Essays on patenting, R & D and technological innovation

Na Cheng

University at Albany, State University of New York, ncheng@albany.edu

The University at Albany community has made this article openly available. Please share how this access benefits you.

Follow this and additional works at: https://scholarsarchive.library.albany.edu/legacy-etd

Part of the Economics Commons

Recommended Citation
https://scholarsarchive.library.albany.edu/legacy-etd/1582

This Dissertation is brought to you for free and open access by the The Graduate School at Scholars Archive. It has been accepted for inclusion in Legacy Theses & Dissertations (2009 - 2024) by an authorized administrator of Scholars Archive. Please see Terms of Use. For more information, please contact scholarsarchive@albany.edu.
ESSAYS ON PATENTING, R&D AND TECHNOLOGICAL INNOVATION

by

Na Cheng

A Dissertation
Submitted to the University at Albany, State University of New York
in Partial Fulfillment of
the Requirements for the Degree of Doctor of Philosophy

College of Arts and Sciences
Department of Economics
August 2016
ESSAYS ON PATENTING, R&D AND TECHNOLOGICAL INNOVATION

by

Na Cheng

© Copyright 2016
To my parents Yongsheng Cheng, Xinli Wang, my husband Shuai Fu,

and my daughter Olivia C. Fu
The innovative activities by firms are considered to be the main driving force of the economic growth. The patent data has been widely used to investigate the relationship between R&D and patents which are taken as an output indicator of the new invention or innovation. On the other hand, empirical evidence supports that teamwork of inventors fosters innovation and technological progress. Motivated by the importance of understanding the patenting behavior of firms and collaboration of inventors, we exploit the continuing applications of patents, examine the R&D and patents relationship, and investigate the effects of teamwork on innovation output.

Chapter 1 studies the role of continuing patent applications across six main technological fields in the United States during 1976-2012. The research represents the first effort of which we are aware to analyze patent application data on continuing histories, providing information on the effects of 1995 GATT legislation changes in the US patent law intended to curb “submarine patenting”. We employ novel data on filings of three types of continuing applications - the continuations, the continuation in part, and divisions to distinguish among the motives for continuing patents. We find that the GATT legislation reduced continuing applications overall and mitigated submarine patenting. The disparate impact of the GATT change will be most acute on pharmaceutical and chemical industries.

Chapter 2 studies the relationship between R&D expenditures and patenting in the context of the U.S. chemical and pharmaceutical industries during the 1981-2003 period. Our empirical analysis differs from previous work that investigates the patents-R&D relationship in two main aspects. First, we substitute priority date of a patent for application date as a proxy for the
invention date which ties the timing of innovation to patenting. Our estimation results exhibit the change of marginal effect on R&D by using priority date compared to application date. Second, we develop a new transformation of the quasi-differenced GMM estimator that allows for endogenous regressors with the dynamic linear feedback model. Our empirical results from various count panel data models show that the correlation between R&D and patents is higher by using priority date than application date for the pre-GATT period and lower for the post-GATT period. This implies that the 1995 legislation change in patent term has reduced continuing patent applications and effectively alleviated the submarine patent problems particularly in the chemical and pharmaceutical industries.

Chapter 3 studies the impact of inventor team characteristics on the quality of innovation outputs. We investigate the impact of two attributes of teamwork - team composition and persistence on the quality of patentable inventions, using the number of claims and forward citations as proxy for patent quality. The analysis on team composition is conducted at inventor-team level and we show that both the skillset diversity of team members and the presence of generalist inventors have positive impacts on the quality of patents. We also conduct a firm level analysis on team persistence and use firm dissolution as an instrumental variable to control for endogeneity in our estimation. We show that the team persistence has a positive impact on patent quality.
CONTENTS

ABSTRACT ........................................................................................................................................ iv

I Continuing Patent Applications and the Impact of GATT Legislation on Patents 1

1. Introduction ...................................................................................................................................... 1
   1.1 Continuation, Continuation-in-part, and Divisional applications ........................................... 3
   1.2 Patent Term and Provisional Rights ....................................................................................... 7
2. 1995 GATT Change and Effects on Continuing Applications ..................................................... 8
3. Data .................................................................................................................................................. 11
   3.1 Data Description ..................................................................................................................... 11
   3.2 A Patent Example ................................................................................................................... 12
4. Descriptive Analyses ...................................................................................................................... 13
   4.1 Trends in Filing Continuing Applications ............................................................................... 13
   4.2 Application vs Priority Dates ................................................................................................. 16
   4.3 Quality of Continued Patents .................................................................................................. 17
   4.4 Patent Clustering Analysis ..................................................................................................... 19
5. Conclusion ....................................................................................................................................... 21

II The R&D-Patenting Relationship Revisited: An Empirical Analysis on the U.S. Chemical and Pharmaceutical Industries 44

1. Introduction ...................................................................................................................................... 44
2. Theoretical Background and Hypotheses ...................................................................................... 47
III Impacts of Inventor Team Composition and Persistence on Value Innovation

1. Introduction .................................................................................................................. 74
2. Theory and Hypotheses ............................................................................................... 77
   2.1 Team Composition .................................................................................................. 77
      2.1.1 Skillset Diversity Index ...................................................................................... 77
      2.1.2 Generalist Inventor Indicator ............................................................................. 79
   2.2 Team Persistence ..................................................................................................... 81
3. Data, Variables and Empirical Strategies .................................................................... 82
   3.1 Data ....................................................................................................................... 82
   3.2 Variables ............................................................................................................... 84
3.2.1 Dependent Variable ................................................................. 84
3.2.2 Independent Variable ............................................................. 85
3.2.3 Control Variables ................................................................. 85
3.3 Empirical Strategies ............................................................... 86
  3.3.1 Patent- or Team-Level Analysis on Team Composition .......... 86
  3.3.2 Firm-Level Analysis on Team Persistence ............................. 87
4. Results ....................................................................................... 89
5. Concluding Remarks .................................................................. 91
CHAPTER I

Continuing Patent Applications and the Impact of GATT Legislation on Patents

1 Introduction

If you have filed a patent in the United States Patent and Trademark Office (USPTO), one or more applications can later be filed that claim the priority date of the earlier patent, which is referred to as a parent application. The applications later filed are known as continuing applications. The continuing applications are prevalently used by U.S. firms and they play an important role in U.S. patent policy reforms.

Filing a continuing application can be extremely useful as it offers inventors the ability to modify the original application and claim the benefit of the filing date of the earlier application. The filing date of the parent application is referred to as the “priority date” of the continuing application. The priority date is important because it determines the references that the examiner can use to reject a patent claim but also defines the scope of prior art available to potentially invalidate the patent claims.

In order for the continuing application to be entitled to claiming for the priority date of the parent application, several conditions must be met and these rules apply to all three continuing types. The first condition is that the continuing application must be filed when the parent application is “pending”, that is during the time between its filing date and the date that it is either

---

1 There are very few economics literature on continuing applications. For example, Lemley and Sampat, 2008, 2010, Quillen and Webster, 2001, 2006, 2009, etc.
abandoned or issued into a patent. According to 35 U.S.C. §119 & §120, at any time during this period of pendency, the applicant has the right to file a continuing application, which includes a claim for “priority” to that original application. The second condition is that there must be at least one inventor listed under the continuing application that was listed under the parent application. The third condition is that the claims in the continuing application must be fully supported in the parent application. If not, the claims will not be entitled to the priority date of the parent application but to the later filed application date. The problem with non-supported claims often comes about with the continuation-in-part applications. A more detailed description of CIP type will be discussed in the following section.

There are two major benefits to the claim for the priority date of the parent application. First, it can prevent using any patents, published patent applications, non-patent publications, and certain other prior art that have become public between the filing date of the parent application and that of the continuing application as prior art against the continuing application. Second, it can prevent using the parent application itself which has to be published 18 months after filing as prior art against the continuing application. However, as the standard patent term is calculated as 20 years after the application’s earliest priority date, the claim for priority will likely shorten the patent term.

This chapter provides a detailed introduction into continuing applications and discusses its implications for use of patents to study innovation. It constitutes a foundation that our patent production function analysis presented in the second chapter builds upon. In this chapter, we use data on filing histories of patents issued in the United States during 1975-2012 to investigate the influence of uses and characteristics of continuing applications. We examine the motives for filing the three different types of continuing applications - continuation (CAP), continuation-in-part (CIP)
and divisional (DIV) applications across various technological fields. We compare the values of patents issued from continuing applications with those that are not associated with continuing applications. We also analyze the effects of the 1995 General Agreement on Tariffs and Trade (GATT) change in patent term on strategic uses of different types of continuing applications. We aim to document relevant facts and statistics on the U.S. patenting associated with continuing applications and suggest possible new avenue of inquiry that continuing applications and relevant patent policy open up.

This chapter is organized as follows. The remainder of this section describes three types of continuing applications and relevant concepts. Section 2 discusses the GATT policy change in patent term and its impact on continuing applications. Section 3 describes data and provides a specific example to illustrate the timeline of filing continuing applications and issuance of patents. Section 4 presents and discusses the results of our analyses. Section 5 concludes.

1.1 Continuation, Continuation-in-part, and Divisional Applications

In this section we discuss the three types of continuing applications and in which situations each type will be used.

A continuation (CAP) allows an inventor to amend, add or remove the claims without changing the specifications or adding any new subject matter. According to the Manual of Patent Examining Procedure (MPEP), “the continuation should not include anything which would constitute new matter if inserted in the original application.” The new claims in the continuing application must be fully supported by the specifications as they were written in the parent application. Thus, applicant will file a CAP when he discovers a potentially patentable invention that has been disclosed but not claimed in the original application.
CAP applications can be useful when the examiner allows certain narrow claims in an application but rejects other broader claims. Suppose that you have a pending application and you have received an action from the USPTO that “the application will be approved if some claims are removed”. Since you are filing the application in order to get patent protection for a particular product to block a competitor from infringing, it will be advantageous to remove the rejected claims and then file a CAP to further pursue protection on broader claims that were rejected, and you can get a patent issued within another two months on the allowed claims.

Unlike the CAP, a continuation-in-part (CIP) application allows an inventor to modify the specification or add new material required to support the new claims. The MPEP describes a CIP as “an application repeating some substantial portion or all of the earlier nonprovisional application and adding matter not disclosed in the said earlier nonprovisional application.” This allows inventors to add new material to describe new improvements as long as the majority of the specifications remain unchanged.

As mentioned earlier, it is important to be aware that only the claims that are supported by the disclosure in the parent application are entitled to the priority date of the parent application. According to MPEP §211.05 (I) (B), claims may extend back through “a continuous chain of copending nonprovisional applications” that disclose the claimed subject matter. However, claims that are related to the new subject matter introduced in the CIP are not entitled to the priority date but to the filing date of the CIP. That is, in CIP applications, priority date is determined on a claim-by-claim basis, and a CIP application can have various priority dates. Take the case Santarus v. Par

---

2 Provisional application is a legal document filed in the USPTO under 35 U.S.C. §111 (b). A provisional application has a pendency lasting 12 months from the date the provisional application is filed. A provisional application secures a priority filing date if a nonprovisional patent application for the same invention is filed within the 12-month pending period. If an inventor does not file a nonprovisional application within that timeframe, the provisional application is deemed abandoned.

3 ”Copending” means that the applications are pending, i.e. neither abandoned nor granted.
Pharma [Fed. Cir. 2012] for example: in its decision, the Court of Appeals for the Federal Circuit ruled that some of the disputed claims were sufficiently disclosed in the parent filing and therefore properly claimed priority to that original filing date. Another set of claims, however, included matter that was added in the CIP. Those claims with new matter were not given priority to the original filing and thus had to rely on the filing date of the CIP application.

In a CIP application, there is a trade-off between an earlier priority date for some claims and a shortened patent term for all claims. The advantages of CIP applications are obvious: first, advantage is CIPs allow inventors to add new subject matter as well as expand the descriptions; second, CIPs can be considered to economize on filing fees by combining existing and new disclosures into one application; third, claims that are supported by the disclosure in the parent application can inherit the priority date of the parent application. However, as any type of continuing applications, the patent term of a CIP is shortened to its parent application’s term. This likely reduction in patent term applies to all claims in the CIP application, even those that are not entitled to the earlier priority date. Therefore, it will be advantageous to file a CIP only if the priority date of the parent application will be eligible. If not, it would result in loss of a portion of the life of the patent without obtaining the benefit of the earlier priority date. According to 35 U.S.C. §102 (b), “for the claims with new matter any patent issued or document published more than one year before the CIP filing date would count as prior art”. That is, if the CIP’s claim extends even slightly past what a parent application disclosed more than a year earlier, the parent can be used as prior art against the CIP claim. Figures 1.14 – 1.16 demonstrate the fact that most CIP applications are filed within a shorter time lag after the parent application than the other two types.

The previous two types of continuing applications are usually voluntarily filed by the
applicant, while the divisional (DIV) application is usually due to a restriction requirement under 35 U.S.C. §121 by the examiner⁴. According to MPEP, each patent application must be strictly focused on one invention or on “a group of inventions so linked as to form a single general inventive concept”, which is known as “unity of invention”. If an inventor files an application that includes multiple inventions, the examiner may ask the applicant to remove the extra inventions into DIV applications that each contain a single invention. Similar to CAPs, no new subject matter can be added in a DIV application. According to MPEP, the DIV is “carved out of a pending application and disclosing and claiming only subject matter disclosed in the earlier or parent application.” That is, a DIV application simply takes a subset of the parent application’s claims and moves them into one or more than one separate applications that would claim priority to the parent application. Thus, inventors can pursue protection of some or all aspects of the claims through filing DIVs.

In addition to the description of each type of continuing applications and their uses, there are some other situations when the continuing applications can be strategically useful. One scenario is identified as “Emergency”. For example, suppose you have a pending application, and have decided to abandon it after you have received a rejection action from the USPTO. By law 35 U.S.C. §133, you have six months from the mailing date of the Office Action to respond to it. But you may change your mind when it is very close to the six-month deadline, and then a continuing application can be filed promptly and retain your rights alive to be further pursued. Another strategic use of continuing applications is to “keep it alive”. Suppose you are nearing the issuance of a patent but you have not determined whether or what additional subject matter you want to pursue for protection later, and then it can be useful to file a continuing application to keep

⁴ According to MPEP §201.06, "a divisional application is often filed as a result of a restriction requirement made by the examiner".
your options flexible and block the potential competitor in the future.

U.S. patent applicants would also pursue the benefit from “submarine” patents by delaying the expiration date of a patent through filing a succession of continuing applications (Graham and Mowery, 2004). Beard (2008) propose that patent applicants use submarine patents to gain a competitive advantage over competitors by filing continuing applications to keep the patent “submerged” and unpublished, until a product is infringed by competitors. Empirical evidence shows that although the submarine patent is relatively rare it poses a significant risk to inventors, such as a payment of royalties in order to avoid a costly legal battle, an award of higher royalties, or worse, a complete bar to sale of the new product.

