Firm-Inventor Links and the Composition of Innovation

Jeremy Pearce*

University of Chicago

December 21, 2019

Abstract

Innovative firms often operate across many technological areas. In order to move into these areas, firms require inventors with potential to add specific expertise to the firm. This paper explores factors that drive firms to innovate in different fields and the role of their inventors in determining the direction (which field in technology space) and quality (how much innovation impact) of their idea production. The paper explores these factors in three steps. First, I construct measures that illustrate a firm or individual's connection to a given technology area in USPTO patent data. Second, I use these measures to evaluate which features of the firm drive the direction of the innovative activity. A firm's history and its scientists' history are each crucial in determining firm's direction. However, the history of the specific scientists listed on the patent that provide the best predictor of the technology space on innovation and innovation quality as measured in citations. Third, I explore the role of the match between inventors and the firm in determining direction and quality of innovation. The *human capital match*, e.g. the inventors at the firm matching with each other, is more important than the *firm inventor match* in explaining both direction and quality of ideas produced. When innovating in new fields, firms have higher quality patents when their new inventors are more distant from their existing portfolio, suggesting the matching of distant spheres enable more radical innovation. Overall, this paper indicates the importance of understanding the composition of inventors at firms as a key issue for innovation.

1 Introduction

Most important advances in technology come from firms. Yet human capital and individual expertise is the primary building block of idea production and long-run economic growth. While most work in endogenous growth has addressed idea production through the lens of creative destruction, there has been less investigation into the factors that determine the details of a

*Contact: jgpearce@uchicago.edu.

firm's innovation choice. In particular, few papers have focused on the interaction of individual domain specific human capital and firm-specific knowledge in producing new ideas.

This paper explores the dual role of firms and inventors in innovation.[1] A key underlying message is that the match between firms and inventors determines the direction and quality of innovation. Further, due to data on a firm and inventors patent histories across technologies, there are empirical methods that can speak to the structure of this match. Constructing firm-inventor match histories enable a textured discussion of the composition of innovation.

The goal of this paper is to document the factors that shape firms' decisions on where to innovate (e.g. what technological area) and how these factors affect their quality of innovation (e.g. innovation impact as measured by citations). A key contribution of this paper is linking this decision to the inventor stock of the firm and the inventors on a firm's patent. I ask two main questions in this paper. First, I ask what is most predictive of the types of technology a firm innovates in next – (i) the firm's technological domain, (ii) the scientists at the firm's technological domain, or (iii) the specific inventors on the patent? Second, what role do the inventors play on the patent more specifically? For instance, how does the inventor-firm matching determine both the quality and direction of innovation?[2]

In order to answer these questions, I construct measures of firm history, worker history, and firm-worker match. First, I discuss technology classes, where the USPTO groups patents according to their function, and construct measures that illustrate a firm or scientist's connection to a given technology class, using patent citations as an indicator of connection between classes. I discuss firm expansion into fields they have not worked in and how close their new class is to their past activity, proxying for direction of innovation by USPTO patent class. I then construct measures of firm-scientist match, which is determined by their overlap in operating in classes that pull on the same knowledge. I use patent citations to proxy for patent quality. My goal is to use these measures to attempt to answer the questions in previous paragraph.

Through patents, this paper makes use of the idea of *realized regions* of production of ideas and *potential regions* where firms and scientists operate. Firms produce in certain areas at a given time t and plan to produce again in period $t+1$. This paper provides tools to inform where a firm might go in technology space given information on past production.

Diversification across technologies is important for a firm's growth and survival. Without diversification, firms expose themselves to greater chance of failure given a strong reliance on specific markets. However, expanding into new areas is challenging. Companies have a dual issue of building domain expertise outside their primary area and deciding the proper direction to build at the firm level. Quantifying the forces behind this decision has heretofore been difficult, as the literature has lacked compelling measures to speak to potential fields a firm may become

[1] I refer to inventors and scientists interchangeably in this paper.

[2] Direction is denoted by the patent class a firm innovates in, relative to the patent classes they have been active in. Quality is measured as patent impact through citations, but also employs a measure from Kogan et al. (2017), who use stock market surprise measures as indicators for patent monetary value.

active in.

I find that a firm's direction into technology classes is dependent on both their technological history and their scientists' technological history. The predictive power is strongest in response to the scientists participating on the patent. If a firm looks to expand into new fields, a scientist in a different knowledge space than the firm will be more likely to move the firm into new technologies. Scientists push firms to new frontiers most often early in their tenure. Further, successful matches that allow for diversification require some similarity of new scientists with the firm's existing scientists. This illustrates the vital role of scientists in firm growth and diversification.

Related Literature Ideas are the bedrock of long-run economic growth. Since Romer (1986), the concept of growth as recipes redirecting matter towards specific uses has been the method for thinking about technology production. The fact that new ideas knock off older ideas was classically shown in Schumpeter (1942), who discussed the importance of creative destruction coming from both large and small firms. Creative destruction was later formalized by Aghion and Howitt (1992), who stress the importance of certain ideas replacing others, thus the key role of ideas in firm life-cycles. Because of creative destruction, ideas that generate temporary monopoly power are crucial to the life-cycle of firms.

Empirically, the most effective way of studying ideas in firms is through patent data.[3] Outside of patent data, it is very challenging to understand the role of ideas in firms since they outside of patents ideas are highly ephemeral and subtle. Since patents provide a strong incentive to report the details of an idea, patents generate a great database as noted by Hall et al. (2001).

Klette and Kortum (2004) link the theoretical to data to show how creative destruction at the firm-level can lead to aggregate innovation. This ties firm dynamics closer aggregate growth. However, Klette and Kortum (2004) do not focus on the firm's tradeoff about where to innovate, on their own product lines or outside. All growth comes from undirected search across any new product.

Akcigit and Kerr (2018) bring in the idea of "internal" and "external" innovation, where a firm can build on their existing technology or venture into new technologies. This paper focuses on firm attempts to build outside of their existing product lines, external innovation. Most endogenous growth models tend to treat new innovation as undirected, in the sense that firms attempt to go outside their existing technologies but can land in any technology. This paper expands on the previous literature by attempting to speak to what drives the direction of new idea production. This has been less of a focus than the idea that firms direct their production outward once the returns are large enough.

In introducing a new framework for addressing the forces that shape firm expansion, this paper speaks to a large theoretical and empirical literature on the process of R&D in firms. Since Lucas (1988), how human capital interactions have been an important component of economic

[3]There is a burgeoning field of analysis of scientific publications as well, which is less closely linked to firm dynamics

growth theory. This has been linked to firms theoretically in papers (e.g. Bolton and Dewatripont, 1994). Nelson (1982) noted that R&D efficiency comes from firm knowledge stock, and has modeled this process as one of combining utility with design. Naturally, this question touches on classic questions related to the boundary of the firm (Coase, 1937) and its interaction with human capital.

Empirically, researchers have more often focused on the role of R&D in firms rather than their interaction with human capital. Bloom et al. (2013) explore the role of product markets and their interaction with a firm's technological structure. While important, this element misses the interaction with the individuals' knowledge stock at the firm. More generally, Dimos and Pugh (2016) review the literature of the effectiveness of R&D subsidies focusing specifically on the firm's reaction. The role of the match with the firm's inventors has received less attention.

In addressing the importance of firm-worker matching in innovation, this paper speaks to matching in a more general context, which has received significant attention in the literature at least since **?**. Among others, Jovanovic and Nyarko (1996) and Davis (1997) both investigate the role for learning and its importance in understanding why individuals switch firms. They both show how varying understanding of the structure of teamwork and management implies different matches depending on the sensitivity to worker ability. The role of learning in this model can be proxied for by experience across classes that will determine the skill set the individual brings to the firm.

