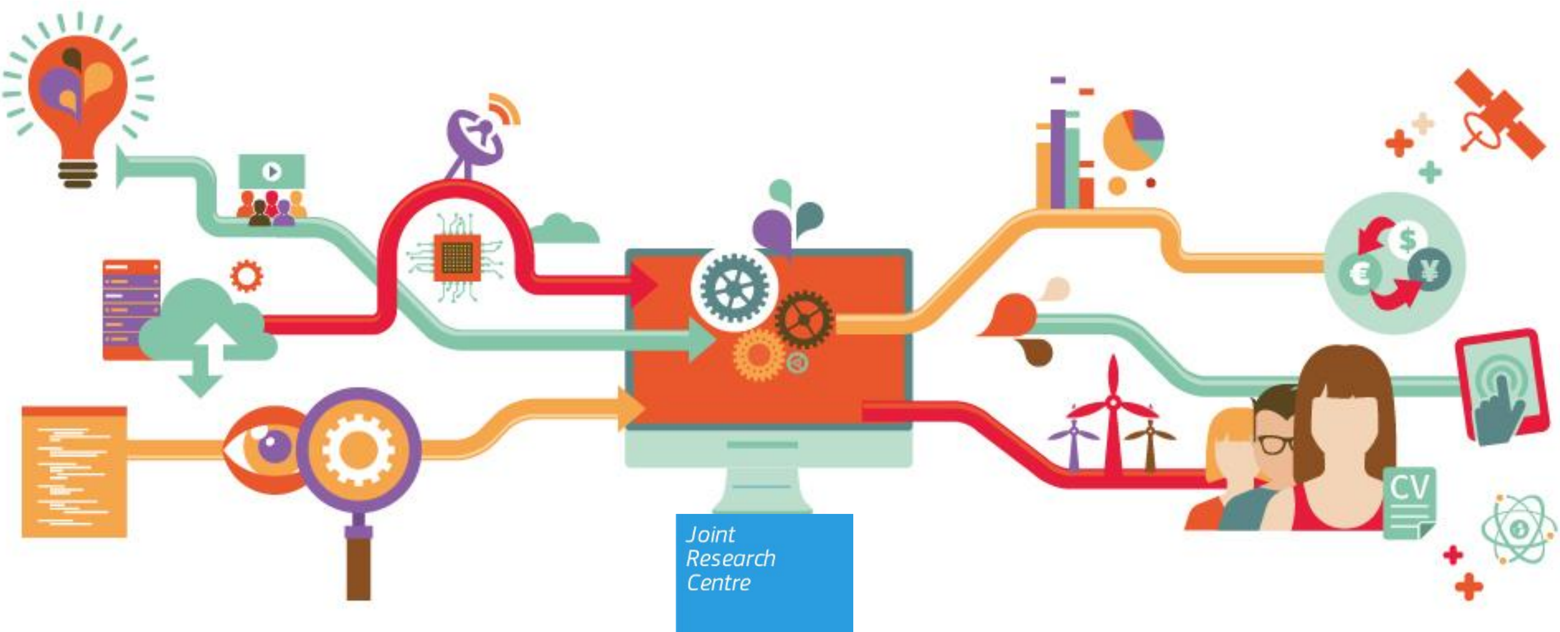
European Commission

JRC TECHNICAL REPORTS

Teaming up with Large R&D Investors: Good or Bad for Knowledge Production and Diffusion?

JRC Working Papers on Corporate R&D and Innovation No 05/2019

Amoroso, S. and Vannuccini, S.



Joint Research Centre

Contact information
Sara Amoroso
Address: Edificio Expo. c/ Inca Garcilaso, 3. E-41092 Seville (Spain)
E-mail: jrc-b3-secretariat@ec.europa.eu
Tel.: +34 954488388
Fax: +34 954488316

EU Science Hub
https://ec.europa.eu/jrc

JRC118081

ISSN 1831-9408 (online)

Seville, Spain: European Commission, 2019

How to cite this report: Amoroso, S. and Vannuccini, S., *Teaming up with Large R&D Investors: Good or Bad for Knowledge Production and Diffusion?*, JRC Working Papers on Corporate R&D and Innovation No 05/2019, Joint Research Centre

Teaming up with Large R&D Investors: Good or Bad for Knowledge Production and Diffusion?*

Sara Amoroso[1] and Simone Vannuccini[2]

[1]Joint Research Centre, European Commission
[2]Science Policy Research Unit (SPRU), University of Sussex Business School, University of Sussex

October 2, 2019

Abstract

The participation of top R&D players to publicly funded research collaborations is a common yet unexplored phenomenon. If, on the one hand, including top R&D firms creates opportunities for knowledge spillovers and increases the chance for a project to be funded, on the other hand, the uneven nature of such partnerships and the asymmetry in knowledge appropriation capabilities could hinder the overall performance of such collaborations.

In this paper, we study the role of top R&D investors in the performance of publicly funded R&D consortia (in terms of number of patents and publications). Using a unique dataset that matches information on R&D collaborative projects and proposals with data on international top R&D firms, we find that indeed teaming up with leading R&D firms increases the probability to obtain funds. However, the participation of such R&D leaders hinders the innovative performance of the funded projects, both in terms of patents and publications. In light of this evidence, the benefits of mobilizing top R&D players should be carefully leveraged in the evaluation and design of innovation policies aimed at R&D collaboration and technology diffusion.

Keywords: Collaboration; public funding; innovation performance; appropriability; top R&D investor

JEL Classification: L24; L25; O33

*We kindly acknowledge comments from Mara Giua (Roma Tre University), Liuchun Deng (Yale-NUS College), Mariagrazia Squicciarini (OECD), the Jena Economic Research Seminar, and one anonymous referee of the SPRU Working Paper Series.

1 Introduction

The existing literature on collaboration in innovation explored the role of different actors with respect to their type (e.g. private firms vs public research institutes), along with project characteristics, size of funding, geographical proximity, performance (Schwartz et al., 2012), and position in the value chain (e.g. suppliers, customers, competitors (Aschhoff and Schmidt, 2008). An less explored dimension of analysis is the uneven distribution of R&D capital among partners of a R&D alliance[1].

The participation of R&D 'star' firms to research consortia and the uneven distribution of R&D capital could generate asymmetric spillovers (Cohen and Klepper, 1992; Atallah, 2005; Busom and Fern/'andez-Ribas, 2008) and may result in a trade-off. On the one hand, there are many benefits in collaborating with top R&D firms, as these are usually large, experienced innovators, endowed with a consistent R&D capital that can lead to knowledge spillovers (Kamien and Zang, 2000; Branstetter and Sakakibara, 2002) and foster technology diffusion (Baptista, 2000; Keller, 2004). There are also reputation–related effects that might help projects that include these top R&D players to be selected and funded.

On the other hand, factors related to the structure of the collaboration (e.g. the distribution of appropriation capabilities and market power) might hinder the overall project's performance, as it has for instance being found by Crescenzi et al. (2018) for the case of a collaborative industrial research programme forerunning the more recent European Smart Specialisation Strategy. Top R&D firms may leverage their incumbent position and exploit winner-takes-most dynamics, as they are able to extract more value from the complementarities generated by the collaboration (Cabral and Pacheco-de Almeida, 2018), and can transform gains in know-how into commercialisation and economic performance faster compared to the other members of the consortium. This is particularly relevant for the European context, where a existence of a 'European paradox', with frontier research not adequately transformed in innovations, is still debated (Dosi et al., 2006; Albarrán et al., 2010; Jonkers and Sachwald, 2018).

In this paper, we investigate the impact of worldwide top R&D firms' participation in R&D programmes on R&D partnerships performance. Specifically, we use a unique dataset that matches data from the Seventh Framework Programme's (FP7) collaborative projects with information on top R&D investors from the EU Industrial R&D Investment Scoreboard (Hernàndez et al., 2018). We present empirical evidence on the role of top R&D companies' participation for both project selection and project innova-

[1]The top 10% of corporate R&D investors account for 60% of IP5 patent families and for 70% of world corporate R&D investments (Daiko et al., 2017; Hernàndez et al., 2018).

tive outcomes, namely projects' patents and publications.

To identify the effect of the inclusion of a top R&D players on consortia's research performance is an important concern for policy makers. Public support programmes to encourage private R&D effort and the development of research partnerships between private firms and public research organizations aim at maximising the production of scientific and technological output and knowledge spillovers. However, if a firms objective is to find complementary assets and skills, it will tend to form asymmetric partnerships, leading to reduced technological spillovers.