Lastly, continuing applications can also be considered as a long-term strategy to make it valuable to potential investors and/or licensees of your patents.

1.2 Patent Term and Provisional Rights

Under the current U.S. patent law, the patent term is 20 years from the earliest filing date if the patent application is filed on or after June 8, 1995. According to 35 U.S.C. §154 (a) (2), the patent term is “beginning on the date on which the patent issues and ending 20 years from the date on which the application for the patent was filed in the United States or, if the application contains a specific reference to an earlier filed application or applications under section 120, 121, or 365 (c) of this title, from the date on which the earliest such application was filed.” In other words, the patent term is calculated as 20 years from the priority date as we have defined before if the patent results from a continuing application filed on or after June 8, 1995.

However, it is important to point out that although the patent term starts from the date of grant under the current rule, the monopoly period for a patent owner can start as early as the date
of publication of the application (usually 18 months after filing). According to 35 U.S.C. §154 (d), patent applicants are given “the right to obtain a reasonable royalty from any person who makes, uses, offers for sale, or sells the invention ... during the period beginning on the date of publication of the application, ... and ending on the date the patent is issued.” These rights are referred to as “provisional rights”. For example, if an inventor has a patent pending for an invention and a competitor starts copying that invention while the patent application is pending, the inventor can sue the competitor for patent infringement once the patent is eventually issued.

2 GATT Change and Impact on Continuing Applications

In this section we further discuss the 1995 legislation GATT change in patent term and its impact on continuing application behavior. The 1995 GATT change in patent term meant that any use of continuing applications to delay the expiration date of a patent results in a shorter patent term. Given that one motive for the 1995 change was curbing the abuse of continuing applications, we can use GATT to measure the importance of submarine patents. Our descriptive analysis result shows that the fall in continuing applications after 1995 suggest the effectiveness of the policy change. Furthermore, we can use GATT as an instrument to measure the impact on reduction in submarine patents quantitatively.

In 1995 the U.S. patent law changed the patent term from 17 years from the grant date to 20 years from the earliest application date under amended 35 U.S.C. §154 affected by the GATT Uruguay Round implementing legislation (P.L. 103-465). Under the Uruguay Round Agreements Act, patents issuing on applications filed on or after six months from the date of enactment would be subject to the 20-year patent term. In other words, the last chance to file an application to obtain a patent with a patent term of 17 years from grant is to file an application the day before the
The effective date of the 20-year patent term - i.e., before June 8, 1995.

The GATT legislation has effectively changed the length of the maximum patent term associated with a granted patent. For patent grants filed before the June 8, 1995 implementation of the GATT legislation, the maximum patent term was to be 17 years from the date of patent grant. For patent grants filed after the June 8, 1995 implementation of the GATT legislation, the maximum patent term is to be 20 years from the date of earliest filing.

All applications filed on and after June 8, 1995 will be subject to the new 20-year patent term. The change applies to all original applications as well as all continuing applications (i.e., continuations, continuations-in-part, and divisionals). The change does not apply to provisional applications because a provisional application can never be issued as a patent. As we discussed before, continuing applications claim priority from a previously filed application. For many of these continuing applications, the maximum patent term would have been much shorter for continuing applications filed on or after June 8, 1995 because the maximum patent term (under the implemented GATT legislation) would be calculated as 20 years from the filing date of the earliest associated application. Therefore, by submitting the continuing applications prior to the June 8 date, many applicants would be able to maximize the maximum possible patent term for their inventions. This explains why it has been anticipated by a significant increase in filing continuing applications prior to the 1995 date and then a proportionate decrease after the date.5 These facts are exhibited in Figures 1.5 – 1.7.

Furthermore, the disparate impact of the GATT change will be most acute on industries such as pharmaceuticals or chemicals. In software or most electronic industries, where the inventions are often obsolete within a shorter period, a reduction in the patent term by a few years

---

is often less important. However, in pharmaceutical or chemical industries, where the last years of the patent term are often the most commercially valuable (Eisenberg, 2000), a change in patent term becomes more crucial. Figures 1.8 – 1.10 show that the surges in continuing applications around 1995 are most remarkable in chemical and drugs and medical technology fields.

The implementation of the GATT change in patent term has fundamentally changed the strategic behaviors traditionally conducted through filing continuing applications by the U.S. patentees. Most importantly, it has eliminated most of the related submarine patent issues.6 “Submarine” patents, as described earlier, refer to patents that were issued after long pending application periods until the claimed invention became market ready. Then the patents would have a term of 17 years from the issue date before 1995. Thus submarine patents could enable their assignees to threaten legal action against users of widely adopted technologies for infringement. Prior to the 1995 GATT change, U.S. patent applicants would pursue the benefit from submarine patents by delaying the expiration date of a patent through filing a succession of continuing applications. However, due to the legal change in patent term there would be little benefit in postponing the issuance of patent. As a result, it would curtail the former practice of filing continuing applications. In particular, continuation-in-part applications have fallen out of favor after mid-1995. As discussed in the previous section, CIP applications can cause loss of patent term without any benefit when it is not eligible for the priority claim.

In addition, the rules were subsequently modified so that applications filed prior to the June 8 GATT implementation date were later given a maximum patent term of 20 years from the date of earliest filing or 17 years from the date of grant, whichever period is longer. Under 35 U.S.C. §154 (c), “the term of a patent that is in force on or that results from an application filed

---

6 The other legal change that has facilitated the elimination of submarine patents was the USPTO began publishing pending patent applications likely 18 months after filing in 2000.
before the date that is 6 months after the date of the enactment of the Uruguay Round Agreements Act shall be the greater of the 20-year term as provided in subsection (a), or 17 years from grant, subject to any terminal disclaimers.”

3 Data

In this section, we describe the data we use for our descriptive and statistical analyses of investigating the characteristics of continuing applications. We describe our data sources and highlight some relevant facts. We also provide a specific patent example to illustrate the historical timeline of filing a series of continuing applications and issuance of the resulting patents.

3.1 Data Description

Our analysis uses data from the USPTO on utility patents granted between January 1975 and December 2012. The USPTO reported 4,459,694 utility patents granted for the entire period. We gathered the filing continuing application history of each patent from the “related applications data” in the customized Technology Assessment and Forecast (TAF) database.

We calculate the “priority date” as the date on which the first in a series of continuing applications was filed. For patents that were never subject to the continuing applications procedure (referred to below as “original” patents), the priority date is the first and only application date. The grant lag is calculated as the year gap between the priority date and grant date. We retain in our analysis only those patents with priority dates between 1975 and 2008 that were granted during 1975 and 2012. For patents granted from applications with continuing applications, the average grant lag is 4.97 years with a standard deviation (s.d.) of 8.04 and the median is 4 years. The original patents, by contrast, are granted 2.62 years (s.d. of 1.54) after the date of application on
average and the median is 2 years. The longer grant lag for continued patents introduces a bias that over-represents applications that are not continued in the population of granted patents during the later years. Restricting the last year of priority dates to 2008 in all patents granted through the year 2012 minimizes this bias.

### 3.2 A Patent Example

Here we choose a particular example to illustrate the practice of a chain of continuing applications. Consider the issued U.S. patent number 7,629,736 which was applied on December 12, 2005 and granted on December 8, 2009, this application is a CAP of application filed on March 27, 2003, now patent number 7,098,587 issued on August 29, 2006, which is a DIV of application filed July 8, 2002, now patent number 6,712,664 issued March 30, 2004, which is a DIV of application filed September 23, 1998, now patent number 6,417,605 issued July 9, 2002, which is a CIP of application filed August 6, 1997, now abandoned, which is a CPA of application filed October 13, 1995, now abandoned, which is a CIP of application filed September 16, 1994, now abandoned. Its priority date has been identified as the earliest filing date in the continuity chain, i.e. September 16, 1994.

See the diagram below for the related application timeline.
4 Descriptive Analyses

4.1 Trends in Filing Continuing Applications

During the January 1975 to December 2012 period, 23% of the nearly four and a half million patents are granted from continuing applications. Figure 1.1 shows the total number of patents and patents granted from continuations, continuations-in-part and divisions yearly by grant year from 1975-2012 and Figure 1.2 shows the patents by application year from 1975-2007.7

Figure 1.2 shows that the number of patents granted each year from continuing applications

---

7 The figures which are plotted by application or priority year are ended in year 2007 due to the truncation problem. An analysis on truncation problem is shown in Figure 1.27 and Figure 1.28.
climbed steadily in the second half of the 1980s and throughout the early 1990s and there was a sharp rise starting from two years prior to the 1995 implementation of the GATT legislation. It reached a peak at 50,629 patents granted from continuing applications which was 35% of the 144,646 total grants in 1995 and was followed by a proportionate drop after 1995. We exploit the data at date level and find that the surge in continuing applications occurred before the effective date of the GATT change. There was a surprisingly huge increase in continuing applications on the day before June 8, 1995 and a proportionate fall on and after June 8, 1995. Given that the GATT change would shorten the term of patents of which the applications were filed after June 8, 1995, the applicants would be able to maximize the maximum possible patent term for their inventions by submitting the continuing applications prior to the effective date.

The peak in frequency of continuing applications in 1995 has been attributed primarily to a rush in filings prior to the June 1995 change in the U.S. patent term from 17 years from the date of patent issuance to 20 years from the priority date, as required by the Uruguay Round of the General Agreement on Tariffs and Trade (GATT).

Figure 1.6 shows that the fraction of patents that have at least one type of continuing applications grew steadily since the early 1980s and throughout the first half of the 1990s, peaking at 35% in 1995. Thereafter, the fraction declined sharply in the late 1990s and rebounded since 2000, but it did not reach the levels from 1995. The frequency in the filings of three types of continuing applications - the continuations (CAP), the continuations-in-part (CIP) and divisions has also been shown in Figure 1.6. Between 1975 and 2012, 9.8% of patents granted are associated with CAPs, 6.7% with CIPs and 6.8% with divisions. In general, the trends for all types mirrored those for CAPs which account for 42% of all continuing applications for the 1975-2012 period. The CIPs grew steadily in the 1980s and the first half of the 1990s, peaking at nearly 10 percent
in 1995 and declined steadily since then. The divisions leveled off until they increased sharply and peaked at over 10 percent in 1995 and then plunged to the initial levels.

Figure 1.3 and Figure 1.4 show the yearly patents granted from all types of continuing applications by technology category. We used the National Bureau of Economic Research (NBER) classification developed by Hall et al. (2001) to aggregate the USPTO technology classes into 6 categories: chemical; drugs and medical; electrical and electronic; computers and communications; mechanical; and a miscellaneous “other”.

The 1995 peaks across the technology classes in Figure 1.4 suggest that the GATT legislation has greater impact on patents in pharmaceutical and chemical industries than other industries such as electronics and computers.

This is because the patent term is more likely to be valuable towards the end for patents in the fields such as chemicals and pharmaceuticals and the 1995 GATT change which would result in a reduction in the term is far more important to the owners of these patents.

On the other hand, pharmaceutical inventors reportedly perceived the 17-year term to be more advantageous because it ran from the date of patent issuance. The pending period could thus be used to comply with substantial product development and regulatory requirements without reducing the patent term, as would occur once the term ran from the date of application (Adelman and DeAngelis, 2007).

Figure 1.9 shows the yearly fraction of patents granted from all types of continuing applications by technology category.

The trends in filing continuing applications for all technology categories are similar to those for overall continuing applications. They climbed steadily since the early 1980s throughout the first half of the 1990s until they peaked in 1995. Thereafter, they decreased sharply in 1996
and then started to level off or increase slightly since the late 1990s.

Among the six technology categories, continuing applications are more frequently used for patents in the “drugs and medical” and “chemical” technology categories. Continuing applications overall account for 38% of the patents granted in the field of “drugs and medical” for priority years 1975 and 2012, and 30% of those granted in “chemical” during the same period.

Figures 1.17 – 1.19 shows that the distribution of three types of continuing applications varies considerably across the technology categories. The three types of continuing applications are evenly distributed in “chemical”. CAPs are remarkably in the majority for the patents in “computers and communications”. The “electrical and electronic” and “mechanical” patents use CAPs more than the other two types. CIPs are the dominant strategies in “drugs and medical”. Empirical evidence from Hedge et al. (2009) shows that CIPs are mainly filed by R&D-intensive firms that patent heavily, and that they are more common in chemical and biological technology fields. According to Hedge et al. (2009), patents issuing from CIPs cover relatively important inventions and their use relates to a strategy of protecting “pioneering inventions.” In contrast, CAPs and divisions relate to less important patents assigned to capital-intensive firms, particularly in computer and semiconductor fields, and their use is consistent with “defensive patenting strategies”.

4.2 Application Dates vs. Priority Dates

In this section, we distinguish the patents that have at least one continuing application from the original patents. Figure 1.11 – 1.13 show that the yearly lags between the application dates and priority dates vary over time. On average the lag between the application and priority years is 2.58 years (s.d. of 0.69). If we consider a possibly alternative definition of priority date, i.e. the date
on which the last CIP was filed, the lag is 2.36 years (s.d. of 0.67). Between 1975 and 2008, the year lag remained around two years before it jumped to two and a half years in 1995 and then it fell back in 1996 and increased subsequently.

Figures 1.14 – 1.16 show that the lag between application and priority years differ by continuing types. CAPs have the longest lags between the application and priority years throughout the 1975 and 2008 period, CIPs have the shortest, and divisions are in the middle.

Figures 1.17 – 1.19 shows that the lag between application and priority years vary across the technology categories. On average the lag between application and priority years is three and a half years for the patents in “computers and communications” and three years for the patents in “drugs and medical”. The other fields have relatively shorter lags of two and a half years.

The significance of the time lag between application and priority dates or years motivate us to re-examine the R&D and patenting relationship. In prior literature (e.g. Hall and Ziedonis, 2001), the application year has been used as a proxy for invention date and a time placer for patenting. Given that overall about one third of patents are issued from continuing applications, the priority date should be a better proxy for invention date than application date in the context of the R&D-patenting estimation. In Chapter II, we test this hypothesis and show that estimated coefficient in the patent production function is larger when using priority date prior particularly prior to the 1995 GATT change.

4.3 Quality of Continued Patents

We use the number of claims and forward citations to the patents as two measurements for the quality of patents.

The claims in the patent specification determines the boundaries of the exclusive property
rights protected by the patent. Only the technology or aspects covered in the claims can be legally protected and enforced. The patentee has an incentive claim as much as possible in the application but the patent examiner may require that the claims be narrowed before granting. The number of claims determine the technological breadth of a patent as well as the expected market value of the patent.

