Endogenous Matching in University-Industry Collaboration: Theory and Empirical Evidence from the UK*

Albert Banal-Estañol[†], Inés Macho-Stadler[‡] and David Pérez-Castrillo[§]

June 2016

Abstract

We use a two-sided matching framework to analyze collaboration between heterogeneous academics and firms. We consider both horizontal and vertical characteristics, i.e., those related to affinity (e.g., preferences for a type of scientific research) and those related to ability (e.g., capacity to produce high-quality scientific output). We build a unique dataset based on the teams of academics and firms that proposed research projects to the UK's Engineering and Physical Sciences Research Council. Our results are suggestive of positive assortative matching in terms of ability and type while the matching is negative assortative in terms of their interactions. The most able and the most applied academics are the ones that are more likely to propose collaborative as opposed to non-collaborative projects.

JEL Classification numbers: O32, I23

Keywords: matching, industry-science links, research collaborations, complementarity

*We thank Christian Fons-Rosen, Bing Guo, Joan Llull, Ernest Miguelez, Anna Toldrà, the associate editor and two reviewers for very useful comments, as well as Eduard Clariana, Carlos Marcos and Jonás Nieto for excellent research assistance. We are also thankful for the comments received at KU Leuven, Universitat de Barcelona, Ben-Gurion University, Academia Sinica in Taipei, Universitat d'Alacant, Universitat de Valencia, ISER at Osaka, Universidad Católica de Chile, and at workshops on "University-Industry Collaboration" and "Economics of Science and Innovation" in Barcelona, "Incentives and the Design of Mechanisms in Open Science" in Bordeaux, the Roundtable on Engineering and Entrepreneurship Research (REER) in Atlanta, the Conference on Innovation and Entrepreneurship at Northwestern, the Asian Meeting of the Econometric Society, "KID" in Nice, and "CGTiBP" in Leipzig. We gratefully acknowledge financial support from the Ministerio de Economía y Competitividad research grants (ECO2012-31962, ECO2013-43011-P, and ECO2015-63679-P), Generalitat de Catalunya (2014SGR-142), Ramon y Cajal, ICREA Academia, and Severo Ochoa Programme (SEV-2015-0563).

[†]Universitat Pompeu Fabra and Barcelona GSE. Email: albert.banalestanol@upf.edu.

[‡]Universitat Autonoma de Barcelona and Barcelona GSE. Email: ines.macho@uab.es.

[§]Universitat Autonoma de Barcelona and Barcelona GSE. Email: david.perez@uab.es.

1 Introduction

Science and innovation in modern economies often involves collaboration across institutional boundaries. Academic research groups sometimes work independently, but inter-institutional and international collaborations and coauthorships are very common (Wagner and Leydesdorff, 2005). Similarly, while some technologies are developed by one single firm, many others are developed by research joint ventures (Kamien et al., 1992). Fortunately, a substantial body of research in the economics and management literatures has identified the causes and the consequences of inter-institutional collaboration *within* institutional markets, i.e., "one-sided market" partnerships (see Katz and Martin, 1997, and Caloghirou et al., 2003, for reviews).

However, the full transformation of modern societies into knowledge- and science-based economies also requires collaboration *across* institutional markets, i.e., "two-sided market" partnerships. Business-science links through joint research, consulting or training arrangements, for example, are key channels of knowledge transfer between academia and industry according to both academics (Agrawal and Henderson, 2002) and firms (Cohen et al., 2002). As a result, university-industry collaborations are nowadays stronger and more widespread than ever before (Jensen and Thursby, 2001; Perkmann et al., 2013). Unfortunately, in spite of their tremendous importance, we know very little about which groups of which institutions engage in collaboration and which two-sided market partnerships are actually formed.

Take for example the partnership formed in 2007 by Professor Sir Colin John Humphreys of Cambridge University, who specializes in electron microscopy and analysis, and FEI, a world-leading company in the production and distribution of electron microscopes. In this case, a prolific researcher of a top university, whose research is considered basic, collaborated with a research-intensive firm, heavily oriented toward basic research. We ask whether this pattern is the most common: Do top academics collaborate with top firms, whereas less productive researchers collaborate with less productive firms? Do they collaborate because they have similar preferences? Do they choose each other because of individual or institutional characteristics? Are other, less productive and more applied academics and firms more likely to stay independent?

This paper investigates the two-sided market collaboration formation process and, in particular, the characteristics of the resulting partnerships. We study what type of partners on each side of the market are more likely to collaborate with each other, and which characteristics affect the likelihood of collaborating, as opposed to working independently. We consider both "horizontal" and "vertical" characteristics, i.e., those related to affinity (e.g., preferences for a type of scientific research) and those related to ability (e.g., capacity to produce high-quality scientific output). We show that collaboration decisions are affected both by affinity-based and ability-based characteristics, as well as by their interactions.

Collaboration has benefits and costs for participants on both sides of the market. Academics claim that industry collaboration provides them with additional funds and insights (Mansfield, 1995; Lee, 2000), but it might bias their selection of research topics and methodology (Florida and Cohen, 1999). Firms report collaborating with academics to get access to new university research and discoveries (Lee, 2000), although some of these outcomes have little or no commercial value (Jensen et al., 2003). Firms are also concerned with the differences in terms of organizational and institutional structure, and with the existence of the open science culture in academia (Dasgupta and David, 1994).[1]

Overall, previous empirical studies have shown that thanks to collaboration firms obtain better patents, more products, and increased sales (Cockburn and Henderson, 1998; Zucker et al., 2002; Cassiman and Veugelers, 2006). On the academic side, collaboration has recently been linked to a higher number of academic publications (Fabrizio and DiMinin, 2008; Azoulay et al., 2009). But Agrawal and Henderson (2002) do not find significant effects. Goldfarb (2008) even reports a negative effect for researchers who maintain funding relationships with an applied sponsor. Banal-Estañol et al. (2015) suggest that the relationship between industry collaboration and academic output may resemble an inverted U-shaped curve.

Unfortunately, most of the existing empirical evidence on performance provides average effects, across all partnerships. Recent evidence, however, stresses the importance of the characteristics of the matched partners in assessing collaboration outcomes. Banal-Estañol et al. (2013), for example, show that the research projects in collaboration with firms produce more scientific output than those without them if and only if the firms in the project are research-intensive. The rewards from collaboration might thus be highly heterogeneous and depend on own, as well as on potential partners' characteristics. For instance, all academics, but especially those that are more oriented toward basic research, might prefer firms that encourage their employees to publish scientific articles (Cockburn and Henderson, 1998). Similarly, all firms might prefer to collaborate with "star" academics, as their input increases firm performance (Zucker et al., 2002). Research-oriented firms and star academics, however, might not be willing or able to collaborate with all participants on the other side of the market. Given the costs and benefits

[1] Academic researchers' individual characteristics and attitudes, as well as local group norms, have also been shown to play a role in the collaboration decision (Louis et al., 1989). Firms' size, absorptive capacity and the adoption of open search strategies are also important factors in the firms' willingness to collaborate (Mohnen and Hoareau, 2003; Veugelers and Cassiman, 2005; Bercovitz and Feldman, 2007). Geographical proximity between the researchers' university and the firms has also been shown to be important, particularly for researchers in universities with modestly rated faculties (Audretsch and Stephan, 1996).

of collaborating with each potential partner, how do academics and firms mutually choose each other, and which of them decide to work independently?

To understand the mechanisms at work, we use a one-to-one two-sided matching framework of academic researchers and firms developing research projects.[2] Participants on each side of the market are heterogeneous in terms of their scientific ability and the type of research they undertake (i.e., its degree of "appliedness"). Each participant can develop a project on her (its) own, or search for an appropriate partner on the other side of the market to develop a collaborative project. Our specification of value for collaborative projects can accommodate complementarities and substitutabilities in terms of partners' abilities as well as the positive and the negative effects of heterogeneity in terms of knowledge or expertise (Dahlin et al., 2005).

Our theoretical framework underscores the differences of analyzing equilibrium matching in terms of vertical and horizontal characteristics. Indeed, the two main partner attributes, ability and type, have different effects on the value of a partnership. Differences in ability represent vertical differentiation: an academic (or a firm) with higher ability is a better academic (firm) than an academic (firm) with lower ability, independent of the characteristics of her (its) partner. Differences in types of research instead represent horizontal differentiation: an academic (or a firm) with high type, i.e., undertaking more applied research, is neither better nor worse than an academic (firm) with low type, i.e., undertaking more basic research. Joint profits increase when the types of the academic and the firm are, depending on the value of the parameters, more similar or more distant, not when one of them is high or low.

We use a linear functional form of project value that captures the direct effects of, as well as interactions between, vertical and horizontal characteristics. Our specification includes a term based on the cross-product of the abilities, which makes the profit function twice-continuously differentiable with respect to this attribute. A sufficient condition for the equilibrium matching to be positive (resp. negative) assortative in terms of ability is then that the cross-partial derivative with respect to the abilities of both partners is positive (negative). In our framework, as the cross-partial derivative is constant, this condition is both necessary and sufficient and the equilibrium matching is unique. We also include a linear term based on the absolute value of the difference between types of research, which means that the profit function is not twice-continuously differentiable everywhere. Still, we identify sufficient conditions for the positive or negative assortative matching to be one of the equilibrium matchings in terms of type. Finally, we allow for interactions between the main attributes. For instance, distance in types of research

[2] Our framework follows the approach of Shapley and Shubik (1972), Becker (1973), and Legros and Newman (2002).

may reduce the positive impact of ability in the value of the partnership. Distance of types can be considered a vertical characteristic, even if the types themselves are not. In terms of ability-affinity pairs, we also provide necessary and sufficient conditions for the negative (positive) assortative matching to be the unique equilibrium matching.

We build a unique dataset based on the teams of academic researchers and firms that have proposed research projects to the Engineering and Physical Sciences Research Council (EPSRC), the main government agency for research funding for the engineering departments of the UK universities. The EPSRC grants are allocated to teams of academic researchers alone and also to teams that include one or more firms as industry partners. For each of our $5,855$ projects, we identify project participants' past publications, which allow us to construct various measures of ability and affinity. We use, in particular, the normal count, the impact-factor-weighted sum and the (average) impact-factor of the publications as proxies for scientific ability and the proportion of publications in basic or applied journals as a proxy for type of research (Narin et al., 1976; Godin, 1996; van Looy et al., 2006).

