

CREA Discussion Paper 2015-05

Economics

Center for Research in Economics and Management
University of Luxembourg

Individual versus institutional ownership of university-discovered inventions

available online : http://www.fr.uni.lu/recherche/fdef/crea/publications2/discussion_papers

Dirk Czarnitzki, KU Leuven, Belgium,
Thorsten Doherr, CREA, Université du Luxembourg,
Katrin Hussinger, CREA, Université du Luxembourg
Paula Schliesser, KU Leuven, Belgium
Andrew A. Toole, Economic Research Service, Washington, USA

April, 2015

For editorial correspondence, please contact: crea@uni.lu
University of Luxembourg
Faculty of Law, Economics and Finance
162A, avenue de la Faiencerie
L-1511 Luxembourg

Individual versus institutional ownership of university-discovered inventions

Dirk Czarnitzki ^{a,b}, Thorsten Doherr ^{b,c}, Katrin Hussinger ^{c,b,a},
Paula Schliessler ^{a,b} and Andrew A. Toole ^{d,b}

- a) *KU Leuven, Dept. of Managerial Economics, Strategy and Innovation, Leuven, Belgium*
- b) *Centre for European Economic Research (ZEW), Mannheim, Germany*
- c) *University of Luxembourg, Luxembourg*
- d) *Economic Research Service, US Dept. of Agriculture, Washington D.C., United States*

April 2015

Abstract

We examine how the ownership of intellectual property rights influences patenting of university-discovered inventions. In 2002, Germany transferred patent rights from faculty members to their universities. To identify the effect on the volume of patenting, we exploit the researcher-level exogeneity of the 2002 policy change using a novel researcher-level panel database that includes a control group not affected by the law change. For professors who had existing industry connections, the policy decreased patenting, but for those without prior industry connections, it increased patenting. Overall, fewer university inventions were patented following the shift from inventor to institutional ownership.

Keywords: Intellectual property, patents, technology transfer, policy evaluation

JEL – Classification: O34, O38

Acknowledgements: We are grateful for the funding of this research project by the Centre for European Economic Research (ZEW) within the research program "Strengthening Efficiency and Competitiveness in the European Knowledge Economies" (SEEK), and the audience and discussants at EARIE 2014, IIOC 2014, the 2014 Technology Transfer Society Conference, the 2014 Asia Pacific Innovation Conference, the BRICK Workshop on the Organization, Economics and Policy of Scientific Research 2014, TIE Tagung Munich 2014, ZEW PhD Workshop 2014, KOF Research Seminar in Zürich 2014, Department Seminar at the University of Luxembourg 2014, 'Betriebswirtschaftliches Forschungskolloquium' at University of Trier, Wharton Technology and Innovation Conference 2015. We particularly thank Thomas Astebro, Olof Ejeremo, Carolin Haeussler, David Hsu, Jeongsik Lee, Georg Licht, Fabio Montobbio, Henry Saueremann, Christian Rammer and Otto Toivanen for helpful comments. The views expressed in this article are the authors' and do not necessarily represent the views of the United States Department of Agriculture.

1 Introduction

Intellectual property (IP) policies are among the most powerful instruments shaping the incentives that drive the discovery and commercialization of knowledge. For U.S. academic institutions the Bayh-Dole Act of 1980 is perhaps the most influential and far-reaching of these IP policies. The legislation facilitated private institutional ownership of inventions discovered by researchers who were supported by federal funds. Many observers credit the Bayh-Dole Act with spurring university patenting and licensing that, in turn, stimulated innovation and entrepreneurship (The Economist 2002; OECD 2003; Stevens 2004). Based on this perceived success, the Bayh-Dole Act has become a model of university IP policy that is being debated and emulated in many countries around the world including Germany, Denmark, Japan, China, and others (OECD 2003; Mowery and Sampat 2005; So et al. 2008).

The key component of the Bayh-Dole model is granting the university, not the inventor, ownership rights to patentable inventions discovered using public research funds (Crespi et al. 2006; Geuna and Nesta 2006; Kenney and Patton 2009). However, the incentive effects on academic inventors of university versus individual ownership are not well understood. In a theoretical contribution, Hellmann (2007) found that university ownership is efficient when inventors must search for a commercial partner as long as the cost of search is higher for inventors than for the university. Using survey and case study evidence, Litan et al. (2007) and Kenney and Patton (2009) argued that conflicting objectives and excessive bureaucracy make institutional ownership ineffective and suggest an individual ownership system may be superior. Due to a paucity of evidence, however, the U.S. National Research Council recently concluded that “arguments for superiority of an inventor-driven system of technology transfer are largely conjectural” (NRC 2010).

Our analysis uses the framework of Pakes and Griliches (1984) and a quasi-experimental research design to provide the first systematic evidence on how intellectual property rights impact patenting of university-discovered inventions. We examine a fundamental change in German patent law from individual to institutional ownership. Prior to 2002, university professors and researchers had exclusive intellectual property rights to their inventions. This “Professor’s Privilege” allowed university researchers to decide whether or not to patent and how to commercialize their discoveries, even if the underlying research was supported by public funds. After 2002, universities were granted the intellectual property rights to all inventions made by their employees and this shifted the decision to patent from the researchers to the universities. The policy goal was to increase patenting of university-invented technologies which is often used as a surrogate indicator of successful university technology transfer.

By changing the agent who makes the patenting decision, the abolishment of Professor’s Privilege caused a “regime shift” that substituted institutional benefit and cost schedules for those of the individual inventors. The net effect on the volume of patenting depends primarily on the relative costs between the regimes. To identify how the regime shift affected patenting, we exploit the researcher-level exogeneity of the 2002 abolishment of Professor’s Privilege along with the institutional structure of the German research system in which universities and other public research organizations (PROs) co-exist. PRO researchers were not affected by the ownership change and serve as a control group. We use a difference-in-difference methodology and control for the arrival of new patentable discoveries using publications and peer-to-peer matching.

Our analysis shows that fewer university inventions were patented following the 2002 regime shift. For a given discovery, the schedule of benefits to institutional owners, who are the post-change patent decision makers, is lower because the university became an additional party in the negotiations over the split of expected revenues. This partly explains why fewer inventions qualified for patent protection following the regime shift. However, the effect on expected revenues can be offset if institutional costs (broadly conceived) are sufficiently lower than those faced by individual researchers (Hellmann 2007). Our results show that institutional patenting costs were lower for the subset of university inventors who did not have relationships with industry partners prior to the policy change. For those individuals, patenting increased. But, the data also show that most German patenting professors had prior industry relationships. Post-change institutional costs were not low enough to offset the revenue effect for this group. Our results highlight the critical importance of understanding the nature and strength of faculty-industry relationships before undertaking policy initiatives intended to foster technology transfer.

The remainder of this paper is structured as follows. Section 2 summarizes the background and implementation of the law change in Germany. Section 3 describes the Pakes and Griliches (1984) framework and develops our hypotheses. Section 4 presents the empirical approach, the data collection strategy and provides descriptive statistics. Section 5 shows the econometric results, and robustness checks are presented in Section 6. The final section 7 concludes with a discussion of the implications for policy.

2 The regime change: from inventor to university ownership

In February 2002, the German Federal Government launched a comprehensive new program called “Knowledge Creates Markets” to stimulate technology transfer from

universities to private industry for innovation and economic growth¹. The program was largely a reaction to the “European paradox” (European commission 1995). At that time, policymakers believed that Germany had one of the world’s leading scientific research enterprises, but was lagging the United States in terms of technology transfer and commercialization. The new program addressed a wide spectrum of science-industry interactions including processes and guidelines governing knowledge transfer, science-based spin-offs, collaboration, and the exploitation of scientific knowledge in the private sector. The abolishment of Professor’s Privilege was one of the most significant changes from both a legal and cultural perspective. Professor’s Privilege originated from Article 5 of the German constitution that protects the freedom of science and research. The new program repealed Clause 42 of the German employee invention law that had granted university researchers - as the only occupational group in Germany - the privilege to retain the ownership rights to their inventions that otherwise rest with the employer².

Under the new law, German university researchers are required to cull their research findings for inventions and report any inventions to the university – unless the researcher decides to keep his or her inventions secret by not publishing or patenting. The university has four months to consider any submitted inventions for patenting. If the university does not claim the invention, the rights to pursue patenting and commercialization are returned to the researcher. If the university does claim the invention, the inventor receives at least 30% of the revenues from successful commercialization, but nothing otherwise.

Furthermore, the university handles the patenting process and pays all related expenses

¹ Bundesministerium für Bildung und Forschung and Bundesministerium für Wirtschaft und Technologie (2001), Wissen schafft Märkte - Aktionsprogramm der Bundesregierung.

² Gesetz über Arbeitnehmererfindungen in der im Bundesgesetzblatt Teil III, Gliederungsnummer 422-1, veröffentlichten bereinigten Fassung, das zuletzt durch Artikel 7 des Gesetzes vom 31. Juli 2009 (BGBl. I S. 2521) geändert worden ist.

such as processing fees, translation costs and legal expenses. University researchers retain the right to disclose the invention through publication two months after submitting the invention to the university. Prior contractual agreements with third parties also remained valid during a prescribed transition period.³

At the time of the law change, German universities had little experience undertaking technology transfer activities, and only a few universities maintained professionally managed technology transfer offices (TTOs) (cf. e.g. Schmoch et al., 2000). Therefore the government decided to support the commercialization activities by establishing regional patent valorization agencies (PVAs), which was supported with a budget of 46.2 million EUR to be used before the end of 2004 (Kilger und Bartenbach 2002). Universities were free to choose whether to use the PVAs' services or not. To date, 29 PVAs serve different regional university networks and employ experts specialized in these universities' research areas. The PVAs support the entire process from screening inventions, finding industry partners, and determining fruitful commercialization paths. They are also supposed to promote collaboration between their member universities and industry.

To date, a handful of prior studies have examined the effects of abolishing Professor's Privilege on patenting rates and ownership patterns in Germany. Schmoch (2007) found that the number of university-owned patents increased. Based on inventor lists, his data also suggested the most active faculty inventors were discouraged by the abolishment of Professor's Privilege and that non-patenting professors were encouraged, which suggests the law changed the mix of inventors. In a follow-up study, Cuntz et al. (2012) showed that the share of university-owned inventions increased after 2002 while the share of

³ Contracts made before July 18th 2001 were to be treated under the old law until February 2003 (Gesetz über Arbeitnehmererfindungen, § 43 ArbNErfG).

individually or industry- owned university inventions decreased. Von Proff et al. (2012) found that the policy change did not increase university-invented patents. They also suggested an ownership shift from individual and firm-owned patents to universities.⁴ Our analysis extends this work by combining an established economic framework with a stronger research design and a more comprehensive researcher-level database allowing the identification of causal effects of the law change.

3 Economic framework and hypotheses

In economic models, patents reflect the combined influence of an agent's propensity to patent and the arrival of new knowledge through the agent's inventive process.

$$(1) \quad (\textit{patents})_{it} = (\textit{propensity to patent})_{it} \cdot (\textit{new knowledge})_{it}$$

Pakes and Griliches (1984) called this relation the patent indicator function. The propensity to patent can change due to legal or economic conditions that affect the expected benefits and costs of having a patent. It captures the decision to patent. In equation (1), increments to knowledge reflect investments into discovery, which Pakes and Griliches summarized as the "knowledge production function." Their analysis focused on the relationship between new knowledge and the volume of patenting, holding the propensity to patent constant. In this paper, we focus on how the volume of patents responds to changes in the propensity to patent, holding increments to knowledge constant.^{5,6}

⁴ Other studies consider the effects of patent ownership rights in other European countries. Examples are Valentin and Jensen (2007), Lissoni et al. (2009, 2013), Della Malva et al. (2013). For a broader discussion of academic patenting in Europe see Lissoni (2013).

⁵ A substantial literature has emerged that examines how commercial incentives influence the rate, direction, and disclosure of academic research. This literature focuses on the knowledge production function component of equation (1). Some references include: Jensen and Thursby (2001, 2004); Banal-Estanol and Macho-Stadler (2010); Thursby et al. (2007); Lach and Schankerman (2008); Dechenaux et al. (2009); Azoulay et al. (2007, 2009), Czarnitzki et al. (2011, 2014).