Figures 1.20 – 1.22 compare the quality of patents that have at least one continuing applications to the original patents in terms of the number of claims. While the trends in the number of claims in the patents with continuing applications and the original patents have been closely aligned from 1975 to 2010, the total number of claims of the patents with continuing applications exceeds those of the original patents.

Moreover, the comparison of the number of claims differs across the three different types of continuing applications. Patents with CIPs have many more claims than the original patents during the 1975-2010 period. This can be attributed to the functions of filing CIPs. A CIP can be filed when an applicant discovers that a potentially patentable invention that was disclosed in the original application was not claimed. Filing the CIP also allows the inventor to add new materials to describe the improvements that are built upon the original inventions. Patents granted from CIPs cover relatively important inventions and their use appears consistent with a strategy of protecting “pioneering inventions” (Hedge, et al., 2009).

In contrast, divisions consistently have about five fewer claims compared to the original patents during the same period. This is apparently because of the functions of filing divisions. Divisions are often used due to a restriction requirement made by a patent examiner, in which case the claims need to be divided out of the original applications. According to MPEP, each patent application must be strictly focused on one invention. A divisional application simply takes a
section of the parent application’s claims and moves them into its own application that retains the parent’s filing date.

Forward citations, i.e. the citations a patent receives from subsequent patents, are often seen as an indicator of technological quality. Forward citations are related most directly to technological importance. Forward citations over the long term indicate an innovation has contributed to future research. Citations soon after patent application suggests rapid recognition of its importance as well as the presence of others working in a similar area, and thus the expectation of a valuable technological area (Lanjouw and Schankerman, 2004).

Figures 1.23 – 1.25 compare the number of forward citations to the patents with continuing applications to the original ones. During the entire 1975-2010 period, Patents issuing from overall continuing applications, CAPs and CIPs have demonstrated a higher quality in terms of forward citations than the original patents. Compared to the original patents, patents with divisions have fewer forward citations before 1995 but they were matched in the second half of the 1990s and the number of the patents with divisions has exceeded since then.

4.4 Patent Clustering Analysis

The distinction between application and priority dates also inspires us to investigate the cluster of patents from the same inventor which may be interpreted as evidence of a single scientific project that spins off multiple inventions.

We use the Institute for Quantitative Social Science (IQSS) inventor disambiguation data produced by Li et al. (2014) for inventor information. We use the NBER Patent Data Project (PDP) assignee dataset (Hall et al., 2001) for the employment information of the inventors. This enables us to measure the clustering patents and mobility of inventors by priority and application
date and then compare the results.

First, we define clustering patents as at least two patents that have the same inventor filed within 90 days. We consider an inventor as a clustering inventor if he has at least one record of clustering patent. As the NBER assignee data is available from 1976 to 2006, we consider all utility patents issued during that period. During 1976-2006, there were 5,332,069 patent grants in total and 863,570 inventors associated with these patents.

Table 1.1 shows that there were 534,924 (i.e., about 32%) more clustering patents identified when we use priority dates instead of using application dates. Table 1.2 shows that there were 158,637 (38%) more clustering inventors identified by priority dates instead of application dates.

Second, we use assignees as proxy for employers of inventors and calculate the move ratio for each inventor. We identify a “boomerang” when an inventor changes his job to a new employer but subsequently moves back to the original employer. We hypothesize that the move ratio (or the number of boomerangs) is lower when using priority date compared to by application date.

The total number of boomerangs is 155,474 when we use application dates to determine the inventors’ career paths and the number is 137,478 when using priority dates. We find that substituting priority dates for application dates can eliminate more than half of the boomerangs while it creates some new boomerangs. In fact, there were 95,012 boomerangs eliminated by using priority dates instead of application dates.

We also calculate the number of inventors who have at least one boomerang and compare the results indicated by the two different dates. There were 85,536 inventors identified having boomerang records ever in their career histories by application dates and 22,473 (about a quarter) among those can be eliminated when using priority dates.
Third, we examine the frequency of inventors by patent numbers and compare it to the frequency of inventors by numbers of patent clusters. We define a patent cluster as a group of patents for which any two consecutive applications are within 90 days. Our primary finding is that it will increase the number of single patent (or cluster) inventors when we count the patent clusters for each inventor instead of individual patents.

Figure 1.26 shows that 52.36% (i.e. 949,306 out of 1,812,876) inventors have only one patent during 1976-2006, but 61.15% (i.e. 1,108,536 out of 1,812,876) inventors have only one cluster during the period. It implies that 159,230 inventors have more than one patents of which the priority dates are no more than 90 days apart and they have only one patent cluster for the entire sample period. We also calculate the cluster numbers by using application date instead of priority date. We find that the number of single patent inventors is 5.42% less by using application date than by priority date.

5 Conclusion

This chapter provides a description on continuing applications and implications for use of patents to study innovation. Our study represents the first effort of which we are aware to analyze patent application data on continuing histories, providing information on the effects of 1995 GATT legislation changes in the US patent law intended to curb “submarine patenting”. We employ novel data on filings of three types of continuing applications – CAPs, CIPs, and divisions to distinguish among the motives for continuing patents. We introduce the 1995 GATT change in patent term and find that the GATT legislation reduced continuing applications overall and

---

8 We also examine the alternatives by changing the time length to 30 or 60 days and the results show the same patterns.
mitigated submarine patenting. The disparate impact of the GATT change will be most acute on pharmaceutical and chemical industries.

We compare the quality of patents in terms of number of claims and forward citations. Our analysis results suggest that patents issued from continuing applications have better quality than those issued from new applications. Specifically, the CIPs have more claims and more forward citations compared to the other two types of continuing applications as well as to the new applications. We also provide a starting point to analyze the clustering patents and inventor mobility. A possible future work is to develop the analysis following the strategies by Marx (2009).
BIBLIOGRAPHY


Table 1.1: Number of patent clustering

<table>
<thead>
<tr>
<th>Application date</th>
<th>Priority data</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,683,239</td>
<td>2,218,163</td>
<td>534,924</td>
</tr>
</tbody>
</table>

Table 1.2: Number of inventor clustering

<table>
<thead>
<tr>
<th>Application date</th>
<th>Priority data</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>417,191</td>
<td>575,828</td>
<td>158,637</td>
</tr>
</tbody>
</table>
Figure 1.1: Counts of patents granted from continuing applications by grant year

Figure 1.2: Counts of patents granted from continuing applications by application year
**Figure 1.3:** Counts of patents granted from continuing applications across technology fields by grant year

![Graph showing patent counts by grant year across different technology fields.](image)

**Figure 1.4:** Counts of patents granted from continuing applications across technology fields by application year

![Graph showing patent counts by application year across different technology fields.](image)
Figure 1.5: Trends in continuing applications by grant year

Figure 1.6: Trends in continuing applications by application year

Figure 1.7: Trends in continuing applications by priority year
Figure 1.8: Trends in continuing applications across technology fields by grant year

Figure 1.9: Trends in continuing applications across technology fields by application year

Figure 1.10: Trends in continuing applications across technology fields by priority year
Figure 1.11: Lag between priority dates and application dates by grant year

Figure 1.12: Lag between priority dates and application dates by application year

Figure 1.13: Lag between priority dates and application dates by priority year
Figure 1.14: Lag between priority dates and application dates across continuing types by grant year

Figure 1.15: Lag between priority dates and application dates across continuing types by application year

Figure 1.16: Lag between priority dates and application dates across continuing types by priority year
Figure 1.17: Lag between priority dates and application dates across technology fields by grant year
Figure 1.18: Lag between priority dates and application dates across technology fields by application year
Figure 1.19: Lag between priority dates and application dates across technology fields by priority year
Figure 1.20: Compare quality of patents issued from each type of continuing applications with those from original applications - number of claims by grant year
Figure 1.21: Compare quality of patents issued from each type of continuing applications with those from original applications - number of claims by application year.
Figure 1.22: Compare quality of patents issued from each type of continuing applications with those from original applications - number of claims by priority year
Figure 1.23: Compare quality of patents issued from each type of continuing applications with those from original applications - number of forward citations by grant year
Figure 1.24: Compare quality of patents issued from each type of continuing applications with those from original applications - number of forward citations by application year
Figure 1.25: Compare quality of patents issued from each type of continuing applications with those from original applications - number of forward citations by priority year
Figure 1.26: Compare distribution of inventors by numbers of individual patents with patent clusters - cluster defined as 90 days by priority date
Figure 1.27: Truncation problem analysis - plot lag between priority dates and application dates and compare with the non-truncated vs truncated observations
Figure 1.28: Truncation problem analysis - plot ratio between the number of forward citations to patents from continuing applications and patents from original applications and compare with the non-truncated vs truncated observations.
CHAPTER II

The R&D-Patenting Relationship Revisited: An Empirical Analysis on the U.S. Chemical and Pharmaceutical Industries

1 Introduction

The innovative activities by firms are considered to be the main driving force of the growth process in advanced economies. The relationship between R&D and patents which are taken as an output indicator of the new creation or innovation has been widely studied in the literature. Starting with the seminal work by Pakes and Griliches (1980) and Hausman et al. (1984), various count panel data models have been proposed and applied to analyze the relationship between R&D expenditures and patents known as the patent production function. Patent counts are characterized by long upper tails with relatively low medians and some proportion of zeros. The skewness in patent distribution may be attributed to the presence of unobserved heterogeneity such as firm specific propensity to patent, which is generally modelled multiplicatively as an exponential (or log-link) model. Hall et al. (1986) applied fixed- and random-effects Poisson and negative binomial models to analyze the R&D-patents relationship by the U.S. manufacturing firms during the 1970’s, and Hall and Ziedonis (2001) then extended the data set to the 1990’s and used the same models to analyze that for the U.S. semiconductor firms. The consistency of the standard exponential estimates relies on the strict exogeneity of the explanatory variables. Chamberlain (1992) and Wooldridge (1997) developed a quasi-differenced GMM estimator that is consistent

---

9 General discussions of the exponential count panel data models are in Gourieroux et al., 1984, McCullagh and Nelder, 1989, Cameron and Trivedi, 1986, 1998, etc.
for count panel data models with weakly exogenous or predetermined regressors. This quasi-differenced GMM estimator has been applied to the analysis of the R&D-patents relationship by Montalvo (1997), Crep’ on and Duguet (1997), Cincera (1997), and Gurmu and P’erez-Sebastia´n (2008). Windmeijer (2000) proposed a variant of the quasi-differenced GMM estimator that is consistent when a regressor is contemporaneously correlated with the idiosyncratic error term, or we say that the repressor is endogenous. Blundell et al. (2002) extended the quasi-differenced GMM estimators with an application to a dynamic linear feedback model. Blundell et al. (2002) also proposed an alternative estimator, the pre-sample mean (PSM) estimator, based on pre-sample information on the dependent variable. Some Monte Carlo experiments by Blundell et al. (2002) indicate that the PSM estimator performs well compared to the conventional maximum likelihood and quasi-differenced GMM estimators. This PSM estimator was employed by Wang and Hagedoorn (2014) to analyze the lag structure of the R&D-patents relationship. Results were again in line with Blundell et al. (2002).

So far the earlier work in this area, as aforementioned, has applied various count panel data models to investigate the relationship between R&D and patenting using application date as a proxy for the invention date in their estimations. However, no previous work has considered using priority date as a proxy for the invention date on account of the frequent uses of continuing applications. This paper revisits the classic research question regarding R&D-patents relationship in the context of the chemical and pharmaceutical industries in the U.S. during 1981-2003. We focus on chemical and pharmaceutical industries for the following main reasons. First, the

---

10 Also see Windmeijer (2000) and Kitazawa (2013) for other moment conditions that can be used with explanatory variables being predetermined.
11 For a survey of dynamic count panel data models, see Cameron and Trivedi, 2015.
12 In the U.S. inventors can file an application that claim the benefit of the filing date of an earlier application. The application filed later are known as continuing application. See Cheng et al. (2016) for a detailed discussion on this particular type of patent application. The filing date of the last application of a patent is defined as “application date” and the filing date of the earliest application is defined as “priority date”.

45
chemical and pharmaceutical industries as high-technology sectors are characterized by large patent propensity and high R&D intensity. Second, the change of marginal effect on R&D is essential by using priority date instead of application date especially for continuing patent applications, which are most frequently used by chemical and pharmaceutical firms compared to the other technological fields such as computers and electronics. Lastly, we examine the effects of a legislative change in patent term on continuing applications and the impact of the change is more influential to the firms in chemical and pharmaceutical industries compared to the other fields. Our empirical analysis, using both a standard negative binomial count data model and variants of GMM and PSM estimators which relax the strict exogeneity assumption, differs from previous work in three aspects. First, we use priority date instead of application date as a proxy for the invention date to link a firm’s R&D expenditures to patenting. We compare the estimated correlations in two time periods (pre-GATT\textsuperscript{13} and post-GATT) and show the effectiveness of the 1995 GATT change on continuing applications and submarine patents. Second, we develop a variant of the dynamic count panel data model that allows for endogenous regressors based on Windmeijer’s (2000) quasi-differenced GMM estimator and Blundell et al.’s (2002) linear feedback model and apply it to the patent production function estimation. In addition, we compare the estimation results from the fixed- and random-effects negative binomial models to the variant of quasi-differenced GMM for linear feedback model and PSM estimators. Lastly, we estimate using a new panel data set for the U.S. chemical and pharmaceutical industries from 1981 to 2003.

The layout of the paper is as follows. In the next Section, we provide the theoretical background and propose our main hypotheses. In Section 3, we describe the construction of the

\textsuperscript{13}GATT is an abbreviation for the General Agreement on Tariffs and Trade Uruguay Round implementing legislation. See Cheng et al. (2016) for an introduction to the policy change of GATT.
data set and the various variables. Section 4 presents the count panel data models underlying our empirical analysis and discusses the associated estimation techniques, and the empirical results are reported in Section 5. Section 6 concludes our main findings and discusses some possible future lines of work.

