# TEAMING UP WITH LARGE R&D INVESTORS: GOOD OR BAD FOR KNOWLEDGE PRODUCTION AND DIFFUSION?

***Documents de travail GREDEG***
***GREDEG Working Papers Series***

SARA AMOROSO
SIMONE VANNUCCINI

**GREDEG WP No. 2025-41**
https://ideas.repec.org/s/gre/wpaper.html

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

Sara Amoroso†[1] and Simone Vannuccini[2]

[1]DIW Berlin

[2]GREDEG-CNRS, Université Côte d'Azur

September 30, 2025

**Abstract**

The participation of top R&D investors in publicly funded research collaborations is a common, yet largely unexplored phenomenon. It creates opportunities for knowledge spillovers and may increase the chance for a project to be funded. At the same time, the unbalanced nature of such partnerships could exacerbate power asymmetries and hinder the overall performance of such collaborations. In this paper, we examine whether cooperating with top R&D companies affects the innovative performance of publicly funded research consortia.

We build a fit-for-purpose dataset that matches information from the European Union's Seventh Framework Programme (FP7) on R&D collaborative projects and proposals with data on the world's top 2,500 companies with the highest R&D investment (R&D Scoreboard). Accounting for both sample selection and endogeneity in the

*We kindly acknowledge comments from Mara Giua (Roma Tre University), Liuchun Deng (Yale-NUS College), Mariagrazia Squicciarini (UNESCO), Laura Magazzini (Scuola Superiore Sant'Anna), participants to the Jena Economic Research Seminar, and one anonymous referee of the SPRU Working Paper Series, where an earlier version of the paper appeared.

†Corresponding author: samoroso@diw.de

participation of top R&D investors in a two-part count model framework, we find that teaming up with leading R&D companies increases the probability of obtaining funds. However, this comes at the cost of hindering the innovative performance of the funded projects, both in terms of patents and publications. In light of this evidence, the trade-offs 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:** Research collaboration; Public funding; Innovation performance; Appropriability; Top R&D investors

**JEL Classification:** L24; L25; O33

# 1 Introduction

Studies on collaboration in innovation explore the role played by heterogeneous actors (e.g., private firms vs. public research institutes, Coad et al., 2017), 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)). The heterogeneity in levels of R&D capital among partners of an R&D alliance is, instead, a less explored dimension of analysis.[1]

In this paper, we aim to fill this gap by investigating the impact of the world's top R&D investors' participation in publicly funded European R&D programmes on R&D partnerships' scientific and technological performance, measured by publications and patents. The key focus is on the economic mechanisms set in motion by the *R&D asymmetry* between actors involved in collaborations. In particular, the participation of R&D 'star' firms in research consortia and the uneven distribution of R&D capital might generate asymmetric spillovers (Atallah, 2005; Busom and Fernández-Ribas, 2008) and may result in a trade-off. On the one hand, there are many benefits of collaborating with top R&D investors: these are usually large, experienced innovators, endowed with consistent R&D capital that can produce knowledge spillovers (Kamien and Zang, 2000; Branstetter and Sakakibara, 2002) and foster technology diffusion (Baptista, 2000). 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., market power and the distribution of appropriation capabilities) might hinder the overall project's performance. This has been found by Crescenzi et al. (2018), for the case of a collaborative industrial research programme forerunning the recent Smart Specialisation Strategy of the European Union (EU). 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 comple-

[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)

mentarities generated by the collaboration (Cabral and Pacheco-de Almeida, 2018). This tension is particularly relevant for the European context, where the existence of a 'European paradox', with frontier research not adequately commercialized, is an important policy issue (Dosi et al., 2006; Albarrán et al., 2010; Jonkers and Sachwald, 2018). Our study focuses on the EU, given the importance placed on large-scale collaborative research projects.

We use a unique dataset that matches data on collaborative projects from the Seventh Framework Programme (FP7) of the European Community for research and technological development and demonstration activities (2007–2013) with information on the world's 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 innovative outcomes, namely projects' patents and publications.

