The Impact of EU Grants for Research and Innovation on Private Firms' Performance*

Gábor Kátay† Pálma Mosberger‡ Francesco Tucci§

Abstract

The paper assesses the impact of European Union (EU) grants for research and innovation on profit-oriented firms' productivity. Using a unique dataset on both successful and unsuccessful applicants to the EU's 7th Framework Programme and balance-sheet data for firms from 46 countries, we show that the EU grants have had a positive impact on firms' post-treatment productivity.

Keywords: EU funds for research and innovation; firm productivity; fuzzy regression-discontinuity design
JEL classification: C31; G28; H57; O31

1 Introduction

It has been long recognised that without public support, the level of private R&D investments falls short of the socially optimal level (Nelson (1959)). As argued by Arrow (1962), innovation-related knowledge intrinsically entails non-divisibility (half the knowledge of the technology is not worth half the full one), "non-probabilisable" uncertainty concerning its economic outcomes and non-full appropriability even in the presence of patent protection, which hamper firms from

*The paper reflects ongoing work which may be subject to further revisions. Please do not circulate, cite or quote without authors' permission. The views expressed in this paper are those of the authors and not necessarily those of the institutions the authors are affiliated with. This version: Nov. 2018. The authors are grateful to Richard Deiss and Daniel Szmytkowski for giving us access to the Corda dataset and for all the useful information they provided on the the content of the dataset. We aslo thank Alan Cross and Sandra Almeida Dworak for the information on the FP7 selection procedure they have shared with us.

†Corresponding author. European Commission, Joint Research Centre (JRC), Via E. Fermi 2749, 21027 Ispra, Italy. gabor.katay@ec.europa.eu. Tel.: +39 0332 786475

‡European Commission, Joint Research Centre (JRC)

§European Commission, Joint Research Centre (JRC)

undertaking innovation projects. Financial frictions caused by uncertainty associated with R&D and information asymmetries between borrowers and lenders may also lead private firms to engage less in R&D activities (Griliches (1986); Hall (2002)). In these circumstances the social return on R&D spending is greater than the private return and, consequently, direct or indirect subsidies can diminish this market failure and improve social welfare (Spence (1984)).

For this reason, governments in numerous countries and supranational authorities have put policies in place to directly support private innovation activities in the form of grants and subsidies or indirectly through tax credits. These policies aim at reducing the costs of the innovation in order to stimulate R&D investments. One of the largest such funding programmes worldwide is provided by the European Commission (EC) under its Framework Programs (FP) for Research and Technological Development (RTD). The first FP was launched in 1984, and successive framework programmes covered 5-year periods till FP7, which was the first one running for seven years (2007-2013). The actual ongoing FP8, a.k.a. Horizon 2020 covers the period between 2014 and 2020. The budget allocated to FPs gradually increased from about 4 billion EUR for FP1 to over 50 billion EUR for FP7 and 77 billion EUR for FP8.

Despite its strategic and budgetary importance for the EU, the effectiveness of these framework programmes has never been rigorously assessed. To fill the gap, this paper estimates the impact of FP7 grants for research and innovation on profit-oriented firms' post-treatment productivity. Using a unique dataset covering both successful and unsuccessful applicants to the FP7 and balance-sheet data for firms from 46 countries, we rely on a quasi-experimental research design to assess the effectiveness of the programme.

A wide array of previous studies have analysed the impact of other R&D public support – both in the form of direct subsidies and tax credits – on multiple dimension of firms' performances. As documented by Zúñiga-Vicente et al. (2014) in a comprehensive survey of the topic[1], economists have tried to assess empirically the effectiveness of such programmes since the 60's, with very mixed results. Empirical evidence depends on the time span under scrutiny, on the type of R&D projects covered, on the subsidy amount and on previous grants received.

Early studies mainly relied on matching methods based on a set of observables for building control groups. The main drawback of such approach is that the unobserved characteristics of firms that have applied for financing R&D may be too far from the characteristics of a random sample of firms that may not even be willing to commit to R&D, which makes the treatment-control comparison misleading.

More compelling counterfactuals appeared with the availability of data on both successful and rejected applicants within the same support programme. Among the most recent papers in the

[1] See also Becker (2015) for a survey focusing also on tax credits.

field, several made use of this advantage to exploit regression discontinuity design (RDD) as an identification strategy. Bronzini and Iachini (2014), relying on sharp RDD, focus on a regional programme in northern Italy and find a positive effect of the funds on R&D investments (both in tangible and intangible assets) for small firms. The authors argue that their result points in the direction of tighter financial constraints faced by small firms, although they do not provide strong evidence for such a mechanism. In a follow-up paper, Bronzini and Piselli (2016) find a sizeable effect of the programme on the number of patents and in the probability of patenting innovations. Both impacts appear to be larger in magnitude for smaller firms. Howell (2017), relying on the same identification strategy, studies the impact of the Small Business Innovation Research (SBIR) grant programme in the US. Results show that grants positively impact the revenues, the citations-weighted number of patents and the probability of receiving Venture Capital financing, especially for first-time applicants. These latter businesses witness also improvements in their survival figures. The authors also show that with the exception of the effects on revenues, the positive impacts decline with the number of applications and other subsidies received.

