INDUSTRY AND INNOVATION
https://doi.org/10.1080/13662716.2020.1817728

Routledge
Taylor & Francis Group

ARTICLE

Check for updates

PhD trained employees and firms' transitions to upstream R&D activities

Andrés Barge-Gil[a], Pablo D'Este[b] and Liliana Herrera[c]

[a]Department of Quantitative Economics, ICAE and GRIPICO, Campus de Somosaguas (28223), Madrid, Spain; [b]INGENIO (CSIC-UPV), Universitat Politècnica de València, Valencia, Spain; [c]Departamento de Dirección y Economía de la Empresa, Área de Organización de Empresa, Universidad de León, León, Spain

ABSTRACT
This paper investigates the relationship between firms' transition towards upstream-R&D activities and the availability of R&D employees with PhD training. Doctoral trained employees have distinct motivations for research: some have stronger preferences for intellectual freedom and autonomy, while others reveal greater aspirations for targeted research and opportunities for development of new products and processes. These contrasting profiles among PhD trained employees lead to ambiguous predictions about whether a greater presence of employees with a doctoral training enhances the capacity of firms to initiate upstream-oriented R&D. We examine this question by studying a large sample of Spanish manufacturing firms which are active in development activities, and investigate the effect of PhD trained R&D employees on the propensity of firms to initiate upstream-oriented R&D. Our results show that a higher proportion of PhDs in R&D functions has a positive and significant influence on the firm's initiation an upstream-oriented R&D strategy.

KEYWORDS
PhD trained employees; upstream R&D; basic and applied research; R&D strategy; firms' transitions

1. Introduction

The management and economics of innovation literatures argue for the importance of the knowledge creation process as a fundamental driver of firms' innovation performance (Nelson and Winter 1982; Nonaka 1994; Teece, Pisano, and Shuen 1997). More recently, investigations at the micro level have highlighted the role of individuals in the generation of ideas and knowledge in organisations (Felin and Foss 2005; Groysberg, Lee, and Nanda 2008). Interest has increased among innovation scholars in whether knowledge workers and, especially, highly trained employees in the research and development (R&D) workforce, help to explain firm innovation and technological advancement (Gittelman and Kogut 2003; Subramanian, Lim, and Soh 2013; Grigoriou and Rothaermel 2014).

The resource-based view of the firm and the human capital perspective propose that highly trained R&D employees, such as those with doctoral research training, are valuable firm resources. Their advanced education and research training provide firms with inimitable tacit knowledge, which can be a source of competitive advantage (Deeds,

CONTACT Andrés Barge-Gil abarge@ccee.ucm.es Department of Quantitative Economics, ICAE and GRIPICO, Campus de Somosaguas (28223), Madrid, Spain

DeCarolis, and Coombs 2000; Subramaniam and Youndt 2005; Luo, Koput, and Powell 2009). While analyses of the role of highly trained employees in firms' innovation activities provide evidence of a link between them and firms' innovation performance (Herrmann and Peine 2011; Hess and Rothaermel 2012; Tzabbar, Aharonson, and Amburgey 2013), few studies examine whether PhD trained employees significantly contribute to change the firm's R&D orientation in terms of resources devoted to upstream R&D activities. This lack of evidence is in sharp contrast with abundant research showing that the potential benefits of firms' upstream-oriented R&D include boosting new product development (Añón Higón 2016) and productivity growth (Griliches 1986; Czarnitzki and Thowarth 2012), facilitating absorption of external knowledge (Rosenberg 1990) and promoting cooperation with new partners based on better ability to signal scientific competences (Cockburn and Henderson 1998; Cassiman, Veugelers, and Arts 2018). Bringing together these complementary streams of research motivates our research question: to what extent does the availability of R&D employees with PhD training contribute to firms' initiation of upstream-oriented R&D?

This research question emerges from on-going, open debate in the innovation management and economics of science literatures, revolving around the degree of heterogeneity in the 'taste for science' among PhD trained employees and its effect on firms' innovation performance. Taste for science refers to preferences for upstream R&D, freedom in choosing research projects, publishing opportunities and interaction with the scientific community (Stern 2004; Roach and Sauermann 2010). On the one hand, this debate highlights that PhD trained individuals working in academy exhibit very different profiles in terms of upstream research approaches compared to similar PhD trained R&D employees working in industry. However, there is a significant disagreement about whether such differences are mainly due to industrial PhD scientists being willing to sacrifice their taste for science in exchange for better wages and employment conditions (Stern 2004), or whether they are a consequence of industrial PhD scientists self-selecting into development-downstream oriented jobs which better match their personal preferences (Roach and Sauermann 2010; Sauermann and Roach 2014) – or a combination of these two arguments.

On the other hand, a complementary position in this debate argues that differences in research profiles among PhD scientists are often more striking within academe or industry settings, rather than between them (Sauermann and Stephan 2013; Kaiser et al. 2018). For instance, a significant proportion of PhD scientists in industry reveal strong preferences for pursuing autonomous research agendas and a desire to tackle intellectual challenges in their research activities, contrasting with other industry PhD scientists whose preferences are more oriented to secure salary, career advancement or contributions to society through downstream R&D activities (Sauermann and Cohen 2010).

This open debate constitutes the background for our research question, since it suggests contrasting expectations on whether firms that have R&D employees with PhD training might be more likely to initiate an R&D strategy towards upstream-oriented R&D activities, or instead, might reinforce their orientation towards downstream R&D activities. We examine these contrasting perspectives on a large sample of Spanish manufacturing and R&D active firms which conduct downstream-development activities. We employ longitudinal data that covers the period 2006 to 2012 and examine the factors that contribute to adoption of a particular R&D strategy, that is, the decision to initiate upstream-oriented R&D.

We assess firms' R&D strategies from an input-based perspective which explicitly captures both upstream-related expenditure (i.e. basic and applied research) and downstream-related expenditure (i.e. development activities). Since basic and applied research are associated to the production and acquisition of new knowledge, we regard firms that conduct these research activities as engaging in upstream-oriented R&D. In contrast, since development is associated to exploiting available knowledge, we consider development and downstream-oriented R&D as interchangeable concepts. This distinction between upstream (basic/applied R&D) and downstream (development R&D) has been used frequently – see, for example, (Arora, Belenzon, and Patacconi 2018; Barge-Gil and López 2014, 2015; D'Este, Marzucchi and Rentocchini 2018; Herrera and Nieto 2015; OECD, 2005; Sauermann and Cohen 2010). Compared to output-based approaches, which consider inventions and patents, for example, and input-based approaches that do not differentiate between upstream and downstream R&D expenditures, we consider that the approach we propose is appropriate to examine the firm's resource allocation choices in relation to its research and innovation orientation. Finally, our analysis includes all manufacturing industries rather than examining highly R&D intensive industries or the biotech sector, which is the focus of most work on the relationship between highly trained employees in R&D functions and the type of R&D strategy.

Our results suggest that, first, there is a high degree of heterogeneity in the profiles of firms with regards to employment of PhD trained individuals and the implementation of upstream-oriented R&D. In fact, a significant proportion of firms that engage in upstream R&D, has no doctoral trained employees, while there are also many firms that do not conduct upstream R&D despite having PhD individuals among their R&D staff. Second, we find that organisations that have a higher proportion of R&D employees with a doctoral training are more likely to shift towards upstream-oriented R&D strategies.

These results are robust to alternative explanatory factors and we address the problem of reverse causality by adopting an Instrumental Variables (IV) approach. We build an indicator of the exogenous supply of PhD trained R&D employees for each firm, using data on doctoral graduates from each university and scientific area, and the correspondence matrix of scientific areas and industries tested in Abramovsky, Harrison, and Simpson (2007). We construct a second instrumental variable based on the number of PhD graduates working in the industry of each of the firms.

This paper contributes to the debate on the benefits of PhD trained employees for firms' R&D activities, focusing on whether they influence the capacity of firms to initiate a strategy towards upstream R&D activities. We provide new evidence on whether and to what extent researchers with a PhD degree, shape firms' R&D activities and facilitate a change in their orientation towards upstream R&D. Moreover, our focus on the education/training of the firm's R&D employees, highlighting the proportion of R&D employees with a doctoral training and emphasising the composition of the R&D employees skill base, constitutes a novel approach that move beyond analysis restricted to examining the impact of a specific type of scientist (e.g., star scientist) or limited to the scale of R&D resources (e.g., number of R&D employees). Finally, we contend that the positive influence towards upstream-oriented R&D of a higher share of PhD degree holders among the firm's R&D staff, is not restricted only to certain industries, but is pervasive among R&D-active firms. That is, the influence of researchers with a doctoral training on the transition to upstream R&D is pervasive among all downstream-oriented R&D firms, regardless of industry or sector. These results have important implications for both R&D management and innovation policy.

2. Conceptual background

2.1. The transition from downstream to upstream-oriented R&D

There is broad agreement in the literature that upstream R&D is a fundamental component of firms' R&D strategies to enhance opportunities for knowledge recombination and exploration (March 1991; Levinthal and March 1993; Bercovitz and Feldman 2008; Walrave, van Oorschot, and Romme 2015; Swift 2016). Exploratory search involves experimentation aimed at identifying novel goals, in a deliberate effort to move away from current organisational routines and knowledge bases (March 1991; Levinthal and March 1993; Katila and Ahuja 2002). While upstream R&D entails high risks and uncertainties (compared to downstream R&D), it is expected to contribute to the achievement of sustainable competitive advantage by building the technological capabilities required to launch radically new products and enhance new product development (Mudambi and Swift 2014).

Despite the potential benefits of adopting an upstream-oriented R&D strategy, initiating basic and applied research for the first time involves at least two challenges. First, the firm needs to move beyond existing R&D routines and integrate new research practices. Since many firms lack connections to the science-base or a favourable organisational learning environment, they can find it hard to embark on upstream R&D activities (García-Quevedo, Mas-Verdu, and Polo-Otero 2012; Tzabbar 2009; Lowe and Veloso 2015). Second, the firm faces a higher risk of failure, since the fixed-costs might not be fully recovered and outcomes are uncertain (Garcia, Calantone, and Levine 2003; D'Este, Amara, and Olmos-Peñuela 2016; Kim and Kim 2015). Therefore, adopting an upstream-oriented R&D strategy can represent a difficult transition.

