the Objective 2 area, which are complete regional-level statistical units within the objective areas.

We drew administrative data on R&D-support allocation from the Firm Assistance Database (FAD), which provides information on all applications submitted to Tekes during the period 2000-2006. Because Tekes is the only authority allocating R&D subsidies, the FAD information covers all government R&D subsidy recipients. Furthermore, because we had information on all subsidy recipients, we were able to identify whether any firm in any other database had received a subsidy or not. This was particularly important in the construction of the control group with which we compared the performance of the subsidized firms. The FAD contains the following project-level information: the project identifier number, the firm identifier number, the amount applied for, the decisions on the amount of subsidy or loans granted, if any, the date of the decision, and the date and amount of each payment. For each project, we allocated the received amount of subsidy uniformly over each payment period. We then constructed the firm-level data by aggregating the project-level data according to the firm identifier number.

We obtained information on R&D expenditure from the annual R&D survey panel (RDS). The initial version was based on a census conducted in 1995. Since then, the main survey frame has covered all firms reporting positive R&D expenditure in the surveyed year, or expecting to conduct R&D in the following year. It also includes all firms employing at least 100 persons. The main survey frame is augmented by drawing a random sample from the population of firms not included in it. Finally, the survey includes all firms that had applied for Tekes funding.[11] The response rate for the survey was around 85-87 percent during the period under observation.

We drew information on sales and fixed assets from the Financial Statement Database (FSD). If no information was available we used sales figures from the Business Register Firm-Level Database (BRFD), which is based on data provided by the tax authorities. We also obtained administrative information on the age and export status of the firms from the BRDF.

Our identification is based on the regional variation in the government R&D funding and thus it is essential to accurately identify the location of each firm. Our primary source was the RDS, which provides firm-level information on the regional distribution of

[11]Thus the RDS is not a random sample of R&D-inactive and new firms. A potential concern is that the proportion of supported firms in the higher-aid region is overestimated in our sample because of the non-random sampling of applicant firms. The robustness tests presented in Section 5 indicate that such concern is not relevant for our results.

R&D activities on the municipal level: a firm was defined as being located in a region[12] if it conducts all of its R&D there. A secondary source of location information is the Business Register Plant-Level Database (BRPD)[13], which yields the following plant-level information: the plant identification number, the identification number of the parent firm, the postcode for the location, and the number of employees. We used this information in order to calculate firm-level employment by region. The secondary location criterion was that all of the firm's employees are based in the relevant region.

Table 1 reports the descriptive statistics for the analysis sample and the R&D survey panel sample. In the table, year $t-1$ refers to the year before and year $t+1$ to the year after supported firms entered the program.[14] The table shows that the group of supported firms, while slightly younger on average, had higher average sales and fixed assets relative to the unsupported group. In the year before being granted the subsidy they were devoting, on average, 691,451 euro to R&D while the average R&D expenditure in the control group was 335,694 euro. The distribution of R&D expenditure is very skewed and the pre-treatment differences between the treatment and control groups are less pronounced when the 25th, 50th, and 75th percentiles are compared. In the year following a positive grant decision the grantees had increased their average R&D expenditure to 850,024 euro and were receiving an average annual subsidy flow of 70,762 euro. The subsidies covered, on average, 26 percent of the total R&D costs. The average change in R&D expenditure among the supported firms from the year before they accessed the program to the year following it was 158,573 euro, whereas among the unsupported group the corresponding change was 36,794 euro. A simple before-after estimate suggests that the average effect of the program was 158,573 - 36,794 = 121,779 euro. However, this naive estimate is likely to suffer from selection bias because the support was not assigned randomly, and it is unlikely that the treatment and control groups are directly comparable.

[12] An ERDF area or other relevant regional unit used in the analysis.

[13] If a firm has only one plant it is always included in this database. For multi-plant firms the information is based on a questionnaire.

[14] For a detailed definition of treatment and control groups see Section 4.

4 The empirical strategy

This section explains the empirical strategy adopted for identifying the causal effect of the R&D support program. The aim is to assess whether, as a result of becoming a program participant, the firm increased its R&D expenditure. The "treatment group" comprise the firms that started in the support program in year t. We estimate the effect of program participation on R&D effort in the second participation year $t+1$ for two reasons. First, because decisions are given at any point of the year, and it may take some time before the firm gets the subsidized project running after a grant decision, much of the participation effect on annual R&D expenditure is likely to emerge in future years. Second, only a very small proportion of subsidized projects last less than one year.[15] In order to ensure that the firm was supported in year $t+1$, those in the treatment group had to have received payments from the agency in that year. The "control group" comprised firms that did not receive any payments from the agency in the consecutive three years $t-1$, t, and $t+1$. In order to eliminate the possible long-term effects of the program among the control group, it was also required that the firms in it did not fulfill the criteria for the treatment group in other years.

