UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Intellectual Property and Academic Science

Permalink
https://escholarship.org/uc/item/1tq6j8h5

Author
Thompson, Neil

Publication Date
2012

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Intellectual Property and Academic Science

By
Neil Charles Thompson

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy
in
Business Administration
and the Designated Emphasis
in
Computational Science and Engineering
in the
Graduate Division
of the
University of California, Berkeley

Committee in charge:
Professor David C. Mowery, Chair
Professor Lee Fleming
Professor Brian Wright
Professor Bronwyn Hall

Fall 2012

Abstract

Intellectual Property and Academic Science

by

Neil Charles Thompson

Doctor of Philosophy in Business Administration

University of California, Berkeley

Professor David C. Mowery, Chair

Academia's usage of intellectual property (IP) has occasioned passionate debates on both sides. Supporters argue that it speeds the transfer of scientific discoveries to the private sector, and have advocated policies such as the Bayh-Dole Act to promote it. Detractors worry about the collision between the norms of science and the norms of commerce. They fear that the exclusionary rights of patents and licenses may fence off areas of research, making the costs to science outweigh the benefits from increased technology transfer. This dissertation addresses this empirically by testing whether the issuance of a license increases or decreases the flow of knowledge on that discovery, using citations to the focal publication(s) as a measure of knowledge flows.

In order to address this question, this dissertation first develops two intermediate results. The first is a primer on material transfer agreements which shows that they can be used as a proxy for research tools – an area where the conflict between scientific and commercial norms may be more significant. The second is to show that *inventor-based matching*, a new, automated methodology for matching intellectual property to its underlying scientific publication(s), can be used to construct a large dataset for this analysis.

Using these intermediate results and a non-parametric method for building a convincing control group, this dissertation finds two important results. First, it finds that, for most discoveries, licensing *increases* the flow of knowledge on that discovery. This is consistent with the license providing a positive signal about the discovery that benefits the underlying science. In contrast, this paper also finds that for research tools, licensing *decreases* the flow of knowledge on that discovery. This may indicate that licensed materials are being shared less widely amongst scientists, raising concerns about science progressing more slowly in these areas.

Acknowledgements

This dissertation is an expansion of the work contained in Thompson, Mowery, Ziedonis (working paper). It contains excerpts from that work throughout the document. These are not called out individually, but just noted here.

I am grateful to the employees of the UC Office of the President, and the U.C. Berkeley and U.C. Davis Offices of Technology Transfer for their invaluable assistance in gaining access to the intellectual property data employed in this paper, and to Thomson Reuters for their cooperation in obtaining the data on scientific publications. Thanks also to Professors Fiona Murray and Scott Stern of the Sloan School at MIT for sharing their data with us. Earlier versions of this work benefited from the comments of participants in seminars at the Haas School of Business, U.C. Berkeley. This work benefited from the financial support of the Institute for Business Innovation at the Haas School of Business, the Industrial Partnerships Office at Lawrence Livermore National Laboratory, and NSF Grant SMA-1064194.

I am indebted to my advisor, David Mowery, for excellent guidance and mentorship. He has steered me towards promising paths, and pulled me back from thorny ones. I greatly appreciate all he has done for me. I am also grateful to Lee Fleming, Brian Wright and Bronwyn Hall for excellent advice and encouragement. I also want to acknowledge the students of the BPP program, who have seen me present this work innumerable times. Their patience and feedback have helped this paper greatly.

On a more personal note, I want to thank my family and particularly my parents. They have made me the man I am, and without their support, I wouldn't be here.

Finally, thanks to my wife, Kate, for all of her love and support. She is my ever-faithful flossing buddy and we are *Better Together*.

Table of Contents

1 Introduction

The use of patents and other forms of intellectual property to 'protect' academic science has occasioned passionate debate. Supporters argue that it speeds the transfer of scientific discoveries to the private sector, and have advocated policies such as the Bayh-Dole Act to promote it. Detractors worry about the collision between the norms of science and the norms of commerce. They fear that the exclusionary rights of patents and licenses may annex off areas of research, making the costs to science outweigh the benefits of increased technology transfer.

Despite the importance of evaluating this question empirically, relatively little work exists that evaluates how the assertion of a university's intellectual property rights impact academic science. Of the work that does exist, much of it is survey work, eliciting scientists' opinions of, and experiences with, intellectual property. This predominance of descriptive work in this area is likely driven by several data challenges. While patent information is publically available, the data on other forms of intellectual property, such as licenses, are usually proprietary, hindering access. Another challenge is connecting intellectual property to the science upon which it is based. This connection is often unclear and time-consuming to evaluate.

This dissertation attempts to address these issues. Chapter 2 introduces a collection of data, which provides a rich background on the usage of intellectual property by the University of California system, including a particularly detailed view of material transfer agreements. It also highlights how this data can be used to identify research tools, where theory suggests we might expect the most conflict between the norms of science and those of commercialization. Chapter 3 presents *inventor-based matching*, a novel way of matching intellectual property to scientific publications, automating and scaling up this process to allow the evaluation of a large dataset. It then implements this to create a set of *patented publications*, that is, publications whose discoveries have also been patented. Chapter 4 uses this set of patented publications to evaluate the impact of licensing on the flow of scientific knowledge, as measured by changes in the citations to the scientific paper. This is examined both for the general case and for research tools specifically. Finally, the chapter concludes with a discussion of the importance of these results and their welfare implications.

2 Intellectual Property at the University of California

Universities have long been important producers of research, particularly basic research. This role has grown in recent years, as data from the National Science Foundation's *Science and Engineering Indicators* shows (see Figure 1). Contemporaneous with this increased role in research has been a marked increase in university usage of intellectual property, particularly in biotechnology. Figure 2 shows the growth in academia's share of U.S. patents since 1969.

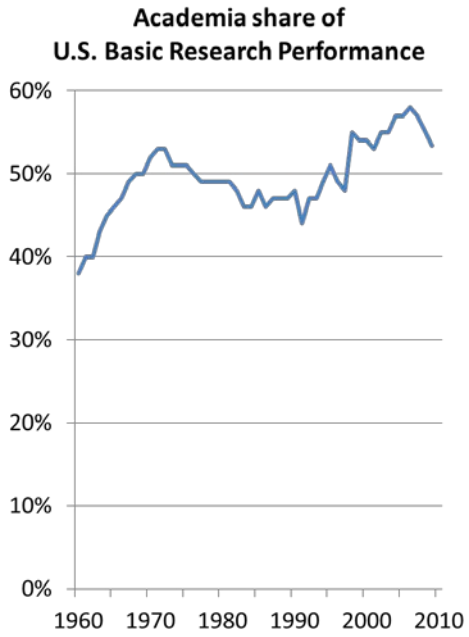

Figure 1[1]

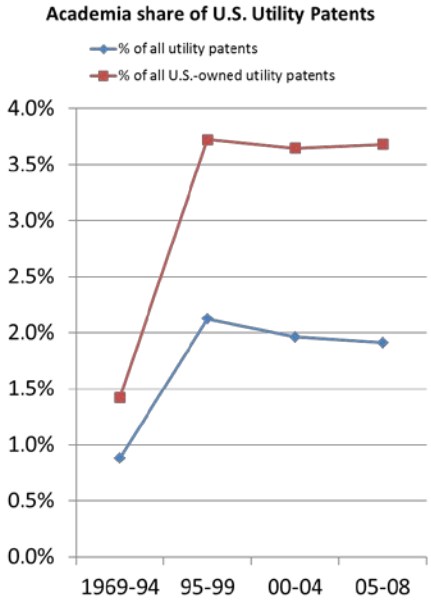

Figure 2[2]

The University of California (UC) system, with ten campuses and nearly 19,000 faculty members, is a major source of research and intellectual property in the United States.[3] In 2009-2010, the system received $4.3 billion in research funding, held nearly four thousand active patents, and received ~1,500 new disclosures of potentially patentable discoveries.[4] The size of UC's impact on research can be understood contextually, for example through the UC biomedical research and education system, which comprises 18 health professional schools and programs, including 5 medical schools, making it "the largest system of health sciences research and training in the nation."[5,6] This

[1] National Science Foundation (2011). National Science Foundation (2012), table 4-4. All research values on a dollar-input basis. Value for 2008 interpolated from 2007 and 2009 data.

[2] USPTO (2012).

[3] Hunter-Davis (2012).

[4] University of California (2012).

[5] University of California (2012b).

dissertation contends that, as a result of its sizable research and intellectual property activities, the UC system is an important case in itself, as well as a reasonable sample for inferring about trends for a larger set of U.S. research universities.

This chapter provides a descriptive account of intellectual property usage at the Universities of California. It provides information on patenting, licensing, and on material transfer agreements (MTAs) – contracts governing the transfer of materials between researchers. Because more is already known about patents and licenses, these are only touched on briefly to provide context. The focus is instead on MTAs, about which less has been written. MTAs are also discussed at length because in Chapter 4 they are used to identify research tools, an area of particular public policy interest. This chapter shows empirically why that usage is reasonable.

2.1 Disclosures, Patents, and Licenses

Since 1963 the patent policy of the University of California has required "that employees and certain others agree to assign inventions and patents to the University or other parties as appropriate, [and] to promptly report and fully disclose potentially patentable inventions."[7] Figure 3 shows that these disclosures by UC employees are rising over time. A simple OLS regression estimates this growth at 7.2% from 1998-2007.[8]

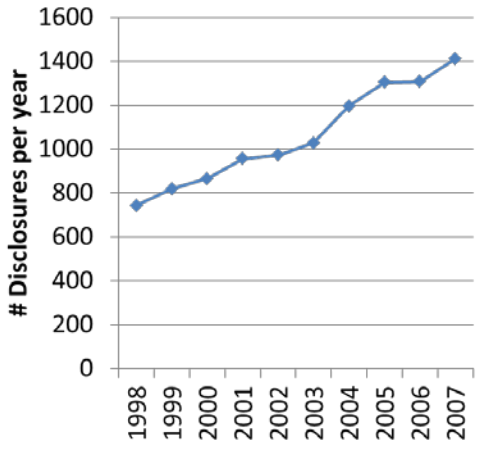

Figure 3: Disclosures at UC Campuses[9]

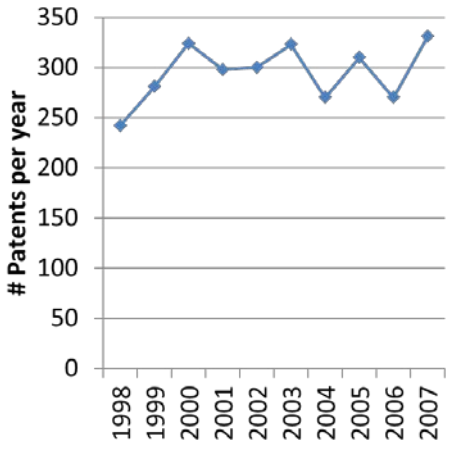

Figure 4: Patents at UC Campuses[10]

[6] University of California (2012).

[7] University of California (2012). Current wording quoted.

[8] Regression specification: $\log(disclosures) = \alpha + \beta * year$

[9] University of California Office of the President (2012b).

[10] University of California Office of the President (2012b).

There may be many reasons for this increase in disclosures. Plausible ones include increased numbers of personnel or research funding at UC, increased interest by faculty in commercializing their existing work, or increased interest in areas of research that are more commercial.

This increase in disclosures has only partially translated into increased patenting, which a similar regression analysis shows has grown only 1.2% per year over this period (see Figure 4). Discussions with OTT officials suggest that this reduction in the percentage of disclosures that are patented is a cost saving measure, with universities seeking licensees for research disclosures (and seeking to have the licensee underwrite all or a portion of the patent prosecution expenses) prior to filing a patent application. This explanation is supported by the trend in UC licensing, which has grown, albeit volatilely, at 6.1% per year over this period. Figure 5 shows this trend.

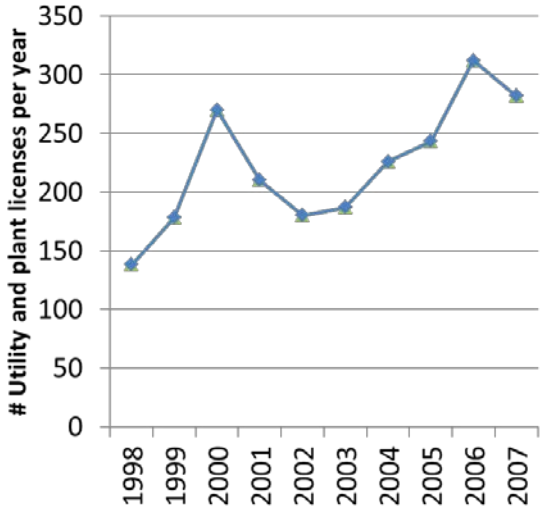

Figure 5: Utility and plant licenses at UC campuses[11]

2.2 University of California Data

This remainder of this chapter uses a collection of data, both quantitative and qualitative, from the University of California system. The qualitative data includes interviews with staff members at the offices of technology transfer (OTT) at UC Davis, UC Berkeley, and the UC Office of the President (UCOP). Interviews were also conducted with scientists at UC Berkeley and UC Davis, of which the UC Davis ones were focused on heavy users of material transfer agreements.

