Firm RD units and outsourcing partners: A matching story

Barge-Gil, Andrés and Conti, Annamaria

Universidad Complutense, Department of Quantitative Economics, College of Management, Georgia Institute of Technology

May 2013

Firm R&D Units and Outsourcing Partners: a Matching Story

Andrés Barge-Gil *and Annamaria Conti †

May 31, 2013

Abstract

We present a theory that examines the optimal match between firm R&D units and external partners for projects that involve problem solving. We have a firm selecting an external partner conditional on the learning costs of its internal R&D unit. We show that there exists a matching equilibrium with property that external partners with low learning costs for a project work with R&D units that also have low learning costs for the same project. Empirically, we use a dataset of Spanish R&D firms and relate their share of R&D outsourcing to universities to the composition of their R&D units, described by the presence of staff with a PhD. Our main finding is that, controlling for endogeneity, firms that employ R&D staff with a PhD outsource relatively more to universities than to firms. We interpret this result as evidence that R&D units with relatively low learning costs for basic projects tend to match with external partners, universities, with relatively low learning costs for the same projects.

*Department of Foundations of Economic Analysis (Quantitative Economics), Complutense University, Campus de Somosaguas 28223, Pozuelo de Alarcón, Madrid, Spain, abarge@ccee.ucm.es

†College of Management, Georgia Institute of Technology, 800 West Peachtree Street NW Atlanta, Georgia 30308, USA, annamaria.conti@mgt.gatech.edu

1 Introduction

The importance of external sources of knowledge for firm innovation has been extensively examined (Hagedoorn, 2001; Thursby and Thursby, 2006; Branstetter, 2005; Alcacér and Chung, 2007). Hagedoorn (2001) documents a steady increase in inter-firm R&D partnerships during the 1980s and the 1990s. Branstetter (2005), Thursby and Thursby (2006), and Alcacér and Chung (2007) provide evidence that firms increasingly rely on input provided by universities and that the quality of this input determines the countries in which firms choose to locate their R&D activities.

The existing literature has focused on the role that the absorptive capacity of firms plays in their choice between conducting their R&D activities internally or in collaboration with external partners (Cohen and Levinthal, 1989; Cassiman and Veugelers, 2002). Other studies have focused on the importance of incoming, source-specific, spillovers for firms' decisions regarding the partners they collaborate with in their R&D projects (Cassiman and Veugelers, 2002; Belderbos et al., 2004; Veugelers and Cassiman, 2005). Another important aspect of firm R&D collaboration that has received little attention is the optimal match between a firm's internal R&D unit and a potential partner on an R&D project, conditional on some characteristics of both partners.

Our analysis intends to fill this gap by extending the theory of hierarchical sorting, developed by Garicano (2000), Garicano and Hubbard (2005), and Garicano and Rossi-Hansberg (2006), to the optimal match between a firm's internal R&D unit and an external partner for projects that involve problem solving and that require the active participation of the internal R&D unit. We test the prediction of our theory using a rich database on R&D in Spanish manufacturing firms.

In the theory, a firm has to choose a project from a menu of projects that differ in the level of some characteristic. We examine projects that are complex enough to require the firm's internal R&D unit to work in collaboration with an external partner. The advantage is that the external partner has a lower cost of learning to solve problems that are relevant to the project and thus can solve a larger proportion of problems than a firm's R&D unit. As we consider a situation of sticky labor, and thus the composition of the internal R&D unit is fixed, the choice of the project depends on the learning cost of the internal R&D unit. We show that there exists a matching equilibrium with property that, for a given project, external partners with low learning costs for that project work with R&D units that also have low learning costs for the same project. This allows the external partner to relatively specialize in the solutions to less common problems, while the internal R&D unit relatively specializes in solutions to the most common ones. This result is interesting in that it suggests that, for joint R&D projects that involve problem solving, it is not profitable for a firm to entirely outsource the solution of problems to an external partner, and have the internal unit specialize on the application of these solutions. On the contrary, firm profits are maximized when also the internal unit actively participates in the solution of problems and when it does so at the lowest possible learning costs. In this way, the external partner avoids spending time assisting the internal R&D unit and dedicates its limited time to solving the least common problems. While our model is an extension of the theory of hierarchical sorting within an organization developed by Garicano (2000), Garicano and Hubbard (2005), and Garicano and Rossi-Hansberg (2006), it differs in two important aspects. First, our model deals with the optimal match between an internal firm's unit and a partner, *outside of the firm's organization*. Yet, even in this case we show that an optimal matching occurs, along the lines of the learning costs of the firm's unit and the partner. Second, we examine a situation with sticky labor and, thus, in our case, the choice of the external partner is conditional on the project a firm has selected, which has to reflect the (invariant) learning cost structure of the internal R&D unit. Finally, relative to the theory of absorptive capacity, our model proposes a complementary explanation of the mechanisms that allow a firm to extract a value from its collaborations. In their seminal paper, Cohen and Levinthal (1990) suggest that the value to a firm from a collaboration increases in a firm's level of investment in those assets that are specific to that collaboration. In our model, such type of investment decreases the learning costs for the problems the partners have to solve, and with that it increases the value of the collaboration to the firm, *relative* to other types of collaborations the firm can engage in.

We test the theoretical model using information from a sample of Spanish manufacturing firms that conducted innovation activities during 2006-2009. We estimate a model in which we relate the share of firms' R&D outsourcing to universities with the compositions of their R&D units, described by the presence of staff with a PhD degree. The logic here is that both universities and R&D units with PhD staff face lower costs in conducting relatively more basic projects than either companies to which a firm may want to outsource or staff who lack PhD degrees (see for example Cockburn and Henderson, 1998; Aghion et al., 2008; Lacetera, 2009; Gruber et al., 2012). We find evidence that R&D units with PhD staff tend to collaborate relatively more with universities than with other partners, even after controlling for endogeneity in the composition of the internal unit. We interpret this result as evidence that R&D units with relatively lower learning costs for basic projects tend to match with external partners, universities, with relatively lower learning costs for the same projects. In contrast to previous empirical studies on incoming spillovers (Cassiman and Veugelers, 2002; Belderbos et al., 2004) and on firms' internal R&D intensity (Leiponen, 2005; Lopez, 2008), our results are novel in that they focus on the composition of firms' R&D units rather than on the information they can access externally or internally. Moreover, while previous studies have examined the role of incoming spillovers on the *likelihood* of collaboration with different categories of external partners, we focus on the *relative importance* of external partners to a firm R&D unit. We argue that the pattern of external collaborations reflects an optimal match of internal R&D units and external partners along the lines of their learning costs for the projects they work on. Finally, in contrast to the work by Cockburn and Henderson (1998), which examines the correlation between a firm's degree of connection with universities and the firm's patent productivity, we look at a firm's decision regarding the partners it collaborate with in its R&D projects, conditional on the composition of its R&D unit.

The remainder of the paper is organized as follows. Section two introduces the model. Section three presents an empirical estimation of the theory. Section four concludes.

2 A matching model of outsourcing

In this section, we present a model in which the manager of an R&D firm needs to select a project from a menu of projects that differ in some characteristic. Our focus is on projects that involve problem solving. In this model, labor is sticky and the firm's project choice depends on the knowledge of its internal R&D unit, the composition of which does not change. This analysis is appropriate to analyze the short-run decisions of firms, especially in contexts, in which the labor market is characterized by important rigidities, such as that in Spain.

Once the firm's manager has made her choice, she needs choose an external partner to collaborate with the internal R&D unit on the project. We examine firms' R&D projects that are complex enough to require the support of an external partner. The fundamental problem facing the firm consists of finding the optimal match between an internal R&D unit and an external partner, according to some characteristics of both. This study is based on models developed by Garicano (2000), Garicano and Rossi-Hansberg (2006), and Garicano and Hubbard (2005) and extends them to the optimal match between firm R&D units and external partners to which the firm outsources a component of its R&D projects.

Intuitively, the advantage of hiring an external partner is that the partner faces lower costs to solve the problems that are generated from a project and, therefore, can solve a larger range of problems than a firm's R&D unit. The comparative advantage at solving problems of the external partner derives either from the experience it has accumulated from working on similar types of projects in the past or because it has either the necessary instrumentation or the appropriate incentives to solve problems at relatively low costs.

2.1 Model setup

A firm needs to choose a project, p, from a continuum of projects, $[\underline{p}, \overline{p}]$. Projects differ in a characteristic, and they are ordered according to increasing levels of this characteristic. Hence, $\underline{p}$ is endowed with the lowest level of the characteristic, and $\overline{p}$ with the highest. The firm can only choose one project. This project involves problem solving. As in Garicano (2000), Garicano and Rossi-Hansberg (2006), and Garicano and Hubbard (2005), problems differ in that some of them occur more frequently than others. The cumulative density function of problems is $F(z)$, with higher values of z being associated with rarer problems. Thus, $f\prime(z) < 0$.