While there is an extensive body of empirical research on the determinants and the impact of research partnerships on innovation, there is limited to no evidence on the role of R&D star firms. Our study contributes to several strands of literature. First, we contribute to the literature on R&D collaboration and performance by considering the effects of top R&D firms on both the project selection probability and the returns from a public R&D partnership programme. Second, by taking into account the role of world leading R&D firms, we contribute to the growing literature and debate on 'super-star' firms and winner-takes-most dynamics (Schwellnus et al., 2018) that is generating a greater unequal distribution of market power and opportunities, widening the gap between frontier firms and followers.

Accounting for sample selection and endogeneity, we find that the hypothesized trade-off is at work: the participation of top R&D firms increase a project proposal's probability of being funded. Among funded proposals, the participation of top R&D firms is, however, negatively related to the number of patents and publications. For patents, a more thorough analysis shows that the negative impact of top R&D firms participation to the research consortium operates on the intensive margin, rather than on the extensive one: teaming up with a R&D star firm increase the probability to patent, but reduces the intensity of patenting when at least one patent is granted.

The paper proceeds as follows. In Section 2 we review the literature on the incentives for the establishment of R&D partnerships. In Section 3 we describe the data and provide summary statistics and descriptive accounts of how projects differ with and without the participation of top R&D investors. In Section 4 we conduct the empirical analysis and comment the results. In Section 5 we conclude, advancing our interpretation of the findings.

2 Existing literature on cooperation and knowledge diffusion

In what follows, first, we overview theoretical research on the reasons for establishing cooperative ventures among actors. We consider both reasons why actors might want to collaborate with top R&D performers and reasons why top R&D performers decide to engage in collaborations. Second, we delve into the research on innovation networks that shows the empirical side of the theoretical arguments put forward. Third, we consider how actors' heterogeneity, in particular when 'star' actors are considered, can influence collaboration patterns and outcomes. After that, we present evidence on collaborations and performance and discuss some theoretical justification for the possibility that collaboration with top R&D performers (the stars of innovative activities) could be detrimental to a project outcomes.

As our interest is to understand how consortia structure and the presence of peculiar actors in a given collaboration result into outcomes, we adopt a perspective mostly micro–economic in nature. We only incidentally touch contributions on clusters theory (Karlsson, 2010) and cluster policy, as research consortia under FP7 are closer to network rather than cluster architectures. Also, while we mention the consequences of spillovers, we do not focus on issues such as policy inducements and R&D additionality deriving, for example, from subsidies (David et al., 2000; Broekel and Boschma, 2011; Dimos and Pugh, 2016). Research on the factors affecting the receipt of R&D subsidies (Busom, 2000; Almus and Czarnitzki, 2003) and on public selection of R&D projects (Lee et al., 1996; Bozeman and Rogers, 2001) is less relevant for what concerns the issue of R&D collaboration and performance; however, it is important to be mentioned as it provides insights for the first prong of our conjectured trade–off, namely the determinants of projects selection.

2.1 Drivers of collaboration

The first issue of interest is 'what do we know' about the reasons for establishing cooperative ventures aimed at producing innovation output. In general, economic theory looks suspiciously at the possibility of voluntary knowledge disclosure, as knowledge is a 'latent' public good (Nelson, 1991).

The problematic nature of knowledge as a commodity would suggest that agents have the incentive to retain their know–how in order to enjoy monopoly rents or first–mover advantages. This idea is at the very core at the so–called Arrow–Nelson paradigm (Nelson, 1959; Arrow, 1962) that informed a good deal of Science and Technology policy styles. There are, however, cases in which the establishment of linkages between actors can be considered a rational decision: for example, in oligopolistic settings where R&D cooperation occurs in a first stage before competition, such cooperation is justified (d'Aspremont and Jacquemin, 1988) as it produces higher equilibrium levels of R&D expenditure and performance. In sum, cooperation between rivals and the voluntary disclosure of knowledge – also known as informal know–how trading (Von Hippel, 1989) – can, under certain conditions, improve payoffs.

R&D consortia have been considered since the Eighties as a tool to internalize R&D externalities (Branstetter and Sakakibara, 2002). Allen (1983) identified early cases of 'collective inventions' relying on and producing positive feedbacks (Cowan and Jonard, 2004). Nowadays, after a period of R&D investments conducted prevalently within large firms, industrial R&D started to drift again in the direction of increasing specialization and inter–firm connections (Mowery, 2009), with strategic alliances growing in importance (see Hagedoorn, 2002 for a review of major trends in inter–firm R&D partnerships). The literature on firms strategy re–opened the Pandora's box of firms collaborations, delving into the practices of open innovation (Berchicci, 2013; Bogers, 2011), make or buy decisions, and markets for technologies (Arora and Gambardella, 2010).

Empirically, the arguments supporting the establishment of firms' collaborations can be several: first, collaborations take–off from the necessity to combining complementary assets (including know–how) when endowments of these are not uniformly distributed in the market; second, collaborations may be started in order to build trusts between partners or to strengthen control; third, collaborations are an useful strategy to reduce costs, both economic (through simple sharing or by exploiting economies of scale and scope) and in terms of uncertainty reduction; fourth, through collaborations spillovers can be internalized. Collaborative or joint projects are only one among the possible ways for firms to exploit assets dispersed in the market, enhance control and reduce costs. In fact, when in need of know–how or skills, firms continuously face the alternative between internalizing spillovers through hierarchy and integration (Bresnahan and Levin, 2012) or to recur to pure arm–length market transactions. A third option is to opt for an informal yet structured form of collaboration structure – an innovation network – where networks represent a form of organization providing both stability (driven by reputation) and plasticity.

If we distinguish by firm type, the specific reasons to collaborate with large R&D performers fall in two main domains: on the one hand, an ease to access to markets, and hence to scale–up or to accelerate commercialization, especially for high–tech products (Nieto and Santamaría, 2010); on the other hand, a reputation effects that can reverberate both in lower barriers to access funding and in signalling of trust reinforced by the tie with prestigious partners. Yang et al. (2014) claim additionally that small firms can gain more from exploitation alliances with large firms, as alliances devoted to exploration activities entail higher risks of appropriation of value from the big actors.

For what concerns large R&D performers, more emphasis goes to the control end of the collaboration (Nieto and Santamaría, 2010), as big companies can use collaborations as tools to explore new technological trajectories and acquire ideas.

Empirically, Cantner and Meder (2007) explain cooperation in innovation with (i) measures of reciprocity (where indicators are built using sizes of firms' patent stocks), (ii) overlap in technology space and (iii) experience in cooperation. Experience (in general or repeated experience with the same partners) has been found to be a crucial factor for the success of collaborations, as according to Belderbos et al. (2015) the inter–temporally persistent collaborations are those providing a systematically positive effect on innovation performance. On the same line, Mora-Valentin et al. (2004) study contextual and organizational factors affecting collaboration between firms and scientific organization. Using Spanish data they find that —- to different degrees with respect to the type of actor — commitment, communication, reputation and previous linkages are drivers of successful cooperative agreements. McKelvey et al. (2015) focus on how research collaboration – in particular, collaboration with universities – favors the creation of innovative opportunities using a case study on university–industry research centers on engineering. All the reasons to establish collaborations just discussed can be summarized by resorting to the concept of proximity, in its various dimensions (Boschma, 2005). The notion of proximity, meant as 'closeness' with respect to a given dimension or space, captures the idea that the incentive to establish a linkage between two actors may be non–linear: actors too close may have nothing to exchange, while actors too far from each other may incur in high costs in order to come together. Empirically, Paier and Scherngell (2011) found that thematic and geographical proximity are main facilitators for the partner choice in FP5 projects. Recently, the notion of proximity has been embodied in a nascent research programme on the 'principle of relatedness', that aims at proposing a unifying framework to study collective learning, collaboration, diversification and co–location of economic activities (Hidalgo et al., 2018).

Finally, among the reasons to establish firms' collaborations we mentioned the possibility to internalize spillovers. Spillover effects, on the one hand, limit the appropriation of rents for the innovative actors, thus reducing their incentive to engage in innovative activities. On the other hand, they generate potential crowding–out of innovation efforts for the recipients of the spillover, if the know–how is a substitute rather than a complement to those efforts.

2.2 Innovation networks and clusters

Another strand of literature useful to understand collaborations in innovation is the one dealing with innovation networks. In fact, the choice to establish a collaboration network aimed at engaging in innovative activities has the advantage to retain flexibility while nurturing reliable ties based on trust and reputation. In a nutshell, innovation networks are a type of architecture capable of easing knowledge exchange and diffusion among actors. More precisely, they can be defined as the networks connecting either innovators (these are usually built using information on patents' co–applications) or inventors, with the latter focusing on inventors mobility across patenting firms (Cantner and Graf, 2006).

