DO BUSINESSES PAY TO DO SCIENCE? THE EFFECTS OF HIRING PHDS ON FIRMS' PATENTING PROCESS

Shawn J. Mushtaq, James M. Sappenfield, and Dongbo Shi*

Preliminary Draft. Please do not cite or distribute without permission

May 5, 2015

Abstract: Previous research studying labor practices among science-oriented firms identifies that 'scientists pay to be scientists', as postdoctoral fellows accept lower wages to work at science-oriented firms even when controlling for ability. We reverse the question, and ask the opposite: do science-oriented firms, who recruit doctorate-level researchers, pay to practice science? We find that they do not. Specifically, we present evidence that the inclusion of doctorate-level researchers in the R&D process leads to an increase in the stated reliance of firm inventors on scientific literature and conferences as sources of knowledge for patent development. This increase in scientific source reliance due to involvement of doctoral inventors pairs with no observable change in commercial motivations for patenting. To reach this result, we analyze PatVal-EU survey data on 6,122 patents involving firm applicants. Prior to analysis, we restore missing values via multiple imputation to create ten balanced panels of 6,769 observations (restoring 9.5% of the original sample). We then sort observations into matched treatment (inclusion of doctoral-level inventors in the patent R&D process) and control (lack thereof) pairs based on coarsened exact matching methods. We do so in order to mimic the ideal experiment where covariates between matched groups are balanced, except along the treatment of interest (including doctoral-level inventors in the firm R&D process) and related post-treatment variables. Following this matching, we utilize probit regression to demonstrate that assignment to the treatment condition increases science-orientation, and has no impact on stated drive to commercialize.

*Mushtaq: Masters Candidate, Harvard University; Sappenfield: Doctoral Student, Harvard Business School; Shi: Fellow, Harvard University / Doctoral Candidate, Tsinghua University; We gratefully acknowledge and thank Myriam Mariani and Paola Giuri for their graciousness in sharing the data necessary to replicate and extend the results of their paper, as well as for their continuing advice on proceeding with the current analysis. We additionally wish to thank several others. Thanks go to Martin Saveski and Abdullah Almaatouq for their review of our initial replication results of Giuri and Mariani (2013), as well as for very helpful suggestions on how to proceed with our extension. We are further indebted to Professor Gary King for a fantastic course of instruction, as well as Stephen Pettigrew and Solédad Prillaman for their excellent review sections and plentiful answers to our numerous methodological questions. All three deserve thanks for incredible responsiveness to questions via email at all times.

1 Introduction

In a well-known examination of labor market outcomes for biological postdoctoral fellows, scientists were found to pay to be scientists. Particularly, the biology postdoctoral fellows in the sample studied received job offers awarding less pay the more scientifically-oriented a firm was (Stern 2004). We are interested in asking a similar but opposite question. Namely, do firms pay to do science and to be science-oriented, as a function of their workforce? In study specific terms, as firms involve individuals with greater scientific orientation and training into their research and development processes, do those firms suffer in terms of their drive to commercialize the fruits of their labor, products and research, through patenting?

The scientific firm's research and development process can be modeled as a black box. On one side, the scientific firm selects a series of inputs (capital and labor with specific types of training and education), and produces a series of outputs (research, products, technologies, science) with the goal of commercializing some or all of those outputs. We make standard microeconomic assumptions going forward, assuming that a given firm is profit-maximizing, and that firms seeks to maximize the rate of commercialization of new technologies, research, and products in order to profit. We consider patenting to be a by-product of the firms R&D processes.

Following this toy-model, we examine two questions in the analysis below:

- **Question 1:** How do changes in the involvement of doctorate-educated inventors in a firm's R&D process lead to changes in the stated scientific orientation of the firm?
- **Question 2:** How do changes in the involvement of doctorate-educated inventors in a firm's R&D process impact the importance of commercial exploitation in as a reason driving a firm to patent?

Key to our analysis, we assume that the composition of the firm labor and talent pool (i.e., the hiring of individuals with doctorate education and thereby scientifically-oriented training) is a discrete choice made by the firm in order to accomplish firm research and development processes. It is the impact of this decision on firm patenting behavior with which we are interested.