A firm's R&D unit divides its time between learning how to solve problems and producing. We assume that the unit spends an amount of time t on production and the remaining time learning how to solve problems. Moreover, for every unit of time the R&D unit spends on production it encounters a problem. We assume that the cost of learning to solve problems on the interval $[0, z_i]$, expressed in units of time, is proportional to the length of the interval and equal to $c_i z_i$. Thus, if an internal R&D unit spends $c_i z_i$ units of time learning how to solve problems, up to z_i, it will dedicate the remaining time $(1 - c_i z_i)$ to producing. The R&D unit can solve a proportion of problems equal to $F(z_i)$ and requires external assistance for the remaining $(1 - F(z_i))$. Essentially, the unit can solve the most common problems and seeks assistance for the less common ones.

The firm needs to pay the external partner a payment, K. The market of external partners is competitive, and, thus, each partner is paid an amount, K, which is equal to the opportunity cost of participating in the firm's project. We assume that the opportunity cost is the same for all external partners. The logic here is that the external partner specializes in solving certain types of problems, and it can either use its knowledge for the project that the firm is interested in working on or for other projects with other partners. During the time that the partner works with the firm's R&D unit, which we normalize to 1, the partner learns how to solve problems on the interval $[0, z_e]$, at a cost $c_e z_e$ and provides assistance to the firm's unit for the problems that the latter cannot solve. The learning costs of the external partner have to be lower than those of the internal R&D unit, $(c_e < c_i)$, otherwise the firm would not find it profitable to hire an external partner to work on a given project. We rule out the possibility that the external partner might spend time producing problems relevant for a firm by assuming that the time cost of producing one problem faced by the external partner is greater than one. The logic here is that the comparative advantage of the external partner lies in problem solving and not in production. This assumption rules out the possibility that a project might be entirely outsourced to an external partner. Similar to Garicano (2000), Garicano and Rossi-Hansberg (2006), and Garicano and Hubbard (2005), assisting the R&D unit involves a time cost, h, with $h < 1$[1], that is incurred by the external partner whenever it assists the firm's R&D unit. Thus, even if the external

[1] If assisting an internal unit were to require an amount of time greater than or equal to 1, the external partner would not have enough time to learn to solve problems that are relevant for a project.

partner does not know the solution to a problem, as this lies within the interval $(z_e, 1]$, it still incurs the cost h, which we can view in this case as the partner's cost of determining whether it knows the solution to the problem. This cost is the same for all external partners.

The time constraint of the external partner is:

$$1 = c_e z_e + h(1 - F(z_i)) \tag{1}$$

Given the time constraint of the external partner, the lower the proportion of problems that the firm's R&D unit can solve, the greater the time the external partner has to spend assisting the unit, and the less time it has left to learn the solutions to problems that are relevant for a project. In this model knowledge is cumulative (Griliches, 1979), and thus, even if the external partner assists the firm's unit to solve problems in the interval $(z_i, z_e]$, it still has to learn how to solve problems in the interval $[0, z_i]$. This setting corresponds to a situation in which the external partner needs to spend some time to acquire the knowledge the firm's R&D unit has accumulated on a project before it is able to assist the unit in expanding its knowledge frontier.

From the external partner's time constraint (1), we derive an expression for z_e, which is equal to:

$$z_e = \frac{1 - h(1 - F(z_i))}{c_e} \tag{2}$$

The external partner's fraction of problems it can solve, which corresponds to its knowledge level, is an increasing function of the knowledge level of the firm's R&D unit, z_i. The greater z_i is, the lower the range of problems for which the R&D unit requires external assistance, and the greater the time the external partner can dedicate to learning how to solve rarer problems. Moreover, z_e is decreasing in c_e. This is because the larger c_e is, the greater the cost of learning how to solve problems in a given range, and the smaller the range of problems solved is.

2.2 The manager's choice of an internal R&D unit and the external partner

The relevant decisions for a firm can be described as follows. The firm selects a project, p, from the interval $[\underline{p}, \overline{p}]$. It then selects an external partner that will collaborate with the internal R&D unit on the project, such that their match maximizes the firm's profits. The match between the firm's R&D unit and the external partner occurs with respect to their learning costs, c_i and c_e, for the project the firm has selected. Conditional on the learning costs of the internal

R&D unit and the resulting match with the external partner, the firm chooses the knowledge levels of the R&D unit and the external partner, z_i and z_e, that maximize the firm's profits. The choice of the internal unit's knowledge level, z_i, has to be such that, for given values of c_i, c_e and h, $z_e > z_i$. If this were not the case, the firm would not find it profitable to hire an external partner.

The firm's revenue, R, derives from commercializing the output that is generated from a project. Production occurs any time that either the internal R&D unit or the external partner knows the solution to a problem. Thus, the firm's revenue is equal to $F(z_e)(1 - c_i z_i)$, given that z_i is a subset of z_e. The inputs z_i and z_e are substitutes in the firm's production function because $\frac{\partial R}{\partial z_i \partial z_e}$ < 0. This is because R depends on the proportion of problems that are solved *and* the resulting production. However, because the internal unit has to divide its (limited) time between solving problems and producing, the firm faces the following trade-off. The more time the internal unit spends solving problems, which allows the external partner to concentrate on the least common ones and increases the total fraction of problem solved, the fewer problems end up being produced. We shall restrict our attention to the case in which the magnitude of substitution is sufficiently small and, specifically, smaller than $\frac{F(z_e)ci}{z_e}$. For the region of the parameters such that this condition holds, the fraction of problems solved by the external and the internal unit is decreasing in c_e. The alternative would be that increasing values of c_e lead the internal unit to make up for the high learning costs of the external partner by solving an higher fraction of problems, and hence reducing the time the external partner has to spend assisting the internal unit. We rule out this alternative as we (reasonably) assume that the primary goal of hiring an external partner is to ease the task of the internal unit, without the latter having to make up for the deficiencies of the first.

The costs the firm has to incur are represented by *i)* the payment owed to the R&D unit $w = w(\bar{c})$, and *ii)* the payment owed to the external partner that takes the form of a lump-sum payment equal to K. The payment, $w(\bar{c})$, the firm offers to the internal R&D unit is a function of the R&D unit's average learning costs for the projects it is required to work on during its life-cycle. Because we are interested in a firm's projects that require some level of outsourcing, we posit that K is sufficiently small to ensure that the firm always finds it profitable to outsource part of the project. Relaxing this assumption would imply that the firm's project choice takes into account the cost of hiring an external partner. Moreover, we assume that K is at least equal to the compensation that the internal R&D unit with the lowest learning cost for p would receive. This assumption rules out the possibility that the external partner might find it optimal to be hired by the firm.

We define the firm's maximization problem as:

$$\max_{z_i, z_e} \Pi = F(z_e)(1 - c_i z_i) - w - K$$

Substituting for z_e, we rewrite:

$$\max_{z_i} \; F(\frac{1-h(1-F(z_i)}{c_e})(1-c_i z_i) - w - K$$

The first order condition is:

$$\frac{\partial \Pi}{\partial z_i} = -F(z_e)c_i + \frac{h}{c_e} f(z_e) f(z_i)(1-c_i z_i) = 0 \tag{3}$$

The pair (z_i^*, z_e^*), derived from the firm's maximization problem, is a local maximum given that $\frac{\partial^2 \Pi}{\partial z_i^2} < 0$.

2.3 Characterization of a matching equilibrium

In this section, we characterize the properties of a matching equilibrium. To this end, we allow for a continuum of firms distributed across projects $[\underline{p}, \overline{p}]$. Moreover, for a given project, p, the density function for the R&D unit's cost of learning problems relevant for project p is $\phi(c_i(p))$ with support $[\underline{c}_i(p)$,$\overline{c}_i(p)]$. Similarly, the density function for the external partners' cost of learning the problems relevant to p is $\varphi(c_e(p))$ and its support is $[\underline{c}_e(p)$,$\overline{c}_e(p)]$.

The first equilibrium condition that needs to be met is that, for every project, $p \in [\underline{p}, \overline{p}]$, the number of external partners that firms find it optimal to hire, $n_D(p)$, be equal to the number of external partners that find it optimal to work for on p, $n_O(p)$.

$$n_D(p) = n_O(p)$$

With $n_O(p) \geq 2$ to ensure that the external partner's market is competitive. The second condition is that the firm should not find it profitable to deviate from the choice of the external partner it has been tentatively assigned. Computing the partial derivative of the optimized profits Π, generated from project p, with respect to $c_e(p)$, we obtain:

$$\frac{\partial \Pi(p)}{\partial c_e(p)} = -\frac{f(z_e)z_e}{c_e}(1 - c_i(p)z_i(p)) \tag{4}$$

Because the sign of the derivative is negative, the optimal choice of the manager consists of hiring an external partner that can provide $z_e^*(p)$ at the lowest learning costs for that project. That is: $c_e^*(p) = \underline{c}_e(p)$. Thus, for $c_e^*(p) = \underline{c}_e(p)$, the firm does not find it profitable to deviate from its choice of external partner.

