Corporate scientists as the triggers of transitions towards firms' exploration research strategies

Barge-Gil, Andres and D'Este, Pablo and Herrera, Liliana

ICAE. Universidad Complutense de Madrid, INGENIO (CSIC-UPV), Universidad Politécnica de, Departamento de Dirección y Economía de la Empresa, Área de Organización de Empresas, Universidad de León

2018

Corporate scientists as the triggers of transitions towards firms' exploration research strategies

Andrés Barge-Gil[1], Pablo D'Este[2] and Liliana Herrera[3]

Abstract

Exploration-oriented research strategies represent a critical factor for firms' innovation performance and long term survival since they influence the capacity of these organizations to develop breakthrough innovations and move up the quality ladder of product development. However, setting in motion and sustaining exploratory research strategies are not straightforward. In this study, we contend that the availability of corporate scientists (i.e. personnel with PhD qualifications in R&D units) is a triggering factor for two critical transitions in firms' R&D strategies: i) initiation of exploration-oriented research activities (*exploration-enacting strategy*); and ii) increased commitment to exploration-oriented research activities (*exploration-deepening strategy*). We conduct our analysis on a large sample of Spanish manufacturing firms. We address endogeneity concerns by using the exogenous supply of PhD graduates as an instrumental variable. Our results show that firms recruiting doctoral graduates (PhDs) to their R&D units increase the likelihood of initiating and strengthening exploration-oriented research strategies. The implications of these findings for innovation management and policy are discussed.

Keywords: Scientists, exploration research, R&D strategies, PhDs, instrumental variables.

[1] Department of Economic Analysis II (Quantitative Economics). Campus de Somosaguas (28223), Pozuelo de Alarcón, Madrid, Spain. abarge@ccee.ucm.es. +34 91 394 2355.

[2] INGENIO (CSIC-UPV), Universidad Politécnica de Valencia, Ciudad Politécnica de la Innovación, Edif 8E 4 Planta, Camino de Vera s/n, 46022 Valencia. pdeste@ingenio.upv.es. +34 963 877 007x77048

[3] Departamento de Dirección y Economía de la Empresa, Área de Organización de Empresas, Universidad de León, Campus de Vegazana s/n, 24071 León, Spain liliana.herrera@unileon.es

1. Introduction

The management and economics of innovation literatures have convincingly argued about the importance of knowledge creation processes as fundamental drivers of firms' innovation performance (Nelson and Winter, 1982; Nonaka, 1994; Teece et al., 1997). More recently, the role of individuals has been highlighted as a micro-foundation of processes oriented to the generation of ideas and knowledge in organizations (Felin and Foss, 2005; Groysberg et al., 2008). Innovation research has become increasingly interested in the role of knowledge workers, and star scientists in particular, to explain the innovation process and firms' successful technological advancement (Gittelman and Kogut, 2003; Subramanian et al., 2013; Grigoriou and Rothaermel, 2014).

However, how scientists contribute to a shift in R&D strategies has not been analyzed in detail. Although we build in part on Lacetera et al. (2004) and Tzabbar (2009) who analyze the effects of recruiting qualified scientists on firms' innovation performance, we focus first on the impact of corporate scientists on the shift towards exploration-oriented research strategies, as opposed to their impact on the adoption of new capabilities and innovative performance. Second, our analysis is not restricted to a specific type of scientist (i.e. star scientist or technologically distant scientist) (Tzabbar et al., 2013); instead we generalize to the 'mean' scientist by measuring the proportion of highly qualified researchers among the firm's R&D employees. Third, our analysis includes all manufacturing industries, including science intensive and non-science intensive industries.

Specifically, we suggest that the availability of highly qualified scientists in R&D units can trigger a critical transition in the firm's R&D strategy. We investigate whether, and to what extent the proportion of PhD graduates (relative to the number of the firm's R&D employees) contributes to the firm's greater involvement in exploration-oriented research strategies - that is, search processes oriented to knowledge creation, which involve a departure from the firm's current knowledge base, as opposed to search oriented to development-related activities aimed at refinement and extension of existing competencies (March, 1991).

We distinguish two types of transition to an exploration-oriented research strategy. We define *R&D exploration-enacting* as the initiation of exploration research activities in organizations not previously involved in exploratory research. We define *R&D exploration-deepening* as the strengthening of commitment to exploration-oriented research via persistent exploration activity over time or intensification of investment in exploration-oriented research.

We test our propositions on a large sample of Spanish manufacturing and R&D performing firms over the period 2006-2012. We find that organizations with larger proportions of PhD graduates in the R&D unit are more likely to embark on exploration-oriented research activity. We find also that the availability of doctoral graduates contributes to increasing persistence in exploration-oriented research activities, and also the proportion of resources devoted to exploratory research over time. These results are robust to alternative explanatory factors, and we address the problem of reverse causality by adopting an instrumental variables approach. More precisely, we build an indicator of the exogenous supply of PhD graduates for each firm, using data on doctoral graduates in each university and scientific area, and a correspondence matrix of scientific areas and industries tested in Abramovsky et al. (2007). Additionally, we construct a second instrumental variable based on the number of PhD graduates working in the industry of each firm.

The rest of the paper is organized as follows. Section 2 presents the conceptual background. Section 3 describes the data and methodology, and provides some descriptive statistics. Section 4 presents the results and some robustness checks, and section 5 provides a discussion and conclusions.

2. Conceptual Background

There is broad agreement in the literature that exploration-oriented R&D is a fundamental component of firms' R&D strategies (March, 1991; Levinthal and March, 1993; Walrave et al., 2015; Swift, 2016). Exploratory search involves the process of experimentation oriented to identification of novel solutions and inventions, through a deliberate effort to move away from current organizational routines and knowledge bases (March, 1991; Levinthal and March, 1993). While this type of research orientation entails high risks and uncertainties (compared to exploitation-oriented research), it is expected that it will contribute to the achievement of sustainable competitive advantage by building the technological capabilities that can lead to the launch of new products and enchance product development.

Embarking on and maintaining exploration-oriented research is not straightforward. On the one hand, broadening the search space by adopting a more exploration-oriented strategy represents a radical transition for many firms. This is because many firms often lack the necessary human capital and internal expertise to undertake these activities (García-Quevedo et al., 2012; Tzabbar, 2009; Lowe and Veloso, 2015), and also because the expected outcomes of an exploration-oriented strategy are uncertain and involve high-risk choices which could lead to innovation failure (Garcia et al., 2003; D'Este et al., 2016; Kim and Kim, 2015).

On the other hand, firms need to be persistent in their R&D efforts in order to be effective. Many firms struggle with the transition from a sporadic approach to R&D activities towards a more continuous, permanent commitment to R&D effort. These difficulties may be especially severe in the case of exploratory-oriented research activities, since firms tend to give priority to exploitative research (at the expense of exploration) in order to increase the short term returns from R&D activity (Swift, 2016). The challenges related to R&D persistence can be particularly strong for small and medium sized enterprises (SMEs), and firms in non-R&D intensive industries (Mañez et al., 2015). Ahlin et al. (2013) examine how in the absence of previous R&D experience, firms tackle the issues involved in the transition to persistent R&D. Many firms engage only in sporadic formal exploration research activities, and rely on external services provided by specialized R&D partners. This can reduce their capacity to build the internal knowledge base required to support product and process upgrading.

Previous research shows that the capacity of an organization to engage in exploration-oriented research activity is deeply rooted in the skills and behaviors of individual organizational members (Mom et al., 2015). However, analyses of the drivers of exploration pay small attention to the role of the human resources involved directly in the processes of generation and absorption of knowledge. Two different approaches characterize research on this topic. The first analyzes the balance between exploration and exploitation, and focuses on the level of environmental dynamism, the executive team and board of directors' strategic vision, and the viability of financial resources (Walrave et al., 2015). The second seeks to understand exploration at the organizational level by focusing on the organization culture and structure as determining factors (Mom et al., 2015).

Other research highlights the important role of corporate scientists in the performance of R&D activities within organizations but does not specifically link it to a shift to exploration-oriented research activity. For instance, some studies show that scientists generate ideas which set in motion innovation processes, and monitor the external environment to take advantage of knowledge generated outside the firm (Herrera and Nieto, 2010). By hiring highly qualified scientists, firms acquire not only the most-up-to-date scientific knowledge but also the skills needed to exploit it and to produce new knowledge (Lee et al., 2010; Herrera and Nieto, 2015). Since much of the knowledge arising from scientific breakthroughs is characterized by excludability, the complexity and tacitness of such knowledge makes its transmission difficult, and requires the active participation of actors with unique backgrounds and skills (Zucker et al., 1998). Consistently, there is a strand of work which points to the link between corporate scientists and innovation performance (Deeds et al., 2000; Rothaermel and Hess, 2007; Herrmann and Peine, 2011).

Against this background, this research investigates the antecedents of shifts in R&D towards exploration-oriented research strategies by looking at the scientific profile of individuals within firms' R&D functions. More specifically, we contend that the availability of highly qualified scientists in R&D units is a triggering factor for two critical transitions in R&D strategies: (i) the decision to initiate exploration activities (exploration-enacting strategy), and (ii) the decision to hold and intensify exploration-oriented research activities (exploration-deepening strategy). The micro-foundations underlying the connections between the recruitment of highly skilled R&D workers and the enforcement of these two types of R&D strategy transitions are discussed in the succeeding sections.

2.1. Starting exploration-oriented R&D activities: towards an exploration-enacting strategy

In order to sustain their competitive position and respond to market demands, firms need to renew their technological knowledge bases. One way to achieve this is through engagement in exploration research activities (Garcia et al., 2003). These activities require a structured process and expert knowledge to determine the research direction and methods. To decide the research direction firms depend on guidance from highly qualified research personnel.

Prior research links the recruitment of scientific personnel to the likelihood to search beyond the firm's existing technological boundaries (Al-Laham et al., 2011). These highly skilled human resources are well-positioned to make decisions regarding the course and orientation of R&D activities, since scientists perform gate-keeping and boundary spanning roles which enable the development of firm capabilities in collecting, assimilating, filtering and applying external knowledge (Rothaermel and Hess, 2007). Due to their exposure to cutting-edge scientific knowledge and market needs, corporate scientists provide firms with substantial competitive advantage through their ability to assess the commercial implications of current research (Zellner, 2003). Firms that hire highly qualified scientists are well placed to engage in exploration oriented research, and assess the commercial potential of the new knowledge they produce.

Several studies show that the presence of scientists increases the firm's propensity to engage in science-driven research activities, and contributes to creating an environment favorable to experimentation. Lacetera et al. (2004) found that star scientists (those with large numbers of publications) act as magnets for other researchers, leading to a sharp improvement in the firm's ability to generate publications and patents. The capacity to attract qualified researchers depends largely on the presence of highly accomplished scientists who ensure credibility and legitimacy of the R&D strategies oriented towards basic research and exploration. In high tech sectors, Ding (2011) demonstrates that the presence of PhD graduates in biotech firms is strongly related to these firms' adoption of an open science policy which encourages personnel to carry out basic scientific research. Employees with a doctoral degree are able to develop a more profound understanding of scientific knowledge and cutting-edge scientific research methods. Consequently, highly qualified scientists contribute to enhancing the firm's appreciation of the specificities of exploratory research processes and its potential benefits.

A complementary stream of research shows that the recruitment of scientists with distant knowledge (knowledge beyond the firm's technological boundaries) leads to engagement in exploratory research activity (Tzabbar et al., 2013). The presence of such scientists contributes also to improving the firm's exploration abilities and competences. For example, Toole and Czarnitzki (2009) show that scientists oriented to the exploration of scientific opportunities (compared to specializing in identifying commercial opportunities) improve firms' research-oriented task performance significantly. Finally, Herrmman and Peine (2011) show that scientists with a heterogeneous knowledge background are able to reorient their R&D focus in order to pursue innovations associated to exploratory research activities.

In summary, the presence of highly qualified scientists in firms' R&D units is likely to be critical for facilitating the experimentation and discovery processes and favor the adoption of an exploration-oriented strategy in the firm's research activity. Thus, we hypothesize that:

Hypothesis 1: *Firms employing highly qualified scientists are more likely to embark on a transition to an R&D exploration-enacting strategy, such that these firms are more likely to initiate an exploration-oriented research activity.*

2.2. Holding and intensifying exploration-oriented R&D activities: towards an exploration-deepening strategy

While current research highlights the innovation-related benefits of persistent R&D efforts, many firms exhibit discontinuous engagement in R&D (Peters, 2009; Antonelli et al., 2012; Triguero and Córcoles, 2013). Since technological knowledge bases rapidly become outdated, firms need to build the capabilities to generate new knowledge, and to recombine existing bodies of knowledge (Kough and Zander, 1992). In order to nurture the capacity to renew their knowledge base, firms need to commit to exploration-oriented research. Several arguments support the idea that corporate scientists play a critical role in the processes of maintenance and renewal of technological competencies.