Germany's abolishment of Professor's Privilege exogenously changed the agent responsible for the decision to patent university-discovered inventions. In terms of equation (1), the law transferred the propensity to patent from the faculty inventor to the university. Under the former Professor Privilege system, faculty inventors would apply for patents on their discoveries when the expected benefits of patent protection were greater than the costs. Since 2002, faculty members no longer make this choice, but instead must disclose any inventions to the university. The university, perhaps with the PVA, decides to apply for a patent based on its assessment of expected benefits and costs. Consequently, the effect of revoking Professor's Privilege on the volume of patents depends on how the expected benefit and cost schedules shift due to the regime change from the individual faculty inventor to the university.

For any set of discoveries, the schedule of expected benefits considered by the university after the regime change is lower than the schedule of benefits faced by any faculty member prior to the abolishment of Professor's Privilege. After the policy change, the share of revenue appropriable by the university is limited by three-way bargaining between the university, the faculty member, and the licensee company. Under reasonable assumptions about bargaining power and recognizing that the university cannot increase the market value of the discovery, the university will capture a smaller share of the expected revenue stream in three-way bargaining than the faculty member would under two-way bargaining (Frank et al. 2007; Hellmann 2007).⁷ If the university and faculty cost

⁶ We recognize the regime shift could have an indirect influence on patenting through the knowledge production function; however, proper analysis of this effect would require a separate model focusing on new knowledge (i.e. publications) instead of patents.

⁷ Under Professor's Privilege, the faculty member also had a stronger bargaining position for obtaining non-pecuniary benefits associated with collaborative research and technology development. These non-pecuniary benefits would further reduce the university's benefit schedule relative to the faculty member.

schedules were the same, the reduction in benefits after abolishment of Professor's Privilege would lead to fewer patents. Put simply, the policy change would decrease the propensity to patent.

At that time, however, policy makers believed the cost schedules faced by universities would be lower than those faced by individual faculty members. They interpreted the small share of university-owned patents in Germany prior to 2002 as evidence that individual researchers could not afford to undertake the costly and time-consuming process of applying for a patent and pursuing potential licensees (Becher et al. 1996). If the costs of patenting for universities were sufficiently lower, the volume of university inventions receiving patents could increase. So, the net effect of the regime shift on the volume of patenting depends on the costs of the universities compared to the pre-policy costs of faculty inventors.

It is important to remember that the propensity to patent incorporates the benefits and costs of patenting that are expected upon commercialization. The expected revenues from commercialization are compared to the expected costs of achieving commercialization both with and without patent protection. The relevant concept of costs is broader than simply the patent application fees and legal fees. It also includes costs from searching for an industry partner for commercialization, development costs, and so forth. While these costs may be close to homogeneous across universities in the post-policy change period, they are likely to be heterogeneous within the population of university inventors before the abolishment of Professor's Privilege.

We can identify two groups in the population of university inventors who faced significantly different costs of patenting under Professor's Privilege. The first group consists

of university inventors who had relationships with one or more industry partners. These individuals already paid the costs of searching for licensee companies and negotiating their pecuniary and non-pecuniary benefits. In these relationships, industry partners would typically pay the application and legal fees, manage the development process and commercialize the product or service. For this group of “low cost” university inventors, the regime shift to institutional ownership almost surely led to a higher cost schedule as the university, possibly through the PVA, had to renegotiate established relationships (Frank et al. 2007; Kilger and Bartenbach 2002). For this group, we expect the regime shift in the propensity to patent led to a lower benefits schedule and a higher cost schedule. Our first hypothesis is:

H1: Faculty members who had established connections to industry partners experienced a decrease in the volume of patenting, ceteris paribus.

The second group consists of university inventors who did not have a relationship with an industry partner. These individuals obtained a patent, but still needed to search for a licensee company and negotiate pecuniary and non-pecuniary benefits. For this group of “high cost” university inventors, the university may have a considerable cost advantage. The cost advantage could stem from many sources. Hellmann (2007) postulates that a TTO (or PVA) may have a comparative advantage in identifying potential industry partners due to the efforts of specialized managers or, on the licensee’s side, a single institutional source may make it easier to find university discoveries (e.g. Debackere and Veugelers 2005; Siegel et al. 2003). For this group, we expect the post-policy cost schedule shifted downward more than the post-policy benefits schedule. Our second hypothesis is:

H2: Faculty members who did not have established connections with industry partners experienced an increase in the volume of patenting, ceteris paribus.

With cost heterogeneity in the population of university inventors, the net effect of the policy change depends on the share of inventors of each type. If the pre-policy inventor population was predominantly low cost faculty inventors, then the net effect of the policy would be to reduce the volume of patents. Whereas, the policy would increase the volume of patenting of university-discovered inventions if faculty inventors were mostly high cost. As discussed in the data section, most patenting professors were in the low cost group before the policy change.

4 Empirical model and data

4.1 Identification Strategy and Estimation Approach

The German policy change provides a unique opportunity to separate the influence of the propensity to patent from the influence of new knowledge on the volume of patenting. The abolishment of Professor's Privilege was an exogenous "shock" to the propensity to patent university inventions. As seen in equation (1), once new knowledge is held constant, this exogenous variation will identify the effect of the propensity to patent on the volume of patenting. In the literature on academic research, publications are the accepted standard for measuring knowledge production. The database compiled for this analysis includes complete publication histories for university inventors and their peers in non-university, public research organizations (PROs) such as the Max Planck, Fraunhofer, and Helmholtz institutes as well as other federal and state research institutions.⁸

⁸ Major research institutions in Germany are not only universities but other public research institutions that have many branches in a variety of different scientific disciplines. For instance, the Fraunhofer Society has 59 institutes in Germany with about 17,000 employees, the Max Planck Society has 76 institutes with about 12,000 employees. The Leibniz Association employs 16,100 people in 86 research centers. The Helmholtz Association has about 30,000 employees in 16 research centers.

We identify the policy effect using a difference-in-difference (DiD) research design with university inventors as the treatment group and PRO researchers as the control group. Like university professors, PRO researchers conduct academic research at publicly funded institutions in Germany. They work in similar academic fields and experience similar changes in research opportunities that affect the discovery of new knowledge. But unlike university professors, PRO researchers did not have Professor’s Privilege and the patent rights to their inventions were always owned by the institution. To further control for changes in research opportunities, we use peer-to-peer matching between university faculty members and PRO researchers based on characteristics such as publications, scientific discipline, and career age before undertaking DiD estimation. Our DiD setup also accounts for common macroeconomic trends and individual-specific unobserved effects that capture an academic inventor’s “taste” for patenting and commercialization.

For the population of German academic inventors, the DiD model takes the following form:

$$(2) \quad PAT_{it} = \beta_0 + \beta_1(Prof_i \cdot NewPolicy_t) + \beta_2(CareerAge)_{it} + \beta_3(CareerAge^2)_{it} + \beta_4(3yrAvgPubs)_{i,t-1} + \delta_i + \gamma_t + u_{it}$$

where PAT_{it} is the volume of patents by researcher i applied for in year t (i.e. researcher-year observations). The policy effect is captured by the coefficient β_1 of the interaction term $(Prof \cdot NewPolicy)$. $Prof$ is a dummy variable that takes the value of 1 when the inventor is a university professor and 0 when the inventor is a PRO researcher. $NewPolicy$ is a dummy variable that takes the value of 1 following the policy change, 2002 onward, and 0 otherwise. A quadratic specification of career age captures inventor life-cycle effects. We use a three year moving average of past research publications, $(3yrAvgPubs)_{i,t-1}$, to capture the arrival of new knowledge. δ_i is a researcher-level fixed effect and γ_t is a vector

of time dummy variables covering 2-year periods. Note that the professor dummy variable gets absorbed into the researcher fixed effects. Similarly, the new policy dummy variable gets absorbed by the general time trend.

As patent counts take only nonnegative integer values, we use the fixed effects Poisson quasi-maximum likelihood estimator (QMLE). As a member of the linear exponential family of distributions, the Poisson QMLE produces consistent estimates of the population parameters as long as the conditional mean is correctly specified (Gourieroux et al. 1984; Wooldridge 1999). We use robust standard errors to account for any over- or under-dispersion.

4.2 Data and descriptive statistics

As the aim of this research is to examine the effects of abolishing Professor's Privilege on the decision to patent university-discovered inventions, we focus on German academic inventors. This population includes all researchers affiliated with a university or PRO who appeared as an inventor on at least one patent submitted to the German or European Patent Offices between 1978 and 2008. Academic inventors are a subpopulation of all academic researchers in Germany. The broader population includes academic researchers who only published. However, the transfer of patent rights to institutional ownership did not impact these researchers as they never participated in the intellectual property system over the entire time period.

We constructed a researcher-level panel dataset of academic inventors following a multistep procedure, which is summarized in Appendix A. This process yielded a sample

with 3,718 professors and 8,294 PRO researchers.⁹ We defined the study period to extend from 1995 through 2008 so that we observed enough time periods before and after the policy change. For each inventor, our data contains the individual's history of patenting between 1978 and 2008 and the individual's history of publications between 1990 and 2008. Beyond patent and publication characteristics, this information allowed us to calculate each researcher's career age which is used to model quadratic life cycle effects in equation (2). Career age starts when we observe the researcher's first publication or patent application and increases incrementally thereafter to a maximum of 35 years after which we assume the researcher retires. To account for earlier exit, we adopted a 5-year rule that has a researcher leaving the panel if he or she had no patenting or publishing activity for five consecutive years.¹⁰ Researcher industry connections were determined from the patent data. An academic researcher is identified as having an industry connection when he or she is observed as an inventor on a company owned patent. This allows us to distinguish high cost and low cost academic inventors prior to the abolishment of Professor's Privilege and to estimate the model on subsamples to test hypotheses 1 and 2. The estimation sample contains 108,263 researcher-year observations. All of the variables used in the analysis are described in Table in Appendix A.

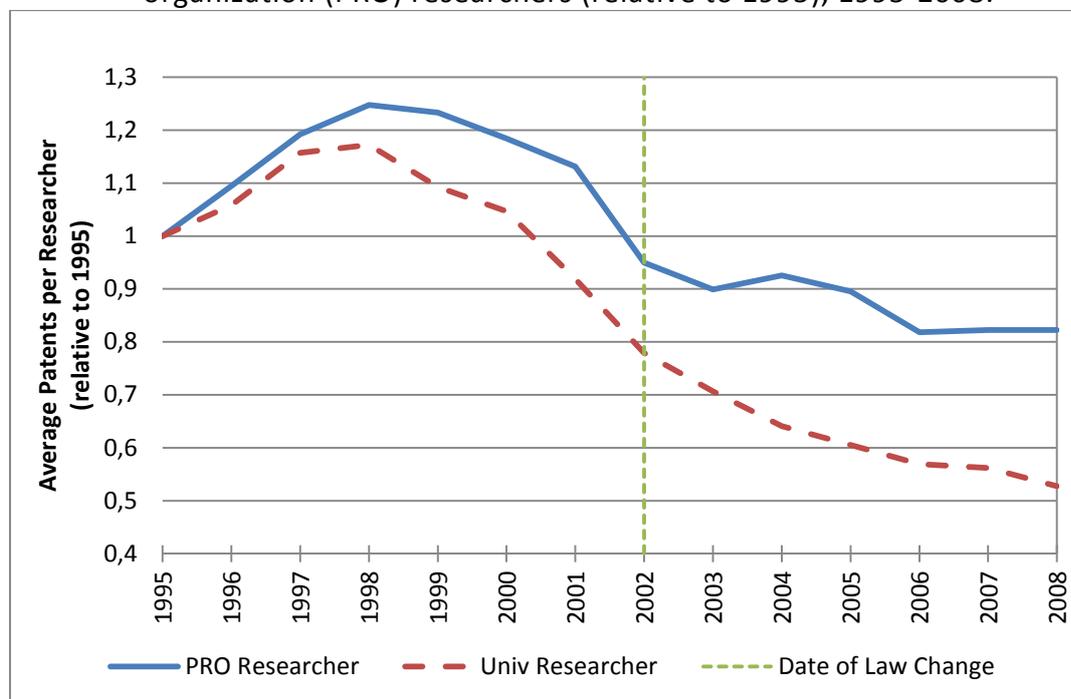
Figure 1 shows the average number of patents per inventor for university and PRO researchers over time. To better compare the trends, annual patents were normalized using 1995 as the reference year (i.e. each data point is relative to 1995). In the years leading up to the policy change, the trends in patents by professors and PRO researchers

⁹ This sample excludes those researchers who were employed at both a PRO and university, as it is not clear which patent regime applied to these researchers.