The third condition is that the project the firm chooses, p, is such that the learning costs of its internal R&D unit for that project maximize the firm's profit. Computing the partial derivative of the optimized profits, $\blacksquare$, with respect to $c_i(p)$, we obtain:

$$\frac{\partial \Pi(p)}{\partial c_i(p)} = -F(z_e(p))z_i(p) - \frac{dw}{dc_i(p)} \tag{5}$$

Where, $\frac{dw}{dc_i(p)} < 0$. The choice of the project has to be such that expression (5) is equal to zero and $\frac{\partial^2 \Pi(p)}{\partial c_i(p)^2} < 0$.

At the optimum, the firm's optimal project choice, p, has to satisfy the following condition: $\frac{d\Pi(p)}{dc_i(p)} = 0$. If we take the full derivative of this condition, we obtain:

$$\frac{\partial^2 \Pi(p)}{\partial (c_i(p))^2} + \frac{\partial^2 \Pi(p)}{\partial c_e(p) \partial c_i(p)} \frac{dc_e(p)}{dc_i(p)} = 0$$

This implies that:

$$\frac{dc_e(p)}{dc_i(p)} = -\frac{\frac{\partial^2 \Pi(p)}{\partial (c_i(p))^2}}{\frac{\partial^2 \Pi(p)}{\partial c_i(p) \partial c_e(p)}}$$

The sign of $\frac{dc_e(p)}{dc_i(p)}$ depends on the sign of the cross-partial $\frac{\partial^2 \Pi(p)}{\partial c_i(p) \partial c_e(p)}$, given that by the second order condition $\frac{\partial^2 \Pi(p)}{\partial (c_i(p))^2} < 0$. The expression for $\frac{\partial^2 \Pi(p)}{\partial c_i(p) \partial c_e(p)}$ is:

$$\frac{\partial^2 \Pi(p)}{\partial c_i(p) \partial c_e(p)} = -f(z_e(p))z_i(p)[\frac{dz_e(p)}{dc_e(p)} + \frac{h}{c_e(p)} \frac{dz_i(p)}{dc_e(p)}] - F(z_e(p)) \frac{dz_i(p)}{dc_e(p)} \tag{6}$$

This results in the following proposition regarding the matching between internal R&D units and external partners.

Proposition 1 *There exists an optimal matching equilibrium such that for a given project, p, both the internal R&D unit and the external partner have the lowest learning costs for that project.*
Proof. *See Appendix.* ■

The result in Proposition 1 shows that it is optimal for a firm that the internal R&D unit and the external partner have a similar comparative advantage in problem solving. This is because it allows the external partner to spend relatively more time solving rarer problems than assisting the internal R&D unit. Hence, the external partner relatively specializes in solving rarer problems, while the internal R&D unit relatively specializes in solving the most common ones.

We have considered a model with sticky labor. This model is a good fit for our Spanish panel dataset, which is characterized by low levels of intra-firm variation in the composition of R&D units. It also fits well with any short-run analysis of a firm's project decision. However, one might wonder whether the results in Proposition 1 also apply to a case in which firms' R&D employees are allowed to move across firms or from a firm's R&D unit to what we have defined as an "external partner". In such a scenario, it is plausible that the firm would select a project that guarantees the highest expected return and choose the composition of the internal R&D unit accordingly. Given expression (4), the firm would still find it profitable to hire an external partner with the lowest learning costs for a project. However, it would now hire an internal R&D unit such that, given the project chosen, its learning costs would ensure that expression (5) equals zero. The R&D employees would either decide to work for firm or be an "external partner", based on the expected compensation they are offered. Finally, the firm would still opt for an internal R&D unit and an external partner that have the lowest learning costs for the project the firm has chosen, provided that the substitution between their knowledge levels is small.

3 Empirical estimation

In the theory section we have shown that the learning costs of an internal R&D unit determine a firms' choice of the project the R&D unit will be working on and, consequently, the type of external partner the unit will work with. Our main result is that for each firm's project, the optimal match between an internal R&D unit and an external partner is such that both the internal unit and the external partner have the lowest learning costs for that project.

In this section we analyze the optimal choice of the firm regarding the type of external partner it works with, conditional on the characteristics of the internal R&D unit to which an R&D project has been assigned. To this end, we employ a rich dataset on Spanish R&D manufacturing firms that conducted some amount of R&D outsourcing. We relate these firms' shares of R&D outsourcing to universities to the composition of their internal R&D units, which we define by the presence of staff with a PhD degree. The logic here is that i) universities should face lower costs in conducting relatively more basic projects than private companies to which a firm might want to outsource; and ii) R&D units with PhD staff should face lower costs in conducting relatively more basic projects than R&D units with staff that lack PhD degrees (see for example Cockburn and Henderson, 1998; Aghion et al., 2008; Lacetera, 2009; Gruber et al., 2012). Hence, for joint R&D projects, if the internal R&D units and their external partners indeed face the trade-offs we have described in the theory, then *ceteris paribus* we should observe that R&D units with PhD staff outsource relatively more to universities than to other companies.

3.1 Description of the dataset

The empirical analysis employs information from the Spanish Technological Innovation Panel (PITEC), which is a statistical instrument designed to study the evolution of the innovation activities of Spanish firms over time. PITEC is a panel survey conducted by the Spanish National Statistics Institute. It is part of a European-wide project known as the Community Innovation Surveys (CIS) that consists of a series of surveys that are administered in every European Union country, including Norway and Iceland, with the goal of collecting information on the innovation activities of firms located in these countries. The information provided by these surveys has been extensively used in the economic literature to address innovation-related research questions (Cassiman and Veugelers, 2002; Raymond et al., 2010; Czarnitzki and Toole, 2011).

PITEC contains information at the firm level from 2003 to 2009, including firms with 200 or more employees and firms with intramural R&D expenditures[2]. For the purposes of our study, we are particularly interested in the information available on firms' R&D outsourcing, including the amount firms have spent on outsourcing activities to universities or to other firms, and in the composition of the firms' R&D units. For this latter aspect, we use information on the share of full-time equivalent workers that hold PhD degrees.

While PITEC was first conducted in 2003, we only use data for the 2006-2009 period because prior to 2006 the number of R&D personnel was not described in terms of full-time equivalent units, and thus it is not possible to compare the composition of R&D unit before and after 2006. As in previous studies that have used CIS survey data (see for instance Cassiman and Veugelers, 2002; Raymond et al., 2010; Czarnitzki and Toole, 2011), we restrict our analysis to innovating firms that operated in the manufacturing sector. Our definition of innovating firms encompasses firms that continuously employed R&D personnel during 2006-2009. Because the focus of our analysis is on firms that have conducted at least some R&D outsourcing on a given year, we further restrict the sample to these firms and denote the corresponding sample sample A. This sample comprises 4,191 firm-year observations and 1,457 firms. As shown in Table 1, sample A decreases from 1,109 firms in 2006 to 984 firms in 2009, due to a reduction in R&D outsourcing by firms. As a check for robustness, we also conduct the analysis at the level of innovating firms, without restricting the analysis to those that on a given year had engaged in outsourcing. We denote the corresponding sample sample B. Sample B is a balanced panel of 2,352 firms and 9,408 firm-year observations. The difference in the number of firms between sample A and sample B is because 895 firms did not engage in outsourcing during our period of interest.

The average annual amount spent on outsourcing by firms in sample A is 641,132 (2006) Euros. The average share of outsourcing to universities is 17.5%; the remaining share of firm outsourcing is mainly directed to other firms. The percentage of firms that employ R&D personnel with PhD degrees is 26.5%; and among these firms, the average percentage of employees with PhD degrees

[2] The database and additional information are available at http://icono.fecyt.es/contenido.asp?dir=05)Publi/AA)panel.

is 21. In sample B, 20% of the firms employ R&D personnel with PhD degrees; and among these firms, the average percentage of employees with PhD degrees is 25. Not surprisingly, the variability in the PhD composition of firm R&D units is higher across firms than within firms. In sample A, the across-firm standard deviation is 0.39, while the within-firm standard deviation is 0.18. In sample B, the corresponding figures are 0.35 and 0.17, respectively. Moreover, in samples A and B, the within-firm variation is generated by only 5% of the sample observations. As expected, the highest concentration of staff with PhD degrees is in the pharmaceutical sector: 73% of pharmaceutical firms in sample A and 80% of pharmaceutical firms in sample B employ R&D personnel with PhD degrees. The chemistry sector is next with 36% of the firms in sample A and 43% of firms in sample B employing R&D personnel with PhD degrees.