2 Theoretical Background and Hypotheses

The relationship between R&D and patents can be estimated by a knowledge production function first introduced by Pakes and Griliches (1980), which relates the number of successful patent applications made by a firm in a given year to its contemporary as well as past history of R&D expenditures, along with other firm characteristics such as size. The annual R&D expenditures are considered as investments which add to a firm’s knowledge stock, and patents are taken as an output indicator of new knowledge creation or innovation (Hall et al., 1986; Hausman et al., 1984). This relationship can be written as

\[ Q_{it} = F(R_{it}, R_{it-1}, ..., \beta, v_i), \]

where \( Q_{it} \) is a latent measure of technological output of firm \( i \) in period \( t \); \( R_{it} \) measures the corresponding R&D investment, \( \beta \) represents the vector of unknown technology parameters and \( v_i \) captures the firm specific propensity to patent. Observed patents, \( P_{it} \), are a noisy indicator of a firm’s technological output, where

\[ P_{it} = Q_{it} + \epsilon_{it} \]

with \( E(\epsilon_{it}|R_{it}, R_{it-1}, ..., v_i) = 0 \). Suppose that historic R&D investments are combined through a Cobb-Douglas technology to produce knowledge stock. Then

\[ Q_{it} = R_{it}^{\alpha_1} R_{it-1}^{\alpha_2} ... v_i. \]

This motivates the standard multiplicative distributed model specification which will be discussed in the estimation section. According to the previous work (Hall et al., 1986; Hausman et al., 1984), no coefficients except for the contemporary R&D variable were statistically significant either in the standard Poisson or negative binomial count panel data models, and therefore the past history of R&D expenditures is ignored in this study.
In prior literature\textsuperscript{14}, the patent application date was considered as a proxy for the invention date and used as the relevant time placer for patents in the context of the relationship being estimated between R&D and patents. However, firms may file continuing applications that can claim the priority date of the earlier patent in the United States Patent and Trademark Office (USPTO). There are two major benefits to the claim for the priority date of the parent application. First, it can prevent using any patents, published patent applications, non-patent publications, and certain other prior art that have become public between the filing date the parent application and that of the continuing application as prior art against the continuing application. Second, it can prevent using the parent application itself which has to be published 18 months after filing as prior art against the continuing application (Cheng et al., 2016).

The continuing applications are prevalently used by U.S. firms for strategic reasons especially in industries such as chemicals and pharmaceuticals where the pace of technology is rapid and firms advance quickly on innovations made by competitors.\textsuperscript{15} This motivates our first hypothesis as follows:

**Hypothesis 1.** The priority date is more closely tied to the timing of the new invention than the application date if the patent is granted from one or a string of continuing applications.

Moreover, the strategic uses traditionally pursued through filing continuing applications by the U.S. firms has been fundamentally changed due to the implementation of the 1995 General Agreement on Tariffs and Trade (GATT) Uruguay Round implementing legislation (P.L. 103-
In 1995 GATT legislation has changed the patent term from 17 years from the grant date to 20 years from the earliest application date. As a result, there was a substantial decline in filing continuing applications by firms since the effective date of the legislation change in 1995. The effectiveness of the 1995 GATT change has eliminated most of the related submarine patent issues.\textsuperscript{16} Beard (2008) propose that patent applicants use submarine patents to gain a competitive advantage over competitors by filing continuing applications to keep the patent “submerged” and unpublished, until a product is infringed by competitors. They find that although the submarine patent is relatively rare it poses a significant risk to inventors, such as a payment of royalties in order to avoid a costly legal battle, an award of higher royalties, or worse, a complete bar to sale of the new product.

Prior to the 1995 GATT change, U.S. patent applicants would pursue the benefit from submarine patents by delaying the expiration date of a patent through filing a succession of continuing applications. However, due to the legal change in patent term there would be little benefit in postponing the issuance of patent and thus it would curtail the former practice of filing continuing applications. The disparate impact of the GATT change will be most acute on industries such as chemicals and pharmaceuticals. Figure 2.1 and Figure 2.2 show that the surges in continuing applications around 1995 are most remarkable in the chemicals and pharmaceuticals among other technology fields. This could be because the patent term is more likely to be valuable towards the end for patents in the fields such as chemicals and pharmaceuticals and the 1995 GATT change which would result in a reduction in the term is far more important to the owners of these patents. The yearly change of fraction of patents granted from each type of continuing applications across all technology categories in Figure 2.3 confirms the fact that the GATT legislation has

\textsuperscript{16} “Submarine” patents refer to patents that were issued after long pending application periods until the claimed invention became market ready. Literature on submarine patents include Blount, 1999; Graham and Mowery, 2004.
greater impact on patents in chemical and pharmaceutical industries than others such as electronics and computers. We therefore propose the following main hypothesis:

**Hypothesis 2.** The correlation between R&D expenditure and patenting is relatively higher by priority date than application date prior to the GATT change and lower after that, which not only implies that the 1995 GATT change has effectively alleviated the submarine patent problems particularly in chemical and pharmaceutical industries, but also suggests that the policy change has affected the patenting strategies such as economizing on the filing fees.

## 3 Data and Variable Construction

The research setting for this study is the U.S. chemical and pharmaceutical industries. We used the NBER classification developed by Hall et al. (2001) which aggregated the U.S. Patent and Trademark Office (USPTO) technology classes into 6 categories (chemical, computers and communications, drugs and medical, electrical and electronics, mechanical, and miscellaneous “other”) to identify the classes in chemical and pharmaceutical fields i.e. categories (1) and (3). Then we used the USPTO concordance between the U.S. Patent Classification System (USPCS) and the Standard Industrial Classification System (SIC) to match the technology classes which have been identified in the chemical and pharmaceutical categories with the SIC codes.\(^\text{17}\)

The universe of the sample is the set of publicly traded U.S. firms in the chemical and pharmaceutical industries as described above that have data on employment, sales, gross capital, market value, and R&D for all years between 1981 and 1995 for pre-GATT analysis or between 1996 and 2003 for post-GATT analysis. The financial and employment data were obtained from Compustat maintained by Wharton Research Data Services (WRDS). The Compustat data file also

\(^{17}\) The SIC codes which represent the chemical and pharmaceutical industries are: 285, 284, 282, 286, 283, 281, 287, 13, 29, 22, 289, 30, 32, 34, 333, 334, 335, 336, 339, 20, 331, and 332.
provides each firm’s SIC code which identifies the technology category and the Global Company Key (GVKEY) as a unique identifier for the firm. All financial data were measured in US dollars and are inflation-adjusted in millions of year 1992 dollars.

A further scrutiny criterion for the sample firms is based on the absence of large jumps during the study period. Following Hall et al. (1986), a jump is defined as an increase in capital stock or employment of more than 100 per cent or a decrease of more than 50 per cent. This test was not applied unless the change in employment was greater than 500 employees or the change in capital stock was greater than two million dollars. We also removed firms which had abnormally small R&D values (less than $10,000) in one of the years.

We followed the NBER Patent Data Project (PDP) to match the patent data to Compustat firms, which overcomes two difficulties: first, the firms may be subsidiaries or have different names on Compustat, and second, although the initial assignees are the initial owners of patents, sometimes ownership of the organization changes through mergers, acquisitions, spinoffs, etc.

For the pre-GATT analysis, we ended up with a balanced panel sample composed of 111 firms covering the 15-year period from 1981 to 1995 for which there is a pre-sample history of patents from 1976-1980, giving 1665 firm-year observations.\(^{18}\) For post-GATT, we had a balanced panel sample of 193 firms covering the 8-year period from 1996 to 2003 for which the pre-sample period is 1991-1995, giving 1544 firm-year observations.

An important source of data for our study is the Technology Assessment and Forecast Database (TAF) maintained by the USPTO. We used the related application data file which contains information on the history of continuing application filings of each patent to identify the

\(^{18}\) Compared to the sample size of 642 firms in Hall et al. (1986), our sample is relatively small because of two reasons. First, we restrict our sample in chemical and pharmaceutical industries whereas Hall et al. (1986) include all manufacturing industries. Second, we exclude firms which do not have a pre-sample history of patents.
priority date for each patent. All patent application data are only based on applications that will be successfully granted as patents. Accordingly, we assigned a granted patent to the year in which the patent was originally applied for. For instance, a patent granted in 2002 but applied for in 2000 which is a continuation of application originally filed in 1998 is considered a 1998 patent.

### 3.1 Dependent Variable

The dependent variable for this study is the number of patents applied for by a chemical or pharmaceutical firm in a given year that are eventually granted, which is taken as an output indicator of firm innovativeness. As displayed in Table 2.1, the standard deviation of the number of patents is approximately twice as large as the mean, indicating the presence of an overdispersion in our patent data. As a result, the negative binomial regression is considered more appropriate than the Poisson regression. In addition, we test the regression on forward citations of patents and provide the results using our benchmark model.

### 3.2 Independent Variable

The main explanatory variable is the logarithm of the value of R&D investments. We measured a firm’s R&D investments by its annual R&D expenditures. The annual R&D expenditures are considered to be investments that add to a firm’s stock of knowledge (Hall et al., 1986). Table 2.1 shows that chemicals and pharmaceuticals are R&D-intensive firms, their average R&D expenditures amounting to approximately 125 (pre-GATT) and 218 (post-GATT) million dollars per year. All R&D expenditures are inflation-adjusted in millions of year 1992 dollars.\(^\text{19}\) When

\(^{19}\) We use GDP inflator for inflation adjustment. The data source is World Bank, World Development Indicators, ERS Estimates, and ERS Baseline Regional Aggregations.
employment (firm size) is also included, we normalize R&D by the number of employees to avoid confounding the R&D effect with the size effect.

### 3.3 Control Variables

The logarithm of employment is included to measure the firm size following the previous studies (Hall et al., 1986; Hausman et al., 1984).

We used the logarithm of pre-sample mean of patents as proxy for the unobserved heterogeneity in firm knowledge production (Blundell et al., 2002). Blundell et al. (2002) showed that the quasi-differenced GMM estimator can be severely biased in small samples. This is particularly the case when regressors are highly persistent and the instruments are therefore weak predictors of the endogenous variables in the differenced model. They proposed an alternative pre-sample mean (PSM) estimator that replaces the fixed effect by the pre-sample mean of the dependent variable. In our study the pre-sample mean is the average number of patents in the 5-year pre-sample period from 1976-1980 (for pre-GATT) and 1991-1995 (for post-GATT). To control for the time-varying effects, for instance, the different levels of backlog\(^{20}\) at a given time, we included annual time dummies for 1981-1995 (pre-GATT) and 1996-2003 (post-GATT) respectively.

### 4 Estimation

As the dependent variable of our estimation, the number of patents by each firm, is a count variable taking only non-negative integer values, we employ various count panel data models to investigate the relationship between R&D and patenting in the U.S. chemical and pharmaceutical

\(^{20}\) Backlog of patent application refers to the patent applications that have been filed and still remain unexamined.
industries. We will compare the estimated results of the correlation between the R&D and patents by using the priority date as the time placer for patents to the results by the application date as used in the previous studies.

4.1 Benchmark Model

Consider a standard multiplicative count panel data model with the conditional mean as

\[ E(P_{it} | \ln R_{it}, \ln E_{it}, w_i, \eta_i) = \exp(\beta_0 + \beta_1 \ln R_{it} + \beta_2 \ln E_{it} + w_i \theta_i + \eta_i) \]

(1)

where \( P_{it} \) are the number of patents originally applied for by firm \( i \) in period \( t \), which is a noisy indicator of a firm’s technological output, \( R_{it} \) measures the corresponding R&D investment, \( E_{it} \) is the level of employment for the observable firm-specific effects, \( w_i \) is a vector of time-specific variables (i.e., year fixed effects), and \( \eta_i \) captures unobserved individual fixed effects (i.e., firm-specific propensity to patent), which are commonly modeled multiplicatively in count panel data models.

The basic reference for estimating the above model is the Poisson and negative binomial estimations (Hausman et al., 1984). As the negative binomial estimation provides a better fit for our data than the Poisson estimation since the assumption of the equality of conditional mean and conditional variance may not hold (Hall et al., 1986; Hausman et al., 1984), we adopt the negative binomial estimation taking account of the overdispersed nature of patent counts. We use both fixed- and random-effects specifications to control for unobserved firm-specific propensity to patent, \( \eta_i \) in the negative binomial estimation. The problem with the random effects specification is that it is inconsistent when \( \eta_i \) is correlated with the explanatory variables. Conditional fixed-effects specification, however, allows for such correlations (Hall et al., 1986; Hausman et al.,
4.2 Endogenous Regressors

The consistency of the standard estimators presented above relies on the strict exogeneity of the explanatory variables. As previously discussed, the patent application tends to occur relatively early in the life of a research project and the bulk of R&D expenditures often occur after the application is made, new patents virtually generate the need for future R&D expenditures (Griliches, 1990; Hall et al., 1986), the R&D expenditures is therefore weakly exogenous or endogenous in our estimation models. When one or more regressors are endogenous, moment condition based on (1) is no longer valid. In this study, we follow Windmeijer’s (2000) quasi-differencing transformation and use the valid moment conditions as below to obtain consistent estimation.

\[
\sum_{i}^{N} \sum_{t=3}^{T} z_{it-2} \left( \frac{y_{it}}{\mu_{it}} - \frac{y_{it-1}}{\mu_{it-1}} \right) = 0 \tag{2}
\]

where \( \mu_{it} = \exp(x_{it}^{\prime}\beta) \), as defined in the conditional mean for a standard multiplicative count panel data model:

\[
E(y_{it} \mid x_{it}, \eta_{i}) = \exp(x_{it}^{\prime}\beta + \eta_{i}) = \mu_{it}v_{i} \tag{3}
\]

where \( y_{it} \) denote the dependent variable, \( x_{it} \) denote a vector of explanatory variables, and \( v_{i} = \exp(\eta_{i}) \) is a permanent scaling factor for the individual specific mean.

The second lag of the endogenous regressor can be used as an instrument. But we could use all further lags of \( x \) as instruments for a more efficient GMM estimator as below:

\[
z_{it-2} = (x_{it-2}, x_{it-3}, \ldots, x_{it}) \tag{4}
\]
4.3 Linear Feedback Model

We now introduce dynamics into the multiplicative distributed models using a dynamic linear feedback model (LFM) proposed by Blundell et al. (2002). In the dynamic linear feedback model, the lagged dependent variable enters the conditional mean specification linearly, which corresponds to the following:

\[ E(P_t | P_{t-1}, \ln R_t, \ln E_t, w_t, \eta_t) = \gamma P_{t-1} + \exp(\beta_0 + \beta_1 \ln R_t + \beta_2 \ln E_t + w_t \theta + \eta_t) \]  

(5)

where \( 1 - \gamma \) estimates the depreciation factor, \( \beta_1 \) and \( (1 - \gamma)\beta_1 \) are the long-run and short-run elasticities of patents with respect to R&D expenses, respectively.

For estimation by GMM, we develop a quasi-differenced estimator for the LFM model based on Windmeijer’s (2000) transformation as follows:

\[ \sum_{i=2}^{N} \sum_{t=3}^{T} z_{it-2} (\frac{y_{it} - \gamma y_{it-1}}{\mu_{it}} - \frac{y_{it-1} - \gamma y_{it-2}}{\mu_{it-1}}) = 0 \]  

(6)

We use the instruments that include all further lags of regressors as below:

\[ z_{it-2} = (y_{it-2}, y_{it-3}, ..., y_{it}, x_{it-2}, x_{it-3}, ..., x_{it}) \]  

(7)

4.4 Pre-sample Mean Estimator

The quasi-differenced GMM estimator is subject to some problems, namely small sample bias and imprecision, particularly when economic series are highly persistent as the instruments are then weak predictors of future changes (Blundell et al., 1995, 2002). As an alternative, Blundell et al. (2002) proposed a pre-sample mean (PSM) estimator using the pre-sample information on the dependent variable and they showed that the PSM estimator is consistent in the presence of endogenous regressors and correlated fixed effects. In view of both highly persistent patents and R&D series in our study, we also use the PSM estimator that replaces the unobserved
individual fixed effects by the pre-sample mean of the number of patents.

