KU LEUVEN

FACULTY OF ECONOMICS AND BUSINESS

Bridging science and technology in the life sciences

How important is the strength of individual industry-science links for high value inventions?

Timpe Callebaut

r0706897

Master's thesis submitted to obtain the degree of

MASTER IN BUSINESS ADMINISTRATION

Promoter: Prof. dr. Samuel Arts
Daily supervisor: Thomas Schaper

Academic Year 2018-2019

Contents[1]

[1] This document might contain (parts of) texts from earlier submitted documents within the same educational programme, related to the Master's Thesis process of the same author as the author of this work.

Abstract

Building upon existing research on Industry-Science Links (ISLs), this thesis explores the importance of individual links between scientists and inventors on high value inventions. Firstly, we argue that the strength of the individual ISLs is positively related to the usefulness and value of the invention derived from the patent. We expect high value inventions to have a stronger degree of connectedness between scientist and inventor. Secondly, we focus on patents citing novel research, which inherently contain more tacit information. This thesis examines the importance of tacit information and the effect it has on the value of the invention, through the inclusion of author-inventors – who already possess this knowledge. In patents citing novel research, we expect the inclusion of author-inventors to lead to a higher probability and degree of high value inventions. Empirical analysis of patent data supports both hypotheses.

1 Introduction

The link between science and industry has been extensively researched and has resulted in many follow-on scientific endeavors. The notion of co-evolution between science and technology, wherein the exponential advancements in science pushes technology forward, is omnipresent in current literature. Recent research focuses on the factors explaining this link. Firms and corporations increasingly use public science as source for their R&D efforts, leading to more technological advancements (Zucker, Darby, & Brewer, 1998). Early research tells us investments in scientific research result in accelerated economic growth (Mansfield, 1972; Rosenberg, 1974; Sveikauskas, 1981); we aim to dive deeper into the connectedness between different actors in the science-to-industry process. The effectiveness and efficiency of scientific implementation on industry is dependent on many factors, leading to larger scale collaboration with academic institutions, focusing on knowledge transfer (Geuna & Muscio, 2009). This knowledge transfer exists in multiple layers, many of which are tacit and are not included in the initial scientific publications. Transferring tacit knowledge is inherently oxymoronic and attempting it requires a repetitive contact between the scientist and the inventor. In the process of bridging the gap between academic researchers and industry inventors, author-inventors have arisen, occupying prominent positions in scientific as well as technological networks (Agrawal, 2006).

In this thesis, we focus on the importance of the individual ties between science and technology for high value inventions. Using data from the United States Patent and Trademark Office (USPTO), we examine patent applications that have varying degrees of scientific backgrounds through sNPRs (scientific Non-Patent References) against the value and impact of that invention, measured by the number of citations the invention has spawned in a five year window. We also elaborate on the notion of tacit knowledge transfer being crucial, especially in novel research. Utilizing PubMed, a database comprising more than 29 million citations for biomedical literature, we track the inventions back to their scientific origins.

When looking at the state of the art, we find that recent research is becoming more focused on uncovering the nature, directionality and magnitude of the knowledge transfer between science and technology (G. Wang & Guan, 2011). For example, literature has explored if personal traits influence the level of citation-weight in applied patents (Arts & Veugelers, 2018), if novel science publications lead to a higher probability of scientific breakthroughs in the long-term (Veugelers & Wang, 2015), or if spillover knowledge is more likely when there is a citation (Jaffe, Trajtenberg, & Fogarty, 2000). However, where we find there to be an existing gap in the literature is how the strength of individual ISLs (Industry Science Links) influences the probability and degree of high value inventions. Examining the link between high value inventions and the science-technology link can push technological advancements more efficiently. The scientist and the inventor can follow a 'blueprint' that leads to having the highest probability of a high value invention within their field. Looking further into the individual connection between scientist and inventor, we discover the importance of tacit knowledge transfer. We argue that this importance increases when the invention is built on novel research.

Multiple growing industries rely heavily on the importance of ISLs, investing billions of dollars into the expansion of efficient knowledge transfer between science and industry. Artificial Intelligence is the fastest growing industry, collecting $12.4bn worth of investments with a projected $232bn by 2025 (Moore, 2018). With such huge investments, it is imperative to closely determine all components affecting this process. In this paper, we focus on the individual links between science and industry.

By conducting research, we hope to further complete the role that science plays in the technological search. The complete research question:

"How important is the strength of individual ties between science and technology for high value inventions?"

2 State of the art

During the state of the art on this topic, we will be discussing 4 themes that play a key role in analyzing our research question: an introduction to industry science links (ISL), the knowledge transfer process, scientist-inventor relation and determining a high value invention.

2.1 Introduction to Industry Science Links

Since long, there has been the notion that scientific research advances technological innovation - it can even be traced back to Adam Smith in the 18th century (Stephan, 1996). However, empirical data from research only started to emerge in the second half of the 20th century. Scientists initially looked at the link from a distance, figuring out if research affected economic growth, which it positively did (Mansfield, 1972; Rosenberg, 1974; Sveikauskas, 1981) or how university research expenditures are positively related to patent rates (Jaffe & Trajtenberg, 1996). It became clear that while there was definitely a link between science and technology, the determining factors needed to be more nuanced, with the use of multiple empirical indicators. A short summary of general conclusions on ISLs:

The body of literature supports a positive relation between science-industry link and commercialization success (Agrawal, 2006), innovative performance, and technological development (Etzkowitz, Webster, & Healey, 1998; Freeman, 1987, 1991; Lundvall, 1992; Richard R Nelson & Rosenberg, 1993). In 1985, research exploiting the link between scientific articles in patents showed the increasing reliance of private technology on public science (Narin & Noma), and the following scientific articles confirm this link (Branstetter & Ogura, 2005; Hicks, Breitzman, Olivastro, & Hamilton, 2001; McMillan, Narin, & Deeds, 2000; Narin, Hamilton, & Olivastro, 1997; Tijssen, 2001). Since the Bayh Dole Act, scientific advancements have experienced exponential growth, as seen by the number of patent applications from USPTO that has risen much faster than the applications of businesses or individuals.

The Bayh Dole Act[2] is US legislation that was adopted in 1980 and permits institutions (academic, non-profit, for-profit…) to pursue ownership of inventions. This act resulted in a standardized, open process for patent applications, making it not only possible for nonprofits to own patent rights, but also for a unified, standardized database of all patents to be built. The reason this is important for our research is that the standardization makes use of citations in the patent applications, so we can determine how research, patents and inventions are interconnected. Conducting research using the database by USPTO, we can find data of all US patent applications since the 1980s, which will be important for the empirical analysis of our hypotheses.

[2] After World War II, more than 28,000 patents belonged to the US government with only 5% finding a commercialization license. The idea that information and research should be centrally managed and used within a country with the intent of gaining an advantage over other countries originated from a Hamiltonian belief (Stevens, 2004). There was no standardized concept of patent application and if one wanted to use this government-owned information, there was a long and chaotic process that one had to endure to eventually receive this patent/information (Mowery, Nelson, Sampat, & Ziedonis, 2001). Institutions could negotiate an Institutional Patent Agreement (IPA) with the Department of Health, Education and Wellfare (HEW) to "obtain assignment of patentable inventions made with federal funding for which the institution had decided to seek patents" (Gambrell, Kayton, & Trucano, 1969). Of course, these IPAs were not the solution for long-term patenting operations, as a result universities complained to senators Birch Bayh and Robert Dole and an arrangement was made in an act that decentralized the control of patents and federally funded inventions.

In more recent times, the focus on ISLs has shifted in a certain capacity from an institutional/ organizational level to a more individual level, with the intent of explaining the intricacies of the individual science-industry relation more in-depth. This shift can be explained by the greater availability of information in databases. Another explanation is scholars being more aware that knowledge transfer happens in more channels than solely the institutional – between university and industry (Breschi & Catalini, 2010). Moreover, a larger focus on individual characteristics may affect the choice to patent altogether (D'Este & Patel, 2007; Owen-Smith & Powell, 2001). This means that knowledge transfer from scientist to inventor cannot happen to its fullest extent using only a written patent application. More on this in *2.2 Transfer of knowledge*.