¹⁰ In section 6 we present an alternative exit rule; however, the results do not change in a meaningful way.

were quite similar. Both series show a peak in 1998 and a downward trend up to 2002. After the abolishment of Professor's Privilege in 2002, the patenting trends diverge with university professors showing a steeper downward trend than PRO researchers. This suggests that abolishing Professor's Privilege led to an overall decrease in the volume of patenting of university-discovered inventions and highlights the importance of using a control group for analyzing the policy change.

Figure 1: Trends in German patenting for university and public research organization (PRO) researchers (relative to 1995), 1995-2008.



Finding a decrease in patents per researcher after 1998 was somewhat surprising because it does not mirror the overall trend in German patent applications over this period. Upon further inquiry, the same pattern for academic patents was found by prior researchers (Cuntz et al., 2012; Schmoch 2007; Von Proff et al. 2012). These authors and others have speculated about the reasons for the decrease. Some suggestions include an increased emphasis on publications in academic performance evaluations, decreased entry into

academic jobs, the end of the New Economy boom, and legal uncertainties surrounding patenting in the field of biotechnology (Cuntz et al. 2012, p.21-22; Schmoch 2007, p. 5-8).

As described in section 3, the overall effect of the policy depends on the composition of university inventors prior to the regime change. If most patenting professors were in the low cost group, the policy would reduce university patenting. The data show that 2,657 (71%) of the university inventors had at least one patent before 2002 and 78% of these inventors had existing industry connections. It is clear that most university inventors were low cost. Among PRO inventors, 5,008 (80%) had patented before the law change and 44% of these inventors had industry connections. The lower percentage of PRO inventors with industry connections probably reflects the institutional ownership system already in place for these researchers.

Table 1 shows descriptive statistics at the researcher-year level for university professors (i.e. the treatment group) and PRO researchers (i.e. the control group) separated into the pre- and post-policy change periods. These groups are further subdivided into those with industry connections in the top portion of the table and those without industry connections in the bottom portion. Looking at academic inventors with industry connections, mean patents by professors declined by 44% after the abolishment of Professor's Privilege while patenting by PRO researchers declined by 27%. Among those without industry connections, mean patents by professors increased 55% after the law change, but only 9% for PRO researchers. These findings are consistent with the hypothesized effects discussed in Section 3. Citation-weighted patents, which partially adjust the raw counts for the "quality" of the inventions, also fell more for professors than PRO researchers among those with industry connections. While the average number of patents by university professors

without industry connections increased by 55%, the citation-weighted patents actually fell by 15%. The differences in career age show that university professors were slightly older than PRO researchers over the whole sample period.

Table 1: Descriptive Statistics for the treatment and control groups
(researcher-year observations)

Prior to law change (1995-2001)					After law change (2002-2008)			
Professors with industry connection								
N = 12508 researcher-years					N = 9141 researcher-years			
Variable	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
# Patents	0.88	2.02	0	64	0.49	1.50	0	28
# Citation-weighted patents	0.67	2.78	0	119	0.27	1.43	0	39
Career age	9.86	6.74	0	34	16.25	6.73	2	35
Avg. publications	2.75	5.51	0	67.33	4.13	6.97	0	67
Control group with industry connection								
N = 13101 researcher-years					N = 9854 researcher-years			
# Patents	1.01	1.98	0	44	0.73	1.70	0	26
# Citation-weighted patents	0.81	2.55	0	55	0.42	1.68	0	41
Career age	8.06	6.02	0	34	14.22	6.43	2	35
Avg. publications	1.21	3.41	0	110.67	2.00	3.95	0	64.67
Professors without industry connection								
N = 6633 researcher-years					N = 8121 researcher-years			
Variable	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
# Patents	0.20	0.59	0	11	0.31	0.84	0	27
# Citation-weighted patents	0.13	0.75	0	15	0.15	0.77	0	19
Career age	5.71	3.89	0	27	9.35	5.92	0	32
Avg. publications	3.03	5.73	0	100.67	3.63	7.07	0	80.67
Control group without industry connection								
N = 19855 researcher-years					N = 29050 researcher-years			
# Patents	0.34	0.76	0	13	0.37	0.93	0	24
# Citation-weighted patents	0.22	0.92	0	16	0.21	1.07	0	61
Career age	4.50	4.06	0	29	7.16	5.53	0	35
Avg. publications	1.12	2.51	0	44	1.32	2.89	0	63.67

Note: Avg. publications are a three-year moving average of publication counts in t-1 for each researcher.

5 Econometric Results

Our baseline results identify the treatment effect of Germany's 2002 policy change that transferred patent ownership rights from inventors to the universities on the decision to patent. Table 2 presents the parameter estimates based on Poisson QMLE with robust

standard errors. The overall treatment effect, which is revealed by the coefficient on $(Prof \cdot NewPolicy)$ in column 2, is negative and statistically significant at the 1% level. This indicates that the overall effect of abolishing Professor's Privilege was to decrease the volume of patents obtained on university-discovered inventions in Germany. It is economically significant as well. Holding the arrival of new knowledge and researcher life cycle effects constant, the coefficient estimate shows the volume of university patents decreased by 18%, on average. At least in part, this result reflects the reduction in benefits appropriable by universities after the abolishment of Professor's Privilege due to three-way bargaining. It would fully describe the effect of the 2002 policy change if university and faculty cost schedules were the same. Turning to the arrival of new knowledge, as captured by a three year moving average of past publications, increases patents by academic inventors with one additional publication boosting expected patents by 14%.

The overall effect, however, masks potential heterogeneous treatment effects due to differences in patent and commercialization costs before the policy change. Even with the reduction in benefits appropriable by the university, the effect of the policy change on the volume of patenting depends on the costs of the university compared to costs of faculty inventors before the transition to institutional ownership. In Section 3, we argued that faculty with prior industry connections were relatively low cost and postulated that the decrease in patent volume due to the policy change would be even larger for this group. As seen in column 3 of Table 2, this hypothesis is supported. In the subsample of academic inventors with industry connections, the expected number of university patents decreased by 26%, holding other factors constant.

For faculty without prior industry connections, we postulated that cost advantages for universities would offset the reduction in benefits and increase patenting. As seen in column 6 of Table 2, treatment effect for this subsample is positive and significant at the 1% level. Holding the arrival of new knowledge and researcher life cycle effects constant, the estimate shows the volume of university patents increased by 39%, on average. For faculty without prior industry connections life cycle effects are statistically stronger while the link between publications and patents is still positive and significant. As seen in the subsample breakout, the overall decrease in patenting of university-discovered inventions reflects the composition of university inventors before the regime change – most inventors had pre-existing connections with industry.

Table 2: Poisson models of patenting output

# Patents	Overall		With industry connection		Without industry connection	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Professor*NewPolicy	-0.184***	(0.053)	-0.262***	(0.067)	0.391***	(0.085)
Career age	-0.028**	(0.014)	-0.030	(0.019)	-0.106***	(0.020)
Career age squared/100	0.002	(0.028)	-0.064*	(0.038)	0.721***	(0.065)
Avg publications	0.028***	(0.005)	0.017***	(0.005)	0.045***	(0.007)
Time dummies (base 1995)						
1996-1997	0.136***	(0.033)	0.160***	(0.039)	0.090	(0.062)
1998-1999	0.210***	(0.052)	0.304***	(0.066)	0.008	(0.086)
2000-2001	0.189**	(0.075)	0.307***	(0.098)	-0.002	(0.113)
2002-2003	0.087	(0.097)	0.184	(0.129)	-0.099	(0.144)
2004-2005	0.094	(0.118)	0.189	(0.156)	-0.117	(0.175)
2006-2007	0.034	(0.139)	0.127	(0.186)	-0.232	(0.203)
2008	-0.068	(0.157)	0.115	(0.210)	-0.446*	(0.228)
# obs.	108,263		44,604		63,659	
# obs. PRO researchers	71,860		22,955		48,905	
# obs. professors	36,403		21,649		14,754	
# obs. Professors after policy change	17,262		9,141		8,121	

Robust standard errors. Significance: * p < 0.1, ** p < 0.05, *** p < 0.01.

Note: Avg. publications are a three-year moving average of publication counts in t-1 for each researcher.

Conditional Difference-in-Difference

One important characteristic of our control group is that they are German academic researchers. Like university professors, these individuals understand the literatures in their disciplines as well as other developments in their fields. Peer-to-peer matching can help control for potential changes in research opportunities. We constructed a matched sample of university professors and PRO researchers by applying caliper matching (caliper threshold = 0.005) to identify the nearest neighbor for each university professor. The inventors were matched based on their career achievements in 1998 (4 years prior to policy change) using their publication count, publication subject field¹¹ and career age. We estimate the DiD specification in equation (2) using observations from 1999 through 2008.

The treatment effects from the abolishment of Professor's Privilege are quite similar in magnitude and significance to those presented in Table 2. The overall treatment effect indicates that patents on university-discovered inventions decreased by 19% instead of 18%, on average. Among those university inventors with prior industry connections, patents decreased by the same magnitude, 26%. The magnitude of the treatment effect for university faculty who were previously high-cost increased by four percentage points and now indicates the policy increased patenting for this group by 43%, on average.

¹¹ The subject fields of the publications have been assigned based on the classification in the ISI Web of Science Citation Index /Science Citation Index. We followed Leydesdorff and Rafols (2009) and defined 18 aggregated publication fields. A researcher has been allocated to one of these aggregated fields by using the field occurring most frequently in his or her publication record.

Table 3: Conditional Difference-in-Difference Poisson models of patenting output

# Patents	Overall		With industry connection		Without industry connection	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Professor*New Policy	-0.19**	0.09	-0.26**	0.11	0.43***	0.13
Time dummies (base 1998-1999)						
2000-2001	-0.10**	0.05	-0.16***	0.05	0.23**	0.11
2002-2003	-0.23***	0.08	-0.34***	0.10	0.18	0.12
2004-2005	-0.31***	0.09	-0.44***	0.11	0.17	0.13
2006-2007	-0.35***	0.10	-0.58***	0.13	0.35***	0.13
2008	-0.38***	0.12	-0.60***	0.16	0.29**	0.14
Observations	33728		18591		15137	

Robust standard errors. Significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

6 Robustness checks

6.1 Citation-weighted patent volume

It is well known that the economic value distribution associated with patents is highly skewed with a very small number of patents accounting for most of the value created through invention. So, even though the German policy change reduced the volume of patents, one might wonder whether the policy change simply eliminated the low value patents and thereby resulted in a smaller quantity of higher quality patents. To address this issue, forward citations are commonly used to weight raw patent counts as a way to partially adjust for the unobserved quality of inventions (Trajtenberg 1990).

Table 4 reports the results from applying the DiD research design to citation-weighted patents. As before, the parameters are estimated using Poisson QMLE with robust standard errors. From column 2, the overall treatment effect from revoking Professors Privilege was to reduce the volume of university citation-weighted patents by 27%, holding the arrival of new knowledge and researcher life cycle effects constant. For university professors who had prior industry connections, university citation-weighted patents fell by 25%, on average.

However, for university professors who did not have prior industry connections, the results are different from those found previously. While the volume of un-weighted patents increased for this group, citation-weighted patents show no significant change. This suggests that while the new policy increased the volume of patenting by professors without industry connections, it did not improve the average quality of these inventions. Among the other covariates, the only notable difference is that new knowledge is no longer significantly related to citation-weight patents among professors with prior industry connections.