I also stress the importance of human capital in innovation. The recent focus in understanding the role of human capital has shifted given the importance of teams in technology (e.g. Wuchty et al., 2007). The role of the individual in requiring a team in order to produce their idea (e.g. Jones, 2009) can be extended to firms, whose expertise may not cover what is necessary to bring an idea to fruition. In addition, it stresses that individuals generally only have a limited breadth of knowledge which induces them to require teams and firms to produce an idea.

2 Data

I rely on being able to specify the technology space firms are active in (USPTO), where an individual has been technologically (Li et al., 2014 and USPTO), and matches between individuals and firms. In addition, it is necessary to have a dataset that can speak to the value of patents. There are two different ways of doing this. The first is looking at impact via patent citations from the USPTO. The second is by using changes in value from patent grant date which comes from Kogan et al. (2017).

Firms or individuals register for patents in order to guarantee a property right to exclusive use of an idea for 20 years. Each patent has a corresponding assignee and set of individuals (i.e. on average a patent has 2.2 co-authors).[4] The patent assignee is generally a firm. The individuals

[4] I drop "garage" patents–patents without firms–in the analysis

generally work for the firm, and I refer to them as inventors or scientists. Each observation corresponds to a $firm \times scientist \times patent$ over the course of 1976-2002, which yields about 6 million observations.

Patents cite each other indicating a flow of ideas. Patent A will cite other patents B and C as a way to signify that A builds on B and C. This is often done by the patent authors or the patent examiner. Citation data provides dual use on both level of impact and specific technological flows. Patent classes can be linked together through frequency of knowledge flows. Individual patents have assignee information as well as first and last names of team members. A dataset from Li et al. (2014) enables me to link the entire career of a scientist to their history. Li et al. (2014) use a Bayesian algorithm which uses individual names, patent classes, location of inventors, their firms, and their corresponding co-authors to break down the names associated with patents to allow for a tracking of the full history of individuals on patents. USPTO patent assignees enable a measure of a firm's history in patent production. This generates a dataset that can speak to the match between scientists and firms.

The data contains patents granted from 1975-2010. I collect citations on all patents, but truncate the end of the analysis since the arrival of the patent is the application year and citations require more patents to follow on in the future. As such, my regressions and graphs cover 1976-2002.

In addition, Kogan et al. (2017) have produced a dataset that assigns patent value to each patent based on the change in a company's stock valuation when the patent is granted. They do an event study the day of the patent grant to determine value. They base this on taking the conditional expectation of patent value prior to grant date and then looking at the change in market capitalization as a result. This measure can get us at the idea of patent value as measured in monetary terms, but it reduces the sample since these are only for publicly traded companies.

I use USPTO patent classes and subcategories as a measure of histories. There are a total of about 430 classes that a patent can belong to, with some classes much more heavily populated than other classes. There are 36 2-digit technology class subcategories set by the USPTO that contain classes at more coarse level. The classes will be used to specify details at the match level, whereas subcategories will be used to discuss individual firms' choices of where to operate. Classes give a more comprehensive indication of the areas that individuals and firms work in. However, in the case of discrete choice models, classes add a lot of unnecessary complexity to computation, while subcategories yield similar overall results.

The class system and citation network allow us to identify technological areas where firms and scientists operate. This paper works at the subcategory level (with 36 subcategories). This is because as the class categories get finer firms are absent in most all classes, delivering less interesting information in discrete choice modeling. Examples of subcategories are Metal Working, Transportation, Optics, Heating, Pipes and Joints, Organic Compounds, Gas, Drugs, Biotech, Power systems, Nuclear and X-Rays.

3 Construction of Measures

There are two types of agents in this economy: scientists and firms. For scientists, I use their history of patent production. For a firm, there are two ways to measure their location: using their scientists' history or using the firm's history itself. Based on these measures, I then construct measures that indicate the "match" between scientists and firms. I primarily explore these three measures throughout: scientist level, firm level as collection of scientists, and firm level as embedded knowledge in the firm.

A scientist or firm has a history across different patent classes. For each scientist on their nth patent, there $n-1$ previous patent productions. Each of the $n-1$ observations are in a patent class. This proxies for the history of production across technologies the inventor has when meeting with the firm. Similar analytics can be done for the firm.

3.1 Realized and Potential Regions

In order to answer some of the questions posed in the introduction, I build measures to enable a framework of predicting the direction of a firm's innovation. In particular, I measure both *realized areas of idea production* along with *potential areas of idea production*. In this section I describe the different measures that are necessary for the analysis.

3.1.1 Realized Areas of Production

I denote the first measure as *Research Concentration*, which indicates the presence of a scientist or firm in a particular patent class. This maps to the *realized regions* of idea production discussed earlier.

$$RC_{j,c} = \frac{\text{\# Patents of } j \text{ in class } c}{\text{Total Patents of } j} \text{ where } j \in \{i, f\} \tag{1}$$

Note that for a fixed firm or scientist a, $\sum_{c \in \mathcal{C}} RC_{a,c} = 1$. This initial measure helps us understand the general areas firms and scientists are working in. Some firms or scientists will be focused in one or two classes, with high research concentration. Others will be dispersed across many classes. When a firm produces in an area with a low values of RC, I consider that firm to be engaged in "external" innovation, advancing into a new field.

3.1.2 Connectivity of Classes

In addition to the Research Concentration measure, I create a measure that quantifies the connectivity of classes through their citations to each other. This enables discussion on the *potential areas of idea production* an individual or a firm might engage in given where they start (for instance, Computer Hardware and Software will have closer links to Information Storage than Agriculture, Food, and Textiles). Then I can speak to where an individual or firm is more likely to produce

given only a couple of observations. These measures serve as the link between the *realized* and *potential* areas of production discussed earlier.

These measures express the connection between class A and class B:

$$c_{uk}(A,B) \equiv \frac{\#\text{Backward Cites on Patents in } A \text{ to } B}{\#\text{Total Backward Cites on Patents in } A}$$

$$c_{dk}(A,B) \equiv \frac{\#\text{Forward Cites on Patents in } A \text{ to } B}{\#\text{Total Forward Cites on Patents in } A}$$

$$c_k(A,B) = c_k(B,A) \equiv \frac{c_{uk}(A,B) + c_{dk}(A,B)}{2} \tag{2}$$

$$c_p(A,B) = c_p(B,A) \equiv \frac{\#A_c \cap B_c}{\#A_c \cup B_c}$$

Where *uk* stands for "upstream knowledge" as in this is A citing B. *dk* stands for "downstream knowledge" as in this is B citing A. The measure c_k in (2) serves as the bedrock for connecting classes in this analysis. c_k indicates an average of how one class is connected to another based on how the classes use each other in both upstream and downstream connections. c_k will be high in cases where classes are highly codependent, often building on each other. However, *k*, *uk*, *dk* do not give very different results.

p stands for proximity, and $\#A_c \cap B_c$ are the number of patents that cite both from class A and B. $\#A_c \cup B_c$ are the number of patents that cite from either class A or B. This measure is taken from Akcigit et al. (2016).[5] All measures are between 0 and 1.

Since I am interested in speaking to firms and scientists' potential connections across classes, these measures allow for reasonable insight into where a firm or scientist could go if they have produced in certain classes, as given by the complementarity and knowledge overlap.

3.1.3 Potential Areas of Production

We can now speak to *potential areas of idea production* from firms and scientists using the classes that the areas of produced ideas are connected to. Here, I can take the individual's experience in each class and multiply it by the connectivity to other classes. This will give us a distribution of potential knowledge that can be expressed at both the firm and individual level. The three main measures:

Knowledge connection of firm f to class B[6] :$K_{f,B}$

$$K_{f,B} = \sum_{c=1,\ldots,C} RC_{f,c} \times c_k(c,B) \tag{3}$$

[5]This measure is used for robustness checks in all the results

[6]Note that I assume there are C classes

Scientist i connection to class B: $S_{i,B}$

$$S_{i,B} = \sum_{c=1,c=1,...,C} RC_{i,c} \times c_k(c, B) \tag{4}$$

Human capital connection of firm f to class B: $H_{f,B}$

$$H_{f,B} = \sum_{c=1,...,C} \overline{RC}_{(i \in f),c} \times c_k(c, B) \tag{5}$$

Where the average in (5) is a weighted average depending on the number of a scientist's individual patents at the firm in the given year, and each i is a collection of individuals at the firm. Here I am using an agent's past production and the connection of that past production to other classes to speak to their potential area of production by using these indicators to proxy for the knowledge they pick up somewhere else.