⟨ Insert Table 1 about here ⟩

3.2 Econometric methodology

To test whether the composition of a firm's R&D unit affects the share of firm R&D outsourcing to universities, we estimate a model that relates the share of firm R&D outsourcing to universities to the composition of their internal R&D units and a number of controls. The equation we estimate is:

$$Out\_UNI_{it} = \beta_0 + \beta_1 PhD_{it} + X'_{it}\gamma + \varepsilon_{it} \tag{1}$$

where Out_UNI_{it} is the share of a firm's i R&D outsourcing to universities (in logs) in year t. The variable PhD_{it} is a dummy that takes the value of one if the R&D unit employs at least one employee with a PhD degree. We also estimate an alternative specification of equation 1 that considers the share of R&D staff that holds a PhD degree (PhD_{it}_Share) in logs. Finally, X_{it} is a matrix of controls that we believe may impact the direction of a firm's outsourcing.

$$Out\_UNI_{it} = \beta_0 + \beta_1 PhD_{it}\_Share + X'_{it}\gamma + \upsilon_{it} \tag{2}$$

Specifically, X_{it} includes firm size (*Size*), which we measure by the log of a firm's number of employees, and firm R&D intensity (*R&D Intensity*), which we define as the log of a firm's total R&D expenditures per employee[3]. These controls have been extensively used in the economic literature (see for instance Cohen and Levinthal, 1989 and Cassiman and Veugelers, 2002) as proxies for the *absorptive capacity* of firms, and they have been shown to affect R&D outsourcing by firms. Additionally, we control for the size of the R&D unit, *Size R&D unit*, with a count of the full time equivalents working in an R&D unit. We also include a dummy that takes a value of one if a firm sells its products overseas (*Export*); a dummy that takes a value of one if a firm is foreign-owned (*Foreign*); a dummy that takes a value of one if a firm is a parent company

[3]Details on the construction of the variables are presented in Appendix B.

(*Parent*); and a dummy that takes a value of one if a firm is the product of a joint venture (*Joint_venture*). These dummies are meant to capture factors that might affect the outsourcing decisions of a firm.

Following Cassiman and Veugelers (2002), we control for a firm's obstacles to innovation. We distinguish between four types of obstacles. The first is a lack of funds available for innovation. In the Spanish survey, firms rated the following options on a scale from one (very important) to four (unimportant): i) lack of funds within the firm or group and ii) a lack of funds from sources outside the firm. To measure the first type of obstacle, we use a dummy that takes a value of one if a firm attached moderate or high importance to either i) or ii). We denote the corresponding variable *Obstacle_funds*. The second obstacle to innovation we control for refers to a lack of qualified personnel assigned to innovation activities. As a measure, we include a dummy that takes a value of one if a firm responded that a lack of qualified personnel was of moderate or high importance. We denote the corresponding variable *Obstacle_qualified*. The third obstacle to innovation is a lack of information. In the survey, using the same scale as above, firms rated the importance they attached to the i) lack of information about technology; ii) lack of information about markets. We include a dummy that the takes a value of one if a firm responded that either i) or ii) was of moderate or high importance. We denote the corresponding variable *Obstacle_information*. Finally, the fourth obstacle we consider is the difficulty of finding partners for innovation projects. We include a dummy that the takes a value of one if a firm responded that this obstacle was of moderate or high importance. We denote the corresponding variable *Obstacle_partner*. The rationale for controlling for obstacles to innovation is that these may affect the type and the organization of projects that are assigned to a firm's R&D unit and, thus, the R&D outsourcing decisions of a firm.

We also control for the degree to which firms can appropriate the results from their innovation projects. Cassiman and Veugelers (2002) and Cassiman and Veugelers (2005) have found that the appropriability of results is likely to be an issue when a firm considers outsourcing to other firms rather than to universities. Because our dependent variable is defined as the share of R&D outsourcing to universities, where the remaining share is represented by outsourcing to firms, we expect appropriability to have an impact on our outcome measure. We follow Czarnitzki et al. (2007) and construct a measure of industry effectiveness at appropriating results from innovation. Hence, we average the scores (from one to four) at the 2-digit NACE industry level that firms reported when answering the following question: "how important are your competitors as a source of information for the innovation process?"[4] We then rescale the total score to a measure that varies between zero and one, and we denote this measure *Appropriability*.

[4] When averaging firms' answers across industry sectors, we exclude the answer provided by firm i.

We also include a measure for the local availability of university knowledge, which is likely to affect the decision of firms to outsource to universities (Jaffe, 1989; Mansfield and Lee, 1996; Audretsch, 2005). This is defined as the total number of scientific articles (expressed in logs) that were published in 2005-2009 by universities located in the same region as the firm and that are relevant to a firm's industry. We denote this variable *Publications*. Finally, we control for industry, region[5], and year fixed effects.

Obtaining evidence on the causal relationship between the composition of a firm's R&D unit and the share of R&D outsourcing to universities is complicated by a number of factors. The decision of a firm to hire a certain R&D unit may be affected by its expectations regarding the availability of external partners with which the corresponding R&D unit could collaborate. Moreover, it may also be affected by the local availability of R&D employees, which are partly supplied by universities. To the extent that we control for the local availability of university knowledge, then this should not be a serious concern. However, it may still be the case that our measure for the availability of university knowledge is subject to measurement error. Indeed, not all university knowledge is contained in scientific articles but it may still be relevant to firms. Further biases could be caused by omitted variables that might be correlated with both the composition of a firm's R&D unit and the firm's decision to outsource to universities. To address these problems, we estimate an Instrumental Variables (IV), continuous, regression model that replaces the variables PhD_{it} and PhD_{it}_Share, in equations 1 and 2, respectively, with their predicted values. We follow Cassiman and Veugelers (2002) and Cassiman and Veugelers (2005) and derive the predicted values by regressing PhD_{it} and PhD_{it}_Share on the industry average of PhD staff employed[6] and the other exogenous regressors.

$$Out\_UNI_{it} = \beta_0 + \beta_1 \widehat{PhD}_{it} + X'_{it}\gamma + \zeta_{it} \tag{3}$$

$$Out\_UNI_{it} = \beta_0 + \beta_1 PhD\widehat{_{it}\_}Share + X'_{it}\gamma + \psi_{it} \tag{4}$$

We estimate our IV models using a two-stage least squares estimator, which delivers consistent estimates and requires few distributional assumptions (Wooldridge, 2002; Angrist and Pischke, 2008). We cluster standard errors at the firm level.

Table 2 presents descriptive statistics for samples A and B.

⟨ Insert Table 2 about here ⟩

[5] We include industry dummies for the NACE-2 industry sectors. We also include dummies for the following NUTS1 regions: Northwest, Northeast, Center, East, and South.

[6] In the count we exclude firm i.

3.3 Results

Table 3 reports the regression results for the impact of the composition of a firm's R&D unit on the share of outsourcing to universities. The first and the second columns display the OLS estimation results, while the third and the fourth columns display the IV estimation results[7]. Columns one and three examine the impact of having at least one R&D employee with a PhD degree (PhD_{it}). Conversely, columns two and four examine the impact of the share of R&D staff that holds a PhD degree (PhD_{it}_Share).

We find that having at least one R&D employee with a PhD degree has a positive impact on the share of outsourcing to universities, all else being equal. The coefficient is positive and statistically significant at the 1% level. Having at least one R&D employee with a PhD increases the share of outsourcing to universities by 90%. We interpret this result as evidence that R&D units with relatively lower learning costs for basic projects tend to match with external partners, universities, with relatively lower learning costs for the same projects. While this result does not rule out the possibility that R&D units with PhD staff collaborate with other firms, it shows that for R&D projects, collaboration tends to be relatively more intense among partners with similar comparative advantages. Collaboration with private companies might still occur for a number of reasons. First, there are some projects, which do not necessarily involve learning new problems, and for these projects our matching story might not apply. Second, some firms might have lower learning costs than universities, for basic projects. Unfortunately, we cannot distinguish in our data the characteristics of firms to whom outsourcing is done. Finally, it might be that the conditions offered by universities, including the price of the collaboration, are more onerous than those offered by private companies, thus inducing a firm to opt for its second-best option.