In order to understand the relationship between firms talent pools, their scientific education and training, and the patented projects produced by the firm innovation process, it is crucial to understand how knowledge diffuses and how diffusion relates to knowledge creation. Knowledge diffusion is important to the generation of new ideas, science, and technologies, as innovation in ideas, science, and technologies among individuals follows a model of recombination of known knowledge (Allen 1977). Knowledge diffusion from one scientist or inventor to another therefore promotes knowledge generation and recombination.

Such knowledge diffusion depends heavily on geographic distance. As potential collaborations become more distant, the frequency of knowledge spillover drops-off exponentially, both in transmission of knowledge in micro-spaces (Allen 1977) and in macro spaces, where knowledge-spillovers

result in geographic clustering as evidenced by patent flows (Fleming, King and Juda 2007, Jaffe, Trajtenberg and Henderson 1993).

Giuri and Mariani (2013) study patterns of knowledge diffusion among inventors in the PatVal-EU survey of European patents and inventors, and find that the presence in patent projects of inventors with advanced education may moderate the impact of distance on knowledge flows. This effect is evidenced by results that patents with inventors possessing advanced education (defined as a university, masters, or doctorate degrees) rely significantly more on distant relative to close unexpected collaborations in order to develop the inventions associated with the patents awarded. This effect is particularly pronounced among patents with inventors who possess doctorate degrees (Giuri and Mariani 2013).

We successfully replicated the results of Giuri and Mariani (2013) and begin our analysis using the data from that study. Please refer to that study for detailed information regarding the construction of the survey and variables utilized in the current analysis. We use many of the same variables and will only explicitly detail variables and their construction where we need to introduce new variables for analysis.[1]

1.1 Organization of the Paper

The paper proceeds as follows. In section 2, we discuss the data and methods used to handle missingness in the data, as well as how we structure the data to investigate our questions of interest, including the creation of a matched sample using coarsened exact matching, where patent observations are matched on whether the observation involves an inventor who possess advanced education or not. Section 3 then discusses our analysis plan. Section 4 presents results, first on the impact of a scientifically-educated talent pool on the science-orientation of a firm's R&D outcomes, and then on the impact of a scientifically-educated talent pool on the importance of commercial exploitation as a reason driving patent behavior. Section 5 concludes.

2 Data and Identification

In order to study how varying a firm's scientific talent pool relates to patterns of collaboration, we begin with 7,527 observations from the PatVal-EU Survey of European Inventors. PatVal-EU is a May 2003 to January 2004 survey directed at inventors of 27,531 patents granted by the European Patent Office from $1993C1997$, focusing on France, Germany, Italy, the Netherlands, Spain, and the United Kingdom. The survey received 9,216 responses covering 9,017 patents (Gambardella, Giuri and Mariani 2005). We utilize the same subset of responses as Giuri and Mariani (2013) and therefore do not include French, Danish, or Hungarian inventors due to systematic and non-

[1] In tables 6 and 7 at the end of the paper we list all variables utilized from the dataset, along with a short description to assist in matching variables with those in Giuri and Mariani (2013)

ignorable missingness of key variables among those countries. This leaves us with 6,769 patent observations, accounting for 6,061 patents where the applicant is a firm.

2.1 Multiple imputation of missing observations

Of the original sample of 7,527 patent observations, approximately 20% involve missing-at-random data, where we can reasonably predict missing cells using observed data. Missing-at-random is a reasonable assumption of the pattern of missingness in the data, as survey response rates are likely a product of observation characteristics, and we have extensive data with which to predict missing observations. We therefore use multiple imputation methods (Rubin 2008) to create a balanced panel for data analysis. Specifically, the R package and algorithms of Amelia II (Honaker, King and Blackwell 2011) are used to bootstrap and impute data for 10 datasets from which we then conduct our analysis.[2] Listwise deletion of missing data among key variables would otherwise reduce the data to 6,051 useable observations for data analysis. Multiple imputation, on the other hand, allows us the use of 6,769 observations in our balanced panels.[3] Table 1 lists key variables and compares descriptive statistics for both the useable observations in the balanced panel created from listwise deletion in the original dataset and the average the multiply imputed data sets, with 6,769 patent observations.