I count a scientist as being at a firm if they have produced at the firm in the corresponding year or in between patents with the firm.[7] This measure is imperfect, since I cannot account for the scientist's role at the firm prior to their first patent or after their last patent. Further work will calculate expected length prior to the first patent to better understand the actual scientist composition.

These measures provide a distribution for each firm of the potential classes they are involved in, via a knowledge connection. K_f and H_f are strongly correlated–overall, the correlation is 0.88.

3.2 Firm Scientist Match

How similar are a firm and scientist in their experience? I evaluate this depending on the definition of the firm–a collection of ideas (i.e. K_f) or a collection of scientists (i.e. H_f). As discussed previously, there are two ways to measure the match. One can look at **closeness of firm ideas and scientist**. Another way is to use the **closeness of firm human capital and scientist**. Further, I can speak to this similarity in *realized regions* or *potential regions*.[8] Since I am interested in understanding the firm's next move, I will use *potential regions* of production to understand the match between scientist and firm.

The construction of the measure of match between inventor and firm follows Jaffe (1986) who uses this measure to illustrate the degree of match between firms depending on their past production.

Define vector $K_f \equiv \{K_{f,1}, K_{f,2}, ..., K_{f,C}\}$, which is a collection of the firm's idea connection to classes $1, ..., C$, vector $S_i \equiv \{S_{i,1}, S_{i,2}, ..., S_{i,C}\}$ which is a collection of the scientist's connection to all classes and vector $H_f \equiv \{H_{f,1}, H_{f,2}, ..., H_{f,C}\}$ which is a collection of the firm's connection, via

[7] This requires a scientist not be producing elsewhere

[8] The results when thinking about the match in terms of realized regions are not significantly different from the results presented in this paper.

scientists, to all classes.

Now I can define the match, based on two different measures of firm technology location[9]:

$$FMATCH_{fi} = \frac{K_f S_i'}{(K_f K_f')^{1/2}(S_i S_i')^{1/2}}$$

$$HMATCH_{fi} = \frac{H_f S_i'}{(H_f H_f')^{1/2}(S_i S_i')^{1/2}}$$

$FMATCH$ is the match of firm f to individual i based on the firm's portfolio history and the individual's portfolio history, as discussed above. $HMATCH$ is the "human capital match" which is the set of scientists at the firm and their collection of patents, matched with an individual. These measures are in $[0, 1]$ which indicates the strength of overlap in similar regions of ideas. Note that FMATCH maps to the firm as a collection of ideas while HMATCH maps to the firm as a collection of scientists.

From their experience in relevant classes, I can speak to the type of expertise firms and individuals are connected to through their past patents. Then I can look at firm and scientist overlap when they join for production. When it comes to a firm-scientist match, matches can be "close" in the sense that firms and scientists operate in similar knowledge networks or "far" in the sense that they tap into different knowledge networks.

3.3 Dependent Variables

As mentioned in the opening, I am mainly interested in the *direction* and *quality* of ideas. I start with a simple procedure for encoding whether a patent is an advancement into a new field by a firm or not. The variable "new" measures whether or not the firm is entering a new field (0 for working on fields in their existing portfolio, 1 for a new field). I define this as the bottom 5th percentile in realized research concentration[10] in technology class of the patent, conditional on firm age. Controlling for firm size allows for a consistent measure of what it means for a firm to expand externally. In this case I only think of new exclusively in relation to the firm.

To understand quality, I simply use patent citations in the next five years after a patent was produced, excluding self-citations from within the firm. More cited patents reflect a higher quality patent (as noted by Kogan et al., 2017), because it serves as a focal point for further research and is strongly correlated with monetary values (see Figure 4 in appendix).

[9] These measures have been used for patent data and technologies in Bloom et al. (2013)
[10] Recall $RC_{f,c}$ defined in (1)

4 Results

I keep in mind the two key questions asked at the beginning as I discuss the results. Which technology direction do firms innovate in? And how does their match with the scientist feature in these innovations? The following bullet-points summarize the main results.

1. Firm expansion into new fields is associated with higher patent value, conditional on knowledge related to the field.
2. A firm's expansion into new fields is strongly related to their past collection of ideas, and the ideas of their scientists. The scientists on the focal patent are the strongest predictor of patent technology class.
3. Firms employ scientists who are technologically distant from the firm's portfolio to expand into new technologies.
4. Higher match between firm and scientist on patent match yields *lower* citation impact
5. Higher human capital match between scientists at the firm and scientists on the patent yields *higher* citation impact and more patents produced by a firm-scientist pair.

Overall, both firm-specific knowledge and human capital at the firm are important in determining direction.[11]. The scientist on the patent is the best predictor of the firm's activity in terms of the quality of its output and its direction. The resulting impact and direction of patents hinge on the firm, the scientist and the existing human capital at the firm.

The results will be presented mostly in regressions of the following form:

$$outcome = W_i'\beta + controls$$

Where the *outcome* variable is generally either some indicator of firm *direction*, i.e. where a firm produces, or *value* as discussed in the initial questions in this paper. In order to address where firms innovate, I stress the *realized area of production* in tables 3-6. When addressing quality, I focus on monetary value (Table 2) or citations (Table 7). W_i' are firm-specific knowledge connections (i.e. H and K) for regressions at the firm level. W_i' are match-specific details (i.e. $HMATCH$ and $FMATCH$). In the following subsection I will discuss the results at the firm-level, i.e. how a firm benefits from diversification and what features of the firm determine the direction of production. In the next subsection, I will focus on the specifics of the match with the scientist, which is crucial to understanding direction and quality of production.

[11] We do not have demand shocks or aggregate shocks in this framework

4.1 Firm Level

At the firm level my first question is what drives the direction of idea production? I start with a discussion of why firms would produce outside of their existing area of expertise and then expand to thinking about where they would want to expand and what shapes the choices of their idea production. I find that a) there are benefits to diversification conditional on knowledge of the relevant field, and b) the decision to operate in a field relies on both scientists at the firm and the firm's past history, but most of all on the scientists specifically on the patent.

4.1.1 Expansion and Value

First, I show in which cases entering new fields can create value for a firm. This can be seen in terms of monetary patent value. I use a dataset from Kogan et al. (2017) dataset which contains the patent dollar values. Whether a firm is "new" to a field, discussed earlier, has implications on the value of the patent, once one controls for the firm's relevant knowledge. This variable has an impact on patent value, as measured by our first regression. The specification is in equation (6), where c(p) signifies the class c that patent p belongs to and X includes year fixed effects, firm size, as well as total people on patent.

$$LogValue_p = \beta_0 + \beta_1 New_{f,c(p)} + \beta_2 K_{f,c(p)} + \beta_3 H_{f,c(p)} + \Lambda' X + u \tag{6}$$

Table 1: Patent Value on Firm and Scientist Knowledge and New Field

	(1)	(2)
	Log Value	Log Value
New	0.00399	0.261**
	(0.08)	(3.20)
$K_{f,c,t}$		0.0371
		(0.53)
$H_{f,c,t}$		0.245**
		(3.14)
Observations	767494	763381
R^2	0.102	0.112

t statistics in parentheses, clustered at the firm level, year fixed effects

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1 provides the intuition of the consequences of new field expansion for a firm. Conditional on a firm's distance to an idea (K – firm specific and H – scientist specific), new technology classes yield higher value patents to the firm. Table 2 indicates that patents produced in new fields are of higher value conditional on the firm's knowledge. Patents in new fields are on average are worth approximately 26% more than patents in "internal" fields, once one conditions on the relevant firm knowledge. Increasing H, the scientist knowledge relevant to the class at the firm, from 0 to 1 also increases value of the patent by 25%. Note that this regression does not control for firm fixed effects, only year and firm size, so it speaks to the cross-section of firms and their pursuit of new fields.

