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

Gábor Kátay *Pálma Mosberger* *Francesco Tucci,*

24 November 2020

Joint
Research
Centre

Objective

- Objective: assess the impact of of EU research grants on profit-oriented firms' performance
- One of the largest funding programmes for R&I worldwide
  - 5-year, then 7-year Framework Programmes (FP) since 1984
  - FP7 runs from 2007 to 2013, total budget of over € 50bn
- We focus on private firms participating in the "Cooperation" programme ($\sim$2/3 of the budget)
  - excluded: academic research ("Ideas"), researchers' mobility ("People"), research infrastructure ("Capacities"), nuclear research
- Performance measures: post-treatment sales, number of employees, labour productivity

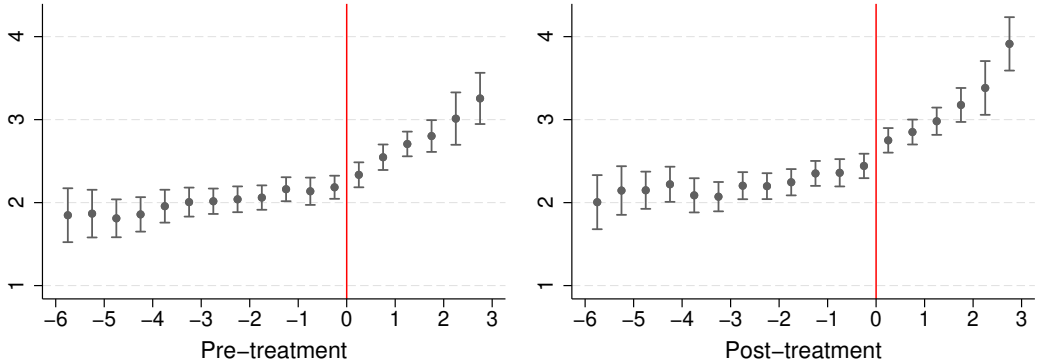
The selection procedure

- 1 or 2-round selection procedure / call
- At least 3 independent expert give scores between [0,15]
- Evaluation meeting of experts → common agreed score between [0,15] (sometimes transformed to [0,100])
- EC decides on a threshold:
  - above the threshold: granted
  - below the threshold: put on reserve list or rejected
- "Non-compliers" (about 6.5% of the sample):
  - Above the threshold, but no contract signed at the end
  - These projects are replaced by others from the reserve list
  - The funding programme can also be extended
- → Perfect setup for a fuzzy RDD

Identification strategy

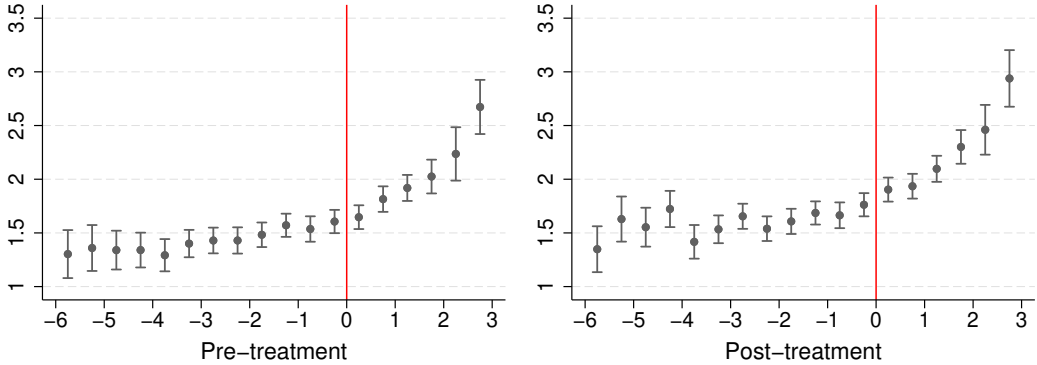
- Score is a proxy for the “quality” of the project proposal. Higher score $\rightarrow$ better and more viable project $\rightarrow$ larger expected impact
- Impact of the fund: discontinuity around the threshold
- $\rightarrow$ RDD: Comparison of firms between “marginal beneficiaries” and “marginal non-beneficiaries”
- $\rightarrow$ Fuzzy RDD because of “non-compliers”

