An Anatomy of the Patent Quality: China vs. US*

Hanming Fang† Zheng (Michael) Song ‡ Hanyi Tao§
Yiran Zhang ¶

February 27, 2021

Abstract

We construct various quality measures to investigate the evolution of the quality of granted Chinese patents, relative to that of the US patents, over the period of 1985 and 2020. Three novel measurements of patent quality for cross-country comparison are developed in this paper: the forward citations standardized by dual-listed patents, the technology gap based on new words, and the cross-country knowledge transfer. All the measures gives consistent results. Standardized forward citations indicates that the US patents' average quality is twice as good as the Chinese. However, the Chinese is catching up. The new words measures suggest that during 1996-2000, 16.72% of new words that appeared in the Chinese patents were words that already appeared in the US; this fraction continuously dropped, and was 7.34% for the new words that appeared in Chinese patents filed in 2011-2015. In terms of cross-country knowledge transfer as proxied by patent citations and new word citations, we find that the knowledge flowed mostly from the US patents to Chinese patents, though in some technology classes, particularly the electrical engineering industry, knowledge started to flow from China to the US from around 2001. We also investigate other measurements, such as patent renewal, originality and generality, and centrality.

JEL No.: O1, O2, O3, O4

Keywords: Patent Quality, Standardized Forward Citation, New Words, Knowledge Transfer

---

*Preliminary and incomplete. All comments are welcome. We thank Kenneth Huang Branstetter Lee and Manuel Trajtenberg for helpful comments.

†Department of Economics, University of Pennsylvania, Perelman Center for Political Science and Economics, 133 S. 36th Street, Philadelphia, PA 19104; ShanghaiTech University, and the NBER. Email: hanming.fang@econ.upenn.edu.

‡Department of Economics, The Chinese University of Hong Kong, Email: zsong@cuhk.edu.hk.

§School of Entrepreneurship and Management, Shanghaitech University, Shanghai, China. Email: taohy@shanghaitech.edu.cn.

¶Department of Economics, The Chinese University of Hong KongEmail: zhangyiran@link.cuhk.edu.hk.

1 Introduction

China has experienced a surge in patent filings in the past twenty years. According to WIPO (2019), the patent applications filed in China increases from 51,906 in year 2000 to 1.54 million in year 2018, and the share of the Chinese patent filings in the world total increased from 3.77% in year 2000 to 46.36% in year 2018. However, the concerns about the quality of Chinese patents always lingered, and the convention wisdom is that Chinese surged in patent *quantity*, but not in patent *quality* (Liang, 2011; Prud'homme and Zhang, 2017).

In this paper, we provide an anatomy of the patent quality in China from 1985 to 2020. By the traditional count-based patent quality measures, forward citation number can indicate the patent quality within a Patent Office. However, citation norms vary across countries. Most of US and Chinese patents registered in different Patent Office gives rise to the problem that traditional measurement based on forward citations cannot be directly comparable across countries. In this paper, we mainly implement anatomy via two ways. The first is, as forward citations within a patent office cannot be compared directly with one another, we use dual-listed patents registered in both countries' patent offices as a bridge and develop a quality measurement called *standardized forward citation numbers*. Given that both countries' non-exported patents can be compared with these bridge patents, which contain the same technology package across countries, our standardized forward citation numbers are comparable across countries. Besides, we introduce the *new words* measure of patent quality, particularly its novelty. The new words method directly reveals who learns from whom, thereby avoiding the aforementioned problem in traditional citation-based measurements. We identify the new words in the abstract of a patent, which have never appeared in patents before, separately for the Chinese and US patents. Based on the yearly new word list, we construct the technology gap between China and US. For the new words appeared in the Chinese patents in each year, we calculate the fraction of words that comes from US existing knowledge, and we call it retrospective technology gap. For the new words in the US patents, we also calculate the share of words that comes from Chinese existing knowledge and we call it prospective technology gap. We use the technology gap to capture the extensive margin of novelty of the Chinese patents. Moreover, we construct the cross-country knowledge flow proxied by the patent forward citations and new word citations. We calculate the cross-country citation share and use it as a measure of intensive margin of the quality of the patents. We also use the measurement based on Herfindahl-Hirschman Index (HHI) to calculate the *originality* and *generality* of each patent. We extend the HHI measurement and calculate the authorities centrality, hubs centrality and betweenness centrality of patents based on the full-blown citation network. The centrality measurements show the importance of the patent in the patent citation network and we treat them as quality indicators. Moreover, we investigate the life of a patent and use patent renewal as a measure of patent quality.