4.1.2 Time Series

Firms that expand into new markets benefit, conditional on their scientist's knowledge. Table 3 looks at their concentration in markets depending on their knowledge connection and human capital connection. With the following regression, I connect their realized region of production (RC) with their potential regions (K and H).

$$RC_{f,c,t+1} = \beta_0 + \beta_1 K_{f,c,t} + \beta_2 H_{f,c,t} + RC_{f,c,t} + \Lambda' X + u \tag{7}$$

Equation (7) covers a yearly time period speaking to how the stocks of knowledge at the firm in year t affect production in year $t+1$. This regression is testing the strength of the response of research concentration in a class to two stock variables (the stock of human capital direction and stock of firm knowledge direction) and one flow variable (research concentration in the previous period). X controls for firm size (100 percentiles), class and year fixed effects, firm fixed effects. Table 2 shows the results when I keep all observations.

All variables are standardized the .596 coefficient can be interpreted as: for a one standard deviation increase in firm specific knowledge as related to class c, there will be an expected .596 standard deviation increase in production in class c in the next period. The effects are strong for both past patent knowledge and human capital in predicting future realized activity. Note that both firm and scientist knowledge matter as well as research concentration in the previous periods.

In Table 3, I keep observations of RC that had 0 concentration in the previous 3 periods. This gets at the idea of being able to predict new areas using the firm's knowledge capital. Note the effect of firm specific knowledge becomes less pronounced, possibly in part because this represents unchartered territory in terms of production activity. There is still a strong response to both firm knowledge and science knowledge in terms of class production. The interpretation of the coefficients is similar to Table 2.

Table 2: Realized Regions of Production and Potential Regions, all obs

	(1)	(1)	(2)	(3)
	$RC_{f,c,t+1}$	$RC_{f,c,t+1}$	$RC_{f,c,t+1}$	$RC_{f,c,t+1}$
$K_{f,c,t}(SD)$		0.596^{***}	0.596^{***}	0.447^{***}
		(92.48)	(91.76)	(48.50)
$H_{f,c,t}(SD)$		0.374^{***}	0.219^{***}	0.440^{***}
		(53.04)	(21.89)	(56.90)
$RC_{f,c,t}(SD)$	0.704^{***}		0.178^{***}	
	(81.07)		(23.69)	
$RC_{f,c,t-1}(SD)$				0.239^{***}
				(39.89)
Observations	12194645	12194645	12194645	8124978
R^2	0.335	0.493	0.498	0.577

t statistics in parentheses, clustered by firm-class

All variables are standardized

X are firm size, class, year, fixed effects

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Realized Regions of Production and Potential Regions, $RC_t \rightarrow RC_{t-3} = 0$

	(1)
	$RC_{f,c,t+1}$
$K_{f,c,t}, (SD)$	0.174^{***}
	(87.92)
$H_{f,c,t}, (SD)$	0.103^{***}
	(45.55)
Observations	11009009
R^2	0.067

t statistics in parentheses, clustered by firm-class

All variables are standardized

X are firm size, class, year, fixed effects

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4.1.3 Logit

Previous regressions show the value of advancement into new fields and the importance of a firm's portfolio and scientist experience in determining the direction. Here, I use a logit model to examine which firm-level characteristics determine the direction of research, examined again at the subcategory level. Using the McFadden (1974) method, I hold the key variables determining class choice constant, but let controls such as firm size and year vary by class.

Denoting $P^f_{c,p} \equiv Pr(\Pi^f_{c,p} > \Pi^f_{c',p} \forall c')$ the probability firm f invents in class c at their next patent p, where Π signifies the payoff. If the error term has a type I extreme value distribution, this model can be estimated as a multinomial logit. I run a multinomial logit (reporting the log odds ratios below) to help explain the variation that determines the firm's direction. Table 4 provides the results, comparing three forces that can shape the firm on the supply side.

$$P^f_{c,p} = \frac{exp(x'_{ij}\beta)}{\sum_{h=0}^{J} exp(x'_{ih}\beta)}$$

$$x_{it} = \begin{bmatrix} \text{Firm Specific Knowledge Connection} \\ \text{Human Capital Knowledge Connection} \\ \text{Scientists on the Patent} \end{bmatrix}$$

Table 4: Multinomial Logit on Firm and Scientist Knowledge

	(1)	(2)
	Realized Category	Realized Category
$K_{f,c}$	0.526***	0.431***
	(431.15)	(210.80)
$H_{f,c}$	0.223***	-.009***
	(178.75)	(-5.72)
$S_{i,c}$		0.664***
		(1119.73)
Observations	49484799	49484799

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The coefficients in Table 4 can be interpreted as a one-SD increase in X in class C increases the log-odds of innovation in class C relative to class 1 by β. If the baseline odds of innovating in class C are 1%, then a 1-SD increase in firm-potential-knowledge will increase the odds of producing next in class C to 1.53%. These odds are significant and indicate the relevance of the firm's past ideas in terms of determining where they can build. Note that this model has a strong assumption on the error terms taking an extreme-value.

In terms of qualitative results, there are two key takeaways from Table 4. First, both the firm's background and their scientists are important for establishing direction. Table 4 suggests that overall firm knowledge is more important than scientists-at-the-firm knowledge. Second, individual scientists on the patent are most predictive of direction. Third, when controlling for the scientists directly on the patent, the collection of scientists at the firm becomes less relevant. Scientists on the patent play a major role in terms of firm direction. This will be seen from another angle in the next section. The magnitudes are less important than understanding the relative weight of individual effects versus general effects and the important role for scientists.

Having established to some degree what determines the firm-level direction, I turn attention to the specific match that the firm uses to expand. Since the scientist's place in technological space seems to be the most crucial element in determining where the firm produces, the match carries significance.

4.2 Match Level

In the previous section, I established the link between the firm's opportunities for expansion and their direction. I noted that scientists play a major role, and in particular scientists on the existing patent. I now delve further into the firm-scientist match to think about how the individual contributes to the firm's goals. This data uses each $firm \times scientist \times patent$ level, which contains around 6 million observations from 1976-2002.

Two measures of match rely on two definitions of the firm–$FMATCH$, the match between ideas produced by the firm and the particular scientist in question, and $HMATCH$, the match between ideas produced by the scientists in the firm and the particular scientist in question, to quantify the relationships between firms and scientists. These measures will inform our analysis on both direction and quality of ideas.

A further match between scientist and firm (i.e. less similar past in terms of realm of ideas) leads to more external expansion. It also leads to benefits in patent impact, but this is conditional on the link between the scientist on the patent and other scientists at the firm. There are benefits to having this scientist close to existing scientists at the firm. The regressions in this section have the same form as discussed in the beginning of section 4.

4.2.1 Direction of Innovation

In this section, I focus on how the match drives advancement into new fields. A further and newer match is more likely to generate a patent in a new field related to the firm. When firms aim to expand, their existing stock of scientists are less likely to know about technologies outside the firm. This will lead to new matches. First I evaluate the role of the match in expansion. The following regression, with results in Table 5, is done at the patent level, where the firm and inventor come together to match on the specific patent. The LHS variable, New, measures whether or not the firm is entering a new field (i.e. equals 0 or 1) and is discussed above. This work can be extended to an intensive margin measure but for now it is binary.

$$New_{c(p),f} = \beta_0 + \beta_1 FMATCH_{p(f,i)} + \beta_2 HMATCH_{p(f,i)} + \Lambda' X + u \tag{8}$$

Where X is a set of controls for firm, firm size (100 percentiles of firm size), individual experience, class, year.