As an empirical strategy, we use both Fox's (2008) "maximum score estimation" method and Gompers et al.'s (forthcoming) "probit-counterfactual" approach. The maximum score method structurally estimates the parameters of a production function whereas the probit-counterfactual approach estimates the probabilities of matching. Fox's (2008) estimates are consistent but, as argued by Akkus et al. (2014), estimation methods based on random utility models, such as the probit-counterfactual approach, are widely understood and applied in other contexts, and thus, they serve as alternative methods. In addition, the probit-counterfactual approach allows us to quantify the impact of several pair-wise characteristics on the probability of being matched.

Our structural estimates suggest that partner abilities are complementary, the effect of the distance between types is negative, and (academic/firm) ability and type distance are substitutes. According to our theoretical framework, this implies that there is positive assortative matching in terms of ability: top academics collaborate with top firms whereas academics of lower ability collaborate with firms of lower ability. Second, there can be positive assortative matching in terms of type: academics with more applied interests collaborate with firms with more applied bias. Finally, the matching is negative assortative in terms of ability-affinity pairs, i.e., the higher the ability of the academics, the closer they are to their partners in terms of type.

Our second empirical strategy, the probit-counterfactual approach, allows us to quantify the impact of the *ranking* of the academics and the firms on the probability of being matched. We show, for instance, that pairs of academics and firms that are in the same quartile of their respective distribution of abilities are between 13.5% and 28.5% more likely to be matched, compared to those pairs that are in different quartiles. Similarly, academics and firms are 33% (39.5%) more likely to be matched if they are both above (below) the median in terms of type. Although positive, the effects of being both above (below) the median in terms of ability are less strong. Overall, our results suggest that matching occurs at the top of the distribution in terms of ability but over the whole distribution in terms of affinity.

We also assess the relative importance of the horizontal versus the vertical characteristics, as well as of the individual versus the institutional characteristics. The horizontal characteristics are relatively more important than the vertical ones, both in terms of magnitude and the significance of effects. Importantly, the characteristics at the individual level are more relevant than those at the institutional level. Top firms form links with top academic institutions only insofar as they include top researchers. This reinforces the view that the fundamental unit of collaboration is composed of individuals, not institutions (Katz and Martin, 1997).

Finally, we investigate the characteristics of the academics that submitted collaborative instead of non-collaborative projects. As in the case of matched pairs, we use both maximum score estimation and a probit analysis. Our maximum score estimation results suggest that the most able academics are those more likely to collaborate but they do not allow us to make a general statement on whether the most applied or the most basic academics are more likely to collaborate. Our probit analysis suggests that academics that are above the median in terms of ability and those who are more applied than the median are 9.1% and 39.5% more likely to propose collaborative projects, respectively. The characteristics at the individual level seem to be again more important than those at the institutional level. Indeed, academics in larger and better-performing universities are neither more nor less likely to submit collaborative projects.

Our leading example fits most of the general properties of the university-industry partnerships. Professor Humphreys, a leading academic, collaborates with FEI, a leading firm. Both of them share the same preferences for the type of research and, as leaders in their markets, they are matched with a partner with similar preferences. As a top academic and a top and basic firm, they end up forming a partnership instead of remaining independent. But, in contrast to one of our results, a basic academic does not stay independent. Finally, our comparison between individual and institutional characteristics suggests that FEI collaborates with professor Humphreys not because he is a professor at Cambridge, but because of his individual characteristics.

This paper provides, to our knowledge, the first framework to analyze the nature of the equilibrium matching in terms of a horizontal as well as in terms of a vertical characteristics. Our framework also enables us to investigate the interaction between a horizontal characteristic in one side of the market and a vertical one in the other. Our paper also includes an analysis of the characteristics of, not only the matched, but also the non-matched agents.

The closest paper to ours is Mindruta (2013), who uses maximum score estimation to identify ability-based characteristics, constructed from publication and patent data, as a source of complementarity in university-industry collaboration.[3] Instead, we consider both ability- and affinity-based characteristics (and their interactions) and show that the latter may be even more important than the former ones. She uses collaboration data between professors at a top U.S. medical school and their industry partners. Using homogeneous data from 40 universities, we are also able to compare individual versus institutional characteristics.

This paper is organized as follows. Section 2 introduces the theoretical framework. Section 3 describes our dataset. Section 4 introduces the empirical strategy, and section 5 the results on matching between matched partners. In section 6, we investigate the characteristics of the academics who collaborate rather than remain independent. Finally, section 7 concludes.

2 Theoretical framework

We consider a market with m academic researchers $\mathcal{A} = \{A_1, A_2, ..., A_m\}$, and n firms $\mathcal{F} = \{F_1, F_2, ..., F_n\}$. Academics and firms can develop research projects on their own, labeled as "non-collaborative," or form academic-firm partnerships and develop "collaborative" projects. Due to time and other constraints, agents have the capacity to develop only one project. That is, the matching is endogenous and we model it as a "one-to-one two-sided matching market."[4]

2.1 Profits for collaborative and non-collaborative projects

Both collaborative and non-collaborative projects are valuable because they may generate new knowledge, academic papers, and/or patents. In this subsection, we describe first the value of a collaborative project between a generic academic A and firm F, and then that of their respective stand-alone, non-collaborative projects. We assume that collaborative projects can include a transfer, or compensation, from one side to the other. Indeed, firms can provide additional

[3] Based on 46 case-study interviews, Carayol (2003) proposes a typology of business-science collaborations and argues that firms involved in (low) high-risk projects are matched with academic teams of a (low) high excellence. Agarwal and Ohyama (2013) study, both theoretically and empirically, the labor market for scientists. The academic and private sectors choose among scientists who differ in their ability and preferences, and scientists choose between academia and industry. Our setup, of course, differs from theirs and includes more than two classes of participants on each side of the market.

[4] A two-sided matching market is one in which there are two distinct sets of agents. It is one-to-one if an agent from one side of the market can be matched only with an agent from the other side or remains unmatched. For an introduction to matching markets, see Roth and Sotomayor (1990).

funding to the academics. Similarly, academics can provide consulting services and/or facilitate access to their laboratories to the firms. The existence of transfers allows us to focus on the *total value* of the partnership, i.e., for both partners. We refer to it as simply "value" from now on.

In our framework, the key determinants of value are the "abilities" of the academic and the firm as well as their "types" of research. The ability parameter $\delta_A \geq 0$ measures the technical and scientific level of academic A. The parameter $\delta_F \geq 0$ measures the scientific level of firm F, i.e., its absorptive capacity or the level of its human capital. We denote by $x_A, x_F \in [0,1]$ the type of research that researcher A and firm F, respectively, undertake. The types of research are indexed in the unit interval by their degree of "appliedness," with 0 representing the least applied type possible (i.e., the most basic one) and 1 representing the most applied type possible. The distance between types, $|x_F - x_A|$, measures the degree of heterogeneity between them.

Our framework posits that the value of a collaborative project may depend on both partners' ability, and that the effects of these relationships can be enhanced or curtailed by the partners' heterogeneity in terms of types of research. Specifically, the expected profits, or value, of a project of a pair (A, F), with characteristics (δ_A, x_A) and (δ_F, x_F), is given by

$$\Pi^c(A,F) = \alpha_a^c \delta_A + \alpha_f^c \delta_F + \gamma_1 \delta_A \delta_F + \gamma_2 |x_A - x_F| + \gamma_3 \delta_A |x_A - x_F| + \gamma_4 \delta_F |x_A - x_F| - C, \quad (1)$$

where α_a^c and α_f^c measure the individual effect of the academic's and firm's ability in a collaborative project and $C > 0$ represents the fixed costs of running a collaborative project.[5]

The other parameters in (1) are interaction terms. The parameter γ_1 measures whether partners' abilities are substitute ($\gamma_1 < 0$), independent ($\gamma_1 = 0$), or complementary ($\gamma_1 > 0$). In the former, a better firm is more valuable for a less-able academic than for a more able one, whereas in the latter better firms are more valuable for better academics. Depending on the value of γ_2, heterogeneity in type can be negative ($\gamma_2 < 0$), neutral ($\gamma_2 = 0$) or positive ($\gamma_2 > 0$) for profits. Indeed, differences in the type of research each partner usually works on can create difficulties for the collaboration, and therefore $\gamma_2 < 0$. Instead, there can be complementarities in terms of knowledge or expertise, as is sometimes claimed in the case of interdisciplinarity, and therefore $\gamma_2 > 0$. Heterogeneity may also be irrelevant, and therefore $\gamma_2 = 0$. In our setup, differences in types of research represent *horizontal* differentiation whereas differences in ability represent *vertical* differentiation. Indeed, profits increase when the types of academic and firm are similar, if $\gamma_2 < 0$, or distant, if $\gamma_2 > 0$, not when one of them is high or low.[6]

[5] For simplicity, and to concentrate on the effects of heterogeneity in terms of ability and type, our model assumes that fixed costs, those that do not depend on partners' abilities or type differences, are common to all collaborative projects. Similarly, the fixed costs of running the project will be assumed to be the same for all non-collaborative projects of the academics (firms).

[6] The resulting profit function in a model in which the type of the project is endogenous would be similar to

Finally, parameter γ_3 (resp. γ_4) measures possible interaction effects between the ability of the academic (firm) and the distance in type. If $\gamma_3 > 0$, the academic's ability and distance are complementary. In this case, the marginal productivity of the distance (which is positive or negative depending on the sign of γ_2) increases with the academic's ability, and the marginal productivity of the academic's ability increases with the distance in types. Instead, if $\gamma_3 < 0$ ability and distance are substitutes.

Academics and firms can also run projects on their own. The value of a non-collaborative project depends on the ability but not on the type. Specifically, the expected profits of a non-collaborative project run by an academic A with characteristics (δ_A, x_A) is given by

$$\Pi_A^n(A) = \alpha_a^n \delta_A - c_a, \tag{2}$$

where $c_a \geq 0$ is the fixed cost of running a non-collaborative project. We allow the individual effect of the ability in a non-collaborative project to be different from that in a collaborative project, that is, α_a^n may be different from α_a^c. This may be because of moral hazard problems that may arise when academics work with firms, in which case $\alpha_a^c < \alpha_a^n$, or because of increased levels of motivation when they work in collaborative projects, in which case $\alpha_a^c > \alpha_a^n$.