Other papers rely on IV methods to assess the effectiveness of funding programmes. Einiö (2014) exploits geographical availability of EU regional aid in Finland as an instrument for the treatment effect. The authors find remarkable positive short-term effects for firms' R&D expenditures, employment and sales, while in the long-run improvements in firms' productivity play a role as well. Wildmann (2017) makes use of jumps in the probability of receiving Austrian Research Promotion Agency (FFG) grants by Austrian firms conditional on the technical score assigned by the evaluation committee. The author reports a remarkable improvement for the propensity of filing a patent application in 3-4 years following the program. More established firms perform better in this dimension and, according to further evidence provided, seem to exploit research grants to undertake ambitious R&D projects at the margin.

We contribute to this strand of the literature by examining the effectiveness of one of the largest R&D funding programme, the EC's 7the Framework Programme. We focus on private companies participating in the core program called *Cooperation*, representing two thirds of the budget. The programme aims at fostering transnational collaborative research consortia among industry and academia in specific thematic areas. The selection process of the granted firms – which involves scoring and ranking of the proposals by a group of external experts and the decision on cut-off score by the EC depending on the total amount of available budget – allows us to implement a quasi-experimental research design. More specifically, we make use of the external experts' scores allocated to each project proposal to compare average post-treatment labour productivity of "marginal beneficiaries" (granted firms with application score slightly above the threshold) and "marginal non-beneficiaries" (unsuccessful firms with score slightly below the threshold) in a fuzzy regression discontinuity framework (see e.g. Imbens and Lemieux (2008)). Our results provide

clear evidence that research grants have a positive and significant impact on the winning firms' post-treatment labour productivity.

2 The 7th EU Framework Programme for Research and Technological Development

Since the European Commission launched the First Framework Programme (FP) for Research and Technological Development (RTD), its strategic objectives have included strengthening science and technology in the interest of the European industry and promoting research activities in support of other EU policies. The program fosters the development of European scientific community and its equipment with appropriate skills and know-how, and supports high quality scientific and technical work conducted through transnational projects.[2] The first FP was launched in 1984, and successive framework programmes covered 5-year periods till FP7, which was the first one running for seven years (2007-2013). The actual ongoing FP8, a.k.a. Horizon 2020 covers the period between 2014 and 2020. The budget allocated to FPs gradually increased from about 4 billion EUR for FP1 to over 50 billion EUR for FP7 and 77 billion EUR for FP8.

The FP7 contains five main building block Programmes. In this paper, we focus on the core programme called *Cooperation*, representing two thirds of the budget. It aims at fostering transnational collaborative research consortia among industry and academia in specific thematic areas. The four other blocks are: *Ideas* program supporting "frontier research" on the basis of excellence in any area of science; *People* program supporting researchers' mobility and career development; *Capacities* program aiming at strengthening the research capacity activities such as research infrastructure, regions of knowledge, research for the benefit of SMEs; and finally *Nuclear research* program that funds activities such as nuclear research, technological development, radiation protection and nuclear safety.

The funding principle of FP7 projects is co-financing. The standard reimbursement rate for research and technological development projects is 50%. The reimbursement rate can reach 75% in case of non-profit public bodies, SMEs, research organizations or higher education institutions. For frontier research actions the reimbursement rate might be even 100%.

The funding in the form of grants is generally allocated through the publication of "calls for proposals". Application to specific calls is open to universities, research institutions, companies, governmental administrations and individual researchers from EU Member States, associate (Norway,

[2]Decision No 1982/2006/EC of the European Parliament and of the Council of 18 December 2006 concerning the Seventh Framework Programme of the European Community for research, technological development and demonstration activities (2007-2013)

Switzerland, Israel, Albania) and candidate countries (Iceland, Macedonia, Serbia, Turkey). Participants from international cooperation countries may also apply in consortium with participants from Member States or associate countries. Project proposals have to be submitted by a certain deadline, should comply with clearly defined themes and have the required partnership structure.