An important role is the function of knowledge provider. Scientists provide firms with intellectual capital derived from their advanced education, training and experience in scientific and technological research (Subramaniam and Youndt, 2005; Luo et al., 2009). The literature links the role of knowledge provider with "learning through hiring" which assumes that a new recruit's tacit knowledge diffuses internally and becomes part of the firm's overall knowledge base (Singh and Agrawal, 2011). Hiring doctoral graduates is important for companies keen to profit from scientific knowledge. By hiring well qualified scientists, firms incorporate not only the most up-to-date scientific knowledge but also the skills needed to produce and exploit scientific knowledge (Lee et al., 2010).

Based on their research training, scientists not only represent a one-time transfer of knowledge but also are able to facilitate continuous access to new discoveries and techniques which otherwise would be unavailable to the firm (Zucker et al., 1998). Several works measure productivity in terms of patents in order to show the influence of knowledge providers on R&D activities. There is evidence showing that firms that recruit scientists increase both their patenting propensity and also the quality of their patents over time (Tzabbar, 2009; Singh and Agrawal, 2011; Al-Laham et al., 2011; Subramanian et al., 2013). Analyses of patent citations show that firms that recruit scientists increase their knowledge bases and exploit and benefit from their ideas (Song et al., 2003; Sing and Agrawal, 2011).

Recent literature suggests that specialized disciplinary knowledge is not the only contribution scientists make to exploratory R&D activities. A shift from sporadic to continuous exploration efforts may require scientific skills, experience and access to knowledge networks in addition to intellectual capital (Luo et al., 2009). Doctoral research equips scientists with various skills which are useful to firms at different stages in the innovation process (Herrera and Nieto, 2015). Some of these skills go beyond the scientist's disciplinary background, and are related to the capacities to formulate, structure and solve problems. Zellner (2003) suggests that these wider skills are more valuable for consolidating the firm's R&D commitment and innovation performance than the skills deriving from a specific disciplinary knowledge background.

Since few organizations are able to generate internally the knowledge they need to support ongoing exploration efforts, they try to source external knowledge through strategic innovation alliances. Scientists are important for enabling such a strategy based on their ability to: (i) integrate different and diffuse knowledge (Hess and Rothaermel, 2012); (ii) connect firms to scientific knowledge required for new R&D activity (Subramaniam and Youndt, 2005); and (iii) attract research partners by providing organizational legitimacy (expert personnel managing R&D), and reducing a potential partner's evaluative uncertainties (Luo et al., 2009). Several works point to the influence of corporate scientists on R&D alliances. Luo et al. (2009) find that increasing the number of scientists in the firm attracted more R&D partners, especially if the partnering firms were located outside the company's primary industry network.

Based on the preceding discussion, we contend that highly qualified scientists are critical to firms that intend to maintain and/or intensify exploratory R&D activities. We hypothesize that:

Hypothesis 2: *Firms employing highly qualified scientists are more likely to embark on a transition to an R&D exploration-deepening strategy. These firms are more likely to: a) exhibit persistent exploration-oriented research activities; and b) intensify research resource commitments to exploration-oriented research over time.*

3. Data and methodology

The discussion in section 2 on the impact of corporate scientists on the firm's research strategy suggests some expectations regarding their influence on the initiation and intensification of exploratory-oriented research. We have argued that although there has been some work in this area (e.g. Lacetera et al., 2004; Tzabbar, 2009), the role of corporate scientists on changes to R&D strategies has not been analyzed in detail. Our study differs from previous research in three important points. First, we do not restrict the analysis to specific types of scientist (star or technologically distant scientists) but generalize it to the 'mean' scientist by measuring the proportion of highly qualified researchers among R&D employees. Second, we do not restrict the analysis to a specific science-intensive industry, such as drugs or biotechnology, but cover all manufacturing industries. Finally, our analysis examines three types of research-related transitions to exploration activities: start exploration, maintain exploration and intensify exploration. In this section, we describe the data used in the analysis, define the key variables and discuss the method employed.

3.1 Data

The empirical analysis employs information from the Spanish Technological Innovation Panel (PITEC). This statistical instrument was developed by the Spanish Institute of Statistics (with advice from a group of university researchers) to study the evolution in Spanish firms' innovation activities over time. The database is available to researchers from an official web site.[4] PITEC information is based on the Spanish Innovation Survey, and structured similar to the Community Innovation Surveys (CIS). CIS-type surveys have been used widely to analyze

[4] http://icono.fecyt.es/Paginas/home.aspx. For confidentiality reasons, an anonymized version of the data is available on the web site. The PITEC anonymization procedure is described on the web page.

innovation-related research questions in economics and management (Cassiman and Veugelers, 2002; Laursen and Salter, 2006; Mairesse and Mohen, 2010).

However, PITEC provides an important value-added: it incorporates data from the Spanish R&D survey which makes it appropriate for the present study for three reasons. First, it provides information on the number of doctoral graduates in firms' R&D units. Second, it provides information on the breakdown of firms' R&D investment in basic, applied and development research activity. Third, PITEC is designed as a panel data survey which allows us to observe firms' R&D transitions over time. The present study would not be feasible using a cross-sectional data. We use data for the period 2006-2012 since prior to 2006 the number of R&D personnel was not defined in terms of full-time equivalent units. Finally, in order to consider a homogenous sample of firms to investigate shifts in research strategies, we restrict our sample to: (i) manufacturing firms, since R&D in services is often governed by different rules compared to manufacturing (Cainelli et al., 2006), and (ii) firms which engage in formal R&D activities - that is, those reporting the presence of a formal R&D unit, since for those units we have information on employment details (specifically, employment of scientists with a PhD degree) and on exploratory-research activities.

3.2 Method

To test the hypotheses, we formulated the following model:

$$Explor_{it} = \beta_0 + \beta_1 PhD_{it} + x'_{it}\gamma + \varepsilon_{it}$$

where *Explor*denotes an indicator of exploration research activity, PhD indicates the presence of scientists in the firm's R&D function (measured by the proportion of doctoral graduates relative to R&D employees), and x is a vector of the covariates. Definitions of the variables are provided in table A1, Appendix 1.

Dependent variables

To construct our dependent variables, we use the percentage of R&D investment that is devoted to firms' exploration activities (in logs). PITEC data allow us to quantify the firm's engagement in the different components of R&D, in a way that is consistent with established measurement of innovation activity (OECD, 2005), differentiating between basic, applied and development investments. According to PITEC guidelines, basic research relates to experimental or theoretical work undertaken to obtain new knowledge on the foundation of phenomena or observable facts, without any particular application in view; while applied research consists of original research work undertaken to obtain new knowledge aimed at a particular objective. In contrast, development is defined as systematic work based on existing knowledge, derived from research or practical experience, that is directed to the production of new materials, products or devices, or to installing new processes, systems and services or to improving those already existing or installed.

Based on these definitions, we construct our dependent variables capturing exploration as measured by the firm's investment of resources in the *basic* and *applied* components of R&D activity (as opposed to the *development* aspect). This construct is based on the fact that (basic and applied) research and development are different activities whose purpose, knowledge bases, personnel requirements and management style have been shown to differ (Barge-Gil and López, 2014, 2015). More precisely, while these activities are associated with the introduction of new or improved products and processes, basic and applied research are directed to the production and acquisition of new knowledge while development refers to the exploitation of the available knowledge (OECD, 2005). Since both applied and basic research are associated with knowledge creation and search processes, we understand basic and applied research as exploration-related activity, and development as capturing exploitation-related activity.[5]

[5] We conducted some case studies in Spanish firms and found that the way the question about basic research was framed was confusing. Some respondents understood 'basic' research as 'easy' or 'plain' research rather than fundamental research. Therefore, in these data the distinction between basic and applied research is not reliable.

Second, we measure transition using three different indicators. The first (*Start_exploration*) measures the shift towards an exploratorion strategy in firms that are R&D active but have not engaged in exploratory research in a given period t. We constructed a dummy variable that takes the value 1 if the firm performs exploration activities in t+1 and zero if the firm is not engaged in exploration activities in t+1. The measure (*Hold_exploration*) captures continuous vs. occasional research activity such that, for those firms that perform exploration activities in period t, is defined as a dummy variable that takes the value 1 if the firm maintains exploration in t+1 and zero if it abandoned its exploration activities. That is, the reverse of this indicator would be the abandonment of research activities. The third measure (*Intensify_exploration*) is a continuous variable which is defined as the log variation between t and t+1 on investment in exploration (normalized by size measured as number of the firm's employees) to capture the escalation in the firms' exploration orientation among those firms already conducting exploration activities.

Explanatory and control variables

Our explanatory variable to measure the availability of corporate scientists (*lsphd*) refers to the proportion of PhD graduates in the firm's R&D unit defined as the number of graduates relative to the number of R&D employees (in log).[6] In a robustness check we used an alternative measure by computing a dummy variable that takes the value 1 if the firm employs at least one PhD graduate and zero otherwise.

Since some characteristics of firms embarking on these transitions may differ from those that do not, we consider a complete set of covariates based on observables from the survey. Were these variables to be excluded from the model, they potentially could be confounding factors and could lead to biased estimation of the role played by doctoral graduates. The covariates included are: firm size (*lsize*), measured by the logarithm of the firm's employee count; firm's R&D intensity (*RD_intensity*), defined by the logarithm of the firm's total R&D expenditure per employee; size of the R&D unit (*lsizeteam*), based on a count of the

[6] In the main analysis, we replace the log of zero with zeros. In a robustness check we used the transformation log(x+1).

full time equivalents working in that unit; sales from new-to-the-market products (*newmer*), defined as the percentage of total sales from new to the market products; exports (*export*) which is a dummy variable that takes the value 1 if the firm sells its products abroad and zero otherwise; *parent* and *joint venture* which are dummy variables that take the value 1 if the firm is the parent in a group or is a joint venture, and zero otherwise; location (*park*) which is a dummy variable that takes the value 1 if the firm is located in a science and technology park and zero otherwise; firm age (*lage*) which we define as the log of the number of years since its birth; and appropriability (*appropriability*) defined following Czarnitzky et al. (2007) as the industry average for responses to the question: "how important are your competitors as a source of information for the innovation process". In this last case the responses were ranked from 1 (high importance) to 4 (no importance). The underlying idea is that appropriability is low in those industries whose firms consider information from competitors to be an important source of information for innovation and high otherwise. We also included two indicators for innovation funding: *pubfun* which is a dummy variable that takes the value 1 if the firm receives public funding and zero otherwise; and *obstacle_funds* which is a dummy variable that takes the value 1 if the firm reports lack of internal or external funds as an obstacle to innovation being moderately or very important. Finally, we include year dummies and the industry dummies in line with the OECD (2011) technological intensity classification.

Descriptive statistics for all the variables in the regressions are reported in table 1, and distinguish between firms that conduct (or not) R&D exploration activities. Tables 2 and 3 provide the correlation matrixes including both the dependent and independent variables, for these two groups of firms.

Table 1. Descriptive statistics

Variable	Not conducting exploration in t (n=5,815)		Already conducting exploration in t (n=11,184)	
	Mean	Std Deviation	Mean	Std Deviation
Start_exploration	0.17	0.37	-	-
Hold_exploration	-	-	80.57	0.40
Intensify_exploration	-	-	-0.873	1.71
Lsphd	0.32	0.97	0.63	1.27
Lsize	4.14	1.28	4.30	1.35
RD_intensity	7.91	1.29	8.01	1.26
Export	0.87	0.34	0.88	0.32
lsizeteam	1.17	1.03	1.43	1.20
Parent	0.07	0.25	0.10	0.30
joint_venture	0.01	0.10	0.01	0.10
Newmer	12.70	24.74	13.56	24.67
Obstacle_funds	0.74	0.44	0.75	0.44
appropriability	2.79	0.16	2.77	0.15
Park	0.03	0.17	0.03	0.18
Lage	3.11	0.66	3.15	0.72
Pubfun	0.49	0.50	0.48	0.50
LMT	0.28	0.45	0.21	0.41
MHT	0.40	0.49	0.38	0.49
HT	0.12	0.12	0.15	0.36