Score & sales



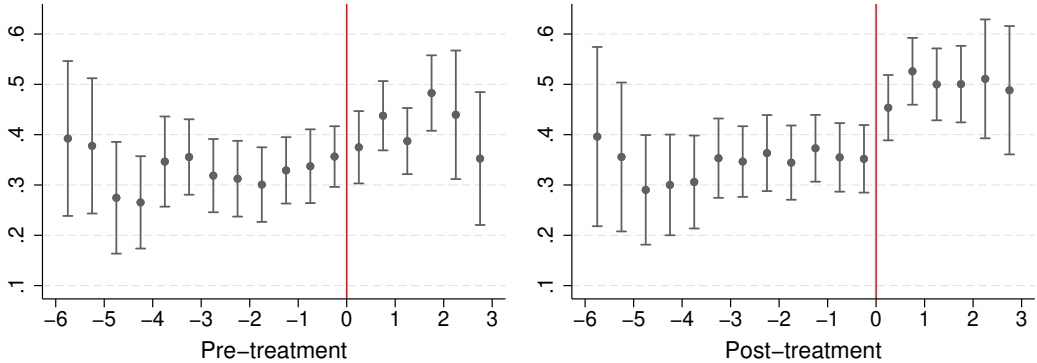
(a) Sales

Score & nb. of employees



(b) Number of employees

Score & productivity



(c) Productivity

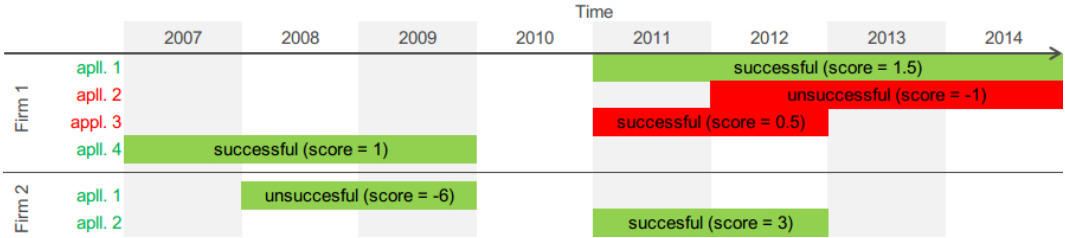
Estimation technique

- First stage regression: the probability of being awarded depends on:
  - the score is above the threshold
  - polynomial of the score (separately below and above the threshold)
  - other controls (potentially including the pre-treatment dep. var.)
- Second stage regression: the outcome variable is regressed on:
  - the predicted probability of being awarded from the first stage
  - polynomial of the score (separately below and above the threshold)
  - other controls (potentially including the pre-treatment dep. var.)

Estimation technique (cont'd)

- Lower weights for observations further away from the threshold (triangular)
- Bandwidth selection: the one that minimises the Mean Squared Error (MSE)
- Optimal order of the polynomial: the one that minimises MSE
- Local average treatment effect (LATE) = coeff. of "awarded" in the second stage
- Coeffs of X_i in the first stage: factors influencing the probability of being selected from the reserve list

Sample selection



- Observed treatment outcome at the firm-level $\rightarrow$ drop parallel projects (marked in red)
  - Firm 1, appl. 2: treated at the same time $\rightarrow$ cannot be used as counterfactual
  - Firm 1, appl. 3: still treated at the end of the project
- Outcome variable: log(sales / nb. of employees / productivity) *after* the end of the project *relative* to the (country $\times$ industry $\times$ time) mean
- Pre-treatment covariates: same measure *before* the call

First stage regressions

	Sales	Nb. of employees	Productivity
Above thold. (score ≥ 0)	0.754***	0.810***	0.788***
Pre-treatment dep. var.	0.002*	0.003	-0.005
Cooperation with:			
Research inst.	0.055***	0.074***	0.076***
Higher edu.	0.018***	0.013	0.019**
Public inst.	-0.022**	-0.035***	-0.032***
Associate country	0.022***	0.032***	0.033***
Candidate country	-0.036***	-0.021	-0.027*
Tiers country	-0.029**	-0.013	-0.019
New member state	0.027***	0.032***	0.028***