After the deadline, all proposals under a call are first checked by the European Commission (EC) for eligibility then undergo a pre-defined selection procedure. The selection process typically consists of either a single-stage or two-stage procedure. The single-stage process is the most widely used selection method: overall, 376 out of 446 calls fall into this category. In this setup, each detailed project proposal submitted to a specific call is first evaluated individually by at least three independent external experts, who assign a score between 0 and 5 to the project based on its quality according to a set of criteria (potential impact, financial and technical feasibility and other relevant aspects). In a second round, the experts discuss the proposals and assign together a final score between 0 and 15 to each of the projects. All evaluated proposals are then ranked according to the final scores. Finally, a threshold is set by the EC based on the total amount of available budget for the specific call. Projects above the threshold are main-listed and the applicants are invited to negotiate the grant agreement. In the course of the negotiation, the EC might ask for additional clarifications or modifications, e.g. regarding the budget structure of the project or the types of actions proposed. Under certain circumstances, the EC may also accept modifications advanced by the applicants, including changes in the composition of the consortium. Non-successful applicants are either rejected or placed on a reserve list. In case if, for any reason, main-listed projects drop out before the contract is signed, they can be replaced by a projects from the reserve list. Consortia can also be selected from the reserve list if the EC decides to increase the amount of fund allocated to the call.

As its name indicated, the key feature of the two-stage selection method is that the submission of proposals takes place in two stages. There are several ways in which the process is carried out. In the most common two-stage submission scheme, applicants are invited to submit a short, partially developed technical proposal with their best solution for fulfilling the goals outlined in the call. The proposals are evaluated, scored and ranked by external experts the same way as in single-stage calls. Applicants with the highest ranked first-stage technical proposal are then invited to submit their full proposal for the second stage. The full proposal must be consistent with the short outline proposal and may not differ substantially. Finally, the full proposals are again evaluated in the similar way as in the single-stage case.

In a simplified two-stage selection framework, applicants from the first-stage are selected to pass to the second stage by the EC without the involvement of external experts. Therefore, no expert score is given in the first-stage. The second stage is similar to the single-stage procedure. In the minority of the cases, the main selection takes place in the first-stage. This modality of the

procedure can be viewed as full expert evaluation (similar to the single-stage case) followed by a discussion with the pre-selected applicants with the purpose of reaching a final agreement.[3]

3 The data

The project involves using two different databases that are linked together. First, data on applicants and signed grants of research funding are drawn from Corda, a database managed by the EC's Directorate-General for Research and Innovation (DG-RTD). Second, firms' performance measures are taken from Orbis, the largest international database for firms maintained by the private company Bureau van Dijk (BvD) containing information retrieved from official business registers, annual reports, newswires, and webpages. After pre-cleaning and harmonising Corda, the two datasets are linked using a combination of direct merge on the basis of the company's VAT number and a similarity score matching based on company name and other company information, such as postal and email addresses and web page. The following sub-sections describe in details the different steps of the data preparation.

3.1 The Corda database

Information on project proposals and on the final contractual grant agreements is kept separately in two different datasets. The *proposals* part of Corda collects information on both successful and unsuccessful applications, such as the starting and ending date of the proposed project, the claimed cost of the project, the fund requested, the amount of grant proposed by the Commission, and various information on the companies such as the company name, address, contact person and others. Additionally, it contains information on the application process, such as the score received from the group of expert evaluators and the final decisions of both the experts and the EC. Once the evaluation of each proposal has been completed and all information is registered, the database is no longer updated.

Information on the contractual agreements between the EC and the successful applicants is registered in the *projects* part of Corda. The database contains continuously updated information on the successful projects after signing the contracts, such as, for example, the final amount of grant received by the successful applicants or the final schedule of the innovation project. Besides containing information on companies as in the proposal dataset, it also includes the VAT numbers.

[3]Even though calls for proposal followed by a single-stage or two-stage selection procedure is the main tool to allocate funds, some exceptions exist. For example, specific FP funds for complex or highly specialised tasks can be allocated through direct invitation.

This paper focuses on FP7 grants, however, we also control for the potential effects of past grants received from the 4th, 5th or 6th Framework Programmes (see Section 5 for the details). Overall, we make use of the *projects* data recorded for all funding programmes starting from FP4 till FP7 and the *proposals* dataset for FP7. In addition, we use the so-called *h20* dataset for harmonisation purposes. Also managed by the EC's DG-RTD, *h20* is a centralized, harmonised metadata on all applicants starting from the beginning of FP7. The dataset is continuously updated, it contains detailed applicant-level information including firms' VAT numbers, duplications and transfer of rights.

The different parts of Corda and the h20 dataset are linked together using a combination of direct matches based upon exact identifiers (firms, projects) and similarity score matching based on company name and other company information. As a first step, projects present at both FP7 proposal and FP7 project datasets are linked: firms within the same project and from the same country are linked if they have the same firm identifier, exactly the same legal name, website or postal address. Second, the combined FP7 database and the previous FP4-FP6 databases are linked to the h20 dataset. Finally, a harmonised Corda dataset is created with information from all previous databases.