The basis of the empirical analysis is the following R&D equation:

$$log(R\&D_{i,t+1}) = \phi_{j(i),t} + \gamma D_{it} + \beta X_{it} + \epsilon_{it}, \tag{1}$$

where i, j and t indicate the firm, industry and time, respectively. The treatment dummy variable D_{it} is one (zero) if firm i was in the treatment (control) group in year t. Because the time patterns of R&D investment may differ across industries, and support allocation across industries may be partly determined by differences in R&D intensity, we allow for industry- and time-specific differences, $\phi_{j(i),t}$, in R&D expenditure. The vector X_{it} includes a set of pre-treatment firm characteristics. Although we have panel data, we are unable to include firm fixed effects because there is no within-firm variation in our instrument, the ERDF-region indicator. However, we include a second-order polynomial of logged R&D in year $t-1$ to control for permanent differences in the levels of R&D expenditure. We also include the log of sales and fixed assets in year $t-1$ in order to

[15]Among the projects ending in the period 2001-2003 only 2.9 percent lasted less than one year, while 64.8 percent lasted between one and three years (Tekes, 2008).

control for firm size and capital intensity, a second-order polynomial of age to control for age effects and whether the firm is a start-up, and a dummy variable indicating whether or not the firm is exporting.

Our parameter of interest is γ, the causal effect of program participation on the log of R&D expenditure. When $\gamma \leq 0$ the program does not induce additional R&D and the firms would have made at least the same R&D investments in the absence of the subsidy. If $\gamma > 0$ the support scheme induces additional R&D. As γ is the coefficient for a binary program-participation dummy and not for the amount of subsidy, it is not directly informative about how many additional euros of additional R&D one subsidy euro induces. In the case where a fixed proportion of R&D that would not have been realized without the subsidy is supported, one subsidy euro will generate additional R&D worth the inverse of the compensation rate (the share of realized R&D costs covered by the subsidy). However, this simple rule is no longer valid if the firm is able to cover R&D costs that would have been realized without the subsidy. However, we can deduce that when the subsidy is the largest possible, i.e. 50 percent of the firm's *total* post-treatment R&D expenditure, a necessary condition for no crowding-out is that γ is at least $log(2) \approx 0.694$.[16] If $\gamma \in (0, log(2))$, we cannot rule out the possibility that some of the support is crowding-out the firm's own R&D investment.

The R&D support program selects the participants partly according to technological potential and expectations of commercial success, which is not observed. Because the firm's investment decisions also depend on these factors, the program-participation status D_{it} and the error term ϵ_{it} in the R&D equation may be correlated, which would result in biases in the least squares estimates of the causal effect of the program γ. In order to overcome this selection problem, we followed the approach taken by Lichtenberg (1988) and Wallsten (2000) in that we exploited the variation in the potentially awardable R&D funding.[17] Instead of relying on variation across the lines of business or industry, we utilized regional variation in public R&D support funding across the ERDF areas. We argue

[16]For example, suppose that R&D expenditure by the firm is z_0 when unsupported and cz_0 when supported. Consider the case in which the firm received the largest possible amount of support, i.e. the whole post-treatment R&D cost cz_0 is eligible for the maximum compensation rate of 0.5. In this case one subsidy euro will generate $(cz_0 - z_0)/0.5cz_0$ euros of additional private R&D. The no-crowding-out condition requires that one subsidy euro induces additional R&D worth at least one euro. This implies $c \geq 2$ for no crowding-out under maximum compensation.

[17]Lichtenberg's potentially awardable R&D contracts are calculated as the total value of government contracts for two-digit Federal Supply Code the firm sold to. Wallsten applies a similar approach to awardable R&D funds.

that this approach has two advantages. First, the government agency may allocate its support partly in line with technological opportunities, which may differ across industries and also affect R&D investment decisions (David, Hall, and Toole, 2000). As a result, industry level variation in R&D support funding is likely to be endogenous. Second, in our empirical setup variation in R&D support funding is based on explicit differences in public policies imposed on comparable firms with well-defined and publicly stated allocation criteria.

We argue that differences in R&D support funding across the ERDF 1 border produce exogenous variation in program participation for two reasons. First, ERDF 1 eligibility is based on the criterion that the population density in the region is "no more than 8 persons per square kilometer" rather than on direct performance measures of the local economy. Second, we are able to control for the effect of population density on R&D expenditure by estimating the following equation:

$$log(R\&D_{i,t+1}) = \phi_{j(i),t} + \gamma_i D_{it} + \beta X_{it} + log(POPDEN_{r(i),t-1}) + \epsilon_{it}, \qquad (2)$$

where we have included the log of the population density in year $t-1$ in NUTS 4 region r.[18] This empirical setup resembles the regression discontinuity framework in that we are able to control for the effect of the factor underlying the eligibility for the treatment on the outcome of interest.[19] Several robustness tests presented in Section 5 verify the exogeneity of the regional support allocation.