Given these challenges, R&D active firms tend to prioritise downstream development approaches in order to increase the short term returns from R&D activity, thus favouring exploitative research at the expense of exploratory one (Swift 2016). Also, upstream R&D can be particularly problematic for firms in non-R&D intensive industries (Máñez et al., 2015), which tend to engage only in sporadic formal research activity or to rely on external services provided by specialised R&D partners (Ahlin, Andersson, and Schubert 2013). This limits the capacity to build an internal knowledge base, which is required to support product and process upgrading and, potentially, can threaten the firm's long run survival.

Finally, it is worth pointing out that an important issue in studies of the influence of PhD trained employees on firms' innovation performance and R&D strategies is reverse causality (Lacetera, Cockburn, and Henderson 2004). In work on strategic management, reverse causality occurs because firms' decisions are not random and involve the mobilisation of resources, which can trigger other decision processes. In the case analysed here, the presence of PhD trained employees could lead to changes to the firm's R&D orientation, but, also, a change in R&D orientation could lead to the decision to hire individuals with PhD degrees. Several studies analyse the first direction of causality using an output approach (Sapsalis, van Pottelsberghe de la Potterie, and Navon 2006; Subramanian, Lim, and Soh 2013) or examine the impact of corporate scientists on the creation of R&D alliances (Stuart, Ozdemir, and Ding 2007; Luo, Koput, and Powell 2009; Spithoven and Teirlinck 2010), but do not explore their impact on a more extended set of the firm's strategic decisions. Our work contributes to this line of research by focusing on the influence of PhD trained employees on the initiation of upstream-oriented R&D; our empirical strategies were selected on the basis of this objective.

2.2. PhD trained R&D employees and initiating upstream-oriented R&D activities

In this study, we extend the discussion on the potential roles of highly trained employees, by focusing on whether the adoption of an upstream-oriented R&D strategy is influenced by the availability and proportion of PhD trained employees in the firm's R&D functions. We examine whether having a higher proportion of R&D employees with a doctoral degree is likely to favour the adoption of an upstream-oriented research approach in firms with no previous similar orientation or strategies. Drawing on previous contributions to the economics of science and innovation management literatures, we acknowledge two contrasting perspectives on the examined relationship.

One perspective contends that firms with PhD trained employees among the firm's R&D staff are more likely to initiate upstream R&D activities. This contention draws on the following three reasons. First, firms with a large share of R&D employees with a doctoral degree might be more sensitive to potential benefits of upstream R&D. The role of PhD trained R&D employees in firms might be particularly relevant during the early stages of a research and innovation process, as their training provides them with the knowledge and skills needed to undertake exploratory research activities (Herrera and Nieto 2015). Having experienced PhD training allows a more profound understanding of scientific knowledge and cutting-edge scientific research methods and, as a result, these researchers can entail for firms an advantage to accurately assess implications of upstream R&D activities to business purposes (Zellner 2003). For example, Ding (2011) shows that the presence of PhDs in biotech firms is related strongly to these firms' adoption of an open science policy, which encourages basic scientific research. We argue, similarly, that a higher share of PhD trained employees in the firm's R&D function is likely to increase the firm's appreciation of the specificities and challenges of exploratory research as well as its potential opportunities and benefits.

Second, a firm's focus on PhD trained R&D employees might be linked to the influence that these human resources might exert on the behaviour and perceptions of other employees within R&D activities. The influence of PhD trained employees could be a result of their academic reputation and their productivity as scientists (McMillan and Thomas 2005; Rao, Chandy, and Prabhu 2008; Cockburn and Henderson 1998; Rothaermel and Hess 2007; Hess and Rothaermel 2012) and/or a consequence of being responsible for a large proportion of organisational resources for R&D activities (Kehoe and Tzabbar 2015). Studies show that scientists may stimulate more favourable perceptions on the use of scientific knowledge and research activities among peers (McMillan and Thomas 2005; Ding 2011; Rao, Chandy, and Prabhu 2008), as well as to increase the scientific productivity of their work team (Furukawa and Goto 2006). Consequently, studies recommend the involvement and interaction of employees with a scientific training with other R&D employees in the firm in order to encourage the use of upstream research and improve firms' R&D performance (Almeida, Hohberger, and Parada 2011; Herrmann and Peine 2011).

Third, a high proportion of PhDs in the firm's R&D function is likely to encourage risk-taking and a pro-active learning culture. An organisational climate that favours learning from both failed and successful innovation is particularly important in the context of upstream R&D activities. For instance, a psychologically safe environment for reporting failure, that is, one where R&D employees are not blamed, but rather are encouraged to report and address non-obvious errors and mistakes, is conducive to a climate that promotes experimentation and learning (Edmondson 2011). Firms with availability of R&D employees with doctoral training are likely to be particularly receptive to organisational environments that provide opportunities for learning and encourage exploration and experimentation.

Based on the above arguments, suggesting a positive relationship between the availability of PhD trained employees and initiation of upstream-oriented R&D activities, we hypothesise that:

H1: *A higher proportion of PhD trained employees among R&D staff increases the likelihood that a downstream-oriented firm will initiate upstream-oriented R&D activities.*

The contrasting perspective builds on the notion that the presence of PhD trained R&D employees does not necessarily result in the initiation of upstream-oriented R&D. Three claims support this perspective. First, academic and industrial science respond to distinct institutional logics, as pointed out in research from the economics of science (Dasgupta and David 1994; Gans and Stern 2010). Firms prioritise the appropriation of the financial returns from knowledge generation processes, which leads them to limit information disclosure and sharing research results with the broader scientific community. According to this perspective, industrial R&D settings are expected to provide a hostile environment for PhD trained individuals willing to disseminate research findings and participate in open science regimes. Therefore, this institutional context may significantly reduce expectations of a transition towards upstream-oriented R&D as a result of employing PhD trained individuals.

Second, firms may recruit PhD trained individuals to strengthen an established orientation towards development and downstream activities. For instance, recruitment of highly qualified researchers may help firms to more effectively adopt and implement the results of research conducted elsewhere, which might improve the firm's current innovation activity, but may not result in a shift in its innovation strategy (Herrera 2020). This might be a particularly prevalent situation in a context in which there is evidence of a decline of science in corporate R&D, where firms use science as a relevant input to their innovation strategies but are increasingly less favourable to invest in internal scientific capabilities (Arora, Belenzon, and Patacconi 2018).

Third, the economics of science literature has pointed out that PhD trained individuals have significantly different preferences with regards both to their favoured type of research jobs and the type of benefits expected from them (Roach and Sauermann 2010). In other words, PhDs might have different preferences for how to capitalise on their PhD training and display a high degree of heterogeneity in their taste for science. When comparing PhD scientists' preferences for jobs in academy and industry, those willing to work in industry tend to have a lower inclination for publishing and a greater preference for access to state-of-art technology and equipments, as well as for conducting development activities (Roach and Sauermann 2010; Sauermann and Roach 2014). The results of these studies substantiate the claim that PhD scientists may self-select into industrial R&D jobs in order to meet their personal preferences for intellectual challenges associated to downstream research activities. In other words, in an organisational environment focused mainly on exploitation, the presence of PhD trained employees can reinforce already existing downstream research activities, without promoting a shift towards more upstream explorative research (Herrera and Nieto 2015). For instance, a high proportion of PhD degree holders in the firm can increase the development of more technologically advanced products and achieve reduced time to commercialisation of close-to-the-market working prototypes (Deeds, DeCarolis, and Coombs 2000; Tegarden et al. 2012). In short, in this perspective, employing researchers with a PhD training does not necessarily lead to the initiation of upstream-oriented R&D; instead, it may reinforce a strategic orientation towards downstream-oriented R&D activities.

In the light of this discussion, we hypothesise that:

H2: *A higher proportion of PhD trained employees among R&D staff increases the likelihood that a downstream-oriented firm will reinforce its focus on downstream R&D activities.*

Finally, in order to examine these competing hypotheses, we take into account potential contingencies associated to sectoral differences. Most empirical work on the effect of highly trained employees on firms' innovation strategies and performance focuses on science-driven and high-tech industries such as biotechnology. This strand of work suggests that the employment of researchers is linked to the capacity to search beyond the firm's existing technological boundaries (Al-Laham, Tzabbar, and Amburgey 2011). Thus, it can be expected that science-driven and high-tech industries will have a higher level of demand for these human resources, making them a suitable case for the positive impact of PhD trained employees on the shift to firms' upstream R&D strategies.

However, there are reasons to contend that the role of PhD trained individuals is relevant, also, for low-tech firms' upstream R&D strategies. Several studies show that the positive impact of scientists on firms' R&D strategies is not confined to high-tech industries, but applies also to medium and low-tech firms' research and innovation strategies (Kim and Marschke 2005; Spithoven and Teirlinck 2010; Teirlinck and Spithoven 2013; Kaiser et al. 2018). Due to their specialised knowledge and skills, the presence of PhD trained individuals might be especially important for firms with fewer knowledge recombination capabilities (Al-Laham, Tzabbar, and Amburgey 2011), young firms that lack the resources required to participate in sophisticated R&D alliances (Tzabbar, Aharonson, and Amburgey 2013) and firms that need to demonstrate scientific and technological competence (Rao, Chandy, and Prabhu 2008). Therefore, we are interested in whether the influence on the firm's initiation of an upstream-oriented R&D strategy, as a consequence of employing PhD trained employees in the firm's R&D function, is restricted to a narrow range of industries (e.g., high-tech industries) or is pervasive among firms regardless of industry.

3. Context, data and methodology

3.1. Context

Spain is an interesting case to study since its low levels of firm R&D activity and weak exploitation of public research from industry pose important challenges on its PhD labour market (Herrera and Nieto 2016; Martinez, Cruz-Castro, and Sanz-Menendez 2016). Compared to other European countries, Spain has limited PhD production in Science and Engineering (S&E) and low PhD employment in firms (Auriol, Misu, and Freeman 2013). According to the Spanish Human Resources in Science and Technology (HRST) survey, in 2009, around 15% of the PhD holders surveyed were employed in the business sector (Herrera and Nieto 2016). These figures contrast with those for other countries, such as Denmark, Belgium and the US, where at least one in three employed PhD researchers works in industry (Auriol, Misu, and Freeman 2013).