Table 4: Poisson models of Citation-weighted patenting output

# Citation-weighted patents	Overall		With industry connection		Without industry connection	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Professor*NewPolicy	-0.274***	(0.086)	-0.254**	(0.104)	0.103	(0.147)
Career age	-0.072***	(0.026)	-0.061*	(0.035)	-0.179***	(0.044)
Career age squared/100	-0.000	(0.045)	-0.052	(0.058)	0.797***	(0.135)
Avg publications	0.014**	(0.007)	0.002	(0.008)	0.026**	(0.011)
Time dummies (base 1995)						
1996-1997	0.111*	(0.065)	0.113	(0.077)	0.097	(0.116)
1998-1999	0.337***	(0.106)	0.373***	(0.130)	0.217	(0.177)
2000-2001	0.099	(0.147)	0.153	(0.184)	-0.019	(0.237)
2002-2003	0.062	(0.195)	-0.003	(0.250)	0.123	(0.308)
2004-2005	0.211	(0.237)	0.134	(0.304)	0.275	(0.372)
2006-2007	0.143	(0.283)	-0.048	(0.356)	0.267	(0.450)
2008	-0.318	(0.318)	-0.389	(0.408)	-0.310	(0.496)
Observations	64,030		32,300		31,730	

Robust standard errors. Significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Note: Avg. publications are a three-year moving average of publication counts in $t-1$ for each researcher.

6.2 Exclusion of pre-policy uncertainty period

As part of our research process, we reviewed the public discussion regarding the abolishment of Professor's Privilege. The possibility of a policy change became public as early as December 1997 when the German Federal Council requested the federal government to review the efficacy and appropriateness of Professor's Privilege. At that time, some policy makers were concerned that only 4% of all German patents originated from universities.¹² As discussed in section 3, they believed professors were not willing or able to invest the time and money for commercialization, but focused instead on publications. After this initial inquiry, Professor's Privilege was debated through March 2001 when the federal government published its action plan for enhanced science-to-industry technology transfer that officially announced the abolishment of the Professor's Privilege. When the final version of the law was published in October 2001, it was clear that Professor's Privilege would be abolished effective February 2002.

To verify that the timing of the policy change does not affect our findings, we exclude this described pre-policy "uncertainty period" from the sample, and compare academic patenting in 1995-1997 (before the law change and before the public discussion has been initiated) with the time period after the law change, 2002-2008. As seen in Table 5, the coefficient magnitudes on the treatment effects are larger. The effect of new knowledge through publications is smaller, but statistically significant across all specifications.

¹² This was discussed in many German newspapers at the time. An example can be found in "Der Spiegel" which is one of the most prominent weekly news magazines in Germany (see <http://www.spiegel.de/wissenschaft/mensch/patentoffensive-bulmahn-will-hochschullehrerprivileg-abschaffen-a-101092.html>). Our data also shows that about 4% of all patents applied for at the German Patent Office and the European Patent Office were university-invented patents. For instance, in 1995 there were 320,000 patents applied for by German inventors at the German Patent Office and the European Patent Office. Out of these, we find 4.7% to be university-inventions. In 2000, there were 460,000 patents out of which 3.3% originated from universities.

Table 5: Poisson models of patenting using only 1995-1997 as pre-treatment time periods

# Patents	Overall		With industry connection		Without industry connection	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Professor*NewPolicy	-0.230***	(0.069)	-0.328***	(0.084)	0.827***	(0.152)
Career age	-0.014	(0.017)	0.012	(0.025)	-0.139***	(0.025)
Career age squared/100	-0.078**	(0.031)	-0.102**	(0.041)	0.531***	(0.069)
Avg. publications	0.030***	(0.006)	0.020***	(0.007)	0.041***	(0.009)
Year dummies (base 1995)						
1996-1997	0.082**	(0.037)	0.084*	(0.046)	0.017	(0.067)
2002-2003	0.245*	(0.128)	0.036	(0.180)	0.549***	(0.179)
2004-2005	0.235	(0.154)	-0.035	(0.216)	0.607***	(0.219)
2006-2007	0.156	(0.181)	-0.166	(0.254)	0.583**	(0.254)
2008	0.055	(0.203)	-0.229	(0.285)	0.468*	(0.284)
Observations	64037		25986		38051	

Robust standard errors. Significance: * p < 0.1, ** p < 0.05, *** p < 0.01.

Note: Avg. publications are a three-year moving average of publication counts in t-1 for each researcher.

6.3 Robustness test on the sample exit rule

For our main analysis we adopted a 5-year rule that has a researcher leaving the panel if he or she had no patenting or publishing activity for five consecutive years. This rule was necessary due to data limitations that prevent us from observing when a researcher retires or leaves academic employment. To verify our results are not driven by this limitation, we imposed a very strict 2-year rule in which researchers are dropped after two consecutive years of inactivity. The results using the strict exit rule are very similar to those found using the 5-year rule (Table 6).

Table 6: Poisson models of patenting using the 2-year exit rule

# Patents	Overall		With industry connection		Without industry connection	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Professor*NewPolicy	-0.172***	0.055	-0.258***	0.070	0.473***	0.085
Career age	0.008	0.015	-0.013	0.020	-0.032	0.021
Career age squared/100	-0.093***	0.033	-0.096**	0.044	0.432***	0.064
Avg. publications	0.020***	0.005	0.011**	0.006	0.033***	0.007
Time dummies (base 1995)						
1996-1997	0.096***	0.034	0.130***	0.041	0.041	0.062
1998-1999	0.212***	0.056	0.327***	0.070	-0.004	0.091
2000-2001	0.188**	0.078	0.322***	0.101	0.004	0.12
2002-2003	0.068	0.103	0.201	0.137	-0.147	0.154
2004-2005	0.079	0.124	0.189	0.163	-0.144	0.187
2006-2007	0.006	0.146	0.118	0.195	-0.300	0.218
2008	-0.078	0.165	0.15	0.221	-0.558**	0.244
Observations	88666		37193		51473	

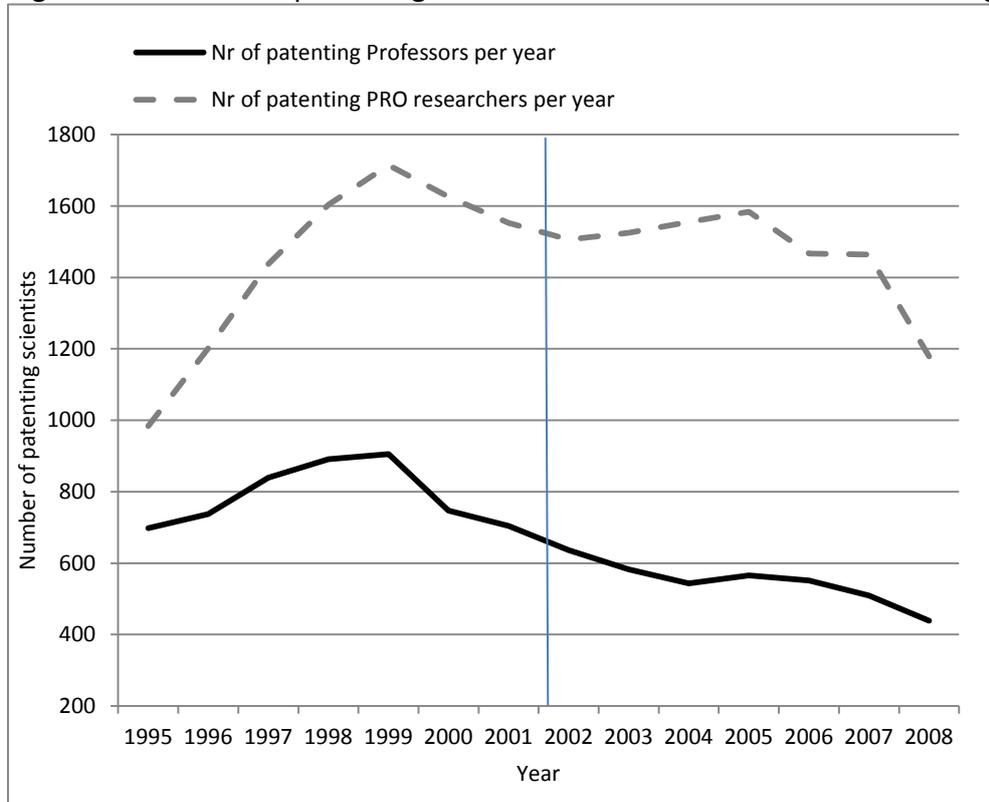
6.4 Number of patenting researchers before and after the law change

The fixed effects regressions presented above estimate the treatment effect of the policy only for scientists that were in the academic system before the policy changed occurred.

This is because of the nature of the fixed effects regressions: the treatment dummy, i.e. the policy change variable only changes from the value zero to one for scientists that were in the sample before 2002. If, however, researchers enter the system after 2002 patent more than earlier cohorts, the fixed effects regressions would not pick this up, as the policy change variable would always be equal to one for these researchers.. If the policy change attracted new entrants into academic patenting, it could have increased the total volume of patents. To check for this possibility, we analyze the trend in the number of patenting researchers before and after the law change in 2002 (see Figure 2). The graph shows that

the number of patenting professors shows a steady decline after 1999. This suggests that the policy change did not attract more professors into patenting.

Figure 2: Number of patenting scientists before and after the law change.



7 Discussion and Conclusion

In this paper we examine how the ownership of patent rights influences the decision to patent in the context of university-discovered inventions. By changing the agent who makes the patenting decision, Germany's abolishment of Professor's Privilege in 2002 caused a regime shift that substituted institutional benefit and cost schedules for those of the individual inventors. Our empirical approach exploits the institutional structure of the German public research system to identify an appropriate control group along with the researcher-level exogeneity of the policy change to implement a difference-in-difference

approach to causal inference. Our analysis shows that fewer university inventions were patented following the 2002 regime shift from inventor to institutional ownership.

The German policy change that abolished Professor's Privilege was based on the presumption that the costs and risks of patenting were so high that professors did not have sufficient incentives to patent their discoveries or pursue commercialization. In retrospect, this presumption appears to be wrong. We find that the treatment effect was heterogeneous among university professors and depended on the costs of the university compared to costs of faculty inventors before the transition to institutional ownership. Post-policy institutional patenting costs were lower for the subset of university inventors who did not have prior relationships with industry partners. For those individuals, patenting increased after the policy change. Yet, most German professors had prior connections with industry partners leading to higher patenting and commercialization costs under institutional ownership. For these professors, patenting decreased substantially.

While these findings reflect the medium-term effects of the law change, it could still be possible that the law change results in higher commercialization in the long-run, that is, when new faculty members enter academe who never experienced the old regime of inventor-ownership. However, trends in the number of patenting researchers until 2008 do not suggest more researchers patented after the law change. On the contrary, the number of patenting professors has declined, at least through 2008.

One possible reason for the miscalculation by German policy makers is a failure to adequately assess the nature and extent of technology transfer and patenting relationships prior to the law change. Informal and formal relationships between university researchers and industry firms had evolved under the Professor's Privilege system. Our results highlight

the critical importance of understanding the nature and strength of faculty-industry relationships before undertaking policy initiatives intended to foster technology transfer.

Our findings provide the strongest evidence to date that an inventor ownership system can produce more university-invented patents, and thereby more technology transfer, than an institutional ownership system. Does this imply that other countries such as the U.S. would increase university technology transfer by adopting an inventor ownership system? Not necessarily. The nature and strength of faculty-industry relationships will differ based on each country's institutions, culture, and historical evolution of networks and trust relationships. Rather than attempting a major policy change as was done in Germany, policymakers in other countries would benefit from a better understanding of current practices. This information could be used to design incremental changes that allow technology transfer processes the flexibility and adaptability needed to fit alternative technologies and markets.

References

- Azoulay, P., Ding, W. and Stuart, T. (2007), The determinants of faculty patenting behavior: demographics or opportunities? *Journal of Economic Behavior & Organization* 63, 599-623.
- Azoulay, P., Ding, W. and Stuart, T. (2009), The impact of academic patenting on the rate, quality and direction of (public) research output, *Journal of Industrial Economics* 57(4), 637-676.
- Banal-Estanol, A. and Macho-Stadler, I. (2010), Scientific and commercial incentives in R&D: research versus development? *Journal of Economics & Management Strategy* 19(1), 185-221.
- Becher, G., Gering, T., Lang, O. and Schmoch, U. (1996), *Patentwesen an Hochschulen: Eine Studie zum Stellenwert gewerblicher Schutzrechte im Technologietransfer Hochschule-Wirtschaft*, BMBF (Hrsg.), Bonn.
- Bundesministerium für Bildung und Forschung (2012), *Bundesbericht Forschung und Innovation*, Bonn.
- Bundesministerium für Bildung und Forschung and Bundesministerium für Wirtschaft und Technologie (2001), *Wissen schafft Märkte - Aktionsprogramm der Bundesregierung*, Berlin.
- Crespi, G., Geuna, A. and Verspagen, B. (2006), University IRPs and knowledge transfer: Is the IPR ownership model more efficient? SPRU Working Paper No. 154, Brighton.
- Cuntz, A., Dauchert, H., Meurer, P. and Philipps, A. (2012), Hochschulpatente zehn Jahre nach Abschaffung des Hochschullehrerprivilegs, Studien zum deutschen Innovationssystem 13-2012, Berlin.
- Czarnitzki, D., Glänzel, W. and Hussinger, K. (2007), Patent and publication activities of German professors: an empirical assessment of their co-activity, *Research Evaluation* 16(4), 311-319.
- Czarnitzki, D., Glänzel, W. and Hussinger, K. (2009), Heterogeneity of patenting activity and its implications for scientific research, *Research Policy* 38, 26-34.