The pre-sample mean estimator solves the following moment conditions:

$$\sum_{t}^{N} \sum_{r=1}^{T} z_{it} (y_{it} - \exp(\beta_{0} + x_{it} \beta + \phi \ln \bar{y}_{ip})) = 0 \tag{8}$$

where $$\bar{y}_{ip} = (1/TP) \sum_{r=0}^{TP-1} y_{i0-r}$$ is the pre-sample mean of $$y$$, $$TP$$ is the number of pre-sample observations, $$\phi$$ is a parameter to be estimated, and a choice of the instruments $$z_{it} = (1, x_{it}, \ln \bar{y}_{ip})$$.

5 Empirical Results

The main findings for the maximum likelihood and GMM estimations are presented in Tables 2.2 – 2.4. Table 2.2 shows results from the standard negative binomial (NB) estimations of the benchmark model. In the pre-GATT period, the coefficient on $$\ln R&D$$ increases from [0.27-0.34] when using priority date instead of application date as a time placer for patents for the fixed-effects NB regression, and increases from [0.32-0.39] using priority date instead of application date for the random-effects NB regression. The coefficients on $$\ln R&D$$ and $$\ln$$ (firm size) are significant for both the fixed- and random-effects NB regressions. In the post-GATT period, the coefficients on $$\ln R&D$$ are lower by priority date compared to by application date for the negative binomial model in both the fixed- and random-effects specifications.

The higher coefficient on $$\ln R&D$$ by priority date compared to by application date during the pre-GATT period suggests that priority date rather than application date should be a better proxy for the invention date before 1995. Accordingly, the results from patent production estimations in prior literature can be refined by using priority date as a substitute for application date to link the timing of the new invention and patent grants. The lower coefficient on $$\ln R&D$$ by priority date than application date during the post-GATT period suggests that application date
should be a more accurate proxy for the invention date than priority date after 1995. This could be due to the impact of the 1995 GATT change on filing continuing patent applications, particularly for firms in the chemical and pharmaceutical industries, where continuing applications were frequently used as a strategy for submarine patents which could enable their assignees to threaten legal action against users of widely adopted technologies for infringement.

The results of regression on citation-weighted patents using the benchmark model are provided in Table 2.3. In the pre-GATT period, the coefficient on lnR&D increases from [0.32-0.40] when using priority date instead of application date as a time placer for patents for the fixed-effects NB regression, and increases from [0.36-0.41] using priority date instead of application date for the random-effects NB regression. In the post-GATT period, the coefficients on lnR&D are lower by priority date compared to by application date for the negative binomial model in both the fixed- and random-effects specifications.

Results for the less constrained GMM, LFM and PSM estimators which relax the strict exogeneity assumption are reported in Table 2.4. The coefficients on lnR&D are significant across all the regression models for both pre- and post-GATT periods. The results that the coefficient on lnR&D is higher by priority date than application date before 1995 and lower after 1995 for all these three models are consistent with the benchmark model. The results of the quasi-differenced GMM estimator assuming that lnR&D is endogenous and its corresponding dynamic LFM estimator are as displayed in columns GMM and LFM respectively. These quasi-differenced GMM and LFM estimates seem puzzling. The coefficients on the lagged dependent variable are negative, whereas the coefficient on lnR&D, although positive and significant, are considerably biased upwards with relatively large standard errors. These results arise due to a weak instruments problem as both the patents and the lnR&D series in our data set are highly persistent over time.
Consider next the results of the PSM estimator that utilizes the available pre-sample information on patents, which are presented in column PSM. We find that for the PSM estimator the estimated coefficients are close to the standard negative binomial estimates, with the coefficient on lnR&D equal to 0.36 by application date and 0.41 by priority date in the pre-GATT period. In the post-GATT period, the PSM estimators seem more stable than the other regressions, with the coefficient on lnR&D equal to 0.36 by application date and 0.32 by priority date. All these results, in terms of the direction of the changes in the estimated coefficients on lnR&D, are the same as the benchmark model.

6 Attenuation Bias Analysis

In this section, we test if the correlation of patents and R&D is sensitive to the time gap between application and priority years. Figure 2.4 and Figure 2.5 exhibit the distribution of gap between application and priority years at patent and firm level respectively. As shown in Figure 2.4, 78.38% of patents granted from 1976 to 2012 have the same application and priority year and 90 percent of patents have no more than two years of gap. At firm level, we calculate the average year gap of the patents received by each firm. Figure 2.5 shows that about one third of firms have an average year gap equal to zero.

Then we exclude those firms whose average year gap is zero and re-estimate the pre-GATT patents-R&D correlation using application and priority year respectively. We hypothesize that the increase in correlation will be more significant using priority year instead of application year after we remove the observations with close year gaps. The re-estimation results of regression on patents are presented in Table 2.5. According to Table 2.5, the coefficient on lnR&D increases

---

21 The distribution is truncated at the 99th percentile in Figure 2.4.
22 In our sample, 25 firms are removed from the list.
from [0.28-0.46] when using priority date instead of application date as a time placer for patents for the fixed-effects NB regression, and increases from [0.34-0.50] using priority date instead of application date for the random-effects NB regression. Table 2.6 shows the results of regression on patent citations. The coefficient on lnR&D increases from [0.30-0.61] when using priority date instead of application date as a time placer for patents for the fixed-effects NB regression, and increases from [0.40-0.63] using priority date instead of application date for the random-effects NB regression. The coefficient differences on both patents and citations are more statistically significant than before we exclude firms with slight application and priority year differences. This empirical evidence further supports our main hypothesis.

7 Conclusion and Discussions

The principal purpose of this study is to investigate whether priority date is a better proxy than application date for the invention date through an examination of the knowledge production function estimations in the context of the U.S. chemical and pharmaceutical industries during 1981-2003. The results from a benchmark negative binomial model show that R&D expenditures are more highly correlated with patenting when using priority date as a time placer for patents instead of using application date before the 1995 GATT change, but are less highly correlated after 1995. This may suggest that the results from previous studies that examined the R&D and patenting relationship can be refined by using priority date as a substitute for application date to link the timing of the new invention and patent grants. Our empirical results imply that the 1995 GATT legislation change in patent term has substantially reduced the continuing applications and alleviated the submarine patents that used to be prevalent in the chemical and pharmaceutical industries prior to 1995. We also develop a new transformation of the quasi-differenced GMM
estimator for endogenous regressors with a dynamic linear feedback model and compare the estimation results from other various count panel data models. Estimation results in terms of the direction of the changes in the estimated coefficients on R&D expenditures by using priority date and application date during the pre-GATT and post-GATT periods are consistent across all empirical specifications. In line with the previous work the pre-sample mean estimator for our data set performs well in comparison to the quasi-differenced GMM estimators.

Our study is subject to some limitations which suggest avenues for future research. Firstly, the number of patents often proxies for the number of inventions which is considered an imperfect measure for a firm’s innovative output. We therefore can use forward citations and claims as measurements for quality of patents as dependent variables in the knowledge production function estimations. The second possible extension is to introduce R&D user cost which incorporates the R&D tax credit incentives for firms as an instrumental variable for the endogenous R&D expenditures in our GMM frameworks.
BIBLIOGRAPHY


StataCorp. (2013) *Stata Statistical Software: Release 13* College Station, TX: StataCorp LP.


Table 2.1: Summary statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>S.D.</th>
<th>Between S.D.</th>
<th>Within S.D.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Pre-GATT) Application year&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patents</td>
<td>38.706</td>
<td>74.084</td>
<td>70.043</td>
<td>24.959</td>
<td>0</td>
<td>685</td>
</tr>
<tr>
<td>Citations</td>
<td>519.285</td>
<td>1059.717</td>
<td>980.851</td>
<td>410.932</td>
<td>0</td>
<td>1528</td>
</tr>
<tr>
<td>R&amp;D ($M) e</td>
<td>122.251</td>
<td>241.383</td>
<td>225.748</td>
<td>87.891</td>
<td>0.039</td>
<td>1543.789</td>
</tr>
<tr>
<td>Employment (1000s)</td>
<td>22.113</td>
<td>29.613</td>
<td>28.471</td>
<td>8.545</td>
<td>0</td>
<td>206.4</td>
</tr>
<tr>
<td>Pre-sample mean</td>
<td>37.65</td>
<td>63.902</td>
<td>64.168</td>
<td>0</td>
<td>0.2</td>
<td>322</td>
</tr>
</tbody>
</table>

| Priority year<sup>b</sup> |        |         |              |             |      |      |
| Patents                   | 40.719 | 76.394  | 71.848       | 26.782      | 0    | 747  |
| Citations                 | 553.455| 1199.964| 1030.37      | 448.987     | 0    | 10365 |
| R&D ($M) e                | 125.725| 242.926 | 226.887      | 89.266      | 0.039| 1543.789 |
| Employment (1000s)        | 22.511 | 29.728  | 28.571       | 8.622       | 0.003| 206.4 |
| Pre-sample mean           | 36.771 | 61.383  | 61.643       | 0           | 0.2  | 308.8 |

| (Post-GATT) Application year<sup>c</sup> |        |         |              |             |      |      |
| Patents                   | 28.405 | 67.153  | 63.681       | 21.731      | 0    | 594  |
| Citations                 | 253.335| 814.989 | 702.543      | 415.732     | 0    | 11053 |
| R&D ($M) e                | 213.233| 635.167 | 575.368      | 271.771     | 0.017| 10433.894 |
| Employment (1000s)        | 16.185 | 33.839  | 33.556       | 4.898       | 0    | 306  |
| Pre-sample mean           | 23.999 | 59.36   | 59.491       | 0           | 0.2  | 508.4 |

| Priority year<sup>d</sup> |        |         |              |             |      |      |
| Patents                   | 26.205 | 62.034  | 57.725       | 23.045      | 0    | 563  |
| Citations                 | 220.073| 711.379 | 605.177      | 376.141     | 0    | 8671 |
| R&D ($M) e                | 218.426| 640.888 | 580.404      | 274.585     | 0.017| 10433.894 |
| Employment (1000s)        | 16.915 | 34.36   | 34.077       | 4.961       | 0.006| 306  |
| Pre-sample mean           | 23.602 | 57.833  | 57.964       | 0           | 0.2  | 465.4 |

<sup>a</sup> 1,695 observations (113 firms), 1981-1995.
<sup>b</sup> 1,665 observations (111 firms), 1981-1995.
<sup>c</sup> 1576 observations (197 firms), 1996-2003.
<sup>d</sup> 1544 observations (193 firms), 1996-2003.
<sup>e</sup> R&D expenditures are inflation-adjusted in 1992 dollars.
Table 2.2: Estimates of patenting propensity (I)\textsuperscript{a}

<table>
<thead>
<tr>
<th>Variable</th>
<th>NB (fixed-effects) App. Year</th>
<th>NB (fixed-effects) Pri. Year</th>
<th>NB (random-effects) App. Year</th>
<th>NB (random-effects) Pri. Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Pre-GATT)\textsuperscript{b}</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(R&amp;D)</td>
<td>0.269** (0.038)</td>
<td>0.343** (0.037)</td>
<td>0.321** (0.036)</td>
<td>0.387** (0.035)</td>
</tr>
<tr>
<td>ln(firm size)</td>
<td>0.360** (0.033)</td>
<td>0.374** (0.032)</td>
<td>0.441** (0.029)</td>
<td>0.444** (0.028)</td>
</tr>
<tr>
<td>Year dummies</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
</tr>
<tr>
<td>Constant</td>
<td>0.436** (0.121)</td>
<td>-0.192 (0.127)</td>
<td>0.184 (0.114)</td>
<td>-0.424** (0.119)</td>
</tr>
<tr>
<td>(Post-GATT)\textsuperscript{c}</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(R&amp;D)</td>
<td>0.132** (0.036)</td>
<td>0.006 (0.038)</td>
<td>0.162** (0.032)</td>
<td>0.095** (0.034)</td>
</tr>
<tr>
<td>ln(firm size)</td>
<td>-0.027 (0.032)</td>
<td>0.072* (0.032)</td>
<td>0.147** (0.028)</td>
<td>0.219** (0.027)</td>
</tr>
<tr>
<td>Year dummies</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
</tr>
<tr>
<td>Constant</td>
<td>0.812** (0.159)</td>
<td>0.148 (0.171)</td>
<td>0.311* (0.149)</td>
<td>-0.394* (0.159)</td>
</tr>
</tbody>
</table>

\textsuperscript{a} \textsuperscript{†} p < 0.1, \textsuperscript{*} p < 0.05, \textsuperscript{**} p < 0.01. Heteroskedastic-consistent standard errors are in parentheses. The method of estimation is maximum likelihood for the Negative Binomial (NB) model.

\textsuperscript{b} The sample is a balanced panel of 113 (111 by priority year) firms covering the period from 1981 to 1995.