As a general rule, first, exact matches are performed in the basis of VAT codes then on the company names. In cases where exact identifiers such as company VAT numbers are missing or are not consistent between the different datasets, firms are linked using similarity score matching technique based on company name. The algorithm – described in details in Raffo and Lhuillery (2009) – assigns to each company in the first dataset a set of potential matches from the second one. A similarity score ranging from 0 to 1 is assigned to each potential match, reflecting the distance between the firms' name entries in the two datasets.[4] We manually set the threshold on the similarity score for two firms to be considered a potential match to 0.5. The final match is then selected among the possible alternatives using additional information on firms, such as postal address (city, postal code and street), website and email address.

After each linkage step, we kept all variants of discrepant information on firms and used them in the successive steps. For example, after linking two records with the same VAT number, but different company names or postal addresses, the new linked record contains both of these name or address variants and are used to find exact matches in the following steps.

3.2 Sample selection

After generating project-level variables (e.g. dummy for the presence research institute, higher education, public institution or various country groups in the project) and the dummies indicating

[4]This step is performed using the Stata command *reclink2*.

whether the firm has received grant from previous framework programmes, we keep only information on profit-oriented applicants to FP7 grants. The reduced database contains 66 895 firms applying with 48 490 project proposals through 446 calls, representing in total 172 258 observations (see line (A) of Table 1). The final estimation sample is further reduced based on specific call, project or applicant related criteria. Table 1 presents the elimination steps where non-italic figures refer to how many calls, projects or applicants were excluded based on specific criteria, while italic figures represents how many calls, projects, firms or applications were indirectly eliminated from the sample based on a particular criterion that is not directly linked to a call, project and/or firm in question within each step.

(a) A total of 134 calls are eliminated due to call specific reasons such as the call was announced for individuals, it was based on invitation, the score variable is non-available or there is not at least one winner and one non-winner firm. With this step of elimination 3 900 firms were entirely excluded from the sample.

(b) A total of 8 184 projects are eliminated due to project specific reasons such as grants received for non-research reasons (support & coordination, networks of excellence), first stage of two-stage application projects, missing project score information or projects ineligible for application. With this step of elimination 5 710 firms were entirely excluded from the sample.

(c) A total of 42 926 applications are eliminated if applicants were in a winner projects, but did not sign the contract, if the project did not end till 2014 or it ended in 2015 but the applicant had no financial information for 2016. With this step 12 344 firms were entirely excluded from the sample.

(d) If a firm took part in at least one successful and one or several unsuccessful applications, these latter applications are dropped from the sample. Since the period after the unsuccessful applications may be influenced by the effect of the intervention related to the successful application, these observations cannot be used as counterfactuals.

Overall, after removing observations that cannot be used, the size of the sample considered is reduced to 58 151 observations corresponding to 19 432 projects and 38 326 firms.

3.3 Matching with Orbis

The data used to retrieve balance-sheet and financial information comes from Orbis, the richest available international dataset at individual firm level. Maintained by BvD, Orbis gathers information coming from a wide array of different sources, including official business registers, webpages, disclosure documents and newswires. This flow of figures is then organized into key financial and

balance sheet variables, covering almost all relevant aspects of firms' activity. In Orbis, firms are uniquely identified by a code assigned to each of them by BvD (so-called BvD Number), which allows to keep track of a firm dynamic over time.

After cleaning firm-level information in order to make them homogeneous and comparable with Corda (for instance the country abbreviation added at the beginning of the VAT code, the web domain removed from the webpage address and so forth), the matching procedure between Orbis and Corda is performed on a country by country basis using the same method as for matching the different parts of Corda: first, a direct merge between the datasets is performed first on the basis of the VAT codes and then on the company names; second, the similarity score matching technique is used to identify potential matches based on company names; third, a selection criteria based on different combinations of secondary variables (postal address, website and email address) is used to find the correct match among the several alternatives; finally false potential matches are manually excluded from the final sample.

The matching success rate is highly dependent on the quality of the Orbis and Corda data, most notably the availability of VAT codes. All countries considered, 81% of firms (30 984 out of 38 326 firms, see last 3 lines of Table 1) were successfully matched.