More specifically, we test the trade-off an R&D project experiences when one of the project participants is a top R&D investor: leveraging private R&D capital and spillovers from an R&D star partner against the risk of asymmetric appropriation of innovation outcomes. To assess the impact of top R&D companies' participation on the scientific output of publicly funded projects, we estimate zero-inflated count models, accounting for sample selection and endogeneity, and do robustness checks using a two-part model. We find that the hypothesized trade-off is at work: the participation of top R&D firms increases 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 produced. For patents, a more in-depth analysis shows that the negative impact of top R&D firms' participation in a research consortium operates on the intensive margin, rather than on the extensive one: teaming up with a R&D star firm increases 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 construct our theoretical arguments building on the literature on collaboration and innovation. 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 on the results. In Section 5 we conclude, advancing our interpretation of the findings.

## 2 On collaboration and innovation

**Drivers of collaboration.** Knowledge as a commodity is a 'latent' public good (Nelson, 1991). Its production generates externalities that reduce private returns to investment in invention and innovation (Nelson, 1959; Arrow, 1962). For this reason, actors are expected to have a strong incentive to appropriate their know-how, in order to enjoy monopoly rents and first-mover advantage. Instead, knowledge exchange between firms, including formal collaboration in research consortia, is the norm rather than the exception. Allen (1983) identified early historical cases of 'collective inventions'. Since the 1980s, R&D consortia have been considered a tool to internalize R&D externalities (Branstetter and Sakakibara, 2002) and have been promoted as a technology policy tool. Recent trends in industrial R&D show an increase in inter-firm partnerships (Mowery, 2009), with strategic alliances (Hagedoorn, 2002), open innovation (Berchicci, 2013; Bogers, 2011) and markets for technologies (Arora and Gambardella, 2010) gaining momentum. Under certain conditions, cooperation in knowledge generation can be considered a rational decision: Von Hippel (1989) and d'Aspremont and Jacquemin (1988) showed that voluntary disclosure of knowledge can improve payoffs. A growing literature (Goyal and Moraga-Gonzalez, 2001; Petrakis and Tsakas, 2018; König et al., 2019; Junlong Chen and Wang, 2024) extended d'Aspremont and Jacquemin's framework to study the stability of R&D networks and the formation of desirable and undesirable collaborations given the strength of competition, the threat of entry, or cost differences.

Several arguments support the establishment of collaboration ties between firms: first, collaborations can be formed to exploit the opportunity (necessity) to combine complementary assets in order to make a (often complex) project technologically feasible; second, they may be started in order to build trust between partners, or alternatively to strengthen control; third, collaborations are a useful strategy to reduce costs, financial, and in terms of uncertainty (e.g., when used as a screening device (Conti and Marini, 2019)); fourth, collaborations allow to internalize spillovers. Experience — in general, and repeated experience with the same partners — is a crucial factor for the success of collaborations: Belderbos et al. (2015) find that persistent collaborations are the ones providing a systematically positive effect on innovation performance. Using Spanish data Mora-Valentin et al. (2004) also find that — to different degrees with respect to the type of actor — commitment, communication, reputation and previous linkages are drivers of successful cooperative agreements between firms and scientific organizations. Overall, all collaborations are driven by some form of proximity (Boschma, 2005) or relatedness (Hidalgo et al., 2018). Paier and Scherngell (2011) found that thematic and geographical proximity are major facilitators of partner choice in the context of the EU's 5th Framework Programme projects.

The structure of collaborations in innovative activities is studied in the literature on innovation and R&D networks. Empirically, innovation networks connect innovators — measured using patents' co-applications — or inventors — exploiting inventors' mobility across patenting firms (Cantner and Graf, 2006; Acemoglu et al., 2016). Research consortia as the ones we are interested in this paper, are a type of innovation network. Proximity is a key force shaping innovation networks: actors' closeness influences tie formation which, in turn, generates knowledge flows (through voluntary disclosure, spillovers, or labor mobility). This feeds back to adjust actors' relative positions in proximity spaces. As a consequence, networks evolve as actors find useful to repeat collaborations until they become too close to their partners, thus reducing the incentive to collaborate. Tomasello et al.