Similarly, the profits of a non-collaborative project run by a firm F with characteristics (δ_F, x_F) is given by

$$\Pi_F^n(F) = \alpha_f^n \delta_F - c_f,$$

where $c_f \geq 0$. We assume $c_a + c_f \leq C$ because of, for example, the "coordination costs" of different institutional cultures (Dasgupta and David, 1994).

2.2 The market equilibrium matching

We now consider the whole market of academic researchers and firms, in which each academic A_i is characterized by a pair (x_{A_i}, δ_{A_i}) and each firm F_j by a pair (x_{F_j}, δ_{F_j}). Academic researchers and firms are heterogeneous in terms of type of research and ability, i.e., it is possible that $x_{A_i} \neq x_{A_{i'}}$, $x_{F_j} \neq x_{F_{j'}}$, $\delta_{A_i} \neq \delta_{A_{i'}}$ and/or $\delta_{F_j} \neq \delta_{F_{j'}}$ for $A_i, A_{i'} \in \mathcal{A}$ and $F_j, F_{j'} \in \mathcal{F}$. In this subsection, we outline the conditions for equilibrium matching *between matched partners*.[7] Still,

the one posited here, as shown in Banal-Estañol et al. (2013). In that paper, the profits of each partner decrease with the distance between the (endogenous) type of the joint project and her (its) "ideal" type of project, or the type they are most productive at (i.e., our x_A and x_F). In equilibrium, the optimal type of the project lies between the types of the participants (see also Pereira, 2007).

[7] In an equilibrium, no academic or firm can improve upon her or its current payoff. That is, no academic or firm in a collaborative project can be better off by developing a non-collaborative project and no academic and firm can be better off by leaving their current partners, if any, and forming a new partnership.

there can be non-matched academics and firms in equilibrium, but we postpone the description of their characteristics, as opposed to those of matched partners, to section 6.[8]

The next three blocks describe, in turn, the conditions under which the equilibrium matching is positive or negative assortative with respect to ability, type, and ability-distance pairs as a function of the parameters γ_1, γ_2, γ_3, and γ_4 of the profit function (1). Of course, the equilibrium partnerships do not depend on the coefficients related to the researcher only or to the firm only (α_a^c and α_f^c), or indeed to the fixed effects to any collaboration, such as the running costs (C).

Ability. Ability is a vertical characteristic: the value function Π^c increases in δ_A (resp. δ_F) independently of the value of δ_F (δ_A). To capture these relationships, we include the cross-product of δ_A and δ_F, which makes the value function twice-continuously differentiable with respect to δ_A and δ_F. As a result, a sufficient condition for the equilibrium matching to be positive (resp. negative) assortative in terms of ability is that the cross-partial derivative with respect to δ_A and δ_F is positive (negative) (see e.g., Legros and Newman, 2002).[9] Since the cross-partial derivative never changes sign (it is in fact constant, and equal to γ_1), the condition is necessary and sufficient and the equilibrium matching is unique. As a result, top academics collaborate with top firms and academics of lower ability collaborate with firms of lower ability if and only if partner abilities are complementary (i.e., if and only if $\gamma_1 > 0$). Similarly, top academics collaborate with low ability firms whereas low ability academics collaborate with top firms if and only if $\gamma_1 < 0$. In this case, a low ability firm is willing to offer a top academic a high transfer to make this collaboration more appealing to the academic. Similarly, low ability academics can use transfers to lure high ability firms into a collaboration.[10]

Type. Type is a horizontal characteristic: the value function Π^c increases or decreases in the distance between x_A and x_F, not in the types themselves. We include a linear term based on the absolute value of the difference between types of research, which means that the value function is

[8] We consider, as is standard in the literature, that each agent has complete information on the characteristics of all the agents, as well as on the value of any potential partnership. We also assume away search costs in finding partners. As we show later, geographical distance is always insignificant in our regressions. This suggests that search costs, at least in terms of geographical distance, are small.

[9] In multivariate settings, the conditions for assortativeness in one characteristic are true as long as all academics, on one side, and all firms, on the other, are homogeneous with respect to the other characteristic. To our knowledge, Lindenlaub (2014) is the first to propose a general definition for positive and negative assortative matching that take into account heterogeneity along two characteristics for twice continously differentiable production functions.

[10] If $\gamma_1 = 0$ then any matching is an equilibrium matching in terms of ability. This will also be the case for the other cases if $\gamma_2 = 0$, $\gamma_3 = 0$, or $\gamma_4 = 0$.

not twice-continuously differentiable everywhere, with respect to x_A and x_F. Still, the positive assortative matching is always an equilibrium matching if $\gamma_2 < 0$. Indeed, $\gamma_2 < 0$ implies that the distance in types is costly and the positive assortative matching minimizes the sum of the distances between the matched academics and firms. In other words, the positive equilibrium matching is *efficient,* in the sense that it maximizes "total surplus," i.e., the sum of the values of all projects in the market cannot be increased by reassigning firms and academics. As is well known (see, e.g., Shapley and Shubik, 1972), efficient matchings are equilibrium matchings in models with transfers.[11] As a result, if the heterogeneity in types is costly (i.e., if $\gamma_2 < 0$), the academics with more applied interests shall collaborate with firms with more applied bias.[12] Similarly, if $\gamma_2 > 0$ the negative assortative matching is always an equilibrium matching. In this case, collaboration takes place between basic firms and applied academics, and vice-versa.

Ability-distance pairs. The value function Π^c is twice continuously differentiable with respect to the ability and the distance, even though it is not with respect to the types themselves. As in the case of ability, the equilibrium matching is unique, and it is also positive (resp. negative) assortative in terms of the academic's ability-distance pair if and only if the academic's ability and distance in terms of type of research are complementary (substitute). In other words, if $\gamma_3 > 0$ the higher the ability of the academics, the further they shall be to their partners in terms of type of research. If, on the other hand, $\gamma_3 < 0$ the higher the ability of the academics, the closer they shall be to their partners in terms of type of research. In this second case, the distance in terms of type of research is more damaging (or less profitable) for the most able academics. A similar argument can be made for the nature of the matching in terms of the firm's ability and distance as a function of γ_4. The equilibrium matching is positive (resp. negative) assortative in terms of the firm's ability-distance pair if and only if $\gamma_4 > 0$ ($\gamma_4 < 0$).

[11] This property derives from competition in the market: as firms compete among themselves for their most preferred academic partner, and academics compete among themselves for their most preferred firm partner, the resulting matching maximizes the total surplus. If this was not the case, an agent, or a pair of agents, would obtain more benefits in an alternative matching.

[12] The equlibrium matching may not be necessarily unique, though. For example, if all academics are more basic than all the firms, the negative assortative matching is also an equilibrium. For this particular distribution of types, the sum of distances in both the positive and the negative assortative matchings are equal.

3 Data and descriptive statistics

3.1 Sample

Our empirical analysis is based on the teams of academic researchers and firms that propose research projects to the Engineering and Physical Sciences Research Council (EPSRC). More than half of the overall research funding for the engineering departments of the UK universities comes from the EPSRC. EPSRC grants are allocated to teams of academic researchers alone as well as to teams of academics and firms (firms alone cannot apply for EPSRC funds). As defined by the EPSRC, "collaborative research grants" are those that involve one or more firms as industrial partners.[13] Industrial partners contribute cash or 'in-kind' services to the full economic cost of the project.

Our initial sample includes all the EPSRC project proposals with the starting year 2005, 2006 or 2007. For each project, we know the holding organization, the principal investigator (PI), the coinvestigators (if any), and the industry partners (if any). We take the projects with at least one academic researcher (not necessarily the PI) in the longitudinal data set in Banal-Estañol et al. (2015), which contains calendar information and publication data on all the academics employed at all the engineering departments of the 40 major UK universities until 2007. Our final sample consists of $5,855$ projects ($1,912$ in 2005, $1,835$ in 2006, and $2,108$ in 2007). As a whole, we have $2,411$ unique academic researchers and $1,735$ firms which are involved in $2,057$ out of the $5,855$ projects. That is, 35% of the projects are "collaborative" projects. The average number of researchers in each project is 2.86, and the average number of firms in each collaborative project is 2.43.

We consider the teams of academics and the teams of firms as the relevant units. We take the view that teams in each side of the market are formed before the two-sided matching market starts. Therefore, the one-to-one matching model is the most suitable framework to analyze our data. For simplicity, in the description of the results, we refer to a team of academics as an "academic" and to a team of firms as a "firm." Therefore, when we talk, for instance, about the ability of the academic in the project, we mean the ability of the team of academics.

[13] Some partners in the EPSRC database are not private firms but are university research centers and schools, large research infrastructures (e.g., the LNCC National Laboratory of Scientific Computing), government and municipal councils, public agencies, public hospitals, charities, and trade associations (e.g., the International Union of Railways). We disregard these partners as we only analyze collaboration with private firms.

3.2 Main variables

To build proxies for scientific ability and type of each partner, we use publication information prior to the start of the project. In particular, we use the publications for the five years prior to, as well as for the starting year of, the project (because they relate to research developed and finished before the start of the project).[14] For example, if the initial year of a project is 2007, then we use the publications of the academics and the firms during the period 2002 to 2007.[15]

Publication data is extracted from the Thomson (formerly ISI) Web of Science (WoS) database (for details, see Banal-Estañol et al., 2015). In total, our $2,411$ academic researchers published $44,399$ articles in the years 2000 to 2007. Similarly, we identify $201,296$ publications for the period 2000 to 2007 for the $1,735$ firms involved in the collaborative projects.

As a measure of scientific ability, we use the normal count of publications, the impact-factor-weighted sum of publications, and the (average) impact-factor per publication, all of them averaged per year.[16] The impact factors are the Science Citation Index (SCI) Journal Impact Factors (JIF), attributed to the publishing journal in the year of publication. The second measure takes into account not only the quantity but also the quality of the publications. Therefore, we use it as our main measure but also report the results for the normal count, which measures quantity, and for the average impact factor, which can be considered a measure of quality.