Several quantitative sources of data were also used for this analysis. The first, 'IP data', is an extract from the technology disclosure database maintained by the Technology Transfer Office within UCOP. This office tracks invention disclosures for all campuses of the University of California. In the period covered by this data, this includes nine of the ten current campuses (not UC Merced), and five medical schools. UCOP also tracks other agreements and intellectual property

[11] University of California Office of the President (2012b). Includes both plant and utility patents.

that relate to these disclosures, examples of which include patents, licenses, and material transfer agreements (MTAs – contracts that accompany material transfers between scientists, explained in more detail below). The database extract used for this analysis spans all invention disclosures from 1997 to 2007. Figure 6 presents a summary of the contents of the database extract.

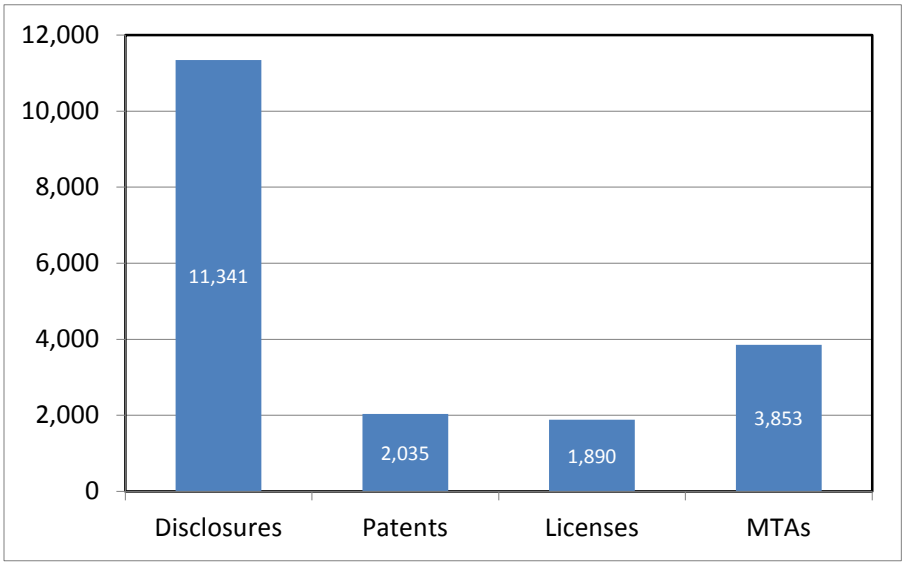

Figure 6: UCOP IP Data

While most, if not all, of the UC patents and licenses from this time are recorded in this database,[12] the same cannot be said of MTAs. Interviews with OTT officials suggest that most MTAs occur without the university's knowledge. Of those that are known, those that relate to invention disclosures are tracked by UCOP. And while it is understood that a greater share of MTAs on disclosures are reported, many are still likely to be missed.

An exception to the rule of most MTAs going untracked is UC Davis, which actively monitors MTAs not related to disclosures.[13] They have generously provided the second quantitative data set for this analysis. Their data covers spans 1999-2008 and includes 3,302 MTAs not linked to invention disclosures. Of these ~40% come from a single organization, the Mutant Mouse Regional Resource Center (MMRRC), which is a biological resource center of the kind discussed in Furman and Stern (2011). Importantly, this database tracks both incoming MTAs (where UC Davis is a recipient of materials) and outgoing MTAs (where UC Davis is sending the materials), whereas the UCOP IP Data contains only outgoing MTAs.

A third source of quantitative data provided for this analysis comes from the UC Berkeley OTT. It includes a record of each incoming MTA, i.e. where a UC Berkeley researcher received materials from another researcher. It includes 671 MTAs from 2004-2008, and includes a record of both the

[12] A probable exception to this would be software that is released with an open source license.

[13] Despite this active monitoring, some MTAs almost assuredly remain untracked by the university. It is difficult to estimate the size of this untracked portion.

material being transferred and its intended usage. Because UC Berkeley's default is *not* to require an MTA when receiving materials, these observations occur when the other party wishes an MTA.

Since most of these data sources are proprietary, care has been taken to present them only in aggregate form.

2.3 Material Transfer Agreements (MTAs)

The aim of this section is to provide a description of MTAs. It does this by discussing the negotiation process of MTAs, the growth of MTAs related to disclosures materials, and the composition of materials being transferred with MTAs. As part of this last analysis, it specifically addresses whether the presence of an MTA is a reasonable indication that a discovery is a research tool, a fact which will be used in chapter 4.

2.3.1 Negotiating MTAs

MTAs are contracts that accompany the transfer of materials from one scientist to another. They are important by virtue of the importance of the material transfers they accompany. Inasmuch as they help or hinder these transfers, they impact the progress of that science, as discussed immediately below.

Work by Walsh, Cho, and Cohen (2007) highlights the importance of material transfers. Using a survey of genomics and proteomics researchers, they find that 75% of these scientists had requested research materials in the previous 2 years, and that, on average, the scientists had made 9 requests – 2 from industry and 7 from academic sources. They further report that in 2003-2004, academics rejected 18% of the material requests and industry rejected almost twice that percentage. They then show that the denial of material requests often delays research by one or more months, or, more seriously, causes scientists to abandon projects. The importance of these transfers is highlighted by the magnitude of these effects; the percentage of scientists forced to delay or abandon their work due to material requests denials is much higher than those forced to delay or abandon their work due to patenting by other researchers.

Historically, materials exchanges were governed by little more than a letter from the source accompanying the materials, requesting acknowledgement and in some cases asking that the materials not be passed on to third parties (See McCain, 1990). Today, however, there exists a formal process whereby both researchers and their institutions sign Material Transfer Agreements (MTAs) which accompany the materials. Despite the existence of this process, many, perhaps the majority of material transfers do *not* use MTAs, with researchers instead preferring to share informally amongst their networks. However, interviews, with campus licensing officials and with scientific researchers suggest that MTAs are much more common for discoveries that are disclosed to OTT officials, and perhaps even more so for those that are patented. This may reflect OTTs' interest in protecting the property rights for these patented disclosures, leading campus licensing

officers to undertake transfers related to these disclosures within the legal protections offered by MTAs.

Typical terms for MTAs include indemnification of the sending party, prohibitions on using the material in human testing, restrictions on sharing the materials with other researchers, and a delineation of which party owns derivative or modified versions of the material.[14] In the Walsh, Cho and Cohen (2007) sample, 29% of finalized MTAs included reach-through rights (i.e. some or all ownership of modifications revert to the original material provider), and 16% of finalized MTAs had royalty demands. Interviews with UC researchers highlight the tension and conflict that is created by these types of terms. One scientist noted that the rights requested by private companies in MTAs can prevent him from pursuing some research. At the same time, however, he wonders whether he should request more rights on the materials that he sends (currently he asks for co-authorship, but no reach-through rights).

The negotiations of the terms of MTAs can be complex, as both the scientists exchanging materials and their institutions need to agree on the terms. In the interviews conducted for this research, this led to examples where the scientist and their institution had different interpretations of what should be signed. For example, one interviewee stated that he takes reach-through provisions "with a grain of salt," since he believes that few are prosecuted and because he isn't focused on commercialization.[15] Not surprisingly, the university was less willing to take on such risks. For UC officials this was not just about liability, but also about protecting academic scholarship. An example of why this is at risk can be seen in the Walsh, Cho and Cohen (2007) sample, where 30% of MTAs had restrictions on the publication of results – which clearly could have career impacts on both the primary investigator and their lab. To address these issues, UC has a series of principles that they use to evaluate whether to sign an MTA. The following are a subset of that list that highlights some of these issues:

[14] See, for example, Association of University Technology Managers (2012).

[15] Anecdotally, this seems to be correct. One OTT official asks this of their colleagues at every meeting, but has yet to find a single instance of a legal dispute centering on an MTA.

Berkeley's principles for approving incoming MTAs[16]

1. No one gets to approve results or restrict reporting to only positive results
2. MTA can't forbid students from being part of research, unless the project really is just the researcher
3. It must be possible for the UC researcher to transfer modified materials, the default level of sharing being those in the NIH sharing rules. This is true even if that researcher isn't on an NIH grant.
4. Research must be publishable
5. Onerous or atypical reporting requirements to the material provider are talked through with researchers
6. Companies do *not* get free licenses on the materials that result from the work, since both the state and federal governments also made significant contributions to make the research possible. However, tiered licensing, where the company gets a research license and a first option on a commercial license, are possible.
7. Researchers, not the company, choose which experiments will be done (although ex-ante discussions between them are fine)

In some cases (in spite or because of these principles) the UC researcher is willing to accept harsher restrictions than the university, creating tension. According to one interviewee "[MTAs] have impacted me very negatively…[and] usually *our* university is the barrier." For him this happens because companies want reach-through rights on his work, which he is fine with: "I don't give a %&@# about those rights, I'm never going to set up a company."

Conflicts about MTAs also occur within companies deciding whether to execute a material transfer with the university. According to a university OTT official, the typical pattern is that the industry scientists has spoken to a university researcher and wishes to complete a material transfer, but the company's business side argues that it jeopardizes market potential, and their legal department argues that the MTA doesn't provide sufficient protection. As a result of these conflicts, MTAs with industry may be slow to complete or may fail altogether.

This is important because the private sector plays a significant role both in sending materials to, and receiving materials from, UC. The UC Davis data show that 31% of incoming MTAs are with the private sector, as are 21% of outgoing MTAs (see Figure 7).

[16] Interview with UC Berkeley OTT staff

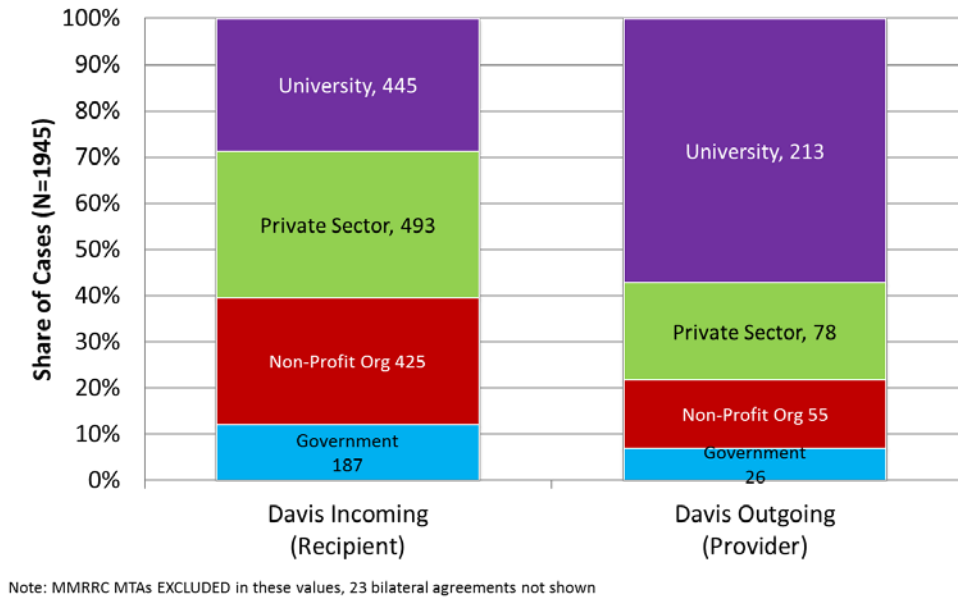

Figure 7: Commercial status of UC Davis counterparties[17]

According to OTT officials, these private sector interactions are significantly more complex and time-consuming than those with academics or non-profit organizations. Three types of cases are particularly difficult. The first are small companies, where the response lag time can be high. The second are international companies where language, cultural, and legal barriers can slow down interactions. The third are companies that receive many requests. To deal with this volume, those companies default to highly restrictive MTAs (which they would be prepared to accept as-is), but which are unacceptable to institutions like the University of California. This necessitates a period of negotiations, which can significantly delay or abrogate the material transfer. Several of these observations are corroborated by the data for UC Berkeley incoming MTAs, where the time to negotiate is tracked (see Figure 8).

[17] The classification for these organizations was inferred based on their names.

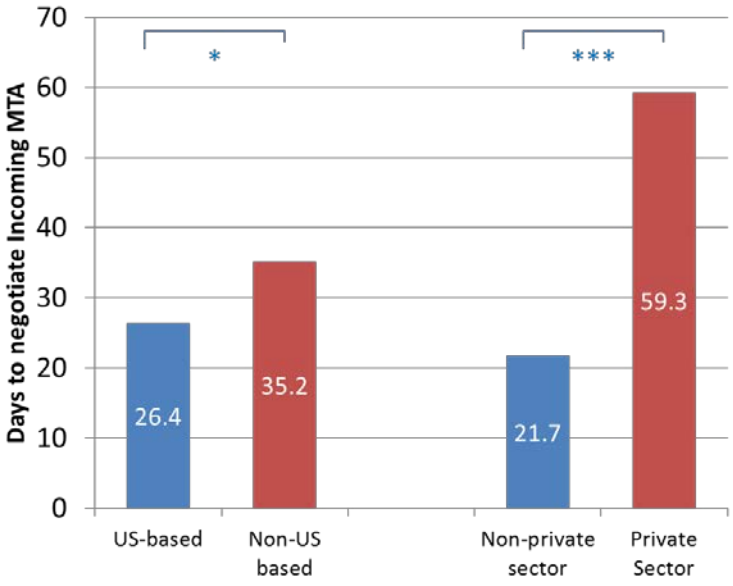

Figure 8: Days to negotiate UC Berkeley Incoming MTAs

This shows that negotiations with US-based entities take 25% less time than non-US-based entities (t-test: 10% significance), and that negotiations with non-private sector entities (governments, non-profits and academia) take 63% less time than those with the private sector (t-test: 1% significance). An analysis of the time taken for individual private-sector MTAs to complete negotiations shows that there is significant dispersion around the mean. Many of these MTAs complete negotiations quite quickly, while almost 40% take 50 days or more to negotiate – a much higher percentage than for the other types of counterparties (see Figure 9).