(2017) identify stylized facts of R&D networks, in particular core-periphery structures and rise-and-fall dynamics,

There are further reasons to seek collaboration, specifically with top R&D performers, which are predominantly large or superstar companies. One motivation to collaborate with superstar companies is the ease of access to markets, and hence to scale-up or to accelerate the commercialization of inventions, especially for high-tech products (Nieto and Santamaría, 2010). Another reason is reputation effects that can help to lower barriers to funding. Yang et al. (2014) claim that small firms can gain from alliances with large firms, in particular those focused on exploitation strategies, while alliances devoted to exploration activities entail higher risks of value appropriation by big actors. Top R&D firms have incentives to collaborate with smaller partners as well. In this case, more emphasis goes to the dimension of control of the collaboration (Nieto and Santamaría, 2010): large companies can use collaborations as tools to explore new technological trajectories and acquire ideas — when they do not opt directly for 'reverse killer acquisitions' (Kokkoris and Valletti, 2020).

**Collaborations and performance.** The effect of collaboration on economic performance is well summarized in Benfratello and Sembenelli (2002). 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 no clear patterns with respect to collaborations affecting labor productivity or price-cost margins. Barajas et al. (2012) find that participation in Framework Programmes has both indirect (through the generation of technological capabilities) and direct effects on labor productivity in a sample of Spanish firms. Aguiar and Gagnepain (2017) exploit data from the 5th Framework Programme and find that participation in joint research ventures produces a strong and positive effect on labor productivity and a small effect on profit margins.

For what concerns the relationship between collaboration and innovation performance, studies usually detect a positive effect (Xie et al., 2023). One of the main predictors of outcomes is the size of the project. 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 members of the consortia. Using a large dataset of subsidized R&D cooperations in the German region of Saxony, Schwartz et al. (2012) find that large firms' participation impacts positively patenting but not publications, while the opposite holds for Universities' participation. Evidence for the university-industry knowledge transfer channel suggests that — at least for publications and in specific fields of knowledge such as engineering — an inverted-U shaped relationship between collaboration size and outcomes exists (Banal-Estañol et al., 2015).

**Collaboration downsides.** There are downsides to collaborations. This is particularly the case when different types of organizations are involved, or when the collaboration occurs across sectors (Vivona et al., 2023). A first issue is that of 'too much' collaboration: larger networks can increase, rather than decrease, uncertainty. As a consequence, the likelihood of coordination failures grows. Managing big consortia consumes relevant resources, including attention and orchestration efforts. The literature on star firms, as well as that on preferential attachment in networks, usually highlights the advantages of being connected with a top-performing actor. For example, interactions with large public institutions (e.g., the Fraunhofer Institutes in Germany) positively affect partners' economic performance (Comin et al., 2019). However, negative effects can also emerge from this kind of collaboration, especially when considering large private actors (Cabral and Pacheco-de Almeida, 2018). We outline six arguments to rationalize the downsides of collaborations with large firms specifically.

A first argument is derived from the literature on firm size and innovation (Cohen, 2010). Large firms have a cost-spreading advantage (Cohen and Klepper, 1996) in performing R&D, but there is no empirical consensus on more-than-proportional returns to size. Lacking scale economies in innovation input, involving a top R&D investor in a joint research project may not directly translate into more innovation output.

A second argument relates to the bargaining-power weight of the actors: if power is unequally distributed among partners, top R&D companies' pressures to appropriate or non-disclose knowledge may prevail, reducing the measurable effect of the collaboration on outcomes.

A third argument looks at the nature of the joint project: if top companies are involved in large and more radical projects, then their presence may produce lagged effects that will not be captured immediately in the data (Hall and Trajtenberg, 2004). In this case, the presence of a top R&D performer in a consortium does generate benefits, but these are delayed in time.

A fourth argument suggests that the beneficial role of top actors may manifest in dimensions other than that of economic and innovation performance. For example, in the 'anchor tenant hypothesis' (Agrawal and Cockburn, 2003), large firms play the role of 'filter' of technology and knowledge diffusion and are crucial for the growth of regions, even though their impact leaves relatively little trace on measured outcomes.

A fifth argument relates to the trend of decreasing scientific publication by large firms (Rafols et al., 2014). 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 (increasing) compared to publications (decreasing). This shift is coupled by increasing specialization and increasing resorting to external sourcing of knowledge generation. We expect this trend to play a role in our context: if that holds, collaborations should experience a rise in patenting and a decrease in publications when top R&D performers are involved.

Finally, a sixth argument relates to a pure economic mechanism: negative outcomes for a research consortium 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.