Table 2: Correlation matrix (Firms not conducting exploration in t, n=5,815)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	Start_exploration																
2	Lsphd	0.05*															
3	Lsize	-0.02*	0.00														
4	RD_intensity	-0.01	0.10*	-0.42*													
5	Export	0.01	0.01	0.22*	-0.06*												
6	Lsizeteam	-0.03*	0.10*	0.53*	0.39*	0.13*											
7	Parent	0.01	0.03*	0.14*	0.00	0.05*	0.11*										
8	joint_venture	0.02	0.00	0.04*	0.01	0.02	0.04*	-0.03*									
9	Newmer	0.01	-0.01	-0.05*	0.14*	0.00	0.09*	0.01	0.01								
10	obstacle_funds	0.03*	0.00	-0.18*	0.09*	-0.03*	-0.09*	-0.02	0.01	0.03*							
11	Appropriability	-0.02	-0.07*	0.15*	-0.24*	-0.01	-0.12*	-0.01	-0.03*	-0.03*	-0.02*						
12	Mediumlow	-0.02	-0.05*	0.13*	-0.14*	0.00	-0.07*	0.04*	0.00	0.00	0.01	0.41*					
13	Mediumhigh	-0.01	0.01	-0.07*	0.10*	0.04*	0.06*	-0.05*	-0.01	0.01	0.03*	-0.39*	-0.50*				
14	High	0.01	0.09*	-0.12*	0.25*	-0.05*	0.15*	0.02	0.00	0.03*	0.00	-0.45*	-0.23*	-0.30*			
15	Park	0.02*	0.04*	-0.07*	0.11*	-0.04*	0.07*	-0.02*	-0.02	0.01	0.02*	-0.12*	-0.07*	0.00	0.18*		
16	Lage	-0.04*	-0.04*	0.31*	-0.17*	0.19*	0.10*	0.07*	0.00	-0.06*	-0.06*	0.05*	0.04*	0.02*	-0.15*	-0.12*	
17	Pubfun	0.00	0.07*	0.08*	0.24*	0.02	0.25*	0.01	-0.02	0.06*	0.05*	-0.02	0.02	0.00	0.02	0.06*	-0.00

*p-value<0.05

Table 3: Correlation matrix (Firms already conducting exploration in t, n=11,184)

		1	2	3	4	6	6	7	8	9	10	11	12	13	14	15	16	17
1	Hold_exploration																	
2	Intensify_exploration	0.96*																
3	Lsphd	0.11*	0.10*															
4	Lsize	0.17*	0.17*	0.11*														
5	RD_intensity	0.09*	0.10*	0.17*	-0.28*													
6	Export	0.08*	0.08*	0.06*	0.18*	0.03*												
7	Lsizeteam	0.2*	0.21*	0.21*	0.62*	0.45*	0.16*											
8	Parent	0.06*	0.06*	0.09*	0.21*	0.01	0.07*	0.18*										
9	joint_venture	0.00	0.00	0.00	0.00	0.02*	0.01	0.02*	-0.04*									
10	Newmer	0.01	0.01	-0.01	-0.07*	0.12*	-0.03*	0.06*	-0.02*	0.02*								
11	obstacle_funds	-0.04*	-0.04*	-0.02*	-0.12*	0.02*	-0.03*	-0.07*	-0.01	-0.02*	0.03*							
12	Appropriability	-0.06*	-0.06*	-0.12*	0.06*	-0.27*	-0.01	-0.21*	0.00	-0.02*	-0.04*	0.01						
13	Mediumlow	-0.05*	-0.04*	-0.07*	0.05*	-0.14*	-0.03*	-0.12*	0.00	-0.01	-0.03*	0.01	0.38*					
14	Mediumhigh	0.02*	0.02*	-0.02*	-0.05*	0.08*	0.06*	0.04*	-0.04*	0.01	0.01	-0.02*	-0.36*	-0.41*				
15	High	0.07*	0.07*	0.15*	-0.04*	0.35*	-0.02*	0.27*	-0.01	0.01	0.06*	0.01	-0.49*	-0.22*	-0.33*			
16	Park	0.02*	0.02*	0.06*	0.00	0.13*	0.03*	0.10*	0.03*	0.01	0.03*	0.02*	-0.13*	-0.05*	-0.03*	0.19*		
17	Lage	0.09*	0.08*	0.05*	0.35*	-0.13*	0.16*	0.18*	0.11*	-0.02*	-0.07*	-0.10*	0.02*	0.01	0.02*	-0.07*	-0.09*	
18	Pubfun	0.05*	0.06*	0.10*	0.15*	0.24*	0.07*	0.29*	0.11*	0.01	0.04*	0.03*	-0.02*	-0.02*	-0.02*	0.07*	0.09*	0.01

*p-value<0.05

Instrumental variables

Although we control for a wide range of covariates, not all the relevant confounding factors may be observable. Even if they were observable, there would remain the possibility that the line of causality between doctoral graduate employees and transitions in R&D strategies would remain undetermined. Since we are interested in the line of causality from employment of highly qualified scientists in R&D departments to the adoption of changes in R&D strategies, we need to address the potential problem of endogeneity. We adopt an instrumental variables approach employing two different instruments.

The first instrument is based on the notion that firms face an exogenous supply of PhD graduates. This exogenous supply should influence their hiring but should be uncorrelated to the firm's transition decision. In other words, it should be a source of exogenous variation. We built an indicator for this supply based on new PhD graduates having studied for their degree in the same region in which the firm is located, and in scientific and technological fields relevant to the firm's economic activity. To determine the scientific and technological fields relevant to the firm's economic activity, we used the matrix provided by Cohen et al. (2002) which links scientific fields to economic industries, and follow the methodology employed in Abramovsky et al. (2007). To build our instrument we exploited University System Statistics provided by the Spanish Institute of Statistics (INE). We matched the supply of PhD graduates to the different manufacturing industries and locations in order to develop an indicator of the firm-specific supply of PhDs (details on the construction of this indicator are provided in Appendix 1).[7]

The second instrument is the industry average of the potentially endogenous variable: that is, the share of doctoral graduate employees in the R&D unit. This type of instrument has been used widely in research based on CIS data (see, e.g., Cassiman and Veugelers, 2002; Veugelers and Cassiman, 2005). The underlying idea is that having controlled for the

[7] In a robustness check we analyse the change in results when we use a different matrix to match scientific fields and economic industries, developed by ourselves and based on data from the PhD Survey (INE).

covariates, the industry average picks up the effect of industry specific attributions uncorrelated to firm specific omitted factors (Veugelers and Cassiman, 2005; Barge-Gil and Conti, 2013).

Estimation method

We start by providing the results in a linear framework. We present the result of the ordinary least squares (OLS) regressions, and perform an analysis of coefficient stability (Oster, 2016). This latter analysis provides a lower bound for the coefficient if the bias due to omitted variables is proportional to the change in the coefficient observed when the covariates are introduced (the method is described in Appendix 2). We then provide two instrumental variables (IV) estimations. The first is a standard IV estimation using two-stage least squares (2SLS), and the second takes account of the fact that our potentially endogenous variable (share of PhDs) is bounded at the extremes of its distribution (i.e. it is a limited variable). Following Wooldridge (2002) we estimate a Tobit model with share of PhD graduates as the dependent variable, and use the values predicted by the Tobit model as the instrument for the IV estimation.

In the two first transitions the dependent variable is binary, which does not favor the use of a linear model since it provides constant marginal effects which are unrealistic as the probability ranges between zero and one. We therefore chose a probit model. First, we provide results taking no account of potential endogeneity of the share of PhD graduates, and then we run two IV estimations: the standard IV estimation, and an IV estimation using the values predicted by the Tobit model.

4. Results

4.1. Transitions and shares of PhDs: basic trends

Table 4 shows that some 17% of the firms in our sample with no exploration-research activities in a particular year (t), invested in exploratory research in the following year (t+1). This proportion ranges from a low value of 16% in low to medium technology (LMT) industries to 18.3% in high tech (HT) industries. A much higher proportion of firms show persistence in exploration activities in t+1 having been already active in exploration research in a particular year t. Specifically, some 81% of firms persisted in their exploratory research activities in two successive years, with the remaining 19% abandoning these investments. Persisting with exploration activities is more frequent in HT industries (87.5%) compared to LT or LMT firms (77%-78%).

Table 4. Proportion (%) of firms starting, holding and increasing their exploration research activities by type of industry

		LT	LMT	MHT	HT	Total
Not conducting exploration in t (n=5,815)	% of firms starting exploration in t+1	18.19%	15.77%	16.57%	18.31%	16.89%
	Share of PhDs	2.57%	2.24%	3.31%	5.08%	3.07%
	Number of firms	1,182	1,604	2,330	699	5,815
Already conducting exploration in t (n=11,184)	% of firms holding exploration in t+1	77.72%	77.05%	81.66%	87.46%	80.57%
	% of firms increasing exploration in t+1	13.51%	13.44%	14.86%	15.16%	14.26%
	Share of PhDs	5.21%	4.33%	5.69%	9.19%	5.81%
	Number of firms	2,819	2,366	4,301	1,698	11,184

LT: Low Tech Industries, LMT: Low-Medium Tech Industries, MHT: Medium High Tech Industries, HT: High Tech Industries. We follow OECD classification (OECD, 2011)

Finally, comparing the proportion of doctoral graduates in R&D employees we find that the share is substantially larger among firms that are exploration-active compared to those that conduct formal R&D but are not involved in exploratory research - 5.8% versus 3%, respectively over the whole period analyzed. The proportion of PhDs is also higher among HT firms (5% for non-exploration active, and 9% for exploration active) than LMT firms (2% for non-exploration active, and 4% for exploration active firms).

The following sub-sections examine the relationship between the proportion of doctoral graduates in R&D units and firms' R&D exploration strategies more systematically. First of all, we analyze the relationship between firms' shares of PhDs and overall levels of investment in exploration-related activities. Second, we delve into this relationship by analyzing the relationship between the proportion of doctoral graduates on each of the transitions to exploration strategies.

4.2. Exploration investments and PhDs

There are 17,699 (firm-year) observations corresponding to companies that conduct R&D. These firms have an average of 4.82% of PhD holders in their R&D units and devote 45.2% of their R&D investment to exploration activities.

The results of the different models are provided in table 5. Column (1) shows the results of the OLS regressions including only year dummies as covariates. Column (2) includes the full set of the covariates in table 2. Columns (3) shows random effects estimates while column (4) shows fixed effects estimates. Columns (5) and (6) present the standard IV results respectively in a linear framework and using the Wooldridge approach.

Column (1) shows that the variable measuring corporate scientists (lsphd) has a positive and statistically significant coefficient of 0.214. That is, a 1% increase in the share of PhD graduates is associated with an increase of 0.21% in orientation to exploration. Column (2) includes all the covariates and provides no evidence of omitted variables bias since the coefficient remains fairly stable, being 0.200. As a consequence, the Oster's lower bound is quite close to the actual value of the coefficient (0.191).

Column (3) and (4) take advantage of the panel structure and estimate using random and fixed effects respectively. The coefficients decrease quite a lot being 0.087 and 0.038, still significant at 0.05 level. Our explanation for such large decrease is that these estimations suffer from attenuation bias due to measurement error. The reason is that data passed through an anonymization process[8] that introduces noise that is specially harmful when within variation is used.

Column (5) presents the result for standard IV. The two instruments are jointly significant (F-statistic=182.7), and clearly satisfy the exclusion restriction (Hansen's J Chi-square=0.038, p-value=0.845). Endogeneity is clearly rejected (Chi-square=148.1, p-value=0.000). The coefficient of 1.06 is much higher than the OLS marginal effects, meaning that a 1% increase in the share of PhDs is associated with a 1.06% of increase in exploration orientation. Our finding of much larger coefficients for the IV compared to the OLS estimations is in line with the literature (see, e.g., Trostel et al., 2002). As we said, working with anonymized data may generate measurement errors in the PhD variable that could result in 'attenuation bias' (although less important than attenuation bias from fixed effects regressions). Other potential explanations include the existence of a heterogeneous treatment effect (IV estimates provide an estimate of the local average treatment effect which may differ from the average treatment effect). In addition, in our model the endogenous regressor is a limited dependent variable (the share of PhDs ranges between 0 and 100). To address this potential problem we apply the 'Wooldridge' approach to IV estimation with a limited endogenous regressor. The results are reported in column (6): the coefficient then becomes 0.87, meaning that a 1% increase in the share of PhDs is associated with an increase of 0.87% increase in research orientation. This is still high but is closer to the OLS estimations.[9]

To summarize the set of models shows the existence a positive and significant relationship between the share of PhDs and orientation to exploration. We suspect that OLS and, specially, fixed effects estimates suffer from attenuation bias. Actually, coefficients from

[8] Details from the anonymization process can be found in López (2011). He concludes that the anonymization process is not problematic when between variation is used: "Effect of microaggregation on regression results: An application to Spanish innovation data," The Empirical Economics Letters, 10(12), 1265-1272, 2011

[9] Wald test of exogeneity: Chi-square(1)=98.5, p-value=0.000).

IV estimates are much larger. Instruments work very well both in satisfying inclusion and exclusion restrictions.