Following multiple imputation, we reduce our sample to firm-relevant observations, in order to limit to only situations where the decision to include a doctoral level researcher in firm R&D operations is conceivably separable from patenting outcomes (as opposed to observing patents driven by lone inventors or by public research institution inventors). We do so by limiting our sample to those observations listing a firm as the patent applicant, focusing the sample on 6,061 patents among 1,813 firms. This sample constitutes our unmatched sample in analyses.

2.2 Matching on the inclusion of PhD level scientists in firm R&D processes

Giuri and Mariani's (2013) emphasis of the important role education plays in determining the flow of knowledge invite a comparison of matched samples to explore how the presence of a scientifically educated and trained talent pool impacts relevant patent-level firm outcomes. Consequently, we create a matched dataset where patents involving doctorate-level inventors (the treatment group) are matched with patents where no inventors received doctorate-level training (control group). We do so via coarsened exact matching on a series of covariates arguably determined prior to treatment.

Specifically, we match upon inventor characteristics (Age, Gender, Inventor Experience Herfindahl Index, Inventor Breadth of Experience, Recent Within and Without NUTS3 Region mobility, Degree in the Country of Patent, log of Experience, breadth of experience, and the share of Far

[2]Following results on power and accuracy analysis of multiply imputed data sets, we determined 10 data sets would be optimal to determine effect sizes (Graham 2007).

[3]Please reference tables 6 and 7 for the list of variables utilized in multiple imputation, as well as those withheld from imputation due to collinearity issues.

Table 1: Descriptive Statistics of Pre-Imputed Data and Post-Imputed Data

	Pre-Imputed		*Post-Imputed*	
	Mean	St.Dev	Mean	St.Dev
Inventor characteristics				
Near	0.31	0.46	0.31	0.46
Far	0.41	0.49	0.41	0.49
Age	44.76	9.61	45.31	9.77
Gender	0.98	0.15	0.98	0.15
High School Degree	0.18	0.39	0.20	0.40
University Degree	0.55	0.50	0.53	0.50
PhD Degree	0.27	0.44	0.27	0.44
Far Past Coinventors	0.60^a	0.38^a	0.30	0.41
Experience	1,989.66	5.49	1,989.69	5.48
Science	2.63	1.86	2.59	1.87
Country of Degree Dummy	0.04	0.19	0.04	0.19
Mobility in Region	0.07	0.25	0.06	0.24
Mobility out Region	0.20	0.40	0.18	0.39
Experience Herfindahl	0.78^b	0.26^b	0.49	0.43
Previous Patents Dummy	0.64	0.48	0.64	0.48
Inventor Past Patents	5.04	12.10	1.07	1.07
Conference	1.72	1.73	1.70	1.74
Applicant controls				
Employees	$89,052.34^c$	$116,337.40^c$	118,187.70	723,011.50
log Employees	7.78	4.53	9.54	2.70
R&D Intensity	0.05^d	0.03^d	0.05	0.03
Number of Inventors	2.36	1.49	2.35	1.51
Commercial Expltitation	3.79	1.55	3.78	1.56
Licensing	2.07	1.52	2.07	1.54
Prevent Imitation	3.81	1.55	3.79	1.57
Regional and other controls				
GDPPC	23,202.06	9,165.44	23,026.45	9,077.62
Pop	737.35	897.27	773.96	957.09
Area	1,519.40	1,934.58	1,584.10	2,052.00
Region Patents	123.72	137.40	122.57	135.97
TOP1% in Technology	0.15	0.36	0.14	0.35
NR Research Universities	0.56	0.78	0.58	0.81
Research University Score	1.20	6.01	1.12	5.77

a.N=3134,b.N=3877,c.N=4850,d.N=2706

Past Coinventors) and characteristics of the firm that are reasonably immutable by the inclusion of individual inventors in firm processes (the country of patent). Bins for coarsening were determined through examining the ranges of variables, as well as comparative histograms and percentiles for treatment relative to control sets in each of the variables used in the matching process. See Table 2 for the covariates used in matching specifications and balance comparisons.

Generally, means are relatively balanced across covariates in the matched dataset. The matched data achieves lower multivariate L1-Imbalance on match covariates than the unmatched observations. However, as some means in the unmatched data are closer than in the matched data, it is possible that we may be able to improve the matching by adjusting the CEM bins further.