Table 1: Summary of sample selection criteria

	calls	projects	firms	obs.
(A) FP7 profit-oriented applicants	446	48 492	60 280	172 199
Exclusions:				
(a) calls (-)	134	8 178	*3 900*	18 677
(b) projects (-)	*49*	8 184	*5 710*	20 214
(c) applicants (-)	*24*	*9 086*	*12 344*	42 926
(d) drop-out from successful applications (-)	*5*	*3 612*	*0*	32 231
(B) Sample considered (A - a - b -c - d)	234	19 432	38 326	58 151
(e) firms not matched with Orbis (-)	*1*	*1 218*	7 342	8 298
(C) Final sample (B - e)	233	18 214	30 984	49 853

Notes:The table summarises the sample selection process of FP7 grants from the Corda database. Line (A) shows the number of calls, projects, firms and observations in the original sample drawn from the Corda database when only profit-oriented applicants are considered. Lines (a) to (d) list the number of observations dropped from the sample based on the criteria explained in the paper. Italic numbers refer to the number of calls, projects and/or firms indirectly eliminated from the sample based on a particular criterion that is not directly linked to the call, project and/or firm in question. For example, by eliminating 134 calls (line (a)), we drop all project (8 178) and observations (18 677) corresponding to these calls. However, these criteria also eliminate those 3 900 firms that only participated in one or several of these 134 calls. Line (B) shows the sample considered after the previous exclusions. Line (e) shows that 7 342 firms in the Corda database could not be matched with Orbis data. Finally, line (C) shows the matched sample.

4 Identification strategy

To assess the effectiveness of the FP7 funding scheme, we rely on Regression Discontinuity Design (RDD) technique. Following seminal contributions by Angrist and Lavy (1999) and Van Der Klaauw (2002), RDD satisfying a reasonable set of validity assumptions is proved to be equivalent to a randomized experiment. A correct specification of such a model requires the existence of an administrative rule or technical score upon which treatment is assigned above a threshold. On top of this, RDD requires no manipulation in the running variable carried out by agents under observation. This requirement seems to be verified in our framework, given the independence of technical experts and of the EU committee. Moreover, the multiple stages in which the procedure is articulated seem to grant the impartiality of the evaluation. *Sharp* RDD is exploited when treatment's assignment happens deterministically, so that:

$$D = 1 \quad \text{if} \quad [S \geq S^*] \tag{1}$$

where S is the score or *running variable* and S^* is the policy threshold. Nevertheless, in our case this condition is not verified. In fact, the EC committee having the final word upon funding might revise experts' evaluations. Thus, following the negotiation of the grant agreement, grants might be denied to main-listed firms above the technical cut-off, while being awarded to some from the reserve list below the threshold. Table 2 shows the magnitude of this confounding process in our study: 5.6 $\%$ of projects that scored below the technical threshold for funding are nonetheless awarded grants by the EC committee, while 1.9 $\%$, in spite of an higher score, are eventually rejected or dropped out for any other reason.

Table 2: Expert scores, granted and non-granted projects: number of observations (and frequencies)

	Non-granted	Granted	Total
Score < threshold	31 573	2 810	34 383
	(63.3%)	(5.6%)	(68.9%)
Score > threshold	935	14 535	15 470
	(1.9%)	(29.2%)	(31.1%)
Total	32 508	17 345	49 853
	(65.2%)	(34.8%)	(100%)

Notes: Own calculations based on the Corda database.

This confounding process in the policy mechanism brings about the necessity for a *fuzzy* RD (FRD) estimation techinque. In order to exploit this method, it is necessary that the probability of treatment exhibits a discontinuity at the cut-off point, that is:

$$\lim_{s \uparrow S^*} P(D = 1 | S = s) \quad \neq \quad \lim_{s \downarrow S^*} P(D = 1 | S = s) \tag{2}$$

In the current framework, assignment into treatment is a function of the score assigned by the technical commission and other observable and unobserved covariates, which are likely to guide the EC committee in the selection process.

As argued by Imbens and Lemieux (2008), in a FRD the unconfoundness assumption is fundamentally violated. Treated and control units, although closely located around the cut-off, differ in some substantial aspect related to assignment of the treatment. Despite violation of unconfoundness, it is still possible to estimate a local average treatment effect (LATE). Exploiting the treatment

discontinuity at the cut-off point, a consistent estimator can be derived:

$$\tau_{FRD} = \frac{\lim_{s\downarrow S^*} E(Y|S=s) - \lim_{s\uparrow S^*} E(Y|S=s)}{\lim_{s\downarrow S^*} E(D|S=s) - \lim_{s\uparrow S^*} E(D|S=s)} \tag{3}$$

Following Hahn et al. (2001), under the monotonicity assumption, this ratio represents the average treatment effect on *compliers*, i.e. those who would take up the treatment upon fulfilment of the requirements, otherwise not. In other words, a *complier* is a unit whose behaviour changes according to his position relative to the program's cut-off for treatment.