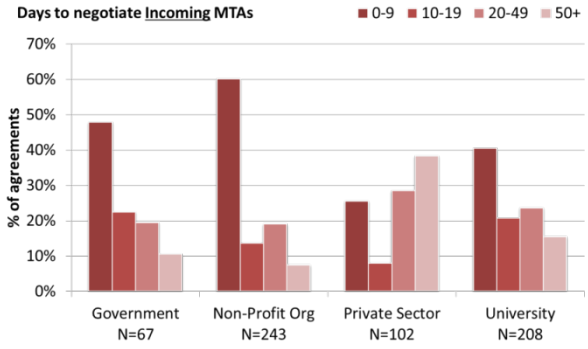

Figure 9: Disaggregation of days to negotiate UC Berkeley MTAs

This differentially longer time to negotiate with the private sector is true for both incoming and outgoing MTAs, which can be seen by looking at the UC Davis MTAs not related to disclosures:

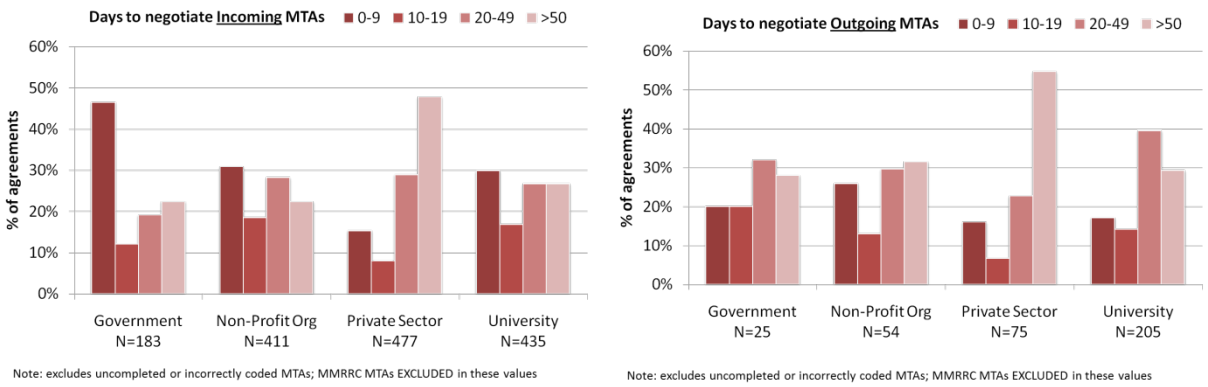

Figure 10: Days to negotiate UC Davis MTAs

For the U.C. Davis data, this finding becomes even stronger if the MTAs from the Mutant Mouse Regional Resource Center (MMRRC) are included. Those MTAs are non-negotiable for both incoming and outgoing materials, and thus can be downloaded and completed very quickly. The bulk of these go to government, university and non-profit institutions, and thus the share of those institutions' agreements that get completed in 0-9 days rises to 80-90% for outgoing MTAs and 35-50% for incoming MTAs, while the negotiation times for private sector MTAs remain virtually unchanged.

The streamlined MTA used by the MMRRC is an interesting institutional development. It suggests that when such a non-negotiable agreement can be created, the cost of transacting with it can be very low – and indeed Furman and Stern (2011) have argued that this formalization has positive impacts on science because it clarifies and regularizes the process of accessing these materials.

An initiative similar to the MMRRC MTA was advocated by the National Institutes of Health in 1995.[18] They promoted the Uniform Biological Materials Transfer Agreement (UBMTA), a "simple letter agreement for transferring nonproprietary biological materials among public and nonprofit organizations. For-profit organizations may also choose to adopt these agreements as well."[19] Discussions with OTT officials suggest that this effort was partially successful; while 487 universities and other groups signed-on to using the UBMTA, few use it unchanged, but instead modify it to suit their own needs.[20]

The contrast between the rapid completion time of MMRRC MTAs and the more protracted negotiations for other MTAs suggests a segmentation of the material transfers. This is valuable since it lowers the transaction costs on agreements whose complexity can be handled by the simpler agreement. A number of on-going initiatives extend this segmentation. For example, the MMRRC has recently moved from requiring an MTA for outgoing material transfer to instead having a short

[18] Association of University Technology Managers (2012).

[19] National Institutes of Health (1995).

[20] Association of University Technology Managers (2012b).

Conditions of Use (COU) statement. Another example is Stanford, who "encourages researchers to share materials with other research colleagues without an MTA when possible."[21] For those that do require MTAs, they have different templates for non-profit organizations and industry.[22] Anecdotal evidence suggests that these policies may be paying off as one UC scientist said that they were collaborating with someone at Stanford because the difficulties of arranging the MTA with a collaborator at another research institution had been too onerous.

2.3.2 Growth of MTAs

There seem to be conflicting trends for MTAs. As Section 2.3.1 argued, material transfers appear to have become more formalized since the late 1970s, with informal letters giving way to contractual agreements. In contrast, institutions like Stanford are opting to exempt certain types of material transfers from requiring an MTA. This makes it unclear whether the net growth in MTAs should be positive or negative. Figure 11 addresses this question. It shows the number of outgoing UC MTAs related to disclosures, aggregated by the year *of the disclosure*. For data truncation reasons, only MTAs in the first 3 years after the disclosure are counted.[23] Hence, the 2001 figure indicates that ~200 MTAs were put in place between 2001 and 2003 on disclosures filed in 2001.

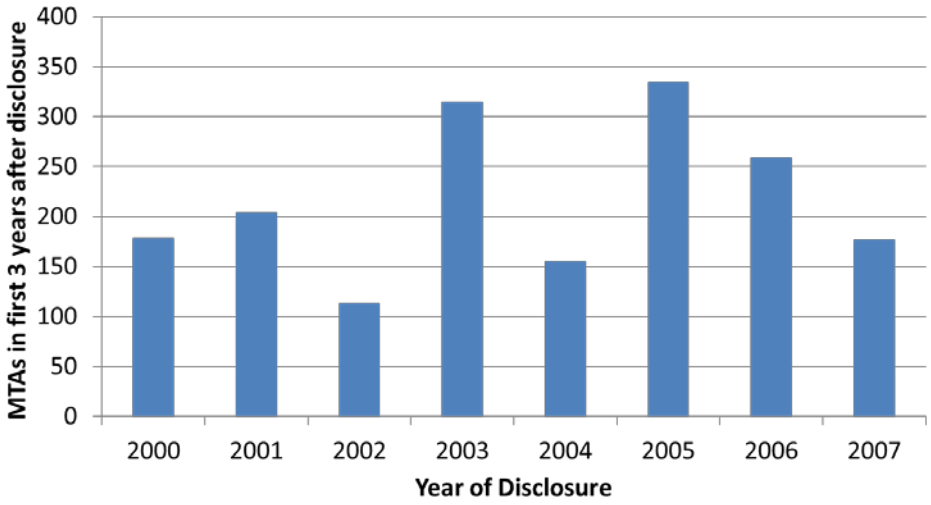

Figure 11: Growth of outgoing MTAs, within 3 years of disclosure, on UC discoveries (all campuses)[24]

This indicates that the may have been some growth in MTAs, but that it is small compared to the year-to-year fluctuations. A simple regression (of the type used in Section 2.1) estimates the growth rate of these MTAs at 4.3% (not statistically significantly different from zero).

[21] Stanford University (2012).

[22] Stanford University (2012).

[23] This ensures even reporting across all years.

[24] In this data, and a single MTA that covers materials from multiple disclosures is counted once for each disclosure.

2.3.3 Composition of MTAs and the connection to Research Tools

As noted by Mowery and Ziedonis (2007), MTAs are disproportionately concentrated in biomedical fields of research. This can be seen both in the groups of scientists using MTAs and in the materials themselves. Figure 12 shows this for UC Davis, with Medicine, Veterinary Medicine, Agriculture & Environmental Sciences and Biological Sciences making up ~85% of all MTAs not linked to disclosures.[25]

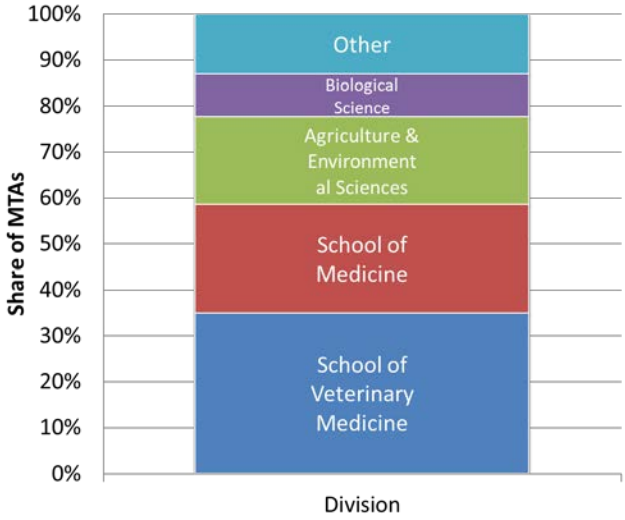

Figure 12: UC Davis division sending / receiving MTAs

A similar pattern can be seen in the list of departments receiving MTAs at UC Berkeley (note: QB3 is the California Institute for Quantitative Biosciences):

[25] This is a slightly different sample than the other results. It includes both initial MTAs and MTA revisions.

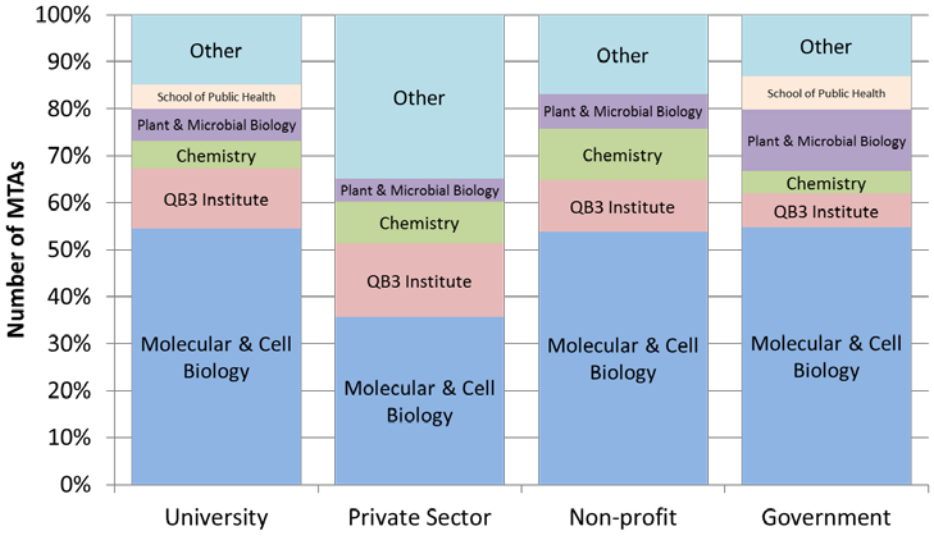

Figure 13: Distribution of counterparties for UC Berkeley incoming MTAs

Having established that MTAs occur primarily in the biomedical disciplines, this section now turns to the question of whether these MTAs are for 'research tools,' which, according to the NIH Working Group Report on Research Tools, can be defined as "the full range of resources that scientists use in the laboratory…the term may thus include cell lines, monoclonal antibodies, reagents, animal models, growth factors, combinatorial chemistry libraries, drugs and drug targets, clones and cloning tools (such as PCR), methods, laboratory equipment and machines, databases and computer software."[26] Because this NIH Working Group definition is broad ("may…include"), a second definition is also useful. Walsh, Cho, and Cohen (2007) define research tools, as indicating "knowledge and material inputs upon which [scientists'] research depends" (p1185).

These definitions suggest that there are two reasonable ways of testing for whether an MTA is on a research tool. The first way is to look whether the material type is listed in the NIH definition. The second, perhaps more compelling, way is to observe whether the material is known to be an input to the research of another scientist. These two ways of identifying research tools are tested for a random sample of 50 MTAs on materials received by UC Berkeley from 2004-2008. Unusually, this data lists both the material being transferred and the usage for which it is intended, allowing both parts of this definition to be tested.

In terms of the materials in sample, 22 (44%) are for DNA/RNA/Plasmids, 16 (32%) are for biological or chemical agents (e.g. cell lines), 8 (16%) are for model animals (e.g. genetically modified mice), 3 (6%) are for data, and 1 (2%) are for something else. These suggests that, based on material composition, 98% 'may' be research tools

[26] National Institutes of Health (1998), p3.

