R&D SUBSIDIES AND COMPANY PERFORMANCE: EVIDENCE FROM GEOGRAPHIC VARIATION IN GOVERNMENT FUNDING BASED ON THE ERDF POPULATION-DENSITY RULE

Elias Einiö*

Abstract—Despite the prevalence of R&D support programs, evaluation studies based on explicit differences in support allocation are rare. In this paper, the identification of the causal effect of R&D support on company performance is based on geographic variation in government funding arising from a population-density rule. I find positive impacts on R&D investment, employment, and sales among the participants who were granted an R&D subsidy as a result of additional aggregate R&D support funding in their region. Although there are no instantaneous impacts on productivity, the study provides evidence of long-term productivity gains.

I. Introduction

ECONOMIC growth depends 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 & Helpman, 1991; Aghion & 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 R&D effort and company performance based on explicit differences in support policies. The main contribution of this paper is to provide such evidence by exploiting geographic variation in public R&D support funding arising from the EU 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 quality 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 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 seldom observed by the researcher. In this case, OLS and other methods based on the assumption that, conditional 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 United States 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, & Toole, 2000).

In this paper, the identification of the effect of R&D support program participation is based on an IV approach that exploits geographic variation in potentially available aggregate R&D-support funding arising from the allocation of European Union Regional Development Funds (ERDF) in Finland. These differences in allocation induce variation in the probability of R&D support program participation, which

Received for publication August 20, 2009. Revision accepted for publication March 11, 2013. Editor: Philippe Aghion.

* VATT (Valtion taloudellinen tutkimuskeskus) and Centre for Economic Performance, London School of Economics.

I thank Oriana Bandiera, Richard Blundell, Chiara Criscuolo, Steve Gibbons, Aki Kangasharju, William Kerr, Erkki Koskela, Teemu Lyytikäinen, Steve Machin, Katariina Nilsson Hakkala, Henry Overman, Tuomas Pekkarinen, Matti Sarvimäki, Mark Schankerman, Otto Toivanen, Janne Tukiainen, Roope Uusitalo, Jouko Verho, two anonymous referees, and numerous seminar participants for helpful comments and suggestions. I also thank Satu Nurmi for her invaluable help with data. The data used in this article are confidential, but my 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] Public R&D grants covered about 7.5% of private R&D in the OECD countries in 2004 (OECD, 2012). For an overview of R&D tax credits, 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 macroeconomic effects of public R&D support, see David et al. (2000). For an R&D-driven growth model incorporating R&D subsidies, see Davidson and Segerstrom (1998).

[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 (2013).

The Review of Economics and Statistics, October 2014, 96(4): 710–728

doi:10.1162/REST_a_00410

facilitates the identification of the causal effects of the program. The advantage of this approach is that it is based on explicit differences in public policies with well-defined publicly stated allocation criteria. The variation in R&D support program participation induced by regional provision of ERDF in Finland is especially suitable for program evaluation purposes because there are regions receiving the highest levels of EU 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, I address the potential endogeneity of this eligibility rule by controlling for the effect of population density on R&D effort and company performance.

I believe that results concerning the Finnish R&D support program may significantly improve understanding about the potential of R&D subsidies to enhance private innovative efforts in technologically advanced economies; Finland has a large export-oriented high-tech sector, and there are no other R&D policy instruments, such as tax exemptions, that would complicate the evaluation analysis. Using firm-level panel data on R&D support program participation and R&D investment constructed by linking a broad R&D survey to administrative information on all recipients of government R&D subsidies over the years 2000 to 2006, I find evidence of the positive effect of the program on R&D expenditure, employment, and sales among firms that were granted an R&D subsidy as a result of higher ERDF funding in their region. Furthermore, the results for the first half of the observation period, in which a larger amount of ERDF funding was allocated and for which the IV strategy provides the most precise estimates, indicate that government R&D assistance did not crowd out private R&D investment. While I find no instantaneous impacts on labor productivity, productivity gains emerge three years after the supported firms entered the program.

The rest of the paper is organized as follows. Section II describes the institutional setting of the R&D support policy in Finland, and section III presents the data. The empirical strategy for identifying the causal effects of the participation in the R&D support program is explained in section IV, section V presents the results and several robustness tests verifying the identification strategy, and section VI concludes.

II. The Institutional Setting

The implementation of the Finnish technology policy is centralized in the Finnish Funding Agency for Technology and Innovation (Tekes), the only authority responsible for allocating government subsidies to R&D projects conducted by private businesses. During the period 2000 to 2005, the agency granted direct R&D subsidies worth 968 million euros. The Tekes budget comes mainly from the Finnish government. Some areas receive funds from the ERDF in addition to national funding.

A. *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 R&D projects and accelerate the completion of ongoing ones. All firms operating in Finland may apply for funding. The main criteria for being selected into the program are 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 those that would not be fully implemented without agency funding are specifically supported.[5]

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% to 50% percent of the realized costs of the project up to an amount that the agency has approved as the maximum cost.[6] 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 EU 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 that the program may be ineffective, however. The agency may be unable to identify the projects for which its funding is of the essence. Also, it is not obvious whether the program is actually abiding by its stated

[5] http://www.tekes.fi, accessed April 18, 2007.

[6] The range for the compensation rate in the period 2000 to 2006 is based on information from the Tekes customer database provided by executive officer Ari Grönroos from Tekes Finance Department. The maximum allowed compensation rate for R&D subsidies in that period was 60% of the total costs (the ceiling for the total funding from the agency, including loans, was 70% of the total costs). However, if more than half of funding for a project conducted by a private company was from a government agency, all procurements for that project had to meet criteria laid down in the Public Procurement Act. For example, the act restricted companies' freedom to choose the most suitable contractor (section 7) and required that tender notices be made public (section 5). As a result, the compensation rate did not exceed 50% of the total costs among the supported private businesses in the observation period.

[7] *Official Journal of the European Union* C 45, 17.2.1996.

selection rules and EU regulations. In its own effectiveness analysis, it relies on figures on the average success of the supported 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 effectiveness of the program. Such an analysis requires a credible source of exogenous variation in the program-participation status. In order to identify the causal effects of government assistance, I exploit regional variation in the available government R&D funding.

B. *Regional Differences in Government R&D Funding*

The Finnish regional policy is based on districts delineated in accordance with EU criteria. Finland joined the EU 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 ERDF program period (1995–1999). Eligibility for the highest level of regional aid (objective 6) was based on population density. The criterion was that the population density of an area does not exceed 8 persons per km^2 in 1992 with the constraint that the whole area had to be connected and should not enclose ineligible areas.[8]

Figure 1 displays areas corresponding to the third (*maakunta*) and fourth (*seutukunta*) levels of the *Nomenclature of Territorial Units for Statistics* (NUTS) and boundaries of the regions eligible for the highest levels of regional aid in the ERDF program periods 1995–1999 (objective 6) and 2000–2006 (objective 1). NUTS3 areas are labeled by the name, while NUTS4 areas are labeled by the numeric area code. The population density rule was first applied to regional units corresponding to the NUTS2 level. This included two areas: Lappi and a region consisting of Koillismaa (178), Kainuu, Pohjois-Karjala, and Etelä-Savo. Because the Finnish NUTS2 areas are larger than in the EU on average, the EU agreed on an additional adjacent region consisting of smaller areas to be included in the objective 6 area. The Finnish government proposed in February 1994 that this adjacent region should cover the NUTS3 area of Etelä-Pohjanmaa and an interlinked area consisting of sixteen NUTS4 areas indicated by an underlined area code in figure 1.[9] This proposition was designed to maximize the population base under objective 6. However, the area was too large for the EU, which constrained the total population base of objective 6 to 840,000 inhabitants.[10] The final adjacent region covered 151,000 inhabitants and excluded Etelä-Pohjanmaa and interlinked NUTS4 areas with the highest population density with the exception that Nivala (176) was included in order to connect Pyhäntä (175) to Kaustinen (161) and Viitasaari (137). The region receiving the highest levels of aid was largely maintained for the ERDF program period 2000–2006 (the name was changed to objective 1).[11]

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. Table 1 shows direct subsidies granted by Tekes in the period 2000 to 2005.[12] The total amount of subsidies grew from 153 million euros in 2000 to 178 million euros in 2005. In 2000, ERDF funding was 2.9 million euros; it picked up 6.8 and 7.0 million euros in the second and third years of the program period, whereas it dropped to 5.0 million euros in 2003 and 4.1 million euros in 2004 and 2005. This is reflected in the proportion of ERDF funding to total subsidies: it decreased from 3.6% in the 2000–2002 period to 2.6% in the 2003–2005 period. Seventy-six percent of total ERDF funding was allocated to the objective 1 area, which received 8% of all subsidies, while it was responsible for 3% of national business sector R&D investment in the period 2000 to 2005. Subsidy funding covered 12.5% of R&D investment in the objective 1 area, compared to the national average of 4.6%. This suggests that ERDF funding may produce substantial variation in the probability of being accepted into the program across areas. Furthermore, the difference in the share of R&D investment covered by grants between the objective 1 area and the rest of the country (objectives 0, 2, and 4) was $13.4 - 4.5 = 8.9$ percentage points in the period 2000 to 2002 while it reduced to $11.7 - 4.3 = 7.4$ percentage points in the period 2003 to 2005, suggesting that variation in R&D support program participation arising from the ERDF fund allocation may be larger in the first three years of the ERDF program period when a larger amount of ERDF funding is allocated. This is verified in the empirical analysis, which indicates that the probability of receiving the subsidy is higher in the objective 1 area compared to areas eligible for lower levels of aid and that this difference is statistically and quantitatively more significant in the period 2000 to 2002.

III. Data

Data used in this study were constructed by linking several administrative data sets to a broad R&D survey, all maintained by Statistics Finland.