The usefulness of connecting industry and science differs across industries (Mansfield, 1995), showcased in the fact that 27 percent of inventions in their sample in the pharmaceutical field required the application of science to avoid delays, as opposed to the 6 percent of inventions in the electronic field. Follow-on research specialized in explaining this phenomenon in the individual invention by explicating one factor: the complexity of the problem that is aiming to be resolved (Fleming & Sorenson, 2004). Empirical analysis conceptualized that when inventions are incremental with a low degree of coupling – meaning that the invention exists through recombination of already existing components – link with science provides little assistance, as search algorithms can locate the most useful combinations. However, when the invention takes place in an increasingly complex space, these algorithms prove insufficient to locate the optimal combination (Fleming & Sorenson, 2001) and science can provide a more directed search for optimal recombining possibilities. The inventions derived from this type of search prove to be more inconsistent, yet more impactful when successful. Fleming and Sorenson describe the technological search as a map with peaks and valleys. Using science can unveil the global peaks and valleys, whereas before there would be an incremental search for local peaks. Having perspective of this landscape may stop inventors from being trapped in a local optimum, and instead find better, more effective outcomes (Fleming & Sorenson, 2004).

2.2 Transfer of knowledge

To understand how the tie between scientist and inventor can potentially influence the value of an invention, we need to look at how (scientific) knowledge is transferred. In his paper on the uneven evolution of human know-how, Nelson (2003) argues one of the reasons for the uneven evolution is that knowledge is cumulative, meaning one learns what others have discovered before them much faster than it would take them to discover the same information. Therefore, this accumulated information can form the basis for what the individual will contribute to human knowledge.

In knowledge, we can distinguish information that is articulate and information that is tacit. While articulate information can be described and communicated, tacit information is harder to articulate (Thompson & Polanyi, 1960). The example Nelson uses is the one of the heart surgeon. In the operating room, they use their theoretical, articulate information to know where to cut, what connects the heart tissue and what the optimal blood pressure is. On the other hand, surgeons use tacit knowledge to control the work of their fingers during delicate procedures. The skill and knowledge cannot be easily explained or written down. Later research renamed these concepts tacit and codifiable knowledge (Katz & Shapiro, 1986).

Until 1995, tacit information remained a one-dimensional concept to define various sorts of information that was deemed ‘hard to explain’, after which a framework was proposed to analyze (tacit) knowledge (Zander & Kogut, 1995). 4 dimensions are represented to categorize knowledge:

1. Codifiability: the degree to which knowledge can be represented by symbols/language
2. Teachability: the degree to which knowledge can be taught

3. Complexity: the degree to which knowledge embodies multiple kinds of competencies
4. System dependence: the degree to which knowledge requires many different experienced people for its application

Other research showed the importance of small world networks to knowledge transfer, specifically tacit knowledge (Newman, 2000a, 2000b, 2001; Wagner & Leydesdorff, 2005). The small world networks are represented by a graph, consisting of various tightly linked cliques where nodes are grouped around. Only a small amount of nodes are interconnecting the various cliques. This model is important to showcase the diffusion of knowledge. Within a clique, individuals have a common language and communication code, supporting the sharing of complex and tacit information among each other. The (fewer) nodes interconnecting different cliques ensure rapid diffusion and recombining capabilities. This model minimizes the risk of lock-in, where scientists are stuck within the limits of their field (Cowan & Jonard, 2004). The risk is palpable when cliques are internally densely connected.

How does this apply to the situation of scientist and inventor? When a scientist publishes a paper or an article, its content is only a fraction of the knowledge that is accumulated throughout the researching process. Information that is deemed unimportant to the eventual outcome – tacit knowledge – is not included. For the inventor however, this information can be crucial to develop the invention efficiently and effectively. In any research, there is a certain level of trial-and-error that happens. Incremental changes to get an optimal model, finding the right variables and control variables, and formulating the hypotheses are a few examples. The trial-and-error process could be codified, but remains uncodified because of the limited added value to the paper. This is called latent information. Agrawal's research (2006) focused on the results of engaging the inventor and the results in commercialization success. He concluded that engaging the inventor favorably influenced the likelihood, as well as the degree of commercialization success. A significant portion of the samples did not engage the inventor at all (36%). Four explanations are put forward for this occurrence by Agrawal. First, firms may undervalue what engaging the inventor can contribute to the commercialization. Second, firms may have in-house experts that could be less inclined to involve the inventor. Third, when the complexity of a problem is lower, it might not even require the uncodified and tacit knowledge derived by the inventor. Finally, the firm might decide against commercializing the invention for strategic reasons.

However, Agrawal notes an important caveat to his research: "the model proposed here offers little insight into the actual functional form of the relationship between engaging the inventor and commercialization success" (2006). Whether the engaging happens through an individual or institutional bond, we do not know. This is important to note as the gap in the literature, for our research to contribute to the state of the art.

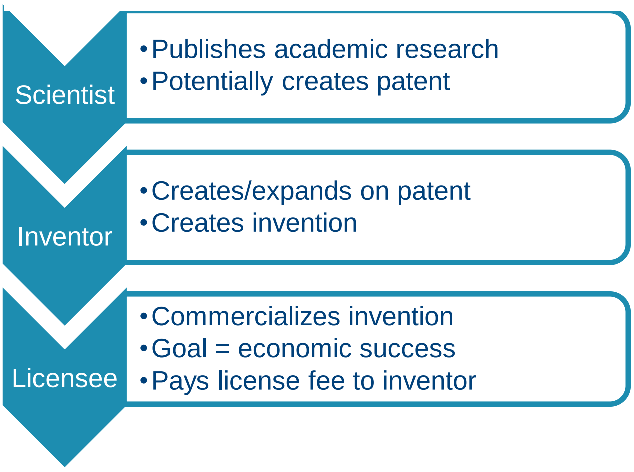
In this paper, we build on Agrawal's framework, deviating from his premise of commercialization probability and success, towards the value of the invention. Whereas previous research focused on the connection between inventor and licensee, this research details the relation between scientist and inventor, and the implication on the value of the invention.

2.3 Scientist-inventor relation

Scientist and technological research environments are typically distinct networks, deriving from different norms of behavior and reward systems. However, research has proven both networks can co-exist and interconnect (Dasgupta & David, 1994). The need for this interconnectedness is suggested in research, noting that the most successful biotech companies are the ones who engage in co-authorship with university professors (Zucker, Darby, & Armstrong, 2002). Gittelman & Kogut (2003) introduce the boundary-spanning 'gatekeepers', individuals connecting the open science community with specific industry firms, biotechnology firms in this research. Other research emphasizes the importance of so-called corporate scientists: enabling flows of external knowledge to corporate researchers (Furukawa & Goto, 2006).

To illustrate the technology transfer process in its entirety, we conclude with the inventor-licensee link. In regards to this relation, Agrawal's research (2006) argues four reasons why science and industry should be linked. Even though the link comprises inventor to licensee, we can make certain parallels to the scientist-inventor relation. See Figure 1: from research to commercialization. Note that while this is not necessarily vital to our research, it is the final part of the technology transfer process and thus worth discussing.

Figure 1: from research to commercialization



1. Latent knowledge from research (codifiable information that is not included in the paper/patent application) can be effectively transferred by inventor.
2. The inventor has mastered the complexity of the problem, and can thus explain the different components to the licensee.
3. The trial-and-error process can be significantly reduced and be made more efficient by intuition gained by the inventor during the experiment phase of the invention.
4. In later developmental stages, the licensee will discover what information needs to be transferred (codified or not). The licensee does not know yet what information they will need throughout the entire process at the outset of the licensing. Therefore, it is more helpful to be able to communicate with the inventor throughout the entire licensing/commercialization process.

In most cases, the parallels from inventor-licensee to scientist-inventor can be made. However, it is important to acknowledge the differences as well, the main ones being the goal and the knowledge transfer process. The licensee's goal is to create economic success and engaging the inventor is a way to optimize their chances at said success. The reason for an inventor to involve a scientist is to create a theoretically sound patent application, thus making the practical applications more likely to succeed. Another fundamentally different component is derived from the divergent reasons clarified above: the method of knowledge transfer. A scientist wants their research to be published and diffused widely, hoping others will build on it. On the other side of the spectrum, a licensee wants the expansion on a patent to be secretive, as the aim is to gain a competitive advantage on the competitors. This leads to a different reward system. Generally, scientists are rewarded through recognition (Ioannidis & Khoury, 2014), while inventors receive monetary rewards (Toivanen & Väänänen, 2012).