For the type of research, we use the Patent Board classification (Narin et al., 1976; Hamilton, 2003). Based on cross-citation matrices, it classifies journals into four categories: (1) applied technology, (2) engineering and technological science, (3) applied and targeted basic research, and (4) basic scientific research. The first two categories are considered to be "technology" and the last two "science" (see Godin, 1996; van Looy et al., 2006). We follow this distinction and define the type of research of a set of articles as the number of publications in the first two categories divided by the number of publications in all four categories.

In our regressions, we use the normal count and the impact-factor-weighted sum of publications per year, as well as the average impact-factor, of the whole team of academic researchers in the project, but we also report results for the PI only. Similarly, we use the normal count,

[14]We present our results using the variables that include information for six years. But our results using information for two or four years are very similar.

[15]Our measures of the firms are noisier than those of the academics because we do not know the name of the researcher of the firm involved in the project. In addition, firms may have different policies with respect to publication activities.

[16]Some of the academic researchers are not in our sample for the whole period, for example, because they are junior or come from abroad. We take this into account by computing the average count and impact-factor-weighted-sum per year, using the years available (out of the six years prior to the start of the project).

the impact-factor-weighted sum of publications per year, and the average impact-factor, of the whole team of firms participating in the project (there is no equivalent to the PI for the firms). Since the distributions of these variables are highly skewed, we use the logarithm of the variables plus the unit and we refer to these measures as the "count," the "impact," and the "average impact" of each of the participants. Similarly, we use the average type of the team of academic researchers, the type of the PI, as well as the average type of the team of firms in the project, and refer to them as the "type" of each of the participants.

3.3 Institutional variables

We obtain information on the size and performance of the departments of our academics from the 2008 "Research Assessment Exercise (RAE)" results. The RAE provides aggregate information on the number of active academics and research funds in each department for the period 2001 to 2007. We assign to each project the information of its holding institution.

We construct variables related to firm size and performance, using the FAME and ORBIS databases. In particular, we compute the average (per firm in the team and per year) number of employees and turnover (in millions of pounds). We also assign to each collaborative project the one-digit Standard Industrial Classification (SIC) code of the firm(s). If the activity of the firms in the project spans several sectors, we randomly assign one of them.[17] Moreover, we assign to each non-collaborative project the one-digit SIC code of the firms that collaborate with the academics in other projects (if they have collaborated with firms associated to several SIC codes, we randomly select one of them).

Finally, for the collaborative projects, we assign a variable which measures the geographical distance between the partners. We retrieved the postal codes of the universities and the headquarters of the firms and, using an application of the UK government,[18] we computed the distance in miles between the postcode of the holding university and the UK-based firm in the project. For firms based outside the UK, we define a fixed distance of $1,000$ miles. When there are several firms, we use the minimum distance between the university and the firms.[19]

[17] We classify firms' activity according to the 10 US SIC division structure. For those firms for which we only have a UK SIC code, we make a simple translation from the two-digit UK SIC codes to US divisions. We assign the "artificial division" 11 to those projects for which we cannot associate a division to any of the firms.

[18] http://www.education.gov.uk/cgi-bin/inyourarea/distance.pl?

[19] This reflects the fact that the closest firm is the one that establishes the link with the university. We have also run the regressions using the average distance of the firms, using UK firms only, and we obtain similar results.

3.4 Descriptive statistics

As shown in table 1, the academic teams in our database published 10.6 articles on average per year, with an impact of 16.9 and an average impact per publication of 1.392, during the six-year period prior to the initial date of the project. Not surprisingly, the count and the impact of the academics in the projects are highly correlated (0.889). For both variables, the mean is greater than the median (for the impact, 16.9 vs 7) because the distributions are negatively skewed, with many academic teams of low impact and count and some of high impact and count (the maximum impact is 526.1). The average PI published 3.6 articles per year, with an impact of 5.6, which are highly correlated with the impact of the whole academic team (0.534 and 0.636, respectively). The average count, impact and average impact for the firms is 749, $1,448$ and 1.21 respectively. The variance of the publications of the firms is much larger than that of the academics, both in absolute terms and relative to the means. In comparison to those of the academic teams, the distributions of the firms are even more skewed. The count, the impact, and the average impact of the firms are positively correlated to the impact of the academics in the same project, 0.223, 0.219, and 0.155, respectively. These correlations provide initial preliminary evidence of positive assortative matching in terms of ability.

[Insert Table 1 here]

We can define the type for the $5,519$ academic teams, $4,674$ PIs and $1,563$ teams of firms with at least one publication in a journal included in the Patent Board classification. The average type of the firms is more basic (0.579) than the average type of the academics (0.656) and the PI (0.666). This is probably because firms do not give their employees an incentive to publish or do not allow them to publish their most applied discoveries, which may be directly profitable for the firms.[20] This fact suggests that the "true" type of a firm is more applied than the one observed in the data. In our empirical analyses, we use an adjusted measure of the distance between an academic of the firm where the type of the firm is increased by a fraction of 0.1.[21] The correlation between the types of the academics and the firms (0.368) provides initial preliminary evidence of positive assortative matching in terms of affinity.

[20] For academics the mean is lower than the median (0.656 vs 0.750) because the distribution of types is positively skewed. In contrast, for firms the mean is similar to the median (0.579 vs 0.600) because the distribution has firms spread over all the range of types.

[21] This adjusted distance leads to qualitatively similar but stronger results than by defining the distance as the difference between the type of the academic and the firm. We have also tried adding other amounts, and the results are similar, but an increase of 0.1 gives the best estimates.

We include the correlations between type and impact, which are all very significant: -0.343 for academics, -0.396 for the PIs and -0.123 for the firms. The (unreported) correlations between type and count are also negative, but the magnitudes are smaller. These correlations indicate that more applied researchers, PIs, and firms publish less in high-impact journals.

Finally, in table 1 we also report the descriptive statistics of the institutional variables for universities and firms.

4 Empirical strategy

We use both Fox's (2008) "maximum score estimation" method and Gompers et al.'s (forthcoming) "probit-counterfactual" approach. The maximum score method structurally estimates the parameters of the production function which determine the properties of the matching. This approach relies on a "rank order" property, which postulates that matchings that generate more surplus in a deterministic setup are more likely to be observed. The probit-counterfactual approach does not attempt to estimate the production function. It assumes that the choices that generate more utility are more likely to be realized. For a given collaborating academic or firm, it estimates the likelihood of this agent ending up with her actual partner rather than with an alternative counterfactual partner.

The main advantage of the maximum score estimation method over the probit-counterfactual approach is that it explicitly accommodates rival choices. Random utility models, on which the probit-counterfactual approach is based, assume instead that agents on each side of the market make their partner choices independently. But, as discussed earlier, observed partner choices may result from decisions of all agents on both sides of the market and, thus, they are determined interdependently. Still, the maximum score method has several drawbacks: the precision of the estimates can only be estimated using a subsampling procedure, and it is less efficient and more challenging computationally than the probit-counterfactual approach.

4.1 Maximum score estimation

Fox (2008) proposes the use of maximum score estimation to estimate a local value maximization condition, which ensures pair-wise stability. As is well known, pair-wise stability is equivalent to total surplus maximization, which, as discussed in section 2, is a necessary and sufficient condition for equilibrium matching. For the set of matched pairs, the local maximization condition states that the sum of value from two observed matches is greater than the sum of value if they exchange partners:

$$\Pi^c(A_i, F_j) + \Pi^c(A_{i'}, F_{j'}) \geq \Pi^c(A_i, F_{j'}) + \Pi^c(A_{i'}, F_j) \tag{3}$$

where (A_i, F_j) and $(A_{i'}, F_{j'})$ are two realized pairs in a given market, and therefore $(A_{i'}, F_j)$ and $(A_i, F_{j'})$ are unrealized counterfactual pairs in that market. In our empirical analysis, we assume that there is a market in every sector and in every year, and that markets in different years and/or different sectors are independent of each other. In total, we have complete information on $1,240$ pairs spread over 21 markets.

The crucial part of the maximum score estimation is the specification of the profit function, $\Pi^c(\cdot, \cdot)$. We use the linear production function introduced in the theoretical framework:

$$\Pi^c(A, F) = \alpha_a^c \delta_A + \alpha_f^c \delta_F + \gamma_1 \delta_A \delta_F + \gamma_2 |x_A - x_F| + \gamma_3 \delta_A |x_A - x_F| + \gamma_4 \delta_F |x_A - x_F| + \iota_A \iota_F - C + \varepsilon. \tag{4}$$

As discussed in other papers (Yang et al., 2009; Akkus et al., 2014), coefficients related to the researcher only or to the firm only, or fixed effects to any collaboration, such as the running costs, cannot be identified because they will cancel out in the local maximization condition (3). We can only estimate interaction terms. The scale of the production function cannot be identified either. Any increasing and affine monotonic transformation of the production function will produce the same results. Following Matzkin (1993), to ensure the identification of the parameters in the production function we include an interaction variable, $\iota_A \iota_F$ where ι_A is the total research funds awarded to the department of the academic and ι_F is the turnover of the firm. We normalize the coefficient for this variable to be ± 1, and, following standard practice, we report the regression results of the one with the maximum score. In our estimations, it is always positive.

From the local profit-maximization conditions (3), we derive a system of inequalities. We apply maximum score estimation (Manski, 1975) and find the production function that maximizes the total number of inequalities. The objective is to maximize the following score function,

$$\max_{\Omega} \frac{1}{H} \sum_{h \in H} \left\{ \sum_{\{A_i, F_j, A_{i'}, F_{j'}\} \in Q_h} \mathbf{1} \left[\Pi^c(A_i, F_j) + \Pi^c(A_{i'}, F_{j'}) \geq \Pi^c(A_i, F_{j'}) + \Pi^c(A_{i'}, F_j) \right] \right\} \tag{5}$$

where H is the number (and the set) of markets, $\mathbf{1}$ is the indicator function that is equal to 1 when the inequality in the bracket is true, Q_h is the set of all quartets consisting of two realized pairs in market h, and $\Omega = \{\gamma_1, \gamma_2, \gamma_3, \gamma_4\}$ the set of variables over which to maximize. The maximum score estimators will be a set of parameters Ω that maximize the score function.

As the objective function of the maximum score estimator is a step function, there are many local optima. Therefore, we apply a global optimization routine, the differential evolution method (Storn and Price, 1997), to estimate the parameters. We use Santiago and Fox's (2008) template in Mathematica, adapted to our production function.