3 Empirical Analysis: Model Specifications

Given the assumption that the pre-treatment variables specified in section 2.1 predict treatment fully, we are able to investigate the feasible sample average treatment effect on the treated (FSATT). Here, this is the effect on patent-level outcomes of involving individuals with doctorate education and vis-a-vis individuals with a scientific orientation in the firms research and development process. We are comfortable making this assumption given the broad-spectrum of our pre-treatment variables.

To analyze the impact of firm selection of a scientifically educated talent pool on science orientation, and then the impact of the treatment on commercial exploitation as a driving reason for firm patenting, we utilize probit regression involving a series of outcome and control variables.

3.1 Outcomes of Interest: Reliance on Science and Conferences or Workshops

In the first probit regressions, we use outcome variables from source of knowledge questions in the PatVal-EU survey requiring the respondent to indicate whether a patented invention relied on scientific literature (0−not used to 5−very important) and conferences and workshops (0−not used to 5−very important) as sources of knowledge. Following Giuri and Mariani(2013), we restructure these variables as dichotomous (0/1) variables to indicate whether scientific literature or conferences/workshops were used as a source of knowledge in the inception of a patent.

3.2 Outcomes of Interest: Commercialization as a Reason for Patenting

We additionally are interested in analyzing the importance of commercialization for firm patenting, relative to other potential reasons, as a result of changes in the presence of doctoral educated inventors in a firm's research and development processes, as well as the consequent changes in the scientific orientation of the firm, measured by reliance on scientific literature or reliance on conferences or workshops. We conduct this analysis by running a regression involving variables formulated from reasons for patenting questions in the PatVal-EU survey requiring respondents to

Table 2: Descriptive Statistics of Unmatched Data and Matched Data

	Unmatched				*Matched*			
	Mean		*St.Dev*		*Mean*		*St.Dev*	
	Teated	Control	Teated	Control	Teated	Control	Teated	Control
Age	44.76	49.29	9.31	9.82	44.56	44.60	9.19	9.27
Gender	0.98	1.07	0.15	0.14	0.99	0.99	0.11	0.08
Experience Herfindahl	0.57	0.62	0.38	0.43	0.54	0.42	0.42	0.45
log Inventor Past Patents	1.55	1.65	1.22	0.98	1.26	0.74	1.16	0.91
Experience Breadth	0.46	0.49	0.50	0.45	0.34	0.17	0.47	0.38
Far Past Coinventors	0.46	0.48	0.42	0.40	0.41	0.25	0.43	0.39
Country of Degree Dummy	0.06	0.06	0.24	0.17	0.02	0.01	0.15	0.11
Mobility in Region	0.05	0.05	0.21	0.25	0.03	0.02	0.17	0.15
Mobility out Region	0.22	0.24	0.42	0.38	0.18	0.16	0.38	0.37
Observations			6122				3910	
L1-Imbalance			0.767				0.657	

indicate the importance of commercial exploitation, licensing, and prevention of imitation (all 0 to 5) to the decision to patent. From these variables, we construct a ratio (Com_ratio) of the stated importance of Commercial Exploitation relative to the sum of all three stated levels of importance, resulting in outcomes of interest bounded between 0 and 1[4].

3.3 Covariates: Doctoral Educated Talent Pool and additional controls

The key covariate in which we are interested in is presence of inventors on a patent, and within a firm's R&D process, who possess doctorate level training. This is indicated by a dichotomous (0/1) variable, where observations receive a 1 if at the time of invention an inventor had obtained a doctorate degree (or equivalent) and 0 if otherwise. We also include a series of control variables at the inventor, firm, and region levels, in line with the specifications in Giuri and Mariani (2013). We regress both sets of specification against the outcome variables discussed above in probit regressions. For the full model specification, we refer you to our results in table 3 (un-matched, Q1 analysis), table 4 (matched, Q1 analysis) and table 5 (Q2 analysis).

We run specifications in both analyses against the unmatched data, and then against the matched data, using the R package Zelig to calculate results for probit regressions across the 10 multiply imputed data sets (Kosuke, King and Lau, 2007; Kosuke, King and Lau, 2008).

4 Results

In this section, we first address the results of our analysis of the reliance on scientific literature and conferences. We then address results pertaining to the relationship of the relative importance to the firm of commercializing patents. We currently present probit regressions. In any future iterations of this paper, however, we plan to present the more interpretable marginal effects alongside the current probit coefficients.