The LATE in a FRD is usually obtained through a two stages procedure. The approach is similar to an IV setting. In the first stage, the treatment dummy is regressed over an indicator for being above the cut-off, the score obtained and a set of covariates that are likely to affect assignment into treatment. Once the vector of propensity scores is obtained, the outcome variable is regressed over the propensity score, the running variable and the same set of covariates as in the first stage. Under the continuity assumption for the covariates at the cut-off, this two stages procedure is consistent (see Cameron and Trivedi (2005)). Our estimated model is therefore the following:

$$D_i = \gamma 1[S_i \geq S^*] + \sum_{j=1}^{3} \beta_j S_i^j + \delta \mathrm{X}_i' + \epsilon_i \tag{4}$$

$$Y_i = \alpha_0 \hat{D}_i + \sum_{j=1}^{3} \alpha_j S_i^j + \delta \mathrm{X}_i' + u_i \tag{5}$$

where the treatment equation (1st stage) is given by (4), and (5) is the outcome equation. $1[S_i \geq S^*]$ is an indicator function equal to 1 when the expression in squared brackets is verified, S is the technical score (our running variable), X_i' is a set of firm and project level covariates. Inclusion of covariates in RDD does not affect the validity of the identification strategy (Imbens and Lemieux (2008); Lee and Lemieux (2010)). However, it can remarkably improve estimation preciseness. This is particularly true for the pre-treatment realization of the outcome variable, which ususually highly correlated with the dependent variable. Hence, in the set of exogenous variable, we naturally include the pre-treatment value of Y. Moreover, we control for the legal nature of the institutions involved in the research consortium (with dummies for research institution, higher education centre or publicly-owned entity). We also include country group indicators to assess two different phenomena. On the one hand we analyse whether grants are more effective in countries where we expect tighter financial constraints, like New Member States or associate countries. Furthermore, we are interested in assessing whether the nationality of the institutions involved in a consortium is among the determinants of funding's award by the EC committee.

5 Estimation results

We fit the model given by (4) and (5) using local-polynomial regressions. Our baseline option is to adopt the MSE-optimal specification (Imbens and Kalyanaraman (2012)), allowing for different bandwidths on the two sides of the cut-off.

Table 3 presents the fuzzy RDD estimates. Column (1) shows the results of the first-stage regression, while column (2) displays the results of the second-stage regression. The main variable of interest is the difference between the firm's post-treatment 3-year average labour productivity and the average labour productivity level in the same sector and country during the same period of time. More precisely, it is constructed as follows: first, aggregate shocks are removed from the firm-level labour productivity, defined as log(turnover / employment), by regressing out the effects of country$\times$sector$\times$year dummies using OLS on the whole sample of firms in Orbis. Second, depending on data availability we took the one, two or three year average values during the three-year window following the end of the project. Similarly, we construct the pre-treatment labour productivity variable for years before the start of the project.

The first-stage regression is a linear probability model where the dependent variable is a dummy for winning a research grant, hence coefficients can be interpreted as percentage point changes in the probability of being selected. Controls include before treatment log-labour productivity, third order local-polynomial, a dummy indicating whether a university, research organisation or public body has been part of the consortium and a dummy indicating whether an associate, candidate, tiers or new member state has been part of the consortium. The coefficients of above threshold dummy is positive and significant, having a score higher than threshold explains well the probability of winning as the number of non-compliers is small (see Table 2).

Table 3: Fuzzy RDD estimate of the effect of FP7 research grants on profit-oriented firms' post-treatment labour productivity

	First-stage regression	Second-stage regression
	(1)	(2)
LATE		0.19***
		(0.05)
above thold (score>0)	1.15***	
	(0.03)	
pre-treatment lab. prod.	0.01***	0.67***
	(0.00)	(0.02)
presence of research inst.	0.04***	0.06***
	(0.01)	(0.02)
presence of higher edu.	0.04***	0.03
	(0.01)	(0.02)
presence of public inst.	-0.03***	-0.04**
	(0.01)	(0.02)
presence of associate country	0.04***	-0.03*
	(0.01)	(0.01)
presence of candidate country	-0.04***	0.00
	(0.01)	(0.02)
presence of tiers country	0.01	0.01
	(0.01)	(0.02)
presence of new member state	0.01**	0.01
	(0.00)	(0.01)
nobs.		18 372
eff. nobs. below thold		7 721
eff. nobs. above thold		6 689
order est.		3

*Notes: The Table shows the fuzzy RDD estimates. The data is drawn from Corda and Orbis. The dependent variable is the post-treatment 3-year average labour productivity. More precisely, it is constructed as follows: first, aggregate shocks are removed from the firm-level labour productivity, defined as log(turnover / employment), by regressing out the effects of country×sector×year dummies using OLS. Second, we took the one, two or three year average values during the three-year window following the end of the project. Column (1) shows the results of the first-stage regression, while column (2) displays the results of the second-stage regression. LATE is the local average treatment effect. For the detailed explanation of the regressors, see the paper. Nobs is the number of observations used in the regression. The table also reports the number of observations below and above the threshold in the relevant sample. The last line specifies the order of the local polynomial used to construct the point estimator. *** significant at 1%, ** significant at 5%, * significant at 10%.*