With those measurements in hand, we investigate the quality change over time. Since the patent data has both left- and right-censored problem, we use the patents in the US as the benchmark and calculate the relative quality of Chinese patents over time. Moreover, we use the time of new words' first appearance as the indicator of technology frontier in China and US separately and show the technology gap between China and US. We find that the relative quality of Chinese patents to US ones is increasing over time. Meanwhile, the technology gap between China and US is narrowing. We find that the knowledge flows mostly from the US patents to Chinese patents, though in some technology classes, particularly electric communication technique (H04) and basic electric elements (H01), knowledge starts to flow from China to the US from around 2001, though in 2017, the knowledge transfer from China to US accounts for only 1.1% of the total China-US knowledge transfers.

Related Literature. Our paper builds on a large literature on measuring the quality of patents. There are two kinds of quality measurement in recent literature. One is patents' economic value which are measured by patents' price or stock price increment induced by new invention. The other is patents' scientific value which is usually constructed by patent citation database. In this paper, we concentrate on the latter.

Kogan et al. (2017) use the stock market price variation to identify the patent value in their QJE paper [1], and furtherly they take the stock return induced by the patent issue as the economic value of patent, and check its' relationship with patent scientific value (citation-based), TFP or output growth. Their results show a positive correlation between patents' economic value and scientific value. However, Abrams, Akcigit and Grennan (2019) document the inverse-U relationship between patent licensing revenue and patent citation number, with fewer citations at the high end of value than in the middle. This paper explains this pattern by dividing the innovation into two types: one is productive innovation and the other is strategic innovation. Productive innovation is to extend knowledge, which leads to the traditional increasing relationship between patent value and citations, and strategic innovation is made by incumbents to protect market

[1] The idea of this paper is that the stock returns in the date with patent issued contains both of the information of the return caused by patent (v) and the return unrelated with the patent (ϵ), while the other date without patent issued merely contains the information of ϵ.

share and discourage follow-on innovations , which leads to the negative relationship between patent value and citations.

The forward citation counts is a long lasting measurement on patent quality, that a patent receiving more forward citations is expected to have higher value (Jaffe and Trajtenberg, 1999; Hall, Jaffe and Trajtenberg, 2005). In the seminal paper Hall, Jaffe and Trajtenberg (2001), originality and generality are proposed to measure the novelty and influence of patents. In this paper, the scaled measure is adopted to deal with the left- and right-censored problem of patent data. Almost all the researchers follow the methods in Hall, Jaffe and Trajtenberg (2001) to calculate the quality of patents(Hagedoorn and Cloodt, 2003; Hu and Jefferson, 2009; Lerner and Seru, 2017; Hu, Zhang and Zhao, 2017; Lin, Wu and Wu, 2020). The forward citation counts, originality and generality measure the patent qualify through the one-layer patent citation network. Liang et al. (2019) extends the one-layer network into a multi-layer patent citation network. They construct the patent citation network in China and use centrality measurements to show the value of patents in the patent citation network. In this paper, we construct the cross-country and cross-technology class patent citation network and quantify the knowledge transfer between China and US.

The use of new word as a measure of patent novelty follows Akcigit, Kerr and Nicholas (2013) and Kelly et al. (2020). Akcigit, Kerr and Nicholas (2013) identify patent with novel technology based on whether the patent helped stimulate a new research field. Kelly et al. (2020) identify the breakthrough innovations based on whether the text of the patent is distinct from prior patents but substantially similar to later patents. In this paper, we identify the new words in the Chinese and US patents, and use these measures to quantify the technology gap between the two countries.