Maximum score estimation does not assume a distribution for the error terms; therefore, we need to use subsampling techniques to calculate the confidence intervals for the estimators. Fox (2008) makes separate consistency arguments for one large matching market and many independent matching markets. In our setting, we have 21 distinct matching markets and view them as a number of independent markets sampled from a larger number of such markets. We follow the procedure proposed in Santiago and Fox's (2008) toolkit. First, from the whole data, we randomly generate 100 subsamples, each containing 19 markets. Second, we apply the above estimation procedure to each of the subsamples and obtain 100 sets of parameter estimates. Finally, for each of the parameters, we use the maximum score estimators of the subsample procedure as its empirical distribution from which we calculate its 95% confidence interval.

4.2 Probit-counterfactual approach

Our second approach follows Agrawal et al. (2008), Hegde and Tumlinson (2014), and Gompers et al. (forthcoming). As in the maximum score estimation, we construct a fixed set of counterfactual collaborations, i.e., collaborations that were possible but were not formed. For a given set of pair-wise characteristics of a collaboration, we estimate the likelihood that this collaboration is an actual rather than a counterfactual collaboration. To that purpose, we run probit regressions on the likelihood of forming a partnership, using a dependent variable which has a value of 1 if the partnership is an actual pair and a value of 0 if it is a counterfactual pair.

In this case, the set of counterfactuals is constructed as follows. We take the set of academics and firms that have a collaborative project. A pair formed by any of these academics and any of these firms is a potential counterfactual pair if they do not form an actual collaboration but have a collaborative project with other partners in the same market (i.e., in the same year and in the same sector). For each actual collaborative project, we select four of these counterfactual pairs in the following way. We randomly choose one counterfactual in which the academic coincides with the one in the actual project; then one counterfactual in which the firm corresponds to the one in the actual project; then another one for the academic and another one for the firm. We alternate the choice in order to have a more balanced set of counterfactuals. We avoid repetitions so that a counterfactual pair can only appear once. We add the resulting $5,838$ counterfactual pairs to the $1,485$ actual pairs of which we have information on ability and type.[22,23]

[22] For a limited number of collaborative pairs, it is not possible to find four counterfactual pairs. This is because there are too few collaborations in that year and in that sector.

[23] We have also run all the regressions of the probit counterfactual-approach on a dataset containing two coun-

We test the properties of the matching between academics and firms using the actual formed partnerships, as well as the non-formed counterfactual partnerships we have constructed. In all the regressions, we use pair-wise horizontal and vertical characteristics simultaneously. As each academic and firm in an actual pair also appear in the same number of counterfactual pairs, the individual characteristics have no impact on the likelihood of forming a partnership, as in the maximum score estimation. We include year and sector fixed effects, and report robust standard errors clustered at the academic researcher level, in all the regressions. We also include, as a control, the variable that measures the geographical distance between the academics and the firms in the projects. This variable, however, turns out to be insignificant in all the regressions.

5 Empirical results

5.1 Maximum score estimation

The first four columns of table 2 describe the parameter estimates $\Omega = \{\gamma_1, \gamma_2, \gamma_3, \gamma_4\}$ of function (4). In the first specification we use, as measure of ability, the impacts of the academic team and the firm and, as a measure of appliedness, the types of the academic team and the firm. In the second and third columns we use count and average impact, instead of impact, as a measure of ability. In the fourth, we use the measures of the PI instead of those of the academic team.

[Insert Table 2 here]

In every specification parameter γ_1 is statistically greater than zero at a 95% confidence level, suggesting that partner abilities are complementary. This implies that there is positive assortative matching in terms of ability: top academics collaborate with top firms whereas academics of lower ability collaborate with firms of lower ability.

Parameter γ_2 is always negative and, in all but one specification, statistically significant, which implies that the effect of distance is negative. This suggests that there are no benefits from participant heterogeneity or that they are outweighed by the costs in terms of divergent interests. Divergent interests may, for example, reduce the effort supplied and the resulting value of the project, e.g., the number of academic papers or patents. Our estimator on γ_2 suggests that there can be positive assortative matching in terms of appliedness: academics with more applied interests collaborate with firms with more applied bias.

terfactuals per project. The results are very similar to those obtained in the case of four counterfactuals. The sign and the significance of the coefficients are almost the same, and the magnitudes of the effects are similar. These results suggest that an increased (or reduced) number of counterfactuals does not alter our results.

As mentioned earlier, the estimator on γ_2 provides some support for the matching being positive assortative in terms of type. As an additional test, we also run a linear regression on the product of types, as well as on the product of abilities (even if this specification implicitly assumes that type is a vertical characteristic). Column 5 provides more evidence of positive assortative matching in terms of type, as it shows that the estimate of the cross-partial derivative of the profit function with respect to the types is positive.

The coefficients for γ_3 and γ_4 are all negative and all but two are significant, thus providing support for the matching being negative assortative in terms of academic ability-distance pair and firm ability-distance pair, respectively. Thus, the higher the ability of the academic, the closer she is to her partner in terms of type. Similarly, we have that the higher the ability of the firm, the closer it is to its partner in terms of type.

We also test the properties of the matching using institutional measures of the universities and the firms. We are particularly interested in determining whether firms select academic researchers because of their individual characteristics or because of the characteristics of the university they work for. As shown in columns 6 and 7, the coefficients associated with university research funds and firms' turnover (performance aggregate), and with university total number of researchers and firms' employees (size aggregate), are all positive but very small and all but one are insignificant. We also run (unreported) regressions using other university characteristics such as number of top quality papers, income, number of undergraduate and graduate students, and expenses. Similarly, we consider other firms' characteristics such as profits and assets. The coefficients associated with all of these variables also turn out to be insignificant.[24]

In terms of performance, our maximum score estimators in the structural equation satisfy between 64.3% and 64.9% of the total number of inequalities in the objective function (5), in the case where we use the measures of the academic team, and around 62.2% in the case where we use the measures of the PI (bottom of table 2). These results are in line with those of previous applications in the literature (e.g., Yang et al., 2009; Mindruta, 2013).[25]

[24] We used the RAE data to construct the variables that measure the number of top quality publications of all the engineering departments at each university. Using data from the Higher Education Statistics Agency, we obtained information related to the general characteristics of the universities: number of undergraduate and graduate students and the university's income and expenses. We also used the FAME and ORBIS databases to compute the average (per firm in the team and per year) tangible assets and profits before tax.

[25] We also run regressions including (i) only ability and distance, as well as (ii) only ability and (iii) only distance. The number of inequalities satisfied are 64.0%, 55.4% and 63.8%, respectively. These results suggest that the horizontal characteristics have more explanatory power than the vertical characteristics, as we will confirm in the probit-counterfactual approach.

5.2 Probit-counterfactual approach

The probit counterfactual approach estimates the probability of matching directly. We use this well-known model, first, as an alternative test to the maximum score estimation results. Indeed, we checked for the consistency of the two empirical approaches by regressing the likelihood of being an actual match, as opposed to a counterfactual match, over the same (absolute) variables we used in the linear specification of the maximum score estimation. The results on the cross-partial effects of the probability of matching are the same as those of the cross-partial derivatives of value obtained in the previous section.

Second, and more importantly, this approach allows quantifying the impact of the *ranking* of the academics and the firms. The assortative nature of the matching is an ordinal, not a cardinal, property. So instead of the absolute values of ability and type, we use the rankings of the academics and the firms in the two characteristics. We construct several categorical and continuous variables accounting for the relative position of each agent in each side of the market. Table 3 presents the average marginal effects across all the observations, measured as discrete changes in the case of the categorical variables and as derivatives for the continuous ones.[26]

[Insert Table 3 here]

We start by using two dummy variables that take the value of 1 if both the academic and the firm are above the median of their respective distribution of ability and type, respectively, and two other dummy variables that take the value of 1 if they are both below the median in terms of impact and type, respectively. According to the results in column 1, the pairs of academics and firms that are both above (below) the median of their respective distribution of abilities are more likely to be matched, compared to those pairs in which one is above the median and the other is below. Albeit positive, these effects are not significant. The effects for the types are instead very significant. The academics and the firms are 6.9% and 8.4% more likely to be matched if they are both above the median and both below the median in terms of type, respectively. Given that the unconditional probability of being matched is 20%, this represents 34.5% and 42% in terms of the unconditional probability.

To further understand the effect of the relative position of academics and firms, we divide the set of academics and the set of firms in quartiles with respect to each characteristic. Column 2 shows that a pair is more likely to be an actual match, rather than a counterfactual match, if both the academic and the firm are in the same quartile in terms of ability. This is especially the case

[26] The results of the marginal effects at the mean of the variables are similar.

for the top quartile, in which the conditional probability is 6.1% higher and the unconditional one 30.5% higher. In the case of the bottom quartile, the effect is not significant.

As a next step, we use the difference in terms of quartiles. We assign to each academic and firm in a match the value of their quartile (from 1 to 4), and compute the pair's quartile difference, defined as the absolute value of the difference between the quartile of the academic and the quartile of the firm. Column 3 shows the difference in predicted probabilities for the case in which the quartile difference is one, two, and three, as compared to the case of a difference of zero, for both the ability and the type. Again, the probability of being matched is lower if the difference is positive. Still, the effects of the ability appear to be rather flat, whereas the effects of the type are increasing in the difference and, overall, stronger.

We also use two variables with the full ranking, with respect to impact and type, for the academics and the firms.[27] To be able to appreciate the magnitude of the effects, we divide the ranking by one thousand. Column 4 reports the average marginal effects. As expected, a higher distance in the ranking in impact and type leads to a lower likelihood of matching. In terms of magnitudes, increasing the difference between the ranking, in terms of impact, of the academic and that of the firm by one thousand positions lowers the conditional probability of being matched by 2.2% (11% of the unconditional probability), whereas the effect in terms of type is 10.1% (50.5% of the unconditional probability).

In sum, table 3 suggests that the positive nature of the matching is stronger for the type, which is a horizontal characteristic, than for the impact, which is a vertical attribute. The coefficients for the joint variables in types are three to four times higher than the coefficients for the corresponding variables for impact. The numbers are meaningful because the variables reflect the relative positions of the agents, which allow for the comparison of the two characteristics.