To deeply understand the scientist-inventor relation, we take a look at how this relation was researched empirically. Early research used empirical approaches based on patent citation analysis (Narin & Noma, 1985). Specifically NPRs (non-patent references) are seen as the most important indicator to find linkage between science and technology. However, the underlying assumption for approaches like these, is that patent-to-patent references is proof of a technology being directly descendent from science, through research papers. By challenging this bibliometric approach, Meyer (2000) finds there are actors who are predominantly responsible for the link between nanoscience publications and nanotechnology patents – the author-inventor scientist. He defines the author-inventor as a scholar who is involved in both publishing and patenting. Focusing on the individual researcher, active in both scientific and technological fields, can form a broader basis in explaining the complexity of interactions between science and technology (G. Wang & Guan, 2011). Studying the Chinese nanotechnology industry, Wang and Guan aim to establish how science and technology overlay in a rapidly evolving field. Firstly, they capture the interlay in the classical approach, by using the NPRs cited by patents. Those patents have a significantly higher citation rate of scientific references and a higher proportion of self-citations. Secondly, they establish the author-inventor link in the Chinese nanotechnology field through matching the data of publications and patents. They concluded that author-inventors are at the top of the most prolific and highly cited researchers. Lastly, using social network analysis (SNA), Wang and Guan explore the characteristics of both scientific and technological networks to investigate the role of author-inventors in these research networks.

Ideally, our research would follow the same structure as Wang & Guan, by capturing the industry-science connection through NPRs cited in the inventions, matching data of publications and inventions to establish author-inventor link and with this matching data, perform social network analysis to map the actors connecting the different groups. More realistically, however, our available resources confine us to patent citation analysis (see 4: Methodology).

2.4 Determining a high-value invention

Within this thesis, the dependent variable for both hypotheses is the value of the invention. Hence, it is important to discuss the determining factors that add value to an invention. As research finds strong support for the model using patent citations as a measure of invention usefulness (Fleming & Sorenson, 2001), we will use this method. There are however some caveats attached to this imperfect method, which are important to keep in mind. Some companies tend to avoid applying for patents for strategic reasons and the propensity in patenting varies across industries (Levin et al., 1987). Moreover, many inventors do not patent all of their inventions, often limiting their repertoire to their most successful inventions. This may lead to a distorted view of reality wherein patents have a higher value than the underlying amount of (largely unpatented) inventions would suggest. Not losing sight of these imperfections, patents still accurately represent the process of an invention. Using patent data also offers us the possibility of quantitative analysis across a wide range of variables, accessible through the USPTO database.

More research supports this methodology, demonstrating that the number of citations a patent receives, has a high and direct correlation to the technological importance of the patent, as measured by expert opinions, social value and industry awards (Albert, Avery, Narin, & McAllister, 1991; Trajtenberg, 1990). Furthermore, economic value is another important aspect that is highly correlated to the number of citations received by a patent. (Harhoff, Narin, Scherer, & Vopel, 1999; Jaffe et al., 2000).

In conclusion, although it has its imperfections, using patent citation data is still considered the most accurate method in portraying an invention’s value and usefulness.

3 Hypotheses

After a thorough literature review, we formulate two hypotheses that we will either confirm or disprove with our empirical analysis.

“H1: High value inventions have stronger scientist-inventor ties.”

We differentiate three types of scientist-inventor ties, according to the strength of the link. The first one is an indirect tie, in which the inventors are not active scientific publishers, but are reliant on science. In practice, this is indicated by patents that cite at least one sNPR (scientific non-patent reference). Secondly we have a direct tie, wherein at least one of the inventors of the patent has recently published scientific material – the inventor is active in science. The strongest tie is found when at least one of the inventors has published a scientific article that is directly related to the patent, i.e. relevant prior art. It entails that the invention is an extension of the research conducted by the same individual. The expectation is that high value inventions have a stronger scientist-inventor tie.

With billions of euros of public and private money being invested in Europe in fostering ISLs (Lanxon, 2017) it is imperative to closely examine all determinants in the transfer process from science to industry. Seeing as this process is evolving from an institutional point of view to an individual one, this research contributes greatly to existing literature and consequently, practical applications.

“H2: Patents citing novel research have a higher probability of developing into a high value invention when author-inventors are involved.”

Since novel research includes a new path formed by the scientist, it contains a vast amount of tacit knowledge. Consequently, anyone wanting to work on that research towards an invention lacks a large part of the knowledge created by the scientist. This knowledge is largely not codified in the research papers and can thus only be acquired in an environment of closer connectedness between the scientist and inventor. In including author-inventors, we know that the inventor and the scientist are mostly the same individual. Therefore, we know they possess the tacit information acquired in the research process – a clear distinction to the inventor that uses the codified information in novel research towards an invention.

The potential of novel research is much more volatile than established paths of research, meaning that practical applications can be extremely valuable, worthless, or anywhere in between. It also implies that if we are able to identify and remove any paths that lead to sub-optimal results, inventors can focus their efforts in developing high value inventions, derived from novel research. It would mean a significant contribution to the literature.

For both hypotheses, we use the framework Fleming & Sorenson (2004) and Agrawal (2006) provided and build on their research. In conclusion, we aim to add to the existing literature by providing empirical proof of individual links between science and industry.

4 Methodology

As reviewed in the state of the art, patent data does not offer perfect measures for the invention and this provides a notable caveat to the paper: preferably, this research is performed in combined empirical analysis of Social Network Analysis (SNA), intensive case studies and patent citation data. With the limited resources available, however, we choose to use patent citation data.

Ideally, we want to achieve causality through a randomized trial. Unfortunately this is impossible given the models we estimate, as we would be in violation of the exogeneity assumption of the Gauss-Markov theorem. The reason for the violated theorem is found in the scientist-inventor variable, which is inherently endogenous. This means that the explanatory variable is correlated with the error term, making the eventual estimate of the regression biased and consequently, confounding the causal effect. The type of endogeneity we encounter is caused by simultaneity and omitted-variable bias. To have causal inference, we would need to create a randomized trial wherein the inventions of a representative group of inventors with a strong scientist-inventor tie would be compared to those of a group of inventors that do not have a strong scientist-inventor tie. Such an approximation would require a natural experiment or an instrumental variable technique, both of which we do not have. Thus, we have tweaked our first hypothesis from the initial: *'a strong individual scientist-inventor tie leads to a higher value invention'* to: *'high value inventions have stronger scientist-inventor ties'* – effectively removing the causality problem.

Data sources

Data is extrapolated from two sources. Firstly, we look at patent data from the USPTO database. All patents from 1980 onwards are available through a standardized application process, making comparison and analysis possible. Among others, the application includes summary of invention, technical problem, solution to problem, industrial applicability, reference sign list, citation list, patent and non-patent literature and sequence listing (USPTO, 2007). More information on the chosen variables is found in 5.1 Defining parameters.

Secondly, we need to trace back the patents from USPTO to the scientific literature. To do so, we utilize PubMed, a database comprising more than 29 million citations for biomedical literature from MEDLINE, life science journals, and online books (NCBI, 2018). PubMed contains biomedical literature, meaning we will largely focus our empirical approach on this industry. Biomedical patents will evidently mostly be linked to biomedical research papers, articles, journals and books.

However, PubMed does not uniquely identify individual scientific authors in its database. This entails that when two individuals share the same last name and first initial, they are viewed as the same person, which makes it difficult to prove correct scientist-inventor links in our research. To circumnavigate this issue, the author name disambiguation model by Torvik et al. (2005) and Torvik & Smalheiser (2009) is used. The model contains a simple yet powerful method to identify author disambiguity, i.e. by constructing a similarity profile between a pair of articles, based on title journal name, co-author names, MeSH terms, language affiliation and name attributes. With the help of the free and publicly available service "Author-ity" by Torvik et al. (2005) and linking it to the disambiguated inventor database (Li et al., 2014), we can retrieve a continuum from publications to patents of each of the scientist-inventors.

Dependent variable

Our dependent variable is a constant throughout our research: probability and degree of value of the invention. We use the number of forward citations in a five year window following the patent's grant date as our dependent variable – as more thoroughly explained in *2.4 Determining a high-value invention*.

Control variables

In the attempt of approximating a randomized trial as much as possible, control variables are introduced. They isolate different factors influencing the empirical results. Our aim is to keep the control variables constant, in order to assess the relation between dependent and independent variables.