A research consortium as the ones we are interested in is nothing else than a network. Hence, understanding what drives network performances can be helpful to assess innovative performance of consortia including top R&D performers. Usually, in the study of network performance, one is interested in two sets of question; the first related to network structure, the second to network dynamics. For what concerns network structure, studies have found that different structures (e.g. small–world networks) are better performing in knowledge diffusion (Cowan and Jonard, 2004) and that relevant dimensions of analysis have to do with the centrality of given nodes and with the importance of strong versus weak ties and full network structures versus structural holes.

Turning to network dynamics, relevant dimensions have to do with the mechanisms of network growth (node addition) and preferential attachment (Barabási and Albert, 1999; Wagner and Leydesdorff, 2005), according to which new linkages lead to certain nodes in proportion to their existing amount of ties – a dynamics generating scale–free networks and characterized by increasing returns. Network structures interacts dynamically with the notion of proximity: actors closeness influences tie formation in given structures, that in turn generates knowledge flows (through voluntary disclosure, spillovers, or labor mobility), which feed–back on relative actors positions in proximity spaces. As a consequence, networks evolve as actors find useful to repeat collaborations, until they become too close with their partners, thus reducing the incentive to collaborate. The continue reconfiguration of innovation networks can lead to different paths, from a a more open architecture to specialization, centralization, inward–looking and lock–in outcomes (Cantner and Graf, 2006; Cantner and Vannuccini, 2017).

2.3 Heterogeneity and star actors

Certain actors are simply better positioned than others in signalling competences, absorbing knowledge and translating it into higher performance. Hence, the establishment of a collaboration with these actors should lead to differential (superior or inferior) performance. In this context, two dimensions of analysis are, on the one hand, the distribution of characteristics and performance (and the skenewss of this distribution) among actors and, on the other hand, the role played by star actors.

The heterogeneity (and the technology gaps) among actors has become a relevant explanatory variable in studies of productivity dynamics. In particular, Andrews et al. (2016) identify in the frictions to technological diffusion between global frontier and non–frontier companies one of the industry–specific determinants of the phenomenon of productivity growth slowdown (Fernald, 2015).

The peculiar role of large, leader or star actors has been dissected in two different research trajectories: first, studies assessing the so–called 'granularity hypothesis' (Gabaix, 2011); second, studies on the role played by stars scientists (Hohberger, 2016) and superstar firms. According to the granularity hypothesis, idiosyncratic shocks to a small number of (large) firms can explain a good share of variability in aggregate outcomes, thus reiterating the claim that few actors have a sizable influence on the whole economy. The same general mechanism holds with the notion of star scientists, according to which few scientists or inventors are associated with a disproportionately large share of new know–how production. For example, contributions from star scientists are found to be positively related to subsequent innovation performance of non–star scientists (while holding less true for star self-referencing). Dorn et al. (2017) document the rise of superstar firms and their contribution to the fall of labor share in the US economy.

The same mechanism of unevenly distributed capabilities to generate and channel know–how can be at work in our case, where top R&D firms act as star partners and influences innovative performance in collaborative research endeavors.

2.4 Collaborations and performance

Up to this point, we outlined four main literature trajectories useful to rationalize the reasons for the establishments of R&D collaborative ventures, in general and specifically when stars are involved.

Proximities, the willingness to internalize or capture spillovers, to create flexible ties in networks and to reap the benefits by collaborating with actors that, alone, have the star weight to influence economic dynamics, all are important explanatory variables that influence the formation of R&D consortia. Most of the literature focus, therefore, on such determinants. Less attention has instead be spent on the performance side of the collaboration story.

For what concerns collaboration performance, we can distinguish research focusing on economic or on innovative performance. Studies measuring the economic performance of R&D consortia use dependent variables like labor productivity or price–cost margins (for firm–level studies) or performance indicators at more aggregated level, for example at the regional one. Studies relating collaborations and innovation–related performance usually capture performance with innovation–output measures like patents or publications.

The effects of collaborations on economic performance is summarized in Benfratello and Sembenelli (2002) who note, on the one hand, that growing evidence suggest the presence of generalized benefits for actors participating in research joint ventures, while on the other hand that the bulk of this evidence is tied to American or Japanese collaboration projects. Scant results are available for European collaborations. Focusing on data on joint research projects under the third and fourth European Framework Programme (1992–1996) and the EUREKA Framework (1985–1996), they find that for joint research projects under Framework Programmes 3 and 4 no clear patters are emerging with respect to labor productivity or price–cost margin. Barajas et al. (2012) find that participation in the Framework Programme has both indirect (through the generation of technological capabilities) and direct effects on labor productivity in a sample of Spanish firms. More recently, Aguiar and Gagnepain (2017) exploit data on the Fifth Framework Programme and find that participation in joint research ventures produce strong and positive effect on labor productivity and small effects on profit margins.

For what concerns collaboration and innovative performance, in general one of the main predictors of outcomes is the size of the project. A positive effect of collaborations, leading to more patents or to patents' higher quality (measured, for instance, using quantity of claims contained in the patent application – see Beaudry and Schiffauerova (2011) for nanotechnology), is usually detected. For example, analyzing patenting performance of Japanese firms involved in government–sponsored research consortia, Branstetter and Sakakibara (2002) find empirical support for the theoretical argument that collaboration fosters innovation while mildly reducing the degree of product market competition among the members of the consortia. Schwartz et al. (2012) find for a large dataset of subsidized R&D cooperations in the german region of Saxony that large firms participation positively impact patenting but not publications, while the opposite holds for Universites' participation. Interestingly, evidence for the university–industry knowledge transfer channel suggests that — at least for publications and in specific knowledge fields like engineering – an inverted–U shaped relationship exists (Banal-Estañol et al., 2015). Non–linearities – that is, in this case, decreasing returns – in the effects of collaborations after a certain threshold or mass are reached, point out that, as suggested theoretically earlier–on when discussing proximity, a 'moderate' degree of collaboration could be the one entailing larger benefits.

Taking stock from the above discussion, and even considering the scattered empirical evidence existing, it can be reasonably expected collaborations to produce economic and innovation–related returns when compared with the same activities conducted in isolation. At this point, it is thus interesting to consider the potential downsides of collaborations. In general, given the existence of a multifaceted non–linear relationship between collaboration and performance, it can be claimed that the possibility of 'too much' collaboration is not too farfetched; too much collaboration may increase uncertainty rather than decrease it, produce coordination failures and consume relevant resources that are rather scarce for firms, for instance attention.

As already mentioned, the literature on stars as well as that on preferential attachment in networks usually highlights the advantages of being connected with a top–performing actor. Indeed, interactions with public large institutions, like Fraunhofer Institutes (Comin et al., 2019) positively affect partners' economic performance. Though, there may be negative effects emerging from this kind of collaboration, especially when considering large private actors (Cabral and Pacheco-de Almeida, 2018). While there is yet no evidence besides the present study on the performance of R&D consortia involving top R&D scoreboard companies, a first explanation can resort to standard arguments developed in the literature on firm size and innovation (Cohen, 2010).

Indeed, large firms have a cost–spreading advantage (Cohen and Klepper, 1996) in performing R&D, but empirical evidence has not identified much more than a linear relationship between size and innovation input measures, indicating the absence of increasing returns to size. Hence, involving a top R&D spender in a joint research project may not directly translate into an ease of innovation.

A second argument can be related to the weight such actors place on the collaborative relationship: if power is unequally distributed among partners, then reasons to appropriate or non–disclose knowledge from the side of the top R&D companies may prevail, thus reducing the measurable effect of the collaboration on outcomes.

Third, if top companies are involved in large and more radical projects, then their presence may produce lagged effects (Hall and Trajtenberg, 2004) not immediately captured by the data; in this case, however, the presence of a top R&D performer in a consortia does generates benefits, that are just delayed in time.

Fourth, the beneficial role of top actors may manifest in dimensions other than that of economic and innovation performance. In the 'anchor tenant hypothesis' (Agrawal and Cockburn, 2003), for example, large firms play the role of filter of technology and knowledge diffusion and, thus, are crucial for the growth of regions even though their effect leave relatively little trace on measured outcomes. In this sense, top–performing firms and firms with high network centralities (Beaudry and Schiffauerova, 2011) may undertake an 'orchestrating' or 'facilitating' role in the consortia they are part of (Mazzucato and Robinson, 2016).

Fifth, a general trend (that is, not specifically focused on collaborative ventures) has been found regarding the decreasing amount of scientific publication produced by large firms. This is the case for the pharmaceutical industry (Rafols et al., 2014) and for overall manufacturing. Arora et al. (2015) suggest that this trend may not be driven by a reduced importance of science for companies, but by a shift in their valuation of patents (that are indeed increasing) compared to publications — this coupled with increasing specialization and with resorting to external sourcing of knowledge generation. In this sense, we may expect such trend to hold also in the restricted context of collaborations that should experience a rise in patenting and a decrease in publications when top R&D performers are part of a consortium.