Similarly, the described usages for the material also indicate that these materials are almost all research tools. Of the 50 descriptions provided by scientists for their use of the material, 47 (94%) explicitly describe the material as an input for their research, using phrases such as:

- "[Material] for use in cell culture experiments"
- "The mouse strain will be crossed with mutants…and phenotypic effects explored"
- "The cells will be used as a source of [material] to study the in vitro [cell function]"

A further 2 descriptions (4%) implicitly suggest that they will be used as research inputs, as identified by phrases such as:

- "Will be used to prepare…proteins as described in published literature"

Of the 50 observations in the sample, only 1 (2%) suggested a usage that was not as a research input. Figure 14 summarize these finding from the random sample of UC Berkeley incoming MTAs:

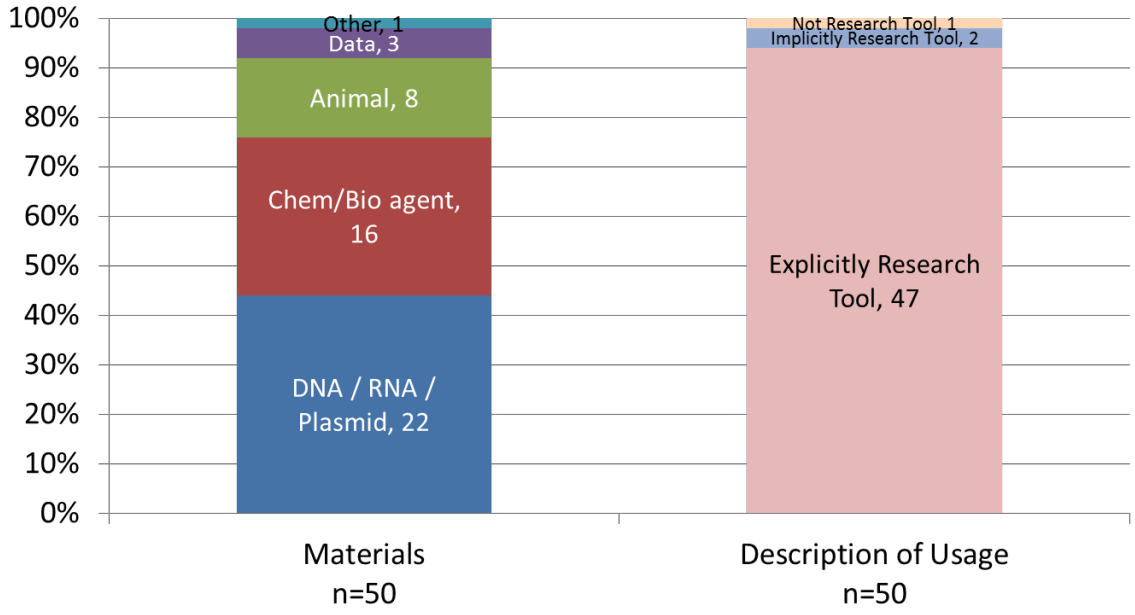

Figure 14: Materials and Usage from a random sample of UC Berkeley MTAs (2004-2008)

This analysis suggests that, according to either a usage definition or a materials definition, the incoming MTAs to UC Berkeley are predominantly for research tools. There are, however, two important caveats to this observation. The first is that the UC Berkeley incoming MTAs may not represent a random sample of MTAs. In particular, since Berkeley does not require MTAs on incoming materials unless the other party wishes one, it is likely to represent a set of materials where these agreements are more important. If UC's policies on this are representative, then the Berkeley sample will over-represent materials that have been disclosed and/or have intellectual property. For the inference in chapter 4 this is a positive thing, since it suggests it more closely represents the

sample being used. However, it also means that these results may not be representative of undisclosed MTAs.

The second caveat to this finding is that while having an MTA strongly suggests that a material is a research tool, it is by no means an exhaustive definition. There may also be many research tools which would not be transferred with MTAs. Thus, this paper takes using MTAs to identify research tools as a definition that is *sufficient* to identify a predominantly research tool group, but at the same time one which is not *necessary*. In other words, the use of an MTA to identify a research tool is relatively restrictive, almost certainly excluding a number of research advances that are research tools but for which no MTA was negotiated with the involvement of a campus OTT.

2.4 Summary

This section has presented a series of datasets from the University of California system. It has attempted to paint a broad picture of intellectual property usage at UC, and a more-detailed one on material transfer agreements, and area about which less has been written. It presents findings that private sector MTAs take longer to negotiate than ones with the government, non-profits or universities, and that international MTAs take longer to negotiate than domestic ones.

Echoing previous research, it finds that biological and chemical materials make up the majority of materials transferred with MTAs. Finally, and most importantly for this dissertation, it argues that the presence of an MTA is a strong indication that a material is a research tool. This is shown using a sample of UC Berkeley incoming MTAs using the definition of the NIH Working Group on Research Tools. This analysis shows that not only would the materials being transferred likely be considered research tools, but that the intended usage of 94-98% of this materials is either explicitly or implicitly for use as a research tool.

3 Patent-Publication Matching

As already mentioned in chapter 1, U.S. universities' have greatly increased their patenting in the past 30 years. This has led some authors, for example Heller and Eisenberg (1998), to worry that there will be negative effects from university patenting on the production of science. If true, this would argue against efforts to maintain or expand university patenting. If not true, evidence demonstrating this might mitigate fears that are delaying further growth in these areas. In either case, the result could have important public policy ramifications.

To evaluate this question, it is important to be able to link the intellectual property being used to the scientific discovery which it is purportedly impacting. In practice this has been a difficult problem. In their paper, Murray and Stern (2007) matched patents to articles published in the journal in *Nature Biotechnology* by reading both patents and the academic articles and using the expert judgment of the reader to link them. In addition to being time consuming, this approach also requires strong expertise on the part of the reader – a problem which escalates if patents from a broad range of disciplines are to be matched.

This chapter presents an alternative, scalable method for matching publications and patents. It then shows that this is a maximum likelihood estimator, and tests its effectiveness on the same sample used by Murray and Stern.

3.1 Data

The patent data used for this analysis is described in Chapter 1, on page 3.

The second source of data, 'Publications data', comes from Web of Science, an internet-based service that tracks the bibliographic information and the citations to and from articles published in 10,000 of the highest-impact journals across 256 disciplines. This Web of Science extract includes the title, author names, journal, publication date, and as well as a number of well-accepted measures of journal quality. The most prominent of these is the 'impact factor', which measures the average number of times an article in that journal is cited in its first two years. This is the measure used in this analysis.

Web of Science also tracks 'forward citations', citations from later works to that publication. This data was extracted for all publications in the data through the end of 2009.

3.2 Inventor-based matching

An alternative to matching based on the content, as done by Murray and Stern (2007), is to match based on the inventors and the timing of their discoveries. This is formalized here as *inventor-based matching*.

Inventor-based matching is based on two assumptions. First, the technique assumes that the inventors listed on patent are likely to be the authors listed on related publications. Since all inventors are legally required to be included on the patent, this assumption relies on the self-interest of those in science to want to be listed on publications, and the integrity of those writing articles to include all who contributed.

The second assumption is that the patent *application* date is likely to occur near the publication date of the academic article.[27] This assumption relies on both the patent and publication coming from a single 'discovery', and that there are commercial and scientific interests in instantiating these products shortly after that discovery. Examples of these incentives might include obtaining an early patent priority date or the fear that a scientific discovery will be published first by another group.

Derived from these principles, it is possible to construct a maximum-likelihood estimator for the publication(s) that best matches a particular patent. This is a multi-stage process, implemented as follows:

1. Start with the patent to be matched, and all of its inventors
2. Search for the publications by each of these inventors in the 2 years before or after the patent was applied for[28,29]
3. Look for overlap in the publication sets of the inventors
4. Choose as a match those publication(s) that have the most inventors listed as authors, conditional on at least having some minimum number of the authors (in case of this analysis: 3)

Thus, if a patent has three inventors, three publication sets (one for each inventor) would be produced and the best match(es) would be those common to all three inventors. Figure 15 schematically outlines this process for the three-inventor case.

[27] In contrast to fields such as economics, the time between submission and publication in biomedical research is often no more than a few months (Murray and Stern (2007))

[28] Because of inconsistencies in the way authors record their first names, testing has determined that the best match is achieved by searching for the combination of first initial and last name.

[29] This ± 2 year period reflects the relative timing of publications and patents observed in the Murray and Stern matching and the replication of it here (described below).

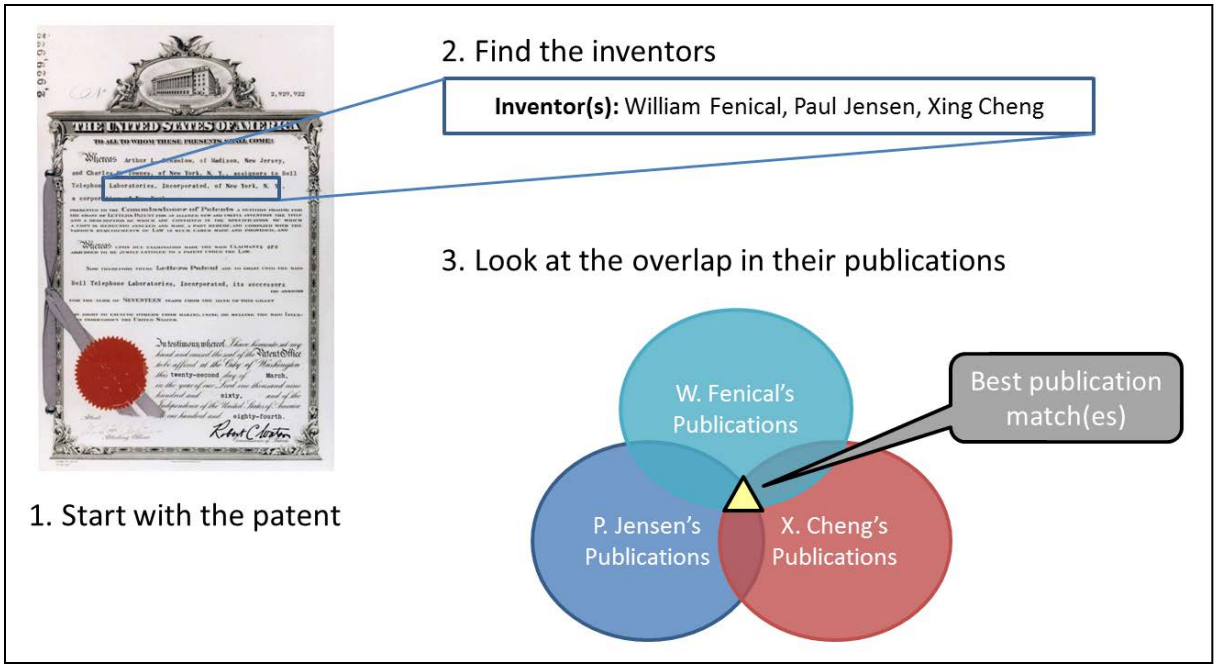

Figure 15: Schematic of Inventor-based Matching

This approach can select multiple publications as 'best' matches, in contrast to the Murray and Stern approach that matches one publication to one patent. The value of this generalization can be seen in an example from the UCOP dataset: a patent on adhesives inspired by the design of gecko feet produces the following matches:

- Adhesive force of a single gecko foot-hair (*Nature*)
- Evidence for van der Waals adhesion in gecko setae (*Proceedings of the National Academy of Sciences*)

Both of these are related to the patent, so comparing the impact of a license on the forward citations of both of them provides additional statistical power.

Using this matching technique does not restrict the sample to only instances where all the inventors are listed as authors on the publication. For example, in the case above, this would occur if a lab technician had also been included on the patent, but had not been listed on any of the academic publications. Under such circumstances the inventor-based matching algorithm will choose the publication(s) with the maximum overlap possible – in this case a publication listing three of the four inventors since there are no four-out-of-four-inventor matches.

This 'best-available' property is a general property of maximum-likelihood estimators, of which this is one. That this is a maximum likelihood estimator can be seen by formalizing the assumptions that defined the matching algorithm. First, assume that a publication and a patent are more likely to be a match if they share an author, i.e.:

$$p\left(match_{pub_i \,\&\, patent_j} \middle| author_k \in \left(authors_{pub_i} \cap inventors_{patent_j}\right)\right) \geq p\left(match_{pub_i \,\&\, patent_j}\right)$$

Here $authors_{pub_i}$ and $inventors_{patent_j}$ are the sets of authors for publication i and the inventors for patent j respectively. Combining this over all inventors yields that $pub_m$is a 'match' for $patent_j$ if

$$m \in \underset{i}{\text{argmax}} \prod_{i,k} p\left(match_{pub_i \,\&\, patent_j} \middle| author_k \in \left(authors_{pub_i} \cap inventors_{patent_j}\right)\right)$$

As highlighted above, an implication of this method is that a single patent can be associated with more than one publication. This occurs precisely when multiple publications share the same level of overlap between the inventors, and when no publications have a greater overlap.

3.2.1 Limits of Maximum Likelihood Estimators

As with all maximum-likelihood estimators, just because an estimate is 'best' doesn't mean it is precise.[30] In this case, the precision will be low if there are many publications that are not a match, but which have a similar level of inventor overlap to the correct matches – as might happen if the inventor overlap on the correct publication is very low. For example, the matching algorithm would produce many incorrect matches if, on a four-inventor patent, no related publications listed more than a single inventor among the authors. In this case, the algorithm would theoretically identify as matches *all* publications by *all* of the inventors in the relevant five-year window.[31] To avoid such errors, matches are restricted to only high-precision estimates. This is accomplished by adding the proviso to Step 4 of the algorithm that includes only publications that list three or more of the inventors. The logic behind this criterion is illustrated in Figure 16, which portrays the number of publications matched to each patent in the dataset. The number shown in each bar is the number of inventors listed on both the publication and the published paper.[32]

[30] See, for example, Casella and Berger (2002) for a discussion of maximum likelihood estimators

[31] The five-year window includes the year of the patent application, the two years prior and the two years afterwards.