To investigate deeper and to take more controlling factors into account, we perform a multivariate analysis through fixed effects negative binomial models of forward citations. The reason this MLE is preferred over OLS is that the dependent variable (citation counts) only registers whole number values – using a linear regression can thus yield inconsistent, inefficient and biased estimates. According to research in 1998, count models can help to avoid these issues (Cameron & Trivedi). We expect the variance to exceed the mean – overdispersion – for which a variance of count models is appropriate: the fixed effects negative binomial (Hausman, Hall, & Griliches, 1984). What advantages precisely does this model provide? In general, it allows more control for a variety of factors. Over technological domains, there are differences in standards regarding citations. This array of heterogeneity makes it unwise to directly compare patents of different domains. In line with Fleming & Sorenson's framework (2004), *technology control* is introduced, which refines the fixed effects. First, it measures the average amount of received citations per patent per USPTO class (technology domain) within our time frame (equation 1), weighting these parameters conforming to the patent's class assignments (equation 2). Expending on technology control, a number of other variables are incorporated. The first one is *number of major classes*. Similar to science, wherein research that covers multiple literatures generally has a higher probability of being cited more, patents that cover a wider range of technologies have a higher tendency of being cited. This will account for the interdisciplinary coverage of the patent. Second, with the aim to account for differences in citation patterns from patents with 'cutting edge' technologies (Katila, 2002), *recent technology* and *technology mean* are introduced. Third, we decide to include the variable *coupling*, as it controls for the degree of re-combinatorial challenge that an invention overcomes. *Number of subclasses* measures the number of components or problem domains in an invention. Patents belonging to only one subclass account for eight percent of all patents, and although these inventions most likely arise from a recombination process, this combination occurs at a finer grain than we can measure. For this reason, we add *single subclass dummy* variable to discover differences between these inventions and those assigned to multiple subclasses. Other control variables include: *university patent, number of inventors, and prior art count.*

$$(1)\ \text{Average citations in patent class } i = \mu_1 = \frac{\sum_{j \in i} \text{citations}_j}{\text{count of patents}_j \text{ in subclass}_i}$$

$$(2)\ \text{Technology mean control patent } k = M_k = p_{ik}\mu_i$$

Why use a Fixed Effects Model (FEM) as opposed to Random Effects Model (REM)? The underlying assumption of FEM is that non-included explanatory variables are correlated with included explanatory variables. Thus, FEM is able to capture the heterogeneity through the estimation of individual dummies. This can be seen in the inclusion of 316 mainclass dummies and 39 application year dummies in the estimation models. FEM will always be consistent, while this is not necessarily the case in REM.

Hypothesis 1

In establishing the independent variables, we dissect the hypothesis '*High value inventions have stronger scientist-inventor ties'*. With high value inventions being the dependent variable, what remains is 'stronger scientist-inventor ties'. How do we measure the link empirically? As discussed in: 3 Hypotheses, we distinguish the degree of scientist-inventor link in three levels. We create a new variable *strength of tie* with four possible inputs. Firstly we have the non-existing tie, patents that do not have a single sNPR, but possibly have cited other patents (1). This is our baseline. Secondly we have the indirect tie, patents that cite at least one sNPR (2). Next up is the direct tie, where at least one of the inventors has been recently (last 3 years) active in the scientific field (3). Lastly, the strongest tie is when the patent cites research that has been published by one of its own inventors – the author-inventors (4). It entails that the invention is an extension of the research conducted by the same individual.

Hypothesis 2

In the second hypothesis, we explore tacit knowledge transfer between novel research and subsequent inventions. In identifying novel research, methods were inspired by MedMeSH[3] and the philosophy of mining tacit knowledge from biomedical literature. Preceding research has indicated using MeSH (Medical Subject Headings) terms results in accurate representation of the entire text if screened appropriately (Bhattacharya, Ha−Thuc, & Srinivasan, 2011). Using MeSH terms reduces computing time, since screening full texts of more than 29 million pieces of literature is very strenuous and time-consuming. Moreover, MeSH terms are always available through PubMed's services, as opposed to full texts often requiring a subscription, and the use of MeSH terms enable higher dataset throughput (Shan, Lu, Min, Qu, & Zhang, 2016). How can we use this method to distinguish novel research? Novel research is conceptualized as research which recombines existing knowledge components in a previously unexplored fashion, and has a higher probability of scientific breakthroughs (Uzzi, Mukherjee, Stringer, & Jones, 2013). Wang, Veugelers and Stephan (2015) find that only 11% of scientific papers make novel combinations (of Web of Science articles published in 2001).

In our dataset, we make use of the approach suggested by Boudreau et al. (2016) to determine scientific novelty, based on the MeSH lexicon. We look for novelty in MeSH term combination in relation to the existing literature. Comparing the MeSH term combinations of a research proposal with combinations that appear in the PubMed database in its entirety, we examine all possible pairs of MeSH terms and determine what fraction of these pairs for a given proposal had not previously appeared in the literature. The advantage of this method is that it results in a variable '*Novelty'* expressed as a percentile, with 1% being the least novel and 100% being the most novel. The variables in the dataset are represented by dummy variables for patents citing the top 1%, top 2%, and top 5% novel research over all fields.

In the second part of the hypothesis, author-inventors arise. In the first hypothesis, we measure the degree of scientist-inventor link through 4 levels. Here, we only care for the distinction between involvement of author-inventors or no involvement. Hence, we use a dummy variable that is parallel to the strongest degree of scientist-inventor tie: when at least one of the inventors has published a scientific article that is directly related to the patent, it entails that the invention is an extension of the research conducted by the same individual, making the creator an author-inventor. The variable for this is own_science.

[3] MedMeSH is the National Library of Medicines' (NLM) controlled vocabulary thesaurus specified for indexing articles from PubMed. Keywords are not assigned by authors, but rather by professional science librarians trained specifically to perform this task (Boudreau, Guinan, Lakhani, & Riedl, 2016).

5 Empirical analysis

5.1 Defining parameters

In order to provide empirical proof of the hypotheses, we need to establish all the necessary parameters required to wholly estimate the models as accurately as possible.

Firstly, the length and timing of the time frame of the data should be carefully chosen, with consideration to sufficient lag and lead time both before and after the sample to provide the optimal dataset. We chose all USPTO patents granted between 1995 and 2004, assigned to at least one of the sixteen three-digit USPC classes of category 3 (“Drugs&Med”) in the NBER patent classification from 2006. The reasoning behind this category lies in the fact that our access to scientific publications is limited to PubMed. The dataset we have acquired contains 176,830 distinct patent-level observations.

We differentiate sources in our dataset. Most of the patent related information is derived from EPO PATSTAT 2018, Spring edition. This includes, amongst others, application & publication years, patent citation links, assignee type, and technology control. Information on the USPC (sub)classes comes from USPTO flat files and bulk data as well as the Fung Institute at Berkeley patent data files. Inventor disambiguation information was used through the Li et al. (2014) database. Author disambiguation information comes from the Author-ity database. sNPRs (from USPTO patents to all PubMed articles) are based on research from 2014 (Agrawal, Lincoln, Cai, & Torvik)– the Patci tool. The author-inventor link disambiguation being used is based on the Author-ity exporter (Tuomela, Fegley, & Torvik, 2016).

Table 1: Legend of variables in dataset

fwd_cit_5y	Number of forward citations to focal patent, over a 5-year window
appyear	Year in which the patent application was received by the USPTO
mainclass	Primary 3-digit class of focal patent
univ_patent dummy	= 1 if at least one assignee of focal patent is a university or public research institute
science_link dummy	= 1 if focal patent includes at least one sNPR to a PubMed article
snpr_count	number of sNPRs to PubMed included in focal patent
invt_count	number of inventors listed on focal patent
*_snpr_nov variables	sum, max, min, std.dev. of scores of scientific novelty across all sNPRs listed on the focal patent.
nov_snpr_top* dummy	= 1 if at least one sNPR of the focal patent is in the top 1%, 2% or 5% based on novelty scores of all articles in the specific year of publication, and respectively in the top 1%, 2%, 5% in the specific field and year (in case “_f”)
active_au dummy	= 1 if at least one inventor of the focal patent is active as author on scientific publications within the last 3 years prior to application date of focal patent + number of active authors on focal patent.
own_science dummy+count	= 1 if focal patents lists as sNPR at least one of the active authors own publications + count of such links among sNPRs.
recent_tech + tech mean	average age of prior art cited
prior_art_count	number of prior art references
coupling	degree to which an invention's components have been previously recombined.

5.2 Hypothesis 1

Inherently, count data models do not fare well in an OLS-model, as the overdispersion – as well as the endogenous scientist-inventor variable – confound the causal effect. Hence, we rely on robustness checks over different estimates to prove the legitimacy of the linear regression. If we want to examine the effect of ISL strength on the value of the invention in a linear regression, we need to log-transform all data with the exception of binary variables. Count data need to be transformed in order to perform OLS, while the other variables are log-transformed to account for their skewed distribution. The natural logarithm of 0 is undefined, thus we add +1 to data with 0 as a base value, making the base value 1.

5.2.1 OLS – linear regression

To determine whether high value inventions have stronger scientist-inventor ties, we use the dummy variables of each of the four strengths, minus one for the first degree, which we use as our base level. The dummy variables are (in order of strength of link): science link, active author and own science. To account for any differences influencing the empirical results, we have isolated following control variables: recent tech (log), prior art (log), coupling (log), subclass (log), single subclass, inventor count (log), university patent, tech mean (log). To include fixed effects in terms of application year and the main class to which the patent belongs, dummy variables were created for each of the values in these variables, giving us an additional 316 dummy variables for mainclass and 39 dummy variables for app_year.