Finally, a type of negative outcome for a research consortia that includes a top R&D company might be the 'intertemporal cost' to be paid in a trade for higher chances of project acceptance at the selection stage.

3 Data

To empirically investigate relationship between innovation performance and the participation of top corporate R&D investors to R&D collaborative projects, we construct a unique dataset by matching information on publicly funded collaborative projects and proposals with data on the top corporate R&D investors from the EU Industrial R&D Investment Scoreboard (SB).

Data on R&D collaborative projects retained or funded by the 7th Framework Programme (FP7, European Commission) is extracted from CORDA (COmmon Research DAta Warehouse) and contains information on applicants (name, type of organization, location, VAT number, financial contribution to the R&D project, business sector if the organization is a company), and on projects (project duration, thematic, output of the project such as publications and patents).

Data on FP7 funded projects contain information on 28,454 distinct organizations participating to 24,502 FP7 projects. Each organization can participate to more than one project. The average organization participates in 4.06 projects.

Table 1 shows the number of distinct participants by type. The majority of participants are firms (65%), while higher or secondary education and research institutes constitute only the 8 and 13 percent of the total number of participants, respectively.

Table 1: Participants by organization type

Organization type	N. organizations	
HSE	2,394	8%
PRC	18,313	65%
RES	3,688	13%
PUB	1,942	7%
OTH	2,126	7%
Tot. parts	28,463	100%

Note: HSE: higher or secondary education
PRC: private for profit
RES: research organizations
PUB: public bodies
OTH: other

Despite the predominance of firms, higher or secondary education institutes are repeatedly involved in many more projects compared to firms. We exclude from the analysis projects with only one participant (approximately 60% of all retained and funded projects which are mainly European Research Council and Marie–Curie Actions) and funded projects in research fields that have less than 100 projects or that produced less than 100 publications or patents overall. The final number of projects is 8,380.

Table 2 reports the number of projects, total numbers of publications and patents by research area and by participation of top R&D companies produced during and after the FP7. The participation of SB companies is concentrated especially among projects in Nano–sciences, Transport and Aeronautics, Energy, Security and Space, where top R&D companies participated to at least 25% of the projects. Despite Marie–Curie funding scheme ranks first in total number of projects, it is not the most productive in terms of scientific publications and patents. The majority of publications and patents come from the research projects in the field of Health. The largest concentration of SB companies, however, is found among Nano–sciences, Energy, Transport and Aeronautics, Nuclear Fission and Space.

Table 3 shows the distribution of patents and publications at project level. We group the number of patents and publications in 5 categories (0, 1, 2–10, 11–100, >100) and

Table 2: Number of scientific publications and patents by research area and top R&D firms participation

Research Area	Projects (N)	with top R&D (%)	Publications (N)	with top R&D (%)	Patents (N)	with top R&D (%)
Health	4977	10.4	57535	15.2	41719	17.4
Research Infrastructures	6912	9.4	24821	4.1	18561	1.3
Nanosciences	3823	28.6	18999	52.1	15093	54.1
Food, Agriculture & Fish	3441	12.4	11922	12.1	6122	13.3
Marie-Curie	16557	8.2	12680	4.9	4843	10.8
Energy	2200	31.6	4748	44.7	3332	52.3
Security	2100	23.7	2254	13.7	1615	10.8
Transport	3819	28.5	1839	55.1	790	67.0
Research for SMEs	1166	13.0	1220	6.6	777	4.9
Environment	3171	9.4	8393	4.6	643	13.2
Nuclear Fission	400	36.3	1195	20.6	367	27.0
International Cooperation	734	4.2	558	0	281	0
Space	1243	12.0	2233	18.3	156	25.6
Science in Society	964	4.7	561	2.5	9	0.0
Socio-economic sciences	2942	3.8	2150	0.0	0	0.0
Total	54449	13.3	151108	17.4	94308	20.9

report the cross-tabulation of the percentages of projects per category of patenting or publishing intensity. More than 50% of the projects did not produce any publications or patents (during the period 2007–2019). Also, among the project that publish, the majority did not file any patent, especially for categories with only 1 or 2–10 publications. However, the majority (64%) of projects with more than 100 scientific papers or articles are also filing more than 100 patents.

Table 3: Share of projects (%), by patent and publication intensity

		Publications					
		0	1	2-10	11-100	>100	Total
Patents	0	51.24	5.01	15.86	13.54	1.01	86.67
Patents	1	1.87	0.64	0.00	0.00	0.00	2.52
Patents	2-10	1.66	0.04	2.96	0.02	0.00	4.68
Patents	11-100	0.05	0.00	0.04	4.11	0.02	4.21
Patents	>100	0.00	0.00	0.00	0.02	1.90	1.92
	Total	54.82	5.69	18.85	17.70	2.94	100.00

Out of the 2,500 top R&D companies, 383 applied to FP7 funding. Of these 383 top

R&D firms, 351 participated to funded projects, 32 to retained proposals that did not receive the funding. Table 10 in Appendix A, reports and compare basic characteristics of top R&D investors by their level of participation to FP7. The 351 SB companies (675 counting the number of distinct subsidiaries) participated to 2,360 different projects, namely, around 10% of the projects included a SB company.

Table 4 reports basic information on the projects with or without the participation of SB companies. Projects that include a top R&D company have on average less publications, but more patents. Also, collaborations with top R&D firms are larger in terms of team size in the average cost of the project, EC contribution, and in the average duration of the project.

Table 4: Project information by top R&D firm participation

		Mean	Median	SD	Min	Max
Without top R&D	N. publications	18	0	139	0	5733
	N. patents	11	0	134	0	5590
	N. participants	9.9	9	6.4	2	56
	Tot. project cost (€mln)	3.5	2.4	5.1	0.01	224.6
	EC contribution (€mln)	2.6	1.9	2.8	0.01	93.0
	Project duration (months)	38	36	12	4	84
		Mean	Median	SD	Min	Max
With top R&D	N. publications	17	0	106	0	2353
	N. patents	13	0	105	0	2353
	N. participants	14	12	8.7	2	71
	Tot. project cost (€mln)	6.7	4.5	7.6	0.1	85.2
	EC contribution (€mln)	4.5	3.4	4.5	0.1	41.8
	Project duration (months)	41	42	10	6	96

4 Empirical strategy

To assess the impact of top R&D companies' participation on the scientific output of publicly funded projects, we estimate a count model, where the main variables of interest are the the number of publications or patents, y; the participation of one or more top R&D companies to funded or retained projects, T (dummy variable); and the selection variable, S (dummy variable), which indicates whether or not a project has been funded.

In estimating the count model, we control for both sample selection and endogeneity, separately and simultaneously (Terza, 1998; Bratti and Miranda, 2011). The sample selection bias derives from the fact that the number of publications or patents, y, is missing when the project proposal is not selected, and the selection dummy S takes on value zero, while it is observed when the project receives funding. Therefore,

$$y = \begin{cases} \text{missing} & \text{if } S = 0 \\ 0, 1, 2, \ldots & \text{if } S = 1 \end{cases} \quad \text{and} \quad ln(\mu) = x'\beta + \delta T + \epsilon \tag{1}$$

$$S = 1(S^* > 0) \quad \text{where} \quad S^* = w'\theta + \phi T + \nu \tag{2}$$

where $\mu \equiv E[y|x, T, \epsilon]$ is either the mean number of patents or the mean number of publications. The vector $x = $ (*n.participants, % of firms, duration, size*) contains a set of available variables that may influence the average number of patents and publications, such as the size of the team (number of participants), the ratio between firms and public/private research organizations, the duration of the project in months, the size of the project (in EUR mln). In addition, we control for research sector and project starting year fixed effects. The vector w is a set of explanatory variables for the selection of a project, in our case we consider size and duration of the project, as well as the share of total (granted and non-granted) funding per scientific sector $w = $ (*size, duration, share funding per sector*).

The participation of top R&D investors T is likely to be endogenous, as it may be correlated with unobserved project characteristics that make them more likely to be selected and to attract the participation of such big investors. To correct for these two sources of endogeneity—sample selection and omitted variables—we use Heckman's sample correction method (Heckman, 1979) in the first case, and both a control function (2-stage residual inclusion, 2SRI; Terza, 2018) and instrumental variable approach for the second. To select instruments, we estimate a mixed-effects probit

$$T^*_{ij} = z'_{ij}\alpha + u_j + \xi_{ij} \tag{3}$$

where the probability for a project proposal to attract a top R&D investor depends on project i and scientific sector j characteristics $z = (\%\ of\ firms,\ n.participants,\ share\ funding\ per\ sector)$, i.e. share of firms, team size and sector funding intensity, a random intercept for each scientific sector (u_j), and a random intercept for each project nested in scientific sectors (ξ_{ij}). Tables 11 and 12 in Appendix B report the estimated random intercepts $\hat{u_j}$ and residuals $\hat{\xi_{ij}}$, and the correlation of the residuals with T, S and y. Our identification assumption is that the estimated intercepts and residuals are the unobserved project characteristics related to the participation of top R&D firms, and they are therefore our instrumental variables of choice.