Table 5. Exploration orientation and PhDs. Results

	(1) SimpleOLS	(2) Multiple OLS	(3) RE	(4) FE	(5) IV	(6) IV Wool
lsphd	0.214***	0.200***	0.087***	0.038**	1.057***	0.871***
	[0.019]	[0.020]	[0.016]	[0.019]	[0.087]	[0.079]
lsize		-0.103***	-0.063**	-0.104	-0.172***	-0.157***
		[0.039]	[0.031]	[0.076]	[0.043]	[0.042]
RD_intensity		-0.083**	-0.049*	-0.045	-0.179***	-0.158***
		[0.033]	[0.025]	[0.029]	[0.038]	[0.037]
export		0.009	0.007	-0.016	-0.051	-0.038
		[0.078]	[0.067]	[0.090]	[0.086]	[0.083]
lsizeteam		0.209***	0.158***	0.118***	0.174***	0.181***
		[0.045]	[0.035]	[0.042]	[0.050]	[0.048]
parent		0.210**	0.045	-0.052	0.035	0.073
		[0.084]	[0.070]	[0.089]	[0.099]	[0.093]
joint_venture		0.038	-0.027	-0.048	0.022	0.025
		[0.227]	[0.184]	[0.210]	[0.250]	[0.241]
newmer		0.000	0.001	0.001	0.001	0.001
		[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
obstacle_funds		0.065	0.069	0.061	0.081	0.078
		[0.058]	[0.046]	[0.054]	[0.064]	[0.061]
appropriability		-0.414*	-0.042	0.147	-0.183	-0.234
		[0.230]	[0.162]	[0.182]	[0.244]	[0.238]
mediumlow		-0.471***	-0.454***	-0.188	-0.386***	-0.404***
		[0.080]	[0.074]	[0.276]	[0.087]	[0.084]
mediumhigh		-0.423***	-0.322***	-0.158	-0.388***	-0.396***
		[0.085]	[0.075]	[0.254]	[0.093]	[0.090]
high		-0.385***	-0.226**	-0.244	-0.571***	-0.531***
		[0.118]	[0.101]	[0.284]	[0.132]	[0.127]
park		-0.061	-0.175	-0.317	-0.165	-0.142
		[0.148]	[0.140]	[0.212]	[0.181]	[0.168]
lage		0.037	0.031	0.065	0.034	0.035
		[0.042]	[0.040]	[0.162]	[0.047]	[0.045]
pubfun		-0.228***	-0.098***	-0.047	-0.279***	-0.268***
		[0.048]	[0.033]	[0.037]	[0.053]	[0.051]
_cons	2.551***	4.727***	3.319***	2.840***	4.813***	4.796***
	[0.044]	[0.744]	[0.542]	[0.844]	[0.797]	[0.775]
N	17,699	17,699	17,699	17,699	17,699	17,699

Coefficients reported with clustered standard error between brackets. All models include year dummies and sectoral dummies according to OECD classification. ***p-value<0.01; **p-value<0.05, *p-value<0.1

4.3. The analysis of transitions to exploration research strategies

In the previous section we analyze the relationship between corporate scientists and exploration investments. The main purpose of this section is to decompose the different decisions regarding exploration: start exploration (when it was not done), hold exploration (when it was already done) and intensification in exploration orientation (for those already doing it)[10].

4.3.1. Start exploration

There are 5,815 (firm-year) observations corresponding to companies that conduct R&D but do not engage in exploration activities in year t. These firms have an average 3.07% of PhD holders in their R&D units. About 17% of these firm-year observations (i.e. 982 cases) make the transition to exploration activities in t+1. Those firms, on average, have 4.35% of PhDs in the R&D team compared to 2.81% on average among those not making the transition.

The results of the different models are provided in table 6. Column (1) shows the results of the OLS regressions including only year dummies as covariates. Column (2) includes the full set of the covariates in table 2. Columns (3) and (4) present the standard IV results respectively in a linear framework and using the Wooldridge approach. Column (5) presents the results of the probit model with covariates. Column (6) presents the results for the 'standard' instrumental variables probit and column (7) includes the probit IV results using the Wooldridge approach.

Column (1) shows that the variable measuring corporate scientists (lsphd) has a positive and statistically significant coefficient of 0.017. That is, a 1% increase in the share of PhD graduates is associated with an increase of 0.017 probability points in the likelihood of a transition to initiate exploration activity. If we apply this effect to the average probability of a transition (0.169) then a 1% increase in the share of PhDs is associated with a 0.1% increase in this probability. Column (2) includes all the covariates and provides no evidence of omitted

[10] Note that as we are using the panel dimension to define the changes in firms' decisions. Once we do that, we cannot use the within variation to estimate the effects as firms do not start research or abandon it several times in the period analysed.

variables bias since the coefficient remains fairly stable. The observables are mostly non-significant with the exception of *obstacle funds* (i.e. firms experiencing funding constraints) and *lage* (firm age).

It is clear that the Oster's bound in this case is equal to the actual coefficient since there are no signs of omitted variables (the coefficient does not change if the observables are included). Column (3) presents the results for the standard IV. The two instruments are jointly significant (F-statistic=28.3), and satisfy the exclusion restriction (Hansen's J Chi-square=2.1, p-value=0.147). Endogeneity is clearly rejected (Chi-square=23.29, p-value=0.000). The coefficient of 0.133 is much higher than the OLS marginal effects, meaning that a 1% increase in the share of PhDs is related to an increase of 0.133 points for the probability of a transition to exploration. Again, taking the average probability of a transition as the reference, the results of the IV regressions mean that a 1% increase in the share of PhDs is associated with an increase of 0.79% in the likelihood of making the transition.

Table 6. Start exploration. Results

	(1) SimpleOLS	(2) Multiple OLS	(3) IV-GMM	(4) IV-GMM Wool	(5) Probit	(6) IV Probit	(7) IV Probit Wool
lsphd	0.017***	0.017***	0.133***	0.080***	0.064***	0.460***	0.279***
	[0.006]	[0.006]	[0.032]	[0.028]	[0.021]	[0.090]	[0.086]
lsize		-0.005	-0.004	-0.004	-0.020	-0.013	-0.016
		[0.008]	[0.009]	[0.009]	[0.033]	[0.032]	[0.033]
RD_intensity		-0.009	-0.013	-0.011	-0.035	-0.043	-0.040
		[0.007]	[0.008]	[0.008]	[0.029]	[0.028]	[0.029]
export		0.018	0.013	0.015	0.072	0.048	0.060
		[0.016]	[0.018]	[0.017]	[0.067]	[0.067]	[0.067]
lsizeteam		-0.002	-0.009	-0.006	-0.012	-0.038	-0.027
		[0.010]	[0.011]	[0.010]	[0.039]	[0.039]	[0.039]
parent		0.025	0.013	0.018	0.099	0.046	0.070
		[0.021]	[0.024]	[0.022]	[0.081]	[0.085]	[0.083]
joint_venture		0.071	0.064	0.069	0.256	0.210	0.238
		[0.055]	[0.060]	[0.057]	[0.179]	[0.186]	[0.182]
newmer		0.000	0.000	0.000	0.001	0.001	0.001
		[0.000]	[0.000]	[0.000]	[0.001]	[0.001]	[0.001]
obstacle_funds		0.025**	0.026**	0.025**	0.102**	0.099**	0.102**
		[0.012]	[0.013]	[0.012]	[0.050]	[0.050]	[0.050]
appropriability		-0.070	-0.061	-0.068	-0.301	-0.250	-0.277
		[0.045]	[0.047]	[0.045]	[0.187]	[0.182]	[0.185]
mediumlow		-0.024	-0.021	-0.022	-0.092	-0.072	-0.084
		[0.017]	[0.018]	[0.017]	[0.065]	[0.064]	[0.065]
mediumhigh		-0.028	-0.031*	-0.030*	-0.115	-0.120*	-0.119*
		[0.018]	[0.019]	[0.018]	[0.070]	[0.068]	[0.070]
high		-0.027	-0.042	-0.038	-0.109	-0.177*	-0.149
		[0.026]	[0.028]	[0.027]	[0.100]	[0.101]	[0.101]
park		0.043	0.026	0.038	0.156	0.116	0.137
		[0.038]	[0.042]	[0.039]	[0.131]	[0.139]	[0.134]
lage		-0.019**	-0.014	-0.016*	-0.074**	-0.052	-0.063*
		[0.009]	[0.010]	[0.009]	[0.035]	[0.035]	[0.035]
pubfun		0.001	-0.009	-0.004	0.005	-0.030	-0.014
		[0.011]	[0.012]	[0.011]	[0.044]	[0.043]	[0.044]
_cons	0.170***	0.502***	0.467***	0.487***	0.395	0.242	0.315
	[0.014]	[0.155]	[0.161]	[0.156]	[0.630]	[0.610]	[0.623]
N	5,815	5,815	5,815	5,815	5,815	5,815	5,815

Coefficients reported with clustered standard error between brackets. All models include year dummies and sectoral dummies according to OECD classification. ***p-value<0.01; **p-value<0.05, *p-value<0.1

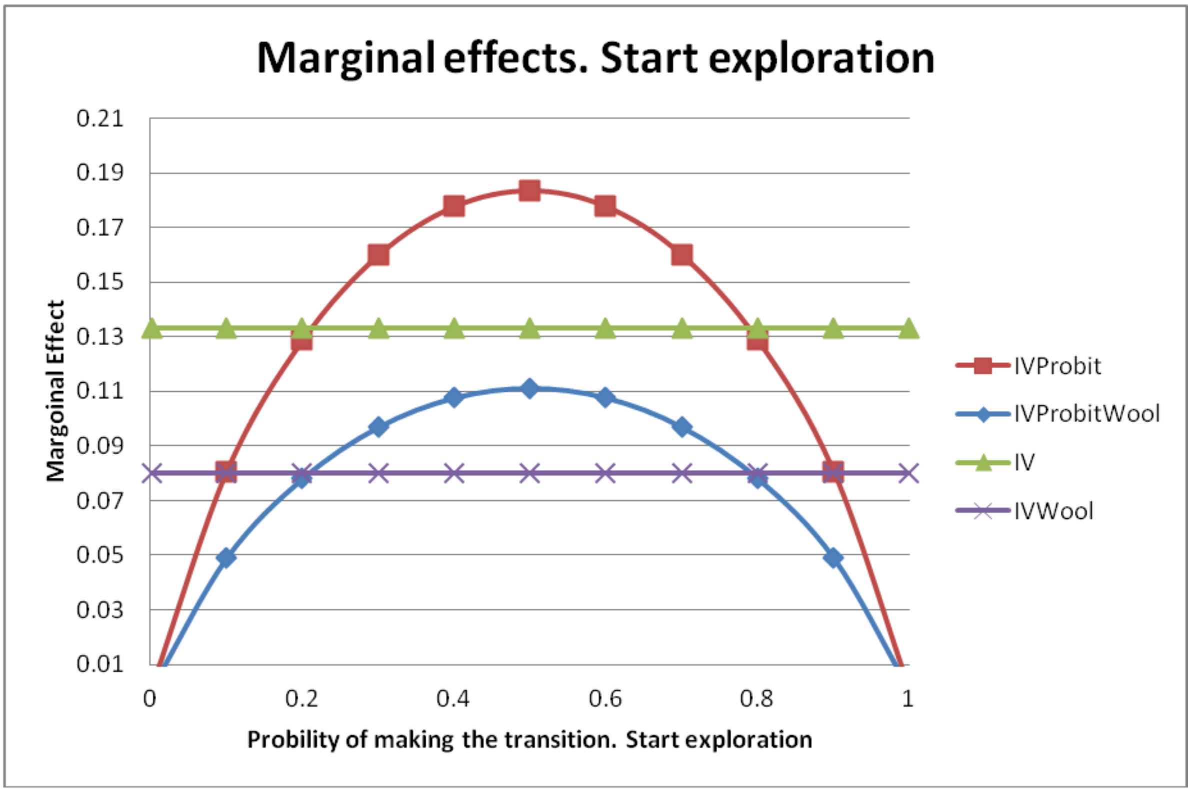
Again, results are much larger when using IV. Results from the 'Wooldridge' approach acknowledging the limited nature of our endogeneous regressor are provided in column (4): the coefficient then becomes 0.08, meaning that a 1% increase in the share of PhDs is associated with an increase of 0.08 points (a change of 0.47% in the probability if the average probability is used as the reference). This is still high but is closer to the OLS estimations.[11]

To summarize, the set of linear models show a positive and significant effect of the share of PhDs on the likelihood of a transition to enact an exploratory research strategy. However, the dependent variable is binary so the marginal effects, by definition, should be non-linear. To analyze these effects, we employed probit models (table 5 columns (5)-(7)). Column (5) provides the results for the probit model assuming exogeneity of PhDs. The marginal effects at the means are very similar to the OLS coefficients. Columns (6) and (7) provide the for the IV approach. Again, the marginal effects at the means are close to (but a bit lower) than the OLS coefficients. The main advantage provided by a probit model is that it allows us to plot the full set of marginal effects.

Figure 1 plots the range of marginal effects for the models in columns (3), (4), (6) and (7) for firms with different values for the probability of a transition.[12] We observe that the maximum marginal effect is 0.18 according to the standard IV probit model (1% increase in the share of PhDs increases the probability of transition by 0.18 probability points), and 0.11 according to the alternative IV probit model (1% increase in the share of PhDs increases the probability of transition by 0.11 probability points).

[11] Wald test of exogeneity: Chi-square(1)=8.66, p-value=0.003).
[12] Note that we plot the marginal effects according to the probability of making the transition, instead of reporting them according to one specific explanatory variable and choosing some specific values for the rest of covariates.