[32] The data in Figure 4 include only papers and patents linked by at least two inventor or author names.

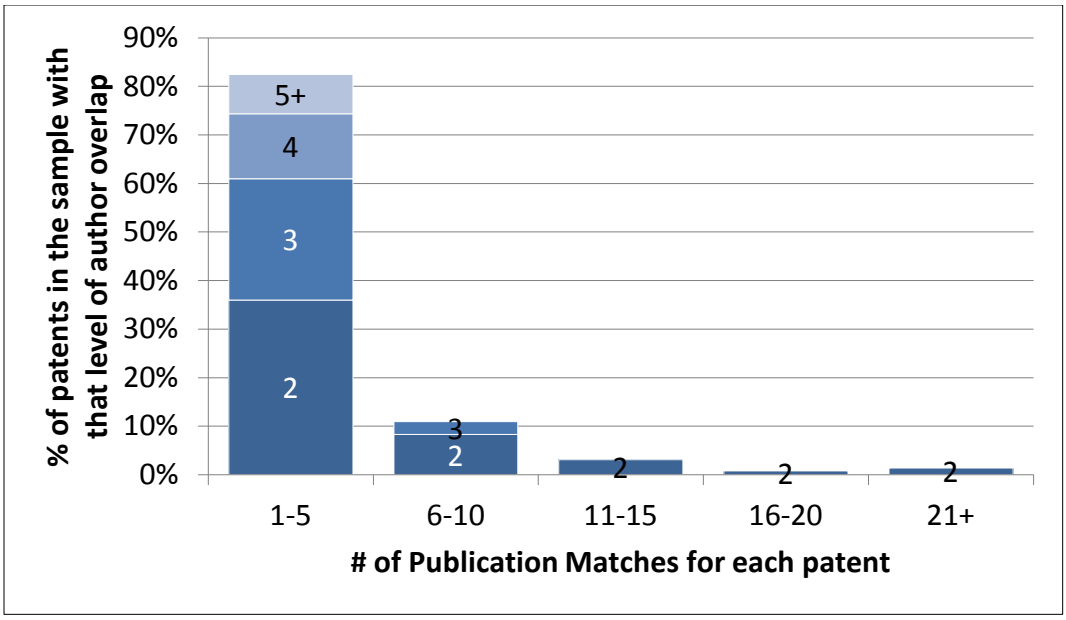

Figure 16: Publications matched to each patent

Figure 16 shows that 82% of the patents in the sample are linked by the 'inventor-based matching' algorithm with 1–5 publications, while the remaining 18% of the patents are associated with six or more publications.[33] One likely cause for the large number of publications associated with each of the patents in this 18% is common scientist names (e.g., "Professor J. Smith"). Figure 16 also demonstrates that the majority of these patents with more than 5 publication matches have only two names that are common to both the inventor list and the author list. This suggests that these observations may be low precision ones that should excluded. The statistical implications of dropping publications with an author overlap of '2' is to restrict the sample to a higher expected probability of a match[34], that is, to exclude publications whose expected probability of a match is 'too low'. It is tempting to argue that '3' author overlap papers should also be excluded based on this same rationale. This is not pursued here because it would decrease the sample size considerably, but it would be an interesting extension for someone a larger dataset. Nevertheless, the sample statistics that would result are presented to explain the logic used in restricting the dataset.

[33] Note that publications listing only a single inventor are already excluded from this analysis

[34] This follows directly from the definition of the maximum-likelihood estimator above

Samples		
Inventor Overlap	3	4
Sample Size		
Publications (000)	1.7	0.6
Patents (000)	0.7	0.3
Publications with MTAs	261	79
Publications / Patent	2.4	1.8
Observations in Life Sciences	49%	44%
Sample Statistics[#]		
Citations per year	11.4	16.2
	(26.1)	(36.3)
Average Impact Factor	8.7	11.0
	(8.4)	(9.8)
Publication year	2000.7	2000.4
	(2.6)	(2.8)
Publication Age	3.2	3.3
	(2.6)	(2.7)
Age at MTA issuance	2.4	2.6
	(2.7)	(2.6)
Age at Patent issuance	3.5	3.5
	(2.0)	(2.0)

[#] values in the parentheses is 1 standard deviation

Table 1: Sample Summary Statistics

Table 1 examines the effect of restricting the sample to higher levels of inventor-overlap by comparing the sample statistics of 3-inventor overlap with 4-inventor overlap. The average year of publication, age of publication when the citations are observed, and the proportion of the papers in the Life Sciences (principally Biology, Biochemistry, and Medicine) are relatively stable across the samples. This suggests little or no introduction of bias along these dimensions.

In contrast, two measures of publication quality: the number of citations per year and the average impact factor of the publication's journal do rise in samples with less noise. This is consistent with the observation that patented publications are of higher quality (i.e. receive more citations) than unpatented ones. Since correct matches are patented publications, but incorrect matches are random additions from the general pool of publications, the effect of a removal of incorrect matches should be to increase average publication quality. Thus, these results are consistent with the argument that higher-overlap specifications reduce noise in the sample, although it is not definitive. Summary statistics for a Life Sciences sub-sample (biology, biochemistry and medicine) show these same trends.

After restricting the data to 3+ inventor overlap, 728 patents and their corresponding publications remain in the sample.

3.2.2 Testing the quality of the patent-publication matching

Having established the methodology of inventor-based matching, it is important to assess its validity. As mentioned earlier, Murray and Stern's patent-paper pair sample was developed through hand-matching publications and patents based on the scientific content of each. As such, their matches provide a useful benchmark against which to measure the performance of inventor-based matching. The gracious cooperation of Murray and Stern makes this comparison possible.[35] As Figure 17 shows, of the 170 patent-publication matches found by Murray and Stern, the new method identifies the identical 'best' publication match for 95% of their sample. In an additional 4% of these instances, the Murray-Stern match was found, but another publication was deemed a better match.[36] In the remaining two cases the algorithm failed because of missing data on the inventor's name or because an author sometimes used his first name and other times used his middle name on publications.

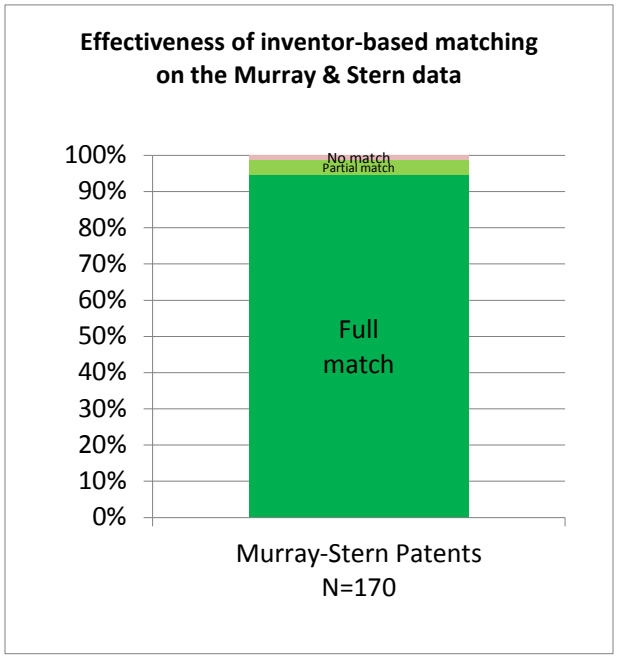

Figure 17: Validating Inventor-based Matching

This benchmarking exercise demonstrates the ability of inventor-based matching to identify the best publication matches for patents. Having established its effectiveness, it is worth noting several other advantages of this method over that employed by Murray and Stern (2007):

[35] Thanks to Professors Fiona Murray and Scott Stern for sharing their dataset.

[36] The difference here is likely because of the direction of matching. Murray and Stern began with a set of publications and found the most-similar patent, whereas this analysis begins with a patent and finds the most-similar publication. This means that it is possible that the patent identified by Murray and Stern to have been the best patent match for that publication, but also that another publication existed which is an even better match. A non-expert review of the 'better' matches suggests that they are well matched.

1. It does not impose a simple one-to-one relationship between patents and publications;
2. It makes it possible to analyze many scientific fields without requiring a domain expert in each;
3. It is a transparent and reproducible matching procedure that does not rely on the gifted intuition of a small group of researchers;
4. It allows the matching process to be automated, making much larger sample sizes feasible.

3.3 Patented Publications

Having outlined how inventor-based matching works and validating its effectiveness on the Murray and Stern data, this section applies this method to construct a set of *patented publications* – that is, publications where the underlying discovery also resulted in a patent.

As mentioned above, to get a more-precise estimate of correct patent-publication matches, a publication must have 3+ of the inventors listed as authors. This imposes two restrictions on the sample. First, a patent must have at least three inventors. Second, the associated publication must list at least three of those inventors as authors. Figure 18 summarizes the impact of these restrictions on the sample:

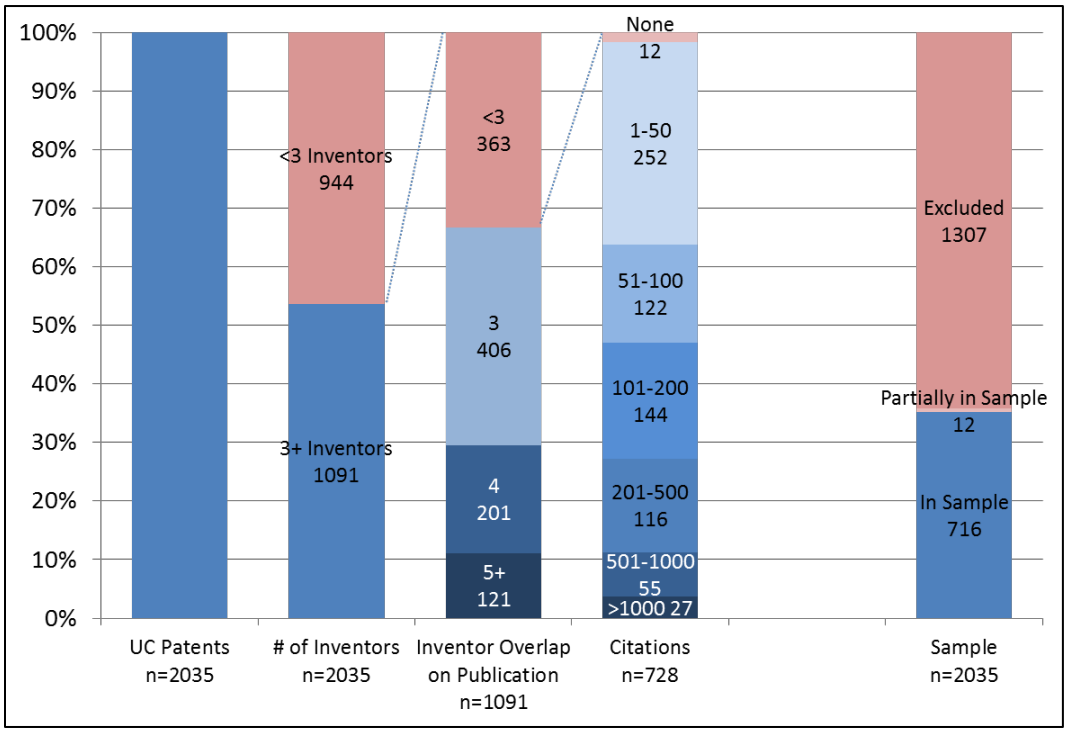

Figure 18: Sample composition

Thus the 3+ inventor restriction excludes 944 patents that have only one or two inventors listed. An additional 363 patents are excluded because, even though they have 3+ inventors, fewer than 3 are listed as authors on any publication. The fourth bar of Figure 18 shows the number of citations per patent, i.e., the number of journal citations for all publications that are matched (using the three-

name overlap restriction) to that patent. Notice that some have no publication citations. While these are included in the sample, they will not exhibit any changes in their citation patterns due to intellectual property, and thus in the final bar they are designated as "Partially in Sample".

The example shows that, despite the strong restriction of having 3+ inventors listed as authors on the publication, more than one-third of patents can be matched using inventor-based matching. This highlights the value of doing this matching automatically, since this sample is more than four times as large as the Murray and Stern hand-matched sample, and covers not just one journal in one discipline, but the 10,000 journals of Web of Science across many different disciplines.

3.4 Summary

Despite the importance of looking at the impact of intellectual property on academic science, it has been difficult to analyze these effects since the matching of intellectual property to scientific publications has been a laborious process requiring field-specific expertise.

This chapter introduces an alternative method of forming these patent-publication matches, called *inventor-based matching*. This method is tested against the hand-matched sample of Murray and Stern, and shows a high level of success in reproducing their matches. It also generalizes from their requirement of a one-to-one match.

Finally, by automating this process, inventor-based matching allows the processing of larger datasets of patents without the intervention of domain experts. Using this method on data from the UC system and Web of Science produces a set of over 700 patented publications that can be used to examine the effects of intellectual property on citations to academic publications.