When we consider the share of R&D staff with a PhD degree, its impact on the share of firm outsourcing to universities is still positive and highly significant. The elasticity is 0.22. Having taken the endogeneity of PhD_{it} and PhD_{it}_Share into account, the results still hold. Their coefficients are still positive and statistically significant at the 1% significance level. This suggests a causal relationship between employing internal R&D staff with relatively lower costs for conducting basic projects and selecting external partners with a similar cost structure.

The results regarding the control variables reveal additional, interesting insights. The impact of firm size on the share of outsourcing to universities is negative, although only statistically significant at the 10% level, as presented in column two. Moreover, firms that reported a lack of qualified personnel as

[7] The results of the first-step regressions from which we derived the predicted values of PhD_{it} and PhD_{it}_Share are reported in Appendix C.

an important obstacle to innovation tend to outsource relatively less to universities. Interestingly, firms that indicated that a lack of information on either technology or external markets was an important obstacle to innovation tend to outsource relatively more to universities, suggesting that universities play an important role in filling information gaps of firms' R&D units. Finally, firms that had indicated difficulty in finding external partners tend to collaborate relatively more with universities. This last result might point to an additional role of universities, one of providing knowledge when it cannot be offered by other partners. Finally, as expected, the local availability of university knowledge is an important driver of the firms' decisions to outsource to universities.

⟨ Insert Table 3 about here ⟩

4 Robustness check analysis

We conduct a series of robustness checks to ensure that our results of interest hold using different model specifications. First, we estimate a Heckman selection model where we model the likelihood that a firm has outsourcing expenditures as a function of the covariates that we used in the previous regressions and an additional variable, whether the firm is located in a technology park. We use sample B for this analysis. In Spain, technology parks are created by either the federal or regional governments with the aim of attracting innovative firms and universities and encouraging intra-park collaboration (Vásquez-Urriago et al., 2011). While locating in one of these parks might facilitate external collaboration it should not bias collaboration with either firms or universities, as both are present in these parks. To validate our conjecture, we included a dummy for whether a firm is located in a technology park in equations one and two. The coefficient on the dummy was not statistically significant in either case. However, being located in a technology park has a positive and significant impact on the likelihood that a firm outsources. The results are reported in Table 4. The coefficients of PhD_{it} and PhD_{it}_Share are still positive and statistically significant at the 1% level. This suggests that had firm R&D units that conducted no outsourcing engaged in outsourcing, they would have matched with external partners based on their learning costs.

In Table 5, we present the results of estimations of equations 1-4, using a one-year lag for our independent variables of interest. Using lags controls for the possibility of a time gap between the point at which a firm R&D unit begins working on a project and the time when it begins working with an external partner. As expected, the results are robust to using a one-year lag.

Finally, in Table 6 we present the results of estimations of equations 1-4 using a random effect (RE) model[8]. The results for the impacts of PhD_{it} and PhD_{it}_Share on the share of outsourcing to universities hold also in this case.

⟨ Insert Table 4 about here ⟩
⟨ Insert Table 5 about here ⟩
⟨ Insert Table 6 about here ⟩

[8] We refrain from estimating a fixed-effect model because intra-firm variability is only generated by 5% of the sample.

5 Concluding Remarks

Our study makes two important contributions to the understanding of the optimal organization of R&D firms and their choice of outsourcing partners. The first contribution consists of extending the theory of hierarchical sorting within an organization developed by Garicano (2000), Garicano and Hubbard (2005), and Garicano and Rossi-Hansberg (2006) to the optimal matching between firms' internal R&D units and external outsourcing partners.

Our theory examines a setting in which the manager of an R&D firm needs to select a project from a menu of projects that differ in some characteristic. In our model, labor is sticky, and the firm's choice depends on the knowledge of its internal R&D unit, the composition of which does not change. This analysis is appropriate to analyze the short-run decisions of firms, especially in contexts in which the labor market is characterized by important rigidities. Once the firm manager has made her choice, she needs to select an external partner that will collaborate on the project with the internal R&D unit. Our focus is on projects that involve the creation of new knowledge and that require at least some level of outsourcing. This last characteristic is typical of projects that are complex enough that a firm is induced to hire an external partner, which will assist the internal R&D unit with problem solving. The rationale for hiring an external partner is that the partner can obtain a certain level of knowledge, on a project chosen by the firm, at lower costs than the firm's internal R&D unit.

Our main result is that in an economy with a continuum of firms distributed across a project characteristics, firms will find it optimal to select a project for which the internal R&D unit and the external partner will have a similar comparative advantage in problem solving. This is because it allows the external partner to spend relatively more time solving rarer problems than assisting the internal R&D unit. We believe this is an interesting result in that it shows that a profit-maximizing strategy for an firm consists in having the external partner relatively specialize in solving rarer problems, and have the internal R&D unit relatively specialize in solving the most common ones. Indeed, it is not profitable for a firm to outsource the solution to problems to an external partner, and have the internal unit specialize on the application of the relative solutions.

The second contribution we make is to test the predictions of our theory using a rich dataset of Spanish manufacturing firms that conducted innovating activities during the 2006-2009 period. Within this sample, we focus on those firms that conducted at least some level of R&D outsourcing. For these firms, we relate the share of R&D outsourcing to universities to the composition of their R&D units, described by the presence of staff with a PhD degree. The logic is that both universities and R&D units with PhD staff members have lower relative costs of conducting relatively more basic projects than companies to which the firm might want to outsource or with staff without PhD degrees.

We find evidence that R&D units with PhD staff tend to collaborate relatively more with universities than with other partners. We interpret this result as evidence that R&D units with relatively low learning costs for basic projects tend to match with external partners, universities, with relatively low learning costs for the same projects.

We believe our study contributes to uncovering the mechanisms through which R&D collaboration generates value for a firm. The patterns we have highlighted are typical of projects that involve the creation of new knowledge. These are projects in which the value added is measured in terms of the number of problems that are solved. For these projects, we analyzed the properties of an optimal match between firms' R&D units and their external partners. The reality, of course, is more complex than the one we have described. Indeed, firms not only can choose projects, conditionally on the characteristics of an R&D unit, but they can also hire new R&D units conditionally on the characteristics of a project they intend to undertake. Even in these cases, the problem of finding an optimal match between an internal R&D unit and an external partner remains relevant. By finding an optimal match, the firm achieves the optimal allocation of the internal unit's time between problem solving and production, as well as the optimal allocation of the external partner's time between problem solving and assisting the internal unit.

Acknowledgments

Andrés Barge-Gil acknowledges financial support from Ramón Areces Foundation. Usual disclaimers apply.

Appendix A: Proof of Proposition 1

To prove Proposition 1 we need to show that $\frac{dz_e(p)}{dc_e(p)} < 0$ and $\frac{dz_i(p)}{dc_e(p)} < 0$. We begin by computing $\frac{dz_i(p)}{dc_e(p)}$. Applying the implicit function theorem, we obtain that $\frac{dz_i(p)}{dc_e(p)} = -\frac{\frac{\partial^2\Pi}{\partial z_i \partial c_e}}{\frac{\partial^2\Pi}{\partial z_i^2}}$. By inspection, $\frac{\partial^2\Pi}{\partial z_i^2}$ is negative and, thus, the sign of $\frac{dz_i(p)}{dc_e(p)}$ depends on the sign of $\frac{\partial^2\Pi}{\partial z_i \partial c_e}$. This last expression is equal to:

$$\frac{\partial^2\Pi}{\partial z_i \partial c_e} = -\frac{z_e}{c_e}[\frac{h}{c_e} f'(z_e) f(z_i)(1 - c_i z_i) - f(z_e) c_i] - \frac{h}{c_e^2} f(z_e) f(z_i)(1 - c_i z_i)$$

Using the first order condition for z_i, we can rewrite the expression above as:

$$\frac{\partial^2\Pi}{\partial z_i \partial c_e} = -\frac{z_e}{c_e}[\frac{h}{c_e} f'(z_e) f(z_i)(1 - c_i z_i) - f(z_e) c_i] - \frac{F(z_e)}{c_e} c_i$$

Noting that $\frac{\partial^2 R}{\partial z_i \partial z_e} = \frac{h}{c_e} f'(z_e) f(z_i)(1 - c_i z_i) - f(z_e) c_i$,we rewrite:

$$\frac{\partial^2 \Pi}{\partial z_i \partial c_e} = -\frac{z_e}{c_e}[\frac{\partial^2 R}{\partial z_i \partial z_e}] - \frac{F(z_e)}{c_e} c_i$$

This expression is < 0 if $\left| \frac{\partial^2 R}{\partial z_i \partial z_e} \right| < \frac{F(z_e)}{z_e} c_i = \overline{k}$, which we have assumed in the text.