4.1 Results of First Analysis: Reliance on Science and Conferences

Table 3 shows results of regressions of 'Science' and 'Conferences' dichotomous variables against covariates, including fixed effects for the application year and NUTS2 Region (model 1 and 3), as well as applicant year, NUTS3 Region and NUTS3 Region-related controls. These are regressions on the unmatched data set. In these observations, the presence of an inventor with a PhD Degree on a patent project has a positive and significant effect on the reported reliance by the invention team on scientific literature and conferences in developing the patented invention.

The effect is the third largest of all positive coefficients in both science and both conference specifications. The effect of R&D intensity is substantially higher, indicating that the presence of

[4]Some outcomes exceeded 1 or fell below 0 due to multiple imputation. In this situation, we censored the ratio at 1 and 0 respectively

Table 3: Probit Estimation of Unmatched Data:Effect of PhD Degree on Conferences

	Science		*Conferences*	
	Model 1	Model 2	Model 3	Model 4
Near	0.461***	0.500***	0.467***	0.488***
	(0.057)	(0.06)	(0.049)	(0.051)
Far	0.348***	0.348***	0.555***	0.574***
	(0.05)	(0.053)	(0.044)	(0.047)
Age	-0.013***	-0.013***	-0.007***	-0.007***
	(0.002)	(0.002)	(0.002)	(0.002)
Gender	-0.163	-0.124	-0.198	-0.133
	(0.162)	(0.17)	(0.127)	(0.13)
PhD Degree	0.395***	0.442***	0.306***	0.350***
	(0.06)	(0.065)	(0.049)	(0.053)
Far Past Coinventors	0.115	0.112	0.001	0.014
	(0.075)	(0.081)	(0.067)	(0.072)
Coinventors Dummy	0.005	0.025	0.161**	0.148**
	(0.075)	(0.081)	(0.07)	(0.074)
Experience	-0.005	-0.007	-0.005	-0.006
	(0.006)	(0.006)	(0.005)	(0.005)
Country of Degree Dummy	0.190*	0.215*	0.192*	0.178*
	(0.113)	(0.124)	(0.099)	(0.104)
Mobility in Region	0.110	0.132	0.070	0.112
	(0.085)	(0.09)	(0.077)	(0.08)
Mobility out Region	-0.006	-0.010	0.059	0.066
	(0.054)	(0.058)	(0.049)	(0.051)
Experience Herfindahl	-0.222**	-0.103	0.032	0.080
	(0.107)	(0.114)	(0.094)	(0.099)
Previous Patents Dummy	0.163	0.058	-0.057	-0.101
	(0.123)	(0.132)	(0.109)	(0.115)
log Employees	0.013	0.015	0.011	0.009
	(0.011)	(0.012)	(0.008)	(0.009)
R&D Intensity	2.442***	3.019***	1.507**	2.078**
	(0.889)	(0.962)	(0.745)	(0.801)
Inventor Past Patents	0.088***	0.099***	0.068***	0.072***
	(0.016)	(0.017)	(0.013)	(0.014)
Application Year FE	YES	YES	YES	YES
NUTS2 Region FE	YES		YES	
NUTS3 Region FE		YES		YES
Regional Controls		YES		YES

N=6122,$^{*}p < 0.1$,$^{*} * p < 0.05$,$^{*} * *p < 0.01$

R&D funding plays more of a role than that of a doctorate-educated inventor in both relationships. Additionally, the presence of near inventors has only slightly more impact on project reliance on scientific literature than the presence of a doctorate-level inventor. This may potentially be explained by the importance of clustering in scientific knowledge flows mentioned earlier. Far collaborators are relatively more important for conferences, indicating a potential logical link between conference attendance and the reliance on distant collaborators.

Table 4 shows these same model specifications, except on the matched data set. Here we see that the positive impact of the inclusion of doctoral inventors on reliance on scientific literature is even more pronounced, and is essentially equivalent to that of near collaborators. The effect of including doctoral inventors on patent projects for conferences likewise becomes more pronounced, but less so, and still remains the third most impactful covariate in determining reliance on conferences as a source of knowledge.