4 The Effect of Licensing on the use by Researchers of UC Patented Publications[37]

This chapter seeks to empirically address the question of whether licensing has a positive or negative effect on scientific communication of the underlying idea. In keeping with the work of Murray and Stern (2007), this chapter operationalizes this 'flow of scientific knowledge' as the number of times a scientific publication is cited by subsequent articles in scientific journals.

Previous research on academia's use of intellectual property has included examinations of patenting, licensing and of material transfer agreements.[38] Some work has examined the effects of patenting on biomedical researchers' willingness to share information on their work (Blumenthal et al., 1997; Campbell et al., 2002). More recent research has analyzed the effects of patenting biomedical discoveries that are also disclosed in scientific papers. Some of this work finds that the issuance of a patent results in modest but significant declines in citations to the research papers related to the patent (Murray and Stern, 2007; Sampat, 2005). Other research, however, finds that biomedical researchers report rarely, if ever, searching to determine whether a prospective research project or experiment will infringe on patents (Walsh, et al., 2005; Lei et al., 2009).

Licenses on university-patented discoveries affect commercialization efforts by companies, and therefore may influence corporate R&D in related areas. Unlike patents, licenses are not published or otherwise subject to mandatory disclosure, and in many cases the identity of licensees is treated by university technology transfer offices as confidential. Why might licenses affect the behavior of academic researchers in formulating their research agenda?

This issue has received little empirical attention from scholars. A recent analysis of patent citations to University of California patents that were licensed (Drivas, et al., 2011) found that citations to these patents by non-licensees increased after exclusive licenses (either by geographic area or field of use) were issued on these patents. Drivas et al., interpret the increase in citations as a reaction by other patent applicants to the demonstration of potential commercial value signaled by the negotiation of a license for the patent. A similar signaling effect could increase citations to patented publications. In this case, the issue of a license "demonstrates" that a particular area of research has potential scientific or commercial value, leading other investigators to pursue work in closely related fields. Indeed, it is plausible (as various scholars have speculated, with limited evidence thus far) that in the wake of the Bayh-Dole Act, academic researchers may choose research areas based on

[37] Because of the strong contributions of co-authors to this introduction, parts of which are drawn directly from Thompson, Mowery, Ziedonis (working paper), the author would like to acknowledge them explicitly.

[38] Material Transfer agreements are bilateral contracts that accompany some transfers of research materials. They are typically signed by the sender of the materials, the receiver of the materials, and both of their institutions. An example would be a contract accompanying the transfer of a genetically modified mouse from its designer to another researcher, perhaps someone involved in clinical trials. The terms of MTAs can be simple and straightforward (e.g. the UBMTA), or complicated and onerous (see, for example, those discussed in Walsh, Cho, and Cohen (2007).

their potential for private profitability. Regardless of whether a license "signal" operates through its effects on perceptions among researchers of scientific or commercial potential, this argument predicts an increase in citations to patented publications following the negotiation of the license.

Equally plausible arguments, however, can be developed to predict a chilling effect of licensing on scientific communication. Reactions by university technology licensing offices and/or their licensees to any evidence of patent infringement (even for research purposes, inasmuch as the research exemption from such infringement suits remains unclear) may be swifter and stronger in the case of patents that are licensed.

As a result of these competing effects, this chapter is agnostic on the likely sign of any effect of licenses on scientific communication associated with publications linked to licensed academic patents. Indeed, both effects may be present for papers in various fields of research, and hopefully this work will shed light on the net magnitude of any offsetting effects.

Despite plausible arguments for increases or decreases in citations following a license in general, there is an area where the negative effects might be more pronounced: research tools. Research tools are discoveries where the output of one scientist is the input for another. Fiona Murray (2006) provides an excellent description of what happens with one research tool, the Oncomouse, and why the licensing of research tools can be problematic:

> "In 1984, scientists at Harvard University carefully engineered a new mouse to have a predisposition to cancer, the Oncomouse…The Harvard researchers…patented their creation and subsequently licensed this patent to DuPont...
>
> [DuPont] set a high price per mouse…placed restrictions on breeding programs…demanded publication oversight… [and] insisted upon a share of any commercial breakthroughs made using the Oncomouse."

It is easy to imagine these types of restrictions could have a chilling effect on follow-on researchers wanting to use the Oncomouse. This is made worse because licensees may have an explicit incentive to restrict access to these materials. The NIH Working Group on Research Tools (1998) outlines the rationale behind this restriction:

> "If the sponsor or licensee plans to develop the research tool as a commercial product for sale to researchers, it may be unwilling to permit the university to undercut its position in this particular market by giving the tool away to potential paying customers."

Thus, for research tools, there is a stronger case that licensing may have a negative effect on citation rates.

4.1 Methodology

This section describes how the sample is constructed, and how research tools are identified. It then outlines the identification strategy used to test the effect of licenses on the flow of scientific knowledge.

4.1.1 Sample Construction

This analysis restricts attention to "patented publications". These are publications that have value to science, and thus get published in scientific journals, but where the underlying discovery also has commercial potential, and thus gets patented. A patented publication, therefore, is a scientific publication based on an invention disclosure that produced a successful patent application. Rather than repeating this at each instance, the shorter moniker of "patented publications" is used.[39] Within this group of patented publications are those that also received licenses, and those that didn't.

The advantage of restricting the comparison to within patented publications is that they are likely to be more similar in quality and other characteristics than would be true of a broader sample of publications. In particular, because these discoveries are all patented, differences in commercializability are considerably lower than they would be in a general sample. Despite this greater similarity, one might imagine that the existence of a license suggests other unobserved differences as compared to patented publications lacking them. For example, patented publications with licenses might be of a higher research quality, which could lead to more citations. Controlling for these differences is subtle and is discussed at length in the identification strategy section.

While restricting the scope of this analysis to patented publications helps ensure a more-homogeneous sample, it also means that the estimates measure the effect of a license issuance on the citations of patented publications, rather than on publications in general. Future work is planned to investigate this difference between patented and unpatented publications.

Building a sample of patented publications with and without licenses requires a number of steps. Firstly, the connections between patents, MTAs and licenses are needed. Fortunately this is tracked directly by the University of California on system-wide "invention disclosures," i.e., the declaration by the university researcher of a potentially patentable advance. Secondly, patents needed to be connected to publications on the same discovery. This is done via inventor-based matching, as described in Chapter 2. Finally, the connection between a publication and those citing it is needed. This is extracted from Web of Science, which tracks it.

[39] Patented publications were inspired by the patent-paper pairs of Murray and Stern (2007), and are very similar. However, their term is not used because they focused on one-to-one relationships between patents and publications, while this work allows multiple matches - and thus the data are not 'pairs.'

4.1.2 Identifying research tools

As previously discussed, there is some basis for believing that research tools may experience a negative impact from being licensed. Testing this requires identifying which patented publications are research tools, and which are not. This has proven to be a difficult question since, in theory, almost any research output could be the input to another scientist's research. For Murray and Stern (2007), their solution to this dilemma is to focus on the 3-digit patent classes for "Chemistry: Molecular Biology and Microbiology" (class 435) and "Multicellular living organisms and unmodified parts thereof and related processes" (class 800)[40]. This definition is tested in Section 4.3.3. This paper adopts an alternative definition for identifying research tools. It argues that the presence of a material transfer agreement is a strong indication that the material is a research tool. This is supported by the analysis in Chapter 2, which shows that the materials being transferred, and their intended usage, both correspond to the NIH Working Group on Research Tools' definition. Because, for example, a genetically-modified mouse is likely to be a research tool even before its first MTA is observed, an observation is deemed a research tool during all periods if an MTA is observed at any point.

4.2 Building the right treatment and control groups

4.2.1 Specification challenges

Designing the right specification to analyze the impact of licensing on scientific communication is challenging, owing largely to the complicated shape of citation curves. Publications all begin with zero citations and then begin to accumulate them, but this may be the only universal trait. Typically, citations will peak and then return to zero. But for some publications, the number of citations it receives continues to grow throughout the period of the data. For those that do peak and then fall, the timing and rate of the ascent and descent may vary. As a result of these differences, constructing a suitably flexible parametric model is difficult, leading to the worry that a poorly-modeled group of observations could bias the results.

The extent of this modeling challenge can be understood by considering just some of the factors that would be involved in a reasonable parametric model. A starting place is the assumption that publications in more highly cited journals would accrue more citations, which would argue for including *Journal Impact Factor* as a control variable. Similar arguments could easily be made that the academic discipline (hereafter *Journal Subject*) could also drive citation patterns, as could how much better cited a publication is prior to the license (*Citations in t-1, Citations in t-2*). Of course, for each of these the effects could be non-linear, which would argue for the inclusion of higher-order terms (e.g. *Citations in t-1 squared*). Interaction terms between these variables would also be important, since for example, effect of a journal impact factor rating is likely to be different on disciplines with more compressed citation patterns. And, since the effect of any of these things will

[40] USPTO (2011).

impact a publication differently 3 years after publication than 10 years afterwards, each of these would need to be interacted with age fixed effects. The combination of all of these factors would result in an enormous number of parameters for the model, far more than the data could reasonably estimate. This argues against using a typical linear, or generalized linear, framework, in favor of a *non-parametric* approach. This weakens the linearity assumptions needed for the model to be correct, and helps account for the many plausible interaction effects that could be present.

4.2.2 Nearest-neighbor matching

The specific non-parametric method used is *nearest neighbor* matching. This method begins with a potential treatment observation, in this case a patented publication that gets licensed. It then searches through those patented publications that never receive a license to find one that looks similar to it in terms of pre-treatment covariates. If such a match is found, the licensed observation is added to the Treatment Group, and the match is added to the Control Group. If multiple matches are found, the one most resembling the treatment observation (i.e. the 'nearest neighbor') is chosen.[41] Finally, if no match is found, the potential treatment observation is discarded and not used in the analysis.[42] Since this method explicitly builds the sample with treatment and control observations with similar observable characteristics, it should produce good covariate balance on these dimensions. This is tested explicitly in the Results section.

Because this matching process begins with the treatment observations, the coefficient should be interpreted as an average treatment effect on the treated (ATT), that is, it is the average effect on citations from having a license issue *on patented publications that are like the ones that actually do receive licenses* (in contrast, for example, to what would happen to an average publication or even an average patented publication). By restricting the sample to control observations that 'match' the treatment observations, those control observations that are dissimilar are excluded, thus decreasing the sample size.[43]

---

[41] In early analyses an n-nearest neighbors approach was also tested, where the n best control observations that match the treated observation were chosen and assign each a weight of $\frac{1}{n}$. This produced no notable change from the single nearest neighbor approach.

[42] While discarding potential observations might seem to be a bad thing, here it is being done precisely because there are no counterfactuals observed in the data which would allow for a reasonable comparison. Statistically speaking, the observation lies outside the range of *common support*, and should thus be excluded. Despite this sound statistical footing, it is important to note that a side-effect of doing this is that, when the pool of "control" observations becomes small, the possibility of finding a match becomes less likely. This constrains how detailed this analysis can examine subsamples of this data.

[43] The effect that this has on the precision of the estimates is ambiguous. Smaller samples will tend to reduce precision (making the standard errors larger), but there may also be a countervailing impact because the smaller sample will likely be more homogenous, which could increase precision.

The actual matches are found using the R statistical library "Matching" by Jasjeet Sekhon (Sekhon, 2011), using the functionality that searches for the best matches using a genetic algorithm.[44]

4.2.3 Criteria for Matching

For the nearest neighbor matching two sets of variables are used, one set where the control observation's characteristics must have an exact match with the treatment observation's, and another where a nearby match is sufficient. The set requiring an exact match are:

- Publication Age: number of years since the paper was published
- Journal Subject: academic discipline of the journal
- Patent Granted (Yes/No): whether the related patent has been granted
- MTA Issued (Yes/No): whether the paper has an associated MTA at the time of the license

Even these restrictions imply a great deal of similarity between the treatment and control observations. For example, a treatment observation in the life sciences with an issued patent and no MTA would be compared with a control group observation in the life sciences with an issued patent and no MTA, and the comparison would be during the exact year that the treatment observation had received the license (e.g. 3rd year after publication).

For those treatment observations that match on the 'exact' characteristics, the match is chosen by picking the nearest neighbor based on their relative proximity in the following five characteristics:

- Journal Impact Factor
- Publication Year
- Citations in $t-2$: citations 2 years before the treatment
- Citations in $t-1$: citations 1 year before the treatment
- Slope of Citation curve between $t-2$ and $t-1$

In each of these dimensions the furthest a 'nearest' neighbor is allowed to be is one standard deviation.[45] Beyond that, the observation is judged to be ineligible. Of course, virtually all matches will be closer than one standard deviation, since the nearest one is being chosen.

Collectively, these restrictions mean that for each observation in the treatment group, there is one in the control group that matches it, either exactly or within one standard deviation, on each of these nine important characteristics.

---

[44] According to Sekhon (2011), "GenMatch dominates the other matching methods in terms of MSE [Mean Squared Error] when assumptions required for EPBR [Equal Percent Bias Reduction] hold and, even more so, when they do not".

[45] The distance limit of an acceptable match thus is the "caliper" of the Matching. Using a caliper helps exclude both observations whose observable covariates would make them outliers and those which would make them inliers, that is observations that are in the 'middle' of the data, but nevertheless lack a comparable control observation (Sekhon, 2011).