Figure 1. Marginal effects. Start exploration



4.3.2. Hold exploration

There are 11,184 observations of firms conducting R&D exploration activities. They have an average of 5.81% of PhD graduates in their R&D unit employment. 80.57% of firms invest in exploration activities in t+1, with an average of 6.28% of PhDs in their R&D units, higher than the 3.88% of those abandoning their exploration activities. We interpret this as transition in the sense that firms display a persistent commitment to exploration activities by maintaining them in two subsequent periods, as opposed to abandoning this activity.

Table 7 presents the results of the models. Column (1) presents the OLS results using only year dummies as covariates. Column (2) presents the results for the OLS estimates that include the full set of the covariates. Column (3) presents the standard IV results in a linear framework and column (4) presents the same using the Wooldridge approach. Column (5) presents the results of the probit model with covariates and column (6) presents the 'standard'

IV probit model. Finally, column (7) presents the results for the probit IV using the Wooldridge approach.

In column (1) the coefficient is positive and statistically significant at 0.033. Column (2) includes all the covariates and the coefficient drops to 0.017 suggesting that in this case, there is omitted variables bias. This is because, in contrast to the effect on the decision to start exploration, many of the covariates have an effect on the decision to continue exploration activity. Larger firms, R&D intensive firms, exporters, HT firms and older firms are more likely to maintain exploratory R&D. Accordingly, we apply the Oster method to estimate the lower bound of this effect which is 0.011. If we calculate the δ required for the marginal effect to be zero this is 2.758 well above 1, suggesting that despite omitted variables bias the true effect is greater than zero.

Table 7. Hold exploration. Results

	(1) SimpleOLS	(2) Multiple OLS	(3) IV-GMM	(4) IV-GMM Wool	(5) Probit	(6) IV Probit	(7) IV Probit Wool
lsphd	0.033***	0.017***	0.065***	0.030***	0.079***	0.339***	0.383***
	[0.003]	[0.003]	[0.010]	[0.009]	[0.014]	[0.042]	[0.063]
lsize		0.051***	0.045***	0.049***	0.193***	0.151***	0.149***
		[0.007]	[0.007]	[0.007]	[0.026]	[0.027]	[0.028]
RD_intensity		0.037***	0.028***	0.034***	0.131***	0.082***	0.079***
		[0.006]	[0.007]	[0.006]	[0.023]	[0.024]	[0.026]
export		0.045***	0.040***	0.044***	0.125***	0.091*	0.081*
		[0.015]	[0.015]	[0.015]	[0.046]	[0.047]	[0.047]
lsizeteam		0.004	0.005	0.005	0.065**	0.064**	0.067**
		[0.008]	[0.008]	[0.008]	[0.031]	[0.031]	[0.031]
parent		0.019*	0.009	0.016	0.096*	0.037	0.038
		[0.011]	[0.012]	[0.012]	[0.056]	[0.059]	[0.060]
joint_venture		-0.009	-0.009	-0.009	-0.057	-0.062	-0.067
		[0.035]	[0.035]	[0.035]	[0.134]	[0.132]	[0.133]
newmer		0.000	0.000	0.000	0.000	0.000	0.000
		[0.000]	[0.000]	[0.000]	[0.001]	[0.001]	[0.001]
obstacle_funds		-0.013	-0.011	-0.012	-0.046	-0.036	-0.034
		[0.009]	[0.009]	[0.009]	[0.036]	[0.037]	[0.037]
appropriability		-0.003	0.020	0.003	0.022	0.103	0.107
		[0.037]	[0.039]	[0.037]	[0.154]	[0.152]	[0.154]
mediumlow		-0.011	-0.006	-0.010	-0.041	-0.012	-0.008
		[0.013]	[0.013]	[0.013]	[0.045]	[0.045]	[0.046]
mediumhigh		0.022*	0.027*	0.023*	0.081	0.097*	0.091*
		[0.013]	[0.014]	[0.013]	[0.051]	[0.051]	[0.052]
high		0.043**	0.033*	0.040**	0.217***	0.184**	0.187**
		[0.018]	[0.019]	[0.018]	[0.076]	[0.077]	[0.078]
park		-0.009	-0.015	-0.010	-0.042	-0.086	-0.081
		[0.022]	[0.023]	[0.022]	[0.096]	[0.101]	[0.103]
lage		0.017***	0.016**	0.016***	0.071***	0.065***	0.064***
		[0.006]	[0.006]	[0.006]	[0.024]	[0.024]	[0.025]
pubfun		-0.013	-0.016*	-0.014*	-0.059*	-0.076**	-0.079**
		[0.008]	[0.008]	[0.008]	[0.032]	[0.032]	[0.032]
_cons	0.795***	0.188	0.195	0.191	-1.498***	-1.331***	-1.346***
	[0.010]	[0.122]	[0.127]	[0.123]	[0.501]	[0.496]	[0.503]
N	11,184	11,184	11,184	11,184	11,184	11,184	11,184

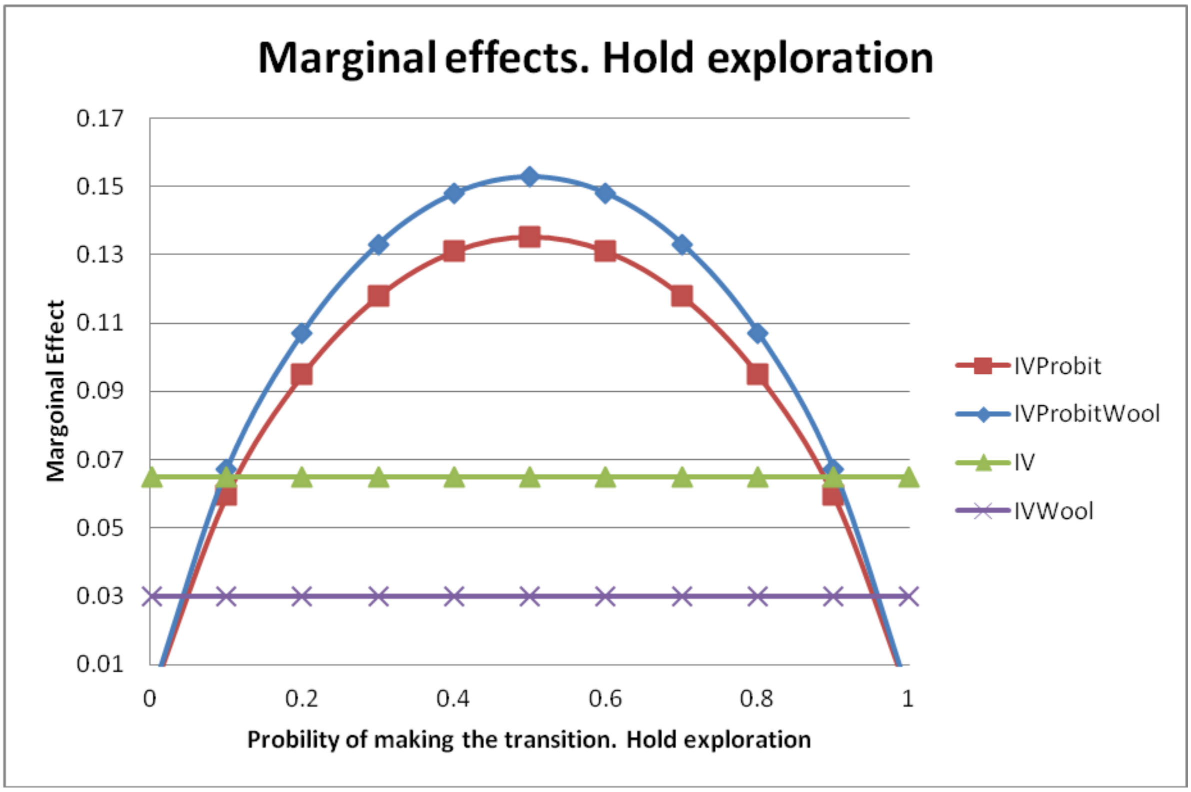
Coefficients reported with clustered standard error between brackets. All models include year dummies and sectoral dummies according to OECD classification. . ***p-value<0.01; **p-value<0.05, *p-value<0.1

Column (3) presents the results for the standard IV. The two instruments are very significant in the first stage (F=156.1), and satisfy the exclusion restriction (Hansen's J Chi-Square=0.678, p-value=0.4102). Exogeneity of PhDs is clearly rejected (Chi-Square=26.58, p-value=0.000). In this case, the coefficient increases less than in the previous analysis. The attenuation bias is now partially compensated for by the omitted variables bias. The IV coefficient is 0.065, meaning that a 1% increase in the share of PhDs results in an increase of 0.065 probability points in the likelihood of continuing with exploration activities. Application of the Wooldridge approach (column (4)) results in the coefficient decreasing to 0.030. Again, this magnitude is closer to the OLS estimation.[13]

To summarize, the set of linear models shows a positive and significant effect of the share of PhDs on the likelihood of maintaining exploration activity. There is some evidence of the existence of omitted variables bias but the lower bound of the coefficient is still higher than zero. In the IV estimations, the coefficients increase suggesting that attenuation bias is higher than omitted variables bias. Again, the dependent variable is binary so that marginal effects should, by definition, be non-linear. To analyze these effects, we employ probit models (columns (5)-(7)). Column (5) provides the probit results assuming exogeneity of PhDs. The marginal effects at the means are slightly higher than the OLS coefficients (0.020 vs. 0.017). Columns (6) and (7) provide the results for the IV approach. Again, the marginal effects at the means are similar to than the OLS coefficient. The main advantage of a probit model is that it allows us to plot the full set of marginal effects. Figure 2 plots the range of marginal effects for the models in columns (3), (4), (6) and (7). The maximum marginal effect is 0.15 for the Wooldridge IV probit, and 0.135 for the standard IV probit.

[13]Exogeneity of PhDs is not rejected in this last model (Chi-Square=2.54, p-value=0.11).

Figure 2. Marginal effects. Hold exploration



4.3.3. Intensify exploration

As in the previous case, we have 11,184 observations corresponding to firms that perform R&D exploration activities. During the period analyzed the weight of exploration in R&D activities decreased by around 8%. For this reason, it is important to examine the factors explaining why certain firms are able to intensify their commitment to exploration by increasing investment in exploration R&D activity.

The results of the models are provided in table 8. Column (1) presents the simple OLS results with only year dummies as covariates. Column (2) presents the multiple OLS results including the full set of the covariates. Column (3) presents the standard IV results in a linear framework, and column (4) presents the same results in a linear framework using the Wooldridge approach. Column (1) shows that there is a positive and statistically significant coefficient of 0.137, meaning that a 1% increase in the share of PhDs is associated to a 0.137% increase in the weight of exploration variation. Column (2) includes all the covariates and the coefficient drops to 0.069, evidence that in this case, there is omitted variable bias. Some of the covariates are significantly related to exploration variation: size, R&D intensity, exporting, belonging to a high tech industry and being an old firm are positively related to exploration variation.

We applied Oster's method to analyze the lower bound of the effect which is 0.043. If we calculate the δ required for the marginal effect to be zero this is 2.595 which is well above 1, suggesting that despite omitted variables bias the true effect is higher than zero. Column (3) presents the standard IV results. The two instruments are jointly significant in the first step (F=156.1), and the exclusion restriction is satisfied (Hansen's J Chi-Square=0.94, p-value=0.33). Exogeneity is clearly rejected (Chi-Square=20.05, p-value=0.000).

In this case the attenuation bias is compensated partially by the omitted variables bias. The estimates in column (3) show that the IV coefficient is 0.247, meaning that a 1% increase in the share of PhD graduates results in a 0.24% increase in the weight of exploration activities. For the Wooldridge approach, this coefficient drops to 0.107 (column (4)). Again, this magnitude is closer to the OLS estimation.[14]

To summarize, the different models all show a positive and significant effect of the share of PhDs on the change in the weight on exploration. It seems that the OLS estimates suffer from omitted variable and attenuation bias. The coefficients derived from the IV methods are larger, suggesting that attenuation bias was greater than omitted variable bias.

[14] The exogeneity of PhDs cannot be rejected according to this last model (Chi-Square=1.187, p-value=0.276).