In equation (2) we also make explicit the fact that the effect of program participation may differ across participants, i.e. $\gamma_i \neq \gamma_j$ for $i \neq j$. Imbens and Angrist (1994) show that when the effects of the treatment are heterogeneous the IV estimate of the treatment parameter is the local average treatment effect (LATE), i.e. the average effect of the treatment among the participants whose treatment status the instrument changes.[20] In the context of this study, LATE is the effect of program participation among the firms entering the program as a result of higher level of funding in the ERDF 1 region.

[18]For the full NUTS 4 classification see http://www.stat.fi/meta/luokitukset/seutukunta/001-2007/.

[19]Our empirical setup differs from what would be considered a policy discontinuity design in two respects. First, the variation in population density we utilize is across a finer regional classification than the one on which the original provisions are based. Second, we control for the current level of population density rather than the pre-program population density on which the original provisions are based.

[20]Abadie (2003) generalizes this result for the case in which covariates may be included in the model and shows that in this case LATE is identified provided that the probability of participation is linear in the instruments.

The relaxation of the assumption that the effect of the program is the same for all participants is especially relevant in the evaluation of R&D support programs because it is likely that the agency selects projects partly according to potential research productivity and expected success. For example, if it selects projects with better research productivity, i.e. those that are likely to be privately profitable and on which the program is expected to have only a minor impact, the average effect of the program among the treated firms is expected to be small. On the other hand, if the agency selects projects with lower potential research productivity that are likely to be privately unprofitable, the average effect of the program among the treated firms is expected to be larger.

5 Results

Table 2 presents the estimation results for the R&D equation (2). The first column reports the OLS estimates whereas the second and third report the first- and second-stage estimates from the two-stage least squares procedure. The OLS estimate of 0.342 suggests that the support program had a positive effect on R&D expenditure. However, if selection into the program is affected by unobservable firm characteristics that also affect the R&D investment decision, the OLS estimate is likely to be biased.

In order to address the selection problem, we use an ERDF region dummy variable as an instrument for participation status. We use a binary area classification in which the first class includes the region receiving the most aid (ERDF 1) and the second class all those receiving less aid (ERDF 0, 2 and 4) . In Table 2, the first-stage estimate for the ERDF dummy variable shows that the probability of program participation is 0.138 points higher in the Objective 1 region, indicating that regional differences in available R&D support funding induce substantial differences in the probability of receiving support. The IV estimate of the causal effect of program participation is 1.391, and significant at the 95 percent confidence level. This suggests that as a result of the program the R&D expenditure among the supported firms was four times larger than it would have been in the absence of the support. As the estimate is clearly larger than $log(2)$, this result also satisfies the necessary condition for no crowding-out derived in the previous section. Furthermore, in the extreme case of maximum subsidy compensation of 50 percent of the total post-treatment R&D costs, this result suggests that one subsidy euro induced at least 1.5 euro of additional company R&D.[21]

Table 3 shows the time patterns of the effect of the support program in the periods 2000-2002 and 2003-2005. The first-stage coefficients of the ERDF instrument show a decreasing trend over time, which is largely attributable to the relatively larger amount of ERDF funding that is allocated at the beginning of the program period. As a result, the differences in the probability of receiving support induced by regional variation in public funding are more pronounced in the first years of the program. These results indicate that it is indeed the additional funding provided by the ERDF that induces higher acceptance probability for applications in the ERDF 1 region. The decreasing strength of the instrument over time is also reflected in the precision of the IV estimates: it is 1.922 and significant with a standard error of 0.874 in the period 2000-2002, and only 0.342 with a standard error of 1.160 in the period 2003-2005.

Robustness tests

We conducted several robustness tests in order to verify our empirical strategy and to assess the validity of the ERDF instrument.

We begin by testing whether the endogenous entry of firms affects our results. Because the ERDF program has been running since 1995 the firms knew about its existence at the beginning of the program period 2000-2006. Thus, entering establishments may have chosen their location partly on account of the fact that the probability of receiving the support was higher in the regions with higher levels of aid. In order to find out whether endogenous entry affects our results we estimated the effect of the program among the firms that started up prior to 1994. Column 1 of Table 4 shows that among these "old firms" the effect of the program is very similar compared to the full analysis sample. This suggests that endogenous entry does not invalidate our results.