The idea of using patent renewal as a measure of patent quality, particularly the perceived commercial value of the patent, first appeared in the seminal paper of Pakes (1986), and subsequently followed in numerous papers in the literature, including Lanjouw, Pakes and Putnam (1998) and Deng (2007).

The remainder of the paper is structured as follows. In Section 2, we provide a brief introduction to China's patent system and contrast that of the US system; in Section 3, we describe the patent data sets used in our analysis; in Section 4, we present the dynamics of various patent quality measures for US and Chinese patents; and finally, in Section 5, we conclude.

2 Patent System in China

In this section, we provide some institutional background of the Chinese patent system.

2.1 Chinese Patent Laws

The patent system in China was established in 1985 when the first Patent Law of the People's Republic of China came in force in April 1985. According to the law, patents are classified into three types: the invention patents, the utility model patents and the design patents. In particular, for each invention patent to be granted, it needs to satisfy three conditions *novelty, inventiveness* and *practical applicability*.[2]

The China National Intellectual Property Administration(CNIPA henceforth), also known as SIPO, issued a detailed guidance about the implementation of Patent Law, and the CNIPA also published several announcements of modification on the regulations related to patents. China's Patent Law has since been amended three times, respectively in September 1992, August 2000 and December 2008, and new versions of guidance were published following each amendment of Patent Law. See Figure 1 for a summary of the specific dates of all the Patent Law amendments and the publication of new guidance, as well as the dates of the CNIPA announcements about the modifications of the patent regulations such as updated patent fees, etc.

[Figure 1 About Here]

The main change in the first amendment of Patent Law in September 1992 is to extend the term of (invention) patent (see below for details); the main change in second amendment in August 2000 is that the CNIPA is required to facilitate the patent examination process, and that the provincial governments are authorized to manage the local patent related issues; the main change in the third amendment in December 2008 is the modification of the definition of novelty from *relative* novelty to *absolute* novelty.[3]

2.2 Patent Examination

The Chinese patent application consists of three components: the first is the application form, which includes the name of patent, as well as the basic information about applicants

[2]See Article No. 22 of the 1985 Patent Law: *Any invention or utility model for which patent right may be granted must possess novelty, inventiveness and practical applicability.*

[3]This modification requires that the patent be novel all around the world, not just in China, which strengthens the patent granting criteria.

and inventors; the second is the illustration document, which describes the patent either in words or in graphical illustrations; and the third is the documents on rights (or priorities) to claim.

The patent examination process consists of two steps: the preliminary examination and the substantive examination. After a patent application is flied, the CNIPA conducts a preliminary examination of the forms and documents. If the application passes the preliminary examination, it will be open to the public within 18 months from the application date.

Whether a patent can be granted depends on the outcomes from the substantive examination. The substantive examination focuses on the priorities claimed in the application, and is conducted within three years from the application date. If the CNIPA disagrees with the applicants in some aspects, it can ask the applicant to provide more information or to make a revision. If the application still can not pass the substantive examination after a revision, the patent application is rejected.

If the CNIPA finds no reason to reject the application, the patent is then granted. Figure 2 depicts the average examination time of patent filings in China conditional on being granted, by patent filing year.[4]

[Figure 2 About Here]

2.3 Term of Patent

The patent system encourages innovation because the patent grants its inventors protection for a limited period of time, referred to as the *term of patent.* The 1985 Patent Law provided invention patents 15 years of protection from the date of application, and 5 years of protection for utility model and design patents. The 1992 amendment of the Patent Law extended the term of the invention patents to 20 years, and that of the utility model and design patents to 10 years.

2.4 Costs of Patent

Although a patent gives the patent holders exclusive rights to produce and sell their ideas for a certain period of time, the patent application and maintenance do incur some costs. There are four types of fees: application fee, examination fee, claim fee, and maintenance fee.

[4]Note that the examination time in Figure 2 for recent years is subject to the right-censored problem.