Table 8. Intensify exploration. Results

	(1) SimpleOLS	(2) Multiple OLS	(3) IV	(4) IVWool
lsphd	0.137***	0.069***	0.247***	0.107***
	[0.012]	[0.013]	[0.043]	[0.037]
lsize		0.229***	0.204***	0.223***
		[0.029]	[0.030]	[0.030]
RD_intensity		0.170***	0.139***	0.163***
		[0.027]	[0.028]	[0.028]
export		0.201***	0.183***	0.197***
		[0.064]	[0.064]	[0.064]
lsizeteam		0.009	0.011	0.010
		[0.034]	[0.034]	[0.034]
parent		0.075	0.039	0.067
		[0.049]	[0.053]	[0.050]
joint_venture		0.004	0.001	0.003
		[0.142]	[0.144]	[0.142]
newmer		0.001	0.001	0.001
		[0.001]	[0.001]	[0.001]
obstacle_funds		-0.046	-0.039	-0.044
		[0.038]	[0.039]	[0.038]
appropriability		0.017	0.100	0.034
		[0.157]	[0.164]	[0.159]
mediumlow		-0.038	-0.020	-0.034
		[0.055]	[0.056]	[0.055]
mediumhigh		0.109*	0.126**	0.112*
		[0.058]	[0.059]	[0.058]
high		0.192**	0.155*	0.184**
		[0.078]	[0.082]	[0.079]
park		-0.018	-0.039	-0.022
		[0.089]	[0.093]	[0.089]
lage		0.066**	0.063**	0.066**
		[0.026]	[0.027]	[0.026]
pubfun		-0.050	-0.059	-0.053
		[0.035]	[0.036]	[0.035]
_cons	-0.854***	-3.675***	-3.651***	-3.669***
	[0.044]	[0.525]	[0.540]	[0.527]
N	11,184	11,184	11,184	11,184

Coefficients reported with clustered standard error between brackets. All models include year dummies and sectoral dummies according to OECD classification. . ***p-value<0.01; **p-value<0.05, *p-value<0.1

4.4. Robustness checks

We performed several robustness checks. First, since the matrix provided by Cohen et al. (2002), and used in Abramovski et al. (2007), is based on US data collected during the 1990s, it may not account adequately for the current link between scientific fields and economic industries in Spain. To address this concern, we build a matrix based on data from the Survey of

Human Resources in Science and Technology, and replicated the analysis using this new matrix. The results were similar and are provided in Appendix 4.

Second, since many firms show no presence of doctoral graduates, we tested the robustness of our results to the inclusion of a dummy for whether the firm employs at least one PhD. The direction and significance of results was very similar to the original results. The robustness check results are provided in Appendix 5.

5. Conclusions

The main aim of this study was to analyze the role of corporate scientists (i.e. PhD employees) in R&D units as a trigger for two critical transitions in firms' R&D strategies: i) initiation of exploration-oriented research activities (*exploration-enacting strategy*), and ii) increased commitment to exploration-oriented research activities (*exploration-deepening strategy*). While previous studies on this topic focus on star scientists and high tech sectors, our analysis is more systematic and includes a wider range of manufacturing industries and all the highly qualified employees in firms' R&D units. We tested our propositions on a large sample of Spanish manufacturing and R&D performing firms for the period 2006-2012.

We found that firms with larger proportions of PhDs in their R&D units were more likely to embark on exploration research activities compared to firms with no or lower proportions of PhD graduates in their R&D function. We found also that the presence of PhDs in R&D units contributes not only to increasing persistence in exploration-oriented research activities but also to increasing the proportion of resources devoted to exploratory research over time.

This study extends existing research in innovation management in three important ways. Conceptually, it provides support for the claim that the role of corporate scientists in R&D units goes beyond the generation and absorption of scientific knowledge. We suggest that highly qualified scientists equip firms with the skills required to make decisions about fundamental changes to the orientation of their research activities. More specifically, we argue that the availability of scientists in the R&D function increases the likelihood that the firm will search beyond its knowledge boundaries, and initiate, maintain and intensify a search strategy oriented to exploratory research.

Empirically, we provide strong evidence for the connection between human capital and the intensification of exploratory research activity. Our analysis included a large set of covariates in order to ensure the robustness of our results to alternative explanatory factors (i.e. appropriability regime, firm age and size, and R&D intensity, among others), and we conducted instrumental variables estimations, building two indicators of the exogenous supply of PhD graduates as instruments. Therefore, the methodology employed makes our findings generally robust to endogeneity concerns.

Finally, the study has some implications for industry practitioners. Our results show that the presence of PhDs has a direct influence on firms' capacities to set in motion, and maintain exploration search strategies. These strategies are critical to achieving sustainable competitive advantage, since firms that neglect exploration activities may constrain their capacity to respond to technological change, and therefore, compromise their long-term survival (Michelino et al., 2005). In this sense, the acquisition of new skills through the enrolment of highly qualified employees in the R&D functions seems crucial to ongoing exploratory activities that contribute to the introduction of new products, and to strengthen the capacity of the firm to continuously adapt the knowledge base to respond to changing customer preferences and novel technologies.

The importance of scientists for exploration activities introduces questions related to the management of highly qualified human resources, and opens new research directions. First, there are issues related to career development opportunities and reward mechanisms, and their effect on scientists' involvement in firms' exploration research activities (Chen et al., 2003; Sauermann and Cohen, 2010; Liu and Stuart, 2014; Balsmeier and Pellens, 2016). Second, we need to know more about the conditions allowing scientists to exploit firm exploration activity to improve innovation performance. Third, further research is required to examine the extent to which corporate scientists enhance the capacity of firms to integrate and exploit external sources of scientific knowledge, in an international context characterized by an increasing reliance on outsourcing, technology markets and an overall decline in firms' engagement to in-house scientific research (Arora et al., 2018). Finally, future research could analyze the heterogeneous effects of PhDs in R&D units by firm type, size, R&D intensity and industry sector.

Acknowledgments

Andrés Barge Gil acknowledgse funding from projects ECO2014-52051-R, S2015/HUM-3417 and ECO2017-82445-R. Pablo D'Este acknowledges funding from project ECO2014-59381-R. Liliana Herrera acknowledges funding from project ECO2015-63880-R.

References

Abramovsky L, Harrison R, Simpson H (2007) University research and the location of business R&D. *Econ. J.* 117: 114–141.

Ahlin L (2013) Implementing an R & D Strategy without Prior R & D - Experience Recruitment as a Source of R & D - related Routines and Capabilitie? *Pap. Innov. Stud.* CIRCLE, Lund University, Sweden.

Al-Laham A, Tzabbar D, Amburgey TL (2011) The dynamics of knowledge stocks and knowledge flows: innovation consequences of recruitment and collaboration in biotech. *Ind. Corp. Chang.* 20(2):555–583.

Antonelli C, Crespi F, Scellato G (2012) Inside innovation persistence: New evidence from Italian micro-data. *Struct. Chang. Econ. Dyn.* 23(4):341–353.

Arora A, Belenzon S, Patacconi A (2018) The decline of science in corporate R&D. *Strategic Management Journal* 39 (1): 3-32.

Balsmeier B, Pellens M (2016) How much does it cost to be a scientist? *J. Technol. Transf.* 41(3):469–505.

Barge-Gil A, Conti A (2013) Firm R&D Units and Outsourcing Partners: A Matching Story. *SSRN Electron. J.*

Barge-Gil A, López A (2014) R&D determinants: Accounting for the differences between research and development. *Res. Policy* 43(9):1634–1648.

Barge-Gil A, López A (2015) R versus D: Estimating the differentiated effect of research and development on innovation results. *Ind. Corp. Chang.* 24(1):93–129.

Cainelli G, Evangelista R, Savona M (2006) Innovation and economic performance in services: A firm-level analysis. *Cambridge J. Econ.* 30(3):435–458.

Cassiman B, Veugelers R (2002) R&D cooperation and spillovers: Some empirical evidence from Belgium. *Am. Econ. Rev.* 92(4):1169–1184.

Chen TY, Chang PL, Yeh CW (2003) The study of career needs, career development programmes and job satisfaction levels of R&D personnel: the case of Taiwan. *Int. J. Hum. Resour. Manag.* 14(6):1001–1026.

Cohen WM, Nelson RR, Walsh JP (2002) Links and Impacts: The Influence of Public Research on Industrial R&D. *Manage. Sci.* 48(1):1–23.

Czarnitzki D, Ebersberger B, Fier A (2007) The relationship between R&D collaboration, subsidies and R&D performance: Empirical evidence from Finland and Germany. *J. Appl. Econom.* 22(7): 1347–1366.

D'Este P, Amara N, Olmos-Peñuela J (2016) Fostering novelty while reducing failure: Balancing the twin challenges of product innovation. *Technol. Forecast. Soc. Change* 113:280–292.

Deeds DL, DeCarolis D, Coombs J (2000) Dynamic capabilities and new product development in high technology ventures: An empirical analysis of new biotechnology firms. *J. Bus. Ventur.* 15(3):211–229.

Ding WW (2011) The Impact of Founders' Professional-Education Background on the Adoption of Open Science by For-Profit Biotechnology Firms. *Manage. Sci.* 57(2):257–273.

Felin T, Foss NJ (2005) Strategic organization: a field in search of micro-foundations. *Strateg. Organ.* 3(4):441–455.

Garcia R, Calantone R, Levine R (2003) The Role of Knowledge in Resource Allocation to Exploration versus Exploitation in Technologically Oriented Organizations. *Decis. Sci.* 34(2):323–349.

García-Quevedo, J., Mas-Verdu, F., Polo-Otero, J. (2012). Which firms want PhDs? An analysis of the determinants of the demand. *Higher Education*, 63 (5), 607-620.

Gittelman M, Kogut B (2003) Does Good Science Lead to Valuable Knowledge? Biotechnology Firms and the Evolutionary Logic of Citation Patterns. *Manage. Sci.* 49(4):366–382.

Grigoriou K, Rothaermel FT (2014) Structural Microfoundations of Innovation. *J. Manage.* 40(2):586–615.

Groysberg B, Lee LE, Nanda A (2008) Can They Take It With Them? The Portability of Star Knowledge Workers' Performance. *Manage. Sci.* 54(7):1213–1230.

Herrmann AM, Peine A (2011) When “national innovation system” meet “varieties of capitalism” arguments on labour qualifications: On the skill types and scientific knowledge needed for radical and incremental product innovations. *Res. Policy* 40(5):687–701.

Herrera L, Muñoz-Doyague MF, Nieto M (2010) Mobility of public researchers, scientific knowledge transfer, and the firm’s innovation process. *J. Bus. Res.* 63(5):510–518.

Herrera L, Nieto M (2015) The determinants of firms’ PhD recruitment to undertake R&D activities. *Eur. Manag. J.* 33(2):132–142.

Hess A, Rothaermel FT (2012) Intellectual Human Capital and the Emergence of Biotechnology: Trends and Patterns, 1974-2006. *IEEE Trans. Eng. Manag.* 59(1):65–76.

Kim N, Kim E (2015) Board capital and exploration: from a resource provisional perspective. *Manag. Decis.* 53(9):2156–2174.

Kogut B, Zander U (1992) Knowledge of the firm, combinative capabilities, and the replication of technology. *Organ. Sci.* 3(3):383–397.

Lacetera N, Cockburn IM, Henderson R (2004) Do firms change capabilities by hiring new people? A study of the adoption of science-based drug discovery. *Adv. Strateg. Manag.* 21:133–159.

Laursen K, Salter A (2006) Open for innovation: The role of openness in explaining innovation performance among U.K. manufacturing firms. *Strateg. Manag. J.* 27(2):131–150.

Lee H fen, Miozzo M, Laredo P (2010) Career patterns and competences of PhDs in science and engineering in the knowledge economy: The case of graduates from a UK research-based university. *Res. Policy* 39(7):869–881.

Levinthal DA, March JG (1993) The myopia of learning. *Strateg. Manag. J.* 14(2 S):95–112.

Liu CC, Stuart T (2014) Positions and rewards: The allocation of resources within a science-based entrepreneurial firm. *Res. Policy* 43(7):1134–1143.

Lowe RA, Veloso FM (2015) Patently wrong? Firm strategy and the decision to disband technological assets. *Eur. Manag. Rev.* 12(2):83–98.

Luo XR, Koput KW, Powell WW (2009) Intellectual capital or signal? The effects of scientists on alliance formation in knowledge-intensive industries. *Res. Policy* 38(8):1313–1325.

Mañez JA, Rochina-Barrachina ME, Sanchis-Llopis A, Sanchis-Llopis JA (2015) The determinants of R&D persistence in SMEs. *Small Bus. Econ.* 44(3):505–528.

March JG (1991) Exploration and Exploitation in Organizational Learning. *Organ. Sci.* 2(1):71–87.

Mairesse J, Mohnen P (2010) Using innovation surveys for econometric analysis. *Handb. Econ. Innov.* 2(1):1129–1155.

Michelino F, Cammarano A, Lamberti E, Caputo M (2015) Knowledge Domains, Technological Strategies and Open Innovation. *J. Technol. Manag. Innov.* 10(2):50–78.

Mom TJM, van Neerijnen P, Reinmoeller P, Verwaal E (2015) Relational Capital and Individual Exploration: Unravelling the Influence of Goal Alignment and Knowledge Acquisition. *Organ. Stud.* 36(6):809–829.

Nelson RR, Winter SG (1982) *An Evolutionary Theory of Economic Change*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press.

Nonaka I (1994) A dynamic theory of organizational knowledge creation. *Organ. Sci.* 1(5):14–37.

OECD, Eurostat (2005) *Oslo Manual Guidelines for Collecting and Interpreting Innovation Data*.

OECD (2011) ISIC REV. 3 *Technology Intensity Definition*. OECD Directorate for Science, Technology and Industry. https://www.oecd.org/sti/ind/48350231.pd

Oster E (2016) Unobservable Selection and Coefficient Stability: Theory and Evidence. *J. Bus. Econ. Stat.*: Forthcoming.