Our key identifying assumption is that conditional on our set of control variables, the ERDF region and R&D expenditure are independent. Suppose that in the areas receiving less aid the projects are of higher quality because of the higher concentration of R&D and the more intensive R&D spillovers. In this case, the firms in this region would be expected to have better R&D productivity and, as a result, to conduct more projects,

[21] For details of this calculation, see Footnote 16.

which would show as higher R&D expenditure. If the differences in R&D spillovers are important, firms in the regions receiving less aid may conduct more R&D compared to those in the high-aid regions in the absence of subsidies. In our study this would result in a negative bias in the IV estimate of the participation effect. However, if the estimated effect were positive, as it is in our case, the analysis would not lose its evidential power: the estimated effect will rather represent a lower bound to the actual effect. Column 2 in Table 4 gives the results for a sample that excludes the less-aided ERDF 0 region, the aim being to asses whether regional differences in the quality of R&D affect our results. The estimate for the participation effect is somewhat larger, but in general it is of the same magnitude as the full sample estimate. This indicates that differences in R&D spillovers do not constitute a problem for our analysis.

Next we analyze whether regional features have an effect by including in the model a set of regional control variables that could affect company R&D on the NUTS 4 level.[22] In order to control for R&D spillovers we include the log of R&D concentration, which we calculate as total R&D expenditure divided by acreage. To capture other effects of the region we include the unemployment rate, the proportion of secondary production, and the logged GDP per capita. The results are shown in Column 4. The effect of being a subsidy grantee is now 1.452, which is slightly larger but similar in magnitude compared to the estimate without regional controls. The finding that R&D concentration and other key performance measures of the local economy do not affect our results gives a strong indication that, conditional on our set of control variables, regional variation in government R&D funding is exogenous.

As discussed in Section 3, the RDS, which is our source of data on R&D expenditure, is not a random sample of R&D-inactive and new firms. We brought up the potential concern that the proportion of supported firms in the region with higher aid levels is overestimated in our sample because of the nonrandom sampling of applicant firms. It is worth noting first, that our analysis includes only R&D-active firms, which are likely always to be included in the sample, allthough we still cannot rule out the possibility that the nonrandom sampling of new R&D-active firms could have affected our results. In order to see whether this constitutes a problem for our analysis we compare the estimated conditional probabilities of receiving support in the high-aid region between the full sample

[22]For the full NUTS 4 classification see http://www.stat.fi/meta/luokitukset/seutukunta/001-2007/.

and the sample of "old firms". An indication of an effect of the nonrandom sampling on our results would be that the coefficient for the ERDF 1 dummy is smaller in the sample of "old firms". The reason for this is because the effect of the nonrandom sampling among the old firms would be expected to have a smaller effect on the probability of receiving support in the high-aid region because a smaller proportion of R&D-active old firms are sampled: firms reporting R&D in the previous year are always included in the RDS sample frame. The first-stage estimate for the ERDF dummy in the sample of "old firms" in Column 1 of Table 4 is 0.169, which is larger than the corresponding estimate in Table 2 (0.133). This finding indicates that the nonrandom sampling of applying firms does not affect our results.

Discussion

As explained in Section 4, we expected the effect of the program to be heterogeneous among the participants, and we have identified a local average treatment effect, i.e. the effect of program participation among the firms that enter it as a result of higher funding in their region. If this is the case we will not be able to rule out the possibility that the effect of the program would be different among the projects that would have received the support even in the absence of the ERDF funding. In fact, we have good reason to believe that the effect may be substantially smaller for a sizable proportion of projects that would also be supported in the absence of ERDF funding. The agency's report documenting the results of a participant survey shows that more than 32 percent of the supported projects would have been undertaken even without its support in the period 1999-2003 (Tekes, 2007). This figure may largely underestimate the actual proportion of projects that would have been implemented even without government assistance because grantees may feel that revealing that government support was not necessary for the completion of the project may reduce the prospects of receiving assistance in the future.

According to the results of the agency's survey, at least one third, and plausibly even more, of supported projects were *a priori* privately profitable. Among these one would expect a large share of government support to crowd out firms' own R&D investments, and as a result that the impact of the support would be small. On the other hand, the large and positive effect revealed in our empirical analysis suggests that projects entering the program as a result of higher funding in their region are mainly privately unprofitable. These observations have two implications. First, they suggest that projects entering the program because of higher funding in ERDF 1 region are also those with lower research productivity. Second, because these projects are unsupported in the absence of ERDF funding and, as argued above, a large proportion of supported projects are privately profitable, it seems likely that the agency will prioritize projects with higher potential research productivity.

6 Conclusions

Government research and development (R&D) support programs are common across industrialized economies, and in several countries there has been political pressure to increase government intervention in the R&D sector.[23] This study provides new evidence of the causal effect of government assistance on private R&D effort. Our identification strategy was based on geographic differences in R&D support allocation in Finland, where several regions are eligible for the highest level of European Union Regional Development Fund (ERDF) aid because the population density is less than eight persons per square kilometer. We controlled for the possible endogeneity of support allocation arising from the population-density rule determining these regional provisions. We have shown that the probability of receiving support is substantially higher in areas that are eligible for the highest level of aid than in regions that are not. These differences provide exogenous variation in program participation, which we used to identify the causal effect of program participation on firm-level R&D expenditure.