We start our initial model with solely the 3 dummy variables as dependent variables and proceed by including interaction variables that encompass the strength of the variables (science link AND author = 1, science link AND author AND own science = 1…). Model 4 and 5 include certain selection variables.

Table 2: OLS Output H1

		Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Model fit	R^2	0.332	0.333	0.333	0.319	0.289
	Durbin-Watson	1.952	1.952	1.952	1.966	1.985
Beta coefficients	science link	**0.106 (0.003)**	**0.120 (0.004)**	**0.122 (0.004)**	Selection variable	Selection variable
	active author	**0.034 (0.003)**	**0.043 (0.003)**	**0.043 (0.003)**	**0.032 (0.004)**	Selection variable
	own science	**0.055 (0.004)**	**0.060 (0.004)**	0.012 (0.016)	**0.048 (0.005)**	**0.047 (0.004)**
	science*author	**/**	**-0.026 (0.005)**	**-0.029 (0.005)**	**/**	**/**
	science*author *own science	**/**	**/**	**0.050 (0.016)**	**/**	**/**

Numbers in bold indicate that the coefficient estimate differs signification from zero with 95% confidence. Standard error shown in brackets.

We start off by examining the model fit of the various models. With an R^2 of 0.332-0.333 and Durbin-Watson of 1.953 it can be stated that we have a model that is sufficiently consistent and significantly more explanative than a null-model. The Durbin-Watson test is consistently close to two, which proves there to be no autocorrelation in the residuals in any of the models.

The beta coefficients – when significant – are used to interpret the data and the effect the independent variables have on the dependent variable. The first model, which only includes the initial dummy variables that define the different industry-science links, provides us with significant data which can be interpreted as follows: when a patent cites at least one sNPR, the number of forward citations in a 5-year window will increase with 10.6%. When at least one of the inventors of the patent has recently published scientific material, the number of forward citations increases with 3.4%. Finally, when a patent references a scientific publication by one of the patent's own inventors, the number of forward citations increase by 5.5%. This data runs parallel to our hypothesis, wherein we state that the stronger the industry-science link, the more useful/valuable the invention will be. We can clearly state that having a science link is by far the more impactful indicator. Moreover, we see that having an author-inventor involved increases the value of the invention by an additional 3.4% and that, when the author-inventor cites his own work in his patent application, there is a final increase of 5.5% in the value of the invention.

Having singular empirical proof of confirmation of the hypothesis in the OLS-model is promising, however it is not conclusive yet. Therefore, we incorporate two interaction terms into the equation. These interaction terms define the stacking degree of ISL strength – namely a dummy variable that is 1 when there is a science link AND an active author and a second interaction term which is 1 when there is a science link AND an active author AND an inventor citing his own work. Respectively, this is represented in model 2 and model 3. First, we discuss the results of model 2: the variables taken over from the previous model (science link, active author, and own science) have all increased positively, while remaining significant. This can be attributed to the inclusion of the new variable, which represents a negative correlation – when both a science link and an active author are present, the number of forward citations decreases with 2.6%.

Fleming & Sorenson suggest an explanation for this decrease in their research in recombinational challenges in high value inventions (2004): inventors that are also active authors might decide to not use science links when the invention is derived from a new combination of uncoupled components. Fleming & Sorenson state that: *"Science offers limited benefits to inventors working with modular components – given the cost of searching and digesting the scientific literature, the price tag for using science likely exceeds its benefits for those working with uncoupled components." (2004, p. 926).*

The third model includes the second dummy variable. After swift inspection however, it is obvious that the own_science variable has become highly insignificant. A short analysis tells us that this new interaction term is 99% identical to the own_science variable. This logically makes sense – when an inventor cites his own work in the patent, there is automatically a science link as well as an author-inventor. Hence, this model yields inaccurate results.

Model 4 includes a selection variable to exclude any observations not citing a single sNPR, to isolate the effect active authors and own science have on patents with science links. Both variables exhibit a positive significant relationship in proportion to earlier models – own science's relation being slightly more impactful than the active author. Finally, model 5 incorporates a selection variable in which only patents that have a science link AND active author are included, to showcase ultimately that the strength of inventor-scientist matters and that high value inventions have stronger scientist-inventor ties.

5.2.2 Robustness checks

Research has established that OLS yields suboptimal results in count data (Gardner, Mulvey, & Shaw, 1995). A common recurrence in models that deal with count data is overdispersion. This means that the standard deviation of the dependent variable exceeds the mean. When dealing with count data, consideration of a maximum likelihood estimate (MLE) is advised. In the case of overdispersion, a negative binomial model is preferred over a Poisson regression, since a Poisson regression can yield inaccurate results. The negative binomial model accounts for the overdispersion and brings the indicator of overdispersion (Value/df) down. Below displays how the standard deviation is twice as large as the mean in the dependent variable.

Table 3: proof of overdispersion in dependent variable

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
fwd_cit_5y	176830	0	2335	26,72	52,829	2790,886
Valid N (listwise)	176830					

To make sure our model works in different estimates, we create a robustness check using the negative binomial model. In this MLE model, the variables stay the same – however, log-transformed data is reverted back to their original state. The output can be used to verify legitimacy of the OLS-model and its interpretation.

The Omnibus test is used to determine whether the model including the predictors represents a significant improvement over an intercept-only or null model, which it does in this case.

Table 4: MLE Omnibus test H1

Likelihood Ratio Chi-Square	df	Sig.
1414,332	3	,000

Dependent Variable: fwd_cit_5y

In judging the goodness of fit of the models, we can see a clear difference between the Poisson model and the negative binomial regression. Indicators of a good model fit are the Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC), which should be minimized. Alongside these criteria, the Value/df represents the overdispersion. Ideally it should be around 1, the higher the value/df, the higher the overdispersion. As can be seen, the negative binomial is the preferred model.

Table 5: goodness of fit MLE

	Poisson regression	**Negative Binomial regression**
AIC	7,291,557	1,311,702
BIC	7,291,596	1,311,752
Value/df	43.594	1.167

The beta coefficients of the negative binomial model are consistent with the OLS-model in terms of correlations between IVs and DV. Output data can be found in *Appendix 1.* With this robustness check, we have verified the legitimacy of the OLS-method over different estimate models.

5.2.3 Conclusion

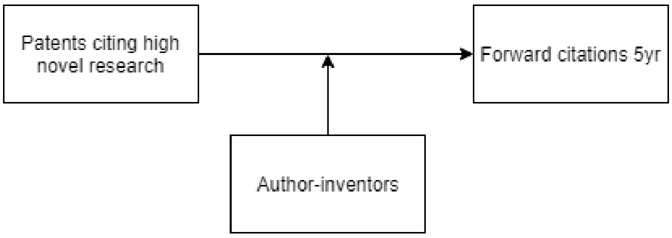
The results provide strong evidence that the strength of industry-science links are in significant positive relation with the number of forward citations in a 5-year window – suggesting that high value inventions have stronger scientist-inventor ties.

5.3 Hypothesis 2

"H2: Patents citing novel research have a higher probability of developing into a high value invention when author-inventors are involved."

A moderating effect is found in this hypothesis, which is represented in the figure below. Empirically, our aim is to prove that author-inventors have an impact on the value of a patent that cites high novel research.

Figure 2: moderating effect H2



Because of this premise, we have to adjust our dataset to only include patents that have a science link – patents that cite at least one sNPR. Including all patents would lead to a confounding effect on other parameters. We implement a selection variable to exclude patents without science link.

5.3.1 OLS – linear regression

Similar to the first hypothesis, we utilize a linear regression in order to be able to interpret the output. The same control variables will be used: binary variables as well as log-transformed non-binary variables. The dependent variable is still the number of forward citations in a 5-year window, log-transformed. The independent variable is the dummy variable that measures whether or not the patent has cited research that is rated in the top 5% of novelty across all fields (nov_snpr_top5). In Model 2, to measure the moderating effect, we have included an interaction term OWN_NOVTOP5, which is obtained by multiplying nov_snpr_top5 with own_science.

The model fit, measured by the R^2 value of 0.321 and Durbin-Watson of 1.96, is adequate and thus we can proceed to the interpretation of the parameter coefficients.

Table 6: OLS output H2

		Model 1	**Model 2**
Model Fit	R^2	0.321	0.321
	Durbin-Watson	1.964	1.965
Beta coefficients	nov_snpr_top5	**0.075 (0.004)**	**0.056 (0.004)**
	OWN_NOVTOP5	**/**	**0.043 (0.006)**

In Model 1 – without moderating effect – the IV has a significant, positive relation to the DV – meaning that when patents cite high novel research, the number of forward citations increase by 7.5%.