6 Collaborating versus staying independent

The equilibrium matching determines not only who collaborates with whom but also who collaborates and who remains independent. This section describes the characteristics that make academics more likely to develop collaborative rather than non-collaborative projects. Unfortunately, we cannot analyze which firms would be more likely to conduct non-collaborative projects because firms alone cannot apply for EPSRC funding and are thus not in our dataset.

As before, we use both Fox's (2008) maximum score estimation method and a probit approach. Again, the main advantage of maximum score estimation over the probit approach is

[27] The ordering treats equal numbers as average ranking. That is, if the impacts were 1, 7, 7, 20, then the associated ranks would be 1, 2.5, 2.5, 4.

that it explicitly accommodates rival choices. Low quality academics may end up not collaborating with a firm not because they do not want to but because the firms prefer other academics. Moreover, the maximum score estimation allows us to estimate another parameter: the difference between the individual effect of the academics in collaborative and non-collaborative projects. But, knowing the parameters of the profit function does not always allow us to make general statements about the characteristics of the academics who collaborate with firms as opposed to those who do not. Therefore, as a second approach, we run probit regressions which, in addition, allow us to quantify the impact of several characteristics on the probability of being matched.

6.1 Maximum score estimation

We extend Fox's (2008) maximum score estimation approach described in section 4 to include, in addition to local value maximization conditions based on matched pairs, shown in (3), local value maximization conditions also involving non-matched academics:[28]

$$\Pi^c(A_i, F_j) + \Pi_A^n(A_{i'}) \geq \Pi_A^n(A_i) + \Pi^c(A_{i'}, F_j) \tag{6}$$

where (A_i, F_j) is a realized pair and $A_{i'}$ a non-matched academic in a given market.[29] As before, we use the econometric specification of the collaborative profits $\Pi^c(A, F)$, shown in (4). We also add an error term to the non-collaborative profits of the academic $\Pi_A^n(A)$ shown in (2). In the estimation, we include, in addition to the $1,240$ pairs used in section 5, $2,842$ non-collaborative academic projects for which we have complete information, spread over the same 21 markets. This allows us to estimate not only the coefficients of the interaction terms $\{\gamma_1, \gamma_2, \gamma_3, \gamma_4\}$, but also the difference between the individual effects of the ability of the academic, $\alpha_a^c - \alpha_a^n$. Still, the lack of information on non-collaborative firms may add some noise to the estimation. As explained earlier, coefficients related to fixed effects to the researcher or to the firm, or fixed effects to any collaboration, such as the running costs, cannot be identified because they will cancel out in the local maximization conditions (3) and (6).

The last two columns of table 2 describe the parameter estimates using, as measures of ability, the impacts and the counts of the academic team and the firm and, as a measure of appliedness, the types of the academic team and the firm. Once again, we report the 95% confidence intervals, obtained from 50 random subsamples. The coefficients of the interaction terms have the same sign as before, although γ_3 and γ_4 are less often significantly different from

[28] We thank an anonymous referee for suggesting this approach to us.

[29] Unfortunately, we cannot include local value maximization conditions involving non-collaborating firms, i.e., $\Pi^c(A_i, F_j) \geq \Pi^n(A_i) + \Pi^n(F_j)$ and $\Pi^c(A_i, F_j) + \Pi^n(F_{j'}) \geq \Pi^c(A_i, F_{j'}) + \Pi^n(F_j)$, where (A_i, F_j) is a realized pair and $F_{j'}$ a non-matched firm, because we do not have information on non-matched firms.

zero ($\gamma_1 > 0$, $\gamma_2 < 0$, $\gamma_3 \leq 0$, and $\gamma_4 \leq 0$). The coefficient for $\alpha_a^c - \alpha_a^n$ is negative (and significant in the second specification), suggesting, going back to the discussion in section 2, that there may be moral hazard problems when academics work in collaborative projects.

We now use the estimated parameters of the value function to analyze the characteristics of the academics who collaborate with firms as opposed to those that do not. Take first a given academic A_i that may form a partnership with a given firm F_j. They will develop a collaborative project rather than two non-collaborative ones if $\Delta_1(A_i, F_j) \equiv \Pi^c(A_i, F_j) - \Pi_A^n(A_i) - \Pi_F^n(F_j) > 0$ where

$$\begin{aligned}\Delta_1(A_i, F_j) &= (\alpha_a^c - \alpha_a^n)\delta_{A_i} + (\alpha_f^c - \alpha_f^n)\delta_{F_j} + \gamma_1\delta_{A_i}\delta_{F_j} + \gamma_2\left|x_{A_i} - x_{F_j}\right| \\ &+\gamma_3\delta_{A_i}\left|x_{A_i} - x_{F_j}\right| + \gamma_4\delta_{F_j}\left|x_{A_i} - x_{F_j}\right| - (C - c_a - c_f) > 0. \qquad (7)\end{aligned}$$

Given our estimated parameters ($\gamma_2 < 0$, $\gamma_3 \leq 0$, and $\gamma_4 \leq 0$), distance between types $\left|x_{A_i} - x_{F_j}\right|$ has an unambiguously negative effect on $\Delta_1(A_i, F_j)$. The ability of the academic has instead conflicting effects on $\Delta_1(A_i, F_j)$. On the one hand, given that $\gamma_1 > 0$, there is a positive effect because of the complementarities between the abilities of the partners. But, on the other hand, as $\gamma_3 \leq 0$, there may be a negative effect due to the substitutability between ability and distance in types. Moreover, as $(\alpha_a^c - \alpha_a^n) \leq 0$, the individual effect of the ability when the academic works in collaboration with a firm can be lower than when she runs a non-collaborative project. Still, if we assume that firms face similar issues to the academics when working in collaborative projects, i.e., $(\alpha_f^c - \alpha_f^n) \leq 0$, the overall effect of ability in the net benefits of collaboration are positive. Indeed, collaboration can only be profitable for some level of ability if $(\alpha_a^c - \alpha_a^n) + \gamma_1\delta_{F_j} + \gamma_3\left|x_{A_i} - x_{F_j}\right| > 0$ and, in this case, $\Delta_1(A_i, F_j)$ is increasing in δ_{A_i}.

Academic A_i may not form a partnership with firm F_j even if $\Delta_1(A_i, F_j) > 0$ because firm F_j may instead collaborate with another (non-collaborating) academic $A_{i'}$ In equilibrium, the pair (A_i, F_j) develops a collaborative project and $A_{i'}$ stays independent instead of $A_{i'}$ and F_j forming a partnership and A_i staying independent if $\Delta_2(A_i, A_{i'}, F_j) \equiv \Pi^c(A_i, F_j) + \Pi_A^n(A_{i'}) - \Pi_A^n(A_i) - \Pi^c(A_{i'}, F_j) > 0$,[30] where we can write

$$\begin{aligned}\Delta_2(A_i, A_{i'}, F_j) \;&=\; \left((\alpha_a^c - \alpha_a^n) + \gamma_1\delta_{F_j} + \gamma_3\left|x_{A_i} - x_{F_j}\right|\right)\left(\delta_{A_i} - \delta_{A_{i'}}\right) \\ &+ \left(\gamma_2 + \gamma_3\delta_{A_{i'}} + \gamma_4\delta_{F_j}\right)\left(\left|x_{A_i} - x_{F_j}\right| - \left|x_{A_{i'}} - x_{F_j}\right|\right). \qquad (8)\end{aligned}$$

Following the same arguments as above, this condition is more likely to hold if the distance between the types of academic A_i and firm F_j is lower and if the ability of A_i is larger.

[30] This is in fact the local maximization condition (6).

We now use the results of a given specific pair to discuss who collaborates, and who does not, when we take into account the whole population of academics and firms. In terms of ability, the message is clear: the higher the ability of an academic, the more likely it is that she will develop a collaborative project. There are two effects working in the same direction. First, as shown in the analysis of a specific pair, a higher δ_A has a direct positive impact on the net benefits of collaboration. Second, given that the matching is positive assortative in terms of ability, an academic with a higher δ_A is matched with a firm with a higher δ_F, which reinforces the first positive effect.

In terms of type, it is not possible to make a general statement on who collaborates and who does not even if, among the collaborating pairs, the equilibrium matching is positive assortative. This part of the equilibrium matching depends not only on the parameters of the profit function but also on the particular distribution of types on each side of the market. For example, if the types of the academics were in general more basic than those of the firms then the analysis of a given specific pair suggests that the most applied academics would collaborate, whereas the most basic academics would remain independent. Indeed, as the relative gains from collaboration with more distant partners would be lower, more basic academics would find it relatively more profitable to stay independent. In contrast, if the types of the academics were in general more applied than those of the firms then the most basic academics would collaborate. Although the former is more likely than the latter, we cannot check if any of these conditions are satisfied in the data because we do not observe the distribution of types of the whole population of firms.

6.2 Probit approach

As a second approach, we run probit regressions on the academics' likelihood of collaborating over several measures of ability and type. We use a dependent variable which has a value of 1 if the academic chooses to submit a collaborative project and a value of 0 if she submits a non-collaborative one. We control for year and university fixed effects, and report robust standard errors.

Table 4 reports the average marginal effects (again in terms of differences in predicted probabilities in the case of the categorical variables). Confirming the results of the discussion in the previous subsection, columns 1 and 2 show that the most able, measured in terms of impact and count respectively, as well as the most applied researchers, are significantly more likely to collaborate. The magnitudes of the effects are easier to compare in the case of the ranking variables. First, column 3 shows that the academics who are above the median in terms of ability and type are 3.3% and 14.1% more likely to collaborate, respectively, than those who are below the median. Given that the unconditional probability of collaborating is 35%, the increases represent 9.4% and 40.3% in terms of the unconditional probability. In column 4, we consider the rank of the academics in terms of impact and type. The regression shows that moving up the rank in any of the two characteristics has a positive and significant effect on the probability of collaboration. In terms of magnitudes, an academic who climbs $1,000$ positions in the ranking in impact increases the conditional probability of collaboration by 10% (28.6% of the unconditional probability), whereas the effect in types is 25% (71.4% of the unconditional probability). With respect to the institutional measures, column 5 shows that university variables, including size and performance, are not significant.

[Insert Table 4 here]

Taken together, our empirical results suggest that the likelihood of collaborating increases with the ability and type of the academics in the project. That is, more able and more applied academics are more likely to collaborate as opposed to staying independent. Our results also suggest that, as is the case in the matching regressions, the effects of the horizontal characteristics are stronger than those of the vertical ones.