Our working hypothesis is that the last mechanism is the one really 'biting' in our context: a trade-off is at work in collaborative innovation involving top R&D investors. In Section 5, we will return to these arguments when discussing the results of our analysis.

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 played by R&D star firms. Our study fills this gap and 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 superstar firms and winner-takes-most dynamics (Schwellnus et al., 2018; Eeckhout, 2022) that is generating a greater unequal distribution of market power and opportunities, widening the gap between frontier firms and followers.

## 3 Data

To investigate empirically the relationship between the participation of top corporate R&D investors in R&D collaborative projects and innovation performance, 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 comes from the 7th Framework Programme of the European Commission. The Seventh Framework Programme (FP7) was a major funding initiative by the EU to support research, technological development, and innovation. It ran from 2007 to 2013 and aimed to strengthen the EU's position as a global leader in science and technology, foster economic growth, and address key societal challenges. FP7 had a budget of approximately €50 billion, making it one of the largest publicly funded R&D programs in the world at the time. FP7 was followed by Horizon 2020 (2014—2020), which expanded its focus on innovation and societal impact while building on FP7's foundation.[2]

FP7 data is extracted from CORDIS (Community Research and Development Information Service)[3] 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). The matching with SB data is done using VAT numbers.

Data on FP7 funded projects contain information on 28,454 distinct organizations participating in 24,502 FP7 projects. Each organization can participate in 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 re-

[2]More details on the FP7 can be found here.

[3]https://data.europa.eu/data/datasets/cordisfp7projects?locale=en. Please note that the data on patents and publications has been rendered public only in 2018. At the beginning of this project, in 2016, we had access to patents applications and publications, as an interservice consultation of the European Commission. Our total number of publications and patents do not match to the final published numbers as many articles took longer to get published and many patents have not been granted.

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 | 1000 | 9.5 | 57535 | 15.2 | 41719 | 17.4 |
| Research Infrastructures | 326 | 7.7 | 24821 | 4.1 | 18561 | 1.3 |
| Nanosciences | 802 | 39.6 | 18999 | 52.1 | 15093 | 54.1 |
| Food, Agriculture & Fish | 509 | 9.0 | 11922 | 12.1 | 6122 | 13.3 |
| Marie-Curie | 1905 | 11.1 | 12680 | 4.9 | 4843 | 10.8 |
| Energy | 359 | 42.0 | 4748 | 44.7 | 3332 | 52.3 |
| Security | 310 | 40.9 | 2254 | 13.7 | 1615 | 10.8 |
| Transport | 713 | 44.5 | 1839 | 55.1 | 790 | 67.0 |
| Research for SMEs | 1026 | 7.6 | 1220 | 6.6 | 777 | 4.9 |
| Environment | 489 | 8.4 | 8393 | 4.6 | 643 | 13.2 |
| Nuclear Fission | 128 | 21.1 | 1195 | 20.6 | 367 | 27.0 |
| International Cooperation | 149 | 0.6 | 558 | 0 | 281 | 0 |
| Space | 259 | 29.7 | 2233 | 18.3 | 156 | 25.6 |
| Science in Society | 155 | 3.2 | 561 | 2.5 | 9 | 0.0 |
| Socio-economic sciences | 250 | 0.4 | 2150 | 0.0 | 0 | 0.0 |
| Total | 8380 | 13.3 | 151108 | 17.4 | 94308 | 20.9 |

peatedly 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. Altogether, these projects account for almost 50% of the FP7 EC budget.

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 in 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 the project level. We group the number of patents and publications in 5 categories (0, 1, 2–10, 11–100, >100) and 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 projects 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 |
| | 1 | 1.87 | 0.64 | 0.00 | 0.00 | 0.00 | 2.52 |
| | 2-10 | 1.66 | 0.04 | 2.96 | 0.02 | 0.00 | 4.68 |
| | 11-100 | 0.05 | 0.00 | 0.04 | 4.11 | 0.02 | 4.21 |
| | >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 in the scoreboard, 383 applied to FP7 funding. Of these 383 top R&D firms, 351 participated in funded projects, 32 in retained proposals that did not receive the funding. Table 11 in Appendix A reports and compares 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 in 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 fewer 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. Finally, Table 13 displays correlation coefficients for the main variables used in the empirical analysis. The variables exhibit low to moderate correlation, suggesting that multicollinearity may not be a problem in our analysis.