As seen in Table 5, when firms enter a new field, they are more likely to do it with someone who has less overlap with the firm. The coefficient measures the increased probability of entering a new field for a firm given a one-standard deviation increase in breadth. Because of firm fixed effects, the variation is occurring in direction and quality is from within the firm. The interpretation of column (1) is that for a one standard deviation decrease in the match there is around a 2.3 percentage point increase in probability of innovation, which, given that the probability of a new field entry is around 5 percent, is quite sizable–an increase from the mean

Table 5: Firm Entering New Field and Match Characteristics

	(1)	(2)	(3)	(4)
	New	New	New	New, 1st patent
FMATCH (SD)	-0.023***		-0.022***	-0.031***
	(-30.99)		(-22.72)	(-20.78)
HMATCH (SD)		-0.018***	-0.001	0.002
		(-27.07)	(-1.01)	(1.33)
Observations	3219083	3140884	3140884	277342
R^2	0.086	0.086	0.087	0.135

t statistics in parentheses, clustered by firm, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

FMATCH: Match between scientist and past ideas from the firm

HMATCH: Match between scientist and human capital at firm

X: Firm, Class, Firm Size, People on Patent, Year

"New"= $\leq$5th percentile in Research Concentration, controlling for firm size

of above 40%. A *low* FMATCH is the key element driving expansion, in particular in column (3), where both HMATCH and FMATCH are included, and (4) where they are both included and only an individual's first patent with the firm is kept.

Figure 1 illustrates two forces. First, the probability of expanding into a new class is decreasing in the patent number of the scientist-firm pair. Second, scientists further from the firm technologically are more likely to bring the firm into new technologies. Newer scientists and further scientists are both more conducive to innovation further from the firm. This complements the work of March (1991), who discussed the key roles that new scientists with different knowledge sets play in allowing a firm to engage in more exploratory innovation.

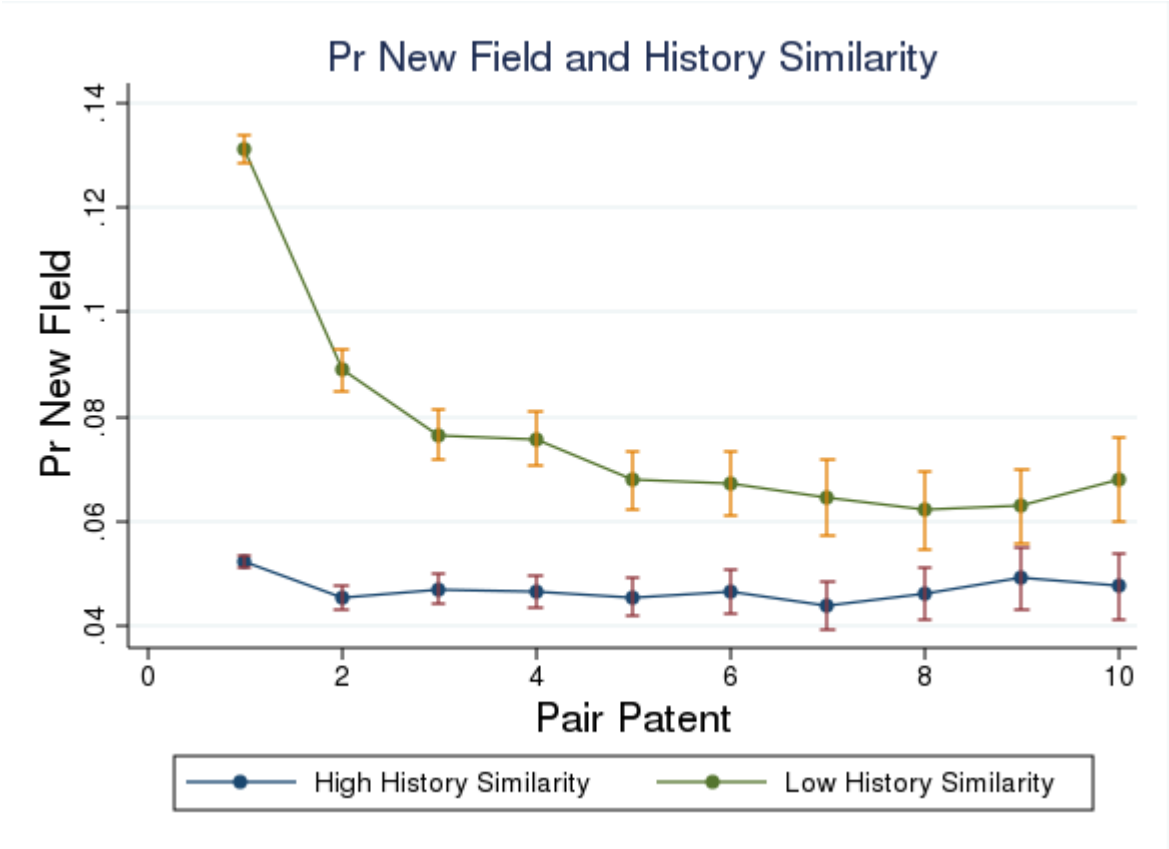


Figure 1: Probability of entering new field decays with the length of the match

4.2.2 Quality of Match

After verifying the connection between expansion and the importance of scientists to broaden the technological expertise of the firm, I look to the quality of the match. Quality can be measured in many ways; I treat this as citations (impact) and number of patents the match produces.

We have seen that the individual scientist is crucial for mapping out the direction of the firm (both from the logit model and at the match level). However, how does the firm ensure that this match will yield *quality*? It is possible that a match will help a firm expand into new fields, but if it produces weak ideas, the expansion will yield little in opening up areas for advancement.

Table 6 shows that patents have higher impact if the FMATCH has less overlap, where the individual is adding new knowledge relative to the firm. This regression uses the match level characteristics, with the scientist's first ten patents at the firm, to speak to the effectiveness of the match on patent quality (e.g. citations). I do not only include the scientist's first patent with the firm because this reduces the number of observations. The following regression encompasses the question of match, looking very similar to equation (8).

$$LogCit_p = \beta_0 + \beta_1 FMATCH_{p(f,i)} + \beta_2 HMATCH_{p(f,i)} + \Lambda' X + u \tag{9}$$

I find that a scientist working on areas distant from the firm's original portfolio will be more effective in idea production at the firm. However, this is conditional on a scientist working predominantly in areas with other scientists at the firm, where good overlap will deliver patent value. A 1 SD decrease in FMATCH increases citations by around 5%, conditional on keeping the scientist-human capital match constant. If that match can increase while FMATCH decreases, a firm can generate big gains in citations.

Table 6: Patent Impact on Match Chacteristics

	(1)	(2)	(3)	(4)
	LogCit	LogCit	LogCit	LogCit
FMATCH (SD)	-0.048***	-0.053***	-0.051***	-0.041***
	(-3.42)	(-4.59)	(-4.65)	(-4.07)
HMATCH (SD)	0.063***	0.047***	0.044***	0.032***
	(5.27)	(4.16)	(4.26)	(3.46)
Observations	1446727	1446727	1446727	1446727
R^2	0.130	0.182	0.185	0.207
Class Fixed Effects	X	X	X	X
Firm Fixed Effects		X	X	X
Firm Size Fixed Effects			X	X
Year Fixed Effects				X

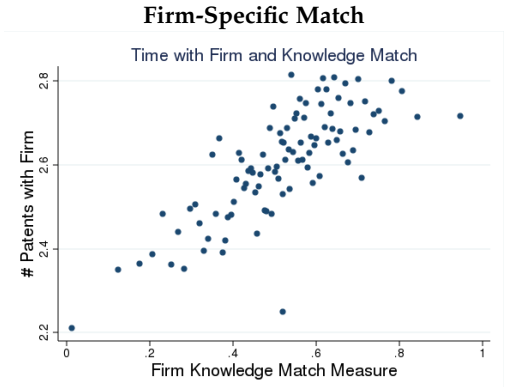
t statistics in parentheses, errors clustered by firm