7 Conclusion

This paper analyzes university-industry collaboration as an endogenous matching problem. We use a two-sided market matching framework of academic researchers and firms that are heterogeneous in terms of ability- and affinity-based characteristics. Depending on the value of the parameters, the matching can be positive or negative assortative in terms of ability and type of research, as well as in terms of their interactions.

Our empirical analysis uses the teams of academic researchers and firms that propose research projects to the EPSRC. We estimate the parameters of the profit function proposed in the theoretical framework using the maximum score estimation method (Fox, 2008). The estimated parameters suggest that partner abilities are complementary and that the joint profit function is decreasing in the distance between types. Moreover, individual ability and distance are substitute attributes. As a result, there is positive assortative matching in terms of ability and type while the matching is negative assortative in terms of ability-affinity pairs. Our results also suggest that the most able and the most applied academic researchers prefer to develop collaborative projects, rather than stand-alone ones.

As an alternative to maximum score estimation, we also use a probit approach which, in addition, allows us to quantify the effects. We show that affinity-based characteristics are relatively more important than ability-based ones. In general, the characteristics at the individual-researcher level are more relevant than those at the institutional level.

In this paper, we address university-industry collaborations in research projects. Still, our approach can also be used to analyze collaboration decisions in other two-sided markets. Examples of such markets could include other channels of knowledge and technology transfer, interactions between suppliers and firms, and relationships between entrepreneurs and venture capitalists. As in our set up, agents in these markets are heterogeneous along horizontal and vertical dimensions and participants can collaborate and, in some cases, stay independent.

By analyzing collaboration across institutional markets (i.e., two-sided market partnerships), this paper complements previous studies about collaboration across institutions within institutional markets (i.e., one-sided market partnerships). A natural next step would be to study the interaction between the two, and identify whether collaboration among academic researchers substitutes or complements collaboration between academic researchers and firms.

References

[1] Agarwal, R. and Ohyama, A. (2013) Industry or Academia, Basic or Applied? Career Choices and Earnings Trajectories of Scientists. *Management Science* 59 (4): 950-970.

[2] Agrawal, A. and Henderson, R. (2002) Putting Patents in Context: Exploring Knowledge Transfer from MIT. *Management Science* 48 (1): 44-60.

[3] Agrawal, A., Kapur, D. and McHale, J. (2008) How do Spatial and Social Proximity Influence Knowledge Flows? Evidence from Patent Data. *Journal of Urban Economics* 64 (2): 258-269.

[4] Akkus, O., Cookson, J.A. and Hortacsu, A. (2014) The Determinants of Bank Mergers: A Revealed Preference Analysis. SSRN W. P.

[5] Audretsch, D.B. and Stephan, P. (1996) Company-Scientist Locational Links: The case of Biotechnology. *The American Economic Review* 86 (3): 641-652.

[6] Azoulay, P., Ding, W. and Stuart, T. (2009) The Impact of Academic Patenting on The Rate, Quality and Direction of (Public) Research Output. *The Journal of Industrial Economics* 57 (4): 637-676.

[7] Banal-Estañol, A., Jofre-Bonet, M. and Meissner, C. (2015) The Impact of Industry Collaboration on Research: Evidence from Engineering Academics in the UK. *Research Policy* 44 (6): 1160-1175.

[8] Banal-Estañol, A., Macho-Stadler, I. and Pérez-Castrillo, D. (2013) Research Output from University-Industry Collaborative Projects. *Economic Development Quarterly* 27 (1): 71-81.

[9] Becker, G.S. (1973) A Theory of Marriage: Part I. *The Journal of Political Economy* 81 (4): 813-846.

[10] Bercovitz, J. and Feldman, M. (2007) Fishing Upstream: Firm Strategic Research Alliances with Universities. *Research Policy* 36 (7): 930-948.

[11] Caloghirou, Y., Ioannides, S. and Vonortas, N.S. (2003) Research Joint Ventures. *Journal of Economic Surveys* 17 (4): 541-570.

[12] Carayol, N. (2003) Objectives, Agreements and Matching in Science-Industry Collaborations: Reassembling the Pieces of the Puzzle. *Research Policy* 32 (6): 887-908.

[13] Cassiman, B. and Veugelers, R. (2006) In Search of Complementarity in Innovation Strategy: Internal R&D and External Knowledge Acquisition. *Management Science* 52 (1): 68-82.

[14] Cohen, W.M., Nelson R.R. and Walsh J.P. (2002) Links and Impacts: The Influence of Public Research on Industrial R&D. *Management Science* 48 (1): 1-23.

[15] Cockburn, I. and Henderson, R. (1998) Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery. *Journal of Industrial Economics* 46 (2): 157-182.

[16] Dahlin, K.B, Weingart, L.R., and Hinds, P.J. (2005) Team diversity and information use. *Academy of Management Journal* 48 (6): 1107-1123.

[17] Dasgupta, P. and David, P. (1994) Towards a New Economics of Science. *Research Policy* 23 (5): 487-522.

[18] Fabrizio, K. and DiMinin, A. (2008) Commercializing the Laboratory: Faculty Patenting and the Open Science Environment. *Research Policy* 37 (5): 914-931.

[19] Florida, R. and Cohen, W.M. (1999) Engine or Infrastructure? The University Role in Economic Development. In Branscomb, L.M., Kodama, F., Florida, R. (eds.), *Industrializing Knowledge: University-Industry Linkages in Japan and the United States.* MIT Press, London, 589-610.

[20] Fox, J. (2008) Estimating Matching Games with Transfers. NBER W. P.

[21] Fox, J. (2010) Identification in Matching Games, *Quantitative Economics*, 1 (2), 203-254.

[22] Godin, B. (1996) The State of Science and Technology Indicators in the OECD Countries. Research Paper, Statistics Canada.

[23] Goldfarb, B. (2008) The Effect of Government Contracting on Academic Research: Does the Source of Funding Affect scientific output? *Research Policy* 37: 41–58.

[24] Gompers, P., Mukharlyamov, V. and Xuan, Y. (forthcoming) The Cost of Friendship. *Journal of Financial Economics.*

[25] Hamilton, K.S. (2003) Subfield and level classification journals. CHI Research N^o 2012-R, January 16.

[26] Hegde D. and Tumlinson, J. (2014) Does Social Proximity Enhance Business Partnership? Theory and Evidence from Ethnicity role in U.S. Venture Capital. *Management Science* 60(9), 2355-2380.

[27] Jensen, R.A., Thursby, M.C. (2001) Proofs and Prototypes for Sale: The Licensing of University Inventions. *The American Economic Review* 91 (1): 240–259.

[28] Jensen, R., Thursby, J. Thursby, M. (2003) Disclosure and Licensing of University Inventions: The Best We Can Do with the S**t We Get to Work with. *International Journal of Industrial Organization* 21 (9): 1271–1300.

[29] Kamien, M., Muller, E. and Zang, I. (1992) Research Joint Ventures and R&D Cartels. *The American Economic Review* 82 (5): 1293-1306.

[30] Katz, J.S. and Martin, B.R. (1997) What is Research Collaboration? *Research Policy* 26 (1): 1-18.

[31] Lee, Y.S. (2000) The Sustainability of University-Industry Research Collaboration: An Empirical Assessment. *The Journal of Technology Transfer* 25 (2): 111-133.

[32] Legros, P. and Newman, A. (2002) Monotone Matching in Perfect and Imperfect Worlds. *Review of Economic Studies* 69 (4): 925-942.

[33] Lindenlaub, I. (2014) Sorting Multidimensional Types: Theory and Application. European University Institute W. P.

[34] Louis, K.S. , Blumenthal, D., Gluck, M.E. and Stoto, M.A. (1989) Entrepreneurs in Academe: An Exploration of Behaviors among Life Scientists. *Administrative Science Quarterly* 34 (1): 110-131.

[35] Manski, C. F. (1975) Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics* 3(3): 205-228.

[36] Matzkin, R.L. (1993) Nonparametric Identification and Estimation of Polychotomous Choice Models. *Journal of Econometrics* 58: 137-168.

[37] Mansfield, E. (1995) Academic Research Underlying Industrial Innovation: Sources, Characteristics and Financing. *Review of Economics and Statistics* 77: 55-65.

[38] Mindruta, D. (2013) Value Creation in University-Firm Research Collaborations: A Matching Approach. *Strategic Management Journal* 34 (6): 644-665.

[39] Mohnen, P. and Hoareau. C. (2003) What Type of Enterprise Forges Close Links with Universities and Government Labs? Evidence from CIS 2. *Managerial and Decision Economics* 24 (2-3): 133-145.

[40] Narin, F., Pinski G. and Gee H. (1976) Structure of the Biomedical Literature. *Journal of the American Society for Information Science* 27 (1): 25-45.

[41] Pereira, I. (2007) Business-Science Research Collaboration under Moral Hazard. W. P. Universitat Autonoma de Barcelona.

[42] Perkmann, M., Tartari, V., McKelvey, M., Autio, E., Broström, A., D'Este, P., Fini, R., Geuna, A., Grimaldi, R., Hughes, A., Kitson, M., Krabel, S., Llerena, P., Lissoni, F., Salter, A. and Sobrero, M. (2013) Academic Engagement and Commercialization: A Review of the Literature on University-Industry Relations. *Research Policy* 42 (2): 423-442.

[43] Roth, A. and Sotomayor, M. (1990) *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Cambridge University Press, New York and Melbourne, Econometric Society Monographs.

[44] Santiago, D., and Fox, J. (2008) A Toolkit for Matching Maximum Score Estimation and Point and Set Identified Subsampling Inference, NBER working paper.

[45] Shapley, Ll.S. and Shubik, M. (1972) The Assignment Game I: The Core. *International Journal of Game Theory* 1 (1): 111-130.

[46] Storn R. and Price, K. (1997) Differential evolution: A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization* 11: 341–359.

[47] van Looy, B., Callaert, J. and Debackere, K. (2006) Publication and Patent Behaviour of Academic Researchers: Conflicting, Reinforcing or Merely Co-existing? *Research Policy* 35 (4): 596-608.