- Czarnitzki, D., Grimpe, C. and Toole, A. (2014), Delay and secrecy: does industry sponsorship jeopardize disclosure of academic research? *Industrial and Corporate Change*, forthcoming.
- Czarnitzki, D., Hussinger, K. and Schneider, C. (2011), Commercializing academic research: the quality of faculty patenting, *Industrial and Corporate Change* 20(5), 1403-1437.
- Debackere, K. and Veugelers, R. (2005), The role of academic technology transfer organizations in improving industry science links, *Research Policy* 34(3), 321-342.
- Dechenaux, E., Thursby, M., and Thursby, J. (2009), Shirking, sharing risk and shelving: the role of university license contracts, *International Journal of Industrial Organization* 27, 80-91.
- Della Malva, A., Lissoni, F., Llerena, P. (2013) Institutional change and academic patenting: French universities and the Innovation Act of 1999, *Journal of Evolutionary Economics* 23, 211–239.
- DER SPIEGEL (2000), Patentoffensive: Bulmahn will "Hochschullehrerprivileg" abschaffen, 3-11-2000, <http://www.spiegel.de/wissenschaft/mensch/patentoffensive-bulmahn-will-hochschullehrerprivileg-abschaffen-a-101092.html>
- European Commission (1995), Green paper on innovation, Brussels.
- Frank, A., Meyer-Guckel, V. and Schneider, C. (2007), Innovationsfaktor Kooperation. Bericht des Stifterverbands zur Zusammenarbeit zwischen Unternehmen und Hochschulen. Gesetz über Arbeitnehmererfindungen (ArbnErfG) in der im Bundesgesetzblatt Teil III, Gliederungsnummer 422-1, veröffentlichten bereinigten Fassung, das zuletzt durch Artikel 7 des Gesetzes vom 31. Juli 2009 (BGBl. I S. 2521) geändert worden ist.
- Geuna, A. and Nesta, L.J.J. (2006), University Patenting and its Effects on Academic Research: The Emerging European Evidence, *Research Policy* 35, 790-807.
- Gouriéroux, C., Montfort, A. and Trognon, A. (1984), Pseudo maximum likelihood methods: application to poisson models, *Econometrica*, 52(3), 701-721.
- Hellmann, T. (2007), The role of patents for bridging the science to market gap, *Journal of Economic Behavior & Organization* 63, 624-647.

- Jensen, R. and Thursby, M. (2004), Patent licensing and the research university, NBER Working Paper 10758, Cambridge.
- Jensen, R. and Thursby, M., (2001), Proofs and prototypes for sale: the licensing of university inventions, *American Economic Review* 91(1), 240-259.
- Kenney, M. and Patton, D. (2009), Reconsidering the Bayh-Dole act and the current university invention ownership model, *Research Policy* 38, 1407-1422.
- Kenney, M. and Patton, D. (2011), Does inventor ownership encourage university research-derived entrepreneurship? A six university comparison, *Research Policy* 40, 1100-1112.
- Kilger, C. and Bartenbach, K. (2002): New rules for German professors, *Science* 298 (5596), 1173-1175.
- Lach, S. and Schankerman, M. (2008), Incentives and invention in universities, *RAND Journal of Economics* 39(2), 403-433.
- Leydesdorff, L. and Rafols, I. (2009), A global map of science based on the ISI subject categories. *Journal of the Association for Information Science and Technology* 60, 348-362.
- Lissoni, F. (2013), Academic Patenting in Europe: A Reassessment of Evidence and Research Practices, *Industry and Innovation* 20(5), 379-384.
- Lissoni, F., Lotz, P., Schovsbo, J., and Treccani, A. (2009), Academic patenting and the professor's privilege: evidence on Denmark from the KEINS database, *Science and Public Policy* 36 (8), 595-607.
- Lissoni, F., Pezzoni, M., Poti, B., Romagnosie, S. (2013), University Autonomy, the Professor Privilege and Academic Patenting: Italy, 1996–2007, *Industry and Innovation* 20(5), 399-421.
- Litan, R., Mitchell, L. and Reedy, R. (2007), Commercializing university innovations: alternative approaches, in: Jaffe, A., Lerner, J., Stern, S. (eds.), *Innovation policy and the economy*, vol. 8, University of Chicago Press, Chicago, IL, 31-58.

- Mowery, D.C., and Sampat, B.N. (2005), The Bayh-Dole Act of 1980 and University-Industry Technology Transfer: A Model for Other OECD Governments? *Journal of Technology Transfer* 30 (1/2), 115-127.
- National Research Council (NRC) (2010), Managing University Intellectual Property in the Public Interest, Committee on Management of University Intellectual Property: *Lessons from a Generation of Experience, Research, and Dialogue*; Stephen A. Merrill and Anne-Marie Mazza (eds.), The National Academy Press, Washington, D.C
- OECD (2003), Turning science into business: patenting and licensing at public research organizations, Paris: OECD.
- Pakes, A. and Griliches, Z. (1984), Patents and R&D at the firm level: a first look, in: *R & D, Patents, and Productivity*, Zvi Griliches (ed.), University of Chicago Press, 55-72.
- Schmoch, U. (2007), Patentanmeldungen aus deutschen Hochschulen - Analysen im Rahmen der Berichterstattung zur Technologischen Leistungsfähigkeit Deutschlands, Studien zum deutschen Innovationssystem Nr. 10-2007, Berlin.
- Schmoch, U. Licht, G., and Reinhard, M. (eds.) (2000), *Wissens- und Technologietransfer in Deutschland*, Stuttgart: Fraunhofer IRB Verlag.
- Siegel, D., Waldman, D., and Link, A. (2003), Assessing the impact of organizational practices on the relative productivity of university technology transfer offices: an exploratory study, *Research Policy* 32(1), 27-48.
- So, A.D., Sampat, B.N., Rai, A.K., Cook-Deegan, R., Reichman, J.H., Weissman, R. and Kapczynski, A. (2008), Is Bayh-Dole Good for Developing Countries? Lessons from the US Experience, *PLoS Biology* 6 (10), 2078-2084.
- Stevens, A. (2004), The enactment of Bayh-Dole, *Journal of Technology Transfer* 29, 93-99.
- The Economist (2002), Innovation's golden goose - The reforms that unleashed American innovation in the 1980s, and were emulated widely around the world, are under attack at home, *Technology Quarterly*, Q4 2002.
- Thursby, M., Thursby, J. and Gupta-Mukherjee, S. (2007), Are there real effects of licensing on academic research? *Journal of Economic Behavior & Organization* 63, 577-598.

- Trajtenberg, M. (1990), A penny for your quotes: patent citations and the value of innovations, *RAND Journal of Economics* 21(1), 172-187.
- Valentin, F. and Jensen, R.L. (2007), Effects on academia-industry collaboration of extending university property rights, *Journal of Technology Transfer* 32, 251-276.
- Von Proff, S. , Buenstorf, G. and Hummel, M. (2012) University patenting in Germany before and after 2002: what role did the professors' privilege play? *Industry & Innovation* 19(1), 23-44.
- Wooldridge, J. (1999), Distribution-free estimation of some nonlinear panel data models, *Journal of Econometrics* 90(1), 77-97.

Appendix A: Data collection procedure

Our relevant starting population for the patent collection are all patent applications filed at the German Patent and Trademark Office (DPMA) and the European Patent Office (EPO) involving at least one German inventor since 1978 up to 2011 using the PATSTAT database. These are 1,682,585 patent documents. Eventually we will collapse the list of relevant patent documents to the number of inventions, that is, we will account for patent families. Between 1978 and 2011 the grand total of patent families amounts to 1,067,753, and in the time period under review in this paper, 1995-2008, the number of different inventions with at least one German inventor amounts to 624,041.

Searching patents invented by university faculty

Unfortunately, no comprehensive list of German university faculty exists. Therefore, we follow another established strategy to identify patents of university professors (see e.g. Czarnitzki et al. 2007, 2009). In Germany, the award of a doctorate and even more of a professor title is considered a great honor. The “Dr.” is an official part of the name and is, for example, even mentioned in the national IDs and passports. The professor title is protected by the German criminal code (article 132a) against misuse by unauthorized persons. Accordingly this title is used as a name affix not only in academic environment, but also in daily life. Thus, we use the inventor records in the database and search for the title “Prof. Dr.” and a large number of variations of this. This initial search identified 69,250 patent documents between 1978 and 2011.¹³ After having obtained an initial list of patent

¹³ One may be concerned that the Professor Doctor title is also given as an honorary title to individuals who are not employed at universities. While the granting of honorary titles seems to be relatively rare, some of these highly qualified individuals may be labeled as professors in our data process. We believe any misclassification error would work against finding a significant policy effect as these individuals are not affected by the policy change.

documents, we then also searched for these inventors again in order to see whether they also patented without the “Prof. Dr.” title. Note that initially we just search for name homonyms of the identified “Prof. Dr.” patents. This step does not involve yet to disambiguate the records in order to find out which of these patents are invented by the same person and which are other inventors with similar names. The actual disambiguation is done at a later stage using cross-referencing to linked publication records. The search for name homonyms added 197,887 (1978-2011) to the 69,250 patent documents. We thus have a raw list of potentially university-invented patent documents of almost 270,000.

Identifying patents by PRO researchers

The identification of patents by PRO scientists is more straightforward, as they can be searched by applicant names as the IP was always subject to institutional ownership. We obtained a list of about 500 PRO institutes existing in Germany from the “Bundesbericht Forschung und Innovation 2012” published by the federal government. These were searched as applicants in the patent documents, and we identified 27,637 (1978-2011) patent documents. As some of these patents involve co-applications with firms, we cannot assume that all inventors listed on the patents are employees of PROs. Therefore, we first omit the co-assigned patents (about 20% of the 27,637). This detour is necessary in order to avoid that e.g. industry researchers whose employer appears as co-applicant on some patents enter our data of PRO inventors mistakenly. We then searched for all patents by the PRO inventors, in order to come up with a comprehensive list of patents filed by PRO inventors. Again, we initially search for name homonyms of these inventors as we did for the university faculty. Note that this step also adds the 20% of co-applied patents back into

the sample. Now, however, we have identified these not by applicants but PRO inventors. This raw list of potential PRO-researcher patents amounts to 195,498 (1978-2011).

Disambiguation routine – Step 1

The two lists of retained patent documents are now pooled (492,340). Note again that this list includes too many patents because of name homonyms. In addition, some inventors may switch between the two groups of institutions and thus appear in both lists. Therefore, we then implemented a first disambiguation routine based on the patent document data. This step determines which patents are clearly not invented by either university faculty or PRO researchers to extent this is possible to infer from the patent data. This initial disambiguation leads to a list of 29,476 unique inventors (either university faculty or PRO researchers) with a total of 174,431 patents (1978-2011).

The reason for the large drop in the overall number of patents is the deliberate oversampling by using the cleaned name (without title) as selection criterion. For example, 979 patents are filed under the common German name “Bernd Müller”, a number much too high for a single person. After the disambiguation procedure 61 distinct persons were identified. Only 3 persons belong to the target group of university faculty or PRO researchers, and these 3 inventors have in total 16 patents.