As explained in Section 2, independent experts evaluate the projects and assign scores, then projects above a given threshold are accepted. If a winner project do not sign the contract, the EC has the possibility to replace it with a project from the reserve list based on their own judgement. The dummy variables in the regression identify systematic choices of the EC. Based on the regression estimation having a research organisation, higher education or research institution, associate or new member country in the consortium slightly increases the probability of being selected from the reserve list, while having a public institution or candidate country slightly decreases it. Nevertheless, all coefficients are small in magnitude, thus the economic importance of the systematic choices is very low.

The second stage result of the RDD estimations for labour productivity are presented in the second column of Table 3. The dependent variable is the post-treatment 3-year average labour productivity. The regression has the same control variables as the first stage. The estimated coefficient of the pre-treatment labour productivity is significant which suggests the persistence of labour productivity in time. The coefficient of treatment variable, the LATE is 0.19 and significant, which provides a clear evidence that receiving co-funding raises post-treatment labour productivity.

Table 4 presents different robustness check results. Each row shows results from a separate fuzzy RDD regression. The first row (a) is the baseline estimate presented in Table 3. In the next two rows (b and c) the order of the local polynomial used to construct the point estimates is changed to 2 and 4 from the baseline 3. Line (d) reports the result of a specification in which the dependent variable is the log-difference between post-treatment and pre-treatment labour productivity. In this regression, the lagged dependent variable is excluded from the regressors. In the specification (e), we restrict the sample to the applications for which 3 years of firm-level data is available before and after the treatment period. The specification (f) excludes firms from the control group which received FP4, FP5 or FP6 grant, while the specification (g) exclude firms from the treated sample which received more than one FP7 grant. Lines (h) and (i) presents RDD estimates using alternative kernel specifications, uniform or epanechnikov instead of triangular used in the baseline specification. The specification (j) uses the coverage error optimal bandwidth selector instead of the mean squared error optimal bandwidth selector. While the specification on log-differences leads to less significant (only at 10 per cent) and lower point estimate, all other specifications result in coefficients close to the baseline.

Table 4: FP7 research grants and post-treatment labour productivity: alternative specifications

	LATE	nobs	eff. nobs. below thold	eff. nobs. above thold
(a) Baseline	0.19***	18 372	7 721	6 689
	(0.05)			
(b) Polynomial order: 2	0.37**	18 372	2 503	6 667
	(0.17)			
(c) Polynomial order: 4	0.21***	18 372	9 053	6 475
	(0.06)			
(d) In diffs.	0.13*	18 372	6 792	6 667
	(0.07)			
(e) 3 years available before/after treatment	0.23***	7 791	3 475	2 531
	(0.07)			
(f) No prev. grants	0.20***	16 917	5 117	6 642
	(0.08)			
(g) Only winner once	0.15***	13 455	9 458	2 629
	(0.06)			
(h) Kernel: uniform	0.17***	18 372	9 577	6 667
	(0.05)			
(i) Kernel: epanechnikov	0.17***	18 372	9 944	6 667
	(0.05)			
(j) Bandwidth select.: CER-optimal	0.21**	18 372	4 779	6 475
	(0.10)			
		Placebo regressions		
(k) Threshold = 0.1	-0.18	18 372	7 499	1 702
	(0.18)			
(l) Threshold = - 0.1	10.71	18 372	8 791	4 357
	(32.96)			

*Notes: Each row collects results from a separate fuzzy RDD regression. The baseline regression (a) corresponds to the estimates presented in Table 3. In (b) and (c), the order of the local polynomial used to construct the point estimates is changed to 2 and 4 from the baseline 3. Line (d) reports the results of a specification in which the dependent variable is the log-difference between post-treatment and pre-treatement labour productivity. In this regrerssion, the lagged dependent variable is excluded from the regressors. In the specification (e), we restrict the sample to the applications for which 3 years of firm-level data is available before and after the treatment period. The specification (f) excludes firms from the control group which received FP4, FP5 or FP6 grant, while the specification (g) exclude firms from the treated sample which received more than one FP7 grant. Lines (h) and (i) presents RDD estimates using alternative kernel specifications. The specification (j) uses the coverage error optimal bandwidth selector instead of the mean squared error optimal bandwidth selector. Finally, lines (k) and (l) present two placebo regressions: the threshold around which the discontinuity is tested is set to -0.1 and 0.1, respectively. Nobs is the number of observations. The table also reports the number of observations below and above the threshold in the relevant sample. *** significant at 1%, ** significant at 5%, * significant at 10%.*

In addition, we also present placebo tests where the threshold of normalised scores for winning applications was modified from true 0 to "fictive" 0.1 and -0.1. With this exercise we estimate whether there is a break-point in the data also at these fictive thresholds. As expected, the results of the placebo regressions are not significant, confirming that the estimated effect is indeed due to receiving research funds.