When the applicant submits a patent to the CNIPA, she needs to pay the application fee and examination fee. If the patent is not granted within two years, the applicant has to pay the application maintenance fee to keep the application valid. In addition, the applicant has to pay a claim fee for each of the claims asserted in the patent application.[5] Moreover, if the applicant claims more than ten priorities, she need to pay additional claim fee for each of the additional claim. If the patent is granted by the CNIPA, the applicant becomes a patent holder, and needs to pay the maintenance fee. The maintenance fee is paid on *yearly* basis. Patent holders pay maintenance fee to protect their patents from expiring up to the maximum of 20 years from the filing date. It is worth noting that the maintenance fee increases as the patent life goes up.

There were several changes in the patent fees. The first document about the cost of applying and holding a patent is the announcement made in April 1985 and it only mentioned that the cost is specified by the CNIPA. The second announcement which is very short-lived is made on October 1992. Since the patent law was amended and the came into force from January 1993, this announcement was soon replace by the No. 36. Currently, the patent fee was specified by the announcement No. 75 issued in 2001. Table 1 lists the main modifications about the cost of applying and holding a patent across years.

[Table 1 About Here]

2.5 Patent System Differences between China and US

We compare the patent system in China and US from the following three aspects: the examination time, the term of patent, and the schedule of maintenance fee.

Differences in Patent Examination Time. The patent examination time in China is on average three years. Figure 2 shows the examination time in China and US from 1985 to 2019. From 1995 to 2010, we observe a great improvement in examination time in China. Before 2004, the patent examination time in China is longer than that in US. However, after 2004, the patent examination time in China becomes shorter than that in US and continues to go down. Since the patent data is right-censored, the patent examination time in later years is downward biased.[6]

[5] In a patent filing, the claims define the extent of the protection conferred by the patent. The purpose of the claims is to define which subject-matter is protected by the patent.

[6] Since the patent data in China is right-censored on June 2020, we do the same censoring on the US data to make the examination time comparable.

Differences in Term of Patent. As mentioned above, the term of patent in China is 15 years from application date in the first patent law and it is extended to 20 years in the amendment in 1992. In US, the calculation method of term of patent was adjusted on June 8th 1995. If a patent is filed or issued after June 8th 1995, the term of patent is the greater of the 20-year term from application date, or 17 years from grant. If a patent is granted before the modification, it has a 17-year term from grant. In addition, according to the United States Patent and Trademark Office (USPTO henceforth), a patent's term may be shortened if it duplicates the claims of another patent which expires sooner.

Differences in Patent Maintenance Fee. In both countries, applying and holding a patent requires application fee, examination fee and maintenance fee. Table 2 shows the comparison of application, examination and maintenance costs of the patents in China and US. The price are indicated in RMB and USD for each country and the dollar value of cost in China are indicated in brackets. In China, the maintenance fee is paid on a yearly basis until the expiration of the patent. In the US, the maintenance fee is paid on the third, seventh and eleventh year from the patent grant date. For example, once the payment is fulfilled in the third year, the patent will enjoy the protection until the 8th year from the grant date. It is worth noting that the maintenance fee payment should be made between 3 to 3.5 years, 7 to 7.5 years and 11 to 11.5 years from the patent grant date. If the patent holder does not pay in time, there would be a surcharge within 6 month from due. If the patent holder still does not pay the maintenance fee, the patent will expire 6 month after the due date.[7]

[Table 2 About Here]

3 Data

The Chinese patent data and US patent data are open to the public, published by CNIPA and the USPTO respectively. In order to properly compare the quality between the Chinese and US patents, we mainly collect the following three pieces of information from the patent data.

Patent Registration Information. We first collect the patent registration data (PRD henceforth) from the CNIPA and USPTO. The Chinese patent registration data covers

[7]Manual of Patent Examining Procedure (MPEP), Ninth Edition, Chapter 2500.

all published patent applications and grants from 1985, when the CNIPA started to accept patent applications, onward until June 2020. It contains detailed information of three categories of patents: the invention patents, the utility model patents, and the design patents. In this paper, we will focus on the category of *invention patents*, which is comparable to the utility patents in the US. Correspondingly, in U.S. patent data coming from the USPTO, we only focus on *utility patents* and delete the design and plant patents [8]. Figure 3 shows the number of fillings of three types of patents in China and U.S. between 1985 and 2019. There is a clear trend that the number of filings is increasing over time.