Regarding $\frac{dz_e(p)}{dc_e(p)}$, this expression is equal to:

$$\frac{dz_e(p)}{dc_e(p)} = -\frac{z_e}{c_e} + \frac{h}{c_e} f(z_i) \frac{dz_i(p)}{dc_e(p)}$$

Standard comparative statics show that $\frac{dz_e(p)}{dc_e(p)} < 0$ if $\left| \frac{\partial^2 R}{\partial z_i \partial z_e} \right| < \overline{k}$.

Appendix B: Variables Construction

⟨ Insert Table B1 about here ⟩

To obtain the publications in scientific fields that are relevant to a firm in a given industry, we follow Abramovsky et al. (2009) and match the scientific fields to industries using data from the 1994 Carnegie Mellon Survey (CNS). This survey reports the importance of the following ten research fields to each industry: biology, chemistry, physics, computer science, material science, medical and health science, chemical engineering, electrical engineering, mechanical engineering and mathematics. We follow Abramovsky et al.'s criterion that a research field is relevant for an industry if more than 50% of the CNS respondents report that the field is moderately or very important for the industry in which the respondent is active. Table B2 shows the matching between industries and scientific fields. Data on publications by university and scientific field for were obtained from website: http://sci2s.ugr.es/rankinguniversidades/. This website provides information on the total number of publications for every Spanish university in different scientific fields using data from Thomson-Reuters Web of Science although the information is not provided on an annual basis.

⟨ Insert Table B2 about here ⟩

Appendix C: First-step regressions

⟨ Insert Table C1 about here ⟩
⟨ Insert Table C2 about here ⟩

References

[1] Abramovsky, L., Harrison, R., Simpson, H. (2009). University research and the location of business R&D. *Economic Journal* 117, 114-141.

[2] Abramovsky, L., Kremp, E., Lopez, A., Schmidt, T., Simpson, H. (2009). Understanding co-operative innovative activity: Evidence from four European countries. *Economics of Innovation and New Technology* 18(3), 243-265.

[3] Aghion, P., Dewatripont, M., Stein, J. (2008). Academic freedom, private-sector focus, and the process of innovation. *RAND Journal of Economics*, 39(3), 617-635.

[4] Angrist, J., Pischke S. (2008). Mostly Harmless Econometrics: An Empiricists Companion, Princeton University Press, Princeton, NJ.

[5] Alcacer, J., and W. Chung. (2007). "Location Strategies and Knowledge Spillovers." *Management Science*, 53(5): 760-776.

[6] Audretsch, D. B., Lehmann, E. E. , Warning S. (2005). University Spillovers and New Firm Location. *Research Policy*, 34(7):1113-1122.

[7] Belderbos, R., Carree, M., Diederen, B., Lokshin, B., Veugelers R. (2004). Heterogeneity in R&D cooperation strategies. *International Journal of Industrial Organization* 22, 1237 – 1263.

[8] Branstetter L. G. (2005). Exploring the Link between Academic Science and Industrial Innovation. *Annales d'Economie et de*

[9] *Statistique* 70, 119-142.

[10] Cassiman, B, Veugelers, R. (2002). R&D Cooperation and Spillovers: Some Empirical Evidence from Belgium. *American Economic Review* 92(4), 1169-1184.

[11] Cassiman, B, Veugelers, R. (2005). In Search of Complementarity in the Innovation Strategy: Internal R&D and External Knowledge Acquisition. *Management Science* 52 (1), 68-82.

[12] Cockburn, I. M., Henderson, R. M. (1998). Absorptive capacity, coauthoring behavior, and the organization of research in drug discovery. *Journal of Industrial Economics*, 46(2), 157-182.

[13] Cohen, W., Levinthal, D. (1989). Innovation and learning: the two faces of R&D. *Economic Journal* 99, 569-596.

[14] Cohen, W., Nelson, R. Walsh, J. (2002). Links and impacts: the influence of public research on industrial R&D. *Management Science* 48(1), 1-23.

[15] Czarnitzki, D., Ebersberger, B., Fier, A. (2007). The relationship between R&D collaboration, subsidies and R&D performance: empirical evidence from Finland and Germany. *Journal of Applied Econometrics* 22, 1347-1366.

[16] Czarnitzki, D., Toole, A. (2011). Patent protection, market uncertainty and product innovation. *The Review of Economics and Statistics* 93(1), 147-159.

[17] Garicano, L. (2000). Hierarchies and the Organization of Knowledge in Production. *Journal of Political Economy*, 108(5): 874-904.

[18] Garicano, L., Hubbard, T. N. (2005). Hierarchical sorting and learning costs: Theory and evidence from the law. *Journal of Economic Behavior & Organization*, 58(2): 349-369.

[19] Garicano, L., Rossi-Hansberg, E. (2006). Organization and Inequality in a Knowledge Economy. *The Quarterly Journal of Economics*, 121(4): 1383-1435.

[20] 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.

[21] Griliches, Z. (1979). Issues in assessing the contribution of Research and Development to Productivity Growth. *Bell Journal of Economics* 10: 92-116.

[22] Gruber, M., Harhoff, D., Hoisl, K. (2012). Knowledge Recombination across Technological Boundaries: Scientists versus Engineers. *Management Science*, forthcming

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

[24] Jaffe, A. (1989). Real Effects of Academic Research. *American Economic Review*, 79(5): 957-970.

[25] Lacetera, N. (2009). Different missions and commitment power in R&D organizations: Theory and evidence on industry-university alliances. *Organization Science*, 20(3): 565-582.

[26] Leiponen, A., (2005). Skills and Innovation. *International Journal of Industrial Organization,* 23 (5–6): 303–323.

[27] Lopez, A., (2008). Determinants for R&D cooperation: Evidence from Spanish manufacturing firms. *International Journal of Industrial Organization* 26(1), 113-136.

[28] Mansfield, E., Lee J-Y. (1996)."The modern university: contributor to industrial innovation and recipient of industrial R&D support." *Research Policy*, 25(7): 1047-1058.

[29] Raymond, W., Mohnen, P., Palm, F. y van der Loeff, S. (2010). Persistence of innovation in Dutch manufacturing: is it spurious? *The Review of Economics and Statistics*, 92(3), 495-504.

[30] Thursby J., Thursby M. (2006). Where is the New Science in Corporate R&D? *Science,* 314, 1547-1548 .

[31] Vásquez-Urriago, A.R., Barge-Gil, A., Modrego, A., Paraskevopoulu, E. (2011). The impact of Science and Technology Parks on firms' product innovation: empirical evidence from Spain. MPRA Working Paper, 30555, University Library of Munich. Munich.

[32] Veugelers, R., Cassiman, B. (2005). R&D cooperation between firms and universities. *International Journal of Industrial Organization* 23, 355-379.

[33] Wolf, A., Mora-Sanguinetti, J.S. (2011). Reforming the Labour Market in Spain. OECD Economics Department Working Papers 845. OECD Publishing.

[34] Wooldridge, J. M. (2002). Econometric Analysis of Cross Section and Panel Data. The MIT Press, Cambridge MA.

Table 1. Number of observations per sample

Year	Sample A	Sample B
2006	1,109	2,352
2007	1,092	2,352
2008	1,006	2,352
2009	984	2,352

Table 2. Summary Statistics

	Sample A					Sample B				
Variable	Obs	Mean	Std. Dev.	Min	Max	Obs	Mean	Std. Dev.	Min	Max
out_uni	4191	1.15	1.78	0	4.61	4191	1.15	1.78	0	4.61
phd	4191	0.26	0.44	0	1.00	9408	0.20	0.40	0	1.00
phd_share	4191	0.70	1.28	0	4.61	9408	0.56	1.21	0	4.61
phd_share_industry_average	4191	0.61	0.53	0	2.40	9408	0.56	0.47	0	2.40
size	4191	4.58	1.38	0	9.23	9408	4.34	1.32	0	9.23
RD_intensity	4191	8.25	1.26	0	12.43	9408	8.03	1.25	0	12.92
size_RD_unit	4191	1.79	1.23	0	6.27	9408	1.44	1.15	0	6.27
export	4191	0.90	0.30	0	1.00	9408	0.88	0.32	0	1.00
foreign	4191	0.53	0.50	0	1.00	9408	0.44	0.50	0	1.00
parent	4191	0.11	0.32	0	1.00	9408	0.09	0.29	0	1.00
joint_venture	4191	0.01	0.12	0	1.00	9408	0.01	0.11	0	1.00
obstacle_funds	4191	0.72	0.45	0	1.00	9408	0.72	0.45	0	1.00
obstacle_qualified	4191	0.45	0.50	0	1.00	9408	0.46	0.50	0	1.00
obstacle_information	4191	0.52	0.50	0	1.00	9408	0.52	0.50	0	1.00
obstacle_partner	4191	0.38	0.48	0	1.00	9408	0.35	0.48	0	1.00
appropriability	4191	2.55	0.17	1.33	3.22	9408	2.56	0.17	1	4.00
publications	4191	6.44	3.30	0	10.16	9408	6.43	3.33	0	10.16

Table 3. Main Regressions. Sample A.