In Model 2 – with moderator own_novtop5 – the moderator measures how much of the initial effect can be attributed towards the inclusion of author-inventors in the patent. As can be seen, the initial effect is still significantly positive, but has decreased to 5.6%. Here, 4.3% of the 5.6% is represented by the author-inventor, or in relative terms 77% of the initial effect is attributed towards of the author-inventor. This initial finding vastly supports our hypothesis. Extensive data output including control variables and significance can be found in *Appendix 2*.

5.3.2 Robustness checks

To verify whether these results hold up under multiple circumstances, we create a number of robustness checks. Before changing the estimate model to MLE, we change the premise of the model from using the top 5% novelty to the top 2% and top 1%, to check for consistency in the results.

Table 7: OLS robustness check H2

		Model 3	**Model 4**
Model Fit	R^2	0.320	0.318
	Durbin-Watson	1.965	1.965
Beta coefficients	nov_snpr_top2	**0.058 (0.006)**	**/**
	OWN_NOVTOP2	**0.039 (0.008)**	**/**
	nov_snpr_top1	**/**	**0.053 (0.008)**
	OWN_NOVTOP1	**/**	**0.043 (0.010)**

The results stay consistent throughout the different models, with the R^2 decreasing ever so slightly as the patents include only the top 2% and top 1% of patents.

As a final robustness check, the estimate is changed from OLS to MLE – from a linear regression to a negative binomial one. The negative binomial models are more suitable to handle count data because they account for the overdispersion which highly often plagues count data. See 5.2.2 Robustness checks. We revert back to using the original data (non log-transformed) but otherwise keep all variables the same.

The model, as measured by the Omnibus test, is different from 0, representing a signficant improvement over the null-model. The negative binomial model is preferred over the Poisson regression, as Poisson does not account for overdispersion.

Maximum likelihood models do not allow for direct interpretation because every value on the estimated curve will contain a seperate marginal effect. When it comes to interaction terms specifically, the correct interpretation can prove highly difficult, and often has an opposite sign (Ai & Norton, 2003). This is the case, as we can see in our moderator (an interaction term), which has an opposite sign. See *Appendix 2* for extensive data output regarding negative binomial models.

5.3.3 Conclusion

The results demonstrate consistently that when author-inventors are involved in patents citing highly novel research, the probability of high value inventions increases significantly. A large portion of the initial effect is explained by the inclusion of author-inventors.

6 Discussion

The results provide strong evidence that individual industry-science links are highly important for high value inventions. Firstly, the strength of the inventor-scientist tie is shown to be an influential factor towards the value of inventions. The biggest percentual increase in forward citations derives from having a science link altogether – patents that have at least one scientific non-patent reference. When the strength of the link increases, so do the number of forward citations, making the invention more useful and valuable – as discussed in *2.4 Determining a high-value invention*. Agrawal's research showcases the importance of engaging the inventor in creating commercial success of a patent (2006). Moreover, because of this importance, boundary-spanning gatekeepers emerge between science and industry. Author-inventors are part of this group, occupying prominent positions in both scientific and technological networks. Our research supports and builds on Agrawal's, by providing evidence that when author-inventors are involved, likelihood and propensity of high value inventions increase, highlighting the importance of these boundary-spanning gatekeepers across networks. Uncovering the nature of other types of gatekeepers between different technological fields, academics and industry, and institutional and individual research can prove vital in increasing the speed of (primarily tacit) knowledge transfer.

Tacit information proves to be more important when dealing with patents citing highly novel research. When author-inventors cite their own novel research, the probability of a high value invention increases significantly; highlighting the importance of tacit knowledge transfer in mostly unexplored scientific fields. Patents derived from novel research have a higher variance in value output, meaning the potential is higher, but so is the risk. If there is a way to remove some of the risks, without endangering the potential rewards, high novel inventions suddenly become more appealing to pursue. In cementing the importance of the tacit knowledge transfer, we may neutralize some of the risks without compromising the potential.

This research was built on the framework of Fleming & Sorenson, utilizing similar control variables and dependent variables to attempt to (dis)prove other hypotheses. The framework is based on the assumption that patent citation analysis is the most accurate form of data analysis for this type of research. However, certain limitations in this framework need to be addressed. We rely on patent data from the USPTO, linked to research data from PubMed. This raises two issues. Firstly, we assume the results from these mostly biomedical fields is applicable over the life sciences. Secondly, many inventors only patent their most successful/highest potential inventions, which may lead to a distorted view of reality wherein patents have a higher value than the underlying amount of (largely unpatented) inventions would suggest. Ideally, a combination of patent citation analysis, social network analysis and intensive case studies would represent the most accurate data, derived from multiple angles.

This thesis suggests that when small world networks become more intertwined with a higher amount of interacting nodes – boundary-spanning gatekeepers – the quality and speed of knowledge transfer increases, which has the potential to lead to higher value inventions.

If increasingly more research highlights the importance of including scientific research and individual ISLs, why is it that only 46% of patents in our dataset cite at least one sNPR? There may be multiple explanations for this phenomenon. Firstly, inventors may undervalue the knowledge science can contribute to one's invention. Moreover, asymmetric ignorance/arrogance can occur, in which the inventor does not know how much they do not know. This can lead to a lower level of engagement with science than if the inventor would know how much knowledge is acquired by science. Secondly, as Fleming & Sorenson's research indicated, science may prove futile when the patent is based on combining relatively independent components (low degree of coupling). Finding useful new configurations and turning them into patents proves relatively easy (Fleming & Sorenson, 2004). Finally, on the other side of the spectrum are patents that include marginal improvements over previous patents, eliminating the need for sNPRs and thus negating a link with science.

In conclusion, we hope to have elaborated on the intricate role science plays in technological advancement and how individual ISLs impact the rate and value of the invention. This thesis adds to the already extensive literature by bringing a more nuanced view of the benefits of including science into technological development.

7 Bibliography

Agrawal, A. (2006). Engaging the inventor: exploring licensing strategies for university inventions and the role of latent knowledge. *Strategic Management Journal, 27*(1), 63-79. doi:10.1002/smj.508

Agrawal, A., Lincoln, M., Cai, H., & Torvik, V. I. (2014). Patci—a tool for identifying scientific articles cited by patents.

Ai, C., & Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics Letters, 80*(1), 123 - 129. doi:https://doi.org/10.1016/S0165-1765(03)00032-6

Albert, M. B., Avery, D., Narin, F., & McAllister, P. (1991). Direct validation of citation counts as indicators of industrially important patents. *Research Policy, 20*(3), 251-259. doi:10.1016/0048-7333(91)90055-U

Arts, S., & Veugelers, R. (2018). Taste for Science, Academic Boundary Spanning and Inventive Performance of Scientists and Engineers in Industry. *IDEAS Working Paper Series from RePEc*.

Bhattacharya, S., Ha−Thuc, V., & Srinivasan, P. (2011). MeSH: a window into full text for document summarization. *Bioinformatics, 27*(13), i120-i128. doi:10.1093/bioinformatics/btr223

Boudreau, K. J., Guinan, E. C., Lakhani, K. R., & Riedl, C. (2016). Looking across and looking beyond the knowledge frontier: intellectual distance, novelty, and resource allocation in science.(Report). *Management Science, 62*(10), 2765. doi:10.1287/mnsc.2015.2285

Branstetter, L., & Ogura, Y. (2005). *Is academic science driving a surge in industrial innovation? Evidence from patent citations*. Retrieved from

Breschi, S., & Catalini, C. (2010). Tracing the links between science and technology: An exploratory analysis of scientists' and inventors' networks. *Research Policy, 39*(1), 14-26. doi:10.1016/j.respol.2009.11.004

Cameron, A. C., & Trivedi, P. K. (1998). *Regression analysis of count data*: Cambridge : Cambridge University press.

Cowan, R., & Jonard, N. (2004). Network structure and the diffusion of knowledge. *Journal of Economic Dynamics and Control, 28*(8), 1557-1575. doi:10.1016/j.jedc.2003.04.002

D'Este, P., & Patel, P. (2007). University-industry linkages in the UK: What are the factors underlying the variety of interactions with industry? *Research Policy, 36*(9), 1295. doi:10.1016/j.respol.2007.05.002

Dasgupta, P., & David, P. A. (1994). Toward a new economics of science. *Research Policy, 23*(5), 487-521. doi:10.1016/0048-7333(94)01002-1

Etzkowitz, H., Webster, A., & Healey, P. (1998). *Capitalizing knowledge: New intersections of industry and academia*: suny Press.