Matching served to reinforce the link already evidenced in the unmatched regressions between inclusion of doctoral inventors in firm patenting projects, and the 'Science' and 'Conferences' dichotomous variables, when controlling for other variables. This is indicative of two key conclusions. First, there likely exists a causal relationship between the inclusion of employees with doctoral training in firm research and development and the reliance of firm projects on scientific literature and on conferences / workshops. Second, these behaviors likely are indicative of an increase in the science-orientation of the firm and its activities.

4.2 Results of Second Analysis: Importance of Commercialization in Patenting

We now move to examine the second outcome of interest, the stated importance of commercial exploitation as a reason for invention patenting at the project level within a firm. Recall that this outcome is structured as a ratio of the stated importance of 'commercial exploitation' over the sum importance of all surveyed reasons for patenting, and thereby provides a relative measure of the importance of commercial exploitation as a reason for patenting. Table 5 presents the relevant regressions.

These regressions produce largely close to zero and insignificant coefficients for almost all covariates. Running the same specifications as before in the matched data against the dichotomous (0/1) variables, we obtain essentially the same result. Only age is significant in its effect, while still relatively much smaller in magnitude than prior specifications.

Assuming the direction of the effect of doctoral-level inventor presence is estimated correctly, despite the relatively large standard errors, would indicate an interesting result. While essentially indistinguishable from zero, the estimated relationship is potentially positively correlated, which may indicate that the presence of doctoral level researchers actually introduces higher rates of stated reliance on commercial exploitation as a reason for patenting. We cannot claim this result, however, without more finely-tuned covariates for the controls and dependent variable.

Table 4: Probit Estimation of Matched Data:Effect of PhD Degree on Conferences

	Science		*Conferences*	
	Model 1	Model 2	Model 3	Model 4
Near	0.495***	0.540***	0.477***	0.498***
	(0.072)	(0.078)	(0.062)	(0.067)
Far	0.354***	0.368***	0.548***	0.591***
	(0.063)	(0.069)	(0.057)	(0.063)
Age	-0.014***	-0.013***	-0.007**	-0.007**
	(0.003)	(0.003)	(0.003)	(0.003)
Gender	0.350	0.488	0.133	0.344
	(0.294)	(0.325)	(0.261)	(0.282)
PhD Degree	0.499***	0.532***	0.393***	0.434***
	(0.074)	(0.08)	(0.06)	(0.064)
Far Past Coinventors	-0.091	-0.058	-0.106	-0.071
	(0.109)	(0.119)	(0.095)	(0.103)
Coinventors Dummy	0.096	0.134	0.180	0.154
	(0.115)	(0.126)	(0.101)	(0.107)
Experience	-0.002	-0.003	-0.002	-0.004
	(0.008)	(0.009)	(0.007)	(0.007)
Country of Degree Dummy	0.334	0.422	0.013	0.005
	(0.269)	(0.301)	(0.202)	(0.213)
Mobility in Region	-0.011	-0.060	0.019	0.087
	(0.166)	(0.177)	(0.152)	(0.159)
Mobility out Region	-0.030	-0.038	0.068	0.074
	(0.074)	(0.081)	(0.067)	(0.072)
Experience Herfindahl	-0.330**	-0.233	0.038	0.044
	(0.159)	(0.173)	(0.138)	(0.148)
Previous Patents Dummy	0.274	0.154	0.009	0.029
	(0.187)	(0.203)	(0.165)	(0.176)
log Employees	0.019	0.026	0.011	0.006
	(0.013)	(0.016)	(0.011)	(0.013)
R&D Intensity	1.764*	2.517**	1.560	2.294**
	(0.965)	(1.109)	(0.97)	(1.075)
Inventor Past Patents	0.076***	0.090***	0.062***	0.069***
	(0.019)	(0.021)	(0.016)	(0.017)
Application Year FE	YES	YES	YES	YES
NUTS2 Region FE	YES		YES	
NUTS3 Region FE		YES		YES
Regional Controls		YES		YES

N=6122,$^{*}p < 0.1,^{*} * p < 0.05,^{*} * *p < 0.01$

Table 5: Probit Estimation of Unmatched and Matched Data:Effect of PhD Degree on Commercialization