[Figure 3 About Here]

Except for application numbers, and application and granted date, patent registration data in both countries include other basic information of patents, such as grant number, publication number, technology class, assignee's nationality and etc..

For the **patent technology class classification**, we mainly International Patent Classification (IPC henceforth) in this paper, which is provided by the World Intellectual Property Office (WIPO henceforth). It has eight main classes as shown in Table 3. In both of two countries' database, the historical IPC records were provided. To guarantee the consistency in the patent technology classification, as we mainly use 3-digit level (class level) IPC code in this paper, we have compare different versions of IPC from version 1 to 8 in class level, there exists no big changes across years.

[Table 3 About Here]

Figure 4(a) shows the composition of granted patent filings in US according to IPC. The share of Physics and Electricity is gradually increasing since 1986, while the share of Performing operations & Transportation and Chemistry is decreasing. The compositional change of the Chinese patents shows a similar pattern but it records a more drastic increase in the share of Physics and Electricity, indicated by Figure 4(b). One distinct feature in the compositional difference is that Chemistry plays an important role in China, while Physics and Electricity accounts for considerable share in US patents.

[Figure 4 About Here]

[8]The patent data from CINPA is right-censored on June, 2020. In order to make the US and Chinese data comparable, we also right censor the US data on this date. Given the examination time, we only analysis the patents applied in and before 2019, conditional on they are being granted by the CINPA or USPTO.

The another important question is that how we can define the Chinese or US domestic patents. In every country's registration database, we use patent's assignee's nationality to define the **patent's nationality** [9]. The detailed clean steps are in Appendix B.1.

Patent Citation Information. We also collect information on the patent citation data (PCD henceforth) from the CNIPA and USPTO. Generally, we usually use the forward citation number the patent received to measure patent's quality. Following Hall, Jaffe and Trajtenberg (2001), we delete the self-citation in this paper. One thing to be noted is that as dual-listed patent takes an important role in this paper, the original forward citation number may be biased if we only calculate the patent citation based on its' application number. Developed countries' patent may be more highly valued, so when making citations, the patent document started with US or WIPO are more popular. If this happens, one problem in our study is that when a Chinese domestic patent cites one U.S. multinational patent exported to China, the USPTO publication number of that patent will be cited, not its' publication number in CNIPA [10]. For this problem, when we calculate the forward citation of the U.S. multinational patent exported to China in CNIPA citation database, we have merged cited publication number in USPTO with its' publication number in CNIPA, if it's a foreign patent publication document and has patent family in CNIPA [11]. In USPTO PCD, we do the same work in case that the cited patent number is this patent's Chinese registered patent publication code, not its' publication code in USPTO.

Patent Family Information. The last piece of information is patent family data (PFD henceforth). By OECD patent database manual's definition(Zuniga et al., 2009), patent family is defined as the set of patents (or applications) filed in several countries which are related to each other by one or several common priority filings, which means that the patents belonging to one family indicating that they include the same or most similar techniques and filed by same assignee. There are different types of patent family, in CNIPA, every patent's simple family member, extended family member and INPADOC

[9]In USPTO, both of inventor's and assignee's information are relatively rich, while CNIPA is lack of inventor's information. So to guarantee the consistency, we use the assignee's nationality to define the patent's nationality.

[10]For example, for the same patent, its' publication number in CNIPA is CHXXXXXX1 with its' USPTO publication number is USXXXXXX2. When a Chinese domestic patent make citation of this patent, USXXXXXX2 possesses higher possibility to be written.

[11](1)Initially, the number of citations between CNIPA registered patents is 2112053 pairs. After merging, 168025 pairs are added; (2) As the lack of patent family information for USPTO patents, we cannot do the same work for USPTO citation data. So Chinese patents' quality may be underestimated if we take Chinese multinational patents exported to USA as a bridge.

extended family information are recorded. Martinez (2010) has introduced the definitions, identification of each kinds of patent family elaborately.