According to our IV estimates, the program has quadrupled R&D expenditure among firms entering it as a result of higher government R&D-support funding in their region. We emphasize that our IV approach identifies the effect of the program only among firms that change their participation status as a result of the higher funding in their region. Thus, our results should not be interpreted as evidence of the aggregate effectiveness of the program. Despite this deficiency, we believe that they are of significance in two

[23] For example, the European Commission has an overall goal of three percent for R&D as a proportion of GDP, with two-thirds of this financed by the private sector (Commission of the European Communities, 2002), and the most recent data show that R&D in the EU was about 1.84 percent of GDP in 2006 (Eurostat, 2008).

respects. First, we have shown that governments may induce additional company R&D given correctly designed public R&D support policies. Second, the impact of R&D support programs may be substantial: a conservative estimate is that one subsidy euro induces 1.5 euro of additional company R&D.

As explained in Section 4, it seems likely that a notable proportion of projects that would have been supported irrespective of ERDF funding were privately profitable. Combining this piece of evidence with our estimated large positive effect of the program, which suggests that a substantial proportion of projects at the margin are privately unprofitable, it seems likely that applicants with higher research productivity are ranked higher by the agency. In this case, enlarging the program may seem advisable as projects that would enter it as a result of increased funding would be those with lower rates of return on R&D, and among which the impact of the program might be expected to be large. However, if the large proportion of government funding crowds out private R&D in the group of already subsidized projects, the most efficient way of improving social returns on public R&D support spending may not be to expand the program; it may be better to restructure the incentives in such a way that the agency will prioritize R&D that would not be pursued without its support.

Our study adds credence to the view that public policies promoting innovative activities in the business sector may have a big impact on private R&D effort. As every program has its own selection rules and managerial practices, results concerning the effectiveness of one are not directly generalizable to others. We believe that the findings of this study are of significance for future empirical work because they suggest that the effect of support may vary substantially among participating projects even within one program. This stresses the importance of applying empirical strategies addressing the potential selection on unobservable attributes, such as research productivity, and implies that misinterpreting instrumental variable estimates as evidential about the average effect of the program may result in misleading recommendations for future policy.

References

[1] Abadie, A. (2003): Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics*, 113, 231-63.

[2] Almus, M., and Czarnitzki, D. (2003): The effects of public R&D subsidies on firms innovation activities: the case of Eastern Germany. *Journal of Business and Economic Statistics*, 21, 226-36.

[3] Arrow, K. (1962): Economic welfare and allocation of resources for invention. In Nelson, R., eds., *The Rate and Direction of Inventive Activity: Economic and Social Factors* (Princeton University Press, Princeton), 609-626.

[4] Aghion, P., and Howitt, P. (1992): A model of growth through creative destruction. *Econometrica*, 60, 323-351.

[5] Bloom, N., Griffith, R., and Van Reenen, J. (2002): Do R&D tax credits work? Evidence from a panel of countries 1979-1997. *Journal of Public Economics*, 85, 1-31.

[6] Commission of the European Communities (2002): More research for Europe: towards 3% of GDP. http://ec.europa.eu/research/era/pdf/com3percent_en.pdf.

[7] Criscuolo, C., Martin, R., Overman, H., and Van Reenen, J. (2007): The effect of industrial policy on corporate performance: evidence from panel data. London School of Economics. http://personal.lse.ac.uk/criscuol/papers/criscuolo.pdf.

[8] David, P., Hall, B., 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.

[9] Davidson, C., and Segerstrom, P. (1998): R&D subsidies and economic growth. *RAND Journal of Economics*, 29, 548-577.

[10] Eurostat (2008): EU27 R&D spending stable at 1.84% of GDP in 2006. Eurostat News Release 34/2008.

[11] Gonzáles, X., Jamandreu, J., and Pazó, C. (2005): Barriers to innovation and subsidy effectiveness. *RAND Journal of Economics*, 36, 930-50.

[12] Görg, H., and, Strobl, E. (2007): The effect of R&D subsidies on private R&D. *Economica*, 74, 215-34.

[13] Grossman, G., and Helpman, E. (1991): Quality ladders in the theory of growth. *Review of Economic Studies*, 58, 43-61.

[14] Howe, J., and McFetridge, D. (1976): The determinants of R&D expenditure. *Canadian Journal of Economics*, 9, 57-71.

[15] Imbens, G., and Angrist, J. (1994): Identification and estimation of local average treatment effects. *Econometrica*, 62, 467-75.

[16] Irwin, D., and Klenow, P. (1996): High-tech R&D subsidies: estimating the effects of Sematech. *Journal of International Economics*, 40, 323-44.

[17] Jaffe, A. (2002): Building program evaluation into the design of public research-support programmes. *Oxford Review of Economic Policy*, 18, 22-34.

[18] Lach, S. (2002): Do R&D subsidies stimulate or displace private R&D? Evidence from Israel. *Journal of Industrial Economics*, 50, 369-90.