This disambiguation algorithm is based on a relation network analysis. Every node within this network is a patent connected to other patents by layers of relations defined by shared applicants, co-inventors, citations and joint sets of IPC codes. The analysis uses a hierarchical approach by first traversing connections of high reliability to define sub-clusters that function as new nodes for the next iterative step. By aggregating information within these ‘hypernodes’ new connections emerge that will also be traversed and so on. As every

sub-cluster describes a part of an inventor career, suspiciously large sub-clusters can easily be identified, rejected and re-traversed with more restrictive requirements for the connections. This method implicitly solves the common name problem. The resulting list of unique individuals and their corresponding patents has been checked manually to the largest extent possible.

Collecting publication data from the Web of Science and disambiguation - Step 2

The retained list of initially disambiguated inventors is now used to perform name searches in the Thomson Reuters Web of Science publication database, 1990 – 2008. We first retrieve all publications from Web of Science that match with respect to the names in our inventor list and have at least one German affiliation. This amounts to 882,702 publications; again including name homonyms. Second, we now use the publication information to disambiguate these authors from Web of Science using cross-referencing information on journals, coauthors, citations and affiliations. 580,448 are identified as being authored by the 29,476 inventors in our sample from 1990 to 2008.

In order to ensure that the match between inventors and authors has a high level of quality we then excluded weak matches. For doing so we only keep a researcher based on author-inventor-link if it is either the only match between author and inventor of the same name or if at least one affiliation matches between inventor and author. This reduced our uniquely identified researchers to 18,092.

Compiling the panel database

The final step of the database construction involves generating a panel of unique academic inventors that includes information on their patents, citation-weighted patents and publications for each year. We count patents at the family level to ensure that patents in

different jurisdictions for the same invention are not counted more than once. The unit of observation is a researcher-year. Some of the professors also appear as PRO researchers at some point in time.

The database is an unbalanced panel identifying 18,092 unique researchers with 99,624 patents and 447,596 publications to originate from a professor or a PRO inventor (overall time span).

The regression sample period that we use in our analysis runs from 1995 through 2008. Note that this sample also contains researchers who patented before 1995 in the sample. This implies that a researcher does not need to have a patent in the 1995 to 2008 period to be in the sample. We defined the study period to extend from 1995 through 2008 so that we observed enough time periods before and after the policy change. For each inventor, our data contains the individual's history of patenting between 1978 and 2008 and the individual's history of publications between 1990 and 2008.

Next, we exclude those researchers who were employed at both a PRO and university, as it is not clear which patent regime applied to these researchers. This reduces the number of observed researchers to 16,291. Beyond patent and publication characteristics, the data on patenting and publication history of every researcher allowed us to calculate each researcher's career age. Career age starts when we observe the researcher's first publication or patent application and increases incrementally thereafter to a maximum of 35 years after which we assume the researcher retires. Dropping researchers after 35 years and defining entry into the panel as either first patent or first publication our observed number of researchers drops to 15,770.

To account for earlier exit from academia, we adopted a 5-year rule that has a researcher leaving the panel if he or she had no patenting or publishing activity for five consecutive years.¹⁴ Researcher industry connections were determined from the patent data. An academic researcher is identified as having an industry connection when he or she is observed as an inventor on a company owned patent. This allows us to distinguish high cost and low cost academic inventors prior to the abolishment of Professor's Privilege and to estimate the model on subsamples to test hypotheses 1 and 2.

As Poisson fixed effects estimations exclude groups with zero outcomes in all periods of the panel, our regression sample excludes those researchers with zero patents in the observed period (they are in the initial sample as they had patented before 1995 and remain in the sample as they had some publishing activity in the last 5 years). Therefore the final estimation sample contains 108,263 researcher-year observations, containing 12,012 researchers (3,718 professors and 8,294 PRO researchers).

¹⁴ In section 6 we present an alternative exit rule; however, the results do not change in a meaningful way.

Table 7: Definition of variables

Variable name	Definition
# Patents	The number of patents applied for in year by an academic inventor
# Citation-weighted patents	The number of citations received by patents applied for in given year in the four subsequent years to the application date
Professor	The academic inventor was professor at some point in his career
Career age	The number of years elapsed since the academic inventor's first patent or publication
New policy	Dummy for years ≥ 2002
Professor*New policy	Interaction of Professor dummy and New Policy
Industry connection	The researcher has at least one patent applied for jointly with a firm applicant prior to 2001
Avg Publications	A moving average of journal publications over the past three years, t-1 to t-3

Appendix B: The inventor mobility index

Usually patent data does not contain any unique identifiers for the patenting assignees or the inventors, as the main tasks of patent authorities is the examination of applications and the administration of the patent documents as public contracts and not the support of the empirical analysis of their data. An inventor in a patent document is identified by his or her name. Depending on the patent authority the full address or parts of it may be included, to further identify this inventor. The goal is to define an inventor mobility index that traces the whole career of an inventor as a living person with all the job switches and relocations and represents the history of the mobility approximated by the patents as potential turning points. The inventor name is the main criteria for this identifier. The inventor address information on the other hand is only of limited use for the definition of a mobility index. The name alone can work for exotic name variants, but for more common names the problem of namesakes gets in the way of identifying individuals by this criteria only. The solution discussed here consists in the construction of a relationship network between inventors with the same name. This network will be created by using all the other information available in the patent data. These could be simple connections like the same applicant or just the same home address, up to more complex connections that are created by the overlapping of colleagues and co-inventors, similar technology fields or shared citations. Traversal of these heuristically weighted networks by using methods of the graph theory leads to clusters representing a person. The applied methodology will implicitly give exotic names a higher degree of freedom regarding the heuristic limitations than the more common names will get.

1 The SearchEngine

The patent offices do not administer special databases for assignees or inventors nor are they obliged to verify the names or addresses. Because of that there may exist multiple variants for a specific inventor or assignee, which can be explained by misspellings, different usage of abbreviations, name or address changes over time. If there is more than the data of one authority involved, this problem increases significantly because of different standards. The solution to this problem discussed here is to create an identifier for every group of variants that belong together with a high probability. This is a virtual cleaning process as the data itself will not be changed nor will there be a "preferred" variant that overwrites the other variants. The tool used for this task is simply called "SearchEngine" for further reference and is under continuous development by Thorsten Doherr, ZEW. It combines many ideas that have their origin in the field of computer science like word based heuristics, phonetic algorithms, fuzzy logic and network analysis.

1.1 The Preparer Gateway

One typical problem the algorithm is designed for is to match two tables from different sources by a combination of fields that share the same - usually fuzzy - characteristics, like name, address, city, zip and so on. A direct SQL join by these fields is of limited use because of abbreviations, misspellings and typing errors, different positioning of words or additional/missing words. The extensive harmonization of both tables by transforming the data to uppercase, replacing special letters to their common (phonetic) representation (i.e.: the German "Ü" to "UE"), suppressing of special characters and the unification of abbreviations will improve the situation for a direct join. These methods are also implemented for the SearchEngine as part of the preparer gateway. The preparer gateway is

responsible for the harmonization of all the data entering the deeper layers of the algorithm. Besides the more or less cosmetic modifications of the data it can also implement more extreme phonetic methods, that destroy the readability of the data but improve the robustness against misspellings and typing errors. Every field can be connected to a different list of preparers, some of them specialized to reflect the context of the associated characteristic. It is even possible to associate more than one preparer list to a single field, creating new entities. These combinations of preparers and fields are further called search types. The outcome of the preparer gateway is a set of words without any specific order separated into subsets by the search types they origin from. From this point on the term "word" describes all the token the preparer gateway returns after applying the harmonizing and/or the more aggressive phonetic preparer like Soundex (RobertRussell, Margaret Odell, 1918), Metaphone (Lawrence Philips, 1990), Kölner Phonetic (Cologne Phonetic) (Hans Joachim Postel, 1969) and n-gramm.

1.2 The Heuristic

The heuristic is based on the assumption that the occurrence of a word is inverse proportional to the identification potential (IP) of this word. Using the internet as an analogy, a quite common word entered into a search engine will result in a large list of results making it difficult to find the intended entry. The resulting list of potential hits for a seldom word is smaller as the identification potential is higher. Because a search usually involves more than one word, the algorithm uses a relative identification potential (rIP). The following section describes the development of this measurement starting with a basic first version:

$$rIP(i) = \frac{occ(i)^{-1}}{\sum_{j \in S} occ(j)^{-1}} \quad (1)$$

with S being a set of words defined by the search term, $i \in S$ and $occ(i)$ returning the occurrence of the word i .

To get the occurrences of the words the SearchEngine needs to be fed with the characteristics of one table, the so called base table. After passing the preparer gateway, the words of the search fields will be registered in a special table, the registry. An entry in the registry consists of a word and a counter for the occurrence of this word. The registry is further organized into chapters, one for every search type, to preserve the context of the words. Every single entry is also linked back to the containing records in the base table by supporting tables. The heuristic is extended by the possibility to put different weights on these search type chapters. These chapter weights are called priorities because they also influence the optimization of the implementation by giving the algorithm an order to work with. Another extension to the heuristic is the introduction of offsets that are added to the word occurrences. These offsets smooth out the relative differences between the words and can also be applied per chapter. The occurrence function now requires two parameters: the word and the search type the word belongs to. With $st(i)$ returning the search type of word i , $pri(j)$ and $off(j)$ returning the priority and the offset of search type j and n being the number of search types, the extended rIP_s can be defined as:

$$IP(i) = \max(occ(i, st(i)) + off(st(i)), 1)^{-1} \quad (2)$$

$$rIP_s(i) = \left(\frac{IP(i)}{\sum_{j \in S} \begin{cases} IP(j) & | \text{st}(i) = \text{st}(j) \\ 0 & | \text{st}(i) \neq \text{st}(j) \end{cases}} \right) \left(\frac{pri(\text{st}(i))}{\sum_{k=1}^n pri(k)} \right) \quad (3)$$

The function $occ(i, st(i))$ returns the average occurrence within the search type $st(i)$, if word i is not found in the registry for search type $st(i)$. The function max returns the numerical highest of the parameters.

Good values for the priorities and the offsets highly depend on the used preparer and the characteristics of the search types. In most cases a match has a clearly dominating characteristic like a company name that should get a higher priority as the supporting characteristics like address, city or zip. In conjunction with a customized cutoff limit it is possible to focus the match and reduce the number of false positives. The usage of offsets is much more experimental. They can be used to like a slider between an occurrence based heuristic and a simple word based metric where every word has the same value. This is especially true if the offset is negative and higher than the highest occurrence of a search type.

For any search term the words of the records found in the base table are compared with the words of the search term. For every shared word the associated rIP_s is summarized to get a measurement for the identity ranging between 0 and 1. An identity of 1 means all words of the search term exists also in the found record. Missing words from the search term result in a lower identity according to their rIP_s . Only found records with an identity above a given limit are considered candidates.

Until now all candidates with the same matching words are equal. In some cases it is desirable to rank these results according to the words of the candidates that are not part of the search term, thus preferring candidates with less additional clutter. The surplus words of

the candidates generate a discount on the identity, called *feedback*. The extent of the discount can be adjusted with the feedback parameter f which has a valid range from 0 to 1. With F being the set of all the words of the found candidate record, the final definition of the rIP_f is available:

$$Jaccard(i) = \frac{\sum_{j \in S} \begin{cases} IP(j) & | st(i) = st(j) \\ 0 & | st(i) \neq st(j) \end{cases}}{\sum_{j \in S \cup F} \begin{cases} IP(j) & | st(i) = st(j) \\ 0 & | st(i) \neq st(j) \end{cases}} \quad (4)$$

$$rIP_f(i) = rIP_s(i)((1 - f) + Jaccard(i)f) \quad (5)$$

The function (4) is called Jaccard because it is the implementation of the Jaccard similarity coefficient (Jaccard, 1901). It measures the similarity of two sets of properties by dividing the number of shared properties by the size of the union of both sets. With a feedback of 1, equal priorities and large enough negative offsets (equalizing all occurrences) for all search types, the identity transforms into a Jaccard index measuring the similarity between two sets of words.