4.2.4 Difference-in-differences Estimator

The effect of a license on the citation pattern of the treatment group is estimated using a difference-in-differences estimator. This compares how the citations to one patented publication increase (or decrease) following the issuance of a license with the changes in citations for a comparable publication that lacks a license. In this case, it is defined as:

$$Treatment\ Effect = (Citations_{t+1} - Citations_{t-1})_{pub\ w/\ License} - (Citations_{t+1} - Citations_{t-1})_{pub\ w/o\ License}$$[46]

The definitions for the treatment effect in period $t+2$ and $t+3$ are calculated similarly, by replacing $Citations_{t+1}$ with citations in the new 'after' period. In all cases, the 'before' period remains $t-1$.

Using a differences-in-differences estimator for this analysis rules out bias from changes that impact the before and after periods similarly. For example, if a particular scientist is well reputed and thus gets 5 more citations per year, every year, than this will be added to both $Citations_{t-1}$ and $Citations_{t+1}$ and the impact on the estimate will be zero. In this way, the estimator accomplishes the same effect as author or publications fixed effects. This also means that the estimate is robust to these types of *unobservable* differences in addition to the *observable* differences controlled for using matching.

Because the distribution of citations is skewed (see Figure 18), there is a risk that outliers that will drive the estimation results. Accordingly, the 2.5% highest and lowest observations are excluded from the analysis. Diagnostic tests are also run after the analysis to check for the results being driven by outliers.

4.2.5 Regression adjustment

Even after matching, it is still possible for the characteristics of the treatment group and the control group to differ slightly in the dimensions where only nearest neighbors were found. For example, if the matched control observations were all in slightly lower impact factor journals than the treatment observation. When such differences remain, *covariate bias adjustment* is used to control for the differences between the groups. This applies a multivariate linear regression (hereafter "regression adjustment") on the post-matching sample (treatment observations plus their matched controls).

To summarize, covariate balance is obtained using two techniques. First, matching produces nearest neighbor control observations for the treatment observations. This non-parametric technique means that fewer assumptions are needed about the parametric form of the effect than would be required if a linear, or generalized linear (e.g. negative binomial) formulation was used.

---

[46] This only conditions on not having had a license *prior* to that time-point. This is because future licensing may be an outcome of either having or not having a license in this period. As such, conditioning on post-treatment licensing could induce bias.

Secondly, a multivariate linear regression (using the same covariates used in matching[47]) is done to adjust for any remaining differences between the groups. Because this is done on the matched sample, the differences between the treatment and control groups are likely smaller, making the linearity assumption embedded in least-squares more plausible, although clearly still subject to the same criticisms leveled above. Rubin (1979) discusses the value of using these (slightly modified) techniques and concludes that "pair-matching coupled with regression adjustment on the matched pairs is a quite effective general plan for controlling the bias due to matching variables, and this combination is clearly superior to regression adjustment" (p.318).

4.3 Results

The results are divided into three sections. The License Effect covers the effect of a license on the full set of matched patented publications in the data. The License Effect on Research Tools covers the effect of a license on a subsample of matched patented publications where all the observations, both treatment and control, are research tools (as identified by their receiving an MTA at some point). A third set of results then shows the effect of using the Murray-Stern research tools definition, rather than the one based on the presence of an MTA.

4.3.1 License Effect

Since the identification strategy depends on a balance on observables, the covariate balance between the treatment and control groups needs to be explicitly tested. If matching has done a good job, the summary statistics for both groups should be similar. Two types of tests are used to confirm this: *t-tests* to compare the means of each group and Kolmogorov-Smirnoff tests (*KS tests*), to compare the entire distributions. The null hypothesis that they are the same for the treatment and control groups, and the alternative hypothesis is that they are different. Table 2 shows the mean of each group for these control variables, and the statistical significance of the test (***, **, *, and '-' implying 1%, 5%, 10%, >10% significance levels, respectively).

[47] License effect: $\Delta Citations = \beta_0 + \beta_1\, Age + \beta_2\, Journal\ Subject + \beta_3\, Patent\ Granted + \beta_4\, MTA\ Issued_{YN} + \beta_5 Journal\ Impact\ Factor + \beta_6\, Publication\ Year + \beta_7\, Citations_{t-2} + \beta_8\, Citations_{t-1} + \beta_9\, CitationsSlope_{t-2\ to\ t-1} + \psi\, License$. The coefficient of interest is ψ, which is what is presented in the results.

	License Sample			
Covariates	Mean Treated	Mean Control	T-Test	KS-Test
Publication Age	2.0	2.0	-	-
Journal Subject[48]	4.3	4.3	-	-
Patent Issued	0.56	0.56	-	-
MTA Issued	0.04	0.04	-	-
Journal Impact Factor	7.3	7.1	***	***
Publication Year	2000.7	2000.5	*	-
Citations in t-1	6.6	6.1	-	***
Citations in t-2	4.3	4.2	-	-
Citation Slope from t-2 to t-1	2.3	1.9	***	***

Table 2: Covariate balance for the full licensing sample

Because the first four variables are matched exactly, it is not surprising that the means are the same and that neither the t-test, nor the KS-test show any difference, since these are true by construction. For the remaining variables, the means for the control and treatment groups are very similar. Despite this, some of these differences are statistically significant, suggesting that it will still be important to check the results from the regression-adjusted analysis.

Because the control and treatment groups are so similar pre-treatment, the main finding is observable just by looking at the evolving citation pattern for the treatment group (blue) and the control group (red), as shown in Figure 19. It shows that patented publications receive more citations starting two years after a license than do a control group that does not receive the license.

[48] Journal subject here is a categorical variable, with each subject mapped to a random integer. Therefore the 4.3 listed has no literal meaning, but the equality between treatment and control is still meaningful.

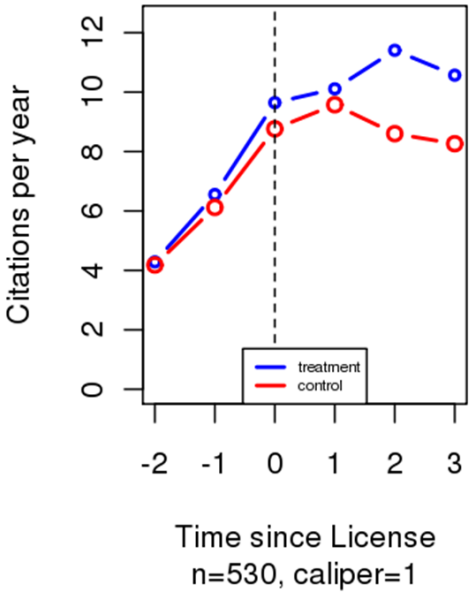

Figure 19: License effect

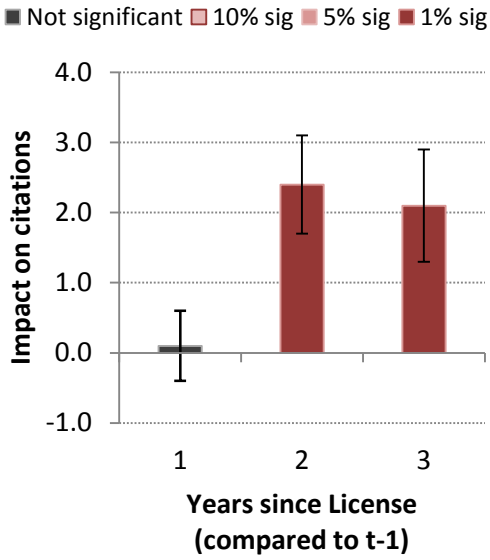

Figure 20: Regression results from the License Effect[49]

This result is confirmed in Figure 20, which shows these results after doing the regression adjustment. It shows a 2.4 citation increase in year 2 and a 2.1 citation increase in year 3, both of which are significant at the 1% level. Notice that these are only slightly different than those implied by the citation trends because of the excellent covariate balance even prior to the regression adjustment. The magnitude of these increases implies that the average publication receives a ~25% increase in citations for these two years as compared to the control group.

It is interesting to note that the increase in citations does not occur immediately, but is delayed 2-3 years after the license issues. This is consistent with licenses being a positive signal to the field, since it takes time for scientists to adjust their research agendas to increase their focus on this area. The delay is also consistent with many licenses not being publicly announced, and thus one would expect it would take time for the news of the announcement to spread through the community.

4.3.2 License Effect on Research Tools

To limit the analysis to research tools, the sample is narrowed to look only at patented publications that are research tools (defined as disclosures associated with a material transfer agreement).[50] Thus, both the treatment and the control observation have MTAs,[51] and the difference remains that

[49] Error bars are ±1 standard error

[50] In principle one might want to restrict this to instances where there is an MTA *prior* to the license, to guard against any reverse causality. That approach is not adopted here for two reasons. Firstly, it is reasonable that even before a material like an oncomouse has its first MTA, it represents a research tool. Secondly, segmenting the sample into MTA before / after leads to results that are similar in direction, although with much smaller sample sizes and thus more noise.

[51] Because exact matching is done on *MTA Issued, the* treatment observations with MTAs prior to license are matched to controls that also already have an MTA, while those that do not yet have one and matched to those that also don't yet

one gets a license and the other does not. The following table presents the covariate balance achieved on this sample:

	Research Tools Sample			
Covariates	Mean Treated	Mean Control	T-Test	KS-Test
Publication Age	1.9	1.9	-	-
Journal Subject	3.9	3.9	-	-
Patent Issued	0.5	0.5	-	-
MTA Issued	0.3	0.3	-	-
Journal Impact Factor	5.2	6.0	-	*
Publication Year	2001.4	2000.5	***	-
Citations in t-1	5.4	5.0	-	-
Citations in t-2	1.2	1.3	-	-
Citation Slope from t-2 to t-1	4.3	3.8	-	-

Table 3: Covariate balance on the research tools sample

Notice that even though the differences between the means of the treatment and control groups are larger in absolute magnitude than for the full sample, the statistical difference is less significant, reflecting more noise in this smaller sample.

As before, both the unadjusted citation curve for the matched samples and the regression results are presented below.

have one.

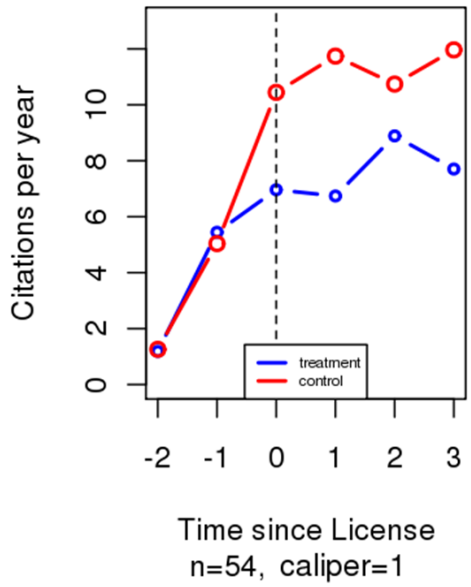

Figure 21: License effect on research tools

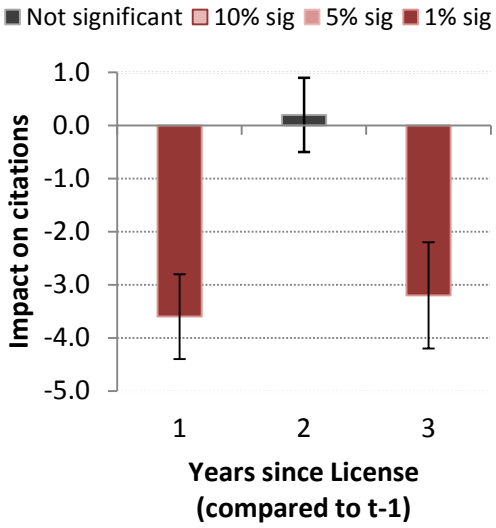

Figure 22: Regression results from the License Effect on research tools[52]

Here the effect is the opposite of that observed in the overall sample. Instead of licensing being associated with an increase in citations, there is a drop. The covariate adjusted values show a highly statistically significant drop of magnitude -3.6 in year t+1 and -3.2 in year t+3, but no statistically significant effect in year t+2.[53] These are big effects, particularly in light of the positive effects observed for the other sample. They represent decreases in citations of approximately 30% for the average publication, although the small sample size argues for caution in interpreting the exact size of the effect too precisely or too broadly.

Both the direction and the onset speed of this effect are consistent with a more direct effect than that observed for licenses in general. One mechanism which would suggest a more direct mechanism is the denial of material requests. This would be consistent with the delays and project abandonment observed in the surveys by Walsh, Cho and Cohen (2007) and Lei et al. (2009), as well as with the interviews conducted for this study.

Of course, if all material requests were denied, then no MTA would occur, and the observation would be excluded from the dataset. However, material sharing behavior could well change over time or across different requesters. For example, material requests might initially be fulfilled (thus an MTA could be observed), but then this might end once discussions with a licensee begin. Alternatively, material requests could be granted asymmetrically, with some being approved, while others to scientists / companies working on commercializable ideas are declined. Indeed, an

[52] Error bars are ±1 standard error

[53] The authors do not believe that there is any economic meaning to the unusual result in t+2, as virtually all of the alternate specifications for this effect show statistical significance in all of years t+1, t+2, and t+3. However, in favor of consistency and not cherry-picking results, the preferred specification is what is presented here.

interviewee for this study who was working with a start-up said explicitly "If another company asked to use our [materials] for [same purpose as our company uses them] we would say 'no'". This comment suggests that rejection of a materials request could occur even if that lab had shared these materials earlier, prior to the creation of the start-up.