Finally, to estimate the effect of top R&D investors' participation to publicly funded R&D projects, we use zero-inflated count models (negative binomial, ZINB and Poisson, ZIP), and a two-part model. The two-part model is based on a statistical decomposition of the density of the outcome into two processes. The first part of a two-part model is a binary outcome equation that models the probability of having at least a patent. The second part uses a linear regression to model $E(lny|y > 0)$. Differently from a Tobit model, the two processes that are generating zeros and positive values are assumed to be different and independent.

4.1 Selection bias

For each estimation model, we compare the basic specification to one that takes into account the selection bias. Results are reported in Tables 5 and 6. Generally, we find that the participation of leading R&D firms to publicly funded R&D consortia is negatively related to both the number of patents and publications. More specifically, in Table 5 the estimated coefficients for T show that the participation of top R&D firms decreases the expected count of patents by a minimum of 1-exp(-0.201)≈20% to a maximum of 1-exp(-0.787)≈50 %. The results differ (statistically) significantly between the ZIP and ZINB. The main difference between the two model specifications is the assumption of variance-to-mean ratio equal to one for the Poisson part of the ZIP. If the data exhibit clusters of occurrences, as in our case (see Table 3), it is likely to be overdispersed, and the ZINB is the preferred estimator. A comparison of log-likelihood, Akaike and Bayes information criteria (AIC and BIC) suggest that the ZINB model provides the best fitting and the most parsimonious specification.[2] To control for the selection bias, we include the inverse Mills ratio as an additional regressor. We find a stronger (larger coefficient) negative relationship between top R&D firms and patents. However, the approach of incorporating selectivity in a count model by including the inverse Mills ratio (Heckman's approach) is inappropriate (Greene, 2006), and therefore, our results are purely speculative.

[2] In the model selection, we compare the goodness-of fit of ZINB, ZIP, Poisson, Poisson hurdle and Negative Binomial hurdle. ZINB is the model that performs the best in terms of these three criteria. Results from model comparison are not reported in the paper, but available upon request.

Table 5: Patents: effects of top R&D companies' participation – Correcting for selection

	ZINB	ZINB†	ZIP	ZIP†	OLS	OLS†	Probit	Probit†
	y		y		$ln(y)\|y > 0$		$P(y > 0)$	
top R&D (T)	-0.577***	-0.787**	-0.201***	-0.231***	-0.257**	-0.723**	0.165***	0.107**
	(0.149)	(0.403)	(0.009)	(0.011)	(0.105)	(0.285)	(0.051)	(0.049)
n.participants	-0.021	-0.021	0.020***	0.010***	0.013	0.013	-0.011**	-0.010**
	(0.014)	(0.014)	(0.001)	(0.002)	(0.011)	(0.011)	(0.004)	(0.004)
% of firms	-0.179***	-0.179***	-0.903***	-0.987***	-0.246***	-0.247***	0.054***	0.049***
	(0.046)	(0.046)	(0.009)	(0.019)	(0.033)	(0.033)	(0.014)	(0.013)
duration	0.026***	0.005	0.024***	0.021***	0.024***	-0.023	0.009***	0.004*
	(0.010)	(0.039)	(0.001)	(0.001)	(0.007)	(0.028)	(0.002)	(0.002)
size	0.177***	0.217***	0.075***	0.076***	0.076***	0.173***	0.023***	0.030***
	(0.026)	(0.077)	(0.001)	(0.001)	(0.018)	(0.0058)	(0.007)	(0.007)
Inverse Mills Ratio		-2.537		-0.642***				
		(4.530)		(0.127)				
$\rho_{\epsilon\nu}$						-0.336***		-0.334***
N	8,248	(52,064)	8,248	(53,920)	1,098	(49,291)	8,248	(53,920)

Note: †Estimation takes into account the selection bias. Significance code: *** p< 0.01, ** p< 0.05, * p< 0.1. The number of observations in parentheses refers to the sample used to calculate the selection probability. The coefficients (of all specifications except Probit) can be interpreted as an approximation of semielasticities: for a unit change in x, the number of patents changes by $\beta * 100$ percent, or as incidence rate ratios by exponentiating the regression coefficient for count models.

Table 6: Scientific publications: effects of top R&D companies' participation – Correcting for selection

	ZINB	ZINB†	ZIP	ZIP†	OLS	OLS†	Probit	Probit†
	y		y		$ln(y)\|y > 0$		$P(y > 0)$	
top R&D (T)	-0.407***	-0.936***	-0.073***	-0.434***	-0.176***	-0.278***	0.017	0.001
	(0.069)	(0.170)	(0.008)	(0.011)	(0.058)	(0.062)	(0.043)	(0.042)
n.participants	0.008	0.009*	0.017***	0.016***	0.016***	0.167***	-0.010***	-0.010***
	(0.005)	(0.005)	(0.001)	(0.000)	(0.005)	(0.005)	(0.003)	(0.003)
% of firms	-0.339***	-0.339***	-0.548***	-0.561***	-0.233***	-0.225***	-0.222***	-0.219***
	(0.023)	(0.023)	(0.007)	(0.001)	(0.024)	(0.024)	(0.019)	(0.019)
duration	0.041***	-0.013	0.025***	0.017***	0.034***	0.021***	0.010***	0.009***
	(0.004)	(0.016)	(0.001)	(0.001)	(0.003)	(0.004)	(0.001)	(0.002)
size	0.128***	0.230***	0.089***	0.173***	0.083***	0.101***	-0.003	0.001
	(0.011)	(0.032)	(0.001)	(0.002)	(0.012)	(0.013)	(0.004)	(0.004)
Inverse Mills Ratio		-6.268***		-3.779***		-1.042***		
		(1.847)		(0.080)				
$\rho_{\epsilon\nu}$						-0.832***		-0.160***
N	8,248	(52,064)	8,248	(52,064)	3,774	(51,967)	8,248	(53,920)

Note: †Estimation takes into account the selection bias. Significance code: *** p< 0.01, ** p< 0.05, * p< 0.1. The number of observations in parentheses refers to the sample used to calculate the selection probability. The coefficients (of all specifications except Probit) can be interpreted as an approximation of semielasticities: for a unit change in x, the number of publications changes by $\beta * 100$ percent, or as incidence rate ratios by exponentiating the regression coefficient for count models.

In the four rightmost columns we report the estimated coefficient of a two-part model. Similar to the results from the count models, we find a negative relationship between the participation of top R&D firms and patents per project (OLS), especially when controlling for the selection bias.[3] On the other hand, top R&D firms' participation increases the probability of patenting, even when controlling for selection. This shows that there are significant differences between the extensive margin (the change in the probability to patent) and the intensive margin (the change in the patenting intensity, conditional on having at least one patent).

Thus, teaming up with large R&D investors make the project more likely to patent, but it also reduces the number of patents. Two mechanisms can be at play behind these results. First, projects including top R&D firms are expected to produce at least a patent, due to the ample experience of these firms. However top R&D companies may have less incentives to disclose their "patentable" ideas with the other consortia partners, and keep some technological development to themselves.

The second reason could be linked to the applied nature of research carried out by top R&D firms together with the advice given by the European Commission in its Guide to Intellectual Property Rights for FP7 projects (p.12) on the results that are capable of industrial or commercial application:

> It might prove advisable to keep the invention confidential and to postpone the filing of a patent (or other IPR) application (and consequently any dissemination), for instance, to allow further development of the invention while avoiding the negative consequences associated with premature filing (earlier priority and filing dates, early publication, possible rejection due to lack of support / industrial applicability, etc.).

Similar results are found for the number of publications. The participation of top R&D investors is associated with a lower number of publications per project, especially when considering the selection bias, while it does not have any effect on the change in the probability to publish. Hence, holding all the other project characteristics constant, projects with or without the participation of a top R&D firm have the same probability to publish scientific articles, however, conditional on publishing, having a top R&D firm on board reduces the number of publications.

While publishing is not the main focus of most firms' activities, it is undoubtedly beneficial for firms to interface with the academia, as they can get access to state-of-the

[3] The correlations $\rho_{\epsilon\nu}$ between the error terms of eqs. (1) and (2), ϵ and ν, are statistically significant, indicating that sample selection is present.

art knowledge, without waiting for it to be published. Also, companies' do not have the same incentives to disclose the results of their research by publishing it. This is confirmed by the negative coefficient of the share of firms (% of firms) on the probability to publish: the higher the share of firms in a project, the lower the probability of publishing, while the higher probability of patenting.