Data on R&D-support allocation were drawn from the Firm Assistance Database (FAD), which provides information on all applications submitted to Tekes during the period 2000 to 2006. Because Tekes is the only authority allocating R&D subsidies, the FAD information covers all government R&D subsidy recipients. Furthermore, because the

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

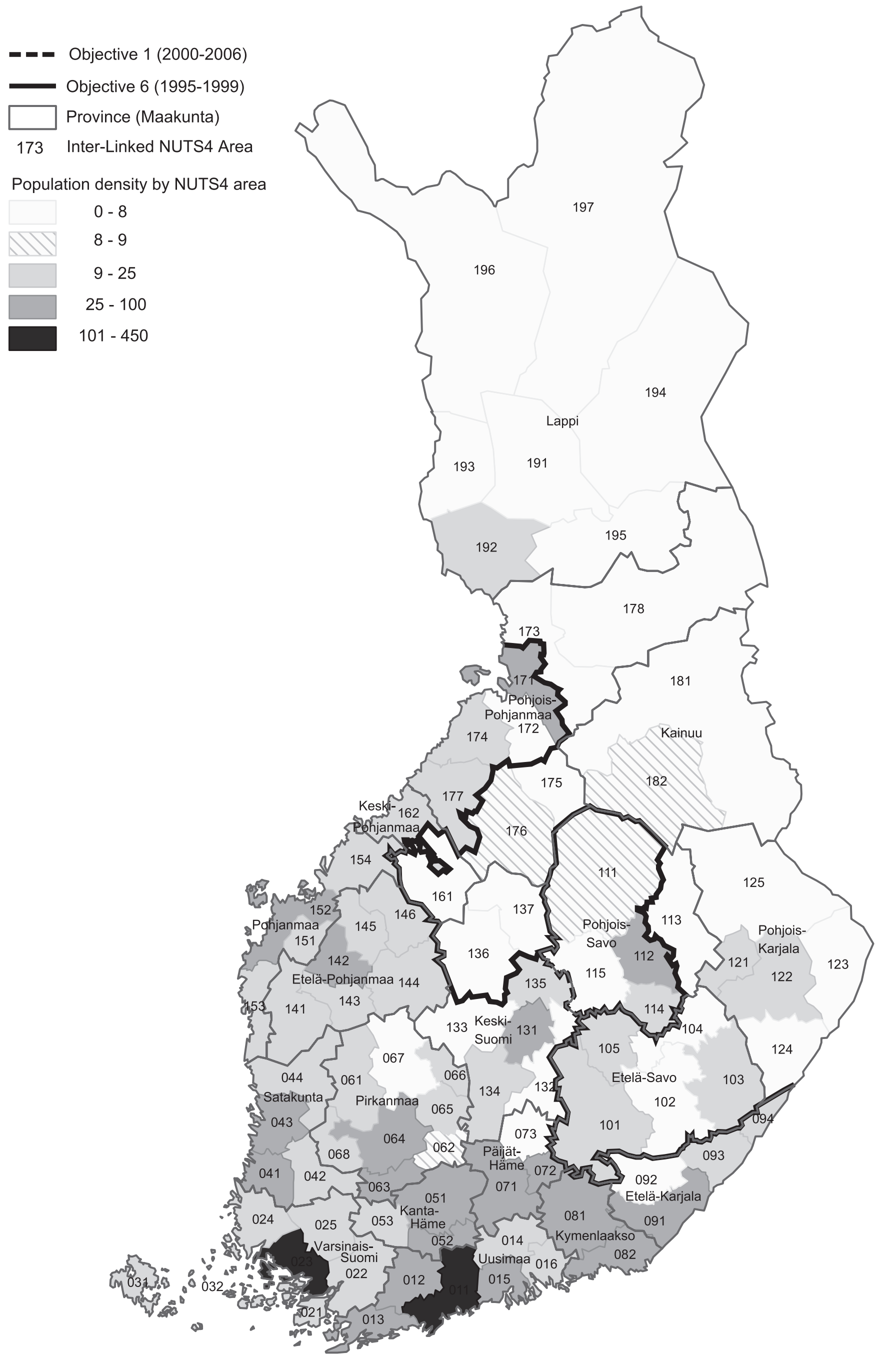
[9] See document EEC0763, February 28, 1994 (Archives of the Finnish Ministry for Foreign Affairs).

[10] See document EEC0905, March 11, 1994 (Archives of the Finnish Ministry for Foreign Affairs).

[11] See article 3 in the European Council regulation no. 1260/1999 on laying down the general provisions on the Structural Funds (*Official Journal of the European Union* L 161, 26.6.1999). The only change to the previous provision was that the region of Pohjois-Savo became eligible for the highest level of aid due to changes in the regional units on which eligibility was based.

[12] These are the years when firms in the treatment group enter the program. For a detailed definition of treatment and control groups, see section IV.

FIGURE 1.—POPULATION DENSITY BY NUTS4 AREA, 1992, AND AREAS ELIGIBLE FOR THE HIGHEST LEVEL OF ERDF FUNDING IN FINLAND IN PROGRAM PERIODS 1995–1999 (OBJECTIVE 6) AND 2000–2006 (OBJECTIVE 1)



NUTS3 areas (*maakunta*) are labeled by the name, while NUTS4 areas (*seutukunta*) are labeled by the numeric area code.

data include information on all subsidy recipients, it facilitated identifying whether any firm in any other database had received a subsidy. This was particularly important in the construction of the control group with which I compared the performance of the subsidized firms. The FAD contains the following project-level information: project identifier number, firm identifier number, amount applied for, decisions on the amount of subsidy or loans granted, if any, the date of the decision, and date and amount of each payment. For each project, I calculated annual subsidy flows by allocating the received amount of subsidy uniformly over each payment period. I then constructed the firm-level data by aggregating the annualized project-level data according to the firm identifier number.

Information on R&D expenditure was obtained from the annual R&D survey panel (RDS). The initial version was

TABLE 1.—R&D SUBSIDIES GRANTED BY THE FINNISH FUNDING AGENCY FOR TECHNOLOGY AND INNOVATION (TEKES), 2000–2005

Years	R&D Expenditure	Granted Subsidies	ERDF Funding of Granted Subsidies	% ERDF Funding of Granted Subsidies	% Granted Subsidies of R&D Expenditure
Whole country					
2000	3,136	153	2.9	1.9	4.9
2001	3,284	160	6.8	4.2	4.9
2002	3,376	157	7.0	4.5	4.7
2003	3,528	155	5.0	3.2	4.4
2004	3,684	165	4.1	2.5	4.5
2005	3,877	178	4.1	2.3	4.6
2000–2005	20,884	968	29.9	3.1	4.6
2000–2002	9,796	470	16.7	3.6	4.8
2003–2005	11,089	498	13.2	2.6	4.5
ERDF objective 1 area					
2000	91	13.9	2.5	17.9	15.3
2001	100	15.5	5.7	36.8	15.5
2002	112	11.2	4.1	36.6	10.0
2003	100	14.1	3.6	25.6	14.1
2004	116	12.1	3.3	27.0	10.4
2005	123	13.4	3.6	26.8	10.9
2000–2005	642	80.2	22.8	28.4	12.5
2000–2002	303	40.6	12.3	30.3	13.4
2003–2005	339	39.6	10.5	26.5	11.7
ERDF objective 0, 2, and 4 areas					
2000	3,045	139.1	0.5	0.3	4.6
2001	3,185	144.5	1.1	0.7	4.5
2002	3,263	145.8	2.9	2.0	4.5
2003	3,428	140.9	1.4	1.0	4.1
2004	3,567	152.9	0.8	0.5	4.3
2005	3,754	164.6	0.5	0.3	4.4
2000–2005	20,243	887.8	7.1	0.8	4.4
2000–2002	9,493	429.4	4.5	1.0	4.5
2003–2005	10,749	458.4	2.7	0.6	4.3

All monetary values are expressed in million euros.
Sources: *R&D expenditure*—Official Statistics of Finland (OSF): Research and development (e-publication); R&D Expenditure by Business Sector; ERDF 1 area figures have been calculated from NUTS3 and NUTS4 level data. *R&D subsidies*—Calculated from municipal level data provided by Tekes.

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 annually 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.[13] The response rate for the survey was around 85% to 87% during the period under observation.

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

The identification strategy of the study 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. The primary source of location was the RDS, which provides firm-level information on the regional distribution of R&D activities on the municipal level. A firm was defined as being located in a region[14] if it conducts all of its R&D there. A secondary source of location was the Business Register Plant-Level Database (BRPD),[15] which yields plant-level information: plant identification number, identification number of the parent firm, postal code for the location, and number of employees. I 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 2 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 treated firms entered the R&D support program.[16] The table shows that the group of supported firms, while 1.6 years younger on average, had slightly lower sales, higher fixed assets, and higher employment relative to the unsupported group in the pretreatment year $t-1$. In the year before being granted the subsidy, they were devoting, on average, 688,581 euros to R&D, while the average R&D expenditure in the control group was 357,597 euros. The distribution of R&D expenditure is very skewed, and the pretreatment differences between the treatment and control groups are less pronounced when the quartiles are compared. In the year following a positive grant decision, the grantees increased their average R&D expenditure to 866,204 euros and were receiving an average annual subsidy flow of 76,309 euros. The subsidies covered, on average, 25.3% 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 177,623 euros (a 25.8% increase), whereas among the unsupported group, the corresponding change was 50,464 euros (a 14.1% increase). A simple before-and-after estimate suggests that the average effect of R&D support was $177{,}623 - 50{,}464 = 127{,}159$ euros (or 0.289 in the average log change of R&D expenditure). In other words, the increase in R&D expenditure was 11.7 percentage points higher for the treatment group. Correspondingly, the increase in sales, fixed assets, and employment was 1.3, 4.1, and 6.6 percentage points higher for the treatment group, respectively. Columns 3 and 4 of table 2 present summary statistics for regions eligible for the lower levels of aid (ERDF 0, 2, and 4) and for the region eligible for the highest level of aid (ERDF 1). The table shows that firms in the less aided region, while having higher average sales and lower average fixed assets, spent more on R&D, on average. The average R&D expenditure increased

[13] 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 because of the nonrandom sampling of applicant firms. The robustness tests presented in section V indicate that such concern is not relevant for the results of this study.

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

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

[16] For a detailed definition of treatment and control groups, see section IV.