Studies of the Spanish PhD labour market reveal that PhD holders traditionally are employed in the public sector (Cruz-Castro and Sanz-Menendez 2005; Herrera and Nieto 2016). However, recently, the private sector has emerged as a potential source of employment. Demand for PhD holders in the private sector is related positively to firm size and firm R&D efforts, which are linked to firms in the medium-high and high technology sectors (Garcia-Quevedo et al., 2012; Herrera and Nieto 2015). The study by Herrera and Nieto (2016) reveals that the chemical and food industries and manufacture of electrical, electronic and optical equipment usually present the highest rates of employment of PhD holders. Other studies of Spanish PhD careers suggest that PhD holders' motivations to work in the private sector are not related to economic factors. The attractiveness of professional experience in the private sector and the absence of stable employment in the public sector are among the most important drivers (Cruz-Castro and Sanz-Menéndez, 2005). With respect to PhD holders' research orientation, Herrera and Nieto (2016) show that PhD holders who engaged in technological development activities during their PhD training period have a higher likelihood of finding a job in the private sector. Similarly, Di Paolo and Mañe (2016) show that PhD holders feel that their taste for science is satisfied by a public sector job. Given the comparatively lower levels of R&D intensity of Spanish firms and the differing expectations about the role of PhD trained employees in firm competitiveness, the Spanish context is a good case to examine whether R&D employees with a doctoral degree contribute to increasing the likelihood that firms will initiate upstream R&D activities.

3.2. Data

The empirical analysis employs information from the Spanish Technological Innovation Panel (PITEC). This statistical instrument was developed by the Spanish Institute of Statistics (INE) (with advice from a group of university researchers), to study the evolution in Spanish firms' innovation activities over time. Access to the database for researchers is facilitated by an official web site.[1] PITEC information is based on the Spanish Innovation Survey and is structured similar to the Community Innovation

[1]https://icono.fecyt.es/pitec. For confidentiality reasons, PITEC anonymises the data using the procedure described on the web page.

Survey (CIS). CIS-type surveys are used widely to analyse innovation-related research questions in economics and management (Cassiman and Veugelers 2002; Laursen and Salter 2006; Mairesse and Mohnen 2010).

However, PITEC provides important added value compared to standard CIS surveys: it incorporates data from the Spanish R&D survey, which makes it appropriate for the present study for three reasons. First, it provides annual information on the proportion of employees in firms' R&D functions who have a doctoral degree. Second, it provides a breakdown of firms' investments in basic, applied and development R&D every year. Third, PITEC is a panel data survey which allows us to observe changes in firms' R&D strategies over time – which would not be feasible using cross-sectional data. We use data for the period 2006–2012 since, prior to 2006, R&D personnel numbers were not defined in terms of full-time equivalent functions. Finally, we restrict our sample to: (i) manufacturing firms – services R&D tends to be subject to different rules compared to manufacturing (Cainelli, Evangelista, and Savona 2006); and (ii) firms that engage in formal R&D activities (i.e., those reporting presence of a formal R&D function). For these firms, we have information on R&D employment (specifically, employees with a PhD degree) and upstream and downstream R&D activity.

From Table 1 we can see the pervasiveness of firms that engage in upstream R&D, but have no PhD degree holders: 51% of firm-year observations correspond to firms engaged in upstream R&D without employing doctoral graduates in their R&D departments, meaning that 78% of 'upstream R&D active firms' do not have PhD degree holders. In addition, 20% of firms whose R&D staff include some PhD trained employee, do not engage in upstream R&D. Therefore, having doctoral graduates in the R&D function is neither a sufficient nor a necessary condition for upstream-oriented R&D activities.

Table 1. Upstream-oriented R&D and PhD degree holders (n = 17,699).

		% PhD degree holders among R&D employees		
		None	> zero	Total
Firms Conducting R&D	No Upstream	5,426 (30.7%)	666 (3.8%)	6,092 (34.4%)
	Upstream doers	9,025 (50.9%)	2,582 (14.6%)	11,607 (65.6%)
	Total	*14,451 (81.7%)*	*3,248 (18.3%)*	*17,699 (100%)*

3.3. Variables and measures

3.3.1. Dependent variable

In the main analysis, our dependent variable is the initiation of upstream-oriented R&D (*Start_upstream*). We construct a dummy variable that takes the value 1 if the firm engaged in upstream-oriented R&D in t + 1 (but not in a given period t) and zero otherwise.[2] This restricts our sample to: (i) those firms observed over at least two consecutive periods; and (ii) firms not involved in upstream-oriented R&D in t,[3] resulting in a final sample of 5,815 observations.

[2]The results (available upon request) are robust to a more restrictive definition of 'upstream starters', including that eligible firms engage in sustained upstream-oriented R&D for 2 years once having embarked on upstream research or requesting that eligible firms had not been engaged in upstream-oriented R&D in the preceding 2 years to the decision to move beyond downstream oriented R&D.

[3]We provide an additional analysis of the relationship between the proportion of R&D employees with a doctoral degree and the degree of orientation to upstream-oriented R&D, for the whole sample of R&D performing firms in section 4.4.

PITEC data allow us to quantify firms' engagement in different types of R&D, consistent with established innovation activity measures (OECD, Eurostat 2005), differentiating between investment in basic, applied and development research. PITEC guidelines define basic research as related to experimental or theoretical work undertaken to obtain new knowledge about phenomena or observable facts, without envisaging any particular application, and applied research as original research undertaken to obtain new knowledge aimed at a particular objective. According to these definitions, our measure of upstream-oriented R&D is built as the sum of firms' basic and applied research expenditure, which captures the production and adoption of new knowledge.

3.3.2. Explanatory and control variables

Our explanatory variable measures the proportion of the firm's R&D staff with a doctoral degree (*Lsphd*), defined as the number of individuals with a PhD degree relative to the total number of R&D employees. We log transform this variable due to the skewness of its distribution and to facilitate interpretation of the estimated coefficients as semi-elasticities.[4]

Since there may be differences in the characteristics of the firms that decide to embark on upstream-oriented R&D compared to those that do not, we consider a complete set of covariates based on observables from the survey. Were these variables not included in the model, they potentially could be confounding factors, which could lead to biased estimation of the role played by PhD trained R&D employees. The covariates included: firm size (*Lsize*), measured as the logarithm of the firm's employee count; firm R&D intensity, which corresponds to development intensity since we use only firms not performing upstream R&D activities in t (*D_intensity*) and it is defined as the logarithm of the firm's development expenditure per employee[5]; the size of the development team (*Lsizeteam*), based on a full time equivalent count of employees; the percentage of sales from new to the market products (*Newmer*); 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*), 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 response to the question: 'how important are your competitors as a source of information for the innovation process'. In this last case, responses are 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 is high otherwise. We also include two indicators for innovation funding: *pubfun*, 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

[4]In the main analysis, we replace the log of zero with zeros. We have conducted several robustness checks (available upon request) to analyse if results were sensitive to this choice. First, we use the transformation log(x + 1). Second, we use the original variable without logs. Third, we remove firms with just one R&D employee. Results from each robustness check are very similar to those presented here.

[5]Development activities are defined by PITEC as systematic work, based on existing knowledge, derived from research or practical experience, which is directed towards the production of new materials, products or devices, or implementation of new processes, systems and services to improve existing materials, products or devices.

that takes the value 1 if the firm reports lack of internal or external funds as an obstacle to innovation – ranked as moderately or very important. Finally, we include year and industry dummies in line with the OECD (2011) technological intensity classification. Appendix 1 Table A1a presents the variables definitions.

Table 2 presents the descriptive statistics for all the variables in the regressions. Appendix 1 Table A1b presents the correlation matrix including both the dependent and independent variables.

Table 3 shows that, although the most likely scenario is no transition towards upstream-oriented R&D (74.3% of firm-year observations), employing PhD trained individuals in the R&D function is not a necessary condition for initiating upstream R&D: 14.6% of firm-year observations embarked on upstream R&D without employing PhD degree holders. Moreover, among firms with PhD employees in their R&D functions that were not engaged in upstream R&D at time t, 80% (515 out of 646) were also not engaged in such research at time t + 1.

It should be highlighted (as shown in Table 3) that only 646 firm-year observations (11.1% of firms without upstream-oriented R&D in t) shows at least one PhD. Figure 1 shows the distribution of the percentage of PhDs in the R&D function for these 646 observations.

Table 2. Descriptive statistics (n = 5,815).

	Mean	Sd	Min	Max
Start_upstream	0.17	0.37	0.00	1.00
SPhD	3.07	12.07	0.00	100.00
Lsize	4.14	1.28	0.00	9.23
D_intensity	7.91	1.30	3.01	12.92
Export	0.87	0.34	0.00	1.00
Lsizeteam	1.17	1.03	0.00	5.95
Parent	0.07	0.25	0.00	1.00
Joint_venture	0.01	0.10	0.00	1.00
Newmer	12.70	24.74	0.00	100.00
Obstacle_funds	0.74	0.44	0.00	1.00
Appropriability	2.79	0.16	1.75	4.00
Park	0.03	0.17	0.00	1.00
Lage	3.11	0.66	0.00	4.71
Pubfun	0.49	0.50	0.00	1.00

Table 3. Transitions to upstream-oriented R&D and PhD degree holders (n = 5,815).

	No upstream in t+1		Upstream doers in t+1	
	% of firms without PhD	% of firms with PhD	% of firms without PhD	% of firms with PhD
No upstream in t	4318 (74.3%)	515 (8.9%)	851 (14.6%)	131 (2.3%)

3.3.3. Instrumental variables

There are several reasons why endogeneity may be a problem in this study (Bascle 2008). First, attenuation bias, stemming from measurement of the PhD variable with error, would mean that the estimated coefficient is lower than the true coefficient.

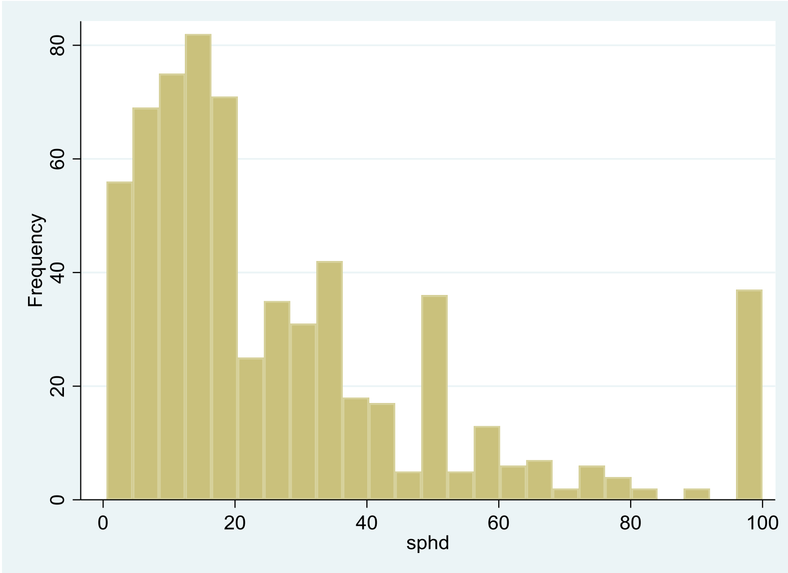

Figure 1. Distribution of the share of PhDs for downstream-oriented firms with at least one PhD (n = 646).