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

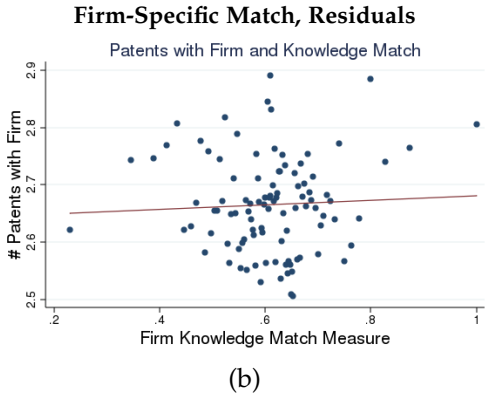
When considering the number of patents with the firm as relevant for match value, a closer match yields more value. Panel (a) in Figure 2 illustrates that the closer the firm and scientist are, the more lasting the match. However, once I control for the match between scientists and the human capital at the firm, this relationship goes away as seen in panel (b). The closeness between scientists at the firm and other scientists at the firm matters, as seen in panel (c).[12] This speaks to similar facets of the regression: the scientist-human capital match is more important for generating quality. A firm's character can more easily change to adapt the existing scientists, but scientists at the firm need a good degree of overlapping communication to produce strong ideas and lasting relationships.

[12]These are residual scatter plots, so in panel (a) I control for firm fixed effects, deciles for firm size, year, class. in panel (b) I control for the same variables except add in the firm-scientist closeness. In (c),I control for firm ideas-scientist match and all other variables.

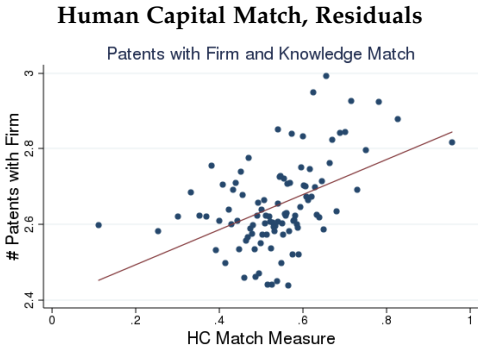
Figure 2: The Role of the Match in # Patents with Firm



(a)



(b)



(c)

Overall, I stress a couple features of the relationship between firm and scientist. The match between firm and scientist is important, but it is more important that there is a positive match between the scientist and the existing scientists at the firm. This yields value at both patent level and in number of patents with the firm. If an individual has a good match with the firm's scientists, they can also yield more value if they operate in different areas from the firm's central product lines.

Going back to the firm-level analysis, these results speak to the importance of the match both in direction and quality of research. The firm, through their stock of scientists and knowledge, seek a direction that requires the scientist on the patent to have some overlap with the firm's knowledge base, but enough difference to expand.

5 Conclusion

This paper illustrated the importance of understanding a firm's paths to expansion through a firm's interaction with their collection of scientists of stock of ideas. I showed that the path of firms into technology classes depends significantly on the firm's connection to patent classes as dictated by its past patent production and its current collection of scientists at the firm. Firms more often produce in new fields with new scientists and these scientists are different than the firm in terms of the knowledge they have connections to – but less different from existing scientists at the firm.

Firm expansion is inextricably linked to both their past and the people they find in order to expand. One can imagine many reasons for the links between firm expansion and their existing technological stock. Firms exist in certain product lines and past production has given them some insight that later a firm realizes can be applied in a certain way. Once firms find new idea potential related to their old ideas, it's crucial that the people operationalizing those new ideas understand the field. In this scenario, it will often be the case that there are gains to firms in finding new scientists who don't overlap so directly with the firm's past ideas. The returns to these scientists in having a close past with the existing scientists at the firm indicates the importance of teamwork. Understanding the firm as a collection of teams to explore these the links between firms and technologies are fruitful avenues for further research.

References

Aghion, P. and P. Howitt: 1992, 'A Model of Growth Through Creative Destruction'. *Econometrica* **60**(3), 323–351.

Akcigit, U., M. A. Celik, and J. Greenwood: 2016, 'Buy, Keep, or Sell: Economic Growth and the Market for Ideas'. *Econometrica* **84**(3), 943–984.

Akcigit, U. and W. R. Kerr: 2018, 'Growth through Heterogeneous Innovations'. *Journal of Political Economy* **126**(4), 1374–1443.

Bloom, N., M. Schankerman, and J. Van Reenen: 2013, 'Identifying Technology Spillovers and Product Market Rivalry'. *Econometrica* **81**(4), 1347–1393.

Bolton, P. and M. Dewatripont: 1994, 'The Firm as a Communication Network'. *Quarterly Journal of Economics* **109**(4), 809–839.

Coase, R.: 1937, 'The Nature of the Firm'. *Economics* **4**(16), 386–405.

Davis, S. J.: 1997, 'Sorting, Learning, and Mobility When Jobs Have Scarcity Value: A Comment'. *Carnegie-Rochester Conference Series on Public Policy* **46**, 327–338.

Dimos, C. and G. Pugh: 2016, 'The effectiveness of R&D subsidies: A meta-regression analysis of the evaluation literature'. *Research Policy* **45**(4), 797–815.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg: 2001, 'The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools'. *National Bureau of Economic Research Working Paper no:8498*.

Jaffe, A. B.: 1986, 'Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value'. *American Economic Review* **76**(5), 984–1001.

Jones, B. F.: 2009, 'The Burden of Knowledge and the "Death of the Renaissance Man" Is Innovation Getting Harder?'. *The Review of Economic Studies* **76**(1), 283–317.

Jovanovic, B. and Y. Nyarko: 1996, 'Learning by Doing and the Choice of Technology'. *Econometrica* **64**(6), 1299–1310.

Klette, T. and S. Kortum: 2004, 'Innovating Firms and Aggregate Innovation'. *Journal of Political Economy* **112**(5), 986–1018.

Kogan, L., D. Papanikolaou, A. Seru, and N. Stoffman: 2017, 'Technological Innovation, Resource Allocation, and Growth'. *The Quarterly Journal of Economics* **132**(2), 665–712.

Li, G.-C., R. Lai, A. D'Amour, D. M. Doolin, Y. Sun, V. I. Torvik, A. Z. Yu, and L. Fleming: 2014, 'Disambiguation and co-authorship networks of the U.S. patent inventor database (1975-2010)'. *Research Policy* **43**(6), 941 – 955.

Lucas, R. E.: 1988, 'On the Mechanics of Economic Development'. *Journal of Monetary Economics* **22**(1), 3 – 42.

March, J. G.: 1991, 'Exploration and Exploitation in Organizational Learning'. *Organization Science* **2**(1), 71–87.

McFadden, D.: 1974, 'Conditional Logit Analysis of Qualitative Choice Behavior,'. Frontiers in Economics, Chapter 4, ed. d. by P. Zarembka, New York: Academic Press.

Nelson, R. R.: 1982, 'The Role of Knowledge in R&D Efficiency'. *Quarterly Journal of Economics* **97**(3), 453–70.

Romer, P.: 1986, 'Increasing Returns and Long-run Growth'. *Journal of Political Economy* **94**(5), 1002–37.

Schumpeter, J.: 1942, *Capitalism, Socialism, and Democracy*. Harper Press.

Wuchty, S., B. F. Jones, and B. Uzzi: 2007, 'The Increasing Dominance of Teams in Production of Knowledge'. *Science* **316**(5827), 1036–1039.