TABLE 2.—DESCRIPTIVE STATISTICS

	Analysis Sample					R&D Survey Sample[a]
	Controls (1)	Treated (2)	ERDF 0, 2, and 4 (3)	ERDF 1 (4)	All (5)	All (6)
R&D Expenditure$_{t+1}$						
Mean	408,061	866,204	504,588	231,253	482,176	1,034,117
25th percentile	46,520	91,100	52,410	43,000	52,000	70,000
Median	144,200	243,000	163,720	105,785	159,280	208,000
75th percentile	395,000	655,295	450,120	259,030	440,000	665,020
Standard deviation	984,505	2,119,033	1,298,469	381,146	1,251,066	3,422,879
Observations	1,513	292	1,657	148	1,805	5,490
R&D Expenditure$_{t-1}$						
Mean	357,597	688,581	431,271	185,772	411,141	938,871
25th percentile	45,200	58,165	47,930	34,012	46,487	68,116
Median	133,080	163,120	146,000	74,160	137,746	194,170
75th percentile	388,110	529,732	434,310	215,500	404,105	600,000
Standard deviation	710,674	1,553,301	941,830	324,042	909,597	3,063,241
Observations	1,513	292	1,657	148	1,805	5,490
Log-Change in R&D Expenditure from $t-1$ to $t+1$						
Mean	0.064	0.353	0.101	0.229	0.111	0.051
25th percentile	−0.318	−0.148	−0.281	−0.356	−0.287	−0.377
Median	0.055	0.311	0.086	0.189	0.092	0.063
75th percentile	0.445	0.860	0.498	0.716	0.511	0.488
Standard deviation	0.879	0.988	0.893	1.009	0.903	0.922
Observations	1,513	292	1,657	148	1,805	5,490
Treated						
Mean	0	1	0.154	0.250	0.162	0.164
25th percentile	0	1	0	0	0	0
Median	0	1	0	0	0	0
75th percentile	0	1	0	1	0	0
Standard deviation	0	0	0.361	0.434	0.368	0.370
Observations	1,513	292	1,657	148	1,805	1,858
Subsidy$_{t+1}$						
Mean	0	76,309	12,239	13,527	12,345	51,872
25th percentile	0	12,199	0	0	0	0
Median	0	32,672	0	0	0	0
75th percentile	0	78,580	0	294	0	31,562
Standard deviation	0	145,372	66,007	49,463	64,799	185,896
Observations	1,513	292	1,657	148	1,805	5,490
Subsidy intensity$_{t+1}$						
Mean	0	0.253	0.037	0.085	0.041	0.105
25th percentile	0	0.056	0	0	0	0
Median	0	0.134	0	0	0	0
75th percentile	0	0.276	0	0.004	0	0.098
Standard deviation	0	0.559	0.236	0.306	0.243	0.459
Observations	1,513	292	1,657	148	1,805	5,490
Age						
Mean	14.76	13.16	14.53	14.31	14.52	13.08
25th percentile	7	6	7	6	7	6
Median	12	11	12	11	12	11
75th percentile	20	18	20	17	20	18
Standard deviation	11	10.63	10.80	12.64	10.96	10.34
Observations	1,479	269	1,609	139	1,748	5,193
Sales$_{t+1}$						
Mean	38,470,858	38,264,149	39,071,843	31,411,114	38,437,072	47,961,270
25th percentile	1,675,044	1,043,498	1,571,260	1,627,680	1,572,594	875,000
Median	9,074,991	7,274,131	8,943,373	7,934,939	8,898,031	5,350,336
75th percentile	33,164,941	30,996,062	32,191,379	40,336,220	32,674,255	27,489,616
Standard deviation	109,651,573	113,291,022	113,903,958	54,654,568	110,221,967	217,824,744
Observations	1,474	288	1,616	146	1,762	5,383
Sales$_{t-1}$						
Mean	35,213,934	34,589,060	35,473,497	31,064,002	35,113,742	43,533,759
25th percentile	1,567,253	962,000	1,461,796	1,170,025	1,461,063	693,233
Median	8,643,941	7,420,000	8,519,547	7,033,763	8,415,000	4,761,000
75th percentile	29,826,315	26,314,000	27,437,093	42,149,338	28,288,784	24,372,537
Standard deviation	104,210,430	106,946,705	107,744,064	59,135,921	104,623,501	181,654,902
Observations	1,482	283	1,621	144	1,765	5,384

TABLE 2.—(CONTINUED)

	Analysis Sample					R&D Survey Sample[a]
	Controls (1)	Treated (2)	ERDF 0, 2, and 4 (3)	ERDF 1 (4)	All (5)	All (6)
Fixed assets$_{t+1}$						
Mean	17,716,013	28,446,630	19,247,516	21,931,828	19,469,939	45,594,615
25th percentile	171,094	181,317	167,617	195,320	173,141	155,074
Median	1,466,409	1,457,854	1,430,009	2,501,530	1,466,409	977,000
75th percentile	7,546,591	6,883,885	7,175,792	15,159,800	7,452,256	6,504,417
Standard deviation	70,239,420	170,677,568	97,433,317	47,188,739	94,287,273	513,656,923
Observations	1,474	288	1,616	146	1,762	5,383
Fixed assets$_{t-1}$						
Mean	17,259,094	26,647,637	18,549,832	21,180,423	18,764,452	43,370,076
25th percentile	184,567	197,825	185,000	219,097	191,000	136,000
Median	1,529,000	1,837,977	1,587,716	1,766,559	1,593,000	931,553
75th percentile	7,621,468	7,231,000	7,396,562	15,444,648	7,570,000	6,457,791
Standard deviation	65,455,485	143,469,655	85,429,386	49,243,520	83,063,273	469,534,231
Observations	1,482	283	1,621	144	1,765	5,384
Employment$_{t+1}$						
Mean	131.1	146.4	136.9	96.0	133.6	164.0
25th percentile	14.0	10.3	13.5	14.5	13.5	9.2
Median	49.7	46.3	49.7	46.6	49.6	35.0
75th percentile	143.3	169.9	147.6	156.5	148.1	136.0
Standard deviation	320.0	278.6	325.4	113.2	313.6	479.9
Observations	1,513	292	1,657	148	1,805	5,489
Employment$_{t-1}$						
Mean	132.1	138.3	136.3	97.2	133.1	165.5
25th percentile	13.5	8.7	12.5	10.8	12.3	8.0
Median	46.8	43.8	47.0	43.3	46.4	33.0
75th percentile	137.4	162.4	140.0	153.0	141.0	133.0
Standard deviation	332.7	278.1	336.5	120.5	324.4	463.0
Observations	1,511	292	1,655	148	1,803	5,483
Population density$_{1992}$						
Mean	170.0	175.2	184.7	15.9	170.8	172.2
25th percentile	28.3	27.5	36.6	5.9	27.5	28.3
Median	61.0	71.0	96.4	10.3	71.0	71.0
75th percentile	433.6	433.6	433.6	20.7	433.6	433.6
Standard deviation	177.9	181.7	179.8	15.2	178.4	178.8
Observations	1,513	292	1,657	148	1,805	5,484

All monetary values are expressed in euros. 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. Population density is based on the NUTS4 classification.
[a]The sample includes observations for which R&D expenditure is available for both years $t-1$ and $t+1$.

by 24.5% among the firms in the ERDF 1 region, whereas among the firms in the less-aided region, the corresponding increase was 17.0%. In the ERDF 1 area, average population density was 15.9 inhabitants per km^2 at the NUTS4 level in 1992. This is higher than the threshold of 8 on which eligibility for ERDF 1 was based on. This is due to the fact that the population density rule was applied at a coarser spatial scale corresponding to the broader NUTS2 level. As a result, some NUTS4 areas with population density above 8 became eligible for ERDF 1 support. Despite this, the eligibility rule induces a substantial difference in the average population density between the ERDF 1 area and the area eligible for lower levels of aid.

In summary, these descriptive statistics suggest that it is important to control for differences in the levels of pretreatment R&D expenditure and for other firm characteristics when the treatment and control groups are compared. However, even after controlling for observable characteristics, such a comparison is likely to suffer from a selection bias, as it is unlikely that all attributes affecting the selection into the program are observed. In order to address the selection problem, I use an IV approach that exploits geographic variation in government R&D assistance arising from the asymmetric regional allocation of ERDF funds.

IV. The Empirical Strategy

This section explains the empirical strategy adopted for identifying the causal effects of the R&D support program participation. The primary aim is to assess whether, as a result of becoming a program participant, the firm increased its R&D expenditure. The "treatment group" comprises the firms that did not receive any R&D support in the year $t-1$, entered the R&D support program in the year t, and received R&D support in the year $t+1$, while the "control group" comprises the 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 baseline specification compares growth rates of R&D investment between the treatment and control groups:

$$y_{i,t+1} - y_{i,t-1} = \alpha + \gamma D_{it} + \epsilon_{it}, \tag{1}$$

where i and t indicate the firm and year, respectively, and y_{it} is the logarithm of R&D investment. The treatment dummy variable D_{it} is 1 (0) if firm i was in the treatment (control) group in the year t. This model is equivalent to a fixed-effects specification in log levels and controls for any time-invariant unobserved heterogeneity at the firm, industry, and region level in R&D expenditure.[17] I examine changes from the last pretreatment year $t-1$ to the second participation year $t+1$ for two reasons. First, because grant 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 decision, in many cases the support does not completely cover the first year. Second, a larger fraction of the cost is likely to be covered in the second year because only a very small proportion of subsidized projects last less than one year.[18]

OLS estimates based on equation (1) will be biased if omitted attributes correlated with the treatment status affect R&D investment. For example, time patterns of R&D investment may differ across industries, and R&D support allocation may be partly determined by expected industry growth in R&D investment. To account for this, I augment equation (1) with industry-year fixed effects, $\phi_{j(i),t}$. In order to control for observed differences between the treatment and control group, I also include terms for pretreatment characteristics $X_{i,t-1}$ and estimate the following equation,

$$y_{i,t+1} - y_{i,t-1} = \phi_{j(i),t} + \gamma D_{it} + \beta X_{i,t-1} + v_{it}, \tag{2}$$

where j indicates the industry. The vector of pretreatment characteristics $X_{i,t-1}$ includes the log of sales and fixed assets in order to 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 the firm is exporting. To address the potential concern that growth rates of R&D investment may vary by pretreatment levels of R&D investment, I also estimate models including a second-order polynomial of logged R&D in the year $t-1$.[19]

The parameter of interest is γ, the causal effect of R&D support program participation on the log change in R&D expenditure. When $\gamma \leq 0$, the program does not induce R&D investment and the firms would have made at least the same investments in the absence of the subsidy. If $\gamma > 0$, the support scheme induces R&D investment. 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 euros of R&D investment that 1 subsidy euro induces. In the case where a fixed proportion of R&D that would not have been realized without the subsidy is supported, 1 subsidy euro will generate R&D worth the inverse of the compensation rate (the share of realized R&D costs covered by the subsidy). 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, when the subsidy is the largest possible, 50% of the firm's total posttreatment R&D expenditure, a necessary condition for no crowding out is that $\gamma \geq \log(1/0.5) \approx 0.693$, while the corresponding condition allowing for $100 \times h$ percent of public spending to crowd out private R&D investment is $\gamma \geq \log(1/(1-0.5(1-h)))$.[20]

A. The IV Approach

The R&D support program selects the participants partly according to technological potential and expectations of commercial success, which are not observed. Because the firm's investment decisions also depend on these factors, the program participation status D_{it} and the error term v_{it} in equation (2) may be correlated, which would result in biases in the OLS estimates of the causal effect of the program, γ. In order to overcome this selection problem, I use an IV approach that exploits regional variation in public R&D support funding across the border of the ERDF 1 area that received the highest levels of funding and was determined by population density (see section II for a detailed description of the eligibility rule).[21] As presented in section IIB, 76% of the agency's ERDF funding was allocated to the objective 1 area. This resulted in substantial differences in the proportion of R&D investment covered by public R&D funding across the objective 1 border and in a higher likelihood of winning an R&D grant in the objective 1 area relative to other areas. The identification of the impact of R&D support program participation

[17] To see this, consider a fixed-effects specification $y_{is} = \alpha_i + \gamma T_{is} + \omega_{is}$ where T_{is} is equal to 0 for firms in the treatment and control groups if $s = t-1$ and equal to 1 (0) for firms in the treatment (control) group if $s = t+1$. Taking differences from $t-1$ to $t+1$ on both sides yields equation (1).