Fleming, L., & Sorenson, O. (2001). Recombinant Uncertainty in Technological Search. *Management Science, 47*(1), 117-132. doi:10.1287/mnsc.47.1.117.10671

Fleming, L., & Sorenson, O. (2004). Science as a map in technological search. *Strategic Management Journal, 25*(8-9), 909-928. doi:10.1002/smj.384

Freeman, C. (1987). Technology policy and economic policy: Lessons from Japan. *London: Pinter*.

Freeman, C. (1991). Networks of innovators: a synthesis of research issues. *Research policy, 20*(5), 499-514.

Furukawa, R., & Goto, A. (2006). The role of corporate scientists in innovation. *Research Policy, 35*(1), 24-36. doi:10.1016/j.respol.2005.07.007

Gambrell, J. B., Kayton, I., & Trucano, M. (1969). Patent Law (1968-69). *Ann. Surv. Am. L.*, 139.

Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychological bulletin, 118*(3), 392.

Geuna, A., & Muscio, A. (2009). The governance of university knowledge transfer: A critical review of the literature. *Minerva, 47*(1), 93-114.

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

Harhoff, D., Narin, F., Scherer, F. M., & Vopel, K. (1999). Citation Frequency and the Value of Patented Inventions. *Review of Economics and Statistics, 81*(3), 511-515. doi:10.1162/003465399558265

Hausman, J., Hall, B. H., & Griliches, Z. (1984). Econometric Models for Count Data with an Application to the Patents-R & D Relationship. *Econometrica, 52*(4), 909-938. doi:10.2307/1911191

Hicks, D., Breitzman, T., Olivastro, D., & Hamilton, K. (2001). The changing composition of innovative activity in the US—a portrait based on patent analysis. *Research policy, 30*(4), 681-703.
Ioannidis, J. P. A., & Khoury, M. J. (2014). Assessing value in biomedical research: the PQRST of appraisal and reward. *JAMA, 312*(5), 483. doi:10.1001/jama.2014.6932
Jaffe, A. B., & Trajtenberg, M. (1996). Flows of knowledge from universities and federal laboratories: Modeling the flow of patent citations over time and across institutional and geographic boundaries. *proceedings of the National Academy of Sciences, 93*(23), 12671-12677.
Jaffe, A. B., Trajtenberg, M., & Fogarty, M. S. (2000). *The meaning of patent citations: Report on the NBER/Case-Western Reserve survey of patentees*. Retrieved from
Katila, R. (2002). New product search overtime: Past ideas in their prime? *Acad. Manage. J., 45*(5), 995-1010. doi:10.2307/3069326
Katz, M. L., & Shapiro, C. (1986). How to license intangible property. *Quarterly Journal of Economics, 101*, 567.
Lanxon, N. (2017). Record $19 Billion in Investments Shows Europe Tech Is Just Fine. In.
Levin, R., Klevorick, A., Nelson, R., Winter, S., Gilbert, R., & Griliches, Z. (1987). Appropriating the Returns from Industrial Research and Development; Comments and Discussion. *Brookings Papers on Economic Activity*, 783.
Li, G.-C., Lai, R., D'amour, A., Doolin, D. M., Sun, Y., Torvik, V. I., . . . Fleming, L. (2014). Disambiguation and co-authorship networks of the U.S. patent inventor database (1975–2010). *Research Policy, 43*(6), 941-955. doi:10.1016/j.respol.2014.01.012
Lundvall, B.-Å. (1992). User-producer relationships, national systems of innovation and internationalisation. In *National systems of innovation: Towards a theory of innovation and interactive learning* (pp. 45-67): Frances Pinter Publishers Ltd.
Mansfield, E. (1972). Contribution of R&D to economic growth in the United States. *Science, 175*(4021), 477-486.
Mansfield, E. (1995). Academic research underlying industrial innovations: sources, characteristics, and financing. *The review of Economics and Statistics*, 55-65.
McMillan, G. S., Narin, F., & Deeds, D. L. (2000). An analysis of the critical role of public science in innovation: the case of biotechnology. *Research policy, 29*(1), 1-8.
Meyer, M. (2000). Does science push technology? Patents citing scientific literature. *Research Policy, 29*(3), 409-434. doi:10.1016/S0048-7333(99)00040-2
Moore, M. (2018). AI investment will hit $232 billion by 2025. In. ITProPortal: KPMG.
Mowery, D. C., Nelson, R. R., Sampat, B. N., & Ziedonis, A. A. (2001). The growth of patenting and licensing by US universities: an assessment of the effects of the Bayh–Dole act of 1980. *Research policy, 30*(1), 99-119.
Narin, F., Hamilton, K. S., & Olivastro, D. (1997). The increasing linkage between US technology and public science. *Research policy, 26*(3), 317-330.
Narin, F., & Noma, E. (1985). Is technology becoming science? *Scientometrics, 7*(3-6), 369-381.
NCBI. (2018). PubMed - about.
Nelson, R. R. (2003). On the uneven evolution of human know-how. *Research Policy, 32*(6), 909-922. doi:10.1016/S0048-7333(02)00093-8
Nelson, R. R., & Rosenberg, N. (1993). Technical innovation and national systems. *National innovation systems: A comparative analysis, 322*.
Newman, M. E. J. (2000a). Models of the Small World. *Journal of Statistical Physics, 101*(3), 819-841. doi:10.1023/A:1026485807148
Newman, M. E. J. (2000b). Who is the best connected scientist? A study of scientific coauthorship networks. doi:10.1103/PhysRevE.64.016132
Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States, 98*(2), 404.
Owen-Smith, J., & Powell, W. (2001). To Patent or Not: Faculty Decisions and Institutional Success at Technology Transfer. *The Journal of Technology Transfer, 26*(1), 99-114. doi:10.1023/A:1007892413701
Rosenberg, N. (1974). Science, invention and economic growth. *The Economic Journal, 84*(333), 90-108.
Shan, G., Lu, Y., Min, B., Qu, W., & Zhang, C. (2016). A MeSH-based text mining method for identifying novel prebiotics. *Medicine, 95*(49), e5585-e5585. doi:10.1097/MD.0000000000005585
Smalheiser, N. R., & Torvik, V. I. (2009). Author name disambiguation. *Annual Review of Information Science and Technology, 43*, 287-313.
Stephan, P. E. (1996). The economics of science. *Journal of Economic literature, 34*(3), 1199-1235.

Stevens, A. (2004). The Enactment of Bayh–Dole. *The Journal of Technology Transfer, 29*(1), 93-99. doi:10.1023/B:JOTT.0000011183.40867.52

Sveikauskas, L. (1981). Technological inputs and multifactor productivity growth. *The Review of Economics and Statistics*, 275-282.

Thompson, M., & Polanyi, M. (1960). Personal Knowledge. In (Vol. 69, pp. 111).

Tijssen, R. J. (2001). Global and domestic utilization of industrial relevant science: patent citation analysis of science–technology interactions and knowledge flows. *Research Policy, 30*(1), 35-54.

Toivanen, O., & Väänänen, L. (2012). Returns to Inventors. *The Review of Economics and Statistics, 94*(4), 1173-1190. doi:10.1162/REST_a_00269

Torvik, V. I., Weeber, M., Swanson, D. R., & Smalheiser, N. R. (2005). A probabilistic similarity metric for Medline records: A model for author name disambiguation. *Journal of the American Society for Information Science and Technology, 56*(2), 140-158. doi:10.1002/asi.20105

Trajtenberg, M. (1990). A Penny for Your Quotes: Patent Citations and the Value of. *The Rand Journal of Economics, 21*(1), 172.

Tuomela, M. S., Fegley, B. D., & Torvik, V. I. (2016). Introducing the Author-ity Exporter, and a case study of geo-temporal movement of authors.

USPTO. (2007). Common Application Format.

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science (New York, N.Y.), 342*(6157), 468. doi:10.1126/science.1240474

Veugelers, R., & Wang, J. (2015). Novel science for industry? *MSI Working paper*.

Wagner, C. S., & Leydesdorff, L. (2005). Network structure, self-organization, and the growth of international collaboration in science. *Research Policy, 34*(10), 1608-1618. doi:10.1016/j.respol.2005.08.002

Wang, G., & Guan, J. (2011). Measuring science–technology interactions using patent citations and author-inventor links: an exploration analysis from Chinese nanotechnology. *An Interdisciplinary Forum for Nanoscale Science and Technology, 13*(12), 6245-6262. doi:10.1007/s11051-011-0549-y

Wang, J., Veugelers, R., & Stephan, P. (2015). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *MSI Working paper*.