\textsuperscript{c} The sample is a balanced panel of 197 (193 by priority year) firms covering the period from 1996 to 2003.
Table 2.3: Estimates of citation-weighted patenting propensity

<table>
<thead>
<tr>
<th>Variable</th>
<th>NB (fixed-effects)</th>
<th>NB (fixed-effects)</th>
<th>NB (random-effects)</th>
<th>NB (random-effects)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>App. Year</td>
<td>Pri. Year</td>
<td>App. Year</td>
<td>Pri. Year</td>
</tr>
<tr>
<td>(Pre-GATT)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(R&amp;D)</td>
<td>0.323**</td>
<td>0.398**</td>
<td>0.357**</td>
<td>0.415**</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.034)</td>
<td>(0.034)</td>
<td>(0.033)</td>
</tr>
<tr>
<td>ln(firm size)</td>
<td>0.527**</td>
<td>0.503**</td>
<td>0.53**</td>
<td>0.501**</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.024)</td>
<td>(0.022)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>Year dummies</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
</tr>
<tr>
<td>Constant</td>
<td>0.194*</td>
<td>-0.019</td>
<td>0.292*</td>
<td>0.035**</td>
</tr>
<tr>
<td></td>
<td>(0.13)</td>
<td>(0.13)</td>
<td>(0.126)</td>
<td>(0.126)</td>
</tr>
<tr>
<td>(Post-GATT)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(R&amp;D)</td>
<td>0.259**</td>
<td>0.242**</td>
<td>0.3**</td>
<td>0.275**</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.03)</td>
<td>(0.026)</td>
<td>(0.028)</td>
</tr>
<tr>
<td>ln(firm size)</td>
<td>0.401**</td>
<td>0.443**</td>
<td>0.447**</td>
<td>0.469**</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.022)</td>
<td>(0.018)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Year dummies</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
</tr>
<tr>
<td>Constant</td>
<td>-2.008**</td>
<td>-2.642**</td>
<td>-2.038**</td>
<td>-2.651**</td>
</tr>
<tr>
<td></td>
<td>(0.128)</td>
<td>(0.149)</td>
<td>(0.126)</td>
<td>(0.147)</td>
</tr>
</tbody>
</table>

a † p < 0.1, * p < 0.05, ** p < 0.01. Heteroskedastic-consistent standard errors are in parentheses. The method of estimation is maximum likelihood for the Negative Binomial (NB) model.

b The sample is a balanced panel of 113 (111 by priority year) firms covering the period from 1981 to 1995.

c The sample is a balanced panel of 197 (193 by priority year) firms covering the period from 1996 to 2003.
Table 2.4: Estimates of patenting propensity (II) a

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ln(R&amp;D)</td>
<td>0.789**</td>
<td>0.803**</td>
<td>0.797**</td>
<td>0.832**</td>
<td>0.356**</td>
<td>0.412**</td>
</tr>
<tr>
<td></td>
<td>(0.199)</td>
<td>(0.185)</td>
<td>(0.157)</td>
<td>(0.160)</td>
<td>(0.050)</td>
<td>(0.065)</td>
</tr>
<tr>
<td>ln(firm size)</td>
<td>0.708**</td>
<td>0.749**</td>
<td>0.757**</td>
<td>0.804**</td>
<td>0.342**</td>
<td>0.364**</td>
</tr>
<tr>
<td></td>
<td>(0.120)</td>
<td>(0.125)</td>
<td>(0.105)</td>
<td>(0.106)</td>
<td>(0.072)</td>
<td>(0.077)</td>
</tr>
<tr>
<td>Patent-1</td>
<td>-0.392*</td>
<td>-0.447**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.163)</td>
<td>(0.070)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(pre-sample mean)</td>
<td></td>
<td></td>
<td>0.547**</td>
<td>0.487**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.067)</td>
<td>(0.071)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.193</td>
<td>0.325</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.219)</td>
<td>(0.225)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(R&amp;D)</td>
<td>0.919**</td>
<td>0.837**</td>
<td>0.994**</td>
<td>0.816**</td>
<td>0.359**</td>
<td>0.320**</td>
</tr>
<tr>
<td></td>
<td>(0.194)</td>
<td>(0.176)</td>
<td>(0.141)</td>
<td>(0.148)</td>
<td>(0.051)</td>
<td>(0.050)</td>
</tr>
<tr>
<td>ln(firm size)</td>
<td>0.592**</td>
<td>0.662**</td>
<td>0.656**</td>
<td>0.667**</td>
<td>0.461**</td>
<td>0.443</td>
</tr>
<tr>
<td></td>
<td>(0.138)</td>
<td>(0.081)</td>
<td>(0.129)</td>
<td>(0.082)</td>
<td>(0.052)</td>
<td>(0.048)</td>
</tr>
<tr>
<td>Patent-1</td>
<td>-0.220</td>
<td>-0.111†</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.146)</td>
<td>(0.067)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(pre-sample mean)</td>
<td></td>
<td></td>
<td>0.348**</td>
<td>0.374**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.056)</td>
<td>(0.048)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.624**</td>
<td>0.577</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.156)</td>
<td>(0.162)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a † p < 0.1, *p < 0.05, ** p < 0.01. Heteroskedastic-consistent standard errors are in parentheses. GMM is quasi-differenced GMM, LFM is linear feedback model for quasi-differenced GMM, and PSM is pre-sample mean estimation. We include all further lags of xit as instruments in GMM estimation.

b The sample is a balanced panel of 113 (111 by priority year) firms covering the period from 1981 to 1995. The pre-sample mean of patents uses the years 1975-1980.

c The sample is a balanced panel of 197 (193 by priority year) firms covering the period from 1996 to 2003. The pre-sample mean of patents uses the years 1991-1995.
<table>
<thead>
<tr>
<th>Variable</th>
<th>NB (fixed-effects) App. Year</th>
<th>NB (fixed-effects) Pri. Year</th>
<th>NB (random-effects) App. Year</th>
<th>NB (random-effects) Pri. Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln(R&amp;D)</td>
<td>0.278**</td>
<td>0.462**</td>
<td>0.342**</td>
<td>0.502**</td>
</tr>
<tr>
<td></td>
<td>(0.046)</td>
<td>(0.039)</td>
<td>(0.043)</td>
<td>(0.036)</td>
</tr>
<tr>
<td>ln(firm size)</td>
<td>0.333**</td>
<td>0.482**</td>
<td>0.434**</td>
<td>0.555**</td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
<td>(0.034)</td>
<td>(0.035)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Year dummies</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
</tr>
<tr>
<td>Constant</td>
<td>1.534**</td>
<td>1.47**</td>
<td>1.48**</td>
<td>1.372**</td>
</tr>
<tr>
<td></td>
<td>(0.168)</td>
<td>(0.144)</td>
<td>(0.156)</td>
<td>(0.135)</td>
</tr>
</tbody>
</table>

\(^a\) \(p < 0.1\), \(^b\) \(p < 0.05\), \(^**\) \(p < 0.01\). Heteroskedastic-consistent standard errors are in parentheses. The method of estimation is maximum likelihood for the Negative Binomial (NB) model. 
\(^b\) The sample is a balanced panel of 87 (83 by priority year) firms covering the period from 1981 to 1995.
Table 2.6: Re-estimates of citation-weighted patenting propensity\textsuperscript{a, b}

<table>
<thead>
<tr>
<th>Variable</th>
<th>NB (fixed-effects)</th>
<th>NB (fixed-effects)</th>
<th>NB (random-effects)</th>
<th>NB (random-effects)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>App. Year</td>
<td>Pri. Year</td>
<td>App. Year</td>
<td>Pri. Year</td>
</tr>
<tr>
<td>ln(R&amp;D)</td>
<td>0.303**</td>
<td>0.606**</td>
<td>0.392**</td>
<td>0.633**</td>
</tr>
<tr>
<td></td>
<td>(0.041)</td>
<td>(0.036)</td>
<td>(0.039)</td>
<td>(0.035)</td>
</tr>
<tr>
<td>ln(firm size)</td>
<td>0.495</td>
<td>0.672**</td>
<td>0.544**</td>
<td>0.691**</td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.026)</td>
<td>(0.026)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>Year dummies</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
</tr>
<tr>
<td>Constant</td>
<td>0.206**</td>
<td>0.431**</td>
<td>0.329*</td>
<td>0.451**</td>
</tr>
<tr>
<td></td>
<td>(0.149)</td>
<td>(0.127)</td>
<td>(0.14)</td>
<td>(0.122)</td>
</tr>
</tbody>
</table>

\textsuperscript{a} \textsuperscript{b} p < 0.1, \textsuperscript{b} p < 0.05, \textsuperscript{**} p < 0.01. Heteroskedastic-consistent standard errors are in parentheses. The method of estimation is maximum likelihood for the Negative Binomial (NB) model. The sample is a balanced panel of 87 (83 by priority year) firms covering the period from 1981 to 1995.
Figure 2.1: Counts of patents granted from continuing applications across technology fields

![Counts of patents granted from continuing applications across technology fields](image1)

Figure 2.2: Fraction of patents granted from continuing applications across technology fields

![Fraction of patents granted from continuing applications across technology fields](image2)
Figure 2.3: Frequency of each type of continuing applications across technology fields
Figure 2.4: Distribution of time lag between application and priority year at patent level

Figure 2.5: Distribution of time lag between application and priority year at firm level
CHAPTER III

Impacts of Inventor Team Composition and Persistence on Value of Innovation

1 Introduction

Teamwork of inventors has an essential role in technological innovation and economic development. Recent studies have revealed a rising importance for teams rather than solo researchers. Wutchy et al. (2007) show that both scientific research and inventions in teams are increasing. They indicate that teams can generate better outcomes than lone researchers on average. Jones (2009) argues that this trend is the consequence of the growing specialization of innovators. Empirical evidence has also suggested that teamwork fosters innovation and technological progress. Singh and Fleming (2010) examine the productivity of teams of inventors versus that of lone inventors. They show that part of the advantage of teams of inventors with respect to lone inventors is due to the higher knowledge variety encompassed by teams.

If teams can perform better, our next question is what attributes in teams affect performance outcomes. Although there is a huge body of literature on the team-work formation, very little is known about how attributes of inventor teams affect their productivity. The value of the innovation output will depend essentially on the human capital and ability of the innovators to combine their existing pieces of knowledge and implement them to achieve the constructive end. Particularly the knowledge set diversity across team members is potentially an important contribution to the circulation of ideas and creativity of innovation outputs. Prior research suggests that teams that
combine a more varied knowledge set enjoy more room for recombination and more alternative paths to solve problems but they also risk suffering more from malfunctions (Paulus, 2000; Jackson, 1996). Melero and Palomeras (2015) argue that internal distribution of knowledge variety among team members is relevant to the generation of innovations in teams of inventors. They indicate that team-level knowledge variety can then be based on the contribution of some generalist inventor(s) or the combination of specialized contributions. “For a given level of team knowledge variety, the presence of generalists in an innovation team enables a more effective recombination of knowledge and attenuates the typical barriers affecting team-working processes. On the other hand, the lack of specialized contributions in such teams may hamper the process of adapting each recombined component in the search for an innovative solution. Thus, we expect teams including generalists to perform better than otherwise comparable teams in contexts where there is not a well-defined path to combine knowledge and the advantage of specialized contributions plays only a secondary role.” (Melero and Palomera, 2015)

Another stream of literature on collaboration networks has made essential contributions in proving that inventor collaboration fosters technological innovativeness. For instance, Balconi et al. (2004) exploits Italian patent data to construct the network of collaborative relationship with a focus on the role of academia. Singh (2005) uses a social geographic proximity to examine interpersonal networks as important determinants of observed patterns of knowledge diffusion. Lee (2010) explores endogenous formation of collaborative inventor networks. While there is a wide literature on inventor collaborations using data on patent co-inventorship, very few studies focus on team persistence. Kim and Marschke (2015) investigate the determinants of team persistence. They test a dynamic team formation model and find that researcher mobility reduces team persistence with team size
held constant. As stated by Kim and Marschke (2015), there is a wide management literature that examines persistence in collaboration, but we are unaware of any analysis of this aspect of teamwork in the economics literature.

Katz (1982) reports that the R&D team’s longevity’s effect on productivity is possibly quadratic, peaking at two to four years from the team’s inception. Akgun and Lynn (2002) find that in product development teams in R&D-performing firms a team’s longevity has a positive effect on productivity-related outcomes including team learning and cycle time, but not when there is a high degree of market and technical turbulence. The literature on repeated collaboration indicates that repetitive collaborations with the same co-inventor build mutual trust and confidence (e.g. Gulati, 1995; Kogut, 1989). Cowan et al. (2005) argue that past collaborations increase the probability of a successful collaboration. Fleming et al. (2007) find that an inventor’s past collaboration network will strongly influence subsequent productivity. To the best of our knowledge, no studies have analyzed the impact of team persistence on patent quality.

This paper aims at contributing to the literature by analyzing how skillset composition and persistence of teams affect their patenting innovation output. Firstly, we focus on the role of skillset diversity distributed among inventors as well as the role of generalist inventors on the relationship between the team composition and the innovative productivity. Secondly, we explore the impact of team persistence also known as the repeated collaborations of co-inventorship on innovation output quality. We use patent data to identify teams of inventors responsible for the creation of the underlying innovation and use the inventor information on patents as a proxy for compositions of inventor teams. Empirical results support our hypotheses: (1) both the skillset diversity and the presence of generalists increase the team output quality; (2) a higher degree of team persistence also
increases the output quality.

The layout of the paper is as follows. In the next Section, we provide our quantitative measurements for team composition diversity and persistence and propose our main hypotheses. In Section 3, we describe our data and variable construction and the estimation methods. Section 4 presents the empirical results. Section 5 concludes our main findings and discusses some possible future lines of work.

2 Theory and Hypotheses

2.1 Team Composition

Innovation can be considered as the result of a knowledge and technology skills combination process. Prior research suggests that teams that combine a more varied knowledge and skillsets enjoy more room for this recombination process and more alternative paths to solve problems (Paulus, 2000; Jackson, 1996). Team-level variety can be determined by the combination of a diversified skillsets distributed among the team members or the contribution of some generalist inventors in the team. In this section we introduce two measurements for knowledge and skillset variety of the inventor teams and develop the argument for our main hypotheses.

2.1.1 Skillset Diversity Index

First, we construct an aggregate one minus Herfindahl-Hirschman index to measure how varied the skillsets are composed at the team level.

We measure the skillset variety at the inventor team level by tracking each inventor’s
patenting history. We retrieve the number of different primary technological classifications to which the patents held by team members are assigned (Singh and Fleming, 2010)\(^{23}\). Patents at the USPTO are classified into 416 primary technological classes and we group them into 6 categories following Hall et al. (2001) (HJT category 1-6). We aggregate the number of patents held by each inventor in the team for each technological category and calculate the ratio of each category to the team. We compute a squared sum of each share as a concentration Herfindahl-Hirschman index (HHI) and then the team-level skillset diversity index is calculated as (1-HHI). The skillset diversity index will be high when a patent team consists of inventors that belong to a wide range of technological areas, whereas if most inventors are concentrated in a few areas it will be low (close to zero).

\[
\text{Skillset diversity index}_i = 1 - \sum_{j}^6 s_{ij}^2 = 1 - \sum_{j}^6 \left(\frac{N_{ij}}{N_i}\right)^2
\]

(1)

Where \(s_{ij}\) denotes the share of patents by all inventors in team \(i\) that belong to HJT category \(j\), out of 6 technological categories.

Table 3.1 presents an example of a three-inventor team in our sample and shows the calculation of the measurement.

Diversity has long been considered a source of creativity in a teamwork collaboration process although it may cause some obstacles (e.g., Jackson, 1996). The personal characteristics of the co-inventors and their skillset diversity can determine the creativity. The distribution of skillset diversity among team members is relevant for the quality of innovations generated in teams of inventors. Teams composed of innovators with more

\(^{23}\) Singh and Fleming (2010) used the number of different technological classes in which team members patented in the past to capture knowledge variety. The larger the number of different areas in which at least one team member worked in the past, the greater the team knowledge variety.
diversified sets of skills and pieces of knowledge outperform those where the innovators concentrate in narrower areas.