Interestingly the timing of this effect seems to occur even in the year of the license itself, which, given the lags needed for research and publication, might suggest that perhaps researchers or universities limit sharing even during the negotiation of the license.[54]

4.3.3 Comparison to the Murray and Stern definition of Research Tools

In Section 4.1.2 it was pointed out that Murray and Stern (2007) used a different approach for identifying research tools. Their approach focused on identifying research tools through patent classes. Figure 23 and Figure 24 show the results of using this definition on the UCOP data sample.

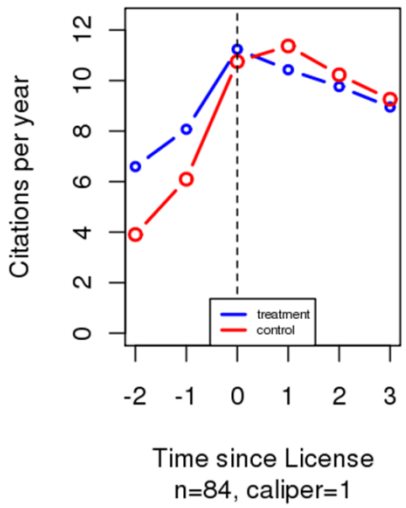

Figure 23: License effect on research tools (Murray and Stern definition)

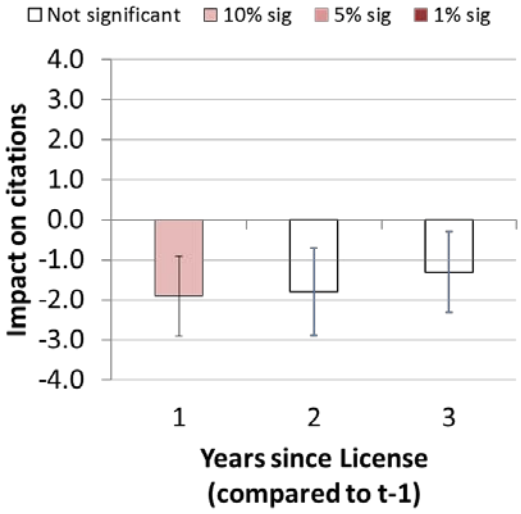

Figure 24: Regression results from the License Effect on research tools (Murray and Stern definition)[55]

These also show a drop in citations arising from the license, but the results are smaller and less statistically significant. This is what one would expect if their definition was an imperfect metric for identifying research tools, since that would imply more 'noise' and a bias towards zero. These results are also less convincing since the pre-treatment citation patterns differ significantly, making the control group a less credible counterfactual to the treatment group.

[54] Note that the identification of any in-year effect is much weaker, since the exact timing of the license during the year is not accounted for in the analysis.

[55] Error bars are ±1 standard error

4.4 Discussion

Ideally this analysis would conclude with strong statements about the overall welfare effects of licensing academic work. However, such statements are beyond the scope of this paper because other important effects that impact the net benefit to society cannot be observed. Examples of these include (i) the amount that firms invest in research and development on the discoveries they license, (ii) the impact on how quickly discoveries get to market, (iii) the utility that consumers get from those products.

Despite these restrictions, this analysis does allow for more modest welfare claims. It argues that, in general, the claims that scientists are being excluded from licensed work are either overblown, or that other positive effects outweigh these exclusions. As such, it provides some reassurance that, in many cases, licensing activities may be complimentary to the flow of scientific knowledge.

At the same time, the finding of the impact of licenses on research tools corroborates the concerns raised by scientists and the NIH on this issue. Not only do citations drop, contrasting starkly with the positive effects seen for most licenses, but the rapid onset of these effects is consistent with a restriction of input materials. And, while it is possible that the incentives created by licensing could lead to private sector research that compensates for these effects, it would be important to show this empirically to alleviate the concerns raised by the findings of this study.

4.5 Summary

This chapter has investigated the effects of licenses on patented publications. It shows that, in general, licenses on scientific work *increase* the number of citations to related publications, but that the opposite effect is observed for research tools, where licenses *decrease* the number of citations.

These results are consistent with other findings in the literature which suggest that licensing may have a positive signaling effect, but that licenses on research tools may lead to restrictions on input materials that are important for follow-on research.

5 Conclusion

This dissertation is intended to contribute to the literature in four ways. Firstly, it provides an overview of material transfer agreements, a seldom observed, but important intellectual property tool used by universities. This analysis confirms previous findings that most MTAs are within the biological sciences. It further shows that MTAs with the private sector are difficult and time-consuming to negotiate, perhaps helping to explain high rejection rates on material transfers that have been seen in the literature. Finally, it argues that MTAs are a good proxy for research tools, a category which has been difficult to define.

The second contribution of this dissertation is to set out a method for finding patents and publications related to the same discovery. It sets out a methodology for doing this that is straightforward, replicable and automated. Moreover, it connects this to the type of estimator it represents, and then validates it on a large set of hand-matched patent-paper pairs, achieving 95%+ success.

Thirdly, in the discussion of the effect of licensing, this paper outlines why a parametric approach to modeling citation curves may create dangers of model misspecification due to their complexity. It then uses a nearest-neighbor approach to produce a set of treatment and control observations with excellent covariate balance.

Finally, this dissertation contributes to the debate on the usage of intellectual property on academic science – an area that is becoming more important as academia's role in both research and intellectual property grows. It shows that on a large group of discoveries, the effect of licensing on scientific communication is positive, perhaps due to signaling effects. In contrast, it also shows that concerns about research tools being negatively affected by licensing seem well founded, with the license producing a rapid and significant fall in the citations accruing to the publication. This may suggest that materials are being shared less widely amongst scientists whose work is being licensed, which raises concerns about the science in these areas being impeded.

In summation, this dissertation attempts to shed light on the important issue of the use of intellectual property on academic science, and develops both methodological techniques and empirical results to do so.

6 References

Association of University Technology Managers (2003). *AUTM Licensing Survey.*

Association of University Technology Managers (2012). "Uniform Biological Materials Transfer Agreement (UBMTA)." *AUTM Website* at www.autm.net/AM/Template.cfm?Section=Technology_Transfer_Resources&Template=/CM/ContentDisplay.cfm&ContentID=2810.

Association of University Technology Managers (2012b). "Master UMBTA [sic] Agreement Signatories" *AUTM Website* at http://www.autm.net/AM/Template.cfm?Section=Technology_Transfer_Resources&Template=/CM/ContentDisplay.cfm&ContentID=8374.

Blumenthal, D., E.G. Campbell, M.S. Anderson, N. Causino, and K.S. Louis (1997). "Withholding Research Results in Academic Life Science," *Journal of the American Medical Association* 277: 1224-1228.

Campbell, E.G., B.R.Clarridge, M. Gokhale, L. Birenbaum, S. Hilgartner, N.A. Holtzman, and D. Blumenthal (2002). "Data Withholding in Academic Genetics; Evidence from a National Survey," *Journal of the American Medical Association* 287: 473-480.

Casella, George and Roger L. Berger (2002). Statistical Inference: Second Edition (United States: Duxbury)

Drivas, K., Z. Lei, and B.D. Wright (2011). "The Role of Exclusive Licensing in Diffusion of Academic Patented Inventions" Working Paper, UC Berkeley.

Eisenberg, R.S. (2001). "Bargaining over the Transfer of Proprietary Research Tools: Is this Market Failing or Emerging?" in R.C. Dreyfuss, D.L. Zimmerman, and H. First, eds., Expanding the Boundaries of Intellectual Property: Innovation Policy for the Knowledge Society (Oxford: Oxford University Press).

Eisenberg, Rebecca. 2003. Patent Swords and Shields. *Science*, Vol. 299. Downloaded from sciencemag.org in May 2008.

Furman, Jeffrey and Scott Stern. 2006. Climbing atop the shoulders of giants: the impact of institutions on cumulative research. NBER Working Paper 12523.

Heller, Michael and Rebecca Eisenberg. 1998. Can Patents Deter Innovation? The Anticommons in Biomedical Research. *Science*, Vol. 280. 698-701.

Hunter-Davis, Terri. 2012. "About the University of California." Website of University of California, Office of the
President, https://wiki.ucop.edu/display/UC101/About+the+University+of+California.

Katz, M.L., and J.A. Ordover. 1990. "R&D Competition and Cooperation." Brookings Papers on Economic Activity: Microeconomics: 137-192.

Lei, Zhen, Rakhi Juneja and Brian Wright. 2009. Patents versus patenting: implications of intellectual propery protection for biological research. *Nature Biotechnology.* Vol. 27.

Marshall, E. (1997). "Need a Reagent? Just Sign Here," 278 *Science* 212.

McCain, K.W. (1991). "Communication, Competition, and Secrecy: The Production and Dissemination of Research-Related Information in Genetics," *Science, Technology, & Human Values 16*, 491-516.

Murray, Fiona. 2006. "The Oncomouse that Roared: resistance & accommodation to patenting in academic science" *working paper.*

Mowery, David C. and Arvids A. Ziedonis. 2007. Academic patents and materials transfer agreements: substitutes or complements? *Journal of Technology Transfer,* 32(3): 157-172.

Murray, F., and S. Stern, 2007. "Do formal intellectual property rights hinder the free flow of scientific knowledge? A test of the anti-commons hypothesis," *Journal of Economic Behavior and Organization* 63, 648-687.

Murray, F., and S. Stern (2008). "Learning to Live with Patents: A Dynamic Model of a Knowledge Community's Response to Legal Institutional
Change". http://fmurray.scripts.mit.edu/docs/Murray.Stern_LearningtoLivewithPatents.pdf

National Institutes of Health (1995). "Uniform Biological Material Transfer Agreement Finalized." *NIH Grant website* at http://grants.nih.gov/grants/guide/notice-files/not95-116.html.

National Institutes of Health (1998). Report of the Working Group on Research
Tools, http://www.nih.gov/news/researchtools/.

National Research Council (2010). *Managing University Intellectual Property in the Public Interest* (Washington, D.C.: National Academies Press).

National Science Foundation (2011). *Science and Engineering Indicators.* http://www.nsf.gov/statistics/seind/.

National Science Foundation (2012). *Science and Engineering Indicators.* http://www.nsf.gov/statistics/seind12/.

Rubin, Donald B. (1979). "Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies," *Journal of the American Statistical Association.*

Rubinstein, E. (1990). "The Untold Story of HUT78," *Science* 248, 1499-1507.

Sampat, B.N. (2004). "Genomic Patenting by Academic Researchers: Bad for Science?" Working Paper.

Sekhon, Jasjeet S. (2011). "Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching package for R." *Journal of Statistical Software.*

Stanford University (2012). "For Researchers – MTA Information." Stanford University website at http://www.stanford.edu/group/ICO/researcher/reMTA.html.

Stern, S. (2004). *Biological Resource Centers: Knowledge Hubs for the Life Sciences* (Washington, D.C.: Brookings Institution).

Stokes, D.E., *Pasteur's Quadrant* (Washington, D.C.: Brookings Institution, 1997).

Streitz, Wendy and Alan Bennett. 2003. Material Transfer Agreements: A University Perspective. *Plant Physiology*, Sept 2003.

Thompson, Neil, David C. Mowery, and Arvids A. Ziedonis (working paper). *Does Licensing Facilitate or Restrict Knowledge Flows Related to Research Tools and Other Outputs of Academic Science?*

Thomson Reuters. Various Pages. *Web of Science.* Downloads in 2010, 2011. www.isiknowledge.com.

University of California (2012). "Research Impact" University of California Research website: http://research.universityofcalifornia.edu/impact/.

University of California (2012b). "UC Health" University of California Health website: http://health.universityofcalifornia.edu/about/.

University of California Office of the President (2012). "Contract and Grant Manual." *Intellectual Property and Related Matters.* Website: http://www.ucop.edu/raohome/cgmanual/chap11.html.

University of California Office of the President (2012b). "Annual Reports / Special Reports." *Technology Transfer.* Website: http://www.ucop.edu/ott/genresources/annualrpts.html.

University of Pittsburgh (2012). "Material Transfer Agreement (MTA) Frequently Asked Questions (FAQ)." *University of Pittsburgh, Office of Research* website at: http://www.pitt.edu/~offres/ResContracts/MTA-FAQ.html.

USPTO (2012). *U.S. Colleges and Universities – Utility Patent Grants 1969-2008.* Website of the United States Patent and Trademark Office at www.uspto.gov.

Wade, N. (1980). "Hybridomas: A Potent New Biotechnology," *Science* 208, 692-693.

Walsh, J.P., and W. Hong (2003). "Secrecy is Increasing in Step with Competition," letter to the editor, *Nature* 412, 801-802.

Walsh, J.P., A. Arora, and W.M. Cohen (2003). "Effects of Research Tool Patents and Licensing on Biomedical Innovation," in W.M. Cohen and S.A. Merrill, eds., *Patents in the Knowledge Based Economy* (Washington DC: The National Academies Press).

Walsh, J.P., Cho, C., Cohen, W.M., 2005. "The view from the bench: patents, material transfers and biomedical research," *Science* 309, 2002–2003.

Walsh, J.P., Cho, C., Cohen, W.M., 2007. "Where excludability matters: Material versus intellectual property in biomedical science," *Research Policy* 36, 1184-1203.

Wasserman, Larry, *All of Statistics* (New York, NY: Springer, 2005).