[48] Veugelers, R. and Cassiman, B. (2005) R&D Cooperation between Firms and Universities. Some Empirical Evidence from Belgian Manufacturing. *International Journal of Industrial Organization* 23 (5-6): 355-379.

[49] Wagner, C.S. and Leydesdorff L. (2005) Network Structure, Self-organization, and the Growth of International Collaboration in Science. *Research Policy* 34 (10): 1608-1618.

[50] Yang, Y., Shi, M., and Goldfarb, A. (2009) Estimating the Value of Brand Alliances in Professional Team Sports, *Marketing Science* 28 (6): 1095–1111.

[51] Zucker, L., Darby M.R. and Armstrong J.S. (2002) Commercializing Knowledge: University Science, Knowledge Capture, and Firm Performance in Biotechnology. *Management Science* 48 (1): 138 - 153.

Table 1. Descriptive statistics

We report the descriptive statistics for the vertical attributes (normal count, impact-weighted sum and average impact of the publications in the previous six years) of the team of academics, the PI and the firms, the horizontal attributes (type of the publications), and two aggregate characteristics for each side of the market (university research funds, university active researchers, firms' turnover and firms' employees).

Vertical attributes	Observations	Mean	St dev	Median	Min	Max	Corr. academics' impact	Corr. PI's impact	Corr. firms' impact
Academics' count (in hundreds)	5855	0.106	0.146	0.059	0.000	2.237	0.889***	0.502***	0.172***
Academics' impact (in hundreds)	5855	0.169	0.309	0.070	0.000	5.261	1	0.636***	0.219***
Academics' average impact	5688	1.392	1.013	1.177	0.000	19.190	0.361***	0.479***	0.212***
PI's count (in hundreds)	5067	0.036	0.037	0.027	0.000	0.350	0.534***	0.854***	0.058***
PI's impact (in hundreds)	5067	0.056	0.081	0.028	0.000	0.920	0.636***	1	0.137***
PI's average impact	4870	1.355	1.055	1.124	0.000	19.190	0.328***	0.501***	0.212***
Firms' count (in thousands)	2057	0.749	1.836	0.080	0.000	17.625	0.223***	0.120***	0.901***
Firms' impact (in thousands)	2057	1.448	5.173	0.066	0.000	58.543	0.219***	0.137***	1
Firms' average impact	1634	1.210	0.942	0.944	0.000	9.620	0.155***	0.148***	0.467***

Horizontal attributes	Observations	Mean	St dev	Median	Min	Max	Corr. PI's type	Corr. firms' type	Corr. respect impact
Academics' type	5519	0.656	0.328	0.750	0.000	1.000	0.947***	0.368***	-0.343***
PI's type	4674	0.666	0.343	0.780	0.000	1.000		0.343***	-0.396***
Firms' type	1563	0.579	0.284	0.600	0.000	1.000			-0.123***

Aggregate attributes	Observations	Mean	St dev	Median	Min	Max	Corr. university active res.	Corr. firms' turnover	Corr. firms' employees
University research funds (in millions)	5933	6.259	3.974	5.484	0.123	12.167	0.908***	0.089***	0.111***
University active researchers (in tens)	5933	14.243	7.220	12.875	0.800	29.310		0.118***	0.106***
Firms' turnover (in millions)	1580	10.830	25.128	1.150	0.000	178.525			0.538***
Firms' employees (in ten thousands)	1549	3.524	6.306	0.819	0.000	47.250			

Table 2. Maximum score estimation for the linear production function.

This table reports the results of the maximum score estimation method for a linear production function of vertical and horizontal pairwise characteristics of academics and firms. Column 1 uses, as measures of ability and type, the impact and the type of the academic and the impact and the type of the firm; Columns 2 and 3 use, instead of the impact, the count and the average impact of the academic and the count and the average impact of the firm; and Column 4 uses the measures of the PI instead of those of the academic team. In all these specifications, we include the product of university research funds and firm's turnover as an additional interaction variable to ensure identification. Column 5 uses the same measures as Column 1 but reports the effect of the product of types instead of the effect of the distance. Column 6 uses, as measures of ability and type, the total research funds and the average type of the university the PI belongs to and the turnover and the type of the firm; Column 7 uses, as measures of ability and type, the total number of researchers and the average type of the university the PI belongs to and the number of employees and the type of the firm. In the last two specifications, we include the impact of the academic and of the firm as an additional interaction variable to ensure identification. Columns 8 and 9 use the same measures as Columns 1 and 2 but include the difference between the individual effects of the ability when the academic works in collaboration and when she does not. We report the 95% confidence interval of each coefficient.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Measure of ability	Impact	Count	Average impact	Impact	Impact	Size	Performance	Impact	Count
Academic unit	Academic team	Academic team	Academic team	PI	Academic team	University	University	Academic	Academic
Academics' ability*Firms' ability	7.33	12.57	27.75	18.64	4.67			14.84	24.43
	(6.49, 10.24)	(7.31, 18.19)	(25.95, 31.92)	(8.19, 25.84)	(3.56, 7.44)			(7.54, 19.47)	(21.57, 33.45)
Type distance	-17.42	-16.40	-3.78	-13.91				-27.45	-24.55
	(-21.10, -15.08)	(-24.73, -13.55)	(-5.84, 2.05)	(-19.37, -12.81)				(-37.80, -23.30)	(-28.15, -22.92)
Academics' ability*Type distance	-6.47	-7.52	-11.05	-51.89				-30.57	-23.01
	(-8.93, -2.45)	(-8.31, 2.78)	(-15.32, -3.40)	(-101.29, -33.71)				(-52.94, 25.57)	(-31.29, 17.70)
Firms' ability*Type distance	-4.39	-5.47	-4.35	-2.74				-3.32	-4.68
	(-7.93, -4.15)	(-10.78, -3.28)	(-14.14, -4.09)	(-6.03, 0.59)				(-9.47, 5.32)	(-12.22, -1.96)
Academics' type*Firms' type					49.83				
					(46.75, 83.28)				
Unis' agg ability*Firms' agg ability						0.14	0.07		
						(0.01, 0.20)	(-0.01, 0.11)		
Unis' type*Firms' type						6.38	7.4		
						(-9.61, 10.42)	(-11.46, 10.01)		
Diff. individual effect academic ability								-4.10	-7.59
								(-31.61, 5.07)	(-30.20, -2.97)
Total inequalities	98701	98701	98701	77586	98701	98701	98701	534626	534626
% satisfied	64.3	64.3	64.9	62.2	64.2	57.0	56.8	62.7	62.6

Table 3. Results of the probit-counterfactual approach for the ranking measures.

This table presents the results of probit regressions for the probability of an academic and firm to partner with each other based on a set of pairwise ranking characteristics. The dependent variable is a dummy variable equal to one if the collaboration between the pair takes place (actual pairs) and zero otherwise (counterfactual pairs). Independent variables are pairwise variables accounting for the relative position of each agent in each side of the market. Both below/above the median in impact/type, and Both in the 1st/2nd/3rd/4th quartile in impact are dummy variables which take the value of one if the condition is satisfied and zero otherwise. Distance of quartiles/rank in impact/type is a discrete variable which measures the absolute value of the difference between the rank order/quartile number of the academic and the firm in terms of impact/type. Columns 1-2 report the differences in predicted probabilities of matching for each of the dummy variables (dummy equal to 1 less dummy equal to 0) while keeping all the others at their mean levels. Columns 3-4 report the marginal effects of each variable while keeping the others at their mean levels. We report robust standard errors clustered at the academic level in brackets. Year and sector fixed effects are included. ***, ** and * indicate that coefficients are significant at the 1%, 5%, and 10% levels, respectively.

	(1) Categories (medians)	(2) Categories (quartiles)	(3) Distances of Quartiles	(4) Distances of Ranks
Both above median in impact	0.014			
	[0.010]			
Both below median in impact	0.012			
	[0.014]			
Both in 1st quartile in impact		0.061***		
		[0.017]		
Both in 2nd quartile in impact		0.028*		
		[0.016]		
Both in 3rd quartile in impact		0.033**		
		[0.016]		
Both in 4rd quartile in impact		0.038		
		[0.076]		
Both above median in type	0.069***	0.069***		
	[0.012]	[0.012]		
Both below median in type	0.084***	0.077***		
	[0.012]	[0.012]		
Distance of quartiles in impact (1 unit)			-0.040***	
			[0.011]	
Distance of quartiles in impact (2 units)			-0.038***	
			[0.013]	
Distance of quartiles in impact (3 units)			-0.031*	
			[0.018]	
Distance of quartiles in types (1 unit)			-0.025**	
			[0.012]	
Distance of quartiles in types (2 units)			-0.080***	
			[0.013]	
Distance of quartiles in types (3 units)			-0.112***	
			[0.015]	
Distance of rank of impact				-0.022**
				[0.011]
Distance of rank of types				-0.101***
				[0.011]
Observations	7323	7323	7323	7323

Table 4. Results of the probit regresssions on the collaboration decision.

This table reports the marginal effects of the probit regressions for the probability that an academic collaborates with a firm based on a set of observable characteristics. The dependent variable is a dummy variable equal to one if the academic collaborates, and zero if she stays independent. Independent variables are characteristics of the academic team. Columns 1 and 2 use, as measures of ability, the impact and the count, respectively. Column 3 displays the marginal effects on the probability of matching for academics above the median, in terms of impact and type, having as reference those that are below the median. Column 4 displays the effects of rank. The last column adds to the specification in Column 1 aggregate measures for the university (university research funds and number of staff). We report robust standard errors clustered at the academic level in brackets. Year fixed effects are included in all regressions, and university fixed effects are included in all the regressions but the last one. ***, ** and * indicate that coefficients are significant at the 1%, 5%, and 10% levels, respectively.

	(1) Impact	(2) Count	(3) Medians	(4) Ranks	(5) Aggregate
Academics' ability	0.092**	0.259***			0.084**
	[0.038]	[0.058]			[0.037]
Academics' type	0.234***	0.231***			0.233***
	[0.022]	[0.021]			[0.021]
Academics above median in impact			0.033**		
			[0.015]		
Academics above median in type			0.141***		
			[0.015]		
Rank academics' impact				0.100***	
				[0.024]	
Rank academics' type				0.250***	
				[0.024]	
University performance					-0.013
					[0.026]
University size					0.048
					[0.030]
Observations	5513	5513	5513	5513	5517