Second, although we control for a wide range of covariates, not all the relevant confounding factors may be observable.[6] Third, even were they observable, there would remain the possibility of reverse causality. Since we are interested in the line of causality from employment of PhD trained individuals in R&D departments, to changes to the R&D strategy, we need to address the potential problem of endogeneity. We adopt an instrumental variable approach employing two different instruments.

The first instrument is based on the notion that firms face an exogenous supply of PhD graduates, which should influence employment of researchers, but should be uncorrelated to the firm's R&D strategy. In other words, it should be a source of exogenous variation. We built an indicator for this supply based on PhD graduates having studied for their degree in the region in which the firm is located, and in scientific and technological fields relevant to the firm's economic activity. To determine which scientific and technological fields are relevant, we use the matrix provided by Cohen, Nelson, and Walsh (2002), which links scientific field to economic industry, and follow the methodology proposed in Abramovsky, Harrison, and Simpson (2007). To build our instrument, we exploit university statistics provided by INE. We match the supply of PhD trained individuals to the different manufacturing industries and locations, to develop an indicator of the firm-specific supply of PhDs (details on the construction of this indicator are provided in Appendix 2).[7]

The second instrument is the industry average of the potentially endogenous variable, that is, the share of doctoral employees in the R&D function. This type of instrument is used widely in research based on CIS data (see, e.g., Cassiman and Veugelers 2002;

[6]It could also be that firms report higher levels of upstream-oriented R&D because they had hired a new PhD.
[7]In a robustness check, we analyse the results using a different matrix (developed by ourselves and based on data from the INE PhD Survey) to match scientific fields and economic industries.

Veugelers and Cassiman 2005). The underlying idea is that, having controlled for the 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).

The identifying assumption in our instrumental variable approach is that, after accounting for firm specific characteristics (size, development-intensity, export, R&D team size, belonging to a group, product innovation, appropiability, obstacle funds, tech content of the industry, being located on a park, age and public funding) the regional supply of PhD graduates (weighted by the field-industry matrix) and the PhD industry average do not influence firms' decision to start upstream R&D activities by a channel different to PhDs. That is, our identifying assumption does not require supply of PhDs to be uncorrelated with the characteristics of firms in the region, as we are controlling by these firm characteristics. Actually, one of the main roles played by covariates in Instrumental Variables regression is to make the instrument exogenous. The key point is that, when using this methodological approach, we are identifying the effects of PhDs on the transition to upstream R&D not by using the actual presence of PhDs in each firm, but the 'expected' PhDs of each firm according to the part of the regional supply of PhD graduates and of the industry average of PhDs which are uncorrelated with the covariates used. This is why this method allows us to address the reverse causality problem of firms actually hiring PhDs because they want to move towards upstream research activities.[8]

In all the specifications we test for the relevance and exogeneity of the instruments using the F-statistic of the first equation and the Hansen's J-statistic respectively. The F-statistics are always well above 10 and the hypothesis of exogeneity of instruments is always clearly not rejected.

3.4. Estimation method

To test our hypotheses, we formulate the following model:

$$Start\_upstream_{it} = \beta_0 + \beta_1 SPhD_{it} + x'_{it}\gamma + \varepsilon_{it}$$

where *Start_upstream* captures initiation of upstream-oriented R&D, *SPhD* denotes presence of PhD trained employees in the firm's R&D function (measured as the proportion of employees with a doctoral degree relative to total R&D employees), and x is a vector of the covariates.

We present the results of the Ordinary Least Squares (OLS) regressions and analyse coefficient stability (Oster 2019), which provides a lower bound of 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 3). In addition, since the dependent variable is binary, we complement the analysis with a probit model.

[8]To further investigate the exogeneity of this instrument we thought about an alternative channel through which the regional supply of related PhDs could influence firm decision to start upstream R&D. A clear candidate would be firm cooperation with universities. That is, it could be that a potential source of endogeneity would be that the instrument proxies for university-firm cooperation. We run an alternative model including firm cooperation with university as an additional regressor and results discard the possibility that this channel would be a source of endogeneity. Note that in the main specifications we do not include cooperation with university as an additional regressor as this could be a channel through which the effect of PhDs on research takes place so that it would be a 'bad' control (Angrist and Pischke 2008).

The first results do not take account of the potential endogeneity of the share of PhD graduates. We then run two IV estimations – first a standard IV estimation using Two-Stage Least Squares (2SLS) and second an estimation that takes account of the fact that our potentially endogenous variable (share of PhD degree holders) 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 it predicts as the instrument for the IV estimation.[9]

4. Results

4.1. Basic trends: starting upstream-oriented R&D by type of industries

Table 4 presents the share of R&D employees with a doctoral degree, number of firms and percentage of firms starting upstream-oriented R&D, for those firms not engaged in upstream R&D in year t, by industry categories. The industry technology categories classification is based on R&D intensity (R&D investments/sales) (OECD 2011).

We observe that, on average, firms in low, low-medium and medium-high tech industries have very similar shares of PhD employees in their R&D functions, with only high-tech firms displaying comparatively higher figures. However, we observe no fundamental differences across low and high tech industries in terms of the probability to embark on upstream-oriented R&D. While low-tech industries devote comparatively fewer resources to R&D, the proportions of firms that decide to undertake upstream R&D in t + 1 are similar across industry categories. This shows that upstream-oriented R&D is not unique to high-tech industries, but occurs in all industries – low, medium and high tech.

The next sub-sections offer a more systematic examination of the relationship between proportion of PhD degree holders in R&D functions and the initiation to upstream R&D.

Table 4. Proportion (%) of PhD degree holders and upstream R&D by industry (n = 5,815).

	LT	LMT	MHT	HT	Total
Share of PhD degree holders	2.57%	2.24%	3.31%	5.08%	3.07%
Number of firms	1,182	1,604	2,330	699	5,815
% of firms starting upstream R&D in t + 1	18.19%	15.77%	16.57%	18.31%	16.89%

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

4.2. Start upstream-oriented R&D and PhD trained R&D employees

There are 5,815 (firm-year) observations, corresponding to companies that conduct R&D, but did not engage in upstream R&D in year t. These firms have an average of 3.07% PhD degree holders in their R&D functions. Some 17% of these firm-year observations (982 cases) make the transition to upstream R&D in t + 1. On average, those firms' R&D teams include 4.35% of employees with a doctoral degree compared to 2.81%, on average, among firms whose R&D strategies do not change.

[9]Note that, although we use the panel dimension to define changes in firms' decisions, we cannot use within-variation to estimate the effects since firms do not either initiate or abandon research activity more than once in the period analysed.

Table 5. Start upstream R&D. Results.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	SimpleOLS	Multiple OLS	IV-GMM	IV-GMM Wool	Probit	IV Probit	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]
D_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 ***p-value<0.01; **p-value<0.05, *p-value<0.1

Table 5 presents the results of the different models. Column (1) presents 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 the covariates. Column (6) presents the results of the 'standard' IV probit and column (7) includes the probit IV results using the Wooldridge approach.

Column (1) shows that the variable measuring the share of doctoral employees (Lsphd) has a positive and statistically significant coefficient of 0.017. That is, a 1% increase in the share of R&D employees with a PhD degree is associated to a 0.017 increased probability of starting upstream R&D activities. If we apply this effect to the average probability of transition (0.169), then a 1% increase in the share of PhDs is associated to a 0.1% increase in this probability. Column (2) includes all the covariates and provides no evidence of omitted variables bias since the coefficient remains fairly stable. With the exceptions of *obstacle funds* (i.e., firms experiencing funding constraints) and *lage* (firm age), the observables are mostly non-significant.

Oster's bound 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 of 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 R&D employees with a PhD degree is related to a 0.133 points increase in the probability of a transition to upstream-oriented R&D. Again, taking the average probability of transition as the reference, the results of the IV regressions mean that a 1% increase in the share of R&D employees with a PhD degree is associated to an increase of 0.79% in the likelihood of a transition.

The much larger size of the coefficients of the IV compared to the OLS estimations is in line with the literature (see, e.g., Trostel, Walker, and Woolley 2002; Jiang 2017) and might be because the IV are weak instruments. Although the F-statistics do not reveal any problems, we need to explore this possibility. Following Bascle (2008), we compare our results with those based on Moreira's (2003) Conditional Likelihood Ratio (CLR). It has been argued that CLR is the test of choice in IV applications (Murray 2006). If the results using IV differ from the results using the CLR method, there is a finite sample/weak instrument problem (Yogo 2004). We employed the CLR method (results available upon request) and found that the coefficients were remarkably similar, suggesting that weak instruments are not the cause of this change in the coefficient.

An alternative reason why IV estimates are much higher than OLS might be the existence of 'essential heterogeneity' (firms respond differently to the treatment and their selection into the treatment is based on unobserved gains). In that case, IV provides a consistent estimate of the Local Average Treatment Effect (LATE), which might differ from the Average Treatment Effect (ATE) (Bascle 2008; Jiang 2017). That is, IV provides an estimation of the effect for 'compliers' (individuals responding, as intended, to the instrument). Complier firms will likely enjoy a stronger effect than the average firm, from having PhD degree holders on their staff, demonstrated by the much larger coefficient in the IV regression. Another reason why the IV coefficients are much higher than OLS is that we are working with anonymised data, which could generate measurement errors and result in 'attenuation bias' in the OLS estimations.

In our model, the endogenous regressor is a limited dependent variable (share of PhD degree holders ranges between 0 and 100). To address this potential problem we apply the 'Wooldridge' approach to the IV estimations, with a limited endogenous regressor. The results of the Wooldridge approach, taking account of the limited nature of our endogenous regressor, are provided in Table 5 Column (4): the coefficient is 0.08, meaning that a 1% increase in the share of R&D employees with a PhD degree is associated to a 0.08 points increase (a change of 0.47% in the probability, taking the average probability as the reference). This is still high, but is closer to the OLS estimations.[10]

To summarise, the linear models show a positive and significant effect of the share of PhD trained R&D employees on the likelihood of transition to an exploratory R&D

[10]Wald test of exogeneity: Chi-square(1) = 8.66, p-value = 0.003.

strategy. However, the dependent variable is binary so, by definition, the marginal effects should be non-linear. To analyse these effects, we employ probit models (Table 5 Columns (5)-(7)). Column (5) provides the results of the probit model assuming exogeneity of PhD degree holders. The marginal effects at the means are very similar to the OLS coefficients. Columns (6) and (7) provide the results of the IV approach. Again, the marginal effects at the means are close to (but slightly lower) than the OLS coefficients. The main advantage of a probit model is that it allows us to plot the full set of marginal effects.