	Unmatched		*Matched*	
	Model 1	Model 2	Model 3	Model 4
Near	-0.034	-0.029	-0.033	-0.021
	(0.045)	(0.046)	(0.055)	(0.058)
Far	-0.005	-0.005	-0.008	-0.009
	(0.042)	(0.043)	(0.052)	(0.055)
Age	0.004***	0.004***	0.006***	0.005***
	(0.002)	(0.002)	(0.003)	(0.003)
Gender	0.015	0.021	0.001	0.037
	(0.116)	(0.118)	(0.234)	(0.249)
PhD Degree	0.002	0.005	0.001	0.002
	(0.045)	(0.047)	(0.054)	(0.057)
Far Past Coinventors	0.004	0.006	0.021	0.022
	(0.063)	(0.067)	(0.088)	(0.093)
Coinventors Dummy	-0.009	-0.015	-0.044	-0.042
	(0.065)	(0.068)	(0.092)	(0.096)
Experience	-0.002	-0.001	-0.002	-0.001
	(0.005)	(0.005)	(0.006)	(0.007)
Country of Degree Dummy	0.018	0.024	-0.003	-0.006
	(0.087)	(0.09)	(0.176)	(0.183)
Mobility in Region	-0.017	-0.016	-0.057	-0.066
	(0.072)	(0.074)	(0.142)	(0.146)
Mobility out Region	0.007	0.007	0.043	0.041
	(0.045)	(0.047)	(0.062)	(0.065)
Experience Herfindahl	-0.020	-0.032	0.011	0.008
	(0.087)	(0.09)	(0.127)	(0.133)
Previous Patents Dummy	-0.009	0.012	-0.012	-0.013
	(0.102)	(0.106)	(0.151)	(0.158)
log Employees	-0.009	-0.007	-0.012	-0.012
	(0.008)	(0.008)	(0.01)	(0.011)
R&D Intensity	-0.467	-0.531	-0.383	-0.396
	(0.59)	(0.626)	(0.721)	(0.787)
Inventor Past Patents	0.005	0.005	0.003	0.004
	(0.012)	(0.012)	(0.015)	(0.015)
Application Year FE	YES	YES	YES	YES
NUTS2 Region FE	YES		YES	
NUTS3 Region FE		YES		YES
Regional Controls		YES		YES

N=6122,$^{*}p < 0.1,^{*} * p < 0.05,^{*} * *p < 0.01$

5 Concluding Remarks

How to best organize the labor of science and research processes remains a crucial question among scientists, public research institutions, and firms. Practitioner decisions in the how to organize for the inventive process frequently take place absent full information, and often come with unexpected consequences.

We analyze one such consequence specific to firms and their employees through investigating the relationship between the maximal level of inventor education involved in an inventing firm's R&D processes and projects and the firm's resultant scientific orientation and patenting behavior. We find that involving a doctorate-level inventor in firm patenting projects leads to increases in reliance on scientific literature and conferences or workshops as sources of knowledge in the research and developing process for that patent project. The consequences of these shifts and the inclusion of a doctorate-level researcher in project R&D does not appear to lead to variation in the relevance of commercialization as a driving force for patenting. Overall, this indicates that inclusion of doctoral researchers in patenting projects creates to increasingly science-oriented firms, without a penalization of (or boon to) commercialization as a reason for firm patenting activity.

A fundamental complication of our analysis is that we only observe intention to commercialize, but not the observed commercialization rates or firm revenues and costs. As a result, we cannot fully derive whether the involvement in firm projects and research and development processes of doctoral-level inventors negatively or positively impacts firm profits, or leads firms to behave in ways that limit profit-maximization goals. If our estimates of the key coefficients are correctly estimated at their true values, this would indicate no effects of the treatment on commercialization and that firms can choose to include the addition of doctoral-trained inventors without fear of lost drive for commercialization. Future investigations may more directly explore this link by more finely estimating the covariates and the dependent variable, as well as examining other ways in which the doctoral-level employee determines the behaviors of the scientific firm.

If this work were to continue, as previously mentioned, we intend to introduce analysis of marginal effects. Additionally, we are considering a number of robustness checks and adjustments, including comparing the performance of CEM matching to propensity score matching and Mahalanobus distance matching, testing our analysis on certain subsamples of the data to see whether results hold (such as for small firms or large firms only), and performing matching on variables that are less determined by the presence of a doctorate-educated inventor which were currently removed from the matching process (such as log of employees).