[19] Lerner, J. (1999): The government as venture capitalist: the long-run impact of the SBIR program. *Journal of Business*, 72, 285-318.

[20] Lichtenberg, F. (1988): The private R&D investment response to federal design and technical competitions. *American Economic Review*, 78, 550-559.

[21] Nelson, R. (1959): The simple economics of basic scientific research. *Journal of Public Economics*, 49, 297-306.

[22] OECD (2006): *Main Science and Technology Indicators*. OECD. Paris.

[23] Romer, P. (1990): Endogenous technological change. *Journal of Political Economy*, 98, S71-S102.

[24] Takalo, T., Tanayama, T., and Toivanen, O. (2008): Evaluating innovation policy: a structural treatment effect model of R&D subsidies. Bank of Finland Research Discussion Papers 7/2008.

[25] Tekes (2007): Tekesin vaikuttavuus: asiakaskyselyjen ja teknologiaohjelmien arviointien analyysi. Teknologiakatsaus 203/2007. Helsinki.

[26] Tekes (2008): Tilastoanalyysi Tekesin vaikuttavuudesta. Tekesin katsaus 229/2008. Helsinki.

[27] Wallsten, S. (2000): 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*, 31, 82-100.

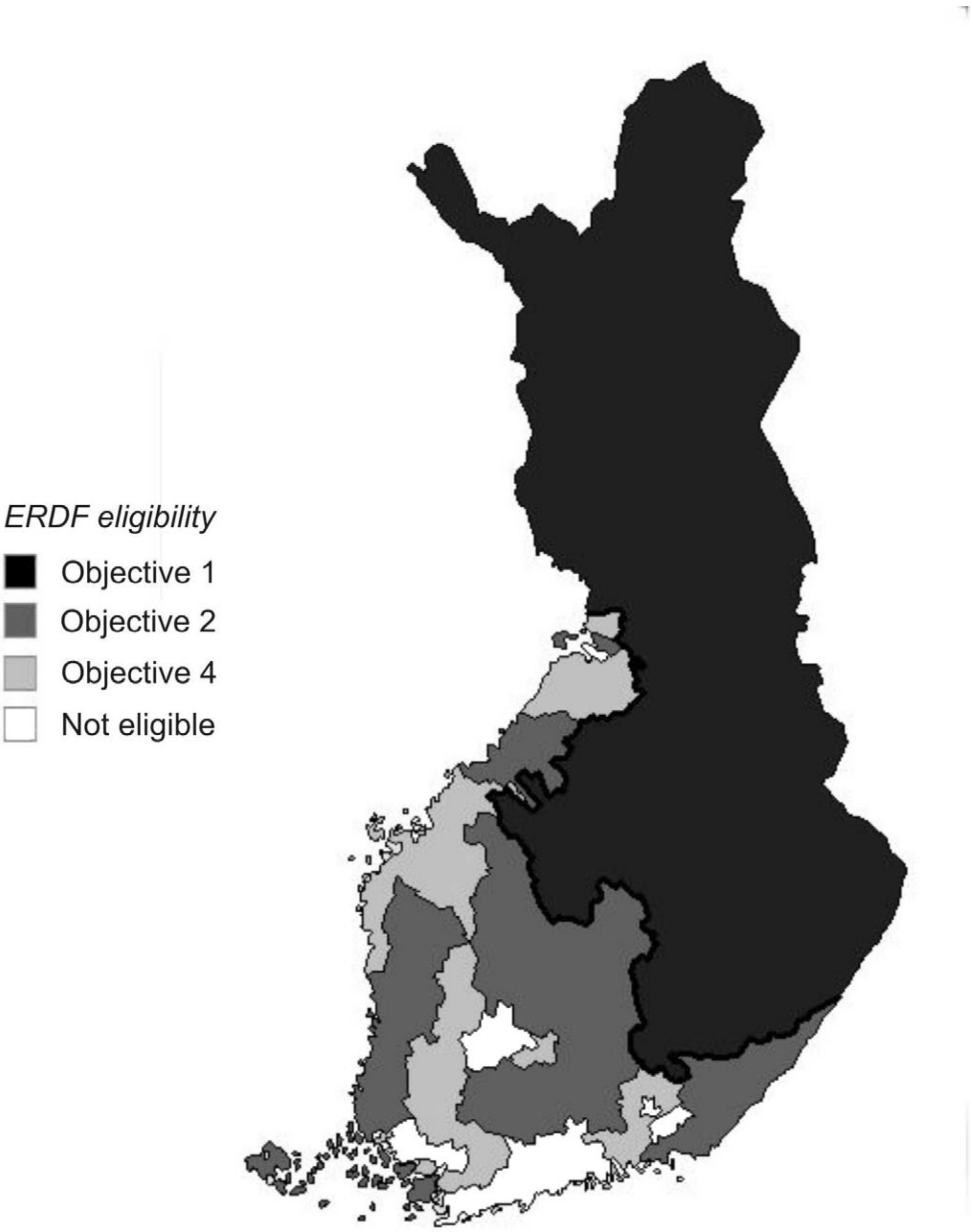

Figure 1: Regional provision of ERDF in Finland during the program period 2000-2006.