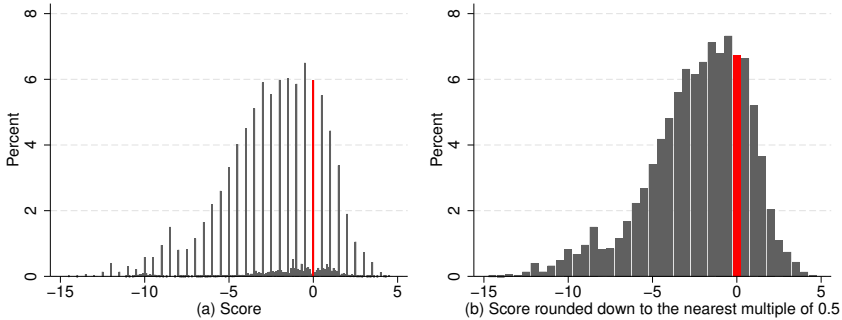
The impact of FP7 grants

	Sales	Nb. of employees	Productivity
Award	0.260***	-0.047	0.186***
Pre-treatment dep. var.	0.870***	0.896***	0.660***
Other controls	Yes	Yes	Yes

- The FP7 grants for R&I have had a significant impact on firms' post-treatment sales and productivity, but no impact on employment.

Manipulation of the running variable

- Individuals / firms may behave in response to rules that allocate resources
- In our case, manipulation is unlikely to be relevant
  - firms cannot manipulate their score around the cut-off
  - external experts don't know the threshold in advance



Discontinuity in the pre-treatment variables

- The treatment (grant) cannot influence variables determined prior to the award decision
- Test the discontinuity in the pre-treatment variables using the same RDD technique

	Sales	Nb. of employees	Productivity
Award	0.120	-0.191	0.023
	(0.125)	(0.209)	(0.054)

Discontinuity at other cut-off points

- We do not expect discontinuities at other points of the assignment variable
- Test the discontinuity at $c = -0.5$ and $c = 0.5$

	Sales	Nb. of employees	Productivity
Cut-off point: -0.5	0.216	2.308	-0.037
	(0.200)	(5.394)	(0.607)
Cut-off point: 0.5	-0.199	0.430	-0.453
	(0.180)	(0.311)	(0.930)

Robustness checks

	Sales	Nb. of employees	Productivity
(a) Polynomial order: 1	0.260***	-0.047	0.186***
(b) Polynomial order: 2	0.370***	-0.051	0.219***
(c) Polynomial order: 3	0.340*	-0.099	0.244***
(d) Epanechnikov kernel	0.241***	-0.051	0.182***
(e) Uniform kernel	0.206***	-0.076	0.193***
(f) CCT Bandwidth	0.305***	-0.024	0.158***
(g) W/o pre-treatment dep. var.	0.210*	0.006	0.164***
(h) W/o covariates	0.194*	-0.008	0.162***
(i) Score rounded	0.308***	-0.008	0.153***

The FP7 Funds

- Objectives:
  1. Cooperation: fostering transnational collaborative research consortia, i.e. funds for innovation (2/3 of the budget)
  2. Ideas: academic research
  3. People: researchers' mobility
  4. Capacities: strengthening the research capacity (e.g. research infrastructure)
  5. Nuclear research
- The standard reimbursement rate is 50%. non-profit public bodies, SMEs: 75%, frontier research: 100%)
- We concentrate on private firms participating in the "Cooperation" programme
  - Analysis covers 46 countries, about 220 calls, 12'000 projects and 24'000 firms