A more diversified skillset in the team can also help to attenuate some problems that are generated by teamwork. For instance, the free-riding problem may occur in working groups when individual inventors’ efforts cannot be monitored or their contributions to the collective output cannot be measured separately. A more varied set of skills and knowledge enable the inventors to enjoy an advantage in performing each subtask as an independent piece of work. Thus, team-level diversity facilitates the labor division and the exertion of individual efforts and it will play an important role against the free riding behavior. In sum, teams that consist of inventors with more diversified skillsets have some advantages and are expected to perform more efficiently and generate the innovation output more effectively.

**Hypothesis 1a.** The skillset diversity in a team of inventors has a positive effect on the quality of the relevant innovation output.

### 2.1.2 Generalist Inventor Indicator

Next, we provide an alternative way to measure the team variety by identifying the presence of generalist inventors in the team.

Our first step is to compute, for each inventor, a one minus the concentration Herfindahl-Hirschman index (1-HHI) for her skillset diversity across technological fields as defined earlier. Then we identify the presence of generalist inventor in a team if the inventor with the highest diversity index (1-HHI) in this team falls within top ten percent of the distribution.\[^{24}\]

\[^{24}\] We also examine and present the results with a more rigorous threshold of top five percent.
We use the earlier example and show the identification of generalist presence or absence in the same three-inventor team in Table 3.2.

For a given level of skillset diversity distributed across inventors, the presence of some generalist inventors in the team can facilitate the knowledge recombination process and help to attenuate the coordination cost or the internal conflict problems. The potential for both individual and collaborative knowledge recombination will increase with the presence of generalist inventors in the team. Knowledge breadth of generalists is particularly valuable for the recombination of knowledge in contexts where the procedures for solving problems are not clearly established (Melero and Palomeras, 2015).

Previous studies suggest a trade-off between the specialization gains and the coordination costs (e.g., Kim and Marschke, 2015). Team members with different specialized knowledge background often speak different jargons, hampering the gains from diversity. The presence of some generalist inventor can facilitate communication and information sharing within the group. The generalist inventors are more likely to have common areas with their co-inventors and then can increase the expected amount of overlapping expertise among other team members.

Another invention teamwork disadvantage has to do with the conflicts that may arise among the co-inventors. Groups that gather heterogeneous knowledge may have especially high levels of internal conflict if their members have strong feelings about their diverse perspectives (Paulus, 2000). The presence of generalist inventors in a team helps to keep conflict intensity at the moderate level at which it may have a positive effect on performance. For example, a broader scope of knowledge can enable a generalist to understand co-inventors’ critiques (Melero and Palomeras, 2015).

To sum up, the presence of generalist play an important role to moderate the adverse
effects with classical teamwork issues such as coordination costs problems and conflicts.

**Hypothesis 1b.** For a given level of skillset diversity in a team of inventors, the generalist inventors make a more valuable contribution to the innovation output.

### 2.2 Team Persistence

We consider the inventor names listed on patents as inventor teams.\(^{25}\) We follow Kim and Marschke (2015) to measure team persistence as the occurrence of multiple patents featuring the same subsets of inventors in a given time window. If the same subsets of inventors appear on multiple patents over time we say the teams persist.

In this paper we focus on the repetition of the same pairs of inventors. When a patent has more than two inventors, the possible pairs from a patent containing inventors A, B, and C, say, are A-B, A-C, and B-C. A higher team persistence means either that the projects in which teams are working are lasting longer, or that the teams are serially working on more projects.

We use Table 3.3 to illustrate the computation of the inventors’ repeated occurrences as proxy for team persistence.

A higher degree of team persistence leads to a more profound research relationship, which may involve a more efficient exchange of information and a greater amount of knowledge transmission and then result in an innovation output of higher quality. Not only should team persistence have a positive impact on the firm’s innovative production, it should also have a positive impact on the scope of patents. Because of the potential benefits that it implies, we thus hypothesize that inventor team persistence has a positive impact on patent

---

\(^{25}\) See Kim and Marschke (2015) for a discussion on co-inventorship.
quality.

**Hypothesis 2.** A higher frequency of team persistence increases the quality of patents.

3 Data, Variables and Empirical Strategies

3.1 Data

We use US patent data for our empirical analysis. Patents are widely used as a proxy for innovation output of teams of inventors. We retrieve the basic patent information from the USPTO Patent Grant Bibliographic Database. Our dataset contains data on all utility patents granted from 1976 to 2006. We use the NBER Patent Data Project (PDP) citations dataset (Hall et al., 2001) for truncation-adjusted citations received by patents through 2006.26 We also follow the NBER classification developed by Hall et al. (2001) which aggregated the U.S. Patent and Trademark Office (USPTO) technology classes into 6 categories (chemical, computers and communications, drugs and medical, electrical and electronics, mechanical, and miscellaneous “other”) to identify technological areas of patents.

We use the inventor “disambiguation” produced by Li et al. (2014) for inventor information.27 By tracking any given inventor’s patenting history and looking in which technological area they are classified, we are able to identify the skillset background of each inventor who participates in an invention team. For our purposes of analyzing inventor teams and their variety and persistence, we restrict our analysis to patents co-invented by at least

---

26 Hall et al. (2001) address the right-side truncation bias in forward citations by standardizing each patent’s citations with respect to the citation distribution of its primary classification in the corresponding application year.
27 We use the disambiguation produced by the “LOWER” parameterization of their algorithm.
two inventors.

Firstly, we analyze the team composition at the patent or team level. The distributions of number of claims and forward citations of patents are presented as histograms in Figure 3.1 and Figure 3.2. The average number of claims across all patents is 15, the median is 12, and the 99th percentile is 60. The average number of forward citations of patents is 12, the median is 5, and the 99th percentile is 103. The distribution of team size across all teams is presented in Figure 3.3. Figure 3.4 presents the distribution of team size excluding the lone-inventor teams. Excluding the lone-inventor teams, the team size is 3 on average and 9 at the 99th percentile level. The distribution of team-level skillset diversity index is displayed in Figure 3.5. The average index is 0.32, the median is 0.35, and the 99th percentile is 0.76. The distribution of the maximum individual diversity index per team is displayed in Figure 3.6. The average maximum index among all teams is 0.38, the median is 0.44, and the 99th percentile is 0.77.

In our estimation, we restrict data to patents created by teams in which at least one inventor has some previous experience, i.e. at least there is a team member who has at least two patents granted during the study period. Our sample end up with 1,668,504 observations on patents granted from 1976-2006.

Secondly, we analyze the team persistence at the firm level. The data construction procedure is briefly discussed as follows. We calculate the multiple occurrences of each pair of inventors in a team in the subsequent years and sum them up for each patent. The distribution of the total number of repetitive occurrences of any inventor-pairs within each team is displayed in Figure 3.7. After excluding those that collaborate only once, all inventor pairs collaborate repeatedly for 20 times on average, 4 times at median, and 244 at the 99th percentile.
Then we match with the NBER data and link our patent-team level dataset to the Compustat database to obtain each firm’s financial information such as R&D expenditure, employment, etc. Our sample for this part of study contain 1,098 publicly traded firms covering period of 1976-2006. The time series of the fractional number of claims and forward citations averaged on patents by firms in every application year is presented in Figure 3.8 and Figure 3.9. Figure 3.10 shows the average number of patents received by firms by application year. Figure 3.11 shows the average number of repetitive collaborations per team of firms by application year. The decline towards the end of the time window in Figures 3.8 – 3.11 is due to the right-censoring problem.

3.2 Variables

3.2.1 Dependent Variable

We measure the quality of innovation output, the dependent variable for our analysis, with (1) the truncation-adjusted number of citations received (i.e. forward citations) and (2) the number of claims to a patent following previous research.\(^{28}\)

The claims in the patent specification determines the boundaries of the exclusive property rights protected by the patent. Only the technology or aspects covered in the claims can be legally protected and enforced. The patentee has an incentive claim as much as possible in the application but the patent examiner may require that the claims be narrowed before granting. The number of claims determine the technological breadth of a patent as well as the expected market value of the patent. Forward citations, i.e. the citations a patent

\(^{28}\) The truncation problem is defined by Hall et al. (2001) as the time series move closer to the last date in the data set, patent data timed according to the application date will increasingly suffer from missing observations consisting of patents filed in recent years that have not yet been granted. Hall et al. (2001) corrected this truncation bias problem for patent citation by scaling up the observed citation total by dividing it by the fraction of the lifetime citations that are predicted to occur during the lag interval that was actually observed.
receives from subsequent patents, are often seen as an indicator of technological quality. Forward citations over the long term indicate an innovation has contributed to future research. Citations soon after patent application suggests rapid recognition of its importance as well as the presence of others working in a similar area, and thus the expectation of a valuable technological area (Lanjouw and Schankerman, 2004).

3.2.2 Independent Variable

The explanatory variable for composition analysis is the team variety. We have discussed the two measurements we develop for the team variety in Section 2. One measurement is a one minus Herfindahl-Hirschman index (1-HHI) for skillset diversity of the team with a range between 0 and 1, approaching 1 when the skillset is highly diversified. The other is an indicator equal to one if at least one inventor in the team can be considered as a generalist, and zero otherwise. The generalist inventor indicator is equal to one when the inventor with the highest diversity index in a team is at or above the 90th percentile of the distribution of all inventors. This corresponds to a value of (1-HHI) greater than or equal to 0.667.29

The main explanatory variable for persistence analysis is the repetitive collaborations of inventors in each team. We sum up the multiple occurrences in the subsequent years of inventor pairs for each patent and then average the number by patents for each firm.

3.2.3 Control Variables

In the estimations on team composition at the patent-team level, we control for team size in the regression analysis since the number of inventors who constitute the team may be related

---

29 The value is 0.713 when we set the threshold at 95th percentile.
to the complexity of the underlying innovation as well as the amount of resources devoted to it.

The quality of inventors’ previous output may affect the value of their subsequent work and it may reflect the team members’ human capital and underlying ability. It is important to include a control variable that accounts for these effects. Specifically, we control for the average quantity as well as the quality of patents granted to team members. We use the earlier example and show the definitions of average quantity and quality of patents in Table 3.4.

The propensity of patenting and citations varies across different technological fields. For example, the propensity is relatively high in the fields such as chemicals and pharmaceuticals. Therefore, we control for the technological category in which the patent is classified in our analysis.

To estimate on team persistence at the firm level, we control for the deflated R&D expenditure normalized by firm size of employment.

In order to control for the time-varying effects, for instance, technological and economic factors that may change over time, we include a set of time dummies to account for the year in which the patent application is filed.

### 3.3 Empirical Strategies

#### 3.3.1 Patent- or Team-Level Analysis on Team Composition

We use Poisson and negative binomial regressions to test our hypotheses. Given that our dependent variable, either the number of forward citations or the number of claims, is a count variable, we consider Poisson and negative binomial regression as appropriate approaches to test the hypothesized relationships.
Furthermore, as displayed in Table 3.5, the variance of the number of forward citations is nearly 40 times larger than the mean and the variance of the number of claims is over 10 times larger than the mean. The distribution of both forward citations and claims is displaying signs of overdispersion in our data. As a result, we use a negative binomial specification for both regressions.

### 3.3.2 Firm-Level Analysis on Team Persistence

As previously discussed, we use the standard multiplicative distributed specification for our count panel data model. On the other hand, we need to determine if we model the data using a fixed or random effect based on our panel data structure. Since we consider that the difference in innovation output across firms should be controlled we utilize the model with firm-level fixed effect in the analysis.

- **Endogeneity and Instrumental Variable**

In this part of the study, the persistence of inventor teams predicts the innovation output, and the innovation output affects the formation and persistence of teams, which results in an endogeneity problem (Wooldridge, 2012). The reason for the endogeneity problem is that better quality of collaboration performance in the past will cultivate mutual trusts and facilitate communications and coordination between team members. Empirical evidence (Inoue, 2015) predicts that the higher quality of patents in terms of citations a team achieves, the longer the team will work together. To address this potential problem, we use an instrumental variables (IV) approach. The idea behind the IV estimation is that there may be some factors affecting the potentially endogenous variable that have no direct impact on the dependent variable and can be exploited as a source of exogenous variation.
In order to verify the existence of an endogenous variable in the regression, we use a Hausman test to test and verify the endogeneity of the team persistence as our main explanatory variable. The small p-value of Hausman chi-square test indicates that there is a significant difference between the instrumental variable and ordinary least squares coefficients, and IV approach is appropriate (Song and Knaap, 2004; Hill et al., 2008).

To eliminate endogenous impact of the explanatory variable, we use a generalized method of moments (GMM) procedure with IV. Stock et al. (2002), among others (e.g. Stock, 2001; Yogo, 2004 recommend the GMM model as an upgrade from traditional methods (e.g. two-stage least squares) to allow for more efficient estimation when the sample size is large enough (i.e. with more than 700 observations, otherwise GMM estimation over-rejects the null; see Ferson and Foerster, 1994).

Then we need to determine the choice of an exogenous instrumental variable. The IV used must be correlated with the independent variable that it is instrumenting for, and have to be uncorrelated with the primary regression’s error term (e.g. Gould and Gruben, 1996; Acemoglu et al., 2000). In our particular setting, we select the dissolution of firms as the IV. The firm dissolution is directly related with the team persistence but weakly related with its patent output. The rationale for using the firm dissolution year is that, the repeated interactions between co-inventors are constrained by the resources available in their company. The co-inventor relationship will be less likely to be sustained if their organization ceases to operate or a dramatic structural change occurs. An additional requirement IV needs to meet is that it can only affect the dependent variable through their effect on the instrumented factors. This means that firm dissolution can only influence the patented output through their effect on team persistence. In other words, we need to assume that there are not important spillovers on performance from dissolution in a given firm.
Although some authors define firm dissolution as occurring when the name of a firm permanently disappeared from the directories without the incidence of an acquisition or name change (Pennings et al., 1998), many others consider merger or acquisition as a distinct type of dissolution apart from closure or liquidation (Aggarwal and Hsu, 2013, Grilli et al., 2010). The NBER matching data assumes that ownership of the organization changes through mergers, acquisitions, spinoffs, etc. They use data on mergers and acquisitions of public companies reported in the SDC database to track these changes. In this study we consider merger, acquisition, spinoffs, and the complete closure as various types of firm dissolution. Hence for firms that are dissolved, we consider the year of dissolution as the final observation year for the firm in our firm-year panel dataset.30

4 Results

Table 3.5 displays the means and standard deviations of the variables included in the team composition analysis. The average patented innovation in our sample has been developed by a team of 3 members. The overall skillset diversity of all members per team has an average (1-H) index of 0.324. This is the independent variable we use to test the impact of team diversity on patent quality. The individual with the highest dispersion in skillset for each team has an average index of 0.376. This is the variable that we use to build the generalist indicator, the presence of at least one generalist in a team.