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

[19] It is worth noting that when pretreatment R&D in the year $t-1$ is included in the model as a control variable, replacing the log-differenced outcome in equation (2) with the log level of R&D in $t+1$, that is, $y_{i,t+1}$, recovers an equivalent estimate of γ.

[20] 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—the entire posttreatment R&D cost cz_0 is eligible for the maximum compensation rate of 0.5. In this case, 1 subsidy euro will generate $(cz_0 - z_0)/0.5cz_0$ euro of private R&D. The no-crowding-out condition requires that 1 subsidy euro induces R&D worth at least 1 euro. This implies $c \geq 1/0.5$ for no crowding out under maximum compensation. The condition for no more than $100 \times h$ percent of public spending crowding out private R&D investment is $c \geq 1/(1-0.5(1-h))$ and can be derived correspondingly by requiring that 1 subsidy euro induces R&D worth at least $1-h$ euro. Statistical testing of crowding out is based on a one-sided test of the null hypothesis that $100 \times h$ percent of public spending crowds out company R&D investment, $\gamma = \log(1/(1-0.5(1-h)))$, against the alternative hypothesis $\gamma > \log(1/(1-0.5(1-h)))$.

[21] Criscuolo et al. (2012) use variation in subsidy intensities across regions and over time in a study evaluating an investment subsidy program in the United Kingdom, while Hyytinen and Toivanen (2005) use regional variation in R&D support in Finland in a study concerning the relationship between financial constraints and innovation. However, they do not exploit regional variation in aggregate potential level of support as a source of exogenous variation in R&D support program participation.

is based on this variation in the likelihood of program participation arising from the asymmetric allocation of ERDF funding.

The analysis is implemented by using a binary indicator for ERDF 1 area as an instrument for the R&D support program participation status D_{it} in equation (2). This approach has two advantages compared to earlier studies relying on industry variation in government R&D funding (Lichtenberg, 1988; Wallsten, 2000). First, industry variation in R&D support funding is likely to be endogenous because 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 et al., 2000). This is not a relevant concern in this study because the IV strategy is not relying on variation across industries as industry-year fixed effects are controlled for in the R&D equation. Second, in this study, variation in government R&D support funding is based on explicit differences in public policies with well-defined and publicly stated rules for eligibility. Furthermore, the 1992 population density on which eligibility for the highest level of regional aid was based on is observed. Thus, I can address the potential concern that eligibility for the highest levels of aid is endogenous (i.e., that firm-level R&D investment is determined by population density) by augmenting the R&D equation with the 1992 population density:

$$y_{i,t+1} - y_{i,t-1} = \phi_{j(i),t} + \gamma_i D_{it} + \beta X_{i,t-1} + \xi \log(POPDEN_{r(i),1992}) + u_{it}, \quad (3)$$

where r indicates area at the NUTS4 level.[22]

The key identifying assumption of the IV analysis is that conditional on population density and firm characteristics included in $X_{i,t-1}$, the ERDF 1 indicator is uncorrelated with the growth rate of R&D investment. In other words, the IV approach recovers an unbiased estimate of the treatment effect provided that conditional unobserved trends are similar across areas eligible and ineligible for ERDF 1 funding. The strength of an empirical setup where the eligibility determining variable is observed is that including it in the model controls for confounding differences that may arise from the allocation rule of regional aid. It is possible, for example, that in more densely populated areas, there is more formal and informal interaction between workers, which may result in larger knowledge spillovers (Glaeser, 1999) and higher expected returns to R&D investment. In this case, the low-density areas selected into the ERDF 1 would be expected to have, on the average, lower growth rates of R&D investment. Controlling for population density will address the potential bias arising from this type of regional selection. Because eligibility for the highest level of ERDF aid is based on population density, unobserved regional attributes that are uncorrelated with population density are also likely to be uncorrelated with the ERDF instrument. However, the possibility that the ERDF 1 areas have some unobserved characteristics that are unrelated to population density and may affect R&D investment cannot be completely ruled out. In the robustness analysis of section V, I test for the common trends assumption and provide evidence that unobserved regional factors are unlikely to confound the results.

Equation (3) also makes explicit the fact that the effect of R&D support program participation may differ across participants, $\gamma_i \neq \gamma_j$ 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), that is, the average effect of the treatment among the participants whose treatment status the instrument changes.[23] In the context of this study, LATE is the effect of R&D support program participation among the firms entering the R&D support program as a result of higher levels of government R&D support funding in the ERDF 1 region.

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—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. 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.

V. Results

This section presents the results of the paper. I first report estimates of the effect of R&D support program participation on R&D expenditure. Then I investigate whether program participation affected employment, fixed assets, sales, and labor productivity and address a number of potential robustness concerns. I also report estimates of the long-term impacts of the R&D support program and conclude the section with a

[22] An alternative approach would be to use a fuzzy regression discontinuity (FRD) approach using a discontinuity at the population density threshold of 8. This would be implemented by replacing the ERDF 1 dummy instrument with an indicator equal to 1 if population density in the area in which the firm is located is less than or equal to 8 and 0 otherwise. However, the population density rule was applied at the spatial scale corresponding to NUTS2 level, and there are only three such areas eligible for the highest level of aid. Furthermore, there are only six eligible areas at the spatial scale corresponding to the NUTS3 level. This is clearly insufficient for FRD estimation. Moreover, the rule does not induce sufficient discontinuity at the threshold at the NUTS4 level because it was applied at a coarser spatial scale and the population base of the area was capped at 840,000 thousand inhabitants, and as a result some ineligible NUTS4 areas are just below the threshold and some eligible NUTS4 areas are just above the threshold (see figure 1 for the spatial distribution of population density at the NUTS4 level).

[23] Abadie (2003) generalizes this result for the case in which covariates may be included in the model and shows that the 2SLS estimator recovers the LATE parameter provided that the probability of participation is linear in the instruments.

discussion of the findings. In all estimations, standard errors are corrected for clustering at the NUTS4 area level.

A. *R&D Investment*

Table 3 reports the estimation results for R&D investment. The table includes panels for the full sample period and for subperiods 2000 to 2002 and 2003 to 2005. Each panel shows treatment effect estimates and coefficients on the log of 1992 population density for the least squares (OLS) and instrumental variables (IV) procedures; coefficients on the ERDF 1 instrument and coefficients on the log of 1992 population density for the reduced-form equation (RF) and the first-stage equation (1st) of the IV procedure; and *p*-values for one-sided tests of 0%, 15%, and 50% of R&D subsidies crowding out company R&D investment where the null hypothesis is $\gamma = 0.693$, $\gamma = 0.553$, and $\gamma = 0.287$, respectively (for details see note 20). In column 1, the OLS estimate of the treatment effect for the the baseline specification in equation (1) is 0.288 and highly significant. The estimates are very similar when industry-year fixed effects (column 2), pretreatment firm characteristics (column 3), and population density (column 4) are included in the model, whereas it increases to 0.340 for the specification controlling for pretreatment R&D investment (column 5).

If selection into the R&D support program is affected by unobservable firm characteristics that also affect the R&D investment decision, the OLS estimates are likely to be biased. In order to address the selection problem, I use an ERDF region dummy variable as an instrument for R&D support program participation status. The instrument is based on a binary area classification in which the first class includes the region receiving the most aid because of low population density (ERDF 1) and the second class includes all those receiving less aid (ERDF 0, 2 and 4).[24] In panel A reporting results for the full sample, the first-stage coefficients on the ERDF 1 dummy variable are between 0.096 and 0.149 and highly significant across specifications, indicating that regional differences in available R&D support funding induce substantial differences in the probability of receiving R&D support. The reduced-form estimate is insignificant for specifications excluding pretreatment control variables (columns 1 and 2), whereas it becomes significant once pretreatment controls and population density are included in the model (column 4). Further controlling for pretreatment, R&D investment has virtually no impact on the estimate but improves the precision of the estimation by lowering the standard error by 11%. The estimate of 0.205 suggests that average R&D investment increased more in the ERDF 1 region, where a larger fraction of firms received an R&D subsidy. However, because some firms in that region did not receive a subsidy and some firms in the regions receiving less aid did receive one, the reduced-form coefficient on the instrument underestimates the treatment effect.

The causal effect of the R&D support program is quantified by the IV procedure where the binary ERDF 1 indicator is used as an instrument for the R&D support treatment. The IV estimate of the treatment effect for the baseline specification is 1.335 with a standard error of 1.285. Including industry-year fixed effects and pretreatment controls increases the estimate to 1.348 and reduces the standard error to 0.817, making the coefficient significant at the 10% risk level. Because population density is correlated with the ERDF 1 instrument as a result of the population-density rule (ERDF 1 region has a lower population density), excluding it from the model will induce bias in the IV estimates if population density also affects growth in R&D investment, because in this case, the ERDF instrument and the second-stage error would be correlated. Including population density in the model further increases the estimate to 1.439 and reduces the standard error to 0.739. The negative bias in the estimate in column 3 compared to the estimate in column 4 arises because population density is negatively correlated with the ERDF instrument and positively correlated with R&D expenditure in this sample. Finally, controlling for lagged R&D investment lowers the estimate to 1.380 and further improves the precision of the estimation, making the estimate significant at the 5% risk level. The comparison of the estimates in columns 4 and 5 suggests that the potential endogeneity of lagged R&D investment is not driving the latter.

The estimate in column 5 is larger than the crowding-out threshold of 0.693, derived in the previous section. In the extreme case of a maximum subsidy compensation of 50% of the total posttreatment R&D cost, this point estimate suggests that 1 subsidy euro induced R&D investment worth 1.49 euros, although it is worth emphasizing that for this sample, the possibility that government funding may have crowded out some private R&D investment cannot be ruled out as the null hypothesis of $\gamma = 0.693$ is not rejected, whereas 15% (50%) crowding out is rejected at the 10% (5%) risk level.[25]

Differences in the probability of receiving R&D support induced by regional variation in R&D support funding are expected to be more pronounced in the first years of the observed ERDF program when a larger share of ERDF funding is allocated (see table 1). Panels B and C of table 3 show the results for two subsamples covering the periods 2000 to 2002 and 2003 to 2005. The first-stage coefficients on the ERDF 1 instrument are larger in the 2000–2002 sample than in the full sample across specifications. This generates larger reduced-form estimates, which are now significant at the 5% risk level for all specifications. The IV estimates of the treatment effect are also significant across specifications, with the smallest and most precise estimate of 1.850 in column 5. This point estimate suggests that 1 subsidy euro induced R&D investment worth 1.68 euros. The null hypothesis of no crowding out is rejected across specifications indicating

[24] For a detailed description of the criteria used for location, see section III.

[25] For details of this calculation and the crowding-out threshold, see note 20.