Overall, regression results are robust to alternative specifications and there are clear signs that the research grant has a positive and significant impact on the winning firms' post-treatment labour productivity.

6 Conclusion

It has been long recognised that without public support, the level of private R&D investments falls short of the socially optimal level. For this reason, governments in numerous countries and supranational authorities have put policies in place to directly support private innovation activities in the form of grants. One of the largest such funding programs worldwide is provided by the European Commission (EC) under its Framework Programs (FP) for Research and Technological Development (RTD). Despite its strategic and budgetary importance for the EU, the effectiveness of these framework programmes has never been rigorously assessed.

To fill the gap, this paper estimated the impact of FP7 grants for research and innovation on profit-oriented firms' post-treatment productivity. To assess the effectiveness of the FP7 funding scheme, we relied on Fuzzy Regression Discontinuity Design (FRD) technique. We used a unique dataset covering both successful and unsuccessful applicants to the FP7 and balance-sheet data for firms from 46 countries. Based on the first stage regression estimation having a research organisation, higher education or research institution institution, associate or new member country in the consortium slightly increased the probability of being selected from the reserve list, while having a public institution or candidate country slightly decreased. But all coefficients are small in magnitude, so the economic importance of the systematic choices was very low. In the second stage of the fuzzy RDD estimations the coefficient of treatment variable, the LATE is 0.19 and significant, which is a clear evidence for that receiving co-funding raised post-treatment labour productivity. Overall, regression results are robust to alternative specifications and there are clear signs that the research grant had a positive and significant impact on the winning firms' post-treatment labour productivity.

References

Angrist, J. D. and Lavy, V. (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *The Quarterly Journal of Economics*, 114(2):533–575.

Arrow, K. (1962). Economic Welfare and the Allocation of Resources for Invention. In *NBER Chapters*, pages 609–626. National Bureau of Economic Research, Inc.

Becker, B. (2015). Public R&d Policies and Private R&d Investment: A Survey of the Empirical Evidence. *Journal of Economic Surveys*, 29(5):917–942.

Bronzini, R. and Iachini, E. (2014). Are Incentives for R&D Effective? Evidence from a Regression Discontinuity Approach. *American Economic Journal: Economic Policy*, 6(4):100–134.

Bronzini, R. and Piselli, P. (2016). The impact of R&D subsidies on firm innovation. *Research Policy*, 45(2):442–457.

Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics*. Cambridge University Press.

Einiö, E. (2014). R&D Subsidies and Company Performance: Evidence from Geographic Variation in Government Funding Based on the ERDF Population-Density Rule. *The Review of Economics and Statistics*, 96(4):710–728.

Griliches, Z. (1986). Productivity, R&D, and the Basic Research at the Firm Level in the 1970's. *American Economic Review*, 76(1):141–154.

Hahn, J., Todd, P., and Van der Klaauw, W. (2001). Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica*, 69(1):201–209.

Hall, B. H. (2002). The Financing of Research and Development. *Oxford Review of Economic Policy*, 18(1):35–51.

Howell, S. T. (2017). Financing Innovation: Evidence from R&D Grants. *American Economic Review*, 107(4):1136–1164.

Imbens, G. and Kalyanaraman, K. (2012). Optimal Bandwidth Choice for the Regression Discontinuity Estimator. *The Review of Economic Studies*, 79(3):933–959.

Imbens, G. W. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635.

Lee, D. S. and Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, 48(2):281–355.

Nelson, R. R. (1959). The Simple Economics of Basic Scientific Research. *Journal of Political Economy*, 67(3):297–306.

Raffo, J. and Lhuillery, S. (2009). How to play the "Names Game": Patent retrieval comparing different heuristics. *Research Policy*, 38(10):1617–1627.

Spence, M. (1984). Cost Reduction, Competition, and Industry Performance. *Econometrica*, 52(1):101–121.

Van Der Klaauw, W. (2002). Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression–Discontinuity Approach*. *International Economic Review*, 43(4):1249–1287.

Wildmann, R. (2017). The effect of government research grants on firm innovation: theory and evidence from Austria.

Zúñiga-Vicente, J. Á., Alonso-Borrego, C., Forcadell, F. J., and Galán, J. I. (2014). Assessing the Effect of Public Subsidies on Firm R&d Investment: A Survey. *Journal of Economic Surveys*, 28(1):36–67.