Figure 2 plots the theoretical 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.[11] We calculated the marginal effects of a 1% increase in the share of R&D employees with a PhD degree for the different firms in the sample: they range from a 0.011 probability points to a 0.18 probability points in the standard IV probit model, and from a 0.011 probability points to a 0.11 probability points in the alternative IV probit model. These results support our expectation that the effects are very heterogeneous across firms, as we anticipated from the difference between ATE and LATE.[12]

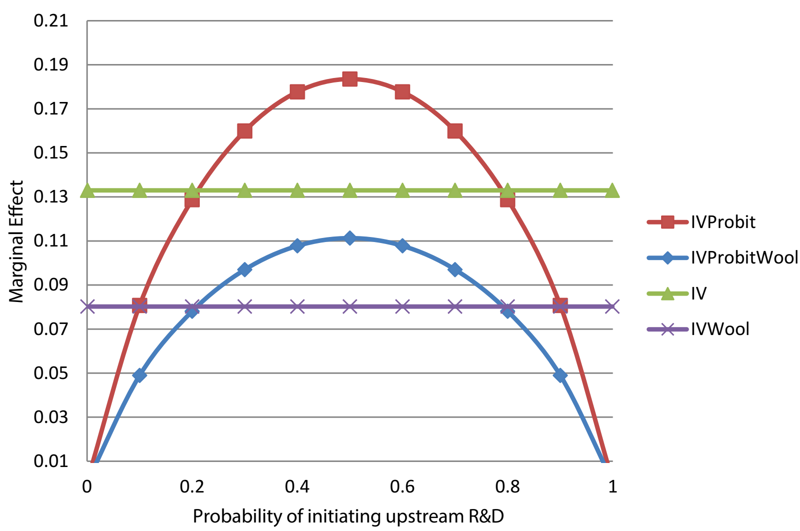

Figure 2. Marginal effects. Start upstream R&D.

4.3. An industry perspective

In this section, we test whether the results are industry-specific or hold across different industries. We estimate the regressions with interaction terms between the share of PhD degree holders and industry dummies, for the decision to engage in upstream-oriented R&D. The interaction terms show the differential effect of PhD trained R&D employees

[11]Note that the marginal effects are plotted based on the probability of a transition to upstream research, rather than on the basis of one particular explanatory variable with specific values chosen for the remaining covariates.

[12]If we focus instead on the effect of the first PhD, we find that the effect ranges from 1.8 to 4.9 probability points, with a mean value of 3.3 probability points for the likelihood of initiating upstream-oriented R&D.

on starting upstream R&D, for different industries, compared to the baseline category (medium-high tech industries). Table 6 presents the results. We use OLS (with clustered errors).[13]

We observe that the effect of the share of PhD degree holders on starting upstream research is similar across industries since the coefficients of the different interaction terms are small in magnitude and not statistically significant. In particular, there is no evidence of higher relative importance of PhD trained R&D employees in high-tech industries, which is something that is not highlighted in previous studies.

4.4. Robustness checks

We performed several robustness checks.[14] First, we conducted an additional analysis of the whole population of firms with R&D activities. We define *Upstream-orientation* as the percentage of the R&D activity that is upstream R&D. This is a continuous variable that ranges between 0 and 100% and has been log-transformed so that results can be interpreted as elasticities. There are 17,699 (firm-year) observations corresponding to companies that conduct R&D. These firms have an average of 4.82% PhD researchers in their R&D functions and devote 45.2% of their R&D investment to upstream R&D. The results are provided in Appendix 4. This robustness check exploits the panel structure of our data by estimating random and fixed effects models. Fixed effects estimations remove any bias caused by correlation between time-invariant firm characteristics and the percentage of PhD holders in the R&D team. Finally, we run two IV estimations. The first is a standard IV estimation using 2SLS; the second uses the Wooldridge method, taking account of the specific shape of the potentially endogenous variable (share of PhDs). The results from this set of models reinforce the existence of a positive and significant relationship between share of R&D employees with a PhD degree and an upstream-oriented R&D.[15]

Second, since the matrix provided by Cohen, Nelson, and Walsh (2002) and applied 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, we built 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 are similar (see Appendix 5).

Third, in the main analysis we jointly consider basic and applied research as upstream-oriented R&D activities. In this robustness check, we focus on the transition to basic research. The sample is composed by firms active in development or applied research in period t which are also observed in period t + 1. The total sample is composed of 15,165 firm-year observations, from which 4.9% firms move towards basic research. Results

[13]Probit models do not allow direct interpretation of the coefficients of the interaction. For ease of reading, we provide only the OLS estimations, although we obtained similar results from the probit and logit models, after the interaction effects are properly computed.

[14]We have also checked if the main results hold when analysing if the percentage of PhD holders reduced the likelihood of abandoning upstream R&D activities for those upstream-active firms. We found that, coherently with the rest of the analysis, a higher percentage of PhD holders reduce the likelihood of abandoning upstream-oriented R&D (results available upon request). We thank one reviewer for the suggestion to complement the analysis in this way.

[15]It is possible that the OLS and, especially, the fixed effects estimates might suffer from attenuation bias. In fact, the coefficients of the IV estimates are much larger. However, we can discard this possibility as being due to weak instruments and conclude that it is likely the consequence of a higher LATE than ATE – 0.87 compared to 0.20.

Table 6. Start upstream R&D by industries Results.

	OLS
Lsphd	0.021**
	(0.009)
Lsphd*low	−0.004
	(0.019)
Lsphd*mediumlow	−0.014
	(0.015)
Lsphd*high	0.001
	(0.017)
Lsize	−0.005
	(0.008)
D_intensity	−0.009
	(0.007)
Export	0.017
	(0.016)
Lsizeteam	−0.002
	(0.010)
Parent	0.025
	(0.021)
Joint_venture	0.071
	(0.056)
Newmer	0.000
	(0.000)
Obstacle_funds	0.025**
	(0.012)
Appropriability	−0.068
	(0.045)
Low-tech	0.029
	(0.018)
Mediumlow	0.007
	(0.016)
High	0.001
	(0.020)
Park	0.042
	(0.038)
Lage	−0.019**
	(0.009)
Pubfun	0.001
	(0.011)
_cons	0.467***
	(0.148)
N	5815

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

fundamentally hold (see Table A6). Note that, although the absolute effect in probability points (pp) is lower (still positive and significant) the relative effect is quite similar. In the multiple regression model a 1% increase in the share of PhDs is related to an increase of 0.4pp in the likelihood of conducting basic research (which is a relative increase of 8.16% when applied to the probability of doing this transition, which is 4.9%). In the main analysis using upstream-oriented R&D a 1% increase in the share of PhDs is related to an increase of 1.7pp in the likelihood of conducting upstream R&D (which is a relative increase of 10% when applied to the probability of doing this transition, which is 17%).

5. Conclusions

The objective of this study was to analyse whether the proportion of PhD trained employees in the firm's R&D function increases the probability of starting upstream-oriented R&D research by firms engaged in downstream R&D activities, or whether instead it reinforces these firms' orientation towards downstream R&D. We tested the validity of these two competing propositions on a large sample of Spanish manufacturing and R&D performing firms during the period 2006–2012 and found that firms with a higher proportion of doctoral trained employees in their R&D functions were more likely to initiate upstream-oriented R&D, controlling for alternative explanatory factors such as scale and intensity of firms' downstream R&D activities.

This study extends the existing innovation management and business strategy research in two ways. First, from a conceptual perspective, we argue that the role of PhD degree holders in R&D functions goes beyond absorption and generation of knowledge and influences the firm's R&D strategy. We suggest that a concentration of research talent instils a taste for upstream R&D among R&D employees, which has a strong effect on the firm's capacity to move beyond the current focus on downstream development activities and favours the adoption of an upstream-oriented R&D. We argue that this shift in the firm's R&D strategy is likely to be an effect of three aspects of changes to the firm's organisational learning culture associated to the availability of PhD trained R&D employees: strong connectivity to the science-base, influence on teamwork research performance and a climate tolerant of risk-taking and failure.

Second, from an empirical perspective, our results highlight two important aspects of the relationship between highly trained employees and firms' R&D strategies. On the one hand, we show that there is no one-to-one relationship between employing PhD trained individuals in the R&D function and conducting upstream-oriented R&D. We found that firms with no doctoral trained employees engage in upstream-oriented R&D and, also, that the presence of PhD trained R&D employees is not sufficient to promote an upstream-oriented R&D strategy. On the other hand, we provide strong evidence of a link between a concentration of highly trained human capital (proportion of employees with a PhD degree among R&D employees) and the probability to initiate upstream-oriented R&D activity. We provide specific estimates of the size of this effect. The average effect of a 1% increase in the share of doctoral employees is a 0.017 probability point increase in the likelihood of initiating upstream R&D, while the local average effect is a 0.066 probability increase. Heterogeneity of the effects is supported by probit models where we find a maximum potential probability increase of 0.18.

Our analysis includes a large set of covariates (e.g., appropriability regime, firm age and size, scale and intensity of firms' downstream activities, among others), which ensures robustness of our results to alternative explanatory factors. Also, we employed instrumental variables estimations, building two indicators of the exogenous supply of PhD graduates as instruments, which renders our findings generally robust to endogeneity concerns. In short, we demonstrate that a higher proportion of PhD degree holders in the R&D function increases the likelihood that the firm will search beyond its knowledge boundaries and invest in exploratory research.

Finally, our study has implications for industry practitioners and policy makers. Our results show that the proportion of PhD trained R&D employees has a direct influence on the firm's capacity to embark on an upstream-oriented strategy. Exploratory research is critical for sustained competitive advantage and ability to respond to technological change (Michelino et al. 2015). In this sense, the acquisition of new skills through recruitment of highly qualified employees for the R&D function, seems crucial for exploratory activities that contribute to the introduction of new products, and strengthening of the firm's capacity to continuously adapt its knowledge base to be able to respond to changing customer preferences and novel technologies. These implications are reinforced by the finding that our results are robust across very different industry and sectoral settings, including firms in low and medium-tech as well as high-tech sectors. Our results show, also, that the other potential drivers of the decision to initiate upstream-oriented R&D, such as the size and scale of downstream-oriented R&D, have no significant effects.