TABLE 3.—ESTIMATED EFFECTS OF R&D SUPPORT PROGRAM PARTICIPATION ON R&D INVESTMENT, OLS, REDUCED FORM, AND IV ESTIMATES

	(1)	(2)	(3)	(4)	(5)
			A. All		
OLS					
R&D subsidy grantee	0.288***	0.283***	0.293***	0.292***	0.340***
	(0.055)	(0.060)	(0.062)	(0.062)	(0.048)
log(Population $Density_{1992}$)				−0.004	0.013
				(0.021)	(0.018)
RF					
ERDF objective 1	0.128	0.114	0.177*	0.202**	0.205**
	(0.101)	(0.100)	(0.100)	(0.100)	(0.089)
log(Population $Density_{1992}$)				0.013	0.029
				(0.025)	(0.022)
1st					
ERDF objective 1	0.096**	0.123**	0.131***	0.140***	0.149***
	(0.044)	(0.046)	(0.044)	(0.048)	(0.048)
log(Population $Density_{1992}$)				0.005	0.003
				(0.007)	(0.007)
IV					
R&D subsidy grantee	1.335	0.930	1.348*	1.439*	1.380**
	(1.285)	(0.833)	(0.817)	(0.794)	(0.641)
log(Population $Density_{1992}$)				0.006	0.025
				(0.025)	(0.023)
Crowding-out test (p-value)					
0%	0.309	0.388	0.211	0.174	0.142
15%	0.271	0.325	0.165	0.132	0.098
50%	0.207	0.220	0.097	0.073	0.044
Observations	1,805	1,805	1,703	1,703	1,703
			B. Years 2000–2002		
OLS					
R&D subsidy grantee	0.258***	0.234***	0.209**	0.207**	0.292***
	(0.090)	(0.086)	(0.100)	(0.099)	(0.076)
log(Population $Density_{1992}$)				−0.014	0.002
				(0.020)	(0.019)
RF					
ERDF objective 1	0.360**	0.306**	0.363***	0.420***	0.332***
	(0.153)	(0.144)	(0.129)	(0.134)	(0.124)
log(Population $Density_{1992}$)				0.028	0.033
				(0.027)	(0.028)
1st					
ERDF objective 1	0.150**	0.159***	0.151***	0.154**	0.179***
	(0.058)	(0.052)	(0.052)	(0.059)	(0.057)
log(Population $Density_{1992}$)				0.002	0.004
				(0.011)	(0.011)
IV					
R&D subsidy grantee	2.402**	1.924**	2.410***	2.724***	1.850***
	(1.127)	(0.841)	(0.882)	(1.043)	(0.672)
log(Population $Density_{1992}$)				0.023	0.026
				(0.037)	(0.031)
Crowding-out test (p-value)					
0%	0.065	0.072	0.026	0.026	0.043
15%	0.050	0.052	0.018	0.019	0.027
50%	0.030	0.026	0.008	0.010	0.010
Observations	871	871	820	820	820
			C. Years 2003–2005		
OLS					
R&D subsidy grantee	0.326***	0.333***	0.384***	0.385***	0.402***
	(0.066)	(0.062)	(0.076)	(0.075)	(0.054)
log(Population $Density_{1992}$)				0.006	0.020
				(0.033)	(0.027)
RF					
ERDF objective 1	−0.116	−0.101	−0.019	−0.012	0.061
	(0.144)	(0.172)	(0.178)	(0.182)	(0.150)
log(Population $Density_{1992}$)				0.004	0.023
				(0.034)	(0.029)
1st					
ERDF objective 1	0.035	0.082	0.105	0.119	0.120
	(0.060)	(0.068)	(0.068)	(0.073)	(0.073)
log(Population $Density_{1992}$)				0.008	0.006
				(0.010)	(0.010)

TABLE 3.—(CONTINUED)

	(1)	(2)	(3)	(4)	(5)
	C. Years 2003–2005				
IV					
R&D subsidy grantee	−3.287	−1.242	−0.185	−0.097	0.504
	(6.872)	(2.366)	(1.610)	(1.444)	(1.118)
log(Population Density$_{1992}$)				0.005	0.020
				(0.031)	(0.027)
Crowding-out test (p-value)					
0%	0.719	0.793	0.707	0.708	0.567
15%	0.712	0.776	0.677	0.674	0.517
50%	0.698	0.741	0.615	0.605	0.423
Observations	934	934	883	883	883
Industry-year fixed effects	No	Yes	Yes	Yes	Yes
Pretreatment controls	No	No	Yes	Yes	Yes
Lagged outcome	No	No	No	No	Yes

The outcome is a log difference from $t-1$ to $t+1$. Standard errors clustered by NUTS4 area are in parentheses. Industry-year fixed effects are based on two-digit NACE classification. Population density is based on 1994 NUTS4 area classification. Pretreatment controls are a quadratic term for age, indicator for exporter, log of employment, log of fixed assets, and log of sales in the year $t-1$. The specification in column 5 includes a quadratic term for the logarithm of R&D investment in the year $t-1$. The 90%, 95%, and 99% confidence levels for a null hypothesis that the coefficient equals 0 are denoted by *, **, and ***, respectively. The crowding-out test is based on a one-sided test of the IV coefficient on the treatment indicator being equal to 0.693, 0.553, and 0.287 for a test of 0%, 15%, and 50% of R&D subsidies crowding out company R&D investment, respectively (see note 20 for details). OLS: ordinary least squares; RF: reduced form; 1st: IV first stage; IV: IV second stage.

that public R&D funding did not crowd out private R&D investment. The first-stage coefficient on the ERDF 1 instrument shows a decreasing trend over time in panels B and C. The reduced strength of the instrument is also reflected in the precision of the IV estimates: for the preferred specification in column 5, the estimate of the treatment effect is 0.504 with a standard error of 1.118 in the period 2003 to 2005. An important finding of this subperiod analysis is that the differences in the probability of receiving R&D support induced by regional variation in public funding are more pronounced in the first years of the ERDF program period, when a larger fraction of ERDF aid is allocated. This indicates that it is indeed the additional funding provided by the ERDF that induces higher acceptance probability for R&D grant applications in the objective 1 area eligible for the highest level of regional aid.

B. *Employment, Fixed Assets, and Labor Productivity*

In order to assess whether participating in the R&D support program has had any instantaneous effects on firm-level inputs and productivity, I estimate specifications corresponding to those in table 3 for employment, fixed assets, sales, and labor productivity. Table 4 presents a panel of results for each outcome. In panel A, the IV estimate of the treatment effect on employment is insignificant for specifications excluding pretreatment control variables (columns 1 and 2) and including pretreatment firm characteristics (column 3). The coefficient increases to 0.712 and becomes significant once population density is included in the model. A positive effect on employment is what would be expected as R&D is a labor-intensive activity and the previous analysis provided evidence that the program increased R&D investment. As in the case of R&D investment, the negative bias in the specification excluding population density arises because it is negatively correlated with the ERDF 1 eligibility as a result of the population density rule, and it is positively correlated with employment growth rate as indicated by the second-stage coefficient on it. Controlling for lagged employment further increases the estimate to 0.848.

In panel B, the IV estimates of the treatment effect on fixed assets are insignificant across specifications. The estimate including all control variables in column 5 is 0.121 with a standard error of 0.520. It is worth noting that in this panel, population density is negatively correlated with growth in fixed assets, and as a result, excluding it from the model induces a positive bias in the IV estimates of the treatment effect. In panel C, the IV estimates of the treatment effect on sales are significant at the 5% risk level across specification. The estimate in column 5 is of a very similar magnitude as the corresponding treatment effect on employment. This suggests that the R&D support program did not have any instantaneous effects on labor productivity. In panel D, the IV estimate of the treatment effect on labor productivity is insignificant for all specifications except the baseline model (column 1). The preferred estimate in column 5 is 0.098 with a standard error of 0.403.[26]

C. *Robustness Analysis*

Tables 5 and 6 present results from several additional specifications in order to verify the empirical strategy to identify the causal effects of R&D support program participation and to assess the validity of the ERDF instrument. The estimations are based on a specification corresponding to column 5 of tables 3 and 4 unless otherwise stated.

I start by reporting the results for R&D investment (table 5), the main target of the R&D support program. The key identifying assumption of the IV analysis is that conditional

[26] I also estimated a similar specification for R&D intensity measured as the ratio of R&D expenditure to sales. The IV estimate (standard error) of the treatment effect was 0.448 (0.762) and 0.290 (0.634) for specifications corresponding to columns 4 and 5 in panel D of table 4, respectively.

TABLE 4.—ESTIMATED EFFECTS OF R&D SUPPORT PROGRAM PARTICIPATION ON EMPLOYMENT, FIXED ASSETS, SALES, AND LABOR PRODUCTIVITY, OLS, REDUCED FORM, AND IV ESTIMATES

	2000–2005					2000–2002
	(1)	(2)	(3)	(4)	(5)	(6)
			A. Employment			
OLS						
R&D subsidy grantee	0.080***	0.073***	0.052***	0.053***	0.067***	0.030
	(0.020)	(0.017)	(0.018)	(0.018)	(0.018)	(0.028)
log(Population $Density_{1992}$)				0.011	0.013	0.007
				(0.007)	(0.010)	(0.011)
RF						
ERDF objective 1	0.007	0.045	0.063	0.103**	0.121**	0.225***
	(0.031)	(0.038)	(0.039)	(0.045)	(0.047)	(0.084)
log(Population $Density_{1992}$)				0.020*	0.024*	0.031
				(0.011)	(0.013)	(0.019)
1st						
ERDF objective 1	0.096**	0.122**	0.136***	0.144***	0.142***	0.168***
	(0.044)	(0.046)	(0.045)	(0.049)	(0.048)	(0.059)
log(Population $Density_{1992}$)				0.004	0.003	0.002
				(0.007)	(0.007)	(0.011)
IV						
R&D subsidy grantee	0.073	0.370	0.465	0.712**	0.848**	1.344**
	(0.311)	(0.308)	(0.291)	(0.347)	(0.346)	(0.653)
log(Population $Density_{1992}$)				0.017*	0.021	0.029
				(0.011)	(0.013)	(0.024)
Observations	1,801	1,801	1,703	1,703	1,703	820
			B. Fixed Assets			
OLS						
R&D subsidy grantee	0.079**	0.078**	0.078**	0.076**	0.092***	0.070
	(0.033)	(0.032)	(0.033)	(0.034)	(0.034)	(0.043)
log(Population $Density_{1992}$)				−0.019*	−0.025**	−0.025
				(0.010)	(0.011)	(0.015)
RF						
ERDF objective 1	0.114	0.064	0.063	0.029	0.017	−0.013
	(0.074)	(0.070)	(0.071)	(0.075)	(0.080)	(0.108)
log(Population $Density_{1992}$)				−0.018	−0.025*	−0.027
				(0.012)	(0.012)	(0.018)
1st						
ERDF objective 1	0.100**	0.127***	0.133***	0.136***	0.140***	0.166***
	(0.045)	(0.046)	(0.045)	(0.049)	(0.048)	(0.060)
log(Population $Density_{1992}$)				0.002	0.004	0.003
				(0.007)	(0.007)	(0.012)
IV						
R&D subsidy grantee	1.143	0.507	0.474	0.212	0.121	−0.078
	(0.736)	(0.485)	(0.472)	(0.492)	(0.520)	(0.597)
log(Population $Density_{1992}$)				−0.018*	−0.025**	−0.027*
				(0.011)	(0.011)	(0.016)
Observations	1,725	1,725	1,677	1,677	1,677	805
			C. Sales			
OLS						
R&D subsidy grantee	0.105***	0.104***	0.077***	0.077***	0.058**	−0.014
	(0.026)	(0.028)	(0.027)	(0.028)	(0.026)	(0.043)
log(Population $Density_{1992}$)				0.002	0.004	0.009
				(0.010)	(0.011)	(0.010)
RF						
ERDF objective 1	0.094**	0.102***	0.104**	0.128***	0.125**	0.193**
	(0.039)	(0.038)	(0.042)	(0.048)	(0.052)	(0.082)
log(Population $Density_{1992}$)				0.012	0.014	0.030*
				(0.015)	(0.016)	(0.016)
1st						
ERDF objective 1	0.104**	0.130***	0.135***	0.140***	0.139***	0.167***
	(0.046)	(0.048)	(0.047)	(0.051)	(0.048)	(0.060)
log(Population $Density_{1992}$)				0.003	0.003	0.002
				(0.007)	(0.007)	(0.011)
IV						
R&D subsidy grantee	0.913**	0.788**	0.774**	0.912**	0.895**	1.159**
	(0.366)	(0.318)	(0.327)	(0.392)	(0.446)	(0.584)
log(Population $Density_{1992}$)				0.010	0.011	0.028
				(0.015)	(0.016)	(0.019)
Observations	1,741	1,741	1,682	1,682	1,682	807