Table 3.6 presents the results of the negative binomial regressions with firm fixed-effects in estimating relationship between the team-level diversity and patent quality. Given that we control for the total number of patents filed and claims or citations to all patents in

30 We can also use the variable of “endyr” in the NBER company-patent matching dataset for dissolution year.
the past by the co-inventors, the effect of skillset diversity in a team is necessarily associated
with the contributions of labor divisions. As displayed in Table 3.6, holding inventor ability
(and all other things) constant, the degree of team-level skillset diversity increases the
probability of receiving one more citation by approximately 3% and adding one more claim
by 0.6%.

Table 3.7 presents the results of the fixed-effects negative binomial regressions in
estimating the effect of individual generalists. The table displays the effect of the presence
of generalists for the two alternative criteria: 90th and 95th percentile threshold. Results in
columns (1) and (2) are based on our definition of teams with a generalist as only those teams
whose highest-breadth member is within the top 10% of the distribution of the knowledge-
breadth measure (1-H). Columns (3) and (4) show the results when the threshold is set at
the top 5% and they are qualitatively similar to those in the first two columns. All the
different specifications in Table 3.7 suggest that having at least one generalist in the team
has a positive and significant effect on the quality of patent.

These findings provide support for our Hypotheses 1a and 1b: the skillset dispersion
and presence of generalists in a team have a positive effect on the value of innovation
patenting output. Most of the control variables in Table 3.6 and Table 3.7 show effects in
the expected direction. The number of inventors has a positive and significant effect on the
quality of patent. The value of the past innovation output has a positive effect while the
number of patents has a negative effect.

Turning now to team persistence, the descriptive statistics of the variables for the
analysis are presented in Table 3.8. The fractional number of claims per firm is around 8 on
average and the average fractional number of forward citations is 7. Overall, the
collaboration of any pair of inventors in teams will repeat 5 times in the subsequent years.
Table 3.9 shows the results of estimating the effect of the team persistence on the quality of patents with the firm dissolution as IV as previously described. The results are derived from the fixed-effects generalized method of moments regressions with instruments. The positive estimated coefficient in column (1) shows that the more any two inventors will collaborate repeatedly, the more the patent to which they have contributed is likely to receive more citations, holding the R&D expenditure and firm size constant. According to column (2), the repetitive collaboration increases the probability of having one more claims for the firm by 5.1%. In sum, the findings provide support for our Hypothesis 2: the team persistence has a positive significant effect on the value of patents.

5 Concluding Remarks

In this study, we argue that skillset composition and persistence of inventor teams have important impacts on value of innovation output. In particular, we propose that teams that consist of a given level of skillset diversified inventors or teams with the presence of individual generalists tend to generate a different innovative outcome (in terms of patent quality) than those based on inventors with narrower technology profiles. The main reason behind this result is that a team with more diversified inventors is more efficient at labor divisions and specializations, and a team with some generalist inventor is more effective at knowledge re-combinations and coordination. We also propose that the persistent collaborations of co-inventors in teams has a positive effect on their productivity.

We test these hypotheses using data on patents from teams of inventors. In line with our predictions, the empirical analysis shows that (1) innovations patented by teams with more diversified inventors or with at least one generalist inventor are more valuable and (2) teams with a higher frequency of repetitive collaborations between co-inventors produce
better quality of patenting output.

We view our research as a first step in the understanding of how the composition and persistence of teams of inventors affect one of the main dimensions of their performance outcomes, i.e., the quality of their patenting output. Our study has several implications. First, it pinpoints that not only the skillset diversity in teams that is obtained by gathering specialists who have deeper knowledge but also the inclusion of generalist individuals who have broader knowledge generate more valuable innovations. More importantly, our investigation on team persistence suggests that teams whose co-inventor relationship sustains in the long term are more productive in patenting output.
BIBLIOGRAPHY


values” *Regional Science and Urban Economics* 34 (6), 663-680.


Table 3.1: Example of calculation for team-level skillset diversity measurement
(Patent ID: 6193841)

<table>
<thead>
<tr>
<th>HJT category</th>
<th>Cat. 1</th>
<th>Cat. 2</th>
<th>Cat. 3</th>
<th>Cat. 4</th>
<th>Cat. 5</th>
<th>Cat. 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inventor A</td>
<td>25</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Inventor B</td>
<td>12</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Inventor C</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Subtotal</td>
<td>40</td>
<td>2</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Total</td>
<td>62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Skillset diversity index = \(1 - HHI\) = 0.53
### Table 3.2: Example of identification for generalist inventor in the team (Patent ID: 6193841)

<table>
<thead>
<tr>
<th>HJT Category</th>
<th>Cat. 1</th>
<th>Cat. 2</th>
<th>Cat. 3</th>
<th>Cat. 4</th>
<th>Cat. 5</th>
<th>Cat. 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inventor A</td>
<td>25</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>(1-HHI) = 0.450</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inventor B</td>
<td>12</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>(1-HHI) = 0.609</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inventor C</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>(1-HHI) = 0.500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Maximum (1-HHI) = 0.609 < Threshold (0.667)

Generalist indicator = 0
Table 3.3: Illustration of calculation for team persistence measurement

<table>
<thead>
<tr>
<th>Firm $i$, Year $t$</th>
<th>Patent 1</th>
<th>Inventor team A, B</th>
<th>Sum of repeated occurrences = $n_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patent 2</td>
<td>Inventor team C, D</td>
<td>Sum of repeated occurrences = $n_2$</td>
<td></td>
</tr>
<tr>
<td>Patent 3</td>
<td>Inventor team E, F, G</td>
<td>Sum of repeated occurrences = $n_3$</td>
<td></td>
</tr>
</tbody>
</table>

Team persistence measurement = $\frac{1}{3}(n_1 + n_2 + n_3)$

$^a$ For patents with more than two inventors, we calculate the sum of repeated occurrences for each pair of inventors.
### Table 3.4: Example of definition for selected control variables in team composition analysis (Patent ID: 6193841)

<table>
<thead>
<tr>
<th>Inventor</th>
<th>Number of patents</th>
<th>Number of citations</th>
<th>Number of claims</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inventor A</td>
<td>35</td>
<td>382</td>
<td>1,098</td>
</tr>
<tr>
<td>Inventor B</td>
<td>21</td>
<td>221</td>
<td>813</td>
</tr>
<tr>
<td>Inventor C</td>
<td>6</td>
<td>39</td>
<td>87</td>
</tr>
</tbody>
</table>

Average ability by patent numbers = $\frac{1}{3} \times (35+21+6) = 21$

Average ability by patent citations = $\frac{1}{3} \times (382+221+39) = 214$

Average ability by patent claims = $\frac{1}{3} \times (1098+813+87) = 666$
Table 3.5: Summary statistics for team composition analysis

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of citations</td>
<td>12.839</td>
<td>22.116</td>
<td>0</td>
<td>2023.908</td>
</tr>
<tr>
<td>Number of claims</td>
<td>15.571</td>
<td>13.496</td>
<td>1</td>
<td>887</td>
</tr>
<tr>
<td>Skillset diversity index</td>
<td>0.324</td>
<td>0.237</td>
<td>0</td>
<td>0.832</td>
</tr>
<tr>
<td>Team size&lt;sup&gt;a&lt;/sup&gt;</td>
<td>3.449</td>
<td>1.582</td>
<td>2</td>
<td>51</td>
</tr>
<tr>
<td>Average ability by patent numbers</td>
<td>19.014</td>
<td>35.231</td>
<td>1.04</td>
<td>1009.5</td>
</tr>
<tr>
<td>Average quality by citations</td>
<td>297.409</td>
<td>933.752</td>
<td>0</td>
<td>32833.641</td>
</tr>
<tr>
<td>Average quality by claims</td>
<td>315.099</td>
<td>782.789</td>
<td>1.091</td>
<td>28701</td>
</tr>
<tr>
<td>Highest individual index</td>
<td>0.376</td>
<td>0.243</td>
<td>0</td>
<td>0.826</td>
</tr>
<tr>
<td>Average individual index</td>
<td>0.239</td>
<td>0.191</td>
<td>0</td>
<td>0.803</td>
</tr>
<tr>
<td>Generalist indicator I&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.088</td>
<td>0.283</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Generalist indicator II&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.050</td>
<td>0.218</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

N=1,668,504

<sup>a</sup> Single-inventor teams has been excluded.
<sup>b</sup> Generalist indicator I equals one when the inventor with the highest diversity index falls within the 90th percentile of the distribution.
<sup>c</sup> Generalist indicator II equals one when the inventor with the highest diversity index falls within the 95th percentile of the distribution.
Table 3.6: Estimates of team-level skillset diversity and value of innovation\textsuperscript{a,b}

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number of citations (Neg. Binomial)</th>
<th>Number of claims (Neg. Binomial)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skillset diversity index</td>
<td>0.030\textsuperscript{**}</td>
<td>0.006\textsuperscript{**}</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Team size</td>
<td>0.026\textsuperscript{**}</td>
<td>0.019\textsuperscript{**}</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Average ability by patent numbers</td>
<td>-0.014\textsuperscript{**}</td>
<td>-0.014\textsuperscript{**}</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Average quality by citations</td>
<td>0.001\textsuperscript{**}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td></td>
</tr>
<tr>
<td>Average quality by claims</td>
<td></td>
<td>0.001\textsuperscript{**}</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.000)</td>
</tr>
<tr>
<td>Constant</td>
<td>-34.406</td>
<td>2.529\textsuperscript{**}</td>
</tr>
<tr>
<td></td>
<td>(24212.121)</td>
<td>(0.012)</td>
</tr>
</tbody>
</table>

\textsuperscript{a} $p < 0.1$, \textsuperscript{b} $p < 0.05$, \textsuperscript{c} $p < 0.01$. Heteroskedastic-consistent standard errors are in parentheses.

\textsuperscript{b} All regressions include year dummies and technological category variables.
Table 3.7: Estimates of the presence of generalist inventors in teams and value of innovation\textsuperscript{a, b}

<table>
<thead>
<tr>
<th>Variable</th>
<th>90th percentile \textsuperscript{c}</th>
<th>95th percentile \textsuperscript{d}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generalist indicator</td>
<td>0.039** (0.004)</td>
<td>0.021** (0.002)</td>
</tr>
<tr>
<td>Team size</td>
<td>0.025** (0.001)</td>
<td>0.018** (0.000)</td>
</tr>
<tr>
<td>Average ability by patent numbers</td>
<td>-0.014** (0.000)</td>
<td>-0.014** (0.000)</td>
</tr>
<tr>
<td>Average quality by citations</td>
<td>0.001** (0.000)</td>
<td>0.001** (0.000)</td>
</tr>
<tr>
<td>Average quality by claims</td>
<td>0.001** (0.000)</td>
<td>0.001** (0.000)</td>
</tr>
<tr>
<td>Constant</td>
<td>-24.504 (6131.134)</td>
<td>2.466** (10081.172)</td>
</tr>
</tbody>
</table>

\textsuperscript{a} \textsuperscript{†} \textit{p} < 0.1, \textsuperscript{b} \textit{p} < 0.05, \textsuperscript{**} \textit{p} < 0.01. Heteroskedastic-consistent standard errors are in parentheses.
\textsuperscript{b} All regressions include year dummies and technological category variables.
\textsuperscript{c} Generalist indicator equals one when the inventor with the highest diversity index falls within the 90th percentile of the distribution.
\textsuperscript{d} Generalist indicator equals one when the inventor with the highest diversity index falls within the 95th percentile of the distribution.
Table 3.8: Summary statistics for team persistence analysis

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>S.D.</th>
<th>Between S.D.</th>
<th>Within S.D.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patents</td>
<td>23.547</td>
<td>109.17</td>
<td>65.099</td>
<td>70.041</td>
<td>0</td>
<td>3508</td>
</tr>
<tr>
<td>Claims per patent</td>
<td>8.424</td>
<td>10.744</td>
<td>6.713</td>
<td>8.431</td>
<td>0</td>
<td>195</td>
</tr>
<tr>
<td>Cites per patent</td>
<td>7.444</td>
<td>13.479</td>
<td>7.835</td>
<td>11.009</td>
<td>0</td>
<td>275.454</td>
</tr>
<tr>
<td>Repeated occurrences</td>
<td>5.102</td>
<td>27.047</td>
<td>14.42</td>
<td>22.306</td>
<td>0</td>
<td>1256.259</td>
</tr>
<tr>
<td>R&amp;D ($M) a</td>
<td>156.971</td>
<td>598.425</td>
<td>417.499</td>
<td>269.626</td>
<td>0.012</td>
<td>10433.89</td>
</tr>
<tr>
<td>Employment (1000s)</td>
<td>19.586</td>
<td>58.422</td>
<td>43.572</td>
<td>15.62</td>
<td>0</td>
<td>876.800</td>
</tr>
</tbody>
</table>

N=15,816, n=1,098

a R&D expenditures are inflation-adjusted in 1992 dollars.
Table 3.9: Estimates of team persistence and value of innovation$^a,b$

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number of citations (Poisson)</th>
<th>Number of claims (Poisson)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln(Repeated occurences)</td>
<td>0.059$^*$</td>
<td>0.051$^*$</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>ln(R&amp;D)</td>
<td>-0.049</td>
<td>0.237$^*$</td>
</tr>
<tr>
<td></td>
<td>(0.056)</td>
<td>(0.033)</td>
</tr>
<tr>
<td>ln(firm size)</td>
<td>-0.890$^{**}$</td>
<td>0.436$^{**}$</td>
</tr>
<tr>
<td></td>
<td>(0.127)</td>
<td>(0.075)</td>
</tr>
</tbody>
</table>

$^a$ $^p < 0.1$, $^* p < 0.05$, $^{**} p < 0.01$. Heteroskedastic-consistent standard errors are in parentheses.  
$^b$ Year dummies are included.
Figure 3.1: Distribution of number of claims across patents

![Distribution of number of claims across patents](image1)

*note: data truncated at 99th percentile*

Figure 3.2: Distribution of number of forward citations across patents

![Distribution of number of forward citations across patents](image2)

*note: data truncated at 99th percentile*
Figure 3.3: Distribution of size of inventor teams

Figure 3.4: Distribution of size of co- or multi-inventor teams

note: only for teams that consist of at least two inventors
Figure 3.5: Distribution of aggregate team-level skillset diversity index across patents

Figure 3.6: Distribution of maximum individual-level skillset diversity index across patents
Figure 3.7: Distribution of repetitive collaborations of any inventor-pair within teams

Note: (a) only for teams that collaborate at least twice; (b) data truncated at 99th percentile
Figure 3.8: Average number of claims per patent received by firms by application year

Figure 3.9: Average number of forward citations per patent received by firms by application year
Figure 3.10: Average number of patents received by firms by application year

Figure 3.11: Average number of repetitive collaborations per invention team of firms by application year