Zander, U., & Kogut, B. (1995). Knowledge and the speed of the transfer and imitation of organizational capabilities: An empirical test. *Organization Science, 6*(1), 76. doi:10.1287/orsc.6.1.76

Zucker, L. G., Darby, M. R., & Armstrong, J. S. (2002). Commercializing Knowledge: University Science, Knowledge Capture, and Firm Performance in Biotechnology. *Management Science, 48*(1), 138-153. doi:10.1287/mnsc.48.1.138.14274

Zucker, L. G., Darby, M. R., & Brewer, M. B. (1998). Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises. *The American Economic Review, 88*(1), 290-306.

Appendix 1

Hypothese 1[4]

Table B1: descriptive statistics H1

	Mean	Std. Deviation	N
FWD_CIT_LOG	1,1549	,50662	152098
science_link	,46	,498	152098
active_au	,55	,498	152098
own_science	,16	,371	152098
SciLink_Aut	,36	,479	152098
SciLink_Aut_Own	,16	,366	152098
univ_patent	,11	,331	152098
single_subclass	,09	,284	152098
RECTECH_LOG	,7752	,39407	152098
SUBCLASS_LOG	,5972	,32251	152098
PRIORART_LOG	1,0501	,41186	152098
INV_CNT_LOG	,3669	,27865	152098
TECH_MEAN_LOG	1,6079	,72330	152098
COUPLING_LOG	-1,4933	1,23688	152098

Table B2: coefficients table H1

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	,185	,239		,773	,439
	science_link	,106	,003	,104	36,636	,000
	active_au	,034	,003	,033	12,685	,000
	own_science	,055	,004	,040	15,413	,000
	univ_patent	,036	,004	,024	10,214	,000
	single_subclass	-,017	,005	-,010	-3,462	,001
	RECTECH_LOG	,234	,004	,182	62,997	,000
	SUBCLASS_LOG	,042	,009	,027	4,706	,000
	PRIORART_LOG	,254	,003	,207	80,977	,000
	INV_CNT_LOG	,121	,004	,067	28,640	,000
	TECH_MEAN_LOG	,358	,003	,511	105,296	,000
	COUPLING_LOG	-,052	,002	-,127	-21,784	,000

[4] Note that in the coefficients table, dummy variables accounting for 3-digit mainclass and appyear are hidden to save space. Complete output can be found in file: *'Output Timpe Callebaut.spv'*

Table B3: Goodness of fit negative binomial H1

	Value	df	Value/df
Deviance	177474,498	152093	1,167
Scaled Deviance	177474,498	152093	
Pearson Chi-Square	403713,534	152093	2,654
Scaled Pearson Chi-Square	403713,534	152093	
Log Likelihood[b]	-655846,039		
Akaike's Information Criterion (AIC)	1311702,077		
Finite Sample Corrected AIC (AICC)	1311702,078		
Bayesian Information Criterion (BIC)	1311751,739		
Consistent AIC (CAIC)	1311756,739		

Dependent Variable: fwd_cit_5y
Model: (Intercept), science_link, active_au, own_science

Table B4: Omnibus test negative binomial H1

Ratio Chi-Square	df	Sig.
1414,332	3	,000

Dependent Variable: fwd_cit_5y
Model: (Intercept), science_link, active_au, own_science

Table B5: Parameter Estimates negative binomial H1

			95% Wald Confidence Interval		Hypothesis Test		
Parameter	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	3,203	,0068	3,190	3,217	220444,285	1	,000
science_link	,232	,0112	,210	,254	427,066	1	,000
active_au	,044	,0103	,023	,064	18,008	1	,000
own_science	-,072	,0185	-,108	-,036	15,155	1	,000
(Scale)	1[a]						
(Negative binomial)	1,347	,0045	1,339	1,356			

Dependent Variable: fwd_cit_5y
Model: (Intercept), science_link, active_au, own_science

Appendix 2

Hypothese 2

Table B6: Descriptive statistics H2

	Mean	Std. Deviation	N
FWD_CIT_LOG	1,1772	,52559	69962
nov_snpr_top5	,37	,484	69962
OWN_NOVTOP5	,1846	,38795	69962
univ_patent	,20	,428	69962
single_subclass	,07	,249	69962
RECTECH_LOG	,6323	,39130	69962
SUBCLASS_LOG	,6636	,32036	69962
PRIORART_LOG	1,0468	,44081	69962
INV_CNT_LOG	,4220	,26755	69962
TECH_MEAN_LOG	1,3920	,62229	69962
COUPLING_LOG	-1,6900	1,19292	69962

Table B7: Coefficients H2

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	,452	,307		1,471	,141
	nov_snpr_top5	,056	,004	,052	12,738	,000
	OWN_NOVTOP5	,043	,006	,031	7,584	,000
	univ_patent	,032	,004	,026	7,975	,000
	single_subclass	-,014	,008	-,006	-1,600	,110
	RECTECH_LOG	,152	,006	,113	25,075	,000
	SUBCLASS_LOG	,038	,012	,023	3,118	,002
	PRIORART_LOG	,284	,005	,238	52,702	,000
	INV_CNT_LOG	,101	,006	,052	16,037	,000
	TECH_MEAN_LOG	,379	,006	,448	68,066	,000
	COUPLING_LOG	-,048	,003	-,108	-14,335	,000

Table B8: Goodness of fit negative binomial H2

	Value	df	Value/df
Deviance	177467,789	152094	1,167
Scaled Deviance	177467,789	152094	
Pearson Chi-Square	397397,085	152094	2,613
Scaled Pearson Chi-Square	397397,085	152094	
Log Likelihood[b]	-655801,891		
Akaike's Information Criterion (AIC)	1311611,782		
Finite Sample Corrected AIC (AICC)	1311611,783		
Bayesian Information Criterion (BIC)	1311651,511		
Consistent AIC (CAIC)	1311655,511		

Dependent Variable: fwd_cit_5y
Model: (Intercept), nov_snpr_top5, OWN_NOVTOP5

Table B9: Omnibus Test negative binomial H2

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
1502,627	2	,000

Dependent Variable: fwd_cit_5y
Model: (Intercept), nov_snpr_top5, OWN_NOVTOP5

Parameter Estimates

			95% Wald Confidence Interval		Hypothesis Test		
Parameter	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	3,270	,0051	3,260	3,280	414275,573	1	,000
nov_snpr_top5	,310	,0180	,274	,345	294,923	1	,000
OWN_NOVTOP5	-,020	,0287	-,076	,037	,467	1	,495
(Scale)	1[a]						
(Negative binomial)	1,347	,0045	1,338	1,355			

Dependent Variable: fwd_cit_5y
Model: (Intercept), nov_snpr_top5, OWN_NOVTOP5

Press Release

From research to invention – the importance of scientist-inventor ties

When scientists and inventors work together towards a patent, high value inventions arise – recent research at KU Leuven finds. The strength of this tie impacts the value of the invention. The inclusion of author-inventors is even more influential in determining the usefulness of the invention when dealing with highly novel research.

ANTWERP, BELGIUM – Since long, the relation between science and industry has been shown to be instrumental in technological advance, dating back to research from the 70s. In the following decades, research aimed to provide context for the relation – initially from an institutional level, later from an individual level. In recent years, literature has evolved to explaining intricate links between scientist and inventor.

The latest research has determined there to be a relation to the value of the invention and the strength of the scientist-inventor tie. When both parties work closer together, the probability and degree of a high value invention increases. The value of the invention is measured by the number of forward citations the patent receives in a five year window. When the inventor includes a reference to a scientific publication, the number of forward citations increase with 10,6%, research shows. When one of the inventors has recently published scientific material, it increases with another 3.4%. Finally, when the inventor cites scientific research that they have published themselves, the number of forward citations increase with an additional 5.5%.

When it comes to patents citing highly novel scientific publications, the inclusion of so-called author-inventors proves to be instrumental in obtaining high value inventions. It was discovered that 4.3% of the initial 5.5% increase (or 77% in relative terms) in forward citations was owed to the inclusion of author-inventors. This cemented the theory that tacit knowledge transfer is far more important in highly novel research.

The research was performed by analyzing 176,830 patent application between 1995 and 2005 and was part of the researcher's master's thesis. The results provide a more nuanced view of the different actors in the science-to-industry process, which has the potential to lead to a more efficient transfer of knowledge. The lead time between scientific breakthroughs and technological advancements can be shortened significantly.

FACULTY OF ECONOMICS AND BUSINESS
CAMPUS CAROLUS ANTWERP
KORTE NIEUWSTRAAT 33
2000 ANTWERP
PHONE. + 32 3 201 18 40
FEB.ANTWERPEN@KULEUVEN.BE