	[1] OLS	[2] IV	[3] OLS	[4] IV
phd	0.642***	1.181***		
	[0.100]	[0.426]		
phd_share			0.217***	0.332***
			[0.034]	[0.119]
size	-0.092	-0.102*	-0.08	-0.08
	[0.057]	[0.058]	[0.057]	[0.057]
RD_intensity	0.001	-0.019	0	-0.013
	[0.050]	[0.053]	[0.050]	[0.052]
size_RD_unit	0.024	-0.03	0.057	0.04
	[0.061]	[0.074]	[0.061]	[0.063]
export	0.118	0.111	0.103	0.091
	[0.127]	[0.125]	[0.126]	[0.125]
foreign	0.029	0.02	0.02	0.011
	[0.099]	[0.099]	[0.099]	[0.099]
parent	0.134	0.107	0.147	0.137
	[0.136]	[0.139]	[0.136]	[0.136]
joint_venture	0.272	0.317	0.259	0.28
	[0.291]	[0.303]	[0.305]	[0.320]
obstacle_funds	-0.025	-0.036	-0.019	-0.023
	[0.087]	[0.088]	[0.087]	[0.087]
obstacle_qualified	-0.402***	-0.374***	-0.412***	-0.399***
	[0.084]	[0.090]	[0.084]	[0.086]
obstacle_information	0.207**	0.216**	0.206**	0.212**
	[0.086]	[0.087]	[0.086]	[0.086]
obstacle_partner	0.172**	0.162**	0.171**	0.164**
	[0.080]	[0.080]	[0.080]	[0.080]
appropriability	-0.207	-0.219	-0.196	-0.197
	[0.277]	[0.270]	[0.278]	[0.272]
publications	0.038***	0.026*	0.037***	0.029*
	[0.013]	[0.015]	[0.013]	[0.015]
_cons	1.27	1.507	1.172	1.269
	[0.947]	[0.951]	[0.951]	[0.946]
Industry FE	YES	YES	YES	YES
Region FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
N	4191	4191	4191	4191
chi2		279.013		276.564
r2	0.127	0.113	0.128	0.122

The dependent variable is the share of firm outsourcing to universities. Sample A includes all innovating firms that on a given year had conducted at least some outsourcing. Clustered standard errors in brackets; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 4. Main Regressions. Sample B. Heckman Model

	[1]	[2]	[3]	[4]
	OLS	IV	OLS	IV
phd	0.613***	1.042**		
	[0.100]	[0.445]		
phd_share			0.210***	0.306**
			[0.034]	[0.131]
size	-0.108*	-0.116*	-0.095*	-0.105*
	[0.058]	[0.060]	[0.058]	[0.059]
RD_intensity	-0.024	-0.045	-0.022	-0.046
	[0.050]	[0.054]	[0.051]	[0.054]
size_RD_unit	0.008	-0.02	0.041	0.036
	[0.062]	[0.072]	[0.061]	[0.063]
export	0.109	0.088	0.095	0.079
	[0.126]	[0.131]	[0.126]	[0.131]
foreign	0.004	-0.023	-0.001	-0.029
	[0.099]	[0.101]	[0.099]	[0.102]
parent	0.132	0.129	0.145	0.147
	[0.135]	[0.137]	[0.135]	[0.136]
joint_venture	0.255	0.265	0.244	0.244
	[0.294]	[0.288]	[0.308]	[0.287]
obstacle_funds	-0.025	-0.025	-0.019	-0.02
	[0.087]	[0.089]	[0.087]	[0.089]
obstacle_qualified	-0.398***	-0.378***	-0.408***	-0.386***
	[0.084]	[0.087]	[0.083]	[0.086]
obstacle_information	0.204**	0.199**	0.204**	0.194**
	[0.086]	[0.087]	[0.086]	[0.087]
obstacle_partner	0.154*	0.151*	0.154*	0.150*
	[0.080]	[0.082]	[0.080]	[0.082]
appropriability	-0.214	-0.179	-0.202	-0.172
	[0.277]	[0.285]	[0.277]	[0.285]
publications	0.039***	0.031**	0.038***	0.033**
	[0.013]	[0.016]	[0.013]	[0.015]
_cons	1.851*	1.972**	1.692*	1.895*
	[0.946]	[0.984]	[0.950]	[0.979]
Industry FE	YES	YES	YES	YES
Region FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
N	9408	9408	9408	9408
chi2	303.142	259.782	302.423	259.782
rho	-0.141	-0.158	-0.126	-0.158

The dependent variable is the share of firm outsourcing to universities. Sample B includes all innovating firms, without restricting the analysis to those that on a given year had engaged in outsourcing. Clustered standard errors in brackets; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Clustered standard errors in brackets; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5. Main Regressions. Sample A. 1 lag

	[1] OLS	[2] IV	[3] OLS	[4] IV
lag_phd	0.611***	1.041**		
	[0.108]	[0.450]		
lag_phd_share			0.012***	0.033**
			[0.003]	[0.015]
size	-0.087	-0.091	-0.067	-0.042
	[0.063]	[0.063]	[0.064]	[0.067]
RD_intensity	0.015	0.002	0.025	0.01
	[0.056]	[0.058]	[0.057]	[0.060]
size_RD_unit	0.014	-0.03	0.068	0.05
	[0.067]	[0.081]	[0.068]	[0.070]
export	0.121	0.115	0.119	0.101
	[0.140]	[0.137]	[0.141]	[0.139]
foreign	0.102	0.097	0.1	0.087
	[0.107]	[0.107]	[0.108]	[0.109]
parent	0.148	0.123	0.179	0.173
	[0.149]	[0.152]	[0.149]	[0.150]
joint_venture	0.214	0.229	0.188	0.179
	[0.322]	[0.329]	[0.340]	[0.398]
obstacle_funds	-0.017	-0.027	-0.01	-0.024
	[0.096]	[0.097]	[0.097]	[0.097]
obstacle_qualified	-0.374***	-0.360***	-0.390***	-0.382***
	[0.092]	[0.095]	[0.091]	[0.093]
obstacle_informati	0.173*	0.181*	0.162*	0.162*
	[0.095]	[0.096]	[0.095]	[0.096]
obstacle_partner	0.202**	0.197**	0.207**	0.202**
	[0.086]	[0.086]	[0.087]	[0.088]
appropriability	-0.029	-0.011	-0.042	-0.021
	[0.326]	[0.319]	[0.333]	[0.328]
publications	0.047***	0.037**	0.054***	0.040**
	[0.014]	[0.016]	[0.014]	[0.016]
_cons	0.508	0.581	0.354	0.264
	[1.072]	[1.063]	[1.097]	[1.104]
Industry FE	YES	YES	YES	YES
Region FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
N	3082	3082	3082	3082
chi2		262.326		245.329
r2	0.131	0.122	0.121	0.093

The dependent variable is the share of firm outsourcing to universities. Sample A includes all innovating firms that on a given year had conducted at least some outsourcing. Clustered standard errors in brackets; * p < 0.05, ** p < 0.01, *** p < 0.001. We allow for a one-year lag between the composition of an internal R&D unit and the share of firm outsourcing.

Table 6. Main Regression. Sample A. Random Effects

	[1]	[2]	[3]	[4]
	OLS	IV	OLS	IV
phd	0.361***	1.245***		
	[0.081]	[0.281]		
phd_share			0.118***	0.368***
			[0.027]	[0.099]
size	-0.027	-0.085*	-0.021	-0.037
	[0.048]	[0.046]	[0.048]	[0.048]
RD_intensity	0.014	-0.011	0.015	-0.002
	[0.038]	[0.042]	[0.038]	[0.040]
size_RD_unit	-0.001	-0.044	0.016	0.004
	[0.050]	[0.056]	[0.050]	[0.051]
export	0.238**	0.136	0.235**	0.188**
	[0.104]	[0.091]	[0.104]	[0.096]
foreign	-0.012	-0.003	-0.014	-0.034
	[0.087]	[0.069]	[0.087]	[0.078]
parent	0.215*	0.112	0.223**	0.186*
	[0.113]	[0.094]	[0.113]	[0.102]
joint_venture	0.195	0.328	0.187	0.261
	[0.232]	[0.226]	[0.233]	[0.217]
obstacle_funds	0.053	-0.011	0.054	0.045
	[0.069]	[0.062]	[0.069]	[0.063]
obstacle_qualified	-0.312***	-0.352***	-0.315***	-0.329***
	[0.071]	[0.062]	[0.071]	[0.061]
obstacle_information	0.132*	0.199***	0.131*	0.152**
	[0.070]	[0.062]	[0.070]	[0.062]
obstacle_partner	0.089	0.146**	0.091	0.109*
	[0.067]	[0.059]	[0.067]	[0.058]
appropriability	0.149	-0.115	0.145	0.098
	[0.218]	[0.221]	[0.218]	[0.222]
publications	0.046***	0.025**	0.046***	0.028**
	[0.012]	[0.012]	[0.012]	[0.014]
_cons	0.024	1.11	-0.008	0.259
	[0.743]	[0.766]	[0.743]	[0.760]
Industry FE	YES	YES	YES	YES
Region FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
N	4191	4191	4191	4191
chi2	302.622	441.253	293.175	287.319