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 |

*Note*: *two-tailed P value is less than 0.001

Table 5: Correlation table

| Variables | N. pubs | N. pats | top R&D | N. parts | % of firms | cost | duration |
|---|---|---|---|---|---|---|---|
| N. patents | 0.99 | 1.00 | | | | | |
| top R&D ($T$) | -0.00 | 0.01 | 1.00 | | | | |
| N. participants | 0.12 | 0.09 | 0.21 | 1.00 | | | |
| % of firms | -0.07 | -0.03 | 0.29 | 0.01 | 1.00 | | |
| Tot. cost | 0.13 | 0.11 | 0.22 | 0.56 | 0.07 | 1.00 | |
| duration | 0.10 | 0.07 | 0.09 | 0.26 | -0.20 | 0.36 | 1.00 |

## 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 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. The level of analysis is that of projects.

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,

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

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

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 ($\xi_{ij}$). Tables 12 and 13 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 in 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 one 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 6 and 7. Generally, we find that the participation of leading R&D firms in publicly funded R&D consortia is negatively related to both the number of patents and publications. More specifically, in Table 6 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) suggests that the ZINB model provides the best fitting and the most parsimonious specification.[4] 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.

[4]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 6: 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 7: 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 the two-part model.[5] 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.[6] 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 makes 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 fewer incentives to disclose their 'patentable' ideas to the other consortium 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.).

[5] We test for multicollinearity in the OLS part, and the VIF values are all below 5, suggesting lack of high collinearity: size (2.71), n. participants (2.38), duration (1.52), % of firms (1.24), top R&D ($T$) (1.12).

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

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 academia, as they can get access to state-of-the-art knowledge, without waiting for it to be published. Publishing is potentially a signaling and commercialization strategy, and a lever to retain industrial researchers (Rotolo et al., 2022). However, companies do not have the same incentives to disclose the results of their research by publishing it as other types of organization. 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 8 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.[7] Despite the tests confirm the endogeneity

[7]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 13 in Appendix B) introduces multicollinear-

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.

ity and, as a result, the estimated coefficients associated with these two variables are both extremely high and of opposite signs.

Table 8: 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.

## 4.3 Endogenous participation in the selection

In this subsection, we consider the case where the participation of a top R&D firm 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 9 displays the results from the selection equation (2), where we compare 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 increases 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 confirms 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 9, compute the inverse Mill's ratio, and use it as an additional regressor in equation (1).

Table 9: Selection Equation $P(S=1)$ - effects of top R&D 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 10: 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 9

Results of a ZINB and two-part model for patents and publications are reported in Table 10 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 of patenting or publishing is negative. Hence, when simultaneously correcting for selection bias and endogeneity, the results suggest that including an 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 whether establishing a collaboration 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 in 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 increases 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 publications and patents. Also, results from a two-part model show that teaming up with an 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 deduced from our empirical analysis, we offer three non-exclusive interpretations.

First, following the anchor tenant hypothesis mentioned in Section 2, the participation of top R&D players has indirect beneficial effects, such as signaling and reputation effects. While these effects are well reflected in the greater estimated probability of obtaining funds when these top firms are included, they may not be 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 of the projects that are 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 projects but represent measures of input to economic performance 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 of collaboration due to corporate incentives and partnership asymmetries. Despite the novel empirical approach and data used, we just began to explore this research trajectory, that has the potential to increase our understanding of the collaboration-performance nexus and to inform policy making.

## References


Acemoglu, D., U. Akcigit, and W. R. Kerr (2016). Innovation network. *Proceedings of the National Academy of Sciences*, 113(41):11483–11488.

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.

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 cooperation—happy 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.

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.

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.

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.

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

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.

Coad, A., S. Amoroso, and N. Grassano (2017). Diversity in one dimension alongside greater similarity in others: evidence from fp7 cooperative research teams. *The Journal of Technology Transfer*, 42:1170–1183.

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 (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.

Conti, C. and M. A. Marini (2019). Are you the right partner? r&d agreement as a screening device. *Economics of Innovation and New Technology*, 28(3):243–264.

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.

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.