In PFD, to identify Chinese patents exported to US, first, in Chinese PFD, we select the patent of which the assignee's nationality is China, and it has family patent member in US; Second, in USPTO PFD, we select the patent of which the assignee's nationality is China, and it has family patent member in China. For those selected patents by these two steps, we know their patent number in CNIPA and USPTO, and they are defined as Chinese multinational patents exported to US.[12] For U.S. patents exported to China, same identification method is taken. The following table exhibits the number of multinational patents in each countries.

[Table 5 About Here]

CNIPA and USPTO report the quantity of the granted patents coming from other countries every year. As shown in Figure 5, we compare the multinational patents identified in the data with the official report.

[Figure 5 About Here]

4 Measurement of Patent Quality: China vs. US

In this section, we use various methods to measure patent quality. First we adopt *counts* based measurement to assess the patent quality, which is widely used in the literature (Hall, Jaffe and Trajtenberg, 2001; Lerner and Seru, 2017; Abrams, Akcigit and Grennan, 2019). As the citation norms various across different countries, the traditional forward citation number measurement may not be comparable across countries. By dual-listed patents, we develop a standardized forward citation measurement to overcome this difficulties. Second, we introduce the *new word* based measurement to gauge the technology gap between China and US. Third, we construct the cross-industry and cross-country knowledge flow to measure the patent quality.

We also study the patent quality with the measurements in the liratature and we put the results in the appendix. We use the Herfindahl-Hirschman Index (HHI) of immediate

[12]Actually, the second step helps us find more Chinese exported patents than the first step, because most of the patent family information are recorded at the time the patent are filed or granted, but the patent are usually registered in the home country earlier than foreigner country. So the home country's patent family database usually doesn't record the patent's family information in foreigner country.

backward and forward citations to create generality and originality measures of patents (Hall, Jaffe and Trajtenberg, 2001). Moreover, we follow Liang et al. (2019) and extend the HHI measure to create network based centrality measures of patents. Finally, we estimate the life of patents and use patent renewal to gauge the patent quality.

4.1 Standardized Forward Citation Number as a Measure of Patent Quality

As citation norms varies across countries. The more forward citation received by US patent may only reflect the compulsory citation regulation in US, while in China, the citation is voluntary. What's more, the increasing trend of average patent's forward citation in China may only reflect some structural change of Chinese patent registrations, saying, the increasing number of granted patents of Chinese domestic patents. The more new patents applied, the higher demand for patent citations. Accordingly, if we only calculate the patent's forward citation number within home country's patent office, this kind of forward citation cannot be comparable across countries. We must find one kind of patent of which the forward citation number can be compared with patent registered in each countries as a bridge. Here, we employ the patents that registered in more than one countries.

Dual-listed Patents. In order to facilitate the analysis, we label the patents with two different names. The first is the *domestic patent*, which is invented by home country and also registered in home country's Intellectual Property Right Office (Hereafter IPO). Among the domestic patents, some patents are also registered in other countries' IPO, which we call as *multinational patent*, aka, the *dual-listed patent*. In this paper, there are two types of multinational patent. One is *US multinational patents exported to CN*, and the other is *CN multinational patents exported to US*.

Figure 6 exhibits the numbers of multinational patents we have identified in the USPTO and CNIPA data, grouped by patent technology class. For US patents exported to China, in terms of time trend, more and more patents exported to China especially after 2000. In the perspective of absolute number of exported patents in every technology class, the number of exported electricity patents is more than other technology classes' patents. However, if we calculate the share of patents exported to China in every technology class, we find that in 2010, about 20% of US chemistry patents were exported to China, and this share ranks first among all the technology classes. For Chinese patents exported to US.

The absolute number of every technology class's patents is increasing over time, especially for electrical engineering patents. The number of electrical patents ranks first among all the technology classes as well.