Peters B (2009) Persistence of innovation: stylised facts and panel data evidence. *J. Technol. Transf.* 34(2):226–243.

Rothaermel FT, Hess AM (2007) Building Dynamic Capabilities: Innovation Driven by Individual-, Firm-, and Network-Level Effects. *Organ. Sci.* 18(6):898–921.

Sauermann H, Cohen WM (2010) What Makes Them Tick? Employee Motives and Firm Innovation. *Manage. Sci.* 56(12):2134–2153.

Singh J, Agrawal A (2011) Recruiting for ideas: How firms exploit the prior inventions of new hires. *Manage. Sci.* 57(1):129–150.

Song J, Almeida P, Wu G (2003) Learning-by-hiring: When is mobility more likely to facilitate interfirm knowledge transfer? *Manage. Sci.* 49(4):351–365.

Subramanian AM, Lim K, Soh PH (2013) When birds of a feather don't flock together: Different scientists and the roles they play in biotech R&D alliances. *Res. Policy* 42(3):595–612.

Subramaniam M, Youndt MA (2005) The influence of intellectual capital on the types of innovative capabilities. *Acad. Manag. J.* 48(3):450–463.

Swift T (2016) The perilous leap between exploration and exploitation. *Strateg. Manag. J.* 37(8):1688–1698.

Teece DJ, Pisano G, Shuen A (1997) Dynamic capabilities and strategic management. *Strateg. Manag. J.* 18(March):509–533.

Triguero Á, Córcoles D (2013) Understanding innovation: An analysis of persistence for Spanish manufacturing firms. *Res. Policy* 42(2):340–352.

Toole AA, Czarnitzki D (2009) Exploring the Relationship Between Scientist Human Capital and Firm Performance: The Case of Biomedical Academic Entrepreneurs in the SBIR Program. *Manage. Sci.* 55(1):101–114.

Trostel P, Walker I, Woolley P (2002) Estimates of the economic return to schooling for 28 countries. *Labour Econ.* 9(1):1–16.

Tzabbar D (2009) When does Scientist Recruitment Affect Technological Repositioning? *Acad. Manag. J.* 52(5):873–896.

Tzabbar D, Aharonson BS, Amburgey TL (2013) When does tapping external sources of knowledge result in knowledge integration? *Res. Policy* 42(2):481–494.

Veugelers R, Cassiman B (2005) R&D cooperation between firms and universities. Some empirical evidence from Belgian manufacturing. *Int. J. Ind. Organ.* 23(5–6):355–379.

Walrave B, van Oorschot KE, Romme AGL (2015) How to counteract the suppression of exploration in publicly traded corporations. *R D Manag.* 45(5):458–473.

Wooldridge JM (2002) *Econometric Analysis of Cross Section and Panel Data.* MIT press.

Zellner C (2003) The economic effects of basic research: evidence for embodied knowledge transfer via scientists' migration. *Res. Policy* 32(10):1881–1895.

Zucker LG, Darby MR, Brewer MB (1998) Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises. *Am. Econ. Rev.* 88(1):290–306.

APPENDIX 1. DEFINITION OF VARIABLES

Table A1. Definition of variables

Variable	Definition
Start_exploration	Dummy variable that takes the value of 1 if the firms perform exploration activities in t+1 and zero if it does not (only for firms not performing exploration activities in t)
Hold_exploration	Dummy variable that takes the value 1 if the firm maintains exploration in t+1, and zero if they stopped exploration activities (only for firms already performing exploration activities in t)
Intensify_exploration	Log variation between t and t+1 on the investment in exploration, normalized by size in terms of firm employees (only for firms already performing exploration activities in t)
Lsphd	Percentage of PhDs relative to the number of employees in the R&D unit (in logs)
Lsize	Firms' employee count (in logs)
RD_intensity	Firm's total R&D expenditures per employee (in logs)
Export	Dummy variable that takes value of 1 if a firm sells products abroad and zero otherwise
Lsizeteam	Number of full time equivalents working in the R&D unit (in logs)
Parent	Dummy variable that takes value 1 if the firm is the parent inside a group and zero otherwise
joint_venture	Dummy variable that takes value 1 if the firm is a joint venture and zero otherwise
Newmer	Percentage of total sales coming from new to the market products
obstacle_funds	Dummy variable that takes value 1 if the firm reports that lack of internal or external funds were an obstacle to innovate of moderate or severe importance and zero otherwise
Appropriability	Industry average of the answer to the following question: "how important are your competitors as a source of information for the innovation process (1-very important, 4-unimportant)"
Park	Dummy variable that takes the value of 1 if the firm is located in a science and technology park and zero otherwise;
Lage	Number of years since birth (in log)
Pubfun	Dummy variable that takes value 1 if the firm received public funding and zero otherwise

APPENDIX 2. CONSTRUCTION OF PhD SUPPLY MEASURE

To build our PhD supply measure we need to determine which scientific fields are relevant to a given industry. We follow Abramovsky et al. (2007) and match scientific fields to industries using data from the 1994 Carnegie Mellon Survey (CMS). This survey asks firms about the importance they attach to ten research fields: biology, chemistry, physics, computer science, material science, medical and health science, chemical engineering, electrical engineering, mechanical engineering and mathematics. As in Abramovsky et al. (2007) we consider a research field to be relevant if more than 50% of the CMS respondents in that industry rank the field moderately or very important.

Data on new PhD graduates by field and university are available at: http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft13%2Fp405&file=inebase&L=0

We compute new PhDs by year, university and scientific field, and match them to each firm based on the firm's region and industry in order to develop a firm-specific measure of PhD supply. More precisely, we compute the number of new PhDs from universities in the same region, and in scientific fields relevant to the firms' economic activity.[15]

Note that for this instrument to satisfy the inclusion restriction we do not need new PhDs to be perfectly immobile. There is a requirement only for some degree of rigidity. Consider two firms (A and B) that are equal in all observable characteristics except firm A is located in a region with a high level of availability of new PhD graduates in relevant scientific fields, and firm B is located in a region with a low level of availability of new PhDs in relevant scientific fields. The inclusion restriction is satisfied since firm A is expected to be more likely than firm B to hire doctoral graduates, due to these firms' distinct locations.

Data from the 2009 Survey of Human Resources for R&D (http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft14%2Fp225&file=inebase&L=0) shows that only 36.9% of PhDs work in another region than their birth region,[16] suggesting that mobility is far from perfect.

[15] The indicator is normalized using the total number of R&D employees in each region.

[16] Ideally, we would know the percentage of PhDs working in the same region as the institution that awarded their degree; however, there are no available public data. Presumably, the figure would be lower than 36.9% because some individuals move region before embarking on their PhD study.

APPENDIX 3. OSTER'S METHOD FOR THE ANALYSIS OF COEFFICIENT STABILITY

The method departs from the key assumption that selection in the unobservables is proportional to selection in the observables. Accordingly, a lower bound of the coefficient can be calculated using the following formula:

$$\beta^* = \tilde{\beta} - \left[\dot{\beta} - \tilde{\beta}\right] \frac{R_{\max} - \tilde{R}}{\tilde{R} - \dot{R}}$$

where :

$\tilde{\beta}$ is the coefficient in the regression with full controls

$\tilde{R}$ is the R^2 in the corresponding regression

$\dot{\beta}$ is the coefficient in the regression without controls (we use just time dummies)

$\dot{R}$ is the R^2 in the corresponding regression

$R_{\max}$ is $1.3\tilde{R}$ (according to Oster's estimation)

Alternatively, the parameter δ can be calculated as the ratio between selection in unobservables and selection in observables required for the coefficient to be zero. If δ>1, selection in unobservables will be higher than selection in observables for the coefficient to be zero.

The method can be implemented using the Stata package: psacalc

APPENDIX 4. ROBUSTNESS CHECK: A DIFFERENT WEIGTHING MATRIX

Table A4a. Start exploration. Results

	(1) SimpleOLS	(2) Multiple OLS	(3) IV-GMM	(4) IV-GMM Wool	(5) Probit	(6) IV Probit	(7) IV Probit Wool
lsphd	0.017***	0.017***	0.121***	0.082***	0.064***	0.388***	0.287***
	[0.006]	[0.006]	[0.032]	[0.028]	[0.021]	[0.108]	[0.086]
lsize		-0.005	-0.004	-0.004	-0.020	-0.014	-0.016
		[0.008]	[0.009]	[0.009]	[0.033]	[0.033]	[0.033]
RD_intensity		-0.009	-0.012	-0.011	-0.035	-0.042	-0.040
		[0.007]	[0.008]	[0.008]	[0.029]	[0.028]	[0.029]
export		0.018	0.014	0.015	0.072	0.053	0.059
		[0.016]	[0.018]	[0.017]	[0.067]	[0.067]	[0.067]
lsizeteam		-0.002	-0.007	-0.007	-0.012	-0.034	-0.028
		[0.010]	[0.010]	[0.010]	[0.039]	[0.039]	[0.039]
parent		0.025	0.012	0.018	0.099	0.056	0.069
		[0.021]	[0.024]	[0.022]	[0.081]	[0.084]	[0.083]
joint_venture		0.071	0.069	0.069	0.256	0.222	0.236
		[0.055]	[0.059]	[0.057]	[0.179]	[0.185]	[0.182]
newmer		0.000	0.000	0.000	0.001	0.001	0.001
		[0.000]	[0.000]	[0.000]	[0.001]	[0.001]	[0.001]
obstacle_funds		0.025**	0.026**	0.025**	0.102**	0.100**	0.102**
		[0.012]	[0.013]	[0.012]	[0.050]	[0.050]	[0.050]
appropriability		-0.070	-0.066	-0.067	-0.301	-0.263	-0.276
		[0.045]	[0.046]	[0.045]	[0.187]	[0.185]	[0.185]
mediumlow		-0.024	-0.022	-0.022	-0.092	-0.077	-0.083
		[0.017]	[0.017]	[0.017]	[0.065]	[0.065]	[0.065]
mediumhigh		-0.028	-0.037**	-0.031*	-0.115	-0.120*	-0.120*
		[0.018]	[0.019]	[0.018]	[0.070]	[0.069]	[0.069]
high		-0.027	-0.043	-0.039	-0.109	-0.169*	-0.150
		[0.026]	[0.028]	[0.027]	[0.100]	[0.102]	[0.101]
park		0.043	0.060	0.038	0.156	0.125	0.136
		[0.038]	[0.042]	[0.039]	[0.131]	[0.136]	[0.134]
lage		-0.019**	-0.015	-0.016*	-0.074**	-0.057	-0.063*
		[0.009]	[0.010]	[0.009]	[0.035]	[0.035]	[0.035]
pubfun		0.001	-0.008	-0.005	0.005	-0.023	-0.015
		[0.011]	[0.012]	[0.011]	[0.044]	[0.044]	[0.044]
_cons	0.170***	0.502***	0.477***	0.486***	0.395	0.276	0.311
	[0.014]	[0.155]	[0.159]	[0.156]	[0.630]	[0.618]	[0.622]
N	5,815	5,815	5,815	5,815	5,815	5,815	5,815

Coefficients reported with clustered standard error between brackets. All models include year dummies and sectoral dummies according to OECD classification.***p-value<0.01, **p-value<0.05, *p-value<0.1

Table A4b. Hold exploration. Results

	(1) SimpleOLS	(2) Multiple OLS	(3) IV-GMM	(4) IV-GMM Wool	(5) Probit	(6) IV Probit	(7) IV Probit Wool
lsphd	0.033***	0.017***	0.113***	0.071***	0.079***	0.410***	0.492***
	[0.003]	[0.003]	[0.016]	[0.014]	[0.014]	[0.047]	[0.076]
lsize		0.051***	0.046***	0.049***	0.193***	0.144***	0.134***
		[0.007]	[0.008]	[0.007]	[0.026]	[0.027]	[0.030]
RD_intensity		0.037***	0.026***	0.031***	0.131***	0.071***	0.060**
		[0.006]	[0.007]	[0.007]	[0.023]	[0.024]	[0.027]
export		0.045***	0.036**	0.041***	0.125***	0.084*	0.062
		[0.015]	[0.016]	[0.016]	[0.046]	[0.048]	[0.050]
lsizeteam		0.004	0.005	0.005	0.065**	0.060*	0.064**
		[0.008]	[0.009]	[0.008]	[0.031]	[0.031]	[0.031]
parent		0.019*	-0.005	0.005	0.096*	0.006	-0.001
		[0.011]	[0.015]	[0.013]	[0.056]	[0.061]	[0.064]
joint_venture		-0.009	-0.018	-0.014	-0.057	-0.080	-0.085
		[0.035]	[0.040]	[0.037]	[0.134]	[0.136]	[0.139]
newmer		0.000	0.000	0.000	0.000	0.000	0.000
		[0.000]	[0.000]	[0.000]	[0.001]	[0.001]	[0.001]
obstacle_funds		-0.013	-0.010	-0.011	-0.046	-0.030	-0.025
		[0.009]	[0.010]	[0.010]	[0.036]	[0.038]	[0.038]
appropriability		-0.003	0.032	0.017	0.022	0.108	0.079
		[0.037]	[0.042]	[0.040]	[0.154]	[0.152]	[0.153]
mediumlow		-0.011	-0.003	-0.007	-0.041	-0.006	0.008
		[0.013]	[0.014]	[0.013]	[0.045]	[0.046]	[0.047]
mediumhigh		0.022*	0.027*	0.025*	0.081	0.096*	0.083
		[0.013]	[0.015]	[0.014]	[0.051]	[0.052]	[0.053]
high		0.043**	0.070***	0.058***	0.217***	0.233***	0.223***
		[0.018]	[0.023]	[0.021]	[0.076]	[0.080]	[0.081]
park		-0.009	-0.021	-0.015	-0.042	-0.063	-0.059
		[0.022]	[0.028]	[0.025]	[0.096]	[0.106]	[0.110]
lage		0.017***	0.016**	0.017**	0.071***	0.060**	0.056**
		[0.006]	[0.007]	[0.007]	[0.024]	[0.025]	[0.025]
pubfun		-0.013	-0.025***	-0.021**	-0.059*	-0.086***	-0.089***
		[0.008]	[0.010]	[0.009]	[0.032]	[0.033]	[0.032]
_cons	0.795***	0.188	0.154	0.159	-1.498***	-1.264**	-1.116**
	[0.010]	[0.122]	[0.137]	[0.130]	[0.501]	[0.497]	[0.512]
N	11,184	11,184	11,184	11,184	11,184	11,184	11,184