TABLE 4.—(CONTINUED)

	2000–2005					2000–2002
	(1)	(2)	(3)	(4)	(5)	(6)
	D. Labor Productivity					
OLS						
R&D subsidy grantee	0.032*	0.033	0.013	0.013	−0.014	−0.035
	(0.018)	(0.021)	(0.021)	(0.021)	(0.021)	(0.031)
log(Population Density$_{1992}$)				−0.007	−0.008	0.004
				(0.006)	(0.007)	(0.009)
RF						
ERDF objective 1	0.064***	0.036	0.038	0.026	0.014	−0.029
	(0.023)	(0.036)	(0.040)	(0.046)	(0.059)	(0.103)
log(Population Density$_{1992}$)				−0.006	−0.008	0.000
				(0.008)	(0.010)	(0.014)
1st						
ERDF objective 1	0.103**	0.130***	0.135***	0.139***	0.138***	0.165***
	(0.046)	(0.048)	(0.047)	(0.051)	(0.048)	(0.060)
log(Population Density$_{1992}$)				0.002	0.004	0.002
				(0.007)	(0.008)	(0.011)
IV						
R&D subsidy grantee	0.619**	0.277	0.285	0.188	0.098	−0.177
	(0.292)	(0.263)	(0.281)	(0.307)	(0.403)	(0.586)
log(Population Density$_{1992}$)				−0.007	−0.009	0.001
				(0.007)	(0.009)	(0.012)
Observations	1,740	1,740	1,682	1,682	1,682	807
Industry-year fixed effects	No	Yes	Yes	Yes	Yes	Yes
Pretreatment controls	No	No	Yes	Yes	Yes	Yes
Lagged outcome	No	No	No	No	Yes	Yes

The outcome is a log difference from $t-1$ to $t+1$. Standard errors clustered by NUTS4 area are in parentheses. Industry-year fixed effects are based on two-digit NACE classification. Population density is based on 1994 NUTS4 area classification. Pretreatment controls include a quadratic term for age, indicator for exporter, log of R&D investment, log of employment, log of fixed assets, and log of sales in the year $t-1$ excluding the term for the lagged outcome (for labor productivity, both sales and employment are excluded). Specifications in columns 5 and 6 include a quadratic term for the logarithm of the level of the lagged outcome in the year $t-1$. The 90%, 95%, and 99% confidence levels for a null hypothesis that the coefficient equals 0 are denoted by *, **, and ***, respectively. OLS: ordinary least squares; RF: reduced form; 1st: IV first stage; IV: IV second stage.

on pretreatment control variables and population density, the ERDF 1 instrument and growth in R&D investment are independent. In order to assess whether regional attributes other than population density are driving the results, I present estimates for three additional specifications. First, I exclude firms located in the least-aided ERDF 0 region, which includes the most economically advantaged areas (mainly the largest cities and their surrounding areas), from the sample. The results are displayed in column 1. The treatment effect for this specification is larger than the corresponding full sample estimate in column 5 of table 3, although it has lower precision due to the smaller sample size. Second, I exclude from the sample the NUTS4 areas of Iisalmi (111), Suonenjoki (115), Varkaus (114), and Kuopio (112), which were not included in the ERDF 1 area in the first ERDF program period but became eligible in 2000. The treatment effect for this specification is 1.971 and significant at the 10% risk level. Third, I analyze whether key regional features have an effect on the results by including in the model a set of regional control variables that could affect company R&D on the NUTS4 level. In order to control for R&D spillovers, I include the log of R&D concentration, which is calculated as gross R&D expenditure divided by acreage. To capture other effects of the region, I include the unemployment rate, the proportion of secondary production, and the logged GDP per capita. I also include distance to the ERDF 1 boundary in order to further control for the location. The results are shown in column 3. The effect of being a subsidy grantee is now 1.645, which is, again, larger than the estimate without additional regional controls. These findings suggest that conditional on the firm-level control variables, controlling for population density is sufficient to remove confounding variation associated with regional attributes correlated with the ERDF instrument, and indicate that endogenous allocation of eligible areas and unobserved regional effects are unlikely to confound the results.

Because the ERDF program has been running since Finland accessed the EU in 1995, the firms knew about its existence at the beginning of the observed ERDF program period, 2000 to 2006. Thus, entering establishments may have chosen their location partly on account of the fact that the probability of receiving the R&D support was higher in the regions with higher levels of ERDF aid. In order to find out whether endogenous entry affects the results, I estimated the effect of the program among the firms that started up prior to 1995. Results in column 4 show that among these "old firms," the effect of the program is larger than in the full sample (column 5 of table 3) and significant at the 5% risk level. This suggests that endogenous entry is unlikely to be a major source of bias.

Column 5 displays results for a specification using pre-ERDF firm characteristics from the year 1994. Because R&D investment and exporter status are not available for that year, the specification does not include them as control

TABLE 5.—ROBUSTNESS ANALYSIS ON R&D INVESTMENT, REDUCED FORM, AND IV ESTIMATES

	Excluding ERDF 0 (1)	Excluding Areas in Pohjois-Savo Entering ERDF 1 in 2000 (2)	Controlling for Additional Regional Characteristics (3)	Excluding Firms That Started Up in 1995 or Later (4)	1994 Control Variables (5)	Second-Order Polynomials (6)	Third-Order Polynomials (7)	Pretreatment Trend Outcome: $y_{t-1} - y_{t-2}$ (8)	Untreated (9)	Treatment Effect Net of Pretreatment Trend Outcome: $y_{t+1} - y_{t-2}$ (10)
RF: ERDF objective 1	0.276***	0.206*	0.252**	0.276**	0.384**	0.207**	0.214**	−0.034	0.087	0.120
	(0.089)	(0.115)	(0.104)	(0.121)	(0.148)	(0.091)	(0.090)	(0.076)	(0.116)	(0.116)
1st: ERDF objective 1	0.094*	0.105**	0.153**	0.167**	0.134**	0.147***	0.140***	0.096*		0.096*
	(0.048)	(0.051)	(0.064)	(0.063)	(0.065)	(0.048)	(0.048)	(0.051)		(0.051)
IV: R&D subsidy grantee	2.935**	1.971*	1.645**	1.653**	2.862*	1.410*	1.531*	−0.352		1.248
	(1.411)	(1.165)	(0.801)	(0.831)	(1.596)	(0.714)	(0.776)	(0.726)		(1.103)
Observations	767	1,670	1,680	1,171	1,130	1,703	1,703	1,326	1,444	1,326
Industry-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pretreatment controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lagged outcome	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Pretreatment year	$t-1$	$t-1$	$t-1$	$t-1$	1994	$t-1$	$t-1$	$t-2$	$t-1$	$t-2$

The outcome is a log difference from $t-1$ to $t+1$, unless otherwise indicated in the column title. Standard errors clustered by NUTS4 area are in parentheses. Industry-year fixed effects are based on two-digit NACE classification. All specifications control for log of population density in 1992 at the NUTS 4 level. Pretreatment control variables include a quadratic term for age, indicator for exporter, log of employment, log of fixed assets, log of sales, and a quadratic term for the log of R&D investment, except in column 5, where exporter status and R&D investment, which are not available for 1994, are excluded. The specification in column 3 controls for the log of R&D concentration, log of GDP per capita, unemployment rate, proportion of secondary production, and distance to the ERDF 1 boundary at the NUTS4 level. The specification in column 6 (7) controls for second- (third-) order polynomials for continuous control variables. The 90%, 95%, and 99% confidence levels for a null hypothesis that the coefficient equals 0 are denoted by *, **, and ***, respectively. RF: reduced form; 1st: IV first stage; IV: IV second stage.

variables.[27] Although lagging controls further back increases standard errors substantially, the reduced-form coefficient on the ERDF 1 instrument and the IV estimate of the treatment effect are larger than and have the same significance as the estimates in column 4 of table 3, which controls for pretreatment firm characteristics in the year $t-1$, excluding R&D investment.

As discussed in section III, the RDS, the source of data on R&D expenditure, is not a random sample of R&D-inactive and new firms. It is worth noting that the analysis includes only R&D-active firms, which are likely always to be included in the sample, although the possibility that the nonrandom sampling of new R&D-active firms that applied for support could have affected the results cannot be completely ruled out. The main concern is that this may lead in overestimation of the conditional probability of receiving support in the highest-aid region where the proportion of supported firms is larger. As explained in section III, the initial version of the RDS was based on a census conducted in 1995, and hence it is representative for firms that started up prior to 1994. In order to assess whether the nonrandom sampling constitutes a problem for the analysis, I compare the estimated conditional probabilities of receiving support in the highest-aid region (ERDF 1) between the full sample and the sample of "old firms." An indication of bias arising from the nonrandom sampling would be that the first-stage coefficient for the ERDF 1 dummy variable is larger in the full sample, which was not completely randomly sampled. The first-stage estimate for the ERDF 1 indicator in the sample of old firms in column 4 of table 5 is 0.167, which is larger than the corresponding estimate in column 5 of table 3 (0.149). This finding indicates that the nonrandom sampling of applying firms is unlikely to drive the results.