Table 1:
Descriptive statistics.

	Analysis sample			R&D survey panel
	Treated	Controls	All	All
R&D expenditure$_{t+1}$				
Mean	850,024	372,488	445,733	701,812
25th percentile	90,000	46,685	51,750	0
Median	227,040	145,000	159,000	82,340
75th percentile	602,000	377,938	413,883	375,000
Standard deviation	2,217,689	733,739	1,112,301	2,881,385
R&D expenditure$_{t-1}$				
Mean	691,451	335,694	390,261	644,675
25th percentile	58,635	45,203	46,512	16,000
Median	150,500	135,485	138,500	95,695
75th percentile	513,390	375,750	388,371	361,000
Standard deviation	1,609,434	629,286	864,665	2,535,446
Subsidy$_{t+1}$				
Mean	70,762	0	10,854	38,529
25th percentile	12,170	0	0	0
Median	30,889	0	0	0
75th percentile	76,287	0	0	18,326
Standard deviation	127,240	0	55,907	155,496
Subsidy intensity$_{t+1}$				
Mean	0.260	0	0.040	0.110
25th percentile	0.058	0	0	0
Median	0.138	0	0	0
75th percentile	0.283	0	0	0.106
Standard deviation	0.574	0	0.243	0.447
Age				
Mean	13.25	14.75	14.52	13.14
25th percentile	6	7	7	6
Median	11	12	12	11
75th percentile	18	20	19	18
Standard deviation	10.44	11.02	10.94	10.54
Sales$_{t-1}$				
Mean	34,640,000	33,680,000	33,820,000	36,337,009
25th percentile	1,039,000	1,518,000	1,444,000	677,933
Median	7,424,000	8,328,000	8,114,000	3,894,000
75th percentile	24,990,000	28,280,000	27,150,000	17,654,561
Standard deviation	110,971,823	102,997,107	104,225,538	168,583,602
Fixed assets$_{t-1}$				
Mean	26,470,000	14,895,954	16,670,000	37,200,000
25th percentile	199,200	169,587	181,000	120,000
Median	1,571,000	1,398,653	1,402,000	786,000
75th percentile	6,444,000	7,202,085	7,154,000	4,841,000
Standard deviation	150,037,223	44,369,169	71,590,466	413,655,751
Observations	254	1,402	1,656	8,354[a]

Notes: All monetary values are expressed in euros. The lower indexes $t-1$ and $t+1$ refer to the year before and the year after the treatment group started to receive the support. Subsidy intensity is the ratio of subsidy to R&D expenditure.

[a]- The number of observations correspond to a sample that included information on R&D expenditure for both years $t-1$ and $t+1$. In that sample, information on subsidy, subsidy intensity, age, sales, and fixed assets were available for 8,354, 6,032, 7,389, 8,166, and 8,166 observations, respectively.

Table 2:
OLS and IV estimates.

	OLS		1st stage		IV	
R&D subsidy grantee	0.342 (0.057)	***			1.391 (0.695)	**
ERDF Objective 1			0.133 (0.037)	***		
$\log(\text{R\&D}_{t-1})$	-0.291 (0.161)	*	-0.239 (0.073)	***	-0.051 (0.284)	
$\log(\text{R\&D}_{t-1})^2$	0.040 (0.007)	***	0.011 (0.003)	***	0.029 (0.013)	**
$\log(\text{Sales}_{t-1})$	0.196 (0.022)	***	-0.026 (0.010)	***	0.223 (0.034)	***
$\log(\text{Fixed assets}_{t-1})$	-0.009 (0.017)		0.021 (0.008)	***	-0.030 (0.026)	
Age	-0.006 (0.005)		-0.004 (0.002)	*	-0.002 (0.007)	
Age^2	0.000 (0.000)	**	0.000 (0.000)		0.000 (0.000)	
Exporter	0.076 (0.064)		0.067 (0.029)	**	0.007 (0.087)	
$\log(\text{Population density}_{t-1})$	0.014 (0.017)		-0.001 (0.009)		0.029 (0.022)	
Intercept	6.516 (1.228)	***	1.256 (0.557)	**	5.199 (1.596)	***
N	1656		1656		1656	

Notes: Industry-year interaction dummy variables are included but not shown.
Heteroskedasticity-robust standard errors in parentheses.
90, 95 and 99 % confidence levels are denoted by "*", "**", and "***", respectively.

Table 3:
OLS and IV estimates, subperiods.