At the policy level, our research findings contribute to the rationale underlying policy initiatives to encourage or facilitate firms' recruitment of PhD graduates, particularly in contexts where the proportion of PhD holders employed in the business sector is comparatively low by international standards. Previous research points to systemic failures in the PhD jobs market (Cruz-Castro and Sanz-Menendez 2005; Martínez et al., 2016; Martínez and Parlane, 2018) due to firms' lack of awareness of the long term benefits associated to PhD recruitment. These failures are due, also, to the difficulty related to attracting PhD trained individuals to work in corporate settings which might fail to provide adequate organisational conditions for the advancement of scientists' professional careers. If increasing exploratory research by firms is a policy target, then it would seem that facilitating the employment of doctoral trained researchers in R&D functions would contribute to address this goal.

The importance of PhDs for upstream R&D raises questions related to the management of highly qualified human resources and suggests some new research directions. First, there are issues related to career development opportunities and reward mechanisms, and their effect on the involvement of R&D employees in firms' exploratory R&D activities, which require further research (Chen, Chang, and Yeh 2003; Sauermann and Cohen 2010; Liu and Stuart 2014; Balsmeier and Pellens 2016). Second, we need to know more about the conditions that allow R&D employees to exploit firm upstream R&D activity to improve innovation performance. Third, more research is needed on the extent to which PhD degree holders enhance the capacity of firms to integrate and exploit external sources of scientific knowledge, in an international context characterised by increasing reliance on outsourcing and overall decline in firms' engagement in internal scientific research (Arora, Belenzon, and Patacconi 2018). Fourth, in addition to using formal education, we need to consider other variables to capture the firm's wider R&D skills base. Fifth, another extension of the work would be to examine the effect of PhD holders outside the R&D team on the initiation of R&D activities. Finally, future research could analyse the potential heterogeneous effects of PhD degree holders in R&D functions, by firm type, for example, firm size and R&D intensity.

Acknowledgments

We are grateful to the Associate Editor Anders Broström and the anonymous reviewers for their excellent advice and constructive feedback throughout the review process. We also thank for the valuable comments received on earlier versions of this article, at the following conferences:

DRUID (2016), EURKIND (2016), ISSC (2016) and EARIE (2016). Andrés Barge-Gil acknowledges funding from projects ECO2014-52051-R, S2015/HUM-3417 and ECO2017-82445-R. Pablo D'Este acknowledges funding from projects ECO2014-59381-R and RTI2018-101232-B-I00. Liliana Herrera acknowledges funding from project ECO2015-63880-R.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Abramovsky, L., R. Harrison, and H. Simpson. 2007. "University Research and the Location of Business R&D." *The Economic Journal* 117 (519): 114–141. doi:10.1111/j.1468-0297.2007.02038.x.

Ahlin, L., M. Andersson, and T. Schubert. 2013. *Implementing an R&D Strategy without Prior R&D-Experience Recruitment as a Source of R&D-Related Routines and Capabilities?* Lund: CIRCLE - Center for Innovation, Research and Competences in the Learning Economy.

Al-Laham, A., D. Tzabbar, and T. L. Amburgey. 2011. "The Dynamics of Knowledge Stocks and Knowledge Flows: Innovation Consequences of Recruitment and Collaboration in Biotech." *Industrial and Corporate Change* 20 (2): 555–583. doi:10.1093/icc/dtr001.

Almeida, P., J. Hohberger, and P. Parada. 2011. "Individual Scientific Collaborations and Firm-level Innovation." *Industrial and Corporate Change* 20 (6): 1571–1599. doi:10.1093/icc/dtr030.

Angrist J. D., and J. S. Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. New Jersey: Princeton University Press.

Añón Higón, D. 2016. "In-house versus External Basic Research and First-to-market Innovations." *Research Policy* 45 (4): 816–829. doi:10.1016/j.respol.2016.01.005.

Arora, A., S. Belenzon, and A. Patacconi. 2018. "The Decline of Science in Corporate R&D." *Strategic Management Journal* 39 (1): 3–32. doi:10.1002/smj.2693.

Auriol, L., M. Misu, and R. A. Freeman. 2013. *Careers of Doctorate Holders: Analysis of Labour Market and Mobility Indicators*. Paris: OECD Science, Technology and Industry.

Balsmeier, B., and M. Pellens. 2016. "How Much Does It Cost to Be a Scientist?" *The Journal of Technology Transfer* 41 (3): 469–505. doi:10.1007/s10961-014-9388-1.

Barge-Gil, A., and A. Conti. 2013. "Firm R&D Functions and Outsourcing Partners: A Matching Story." *SSRN Electronic Journal*. doi:10.2139/ssrn.2210113.

Barge-Gil, A., and A. López. 2014. "R&D Determinants: Accounting for the Differences between Research and Development." *Research Policy* 43 (9): 1634–1648. doi:10.1016/j.respol.2014.04.017.

Barge-Gil, A., and A. López. 2015. "R versus D: Estimating the Differentiated Effect of Research and Development on Innovation Results." *Industrial and Corporate Change* 24 (1): 93–129. doi:10.1093/icc/dtu002.

Bascle, G. 2008. "Controlling for Endogeneity with Instrumental Variables in Strategic Management Research." *Strategic Organization* 6 (3): 285–327. doi:10.1177/1476127008094339.

Bercovitz, J., and M. Feldman. 2008. "Academic Entrepreneurs: Organizational Change at the Individual Level." *Organization Science* 43 (9): 69–89. doi:10.1287/orsc.1070.0295.

Cainelli, G., R. Evangelista, and M. Savona. 2006. "Innovation and Economic Performance in Services: A Firm-level Analysis." *Cambridge Journal of Economics* 30 (3): 435–458. doi:10.1093/cje/bei067.

Cassiman, B., and R. Veugelers. 2002. "R&D Cooperation and Spillovers: Some Empirical Evidence from Belgium." *American Economic Review* 92 (4): 1169–1184. doi:10.1257/00028280260344704.

Cassiman, B., R. Veugelers, and S. Arts. 2018. "Mind the Gap: Capturing Value from Basic Research through Combining Mobile Inventors and Partnerships." *Research Policy* 47 (9): 1811–1824. doi:10.1016/j.respol.2018.06.015.

Chen, T. Y., P. L. Chang, and C. W. Yeh. 2003. "The Study of Career Needs, Career Development Programmes and Job Satisfaction Levels of R&D Personnel: The Case of Taiwan." *The International Journal of Human Resource Management* 14 (6): 1001–1026. doi:10.1080/0958519032000106182.

Cockburn, I. M., and R. M. Henderson. 1998. "Absorptive Capacity, CoAuthoring Behaviour and the Organisation of Research in Drug Discovery." *The Journal of Industrial Economics* XLVI (2): 157–182.

Cohen, W. M., R. R. Nelson, and J. P. Walsh. 2002. "Links and Impacts: The Influence of Public Research on Industrial R&D." *Management Science* 48 (1): 1–23. doi:10.1287/mnsc.48.1.1.14273.

Cruz-Castro, L., and L. Sanz-Menendez. 2005. "The Employment of PhDs in Firms: Trajectories, Mobility and Innovation." *Research Evaluation* 14 (1): 57–69. doi:10.3152/147154405781776292.

Czarnitzki, D., B. Ebersberger, and A. Fier. 2007. "The Relationship between R&D Collaboration, Subsidies and R&D Performance: Empirical Evidence from Finland and Germany." *Journal of Applied Econometrics* 22 (7): 1347–1366. doi:10.1002/jae.992.

Czarnitzki, D., and S. Thorwarth. 2012. "Productivity Effects of Basic Research in Low-tech and High-tech Industries." *Research Policy* 41 (9): 1555–1564. doi:10.1016/j.respol.2012.04.009.

D'Este, P., N. Amara, and J. Olmos-Peñuela. 2016. "Fostering Novelty while Reducing Failure: Balancing the Twin Challenges of Product Innovation." *Technological Forecasting and Social Change* 113: 280–292. doi:10.1016/j.techfore.2015.08.011.

D'Este, P., A. Marzucchi, and F. Rentocchini. 2018. "Exploring and yet failing less: learning from past and current exploration in R&D." 27 (3): 525–553.

Dasgupta, P., and P. A. David. 1994. "Toward a New Economics of Science." *Research Policy* 23 (5): 487–521. doi:10.1016/0048-7333(94)01002-1.

Deeds, D. L., D. DeCarolis, and J. Coombs. 2000. "Dynamic Capabilities and New Product Development in High Technology Ventures: An Empirical Analysis of New Biotechnology Firms." *Journal of Business Venturing* 15 (3): 211–229. doi:10.1016/S0883-9026(98)00013-5.

Di, P. A., and F. Mane. 2016. "Misusing Our Talent? Overeducation, Overskilling and Skill Underutilisation among Spanish PhD Graduates." *The Economic and Labour Relations Review* 27 (4): 432–452. doi:10.1177/1035304616657479.

Ding, W. W. 2011. "The Impact of Founders' Professional-Education Background on the Adoption of Open Science by For-Profit Biotechnology Firms." *Management Science* 57 (2): 257–273. doi:10.1287/mnsc.1100.1278.

Edmondson, A. 2011. "Strategies for Learning from Failure." *Harvard Business Review* 89 (4): 48–55.

Felin, T., and N. J. Foss. 2005. "Strategic Organization: A Field in Search of Micro-foundations." *Strategic Organization* 3 (4): 441–455. doi:10.1177/1476127005055796.

Furukawa, R., and A. Goto. 2006. "Core Scientists and Innovation in Japanese Electronics Companies." *Scientometrics* 68 (2): 227–240. doi:10.1007/s11192-006-0109-x.

Gans, J. S., and S. Stern. 2010. "Is There a Market for Ideas?" *Industrial and Corporate Change* 9 (3): 805–837. doi:10.1093/icc/dtq023.

Garcia, R., R. Calantone, and R. Levine. 2003. "The Role of Knowledge in Resource Allocation to Exploration versus Exploitation in Technologically Oriented Organizations." *Decision Sciences* 34 (2): 323–349. doi:10.1111/1540-5915.02407.

García-Quevedo, J., F. Mas-Verdu, and J. Polo-Otero. 2012. "Which Firms Want PhDs? An Analysis of the Determinants of the Demand." *Higher Education* 63 (5): 607–620. doi:10.1007/s10734-011-9461-8.