Eeckhout, J. (2022). *The Profit Paradox: How Thriving Firms Threaten the Future of Work*. Princeton University Press.

Goyal, S. and J. L. Moraga-Gonzalez (2001). R&d networks. *Rand Journal of Economics*, pages 686–707.

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.

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.

JUNLONG CHEN, J. S., XIAOMIN SUN AND Y. WANG (2024). Generic technology r&d decision with technology spillover, cost difference, and bargaining power under oligopoly competition. *Economics of Innovation and New Technology*, 0(0):1–24.

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.

KOKKORIS, I. AND T. VALLETTI (2020). Innovation considerations in horizontal merger control. *Journal of Competition Law & Economics*, 16(2):220–261.

KÖNIG, M. D., X. LIU, AND Y. ZENOU (2019). R&d networks: Theory, empirics, and policy implications. *Review of Economics and Statistics*, 101(3):476–491.

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.

Petrakis, E. and N. Tsakas (2018). The effect of entry on r&d networks. *The RAND Journal of Economics*, 49(3):706–750.

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.

Rotolo, D., R. Camerani, N. Grassano, and B. R. Martin (2022). Why do firms publish? a systematic literature review and a conceptual framework. *Research Policy*, 51(10):104606.

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.

Tomasello, M. V., M. Napoletano, A. Garas, and F. Schweitzer (2017). The rise and fall of r&d networks. *Industrial and corporate change*, 26(4):617–646.

Vivona, R., M. A. Demircioglu, and D. B. Audretsch (2023). The costs of collaborative innovation. *The Journal of Technology Transfer*, 48(3):873–899.

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

Xie, X., X. Liu, and J. Chen (2023). A meta-analysis of the relationship between collaborative innovation and innovation performance: The role of formal and informal institutions. *Technovation*, 124:102740.

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 in funded projects, 32 in 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 in successfully funded projects, while 32 did not receive any funding.

In Table 11, 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 11: 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% |
| Labor 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. *Labor 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, labor 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; labor 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 in FP7 cooperative projects, irrespective of whether the projects obtained the funding. However, the set of firms that participated in funded project proposals has even larger R&D, sales, profits, and number of employees. Also, the medians of R&D intensity, labor 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 12: 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) |
| $\sigma_{u_j}$ | 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 13: 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

**DOCUMENTS DE TRAVAIL GREDEG PARUS EN 2025**
***GREDEG Working Papers Released in 2025***

**2025-01** Bruno Deffains & Frédéric Marty
*Generative Artificial Intelligence and Revolution of Market for Legal Services*

**2025-02** Annie L. Cot & Muriel Dal Pont Legrand
*"Making war to war" or How to Train Elites about European Economic Ideas: Keynes's Articles Published in L'Europe Nouvelle during the Interwar Period*

**2025-03** Thierry Kirat & Frédéric Marty
*Political Capitalism and Constitutional Doctrine. Originalism in the U.S. Federal Courts*

**2025-04** Laurent Bailly, Rania Belgaied, Thomas Jobert & Benjamin Montmartin
*The Socioeconomic Determinants of Pandemics: A Spatial Methodological Approach with Evidence from COVID-19 in Nice, France*

**2025-05** Samuel De La Cruz Solal
*Co-design of Behavioural Public Policies: Epistemic Promises and Challenges*

**2025-06** Jérôme Ballet, Damien Bazin, Frédéric Thomas & François-Régis Mahieu
*Social Justice: The Missing Link in Sustainable Development*

**2025-07** Jérôme Ballet & Damien Bazin
*Revoir notre manière d'habiter le monde. Pour un croisement de trois mouvements de pensée: capabilités, services écosystémiques, communs*

**2025-08** Frédéric Marty
*Application des DMA et DSA en France : analyse de l'architecture institutionnelle et des dynamiques de régulation*

**2025-09** Éléonore Dodivers & Ismaël Rafaï
*Uncovering the Fairness of AI: Exploring Focal Point, Inequality Aversion, and Altruism in ChatGPT's Dictator Game Decisions*

**2025-10** Hayk Sargsyan, Aleksandr Grigoryan & Olivier Bruno
*Deposits Market Exclusion and the Emergence of Premium Banks*

**2025-11** Lubica Stiblarova & Anna Tykhonenko
*Talent vs. Hard Work: On the Heterogeneous Role of Human Capital in FDI Across EU Member States*