The additional negative effect of the participation of a top R&D firm may stem from additional red tape barriers to publish within these large corporations, or the additional control that these companies may have over the consortium's IP output.

4.2 Endogenous participation

Table 7 reports the results of the ZINB and two-part model for both patents and publications, correcting for the endogeneity of top R&D firms' participation to the FP7. In particular, we assume that there are some unobserved characteristics related to the participation of top R&D firms that are also likely to affect the scientific results of the projects.

To correct for such omitted variables bias, we use a 2SRI approach for the count data model and a 2SLS for the two-part model.[4] Despite the tests confirm the endogeneity of the treatment (statistical significance of the residual random effects $\hat{u}_j$ for the ZINB, significance of the Wald test for the linear model and of the Score test for the Probit model), there are no significant changes from previous estimations in terms of magnitude nor in sign of the coefficients. Even after correcting for the endogeneity of the participation of R&D firms, this is still negatively related to the number of patents and publications, and it affects differently the extensive and intensive margins of patenting and publishing.

4.3 Endogenous participation in the selection

In this subsection, we consider the case where the participation of top R&D is endogenous with respect to the selection of project proposals. In other words, better projects are more likely to both attract top R&D firms and be selected. Therefore, there may be unobserved characteristics (among funded and not funded projects) that are related to the participation of top R&D firms characteristics and to the selection of successful projects. Table 8 displays the results from the selection equation (2), where we compare

[4]In the count model, we only use the research sector specific random effects $\hat{u}_j$, because the high correlation between the project specific residuals $\hat{\xi_{ij}}$ and T (see Table 12 in Appendix B) introduces multicollinearity and, as a result, the estimated coefficients associated with these two variables are both extremely high and of opposite signs.

the estimations from a probit with no endogenous participation, and two probit models that account for endogeneity (2SLS versus 2SRI). The results of all three specifications indicate that the participation of top R&D firms increase the project proposal's probability of being accepted. We also control for project size and duration, and for the share of funding per sector (to impose some exclusion restrictions). As expected, the larger the requested budget (size), the lower the probability of being funded, while longer-term projects are more likely to be accepted. The endogeneity test confirm that there may be unobserved project characteristics attracting top R&D investors that are also related to the selection process of FP7 proposals.

To account for the endogeneity in the selection process, we use the predicted value of the 2SRI of Table 8, compute the inverse Mill's ratio, and use it as an additional regressor in equation (1).

Table 7: Effect of top R&D companies' participation on the number of patents and scientific publications - Correcting for endogenous treatment

	Patents			Publications		
	2SRI-ZINB	2SLS	Probit	2SRI-ZINB	2SLS	Probit
	y	$log(y)\|y > 0$	$P(y \geq 0)$	y	$log(y)\|y > 0$	$P(y \geq 0)$
top R&D (T)	-0.729***	-0.193*	0.123**	-0.427***	-0.103*	-0.014
	(0.154)	(0.106)	(0.052)	(0.073)	(0.058)	(0.043)
n.participants	0.013	0.013	-0.013***	0.015***	0.016***	-0.010***
	(0.016)	(0.011)	(0.004)	(0.006)	(0.005)	(0.003)
% of firms	-0.335***	-0.248***	0.057***	-0.263***	-0.237***	-0.227***
	(0.040)	(0.032)	(0.015)	(0.027)	(0.025)	(0.020)
duration	0.016	0.023***	0.009***	0.033***	0.034***	0.011***
	(0.010)	(0.007)	(0.002)	(0.004)	(0.003)	(0.002)
size	0.120***	0.074***	0.024***	0.152***	0.082***	-0.003
	(0.023)	(0.018)	(0.007)	(0.012)	(0.012)	(0.004)
$\hat{u_j}$	-2.705**			-1.129***		
	(1.346)			(0.321)		
Endogeneity test		0.006	0.000		0.000	0.059
LL (R^2)	-7,462.7	(0.464)	10370.6	-19,922.8	(0.342)	8930.4
N	8,248	1,098	8,248	8,248	3,774	8,248

Note: Significance code: *** p< 0.01, ** p< 0.05, * p< 0.1. Wooldridge's (1995) robust score test is reported for the OLS (2SLS) estimation; Wald chi-squared test is reported for the probit.

Table 8: Selection Equation $P(S = 1)$ - effects of top RD companies participation

	Probit	Probit IV	2SRI
top R&D (T)	0.122**	0.061*	0.127***
	(0.021)	(0.033)	(0.031)
size	-0.024***	-0.024***	-0.025***
	(0.004)	(0.004)	(0.004)
duration	0.012***	0.012***	0.012***
	(0.001)	(0.001)	(0.001)
share funding per sector	12.983***	13.056***	13.027***
	(0.190)	(0.202)	(0.189)
$\hat{u_j}$			0.174***
			(0.029)
Endogeneity test (p-value)		0.008	
LL	-19,481.15	-31,169.4	-19,451.8
N	56,573		

Note: Significance code: *** p< 0.01, ** p< 0.05, * p< 0.1. Wald chi-squared test is reported for the IV Probit

Table 9: Effect of top R&D companies' participation on the number of patents and scientific publications - Correcting for endogenous selection

	Patents			Publications		
	ZINB	OLS	Probit	ZINB	OLS	Probit
	y	$log(y)\|y > 0$	$P(y \geq 0)$	y	$log(y)\|y > 0$	$P(y \geq 0)$
top R&D (T)	-0.799**	-0.717***	-0.272***	-0.875***	-0.650***	-0.323***
	(0.347)	(0.245)	(0.091)	(0.147)	(0.128)	(0.086)
n.participants	-0.021	0.012	-0.015***	0.010*	0.017***	-0.011***
	(0.014)	(0.011)	(0.004)	(0.005)	(0.005)	(0.003)
% of firms	-0.180***	-0.247***	0.051***	-0.339***	-0.233***	-0.223***
	(0.040)	(0.033)	(0.015)	(0.023)	(0.024)	(0.019)
duration	-0.001	0.033	-0.046***	-0.018	-0.025*	-0.032***
	(0.040)	(0.029)	(0.009)	(0.017)	(0.014)	(0.010)
size	0.230***	0.194***	0.142***	0.242***	0.204***	0.085***
	(0.080)	(0.061)	(0.021)	(0.033)	(0.029)	(0.020)
Inverse Mills Ratio	-3.422	-7.028**	-6.264***	-7.047***	-7.053***	-4.848
	(4.083)	(3.471)	(1.073)	(1.959)	(1.691)	(1.078)
LL (R^2)	-7,462.46	(0.467)	-2,809.7	-19,916.5	(0.347)	-4,656.4
N	8,248	1,098	8,248	8,248	3,774	8,248

Note: Significance code: *** p< 0.01, ** p< 0.05, * p< 0.1. Inverse Mill's ratio calculated from the results of 2SRI estimations of Table 8

Results of a ZINB and two-part model for patents and publications are reported in Table 9 and show that the participation of R&D leaders is still associated with a decrease in the number of patents and publications. The main difference is in the probit part of the two-part model, where the effect of top R&D firms' participation on the probability to patenting or publishing is negative. Hence, when simultaneously correcting for selection bias and endogeneity, the results suggest that including a R&D leader helps in obtaining the funds but it is associated with the decrease of both intensive and extensive margins of knowledge production.

5 Discussion and conclusion

In this paper, we ask if establishing a collaborative linkage with a worldwide top R&D company affects the innovative performance of publicly funded research consortia, measured in terms of patents and publications. We test our research question by exploiting a unique matching of European research project (FP7) data with information from the European Commission's EU Industrial R&D Investment Scoreboard on top R&D firms.

In general, one can identify several theoretical reasons why economic actors might want to engage in R&D collaboration—access to complementary assets, trust–building, strengthening control, reduction of costs of knowledge generation and uncertainty, internalization of knowledge spillovers. These reasons may be particularly relevant for alliances where one of the partners is a R&D 'star' firm. However, collaborating with a R&D leading firm can result in a threat of expropriation of unprotected knowledge, reduced bargaining power of the other partners, and asymmetric knowledge spillovers. Therefore, the collaboration with a top R&D firm has nontrivial effects on the innovative performance of a R&D collaborative project.

In estimating the effect of top R&D firms' participation to publicly funded R&D projects, we account for both sample selection and endogeneity, separately and simultaneously. When considering the sample selection and endogeneity separately, we find evidence of a trade-off. On the one hand, the participation of top R&D firms increase a project proposal's probability of being funded. On the other hand, the participation of top R&D firms is negatively related to the number of patents and publications. Also, results from a two-part model show that teaming up with a R&D leader increases the probability to patent or has no effect on the probability to publish, but reduces the intensity of both patenting and publishing among projects with at least one patent or publication.