1.3 The Implementation

The rIP_s is used as the heuristic for the search process that collects the candidate records for a search term. Given that the resources for the algorithm are restricted by computing power it is more profitable to first look for words with a higher rIP_s , until the resources for one search step are exhausted. The maximum size of the candidate list is the main regulator for a healthy balance between performance and completeness. Is the so called search depth too high, the performance will significantly be decreased for little benefit consisting mostly of false positives. A too restrictive search depth can lead to a loss of valuable hits, because

the word with the highest rIP_s may have a higher absolute occurrence. The identity limit will also be considered to further reduce the candidate list for the following steps. A candidate within the list already has a preliminary identity consisting of the words used for filling the list, which are usually just a fraction of the whole search term. If the identity of the unused search words won't push a candidate above the limit, it will be dropped from the list. In the next step the used words of every candidate will be synchronized with the remaining words of the search term to complete the calculation of the identity on the base of the rIP_f . It is this step that requires the majority of the computing power which explains the restrictive selection of the candidates beforehand.

1.4 Handling Misspellings

The SearchEngine is a word based algorithm. If a word cannot be found in the registry it will get a rIP based on the average occurrence for its search type. But the main problem is that there are no connected base table records making this word a dead weight. Another problem are misspelled common words that occur in the registry with a very low occurrence compared to the proper words. These words can misguide the search process in favor of the other misspelled entries. Phonetic preparer like Soundex and Metaphone reduce this problem by creating codes for similar sounding words. The n-gram method uses shifted tokenization (i.e.: $3_gram("DOHERR") = ["DOH", "OHE", "HER", "ERR"]$), creating multiple tokens for one word and thus reducing the impact of misspellings as they concern only a part of the tokens.

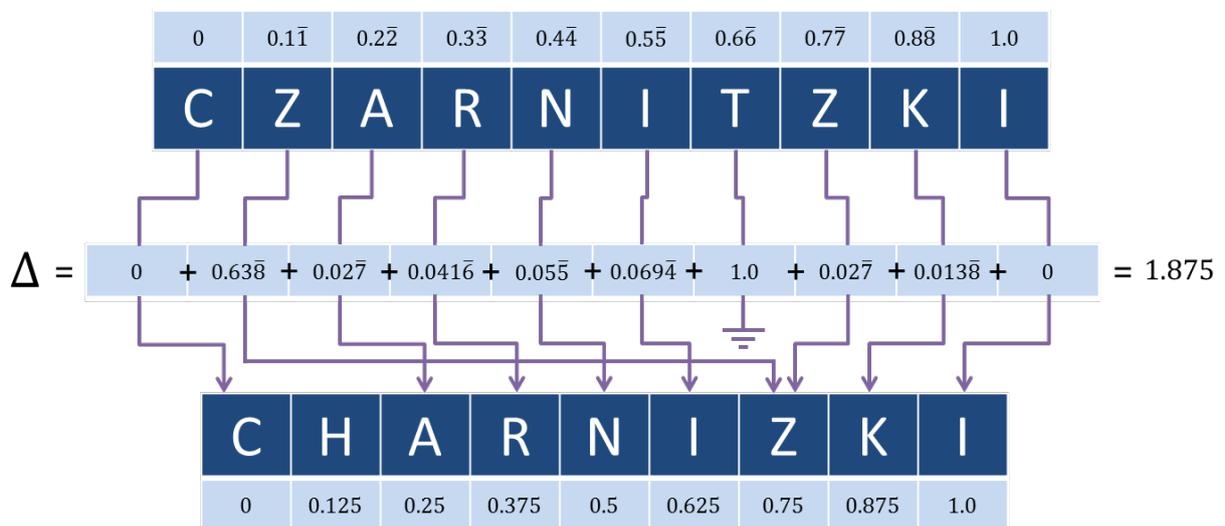
Table B.1: words represented by the same phonetic code

Method	Soundex	Metaphone	Cologne
Code	T652	BRTN	3467
Example 1	TARNOWSKI	BARATON	WAGNER
Example 2	THORENZ	BERTINI	WUCHENAUER
Example 3	TRUNK	BORDIN	WEGENER

There is a price to pay for the gained robustness. The algorithm will return much more false positives because the phonetic representations do not only include the misspellings but also similar "legal" words. In the case of n-grams all entries containing the same tokens have the same identity as the heuristic ignores positioning. The problem is, that phonetic methods are specially designed to retrieve false positives in the hope that the intended result will be within them. Usually these methods are used in an environment where an operator enters single requests into a terminal and examines the retrieved results. For the SearchEngine an additional layer has to be applied that fulfills this task. This layer simulates the operator by applying a string comparison function for every search type that implements phonetic preparer. This function returns a value between 0 and 1 for the similarity of strings, so it can easily be integrated as an equivalent to the identity of a search type. It compares every word of the search term with every word of the found term to identify the pairings with the highest similarities. The final result is the sum of these values divided by the highest possible score based on the term with the most words. An additional score will be calculated that compares the terms as whole strings. The maximum of both scores defines the identity. Both types of comparisons are necessary to guarantee a high flexibility of the measurement against different positioning of words and unclean separated words (i.e. by missing blanks). Because this flexibility requires a large number of comparisons the underlying algorithm has to be very efficient. The method used is called

Least Relative Character Position Deltas (LRCPD). Every character in a string has a relative position between 0 for the first and 1 for the last character. The algorithm searches for every character in the first string the matching character in the second string with the smallest difference between the relative positions. If a character can't be found a maximum delta of 1 is used. The sum of the deltas divided by the length of the first string returns a disparity measure between 0 and 1.

Fig. B. 1: two different spellings of a complicate name compared by the LRCPD method



$$lrcpd(word1, word2) = 1 - \frac{\Delta(word1, word2)}{len(word1)} = 1 - \frac{1.875}{10} = 0.8125$$

The LRCPD heavily depends on the direction of the comparison. For a symmetric behavior the comparison has to be done in both directions using the lower result. Another problem is the reduction of the deltas with increased string lengths. The limes of the average delta approaches zero for the comparison of infinite strings. For this reason the LRCPD implements a search scope around the relative position of the searched character. Starting from this position the search will be carried out in both directions until the character is found or the absolute distance to the start position exceeds the scope. The

delta of a found character will be adjusted as if the string length equals this limit, always resulting in deltas between 0 and 1. The SearchEngine uses an arbitrary default scope of 12 characters in both directions (not including the start position). A higher limit is only recommended for results that will be manually checked.

Search types that implement phonetic preparer somehow distort the idea behind the original heuristic. The codes or fragments returned by the phonetic methods have a different distribution of occurrences than the original words. Through fragmentation or aggregation the number of words stored in the registry is reduced, the average occurrence is increased which leads to more candidate records. This effect is subdued by the LRCPD layer but the main advantage of the original heuristic, finding candidates by the most identifying words, is watered down. Because of that, the SearchEngine supports incremental search steps. Multiple runs with different settings can be merged into one result set. Pairings of previous runs will not be overwritten by following search steps. It is advised to use phonetic preparer for later runs to fetch the candidates that actual have misspellings and to keep the main bulk of the results according to the heuristic.

1.5 Disambiguation

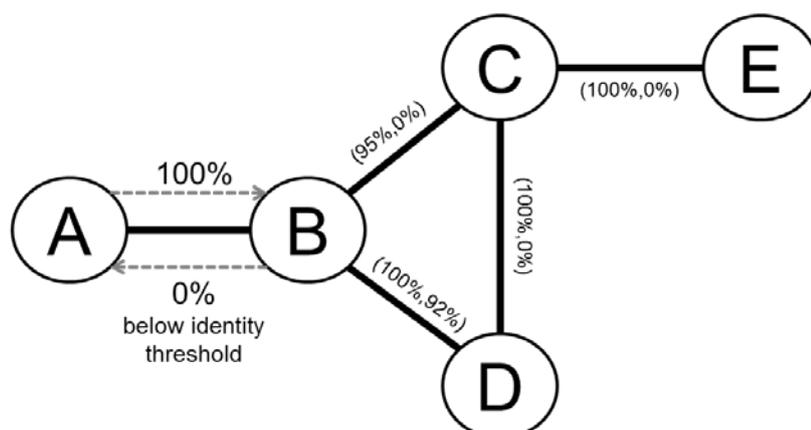
Now that there are all tools and methods in place the actual task of creating identifiers for variant groups of applicants and inventors can be put into focus. The only difference to a common match of two different data sources is that one data source is matched with itself.

There exist many different approaches to disambiguate or match this kind of data.

Trajtenberg et al. (2006) used the Soundex method and introduced a frequency based heuristic. Raffo and Lhuillery (2009) analyzed different cleaning methods for a simple sting based algorithm and compared them to n-gram methods in respect to recall rate and false

positives. Schoen, Heinisch und Buensdorf (2014) combined simple string matching, n-gram and a Jaccard similarity coefficient for their “name game”. All these approaches have in common that all the matching results are transitive, be it by the method used or enforced by following cleaning up. The latter is seen problematic but imposed nevertheless as “the only plausible course of action” (Trajtenberg et.al., 2006). The results of the SearchEngine can also be forced into transitivity by applying a feedback of 1 transforming the identity into a weighted Jaccard index. The advantage of transitive matching in respect to disambiguation is the consistent mapping of entities into groups. Intransitive matching means that the identity of the reversed match can differ from the original identity. The reverse match can even return a value below the identity threshold. If transitive pairs define a network consisting of easy to identify clusters of fully connected subgraphs, intransitive links between nodes create complex directed subgraphs. The connection strength between two nodes can be defined as a tuple consisting of the maximum and the minimum of both identities eliminating the direction of the edges, but the graphs are still not complete.

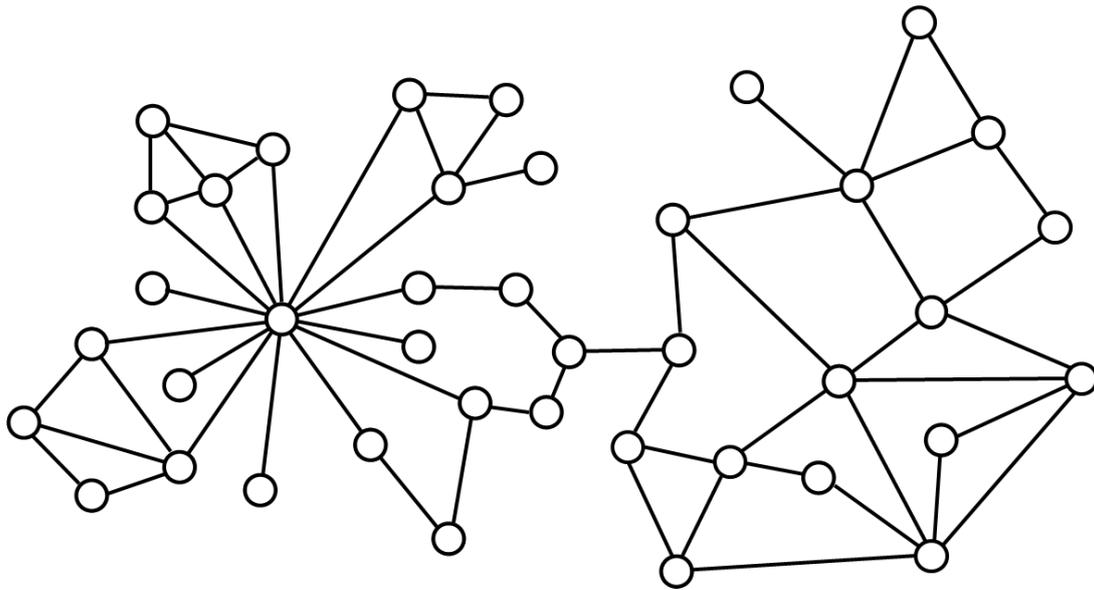
Fig. B.2: undirected graph



It is obvious that a network analysis is required to identify the clusters in such a network. This requires a lot more effort than simply collecting the clusters in the network of complete subgraphs defined by the transitive match. But this effort is justified by the additional freedom of the intransitive match. Intransitive matching allows pairs consisting of over-specified and relatively underspecified search terms to exist as connected nodes. An over-specified search term has additional clutter that distracts from the actual target, i.e. mentioning subdivisions that obfuscate the firm name. As long as the actual target exists in the data in a proper specified form it will collect all the over-specified entries even if these are not able to find the actual target on their turn.

A high identity threshold provides that the single connections in the network are believable. But the size and the structure of a graph can lead to initially unexpected composition of a cluster. Two meta structures can be identified as the main perpetrators in this regard: black holes and thickets. A black hole is a node that has a suspicious number of connections. These are caused by underspecified data artefacts, i.e. a company name consisting only of a legal state or a city name. Luckily these can easily be detected and mitigated before the traversal of the network by cutting all weak connections of a node whose number of connections exceeds an artifact threshold.