References

Allen, Thomas J. *Managing the Flow of Technology.* Cambridge, MA: MIT Press, 1977.

Fleming, Lee, Charles King, and Adam Juda. "Small Worlds and Regional Innovation." *Organization Science* 18 (2007): 938-954.

Gambardella, Alfonso, Paola Giuri, and Myriam Mariani. The Value of European Patents: Evidence From a Survey of Inventors. European Commission, 2005.

Giuri, Paola, and Myriam Mariani. "When Distance Disappears: Inventors, Education, and the Locus of Knowledge Spillovers." *The Review of Economics and Statistics* 95, no. 2 (2013): 449-463.

Graham, J.W. "How many imputations are really needed? Some practical clarifications of multiple imputation theory." *Prevention Science* 8, no. 3 (2007): 206-213.

Honaker, James, Gary King, and MAtthew Blackwell. "Amelia II: A Program for Missing Data." *Journal of Statistical Software* 45, no. 7 (2011): 1-47.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson. "Geographic Knowledge Spillovers as Evidenced by Patent Citations." *Quarterly Journal of Economics* 108 (1993): 557-598.

Kosuke, Imai, Gary King, and Olivia Lau. "Toward A Common Framework for Statistical Analysis and Development." *Journal of Computational and Graphical Statistics* 17, no. 4 (2008): 892-913.

Kosuke, Imai, Gary King, and Olivia Lau. "Zelig: Everyone's Statistical Software." http:GKing.harvard.edu/zelig. 2007.

Rubin, Donald B. Multiple Imputation for Nonresponse in Surveys. New York, NY: John Wiley and Sons, 2008.

Stern, Scott. "Do scientists pay to be scientists?" *Management Science* 50, no. 6 (2004): 835-853.

Table 6: Variables used in multiple imputation

Variable Names	Despription
EDU	Higest degree of the inventor
Location	Country of the inventor
CloseExt	The importance of near collaboration
DistExt	The importance of near collaboration
BirthYear	Birth Year of the inventor
Age	Age of the inventor when applying patents
Age_squared	Age square
Gender	Gender of the inventor (1 if male)
Log of Experience	Log of Past Patents of the inventor
Herfinv1	Herfindahl index for the inverntor's patents across technological classes
Beadthexp01	Breadth Dummy for the inverntor's patents across technological classes
Year_first_patent	Year of first patent
link_diff_nuts_share1_b	Far past collaborators
link_coinv_diff_nuts_share1_b	Far past coinventors
link_coinv_diffsame1	Ratio of far coiventor to near coinventors
SK_ScLit	The importance of scientific knowledge
SK_Tech	The importance of scientific conference
residence_degree	whether working in the country where the inventor get highest degree
YearMobilitybefore	Mobility dummy
reg_mob_nuts3In	Mobility in the region
RDint	R&D intensity of the applicant
lEmployeesMiss	Log of applicant's employees
Ninventors	Number of inventors of the patent
reasonCommExploit	Importance of Commercializatin purpose of the Patent
reasonLic	Importance of lisencing purpse of the Patent
reasonImit	Importance of preventing imitation purpse of the Patent
lgdppop_nuts3	Log of GDP per capita of the region
lpop_nuts3	Log of population of the region
larea_nuts3_km2	Log of area of the region
lav_pat_nuts3_9496	Log of patens of the region
lsharepat_nuts2_tc30_1	Log of patens share of the region
top1_tc	Top 1%tile in Technology region dummy
top5_tc	Top 5%tile in Technology region dummy
shangai_univ	Research University Dummy
shangai_n_univ	Number of Research Universities in NUTS3 regiod (Shanghai Academy Ranking of World Universities)
overall_score	Research University Score
leiden_univ	Leiden-ranked European University in 2008
leiden_n_univ	Leiden Ranking in 2008
cpp_fcsm	Normalized Leiden Ranking Score

Table 7: Variables withheld due to colinearity issues

Variable Names	Despription
id_parent_1	patent grouped ID
no_pastcollab	Weight of the sample
nuts3_d_res	NUTS-3 region
reg_mob_nuts3Out	Mobility out the region
nuts2_d_res_country	NUTS-2
AppYear	Apply year
TechClass	Technical class
Applicant	Applicant Type