The dependent variable is the share of firm outsourcing to universities. Sample A includes all innovating firms that on a given year had conducted at least some outsourcing. Clustered standard errors in brackets; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B1. Description of variables

Variables	Definition
Out_uni	Share of firm's R&D outsourcing to universities (in logs)
PhD	PhD=1, if the R&D unit employs at least one employee with a PhD degree
PhD_share	Share of R&D staff holding a PhD (in logs)
PhD_share_industry_average	Industry average of share of R&D staff holding a PhD (answer from the firm is excluded from the calculation)
Size	Number of employees (in logs)
RD_intensity	R&D expenditures per employee (in logs)
Size_RD_unit	Full-time equivalent R&D workers in the R&D unit (in logs)
Export	Export=1, if the firm has positive sales from exports
Foreign	Foreign=1, if the firm has foreign headquarters
Parent	Parent=1, if the firm is the parent company of the group
Joint_venture	Joint_Venture=1, if the firm results from a joint venture
Obstacle_funds	Obstacle_funds=1, if the firm attaches moderate or high importance to the following obstacles: i) lack of funds within the firm or group, (ii) lack of funds from sources outside the firm
Obstacle_qualified	Obstacle_qualified=1, if the firm attaches moderate or high importance to the following obstacle: lack of qualified personnel
Obstacle_information	Obstacle_information=1, if the firm attaches moderate or high importance to the following obstacles: i) lack of information on technology, (ii) lack of information on markets
Obstacle_partner	Obstacle_partner=1, if the firm attaches moderate or high importance to the following obstacle: difficulty in finding partners for innovatin projects
Appropriability	Industry average of the importance of competitors, according to the answers firms had given to the following question: How important are your competitors as a source of information for the innovation process? [on scale 1 (very imporant) to 4 (unimportant), rescales between 0 and 1]
Publications	Total number of scientific articles published in 2005-2009 by universities located in the same region as the firm and that are relevant to a firm`s industry (in logs). A detailed description is provided in Table B2.
Industry dummies	Chemicals and chemical products; Pharmaceuticals; Machinery and equipement; Office Machinery and computers; Electrical machinery; Radio, television and communication equipment; Medical, precision and optical instruments and Aircraft and spacraft .
Regional dummies	Northwest, Northeast, Center, East, and South.
Park	Park=1, if the firm is located on a Science and Technology Park

Table B2. Matching among industrial sectors and scientific fields

Industrial Sectors	Scientific Fields
Food and beverages	2
Tobacco	2
Textiles	
Wearing apparel; dressing and dyeing of fur	
Leather and leather products	
Wood and wood products	
Pulp, paper and paper products	
Publishing, printing and reproduction of recorded media	5
Coke, refined petroleum product	2, 4
Chemical products (except drugs)	2
Drugs	1, 2, 6
Rubber and plastic products	5
Ceramic tiles and flags	5
Other non-metallic mineral products	4, 5, 8, 9
Basic ferrous metals	
Basic metals (non ferrous)	
Fabricated metal products, except machinery and equipment	
Machinery and equipment	5, 9
Office machinery and computers	4, 5, 8, 9
Electrical machinery and apparatus	8
Electronic valves and tubes and other electronic components	3, 5, 8, 9
Radio, television and communication equipment and apparatus	4, 5, 8
Medical, precision and optical instruments, watches and clocks	4, 5, 6, 8
Motor vehicles	5
Building and repairing of ships and boats	
Aircraft and spacecraft	4, 5, 9
Other transport equipment	
Furniture	
Games and Toys	
Other manufacturing	

Notes: We follow the criteria in Abramovsky et al (2009) that a research field is relevant for an industry if over 50% of respondents reported in Cohen et al (2002) say that the field is moderately or very important for an industry. Scientific Fields are: Biology (1), Chemistry (2), Physics (3), Computer Science (4), Materials Science (5), Medicine (6), Chemical Engineering (7), Electrical Engineering (8), Mechanical Engineering (9) and Mathematics (10). Data on publications by university and scientific field were obtained from website: http://sci2s.ugr.es/rankinguniversidades/. This website provides information on the total number of publications for every Spanish university in different scientific fields using data from Thomson-Reuters Web of Science. Information is available for the whole period, not annually.

Table C1. First Step of IV regressions. Sample A.

	[1] Dummy PHD	[2] Share PHDs
size	0.031***	0.047
	[0.012]	[0.036]
RD_intensity	0.022**	0.060**
	[0.009]	[0.027]
size_RD_unit	0.069***	0.064*
	[0.011]	[0.035]
export	0.017	0.074
	[0.021]	[0.070]
foreign	0.008	0.031
	[0.019]	[0.057]
parent	0.047*	0.077
	[0.028]	[0.082]
joint_venture	-0.102*	-0.241
	[0.057]	[0.157]
obstacle_funds	-0.005	-0.028
	[0.016]	[0.047]
obstacle_qualified	-0.003	0.017
	[0.014]	[0.043]
obstacle_information	-0.014	-0.032
	[0.015]	[0.044]
obstacle_partner	0.01	0.01
	[0.013]	[0.037]
appropriability	-0.011	0.022
	[0.041]	[0.123]
publications	0.006**	0.015*
	[0.003]	[0.009]
phd_share_industry_average	0.215***	0.772***
	[0.022]	[0.066]
_cons	-0.308*	-0.753
	[0.162]	[0.487]
N	4191	4191
chi2	687.478	431.192

Sample A includes all innovating firms that on a given year had conducted at least some outsourcing. Clustered standard errors in brackets; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table C2. Selection Equation. Sample B. Heckman Model

	[1]	[2]	[3]	[4]
	OLS	IV	OLS	IV
phd	0.201***	1.179***		
	[0.054]	[0.246]		
phd_share			0.055***	0.346***
			[0.017]	[0.072]
park	0.205*	0.203*	0.205*	0.203*
	[0.112]	[0.112]	[0.112]	[0.112]
size	0.119***	0.102***	0.121***	0.114***
	[0.034]	[0.035]	[0.034]	[0.035]
RD_intensity	0.170***	0.136***	0.170***	0.134***
	[0.030]	[0.032]	[0.031]	[0.032]
size_RD_unit	0.116***	0.037	0.126***	0.100***
	[0.038]	[0.042]	[0.038]	[0.038]
export	0.051	0.021	0.05	0.01
	[0.062]	[0.062]	[0.062]	[0.063]
foreign	0.152***	0.125**	0.151***	0.118**
	[0.053]	[0.054]	[0.053]	[0.054]
parent	0.023	-0.015	0.026	0.006
	[0.076]	[0.077]	[0.076]	[0.077]
joint_venture	0.111	0.179	0.107	0.155
	[0.161]	[0.159]	[0.161]	[0.159]
obstacle_funds	0.001	-0.011	0.001	-0.006
	[0.047]	[0.047]	[0.047]	[0.047]
obstacle_qualified	-0.023	0.025	-0.025	0.016
	[0.045]	[0.047]	[0.045]	[0.046]
obstacle_information	0.016	0.018	0.015	0.013
	[0.047]	[0.047]	[0.047]	[0.047]
obstacle_partner	0.121***	0.112**	0.121***	0.110**
	[0.044]	[0.044]	[0.044]	[0.044]
appropriability	0.051	0.022	0.054	0.031
	[0.163]	[0.164]	[0.163]	[0.164]
publications	-0.006	-0.024***	-0.005	-0.023**
	[0.008]	[0.009]	[0.008]	[0.009]
_cons	-2.659***	-2.267***	-2.680***	-2.355***
	[0.549]	[0.561]	[0.550]	[0.558]
N	9408	9408	9408	9408
chi2	303.142	259.782	302.423	259.782
rho	-0.141	-0.158	-0.126	-0.158

We estimate the probability that a firm, on a given year, did at least some level of outsourcing. Clustered standard errors in brackets; * p < 0.05, ** p < 0.01, *** p < 0.001.