Theory

- Private R&D investments falls short of the socially optimal level (Nelson, 1959)
- Innovation-related knowledge entails (Arrow, 1962):
  - non-divisibility (half the knowledge of the technology is not worth half the full one)
  - “non-probabilisable” uncertainty
  - non-full appropriability (even if there is patent protection)
- Financial frictions caused by information asymmetries also lead private firms to engage less in R&D (Griliches, 1986)
- Critics: subsidy programs crowd out private investment or allocate funds inefficiently (e.g. Lerner, 2009)

Empirical studies

- Early studies mainly relied on matching methods → mixed results (see Zúñiga-Vicente et al., 2014)
- Studies using quasi-experimental methods:
  - + effect on investment spending, but only for small firms (Bronzini and Iachini, 2014, regional programme in northern Italy)
  - + effect on the number of patent applications, more marked for smaller firms (Bronzini and Piselli, 2016, regional programme in northern Italy)
  - for small high-tech firms: + effects on revenues, number of patents, survival probability, the probability of receiving Venture Capital financing even in the absence of enforcement (signalling effect. see Howell, 2017)

CORDA

1. FP4-FP7 project databases:
   - information on the contract, such as sum of grant awarded, total cost of the project, starting and ending date...
   - regularly updated until the end of the project
   - Firm identifier available: VAT number (+ name of the company, address, web page...)
2. FP7 proposal database:
   - information on the application, such as sum of grant asked, score obtained....
   - not updated later
   - no firm identifier, only name of the company, address...
3. H20 metadata:
   - harmonised database of winners starting somewhere during FP7
   - Firm identifier (VAT number) + name of the company + ...

ORBIS

- Database maintained by Bureau van Dijk
- Largest database for firms containing balance sheet information retrieved from official business registers, annual reports, newswires, webpages...
- Data over several decades until 2015 (or 2016)
- Availability of data varies by year, country, variable...

Matching

All these datasets are matched using:

1. exact identifiers (project ID, VAT number for successful applicants)...
2. ...and similarity score matching using company names:
   - takes into account name misspellings, name variations, ...
   - generates a measure of distance $\in [0, 1]$ with 1 = perfect match
   - information on email address, web page, telephone number and postal address are used to find the correct link among the possible alternatives
   - manual cross-check of dubious matches

- 71% of firms matched; 98% in terms of amount of grants awarded. $\rightarrow$ Non matched are mostly non-successful applicants (control group).

Score, threshold and non-compliants

- Score normalised between [0,1], 0 = threshold
- Threshold: lowest score for “awarded” above the first occurrence of a "reserve list” or “rejected”
- Non-compliers (nb. of obs.):

	Not awarded	Awarded	Total
Score < threshold	8,924	406	9,330
	(72.45)	(3.30)	(75.74)
Score > threshold	398	2,590	2,988
	(3.23)	(21.03)	(24.26)
Total	9,322	2,996	12,318
	(75.68)	(24.32)	(100.00)

Estimation technique

$$\begin{cases} \text{first stage: } D_i = \alpha 1[S_i \geq 0] + \sum_{j=1}^{p} \beta_j S_i^j + \sum_{j=1}^{p} \gamma_j S_i^j 1[S_i \geq 0] + \delta X_i' + \epsilon_i \\ \text{second stage: } Y_i = \theta \hat{D}_i + \sum_{j=1}^{3} \lambda_j S_i^j + + \sum_{j=1}^{p} \mu_j S_i^j 1[S_i \geq 0] + \rho X_i' + \upsilon_i \end{cases}$$

- $D_i =$ awarded dummy
- $S_i =$ score = forcing variable (0 = threshold)
- $1[S_i \geq 0] =$ score is above the threshold
- $X_i' =$ other controls (including pre-treatment dep. var.)

Estimation technique (cont'd)

- Lower weights for the observations further away from the threshold (triangular)
- bandwidth selection: MSE-optimal bandwidth selector
- Optimal p: the one that minimises MSE
- LATE = $\mathrm{E}\left[Y_i\left(1\right) - Y_i\left(0\right) | \mathrm{score} = 0\right]$ = coeff. of "awarded" ($\hat{D}_i$) in the second stage
- Coeffs of X_i in the first stage: factors influencing the probability of being selected from the reserve list