While the structural determinants of our results cannot be fully deducted from our empirical analysis, we offer three non-exclusive interpretations.

First, following the anchor tenant hypothesis discussed in Section 2, the participation of top R&D players has beneficial effects such as signalling and reputation effects. While these effects are well reflected in the greater estimated probability to obtain funds when these top firms are included, they may be not captured by our measures of performance.

Second, innovative performance is negatively influenced by the participation of top R&D firms in consortia because these firms have lower incentives to disclose the outcomes of the R&D projects and may have more leverage over the other project's participants in deciding what and how to disclose information.

Third, it may be too soon to see any innovative outcomes for some projects more related to radical scientific and technological exploration. However, this argument holds mostly for delayed economic outcomes, like productivity growth, rather than for innovative outcomes. In fact, publications and patenting activities are outcomes of collaborative project but represent measures of innovation input and by that they should be less subject to the bias in measurement due to lag.

The results of the analysis are important both for firms' strategic considerations and to derive informed policy recommendations with respect to the promotion of R&D consortia and to fine-tune technology diffusion processes. Firms engaging in R&D collaborations might want to carefully evaluate the trade-offs they will meet when teaming up with top R&D firms. Public institutions should balance the importance of attracting leading R&D investors to their funding schemes and the hindered knowledge production of projects that can affect the overall knowledge diffusion process.

To conclude, while 'standing on the shoulders of giants' or even teaming up with them might be the most effective strategy to speed up technology diffusion and widen the knowledge base, one must also consider the shortcomings due to corporate incentives and partnership asymmetries. Despite the novel empirical approach and data used, we only started to explore this research trajectory that has the potential to increase our understanding of the collaboration-performance nexus and to inform policy making.

References

Agrawal, A. and I. Cockburn (2003). The anchor tenant hypothesis: exploring the role of large, local, R&D-intensive firms in regional innovation systems. *International journal of industrial organization*, 21(9):1227–1253.

Aguiar, L. and P. Gagnepain (2017). European cooperative R&D and firm performance: evidence based on funding differences in key actions. *International Journal of Industrial Organization*, 53:1–31.

Albarrán, P., J. Crespo, I. Ortuno, and J. Ruiz-Castillo (2010). A comparison of the scientific performance of the us and the european union at the turn of the 21st century. *Scientometrics*, 85(1):329–344.

Allen, R. C. (1983). Collective invention. *Journal of Economic Behavior & Organization*, 4(1):1–24.

Almus, M. and D. Czarnitzki (2003). The effects of public r&d subsidies on firms' innovation activities: the case of eastern germany. *Journal of Business & Economic Statistics*, 21(2):226–236.

Andrews, D., C. Criscuolo, and P. N. Gal (2016). The Global Productivity Slowdown, Technology Divergence and Public Policy: A Firm Level Perspective. Technical report, OECD Publishing.

Arora, A., S. Belenzon, and A. Patacconi (2015). Killing the golden goose? The decline of science in corporate R&D. Technical report, National Bureau of Economic Research.

Arora, A. and A. Gambardella (2010). The market for technology. *Handbook of the Economics of Innovation*, 1:641–678.

Arrow, K. (1962). Economic welfare and the allocation of resources for invention. In *The rate and direction of inventive activity: Economic and social factors*, pages 609–626. Princeton University Press.

Aschhoff, B. and T. Schmidt (2008). Empirical evidence on the success of r&d cooperationhappy together? *Review of Industrial Organization*, 33(1):41–62.

Atallah, G. (2005). R&D cooperation with asymmetric spillovers. *Canadian Journal of Economics/Revue canadienne d'économique*, 38(3):919–936.

Banal-Estañol, A., M. Jofre-Bonet, and C. Lawson (2015). The double-edged sword of industry collaboration: Evidence from engineering academics in the UK. *Research Policy*, 44(6):1160–1175.

Baptista, R. (2000). Do innovations diffuse faster within geographical clusters? *International Journal of industrial organization*, 18(3):515–535.

Barabási, A.-L. and R. Albert (1999). Emergence of scaling in random networks. *science*, 286(5439):509–512.

Barajas, A., E. Huergo, and L. Moreno (2012). Measuring the economic impact of research joint ventures supported by the EU Framework Programme. *The Journal of Technology Transfer*, 37(6):917–942.

Beaudry, C. and A. Schiffauerova (2011). Impacts of collaboration and network indicators on patent quality: The case of Canadian nanotechnology innovation. *European Management Journal*, 29(5):362–376.

Belderbos, R., M. Carree, B. Lokshin, and J. F. Sastre (2015). Inter-temporal patterns of R&D collaboration and innovative performance. *The Journal of Technology Transfer*, 40(1):123–137.

Benfratello, L. and A. Sembenelli (2002). Research joint ventures and firm level performance. *Research Policy*, 31(4):493–507.

Berchicci, L. (2013). Towards an open R&D system: Internal R&D investment, external knowledge acquisition and innovative performance. *Research Policy*, 42(1):117–127.

Bogers, M. (2011). The open innovation paradox: knowledge sharing and protection in r&d collaborations. *European Journal of Innovation Management*, 14(1):93–117.

Boschma, R. (2005). Proximity and innovation: a critical assessment. *Regional studies*, 39(1):61–74.

Bozeman, B. and J. Rogers (2001). Strategic management of government-sponsored r&d portfolios. *Environment and Planning C: Government and Policy*, 19(3):413–442.

Branstetter, L. G. and M. Sakakibara (2002). When do research consortia work well and why? evidence from japanese panel data. *American Economic Review*, 92(1):143–159.

Bratti, M. and A. Miranda (2011). Endogenous treatment effects for count data models with endogenous participation or sample selection. *Health Economics*, 20(9):1090–1109.

Bresnahan, T. F. and J. D. Levin (2012). Vertical integration and market structure. Technical report, National Bureau of Economic Research.

Broekel, T. and R. Boschma (2011). Knowledge networks in the dutch aviation industry: the proximity paradox. *Journal of Economic Geography*, 12(2):409–433.

Busom, I. and A. Fern/'andez-Ribas (2008). The impact of firm participation in r&d programmes on r&d partnerships. *Research Policy*, 37(2):240 – 257. ISSN 0048-7333.

Busom, I. (2000). An empirical evaluation of the effects of r&d subsidies. *Economics of innovation and new technology*, 9(2):111–148.

Cabral, L. and G. Pacheco-de Almeida (2018). Alliance formation and firm value. *Management Science*, 65(2):879–895.

Cantner, U. and H. Graf (2006). The network of innovators in Jena: An application of social network analysis. *Research Policy*, 35(4):463–480.

Cantner, U. and A. Meder (2007). Technological proximity and the choice of cooperation partner. *Journal of Economic Interaction and Coordination*, 2(1):45.

Cantner, U. and S. Vannuccini (2017). Innovation and lock-in. *The Elgar Companion to Innovation and Knowledge Creation*, page 165.

Cohen, W. M. (2010). Fifty years of empirical studies of innovative activity and performance. *Handbook of the Economics of Innovation*, 1:129–213.

Cohen, W. M. and S. Klepper (1992). The anatomy of industry r&d intensity distributions. *The American Economic Review*, pages 773–799.

Cohen, W. M. and S. Klepper (1996). A reprise of size and R&D. *The Economic Journal*, pages 925–951.

Comin, D., G. Licht, M. Pellens, and T. Schubert (2019). Do companies benefit from public research organizations? the impact of the fraunhofer society in germany. Technical report, ZEW Discussion Papers.

Cowan, R. and N. Jonard (2004). Network structure and the diffusion of knowledge. *Journal of economic Dynamics and Control*, 28(8):1557–1575.

Crescenzi, R., G. de Blasio, and M. Giua (2018). Cohesion policy incentives for collaborative industrial research: evaluation of a smart specialisation forerunner programme. *Regional Studies*, pages 1–13.

Daiko, T., M. Dosso, P. Gkotsis, M. Squicciarini, A. Tuebke, A. Vezzani, et al. (2017). World top R&D investors: Industrial property strategies in the digital economy. Technical report, Joint Research Centre (Seville site).

d'Aspremont, C. and A. Jacquemin (1988). Cooperative and noncooperative R&D in duopoly with spillovers. *The American Economic Review*, 78(5):1133–1137.

David, P. A., B. H. Hall, and A. A. Toole (2000). Is public R&D a complement or substitute for private R&D? A review of the econometric evidence. *Research policy*, 29(4):497–529.

Dimos, C. and G. Pugh (2016). The effectiveness of r&d subsidies: A meta-regression analysis of the evaluation literature. *Research Policy*, 45(4):797–815.