A1. Appendix: More Figures and Tables

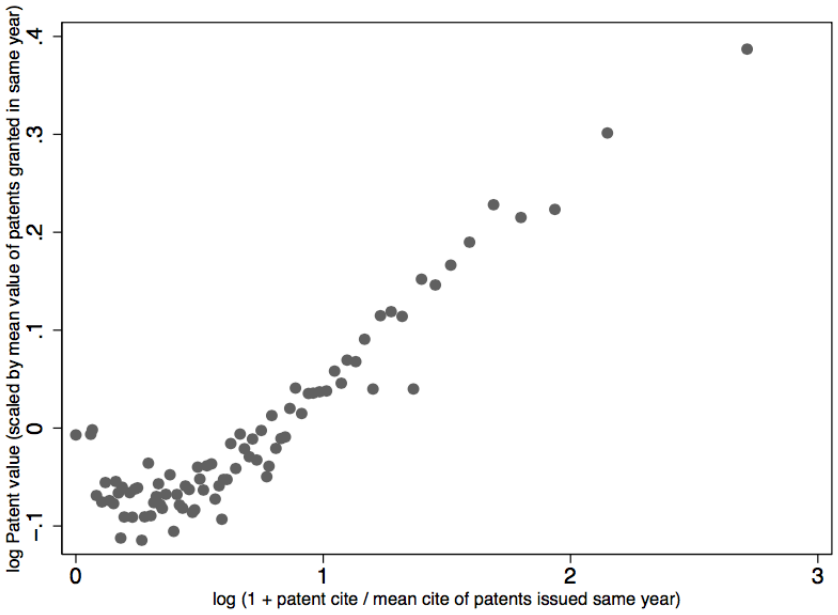


Figure 3: Patent Citations and Patent Value, Kogan et al 2016

In the data, firms take different paths to expansion. Some stay in relatively few technology markets and some expand into many markets. In the model I have not discussed much on firm heterogeneity, but it is clearly an important element of the discussion.

Two types of firms at their 100th patent illustrate how one firm can go a route of greater dispersion across classes and another can stay within certain specific classes. Below is Ford Technologies and International Fragrances and Flavors relative presence across 37 technology categories:

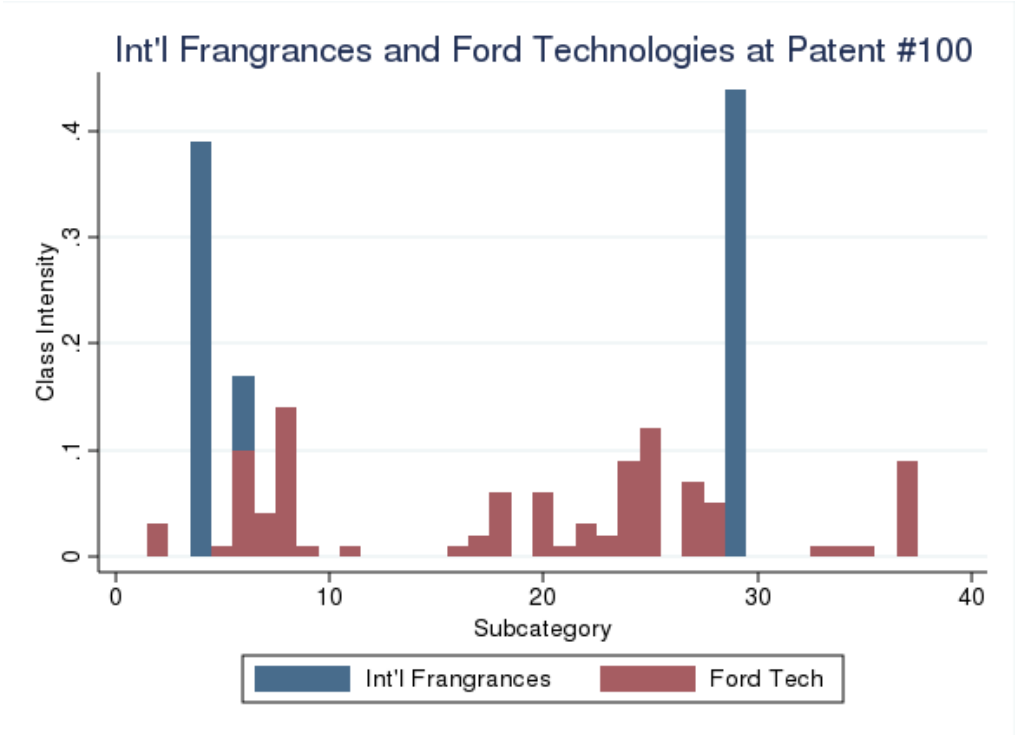


Figure 4: An Example of 2 Firms Class Intensity

Ford Technologies is active in many patent classes, i.e. motors, computers etc. On the other hand, International Flavors and Fragrances stays mostly in sectors like organic compounds. Given the benefits to diversification, a firm like Ford Technologies is more likely to survive and expand over time.

We want to think about what it means for a firm to operate in a new market. Below I categorize a "new" field measures whether or not the firm is entering a new field (i.e. equals 0 or 1). I define this as the bottom 5th percentile in realized research concentration in region of the patent, conditional on firm age. The graph below indicates that the bottom 5th percentile is increasing in firm size (as firms get more dispersed they enter fewer classes where they have 0 research concentration).

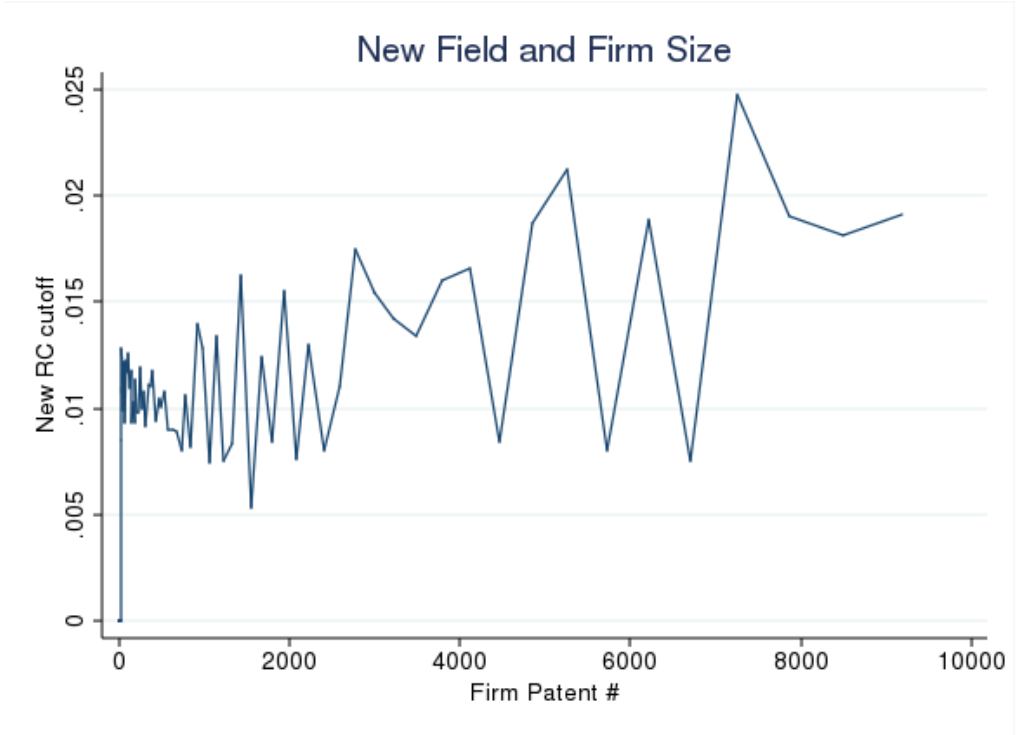


Figure 5: Research Concentration in a "new" field

6 A.2. Sketch of a Model

Entrants attempt to enter the market, grabbing any product line they can. Incumbents also innovate, but must be careful to not cannibalize their own production too much or grab a product line that is not in line with firm valuation. The tradeoff is in getting a product that is close enough to the firm's knowledge base while being far enough that it doesn't simply take from the firm's previous market share.