Fig. B.3: a black hole and a thicket



A thicket can't be identified pre traversal. Its structure is chaotic and only discernible from a healthy cluster during traversal. It is the direct result of under- and over-specified terms that build upon each other. An over-specified term can link to several underspecified terms which open the way to their own subgraphs containing over-specified terms and so on. In this context "over-specified" does not automatically mean "clutter", but also properly specified terms of common words. To solve this problem suspicious large clusters can be traversed again, but now with a limit on the connection strength. A more efficient method to this approach is the cascaded traversal.

1.6 Cascaded Traversal

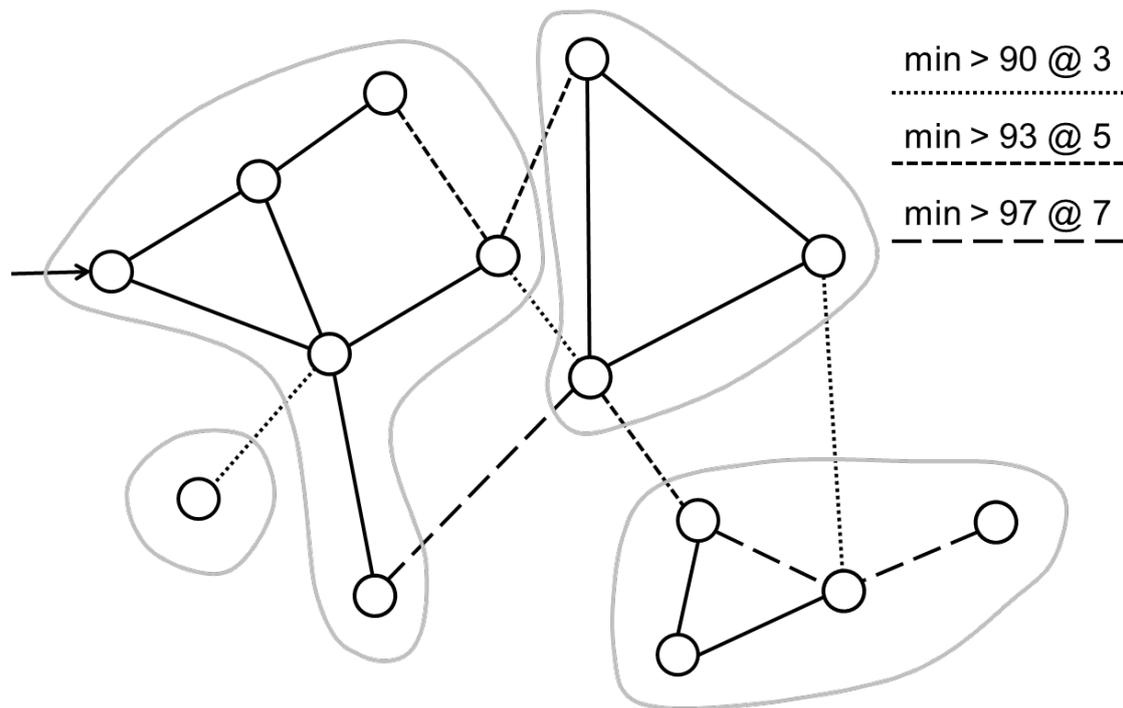
Originally introduced to cut down thickets, cascaded traversal has some additional benefits. The basic idea was to identify thickets during traversal and not after the complete network analysis to save computing time. Every time a cluster reaches a defined node limit the traversal has to start again with a more limiting threshold for the connection strengths. The cluster size limit is a discrete value that can be determined by answering questions like:

“How many misspellings are imaginable for a name?” or “How many different variants of a company name seem to be plausible?”. The answers may be arbitrary but as there are usually multiple cascades with increasingly restrictive conditions in place, the whole process can be adjusted for adaptability. To define a cascade following conditions and steps have to be considered:

- Define a set of rules with increasingly restrictive conditions for the validity of a connection. Any rule has to include the restrictions of the previous rule.
- Attach a maximum cluster size to every rule, i.e.:
unrestricted, min > 90 @ 3, min > 92 @ 5, min > 95 @ 5, min > 97 @ 10
- The rules will be exclusively activated in order of definition. The active rule will be replaced if the cluster size of the following rule is exceeded.
- Every time a new rule is activated, the traversal of the network starts again for a given start node with the new rule in place.
- A valid start node is any node that does not already belong to a cluster created by another start node.

Any rule creates a new virtual network that is a thinned out version of the network defined by the previous rule. As the propagation of this thinning out process is independent from the start node, there is no overlapping of the resulting clusters. As the cluster size limit can grow from rule to rule to reflect the increase of the connection quality of the remaining network, the cascade adjusts itself by being easy on smaller groups and even letting larger groups survive, as long as the connections are strong.

Fig. B.4: a cascade with incremental cluster limits is thinning out a thicket



As manual checking is often not feasible for large numbers of observations the identity threshold for the disambiguation of the applicants and inventor names is quite high to guarantee that the connections in the resulting graph already have a good quality. The maximum value for a connection is always equal or higher than the identity threshold. The minimum value for a connection can be zero if the search in the reversed direction returned an identity below the threshold. For this reason only the minimum is used for the rules. The rule set for the inventor names should be more restricting than the rules defined for the applicants, because the inventor name index is the base for the next step to identify individual inventor careers. The applicant variant index is only used as an instrument to identify these careers in conjunction with the inventor name index. The probability that two inventors with the same name index have patents for two different applicants that are by mistake in the same variant group is quite miniscule. Given the quality of the patent data an additional variant index can be created for the home addresses of the inventors.

2 The Inventor Mobility Index

The inventor name index defines a name space into which all patents of inventors with this name belong. There are two extreme positions imaginable: all patents are from one person only or all patents are from different namesakes. The truth most probably lies in between. By investigating the data available for the patents of this name space it becomes clear that these patents have different types of connections between them. This could be two patents being invented at the same employer (applicant) or inventor home address, sharing some co-inventors, citing each other or are about a similar technology. These pairwise connections span a network with heterogeneous definitions of connection strength. One way to solve this problem is a network analysis of the whole graph using patents as nodes. Because of the high interconnectivity and the disparity of the connection quality the high probability of thickets could be countered with cascaded traversal. But as the criteria for the cluster size limit is the number of patents it becomes obvious that the definition of a rule set will be uncomfortably arbitrary. The number of patents per inventor is an endogenous criteria and the definition of the connection restrictions require some kind of ranking or weighting of the different connection types. The better solution is to define a hierarchical order of the connection types. First only the more trustworthy types are used for traversal. The resulting clusters are the basic milestones of the inventor career. These clusters are also called hypernodes as they become new nodes of a nested-graph model (Alexandra Poulouvassilis and Mark Levene, 1990, 1994) as the different layers of connection types are applied in order of reliability. This approach has two major advantages: it is possible to aggregate the data of the patents within a hypernode to create additional information a single patent could not provide and it is more comfortable to define a cascade using career

milestones. Also, the cascaded traversal implicitly solves the common name problem.¹⁵ One symptom of a common name is a huge name space occupied by dense thickets. Cascaded traversal keeps the number of milestones within a tolerable limit for an inventor career without the explicit knowledge about the commonness of a name.

2.1 Home Address

Having the home address on the inventors is the ideal case. After defining a variant key for the different existing addresses using the SearchEngine it is possible to define the first layer of hypernodes within a name space. A hypernode contains all the patents a person invented at a specific home address. As the nested graph is complete no traversal is required to collect the patents of a hypernode. From this point on the unit of the cluster size limit for a cascade is the relocation.

2.2 Applicant

Usually the applicant is the employer of the inventor. In cases where the inventor address is missing or not exact, i.e. not on street level, the applicant is the substitute for the home address to define a career milestone. In contrast to the home address a patent can have multiple applicants resulting in an incomplete graph. Albeit this is a seldom occurrence network traversal is required. If the home address is already used as the first layer of hypernodes, the different applicant variant keys within a hypernode define the links between the nodes. It is not possible to run a differentiated cascade because the affiliation to an applicant is a binary decision. There are no stronger or weaker connections to discriminate. This is a problem for very large companies in conjunction with common

¹⁵ A more detailed description of this algorithm is described in appendix B.

inventor names. The only solution is a simple cascade that cuts all connections if the cluster size limit is exceeded leaving the hypernodes unconnected. The size limit should not be defined on patent level but on the number of relocations. The consequence is the inevitably overestimation of the number of inventors for large companies. The unit of the cluster size limit for the following connection type is the job change.

2.3 Citations

A citation creates a strong link between two patents. If the same inventor name appears in both documents the probability of them being the same person is very high. In spite of this excellent connecting property this connection type is only at third position as a patent not necessarily has citations that connects back to the same data source. Like the employment the connection by citation is also binary but much more trustworthy. The unit of the cluster size limit for the following connection type is called the environment change.

2.4 Peers

This is finally a connection type that implements the advantages of the hierarchical approach. Using the already defined hypernodes it becomes possible to not only connect single patents by co-inventors but research environments by peers. All co-inventors within a hypernode are the colleagues at a specific career milestone. The strength of a connection is defined by the absolute number of shared colleague names and the Jaccard similarity index based on colleague names between both hypernodes (the weighted mean of both shares of the absolute number of shared colleague names in respect to the corresponding colleague name counts). A cascade will require ascending limits for these values. A higher absolute value should cancel out any restriction based on the relative value as the risk of an invalid connection decreases with the probability to draw the same combination of names by

chance. The minimum can be used in case the absolute number is low and the number of colleagues is high, increasing the potential of accidental drawing the same lean combination by chance. This connection layer often identifies reorganized applicant entities that could not be joined by a variant id because of name and location changes. The unit of the cluster size limit for the following connection type is the community change.

2.5 Expertise

The international patent classification (IPC) codes describe the technologies that define the inventiveness of a patent and thus indirectly the expertise of its inventors. The codes follow a hierarchical system which allows for truncation at specific positions to get a broader view on the involved technology respectively expertise. Only the first 4 positions of the code are used. The ambiguity whether a code is associated with the expertise of the inventor or not increases with the number of co-inventors for a specific patent. To reflect this correlation all codes of a patent get a weight equal to the inverse of the number of inventors. The overall expertise scheme of an inventor within a hypernode is defined as a list of the distinct codes along with their share on the cumulated weights. A connection between two hypernodes is created by the codes they have in common. As the sums of the associated shares will differ the connection strength is defined as a tuple. The first cascades use the higher of both values because of the heterogeneity of the hypernodes. The role of an inventor in a small institution may be much more pronounced than her work in an institution with larger teams. The later cascade levels have to restrict the connections by the lower value of the tuple to keep the cluster sizes under control. The expertise is a relatively weak connection type and should always be the last in the traversal order because larger hypernodes allow

for a better assessment of the work of an inventor. If there would be further cascades the unit of the cluster size limit for the following connection type is the reorientation.

2.6 Final Discussion

A good benchmark for the adjustment of the cascades is the observation of unique names. The more exotic ones can easily be identified by common sense. The more technical approach is to separate the distinct inventor names into first and last name and look for last names that have one or at least only a small number of first names. The name spaces of these inventors are equivalent to their careers. After the final clustering all patents of a benchmark inventor should end up in one hypernode. Naturally this will not happen in all cases as there may be simply no connections between some hypernodes. The adjustments to the cascades should not enforce perfect identification of the benchmark group as this will inevitably lead to false positives within the not controlled cases. There is nothing to say against clustering the benchmark group by their name spaces afterwards.

References for Appendix B

- Trajtenberg, M., G. Shiff, and R. Melamed (2006), The “Names Game”: Harnessing inventors’ patent data for economic research. Working Paper, National Bureau of Economic Research, Cambridge.
- Schön, A., D. Heinisch, and G. Buenstorf (2014), Playing the ‘Name Game’ to identify university patents in Germany, *Scientometrics* 101, 527-545.
- Poulovassilis, A., and M. Levene (1994), A nested-graph model for the representation and manipulation of complex objects, *ACM Transactions on Information Systems* 12(1), 35-68.
- Raffo, J. and S. Lhuillery (2009), How to play the “Names Game”: Patent retrieval comparing different heuristics, *Research Policy* 38(10), 1617-1627.