	OLS		1st stage		IV	
			A. 2000-2002			
R&D subsidy grantee	0.305 (0.089)	***			1.922 (0.874)	**
ERDF Objective 1			0.169 (0.054)	***		
$\log(\text{R\&D}_{t-1})$	-0.002 (0.265)		-0.332 (0.112)	***	0.508 (0.535)	
$\log(\text{R\&D}_{t-1})^2$	0.027 (0.011)	**	0.016 (0.005)	***	0.002 (0.024)	
$\log(\text{Sales}_{t-1})$	0.221 (0.035)	***	-0.017 (0.015)		0.249 (0.052)	***
$\log(\text{Fixed assets}_{t-1})$	-0.022 (0.029)		0.014 (0.012)		-0.046 (0.044)	
Age	-0.009 (0.007)		0.000 (0.003)		-0.009 (0.010)	
Age^2	0.000 (0.000)	**	0.000 (0.000)		0.000 (0.000)	
Exporter	-0.080 (0.110)		0.064 (0.046)		-0.176 (0.152)	
$\log(\text{Population density}_{t-1})$	0.006 (0.027)		0.002 (0.013)		0.033 (0.037)	
Intercept	5.122 (1.779)	***	1.624 (0.752)	**	2.520 (2.939)	
N	791		791		791	
			B. 2003-2005			
R&D subsidy grantee	0.399 (0.073)	***			0.342 (1.160)	
ERDF Objective 1			0.099 (0.052)	*		
$\log(\text{R\&D}_{t-1})$	-0.538 (0.195)	***	-0.156 (0.096)		-0.546 (0.257)	**
$\log(\text{R\&D}_{t-1})^2$	0.053 (0.008)	***	0.007 (0.004)	*	0.053 (0.011)	***
$\log(\text{Sales}_{t-1})$	0.173 (0.027)	***	-0.032 (0.013)	**	0.172 (0.049)	***
$\log(\text{Fixed assets}_{t-1})$	0.005 (0.020)		0.024 (0.010)	**	0.007 (0.035)	
Age	-0.001 (0.006)		-0.008 (0.003)	**	-0.002 (0.011)	
Age^2	0.000 (0.000)		0.000 (0.000)	*	0.000 (0.000)	
Exporter	0.196 (0.076)	***	0.064 (0.037)	*	0.199 (0.109)	*
$\log(\text{Population density}_{t-1})$	0.018 (0.022)		-0.002 (0.012)		0.017 (0.024)	
Intercept	7.990 (1.321)	***	0.743 (0.652)		8.036 (1.499)	***
N	865		865		865	

Notes: Industry-year interaction dummy variables are included but not shown.
Heteroskedasticity-robust standard errors in parentheses.
90, 95 and 99 % confidence levels are denoted by "*", "**", and "***", respectively.

Table 4:
IV estimates, robustness analysis.

	(1)		(2)		(3)	
R&D subsidy grantee	1.459 (0.739)	**	1.657 (1.000)	*	1.452 (0.846)	*
$\log(\text{R\&D}_{t-1})$	0.039 (0.390)		0.298 (0.396)		-0.029 (0.318)	
$\log(\text{R\&D}_{t-1})^2$	0.025 (0.017)		0.012 (0.018)		0.028 (0.014)	*
$\log(\text{Sales}_{t-1})$	0.196 (0.038)	***	0.138 (0.063)	**	0.230 (0.037)	***
$\log(\text{Fixed assets}_{t-1})$	-0.009 (0.033)		0.049 (0.050)		-0.035 (0.031)	
Age	-0.006 (0.012)		-0.005 (0.009)		-0.002 (0.007)	
Age^2	0.000 (0.000)		0.000 (0.000)		0.000 (0.000)	
Exporter	0.113 (0.102)		0.210 (0.136)		0.000 (0.093)	
$\log(\text{Population density}_{t-1})$	0.032 (0.027)		0.010 (0.058)		-0.027 (0.066)	
$\log(\text{R\&D concentration}_{t-1})$					0.018 (0.031)	
$\log(\text{GDP per capita}_{t-1})$					0.096 (0.207)	
$\text{Unemployment rate}_{t-1}$					-0.622 (0.967)	
Share of sec. prod._{t-1}					-0.029 (0.282)	
Intercept	4.780 (2.079)	**	3.445 (2.241)		4.210 (2.484)	*
1^{st} stage regression						
ERDF Objective 1	0.169 (0.046)	***	0.109 (0.046)	**	0.145 (0.050)	***
N	1029		752		1611	

Notes: The estimates in Column (1) are based on a sample that only included firms that started up prior to 1994. The estimates in Column (2) are based on a sample that excludes firms in the ERDF 0 region, and those in Column (3) are based on a specification that includes regional characteristics at the NUTS 4 level. Industry-year interaction dummy variables are included but not shown. Heteroskedasticity-robust standard errors in parentheses.
90, 95 and 99 % confidence levels are denoted by "*", "**", and "***", respectively.