Coefficients reported with clustered standard error between brackets. All models include year dummies and sectoral dummies according to OECD classification.***p-value<0.01, **p-value<0.05, *p-value<0.1

Table A4c. Intensify exploration. Results

	(1) SimpleOLS	(2) Multiple OLS	(3) IV	(4) IVWool
lsphd	0.137***	0.069***	0.460***	0.291***
	[0.012]	[0.013]	[0.069]	[0.056]
lsize		0.229***	0.208***	0.222***
		[0.029]	[0.033]	[0.031]
RD_intensity		0.170***	0.127***	0.149***
		[0.027]	[0.030]	[0.029]
export		0.201***	0.167**	0.187***
		[0.064]	[0.070]	[0.068]
lsizeteam		0.009	0.011	0.012
		[0.034]	[0.037]	[0.036]
parent		0.075	-0.024	0.018
		[0.049]	[0.064]	[0.057]
joint_venture		0.004	-0.021	-0.006
		[0.142]	[0.165]	[0.154]
newmer		0.001	0.001	0.001
		[0.001]	[0.001]	[0.001]
obstacle_funds		-0.046	-0.037	-0.041
		[0.038]	[0.044]	[0.042]
appropriability		0.017	0.146	0.086
		[0.157]	[0.177]	[0.168]
mediumlow		-0.038	-0.006	-0.020
		[0.055]	[0.060]	[0.057]
mediumhigh		0.109*	0.127**	0.116*
		[0.058]	[0.064]	[0.060]
high		0.192**	0.316***	0.270***
		[0.078]	[0.096]	[0.090]
park		-0.018	-0.069	-0.045
		[0.089]	[0.111]	[0.101]
lage		0.066**	0.066**	0.067**
		[0.026]	[0.030]	[0.028]
pubfun		-0.050	-0.101**	-0.084**
		[0.035]	[0.041]	[0.039]
_cons	-0.854***	-3.675***	-3.807***	-3.786***
	[0.044]	[0.525]	[0.584]	[0.555]
N	11,184	11,184	11,184	11,184

Coefficients reported with clustered standard error between brackets. All models include year dummies and sectoral dummies according to OECD classification.***p-value<0.01, **p-value<0.05, *p-value<0.1

APPENDIX 5. ROBUSTNESS CHECK: A BINARY PHD INDICATOR

Table A5a. Start exploration. Results

	(1) SimpleOLS	(2) Multiple OLS	(3) IV-GMM	(4) IV-GMM Wool	(5) Probit	(6) IV Probit	(7) IV Probit Wool
phd	0.038**	0.043**	0.418***	0.301***	0.164**	1.452***	1.048***
	[0.018]	[0.018]	[0.102]	[0.088]	[0.067]	[0.280]	[0.263]
lsize		-0.005	-0.007	-0.005	-0.021	-0.019	-0.020
		[0.008]	[0.009]	[0.009]	[0.033]	[0.032]	[0.033]
RD_intensity		-0.009	-0.012	-0.011	-0.034	-0.039	-0.038
		[0.007]	[0.008]	[0.008]	[0.029]	[0.028]	[0.029]
export		0.018	0.017	0.018	0.074	0.064	0.068
		[0.016]	[0.018]	[0.017]	[0.067]	[0.068]	[0.068]
lsizeteam		-0.003	-0.019*	-0.015	-0.016	-0.075*	-0.057
		[0.010]	[0.012]	[0.011]	[0.039]	[0.041]	[0.041]
parent		0.026	0.017	0.019	0.101	0.060	0.073
		[0.021]	[0.024]	[0.023]	[0.081]	[0.084]	[0.082]
joint_venture		0.072	0.067	0.070	0.257	0.218	0.235
		[0.055]	[0.061]	[0.058]	[0.179]	[0.190]	[0.185]
newmer		0.000	0.000	0.000	0.001	0.001	0.001
		[0.000]	[0.000]	[0.000]	[0.001]	[0.001]	[0.001]
obstacle_funds		0.025**	0.026**	0.025**	0.102**	0.099**	0.102**
		[0.012]	[0.013]	[0.013]	[0.050]	[0.050]	[0.050]
appropriability		-0.070	-0.060	-0.066	-0.301	-0.245	-0.266
		[0.045]	[0.048]	[0.046]	[0.187]	[0.184]	[0.185]
mediumlow		-0.025	-0.022	-0.023	-0.093	-0.078	-0.085
		[0.017]	[0.018]	[0.017]	[0.065]	[0.065]	[0.065]
mediumhigh		-0.028	-0.031	-0.031*	-0.115	-0.119*	-0.119*
		[0.018]	[0.019]	[0.019]	[0.070]	[0.069]	[0.070]
high		-0.026	-0.044	-0.042	-0.107	-0.182*	-0.162
		[0.026]	[0.029]	[0.027]	[0.100]	[0.101]	[0.102]
park		0.043	0.027	0.037	0.156	0.113	0.129
		[0.038]	[0.042]	[0.040]	[0.130]	[0.136]	[0.133]
lage		-0.019**	-0.015	-0.017*	-0.075**	-0.056	-0.063*
		[0.009]	[0.010]	[0.010]	[0.035]	[0.036]	[0.035]
pubfun		0.001	-0.012	-0.008	0.005	-0.039	-0.025
		[0.011]	[0.012]	[0.012]	[0.044]	[0.044]	[0.044]
_cons	0.171***	0.504***	0.479***	0.490***	0.403	0.271	0.312
	[0.014]	[0.155]	[0.165]	[0.160]	[0.630]	[0.618]	[0.623]
N	5,815	5,815	5,815	5,815	5,815	5,815	5,815

Coefficients reported with clustered standard error between brackets. All models include year dummies and sectoral dummies according to OECD classification.***p-value<0.01, **p-value<0.05, *p-value<0.1

Table A5b. Hold exploration. Results

	(1) SimpleOLS	(2) Multiple OLS	(3) IV-GMM	(4) IV-GMM Wool	(5) Probit	(6) IV Probit	(7) IV Probit Wool
phd	0.121***	0.046***	0.218***	0.170***	0.250***	1.125***	1.132***
	[0.008]	[0.009]	[0.034]	[0.031]	[0.045]	[0.139]	[0.148]
lsize		0.052***	0.044***	0.046***	0.193***	0.148***	0.150***
		[0.007]	[0.007]	[0.007]	[0.026]	[0.027]	[0.027]
RD_intensity		0.037***	0.030***	0.032***	0.132***	0.088***	0.090***
		[0.006]	[0.006]	[0.006]	[0.023]	[0.024]	[0.024]
export		0.046***	0.042***	0.043***	0.126***	0.097**	0.095**
		[0.015]	[0.015]	[0.015]	[0.046]	[0.046]	[0.046]
lsizeteam		0.001	-0.009	-0.006	0.054*	-0.005	-0.003
		[0.008]	[0.008]	[0.008]	[0.032]	[0.033]	[0.033]
parent		0.019*	0.007	0.010	0.096*	0.028	0.032
		[0.011]	[0.013]	[0.012]	[0.056]	[0.060]	[0.060]
joint_venture		-0.008	-0.005	-0.006	-0.052	-0.037	-0.040
		[0.035]	[0.035]	[0.035]	[0.134]	[0.130]	[0.130]
newmer		0.000	0.000	0.000	0.000	0.001	0.001
		[0.000]	[0.000]	[0.000]	[0.001]	[0.001]	[0.001]
obstacle_funds		-0.013	-0.012	-0.012	-0.047	-0.039	-0.039
		[0.009]	[0.009]	[0.009]	[0.036]	[0.037]	[0.037]
appropriability		-0.003	0.024	0.016	0.026	0.125	0.119
		[0.037]	[0.040]	[0.039]	[0.155]	[0.156]	[0.157]
mediumlow		-0.011	-0.006	-0.008	-0.041	-0.013	-0.012
		[0.013]	[0.013]	[0.013]	[0.045]	[0.045]	[0.045]
mediumhigh		0.022*	0.028**	0.026*	0.083	0.104**	0.101*
		[0.013]	[0.014]	[0.014]	[0.051]	[0.052]	[0.052]
high		0.044**	0.036*	0.038**	0.219***	0.194**	0.197**
		[0.018]	[0.019]	[0.019]	[0.076]	[0.078]	[0.078]
park		-0.009	-0.016	-0.014	-0.042	-0.093	-0.090
		[0.022]	[0.024]	[0.023]	[0.096]	[0.103]	[0.103]
lage		0.016***	0.014**	0.015**	0.071***	0.058**	0.059**
		[0.006]	[0.006]	[0.006]	[0.024]	[0.024]	[0.025]
pubfun		-0.013	-0.015*	-0.015*	-0.058*	-0.073**	-0.074**
		[0.008]	[0.009]	[0.008]	[0.032]	[0.032]	[0.032]
_cons	0.787***	0.187	0.189	0.189	-1.507***	-1.351***	-1.369***
	[0.010]	[0.122]	[0.130]	[0.127]	[0.502]	[0.507]	[0.509]
N	11,184	11,184	11,184	11,184	11,184	11,184	11,184

Coefficients reported with clustered standard error between brackets. All models include year dummies and sectoral dummies according to OECD classification.***p-value<0.01, **p-value<0.05, *p-value<0.1

Table A5c. Intensify exploration. Results

	(1) SimpleOLS	(2) Multiple OLS	(3) IV	(4) IVWool
phd	0.510^{***}	0.182^{***}	0.827^{***}	0.630^{***}
	[0.036]	[0.040]	[0.144]	[0.130]
lsize		0.230^{***}	0.202^{***}	0.211^{***}
		[0.029]	[0.031]	[0.030]
RD_intensity		0.174^{***}	0.144^{***}	0.153^{***}
		[0.027]	[0.028]	[0.028]
export		0.203^{***}	0.188^{***}	0.194^{***}
		[0.064]	[0.064]	[0.064]
lsizeteam		-0.002	-0.042	-0.029
		[0.034]	[0.036]	[0.035]
parent		0.077	0.032	0.045
		[0.049]	[0.054]	[0.052]
joint_venture		0.008	0.019	0.015
		[0.143]	[0.145]	[0.143]
newmer		0.001	0.001	0.001
		[0.001]	[0.001]	[0.001]
obstacle_funds		-0.047	-0.041	-0.043
		[0.038]	[0.040]	[0.039]
appropriability		0.015	0.116	0.083
		[0.157]	[0.167]	[0.164]
mediumlow		-0.039	-0.020	-0.027
		[0.055]	[0.057]	[0.056]
mediumhigh		0.109^{*}	0.132^{**}	0.124^{**}
		[0.058]	[0.060]	[0.059]
high		0.196^{**}	0.164^{**}	0.175^{**}
		[0.078]	[0.083]	[0.081]
park		-0.017	-0.044	-0.036
		[0.089]	[0.095]	[0.092]
lage		0.066^{**}	0.058^{**}	0.061^{**}
		[0.026]	[0.027]	[0.027]
pubfun		-0.049	-0.058	-0.056
		[0.035]	[0.036]	[0.036]
_cons	-0.887^{***}	-3.682^{***}	-3.672^{***}	-3.673^{***}
	[0.045]	[0.526]	[0.549]	[0.539]
N	11,184	11,184	11,184	11,184

Coefficients reported with clustered standard error between brackets. All models include year dummies and sectoral dummies according to OECD classification.***p-value<0.01, **p-value<0.05, *p-value<0.1