Dorn, D., L. F. Katz, C. Patterson, J. Van Reenen, et al. (2017). Concentrating on the fall of the labor share. *American Economic Review*, 107(5):180–85.

Dosi, G., P. Llerena, and M. S. Labini (2006). The relationships between science, technologies and their industrial exploitation: An illustration through the myths and realities of the so-called european paradox. *Research policy*, 35(10):1450–1464.

Fernald, J. G. (2015). Productivity and Potential Output before, during, and after the Great Recession. *NBER Macroeconomics Annual*, 29(1):1–51.

Gabaix, X. (2011). The granular origins of aggregate fluctuations. *Econometrica*, 79(3):733–772.

Greene, W. (2006). A general approach to incorporating selectivity in a model. Technical report, New York University, Leonard N. Stern School of Business, Department of .

Hagedoorn, J. (2002). Inter-firm R&D partnerships: an overview of major trends and patterns since 1960. *Research policy*, 31(4):477–492.

Hall, B. H. and M. Trajtenberg (2004). Uncovering gpts with patent data. Technical report, National Bureau of Economic Research.

Heckman, J. J. (1979). Sample selection bias as a specification erros. *Econometrica*, 47(1):153–161.

Hernàndez, H., Guevara, N. Grassano, A. Tuebke, L. Potters, P. Gkotsis, A. Vezzani, et al. (2018). The 2018 EU Industrial R&D Investment Scoreboard. Technical report, Publications Office of the European Union.

HIDALGO, C. A., P.-A. BALLAND, R. BOSCHMA, M. DELGADO, M. FELDMAN, K. FRENKEN, E. GLAESER, C. HE, D. F. KOGLER, A. MORRISON, ET AL. (2018). The principle of relatedness. In *International conference on complex systems*, pages 451–457. Springer.

HOHBERGER, J. (2016). Does it pay to stand on the shoulders of giants? an analysis of the inventions of star inventors in the biotechnology sector. *Research Policy*, 45(3):682–698.

JONKERS, K. AND F. SACHWALD (2018). The dual impact of 'excellent' research on science and innovation: The case of Europe. *Science and Public Policy*, 45(2):159–174.

KAMIEN, M. I. AND I. ZANG (2000). Meet me halfway: research joint ventures and absorptive capacity. *International journal of industrial organization*, 18(7):995–1012.

KARLSSON, C. (2010). *Handbook of research on cluster theory*, volume 1. Edward Elgar Publishing.

KELLER, W. (2004). International technology diffusion. *Journal of economic literature*, 42(3):752–782.

LEE, M., B. SON, AND K. OM (1996). Evaluation of national r&d projects in korea. *Research Policy*, 25(5):805–818.

MAZZUCATO, M. AND D. J. ROBINSON (2016). Lost in Space? NASA and the Changing Public-Private Eco-System in Space. Technical report, SSRN Papers.

MCKELVEY, M., O. ZARING, AND D. LJUNGBERG (2015). Creating innovative opportunities through research collaboration: An evolutionary framework and empirical illustration in engineering. *Technovation*, 39:26–36.

MORA-VALENTIN, E. M., A. MONTORO-SANCHEZ, AND L. A. GUERRAS-MARTIN (2004). Determining factors in the success of R&D cooperative agreements between firms and research organizations. *Research Policy*, 33(1):17–40.

MOWERY, D. C. (2009). Plus ca change: Industrial R&D in the "third industrial revolution". *Industrial and corporate change*, 18(1):1–50.

NELSON, R. R. (1959). The simple economics of basic scientific research. *Journal of political economy*, 67(3):297–306.

Nelson, R. R. (1991). Why do firms differ, and how does it matter? *Strategic management journal*, 12(S2):61–74.

Nieto, M. J. and L. Santamaría (2010). Technological collaboration: Bridging the innovation gap between small and large firms. *Journal of Small Business Management*, 48(1):44–69.

Paier, M. and T. Scherngell (2011). Determinants of collaboration in european r&d networks: empirical evidence from a discrete choice model. *Industry and Innovation*, 18(1):89–104.

Rafols, I., M. M. Hopkins, J. Hoekman, J. Siepel, A. O'Hare, A. Perianes-Rodríguez, and P. Nightingale (2014). Big pharma, little science?: A bibliometric perspective on big pharma's r&d decline. *Technological Forecasting and Social Change*, 81:22–38.

Schwartz, M., F. Peglow, M. Fritsch, and J. Günther (2012). What drives innovation output from subsidized r&d cooperation?project-level evidence from germany. *Technovation*, 32(6):358–369.

Schwellnus, C., M. Pak, P.-A. Pionnier, and E. Crivellaro (2018). Labour share developments over the past two decades: the role of technological progress, globalisation and "winnert-takes-most" dynamics. *OECD Economic Department Working Papers*, (1503):0_1–58.

Terza, J. V. (1998). Estimating count data models with endogenous switching: Sample selection and endogenous treatment effects. *Journal of Econometrics*, 84(1):129 – 154. ISSN 0304-4076.

Terza, J. V. (2018). Two-stage residual inclusion estimation: A practitioners guide to stata implementation. *The Stata Journal*, 17(4):916–938.

Von Hippel, E. (1989). Cooperation between rivals: informal know-how trading. In *Industrial dynamics*, pages 157–175. Springer.

Wagner, C. S. and L. Leydesdorff (2005). Network structure, self-organization, and the growth of international collaboration in science. *Research policy*, 34(10):1608–1618.

Yang, H., Y. Zheng, and X. Zhao (2014). Exploration or exploitation? small firms' alliance strategies with large firms. *Strategic Management Journal*, 35(1):146–157.

A Characteristics of top R&D companies

Out of 2,500 top R&D companies (SB), 383 applied to FP7 funding. Of these 383 SB firms, 351 participated to funded projects, 32 to retained proposals that did not receive the funding. The total number of distinct SB firms found among the organizations of retained projects is 67 (out of 7,595 organizations in 10,602 proposals). Out of these 67 companies, 35 have participated also to successfully funded projects, while 32 did not receive any funding.

In Table 10, we compare the characteristics of top R&D firms that did not apply for FP7 funding (no participation) with those of top R&D firms that applied and obtained (funded proposal) or not (retained proposal) the FP7 funding.

Table 10: Medians by top R&D participation to FP7 Projects

	no participation	retained proposal	funded proposal
R&D†	28	69	229
Net Sales†	734	3,192	10,265
EBIT†	52	280	572
log(Employees)	8.3	9.6	11
R&D Intensity	4.3%	3.2%	4.5%
Labour productivity	12.2	12.4	12.3
Profitability	7.7%	10%	8.7%
N. firms	2117	32	351

Note: †Figures are in €mln. *R&D intensity* is calculated as the ratio between the logarithm of R&D expenditure and the logarithm of net sales. Labour productivity is calculated as the ratio between the logarithm of net sales and logarithm of employees. *Profitability* is the ratio between operating profits and net sales.

The table reports the medians of performance indicators such as R&D intensity, labour productivity and profitability—and of the variables used to construct such indicators—for the three groups of companies. R&D intensity is defined as the ratio between R&D spending and net sales; labour productivity is the net sales per employee; profitability (or profit margin) is the ratio between operating profits and net sales. Overall, the median R&D spending, net sales, operating profits and employees are larger for companies that participate to FP7 cooperative projects, irrespective of whether the projects obtained the funding. However, the set of firms that participated to funded project proposals has even larger R&D, sales, profits, and number of employees. Also, the medians of R&D intensity, labour productivity, and profitability are larger (statistically significant difference at 10%) for the groups of companies that applied to FP7.

B Results from mixed effects probit

Table 11: Mixed-effect probit: estimation results

Dep.var: top R&D (T)	coef.	std. err.
% of firms	1.135***	(0.033)
n.participants	0.059***	(0.001)
share funding per sector	1.338***	(0.314)
σ_{u_j}	0.109***	(0.039)
ICC	0.098***	(0.032)
N. obs	56,573	
LL	-19036.6	

Note: Significance code: *** p< 0.01, ** p< 0.05, * p< 0.1. ICC: residual intraclass correlation measures the correlation in the probability of having a top R&D firms for projects within the same sector.

Table 12: Correlation between residuals of eq. 3 and endogenous variable T

	S	$\hat{u_j}$	$\hat{\xi_{ij}}$	y_{Pubs}	y_{Pats}
T	0.0559*	0.1604*	0.4479*	-0.0026	0.0062
S		0.0216*	-0.0055	-	-
$\hat{u_j}$			0.0068	-0.0264	-0.0116
$\hat{\xi_{ij}}$				-0.0259	-0.0166
y_{Pubs}					0.9850*

Note: * Correlation coefficient statistically significant at 1%-level

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

Joint Research Centre

EU Science Hub