Gittelman, M., and B. Kogut. 2003. "Does Good Science Lead to Valuable Knowledge? Biotechnology Firms and the Evolutionary Logic of Citation Patterns." *Management Science* 49 (4): 366–382. doi:10.1287/mnsc.49.4.366.14420.

Grigoriou, K., and F. T. Rothaermel. 2014. "Structural Microfoundations of Innovation." *Journal of Management* 40 (2): 586–615.

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

Groysberg, B., L.-E. Lee, and A. Nanda. 2008. "Can They Take It with Them? The Portability of Star Knowledge Workers' Performance." *Management Science* 54 (7): 1213–1230. doi:10.1287/mnsc.1070.0809.

Herrera, L. 2020. "Effect of Corporate Scientists on Firms' Innovation Activity: A Literature Review." *Journal of Economic Surveys* 34 (1): 109–153. doi:10.1111/joes.12341.

Herrera, L., and M. Nieto. 2015. "The Determinants of Firms' PhD Recruitment to Undertake R&D Activities.".*European Management Journal* 33 (2): 132–142. doi:10.1016/j.emj.2014.10.003.

Herrera, L., and M. Nieto. 2016. "PhD Careers in Spanish Industry: Job Determinants in Manufacturing versus Non-manufacturing Firms." *Technological Forecasting and Social Change* 113: 341–351. doi:10.1016/j.techfore.2015.09.019.

Herrmann, A. M., and A. Peine. 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." *Research Policy* 40 (5): 687–701. doi:10.1016/j.respol.2011.02.004.

Hess, A., and F. T. Rothaermel. 2012. "Intellectual Human Capital and the Emergence of Biotechnology: Trends and Patterns, 1974-2006." *IEEE Transactions on Engineering Management* 59 (1): 65–76. doi:10.1109/TEM.2010.2082550.

Jiang, W. 2017. "Have Instrumental Variables Brought Us Closer to the Truth." *The Review of Corporate Finance Studies* 6 (2): 127–140. doi:10.1093/rcfs/cfx015.

Kaiser, U., H. C. Kongsted, K. Laursen, and A. K. Ejsing. 2018. "Experience Matters: The Role of Academic Scientist Mobility for Industrial Innovation." *Strategic Management Journal* 39 (7): 1935–1958. doi:10.1002/smj.2907.

Katila, R., and G. Ahuja. 2002. "Something Old, Something New: A Longitudinal Study of Search Behavior and New Product Introduction." *Academy of Management Journal* 45 (6): 1183–1194.

Kehoe, R. R., and D. Tzabbar. 2015. "Lighting the Way or Stealing the Shine? An Examination of the Duality in Star Scientists' Effects on Firm Innovative Performance." *Strategic Management Journal* 36 (4): 586–597. doi:10.1002/smj.2240.

Kim, J., and G. Marschke. 2005. "Labor Mobility of Scientists, Technological Diffusion, and the Firm's Patenting Decision." *The RAND Journal of Economics* 36: 298–317.

Kim, N., and E. Kim. 2015. "Board Capital and Exploration: From a Resource Provisional Perspective." *Management Decision* 53 (9): 2156–2174. doi:10.1108/MD-11-2014-0648.

Lacetera, N., I. M. Cockburn, and R. Henderson. 2004. "Do Firms Change Capabilities by Hiring New People? A Study of the Adoption of Science-Based Drug Discovery." *Advances in Strategic Management* 21: 133–159. doi:10.1016/S0742-3322(04)21005-1.

Laursen, K., and A. Salter. 2006. "Open for Innovation: The Role of Openness in Explaining Innovation Performance among U.K. Manufacturing Firms." *Strategic Management Journal* 27 (2): 131–150. doi:10.1002/smj.507.

Levinthal, D. A., and J. G. March. 1993. "The Myopia of Learning." *Strategic Management Journal* 14 (2 S): 95–112. doi:10.1002/smj.4250141009.

Liu, C. C., and T. Stuart. 2014. "Positions and Rewards: The Allocation of Resources within a Science-based Entrepreneurial Firm." *Research Policy* 43 (7): 1134–1143. doi:10.1016/j.respol.2014.04.013.

Lowe, R. A., and F. M. Veloso. 2015. "Patently Wrong? Firm Strategy and the Decision to Disband Technological Assets." *European Management Review* 12 (2): 83–98. doi:10.1111/emre.12044.

Luo, X. R., K. W. Koput, and W. W. Powell. 2009. "Intellectual Capital or Signal? The Effects of Scientists on Alliance Formation in Knowledge-intensive Industries." *Research Policy* 38 (8): 1313–1325. doi:10.1016/j.respol.2009.06.001.

Mairesse, J., and P. Mohnen. 2010. "Using Innovation Surveys for Econometric Analysis." *Handbook Economics Innovation* 2 (1): 1129–1155.

Mañez, J. A., M. E. Rochina-Barrachina, A. Sanchis-Llopis, and J. A. Sanchis-Llopis. 2015. "The Determinants of R&D Persistence in SMEs." *Small Business Economics* 44 (3): 505–528. doi:10.1007/s11187-014-9611-3.

March, J. G. 1991. "Exploration and Exploitation in Organizational Learning." *Organization Science* 2 (1): 71–87. doi:10.1287/orsc.2.1.71.

Martinez, C., L. Cruz-Castro, and L. Sanz-Menendez. 2016. "Innovation Capabilities in the Private Sector: Evaluating Subsidies for Hiring S&T Workers in Spain." *Research Evaluation* 25 (2): 196–208. doi:10.1093/reseval/rvv035.

Martinez, C., and S. Parlane. 2018. *On the Firms' Decisión to Hire Academic Scientists*. Madrid: Instituto de Políticas y Bienes Públicos (IPP) CSIC.

McMillan, G. S., and P. Thomas. 2005. "Financial Success in Biotechnology: Company Age versus Company Science." *Technovation* 25 (5): 463–468. doi:10.1016/j.technovation.2004.10.009.

Michelino, F., A. Cammarano, E. Lamberti, and M. Caputo. 2015. "Knowledge Domains, Technological Strategies and Open Innovation." *Journal of Technology Management & Innovation* 10 (2): 50–78. doi:10.4067/S0718-27242015000200005.

Moreira, M. J. 2003. "A Conditional Likelihood Ratio Test for Structural Models." *Econometrica* 71 (4): 1027–1048. doi:10.1111/1468-0262.00438.

Mudambi, R., and T. Swift. 2014. "Knowing When to Leap: Transitioning between Exploitative and Explorative R&D." *Strategic Management Journal* 35 (1): 126–145. doi:10.1002/smj.2097.

Murray, M. P. 2006. "Avoiding Invalid Instruments and Coping with Weak Instruments." *Journal of Economic Perspectives* 20 (4): 111–132. doi:10.1257/jep.20.4.111.

Nelson, R. R., and S. G. Winter. 1982. *An Evolutionary Theory of Economic Change*. Cambridge, Massachusetts: Belknap Press of Harvard University Press.

Nonaka, I. 1994. "A Dynamic Theory of Organizational Knowledge Creation." *Organization Science* 1 (5): 14–37. doi:10.1287/orsc.5.1.14.

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

OECD, Eurostat. 2005. *Oslo Manual: Guidelines for Collecting and Interpreting Innovation Data: The Measurement of Scientific and Technological Activities*. Paris: OECD Publishing.

Oster, E. 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics* 37 (2): 187–204. doi:10.1080/07350015.2016.1227711.

Rao, R. S., R. K. Chandy, and J. C. Prabhu. 2008. "The Fruits of Legitimacy: Why Some New Ventures Gain More from Innovation than Others." *Journal of Marketing* 72: 58–75.

Roach, M., and H. Sauermann. 2010. "A Taste for Science? PhD Scientists' Academic Orientation and Self-selection into Research Careers in Industry." *Research Policy* 39 (3): 422–434. doi:10.1016/j.respol.2010.01.004.

Rosenberg, N. 1990. "Why Do Firms Do Basic Research (With Their Own Money)?" *Research Policy* 19 (2): 165–174. doi:10.1016/0048-7333(90)90046-9.

Rothaermel, F. T., and A. M. Hess. 2007. "Building Dynamic Capabilities: Innovation Driven by Individual-, Firm-, and Network-Level Effects." *Organization Science* 18 (6): 898–921. doi:10.1287/orsc.1070.0291.

Sapsalis, E., B. van Pottelsberghe de la Potterie, and R. Navon. 2006. "Academic versus Industry Patenting: An In-depth Analysis of What Determines Patent Value." *Research Policy* 35 (10): 1631–1645.

Sauermann, H., and M. Roach. 2014. "Not All Scientists Pay to Be Scientists: PhDs' Preferences for Publishing in Industrial Employment." *Research Policy* 43 (1): 32–47. doi:10.1016/j.respol.2013.07.006.

Sauermann, H., and P. Stephan. 2013. "Conflicting Logics? A Multidimensional View of Industrial and Academic Science." *Organization Science* 24 (3): 889–909. doi:10.1287/orsc.1120.0769.

Sauermann, H., and W. M. Cohen. 2010. "What Makes Them Tick? Employee Motives and Firm Innovation." *Management Science* 56 (12): 2134–2153. doi:10.1287/mnsc.1100.1241.

Spithoven, A., and P. Teirlinck. 2010. "External R&D: Exploring the Functions and Qualifications of R&D Personnel." *International Journal of Innovation Management* 14 (6): 967–987. doi:10.1142/S1363919610002969.

Stern, S. 2004. "Do Scientists Pay to Be Scientists?" *Management Science* 50 (6): 835–853. doi:10.1287/mnsc.1040.0241.

Stuart, T. E., S. Z. Ozdemir, and W. W. Ding. 2007. "Vertical Alliance Networks: The Case of University-biotechnology-pharmaceutical Alliance Chains." *Research Policy* 36 (4): 477–498. doi:10.1016/j.respol.2007.02.016.

Subramaniam, M., and M. A. Youndt. 2005. "The Influence of Intellectual Capital on the Types of Innovative Capabilities." *Academy of Management Journal* 48 (3): 450–463. doi:10.5465/amj.2005.17407911.

Subramanian, A. M., K. Lim, and P.-H. Soh. 2013. "When Birds of a Feather Don't Flock Together: Different Scientists and the Roles They Play in Biotech R&D Alliances." *Research Policy* 42 (3): 595–612. doi:10.1016/j.respol.2012.12.002.

Swift, T. 2016. "The Perilous Leap between Exploration and Exploitation." *Strategic Management Journal* 37 (8): 1688–1698. doi:10.1002/smj.2423.