In order to examine the impact of the functional form on the results, specifications in column 6 (7) add second- (third-) order polynomial terms for all continuous regressors. Although the precision of the estimation reduces slightly, the inclusion of the additional polynomial terms increases both the reduced-form coefficient on the ERDF 1 instrument and the IV estimate of the treatment effect, suggesting that the results are not driven by the choice of the functional form.

The key assumption of the IV strategy is that unobserved trends are similar in the ERDF 1 area and the area eligible for lower levels of aid in the absence of treatment. For example, any other government intervention apart from the R&D support scheme that is asymmetric across the ERDF 1 border and affects conditional company R&D investment—that is, that is correlated with the error term in the R&D equation—may confound the IV estimates. In order to examine the credibility of the common trends assumption and whether firms in the treatment group changed their pretreatment levels of

[27] It is worth noting that the estimates in column 4 of table 3 and this model are not directly comparable, as the former also controls for the exporter status. However, including exporter status in the year $t-1$ in the model does not affect the estimates in column 5 of table 5; hence, only results for the model excluding it are reported.

TABLE 6.—ROBUSTNESS ANALYSIS ON EMPLOYMENT, FIXED ASSETS, SALES, AND LABOR PRODUCTIVITY, REDUCED FORM, AND IV ESTIMATES

	Excluding ERDF 0 (1)	Excluding Areas in Pohjois-Savo Entering ERDF 1 in 2000 (2)	Controlling for Additional Regional Characteristics (3)	Excluding Firms That Started Up in 1995 or Later (4)	1994 Control Variables (5)	Second-Order Polynomials (6)	Third-Order Polynomials (7)	Pretreatment Trend Outcome: $y_{t-1} - y_{t-2}$ (8)	Untreated (9)	Treatment Effect Net of Pretreatment Trend Outcome: $y_{t+1} - y_{t-2}$ (10)
A. Employment										
RF: ERDF objective 1	0.167***	0.083	0.116*	0.159**	0.146**	0.133***	0.131***	0.025	0.074	0.148**
	(0.054)	(0.053)	(0.062)	(0.066)	(0.069)	(0.046)	(0.046)	(0.029)	(0.047)	(0.059)
1st: ERDF objective 1	0.083*	0.098*	0.148**	0.160**	0.130**	0.147***	0.140***	0.088*		0.088*
	(0.048)	(0.051)	(0.063)	(0.063)	(0.064)	(0.048)	(0.048)	(0.052)		(0.052)
IV: R&D subsidy grantee	2.003*	0.844	0.783**	0.995**	1.123*	0.906**	0.937**	0.284		1.675
	(1.050)	(0.590)	(0.374)	(0.445)	(0.617)	(0.374)	(0.388)	(0.314)		(1.022)
Observations	767	1,670	1,680	1,171	1,129	1,703	1,703	1,325	1,444	1,326
B. Fixed Assets										
RF: ERDF objective 1	−0.028	−0.057	−0.062	−0.032	0.022	0.003	0.002	−0.073	0.001	−0.094
	(0.081)	(0.074)	(0.104)	(0.080)	(0.081)	(0.077)	(0.077)	(0.053)	(0.089)	(0.116)
1st: ERDF objective 1	0.082*	0.096*	0.144**	0.158**	0.130*	0.144***	0.138***	0.089*		0.084
	(0.048)	(0.051)	(0.063)	(0.063)	(0.065)	(0.048)	(0.049)	(0.051)		(0.052)
IV: R&D subsidy grantee	−0.337	−0.593	−0.428	−0.206	0.166	0.018	0.012	−0.819		−1.126
	(0.917)	(0.824)	(0.791)	(0.490)	(0.546)	(0.533)	(0.554)	(0.704)		(1.565)
Observations	758	1,644	1,655	1,154	1,105	1,677	1,677	1,301	1,420	1,307
C. Sales										
RF: ERDF objective 1	0.159**	0.117*	0.124*	0.144**	0.148**	0.113**	0.110**	0.044	0.044	0.133*
	(0.068)	(0.059)	(0.064)	(0.058)	(0.063)	(0.051)	(0.054)	(0.045)	(0.050)	(0.073)
1st: ERDF objective 1	0.083*	0.095*	0.143**	0.156**	0.130*	0.144***	0.137***	0.086		0.086
	(0.047)	(0.051)	(0.063)	(0.063)	(0.065)	(0.048)	(0.048)	(0.052)		(0.052)
IV R&D subsidy grantee	1.919*	1.224	0.873*	0.920**	1.141*	0.786*	0.801*	0.511		1.537
	(1.114)	(0.810)	(0.494)	(0.421)	(0.583)	(0.437)	(0.472)	(0.584)		(1.141)
Observations	758	1,649	1,660	1,157	1,116	1,682	1,682	1,306	1,424	1,308
D. Labor Productivity										
RF: ERDF objective 1	0.011	0.043	0.013	0.008	0.002	0.012	−0.001	0.012	−0.023	−0.010
	(0.085)	(0.066)	(0.071)	(0.059)	(0.065)	(0.059)	(0.058)	(0.042)	(0.059)	(0.081)
1st: ERDF objective 1	0.081	0.096*	0.144**	0.153**	0.133**	0.144***	0.134***	0.082		0.082
	(0.049)	(0.051)	(0.062)	(0.064)	(0.065)	(0.048)	(0.049)	(0.052)		(0.052)
IV: R&D subsidy grantee	0.131	0.454	0.089	0.050	0.012	0.086	−0.009	0.149		−0.125
	(0.928)	(0.696)	(0.466)	(0.354)	(0.446)	(0.413)	(0.433)	(0.498)		(0.897)
Observations	758	1,649	1,660	1,157	1,116	1,682	1,682	1,305	1,424	1,308
Industry-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pretreatment controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lagged outcome	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Pretreatment year	$t-1$	$t-1$	$t-1$	$t-1$	1994	$t-1$	$t-1$	$t-2$	$t-1$	$t-2$

The outcome is a log difference from $t-1$ to $t+1$, unless otherwise indicated in the column title. Standard errors clustered by NUTS4 area are in parentheses. Industry-year fixed effects are based on two-digit NACE classification. All specifications control for log of population density in 1992 at the NUTS4 level. In panels A to C, pretreatment control variables are a quadratic term for age, indicator for exporter, log of R&D investment, log of employment, log of fixed assets, log of sales, and a quadratic term for the log of the lagged outcome, except in column 5, where exporter status and R&D investment, which are not available for 1994, are excluded. In panel D, terms for employment and sales are replaced with terms for labor productivity (i.e., log(sales/employment)). The specification in column 3 controls for the log of R&D concentration, log of GDP per capita, unemployment rate, proportion of secondary production, and distance to the ERDF 1 boundary at the NUTS4 level. The specification in column 6 (7) includes second- (third-) order polynomials for continuous control variables. The 90%, 95%, and 99% confidence levels for a null hypothesis that the coefficient equals 0 are denoted by *, **, and ***, respectively. RF: reduced form; 1st: IV first stage; IV: IV second stage.

TABLE 7.—ESTIMATED LONG-TERM EFFECTS OF R&D SUPPORT PROGRAM PARTICIPATION ON R&D INVESTMENT, EMPLOYMENT, FIXED ASSETS, SALES, AND LABOR PRODUCTIVITY, OLS, REDUCED FORM, AND IV ESTIMATES

	R&D Investment		Employment		Fixed Assets		Sales		Labor Productivity	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
OLS: R&D subsidy grantee	0.144	0.230***	0.014	0.050	0.151*	0.172**	0.097**	0.071	0.079	0.024
	(0.091)	(0.079)	(0.047)	(0.045)	(0.079)	(0.079)	(0.046)	(0.051)	(0.051)	(0.058)
RF: ERDF objective 1	0.374*	0.359**	0.187*	0.195*	0.064	0.042	0.280**	0.290**	0.099	0.108
	(0.190)	(0.167)	(0.110)	(0.114)	(0.162)	(0.166)	(0.119)	(0.110)	(0.069)	(0.073)
1st: ERDF objective 1	0.190**	0.212**	0.198**	0.195**	0.241**	0.241**	0.192**	0.195**	0.191**	0.194**
	(0.086)	(0.083)	(0.089)	(0.086)	(0.105)	(0.105)	(0.093)	(0.086)	(0.092)	(0.087)
IV: R&D subsidy grantee	1.974*	1.690**	0.944	1.002*	0.264	0.174	1.456**	1.483**	0.519**	0.555
	(1.086)	(0.803)	(0.643)	(0.586)	(0.612)	(0.611)	(0.673)	(0.665)	(0.262)	(0.390)
Observations	742	742	742	742	737	737	737	737	737	737
Industry-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pretreatment controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lagged outcome	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

The outcome is a log difference from $t-1$ to $t+3$. Standard errors clustered by NUTS4 area are in parentheses. Industry-year fixed effects are based on two-digit NACE classification. All specifications control for log of population density in 1992 at the NUTS4 level. Pretreatment control variables are a quadratic term for age, indicator for exporter, log of R&D investment, log of employment, log of fixed assets, and log of sales in the year $t-1$ excluding the term for the lagged outcome (for labor productivity, both sales and employment are excluded). Specifications in columns 2, 4, 6, 8, and 10 include a quadratic term for the log of the lagged outcome in the year $t-1$. The 90%, 95%, and 99% confidence levels for a null hypothesis that the coefficient equals 0 are denoted by *, **, and ***, respectively. OLS: ordinary least squares; RF: reduced form; 1st: IV first stage; IV: IV second stage.

R&D investment just before entering the program, I present results for three additional specifications. Column 8 displays estimates for a specification where the outcome is the pretreatment change in log R&D investment from the year $t-2$ to the year $t-1$. The IV estimate of this specification provides a test for whether pretreatment trends—trends before the firms receiving an R&D subsidy entered the R&D support program—were different within the complier population in the ERDF 1 area and in the area eligible for lower levels of aid. The IV and reduced-form coefficients are -0.352 and -0.034, respectively, and insignificant, implying that the null hypothesis of common trends is not rejected. Column 9 shows estimates for a reduced-form specification for the sample of untreated firms. This specification compares the trend of untreated noncompliers in the ERDF 1 area to the trend of untreated noncompliers and compliers in the area eligible for lower levels of aid. Although these groups are not directly comparable due to the fact that there are no compliers in the untreated group in the ERDF 1 area, the results are interesting in their own right. The reduced-form estimate is 0.087 and insignificant. Provided that trends in the complier group and the noncomplier group are of a similar sign and magnitude, this estimate complements the results in column 8 supporting the common trends assumption. Column 10 shows results for a specification where the outcome is the change in log R&D investment from year $t-2$ to year $t+1$. This specification quantifies the effect of program participation net of potential pretreatment trends. The IV coefficient on the program participation status is 1.248, and while the precision of the estimation is relatively low, the point estimate is larger than the no-crowding-out threshold of 0.693. Although the reliability of the estimates in columns 8 and 10 is limited due to the reduced strength of the instrument and smaller sample size, the findings are not at odds with the common trends assumption and suggest that pretreatment dips in R&D investment are unlikely to be a major source of bias.