6.1 A2.1. Static Framework

A product line, q_j, has some value to the firm as it holds a degree of monopoly power on its product. The firm's ability to extract resources from this product line is a function of the knowledge base of the firm relative to that product, decreasing in distance, d_j. A firm gets a flow from each product line, depending on technology level and the firm's ability to hold:

$$\Pi(q_j) = \pi \times \underbrace{q_j}_{\text{determined by quality of innovation}} \times \underbrace{(1 - d_j)}_{\text{determined by knowledge at the firm level}}$$

6.2 A2.2. Dynamic Framework

The flow payoffs from holding product lines sets up opportunities for research and development. As in Klette-Kortum and Akcigit-Kerr, firms innovate in step sizes on products. Thus there will be three relevant components relevant for a firm's expansion. First, there is the product lines existing quality at time t, $q_j(t)$. If the firm holds product line j this is partially the result of their past innovation

Thus innovation targeted to product line j delivers:

If $q_j \notin F_f$:

$$\max_{z_e} \left\{ zV(\mathbf{q} \cup (1+s)q_j(1-d_j)) - Cz_e^{\sigma} \right\}$$

If $q_j \in F_f$:

$$\max_{z_i} \left\{ zV(\mathbf{q} \cup (1+s)q_j(1-\alpha d_j) \setminus q_j(1-d_j)) - Cz_i^{\sigma} \right\}$$

With $\alpha < 1$. This gets at the idea that the more a firm innovates on a product line the better it understands the field. The question of whether the firm decides to target an internal or external product line has to do with the relationship of $(1+s)q_j(1-d_j)$ and $(1+s)q_j(1-\alpha d_j) \setminus q_j(1-d_j)$.

We can now evaluate the key tradeoff at the firm level: direction and quality. There are two components of direction. A firm can hire a worker/scientist to work to innovate on an existing product line, or they can hire a worker to go onto another product line. This speaks to the question of direction. This also speaks to another issue. Firms may be more willing to take a hit in quality of innovation when they are expanding into new fields. Will a firm stay put or attempt to advance into new lines? The quality question comes when we think of the step size–what is the scientist adding to the existing q_j?

This simple framework helps motivate us to think about the mechanisms suggested by the paper, but does not attempt to be used for any sort of structural estimation.

A2.3. Profit in a Specific Market

There is a single final good $Y(t)$ which is used for R&D and produce with fixed labor, L, intermediate goods, $k_j(t)$, and technology or quality level $q_j(t)$ across a measure 1 of product lines.

$$Y(t) = \frac{L^{\beta}(t)}{1-\beta} \int_0^1 q_j^{\beta}(t) k_j^{1-\beta}(t) dj \tag{10}$$

And they pay for the intermediate good held by j, $p_j(t)k_j(t)$. =

From this I have the inverse demand for the intermediate good:

$$p_j(t) = L^{\beta}(t) q_j^{\beta}(t) k_j^{-\beta}(t)$$

Let us take firm f with j product lines, dropping t from the rest of the model. Each product

line has quality q_j that can be improved upon with innovation. While attempting to innovate, each individual product line benefits from monopoly power, producing their intermediate good using technology $k_j = \bar{q}l_j$ where $\bar{q} \equiv \int_0^1 q_j dj$. On each product line they decide to price, and intermediate good capital such that:

$$\Pi(q_j) = \max_{k_j, p_j} \left\{ p_j(1-d_j)^{\beta} k_j - \frac{w}{\bar{q}} k_j \right\}$$

$$\Pi(q_j) = \max_{k_j \geq 0} \left\{ (1-d_j)^{\beta} L^{\beta} q_j^{\beta} k_j^{1-\beta} - \frac{w}{\bar{q}} k_j \right\}$$

Solving for k_j and plugging in:

$$k_j = \left[\frac{(1-\beta)\bar{q}}{w} \right]^{\frac{1}{\beta}} L(1-d_j) q_j$$

Plugging in for optimal k_j, I get the profit for firm j.

$$\Pi(q_j) = \pi q_j (1-d_j)$$

Where $\pi \equiv L(\bar{q}/w)^{\frac{1-\beta}{\beta}}(1-\beta)^{\frac{1-\beta}{\beta}}\beta$

A3. Robustness Checks

A3.1 Akcigit et al. (2016) Measure

As discussed earlier, we have another measure of distance to classes from a firm that varies with firm size and is about the complementarity in production of ideas–this is the Akcigit et al. (2016) proximity measure. Using proximity, we get similar results at Tables 3+4, in terms of where a firm will go based on their past production.

We also see the results in citations:

Table 7: Into Fields

	(1)	(2)	(3)
	$RC_{f,c,t+1}$	$RC_{f,c,t+1}$	$RC_{f,c,t+1}$, =0 for 3 past periods
$K_{f,c,t}$	0.613***	0.609***	.188***
	(88.23)	(89.85)	(50.46)
$H_{f,c,t}$	.379***	0.156***	0.055***
	(51.02)	(20.10)	(32.47)
$RC_{f,c,t}$		0.252***	
		(53.34)	
Observations	12174961	12174961	11097858
R^2	0.483	0.484	.022

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8: Citations on Firm and Scientist Knowledge

	(1)	(2)	(3)	(4)
	LogCit	LogCit	LogCit	LogCit
FMATCH (Prox)	-0.172***	-0.172***	-0.167***	-0.138***
	(-3.85)	(-4.39)	(-4.34)	(-3.61)
HMATCH (Prox)	0.277***	0.200***	0.192***	0.152***
	(6.99)	(5.34)	(5.22)	(4.26)
Observations	1446812	1446812	1446812	1446812
R^2	0.130	0.184	0.185	0.207
Class Fixed Effects	X	X	X	X
Firm Fixed Effects		X	X	X
Firm Size Fixed Effects			X	X
Year Fixed Effects				X

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A3.2 Keeping Only US firms

When we only keep US firms, we still get the importance of history for advancement into new fields. This can be seen in Table 10 below.

Table 9: New Field on Firm and Scientist Knowledge, US Only

	(1)	(2)	(3)	(4)
	new	new	new	new
FMATCH	-0.0976***		-0.102***	-0.145***
	(-28.22)		(-16.51)	(-15.38)
HMATCH		-0.0997***	-0.00445	0.0324***
		(-25.04)	(-0.93)	(3.61)
Observations	1098060	1088444	1088444	132301
R^2	0.070	0.096	0.098	0.156

t statistics in parentheses, errors clustered by firm

Controls for firm, firm size, class, year, individual fixed effects

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

But we see the significance of match for citations go away in the specific case where we have fixed effects for class, year, firm, firm size. This can be seen in Table 11 below.

Table 10: Citations on Firm and Scientist Knowledge

	(1)	(2)	(3)	(4)
	LogCit	LogCit	LogCit	LogCit
FMATCH	-0.0746	-0.141**	-0.137**	-0.0860
	(-1.24)	(-2.92)	(-2.87)	(-1.82)
HMATCH	0.149*	0.119*	0.109*	0.0264
	(2.36)	(2.30)	(2.17)	(0.55)
Observations	689392	689392	689392	689392
R^2	0.148	0.209	0.211	0.238
Class Fixed Effects	X	X	X	X
Firm Fixed Effects		X	X	X
Firm Size Fixed Effects			X	X
Year Fixed Effects				X

t statistics in parentheses, errors clustered by firm

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 11: Variables–Useful in Results section

Variable	Mean	SD	[Min,Max]
Firm unique tech-categories at patent 100	12.08	4.57	[1,29]
Citations	4.11	8.18	[0,447]
Patent Value ($)	13.6M	41M	[100,3401M]
Firm Class Connection $K_{i,c}$	0.19	0.21	[0,1]
Scientist Class Connection $S_{i,c}$	0.20	0.21	[0,1]
Research Concentration, firm ($RC_{f,c}$)	0.23	0.24	[0,1]
Research Concentration, scientist ($RC_{i,c}$)	0.51	0.40	[0,1]
Firm ideas-scientist match (FMATCH)	0.62	0.19	[0,1]
Human capital-scientist match (HMATCH)	0.62	0.18	[0,1]