**2025-12** Mathieu Chevrier
*Social Reputation as one of the Key Driver of AI Over-Reliance: An Experimental Test with ChatGPT-3.5*

**2025-13** Rania Belgaied
*L'accessibilité aux médecins généralistes libéraux pour la ville de Nice*

**2025-14** Lorenzo Corno, Giovanni Dosi & Luigi Marengo
*Behaviours and Learning in Complex Evolving Economies*

**2025-15** Grégory Donnat
*Real Exchange Rate and External Public Debt in Emerging and Developing Countries*

**2025-16** Michela Chessa
*Politics as A (Very) Complex System: A New Methodological Approach to Studying Fragmentation within a Council*

**2025-17** Mathieu Lambotte & Benjamin Montmartin
*Competition, Conformism and the Voluntary Adoption of Policies Designed to Freeze Prices*

**2025-18** Leonardo Ciambezi & Alessandro Pietropaoli
*Relative Price Shocks and Inequality: Evidence from Italy*

**2025-19** Matteo Orlandini, Sebastiano Michele Zema, Mauro Napoletano & Giorgio Fagiolo
*A Network Approach to Volatility Diffusion and Forecasting in Global Financial Markets*

**2025-20** Daniele Colombo & Francesco Toni
*Understanding Gas Price Shocks: Elasticities, Volatility and Macroeconomic Transmission*

**2025-21** Simone Vannuccini
*Move Fast and Integrate Things: The Making of a European Industrial Policy for Artificial Intelligence*

**2025-22** Patrice Bougette, Oliver Budzinski & Frédéric Marty
*Revisiting Behavioral Merger Remedies in Turbulent Markets: A Framework for Dynamic Competition*

**2025-23** Amal Ben Khaled, Rami Haj Kacem & Nathalie Lazaric
*Assessing the Role of Governance and Environmental Taxes in Driving Energy Transitions: Evidence from High-income Countries*

**2025-24** Sandye Gloria
*Emergence. Another Look at the Mengerian Theory of Money*

**2025-25** Barbara Brixová, Anna Tykhonenko, Ľubica Štiblárová & Marianna Siničáková
*Middle Income Convergence Trap Phenomenon in CEE Countries*

**2025-26** Grégory Donnat, Maxime Menuet, Alexandru Minea & Patrick Villieu
*Does Public Debt Impair Total Factor Productivity?*

**2025-27** Nathalie Lazaric, Loubna Echajari & Dorota Leszczyńska
*The Multiplicity of Paths to Sustainability, Grand Challenges and Routine Changes: The Long Road for Bordeaux Winemakers*

**2025-28** Olivier Brette & Nathalie Lazaric
*Some Milestones for an Evolutionary-Institutional Approach to the Circular Economy Transition*

**2025-29** Pierre Boutros, Eliana Diodati, Michele Pezzoni & Fabiana Visentin
*Does Training in AI Affect PhD Students' Careers? Evidence from France*

**2025-30** Luca Bargna, Davide La Torre, Rosario Maggistro & Benjamin Montmartin
*Balancing Health and Sustainability: Optimizing Investments in Organic vs. Conventional Agriculture Through Pesticide Reduction*

**2025-31** Jean-Luc Gaffard
*Equilibre général et évolution: un éclairage théorique*

**2025-32** Patrice Bougette, Frédéric Marty & Simone Vannuccini
*Competition Law Enforcement in Dynamic Markets: Proposing a Flexible Trade-off between Fines and Behavioural Injunctions*

**2025-33** Taekyun Kim, Gabriela Fuentes & Simone Vannuccini
*When Robots Met Language (Models): An Exploration of Science and Strategy in the vision-language-action Models Space*

**2025-34** Frédéric Marty & Thierry Warin
*Prévenir les risques de parallélisme de comportements par développeur commun : de la constatation des limites des règles de concurrence au retour du Nouvel Instrument Concurrentiel*

**2025-35** Luca Fontanelli, Mauro Napoletano & Angelo Secchi
*Rethinking Volatility Scaling in Firm Growth*

**2025-36** Charlie Joyez
*Connectivity and Contagion: How Industry Networks Shape the Transmission of Shocks in Global Value Chains*