In order to provide further support for the validity of the empirical setup, table 6 presents results for employment, fixed assets, sales, and labor productivity. Although the precision of the estimation is low and, as a result, some point estimates vary substantially across samples, the results for the specifications excluding ERDF 0 (column 1), excluding NUTS4 areas becoming eligible in Pohjois-Savo in 2000 (column 2), and including additional regional control variables (column 3) are not at odds with those in column 5 of table 4. Furthermore, the estimates are in general very similar for the sample including only firms that started up prior to 1994 (column 4), for the specification using pre-ERDF firm characteristics from the year 1994 (column 5), and for the specifications including second- (column 6) and third- (column 7) order polynomial terms of continuous control variables. Column 8 presents specifications for pretreatment log changes. I do not detect any statistically significant effects on pretreatment trends for the reduced-form or the second-stage IV equations for any outcome, although it is worth noting that the precision of the estimation is relatively low. Also the reduced-form estimates for the sample of untreated firms (column 9) are all insignificant. Finally, reduced-form coefficients on the ERDF 1 instrument in column 10 indicate significant positive effects on employment and sales net of potential pretreatment trends while corresponding estimates for fixed assets and labor productivity are insignificant.

D. *Long-Term Effects*

This section examines the impacts of R&D support program participation three years after the supported firms entered the program. Table 7 presents results for specifications corresponding to those in columns 4 and 5 of tables 3 and 4, but with a four-year log change from $t-1$ to $t+3$ as the outcome. This is of particular interest for outcomes on which the impacts are not expected to be instantaneous. As in the previous sections, the estimated treatment effect quantifies the impact of entering the program in year t and receiving support in the year $t + 1$.[28]

The IV estimates of the long-term treatment effect on R&D investment and employment are somewhat larger, while estimates for fixed assets are of a similar magnitude compared to the corresponding estimates in tables 3 and 4. However, the effect on sales is considerably larger than the corresponding instantaneous effect in table 4. The fact that sales continue to grow while changes in employment are moderate after the first year in the program suggests that labor productivity increases in the long run. Indeed, the treatment effect on labor productivity for the specification in column 9, which does not control for lagged outcome, is 0.519 and highly significant. This estimate is robust to controlling for lagged outcome (column 10), although the precision of the estimation is lower in this specification.[29] These estimates indicate that the program induced labor productivity gains in the long run. Furthermore, because the estimated long-term treatment effects on fixed assets are close to 0, it seems unlikely that these productivity gains are driven by adjustments in the capital stock.

E. *Discussion*

As explained in section IV, the effect of the R&D support program is expected to be heterogeneous among the participants, and the IV analysis identifies a local average treatment effect: the effect of R&D support program participation among the firms that enter it as a result of higher aggregate R&D support funding in their region. If this is the case, I 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, the effect may be substantially smaller for a sizable proportion of projects that would also be supported in the absence of ERDF funding as more than 32% of the supported projects would have been undertaken even without the agency's support in the period 1999 to 2003, according to the results of a grantee survey (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, because dependence on public funding is one of the publicly stated selection criteria.

The results of the agency's survey imply that 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. Yet adjustments at the intensive margin alone are unlikely to generate the large, positive effects revealed in the empirical analysis. This suggests that firms entering the R&D support program as a result of higher funding in their region started new R&D projects as a result of the subsidy. These observations have two implications. First, they suggest that projects entering the R&D support 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.

VI. Conclusion

Government 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.[30] This study provides new evidence of the causal effects of government R&D assistance on company performance. The empirical strategy was based on an IV approach that exploited geographic differences in R&D support allocation in Finland, where several regions are eligible for the highest level of ERDF aid because the population density is less than eight persons per square kilometer. The study controlled for the possible endogeneity of support allocation arising from the population density rule determining these regional provisions. The empirical analysis indicated that the probability of receiving an R&D grant is substantially higher in areas that are eligible for the highest level of regional ERDF aid than in areas that are not. These differences provide exogenous variation in R&D support program participation, which was used to identify the causal effects of the public R&D support on company performance by exploiting firm-level panel data on program participation, R&D investment, inputs, and sales.

The IV analysis suggests statistically and quantitatively significant positive impacts of the R&D support program on R&D investment among firms entering it as a result of higher ERDF funding in their region. Furthermore, crowding out of private R&D investment is rejected for the first half of the sample period (2000–2002) when ERDF funding covered a larger fraction of public spending on R&D assistance and induced more geographic variation in R&D support program participation. The study also finds significant instantaneous impacts on employment and sales, whereas positive effects on productivity are detected after three years of the treatment group entering the program.

It should be emphasized that the IV approach used in this study identifies the effect of the R&D support program

[28] Half of the firms in the treatment group stopped receiving subsidies by the year $t + 3$. An alternative approach would be to compare firms that received support until the year $t + 3$ to those that did not receive support until that year, but this limits the sample size and reduces the precision of the estimation considerably.

[29] I also estimated a similar specification for R&D intensity measured as the ratio of R&D expenditure to sales. The IV estimate (standard error) of the long-run treatment effect was 0.510 (0.923) and 0.161 (0.804) for specifications corresponding to columns 9 and 10 of table 7, respectively.

[30] For example, the European Commission (2002) has an overall goal of 3% for R&D as a proportion of GDP, with two-thirds of this financed by the private sector, while data show that R&D in the EU was about 1.84% of GDP in 2006 (Eurostat, 2008).

only among firms that change their participation status as a result of the higher government R&D support funding in their region. Thus, the results should not be interpreted as evidence of the aggregate effectiveness of the program. Despite this deficiency, they are significant in two respects. First, they suggest 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: the preferred IV point estimate suggests that 1 subsidy euro induces company R&D worth 1.4 euros in the first full support year. Furthermore, the results suggest that the effects may intensify over time.

As explained in section VE, it seems likely that a notable proportion of projects that would have received an R&D grant regardless of ERDF funding were privately profitable. Combining this piece of evidence with the estimated large, positive effect of the R&D support, 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 R&D support 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 a 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 for which its funding is of the essence. It is important to emphasize that high levels of government R&D support funding are not a guarantee of large impacts. A key factor determining the effectiveness is whether subsidies are targeted to R&D activities that will not be pursued in the absence of the support.

The 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 and improve productivity in the long run. As every program has its own selection rules and managerial practices, results concerning the effectiveness of one are not directly generalizable to others. 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

Abadie, Alberto, "Semiparametric Instrumental Variable Estimation of Treatment Response Models," *Journal of Econometrics* 113 (2003), 231–263.

Aghion, Philippe, and Peter Howitt, "A Model of Growth through Creative Destruction," *Econometrica* 60 (1992), 323–351.

Almus, Matthias, and Dirk Czarnitzki, "The Effects of Public R&D Subsidies on Firms' Innovation Activities: The Case of Eastern Germany," *Journal of Business and Economic Statistics* 21 (2003), 226–236.

Arrow, Kenneth J., "Economic Welfare and Allocation of Resources for Invention" (pp. 609–626), in National Bureau of Economic Research, *The Rate and Direction of Inventive Activity: Economic and Social Factors* (Princeton, NJ: Princeton University Press, 1962).

Bloom, Nick, Rachel Griffith, and John Van Reenen, "Do R&D Tax Credits Work? Evidence from a Panel of Countries 1979–1997," *Journal of Public Economics* 85 (2002), 1–31.

Criscuolo, Chiara, Ralf Martin, Henry Overman, and John Van Reenen, "The Causal Effects of an Industrial Policy," NBER working paper 17842 (2012).

David, Paul A., Bronwyn H. Hall, and Andrew A. Toole, "Is Public R&D a Complement or Substitute for Private R&D? A Review of the Econometric Evidence," *Research Policy* 29 (2000), 497–529.

Davidson, Carl, and Paul Segerstrom, "R&D Subsidies and Economic Growth," *RAND Journal of Economics* 29 (1998), 548–577.

European Commission, "More Research for Europe: Towards 3% of GDP," communication from the Commission COM, 499 final (Brussels, September 2002).

Eurostat, "EU27 R&D Spending Stable at 1.84% of GDP in 2006," Eurostat news release 34 (2008).

Glaeser, Edward L., "Learning in Cities," *Journal of Urban Economics* 46 (1999), 254–277.

González, Xulia, Jordi Jaumandreu, and Consuelo Pazó, "Barriers to Innovation and Subsidy Effectiveness," *RAND Journal of Economics* 36 (2005), 930–950.

Görg, Holger, and Eric Strobl, "The Effect of R&D Subsidies on Private R&D," *Economica* 74 (2007), 215–234.

Grossman, Gene M., and Elhanan Helpman, "Quality Ladders in the Theory of Growth," *Review of Economic Studies* 58 (1991), 43–61.

Howe, J. D., and Donald G. McFetridge, "The Determinants of R&D Expenditures," *Canadian Journal of Economics* 9 (1976), 57–71.

Hyytinen, Ari, and Otto Toivanen, "Do Financial Constraints Hold Back Innovation and Growth? Evidence on the Role of Public Policy," *Research Policy* 34 (2005), 1385–1403.

Imbens, Guido W., and Joshua D. Angrist, "Identification and Estimation of Local Average Treatment Effects," *Econometrica* 62 (1994), 467–475.

Irwin, Douglas A., and Peter J. Klenow, "High-Tech R&D Subsidies: Estimating the Effects of Sematech," *Journal of International Economics* 40 (1996), 323–344.

Jaffe, Adam B., "Building Programme Evaluation into the Design of Public Research-Support Programmes," *Oxford Review of Economic Policy* 18 (2002), 22–34.

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

Lerner, Josh, "The Government as Venture Capitalist: The Long-Run Impact of the SBIR Program," *Journal of Business* 72 (1999), 285–318.

Lichtenberg, Frank R., "The Private R&D Investment Response to Federal Design and Technical Competitions," *American Economic Review* 78 (1988), 550–559.

Nelson, Richard R., "The Simple Economics of Basic Scientific Research," *Journal of Political Economy* 67 (1959), 297–306.

OECD, "Main Science and Technology Indicators," *OECD Science and Technology, and R&D Statistics* (2012). doi:10.1787/data-00182-en.

Romer, Paul M., "Endogenous Technological Change," *Journal of Political Economy* 98 (1990), S71–S102.

Takalo, Tuomas, Tanja Tanayama, and Otto Toivanen, "Estimating the Benefits of Targeted R&D Subsidies," this REVIEW 95 (2013), 255–272.

Tekes, "Tekesin vaikuttavuus: asiakaskyselyjen ja teknologiaohjelmien arviointien analyysi," *Teknologiakatsaus* 203 (Helsinki: 2007).

——— "Tilastoanalyysi Tekesin vaikuttavuudesta," *Tekesin katsaus* 229 (Helsinki: 2008).

Wallsten, Scott J., "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 (2000), 82–100.