Teece, D. J., G. Pisano, and A. Shuen. 1997. "Dynamic Capabilities and Strategic Management." *Strategic Management Journal* 18 (March): 509–533. doi:10.1002/(SICI)1097-0266(199708)18:7<509::AID-SMJ882>3.0.CO;2-Z.

Tegarden, L. F., W. B. Lamb, D. E. Hatfield, and F. X. Ji. 2012. "Bringing Emerging Technologies to Market: Does Academic Research Promote Commercial Exploration and Exploitation?" *IEEE Transactions on Engineering Management* 59 (4): 598–608. doi:10.1109/TEM.2011.2170690.

Teirlinck, P., and A. Spithoven. 2013. "Research Collaboration and R&D Outsourcing: Different R&D Personnel Requirements in SMEs." *Technovation* 33 (4–5): 142–153. doi:10.1016/j.technovation.2012.11.005.

Trostel, P., I. Walker, and P. Woolley. 2002. "Estimates of the Economic Return to Schooling for 28 Countries." *Labour Economics* 9 (1): 1–16. doi:10.1016/S0927-5371(01)00052-5.

Tzabbar, D. 2009. "When Does Scientist Recruitment Affect Technological Repositioning?" *Academy of Management Journal* 52 (5): 873–896. doi:10.5465/amj.2009.44632853.

Tzabbar, D., B. S. Aharonson, and T. L. Amburgey. 2013. "When Does Tapping External Sources of Knowledge Result in Knowledge Integration?" *Research Policy* 42 (2): 481–494. doi:10.1016/j.respol.2012.07.007.

Veugelers, R., and B. Cassiman. 2005. "R&D Cooperation between Firms and Universities. Some Empirical Evidence from Belgian Manufacturing." *International Journal of Industrial Organization* 23 (5–6): 355–379. doi:10.1016/j.ijindorg.2005.01.008.

Walrave, B., K. E. van Oorschot, and A. G. L. Romme. 2015. "How to Counteract the Suppression of Exploration in Publicly Traded Corporations." *R&D Management* 45 (5): 458–473. doi:10.1111/radm.12094.

Wooldridge, J. M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT press.

Yogo, M. 2004. "Estimating the Elasticity of Intertemporal Substitution When Instruments are Weak." *Review of Economics and Statistics* 86 (3): 797–810. doi:10.1162/0034653041811770.

Zellner, C. 2003. "The Economic Effects of Basic Research: Evidence for Embodied Knowledge Transfer via Scientists' Migration." *Research Policy* 32 (10): 1881–1895. doi:10.1016/S0048-7333(03)00080-5.

Appendix 1. Definition of variables and correlations

Table A1a. Definition of variables.

Variable	Definition
Start_upstream	Dummy variable that takes the value of 1 if the firms perform upstream-oriented R&D activities in t + 1 and zero if it does not (only for downstream-oriented firms in t)
Lsphd	Percentage of employees with a PhD degree relative to the number of employees in the R&D function (in logs)
Lsize	Firms' employee count (in logs)
D_intensity	Firm's total development 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 function (in logs)
Parent	Dummy variable that takes value 1 if the firm is the parent company 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

Table A1b. Correlation matrix (Downstream-oriented firms in t, n = 5,815).

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	Start_upstream																
2	Lsphd	0.05*															
3	Lsize	−0.02*	0.00														
4	D_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

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, Harrison, and Simpson (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 10 research fields: biology, chemistry, physics, computer science, material science, medical and health science, chemical engineering, electrical engineering, mechanical engineering and mathematics. As in Abramovsky, Harrison, and Simpson (2007), we consider a research field 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/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176744&menu=resultados&idp=1254735573113

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.[16]

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 stability. Consider two firms (A and B) that are equal in all observable characteristics except that 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 shows that only 36.9% of PhDs work in another region than their birth region,[17] suggesting that mobility is far from perfect.

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:

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

where :

β^{F} *is the coefficient in the regression with full controls*

R^{F} *is the* R^2 *in the corresponding regression*

β^{W} *is the coefficient in the regression without controls (we used just time dummies)*

R^{W} *is the* R^2 *in the corresponding regression*

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

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

The method can be implemented using the Stata package: psacalc

[16]The indicator is normalised by the total number of R&D employees in each region.

[17]Ideally, we would want to know the percentage of PhDs working in the same region as the institution that awarded their degree; however, these data are not available. Presumably, the figure would be lower than 36.9% because some individuals move regions before beginning their PhD studies.

Appendix 4. Robustness check: whole sample of R&D performers and orientation to upstream R&D

Table A4. Orientation to upstream R&D and PhDs. Results.

	(1)	(2)	(3)	(4)	(5)	(6)
	SimpleOLS	Multiple OLS	RE	FE	IV	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]
D_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. ***p-value<0.01; **p-value<0.05, *p-value<0.1

Appendix 5. Robustness check: a different weigthing matrix

Table A5. A different weighing matrix for the instrumental variable. Results.

	(1)	(2)	(3)	(4)
	IV-GMM	IV-GMM Wool	IV Probit	IV Probit Wool
Lsphd	0.121***	0.082***	0.388***	0.287***
	[0.032]	[0.028]	[0.108]	[0.086]
Lsize	−0.004	−0.004	−0.014	−0.016
	[0.009]	[0.009]	[0.033]	[0.033]
D_intensity	−0.012	−0.011	−0.042	−0.040
	[0.008]	[0.008]	[0.028]	[0.029]
Export	0.014	0.015	0.053	0.059
	[0.018]	[0.017]	[0.067]	[0.067]
Lsizeteam	−0.007	−0.007	−0.034	−0.028
	[0.010]	[0.010]	[0.039]	[0.039]
Parent	0.012	0.018	0.056	0.069
	[0.024]	[0.022]	[0.084]	[0.083]
Joint_venture	0.069	0.069	0.222	0.236
	[0.059]	[0.057]	[0.185]	[0.182]
Newmer	0.000	0.000	0.001	0.001
	[0.000]	[0.000]	[0.001]	[0.001]
Obstacle_funds	0.026**	0.025**	0.100**	0.102**
	[0.013]	[0.012]	[0.050]	[0.050]
Appropriability	−0.066	−0.067	−0.263	−0.276
	[0.046]	[0.045]	[0.185]	[0.185]
Mediumlow	−0.022	−0.022	−0.077	−0.083
	[0.017]	[0.017]	[0.065]	[0.065]
Mediumhigh	−0.037**	−0.031*	−0.120*	−0.120*
	[0.019]	[0.018]	[0.069]	[0.069]
High	−0.043	−0.039	−0.169*	−0.150
	[0.028]	[0.027]	[0.102]	[0.101]
Park	0.060	0.038	0.125	0.136
	[0.042]	[0.039]	[0.136]	[0.134]
Lage	−0.015	−0.016*	−0.057	−0.063*
	[0.010]	[0.009]	[0.035]	[0.035]
Pubfun	−0.008	−0.005	−0.023	−0.015
	[0.012]	[0.011]	[0.044]	[0.044]
_cons	0.477***	0.486***	0.276	0.311
	[0.159]	[0.156]	[0.618]	[0.622]
N	5,815	5,815	5,815	5,815

Coefficients reported with clustered standard error between brackets. All models include year dummies.***p-value<0.01, **p-value<0.05, *p-value<0.10

Appendix 6. Robustness check: transition to basic research

Table A6. Transition to basic research Results.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	SimpleOLS	Multiple OLS	IV	IV Wool	Probit	IV Probit	IV Probit Wool
Lsphd	0.005***	0.004**	0.025***	0.019***	0.038**	0.220***	0.157***
	[0.002]	[0.002]	[0.006]	[0.007]	[0.015]	[0.050]	[0.049]
Lsize		−0.007**	−0.007**	−0.007**	−0.067**	−0.078***	−0.075**
		[0.003]	[0.003]	[0.003]	[0.029]	[0.029]	[0.029]
RD_intensity		−0.005**	−0.007***	−0.007***	−0.055**	−0.073***	−0.068***
		[0.002]	[0.002]	[0.002]	[0.025]	[0.025]	[0.025]
Export		−0.001	−0.002	−0.001	−0.011	−0.020	−0.017
		[0.006]	[0.006]	[0.006]	[0.058]	[0.057]	[0.058]
Lsizeteam		0.011***	0.009***	0.010***	0.107***	0.098***	0.101***
		[0.003]	[0.003]	[0.003]	[0.033]	[0.033]	[0.033]
Parent		0.011	0.005	0.007	0.094	0.046	0.060
		[0.007]	[0.007]	[0.007]	[0.064]	[0.064]	[0.064]
Joint_venture		0.047*	0.046*	0.047*	0.357**	0.343**	0.350**
		[0.024]	[0.025]	[0.025]	[0.145]	[0.152]	[0.149]
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		0.001	0.002	0.002	0.012	0.015	0.014
		[0.004]	[0.004]	[0.004]	[0.044]	[0.043]	[0.044]
Appropriability		−0.005	0.003	−0.001	−0.031	0.046	0.021
		[0.016]	[0.015]	[0.016]	[0.158]	[0.160]	[0.159]
Mediumlow		−0.001	0.001	0.000	−0.013	0.007	−0.000
		[0.005]	[0.005]	[0.005]	[0.056]	[0.056]	[0.056]
Mediumhigh		0.006	0.007	0.006	0.060	0.068	0.068
		[0.006]	[0.006]	[0.006]	[0.058]	[0.058]	[0.058]
High		0.000	−0.004	−0.002	0.014	−0.022	−0.009
		[0.008]	[0.009]	[0.009]	[0.082]	[0.084]	[0.084]
Park		0.007	0.003	0.005	0.059	0.035	0.041
		[0.013]	[0.013]	[0.013]	[0.113]	[0.113]	[0.113]
Lage		−0.001	−0.001	−0.001	−0.010	−0.009	−0.009
		[0.003]	[0.003]	[0.003]	[0.031]	[0.030]	[0.030]
Pubfun		0.001	0.000	0.000	0.008	−0.002	0.002
		[0.004]	[0.004]	[0.004]	[0.037]	[0.037]	[0.037]
_cons	0.036***	0.104**	0.087*	0.101*	−1.144**	−1.210**	−1.190**
	[0.004]	[0.052]	[0.052]	[0.052]	[0.536]	[0.539]	[0.535]
N	15,165	15,165	15,165	15,165	15,165	15,165	15,165

Coefficients reported with clustered standard error between brackets. All models include year dummies.***p-value<0.01, **p-value<0.05, *p-value<0.10