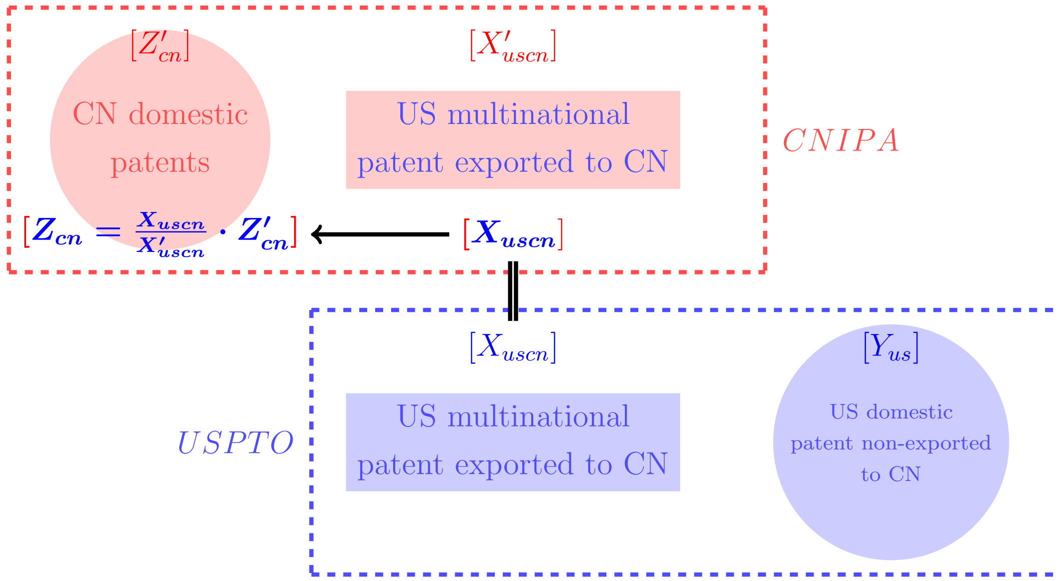
[Figure 6 About Here]

It takes time for patents to export to other countries. We define the year the patent applied at exported country less the year the patent applied at home country as the application lag. Figure 7 demonstrates the application lag of US exported patents and Chinese exported patents of different cohorts. For large proportion of patents in every cohort, it takes about 1 year to registered at foreigner country after applying in home country.

[Figure 7 About Here]

Logic behind Standardization. Multinational patent can be considered as a bridge for comparison. Either the US multinational patents exported to CN or the CN multinational patents exported to US can be taken as the bridge. For example, as shown in logic graph below, by the CNIPA citation data and USPTO citation database, we calculate the forward citation value for Chinese domestic patents, US multinational patents exported to CN and US domestic patents non-exported to CN in red and blue rectangle respectively. And as US multinational patents exported to CN simultaneously registered in CNIPA and USPTO [13], they have two forward citation values in each country's IPO, X_{uscn} in USPTO and X'_{uscn} in CNIPA. If assume the techniques inside the patent document are same across different IPOs, the simple transformation for the quality value of Chinese domestic patents is $Z_{cn} = \frac{X_{uscn}}{X'_{uscn}} \cdot Z'_{cn}$, which can be compared with Y_{us} and X_{uscn} Chinese multinational patents exported to US can be this bridge by the same logic as well. The detailed calculation steps of standardized forward citation numbers are shown in Appendix B.4.

[13] Most patents will registered at their home country, as in CNIPA data, there are 189226 patents' assignee are US, within which, only 2.6% patents registered in CNIPA but not in USPTO, their home country's IPO.



Results. Figure 8 shows the comparison of the Chinese domestic patents and the US domestic patents based on the standardized forward citations. First, no matter which multinational patents we take as bridge, it is obvious that Chinese patents quality approaches US patents quality in recent years, especially for electrical engineering patents. Second, there exists a difference between the results calculated by these two bridges that if we take US patents exported to China as bridge, the Chinese domestic patents' quality is even higher than US domestic patent after 2008. The decomposition of gap calculation results shown in Figure A5 can provide some reasons for this phenomenon. When we take US patents exported to China as a bridge, we can find that US exports the patents of which the quality is lower compared with non-exported patents' quality. This gap remains stable across years since 1995 to 2010. However, compared to Chinese domestic patents, those US exported patents received less and less forward citations in CNIPA in recent years.

[Figure 8 About Here]

4.2 New Words as a Measure of Patent Quality

In this subsection, we utilize the application text of each patent and the new words contained in the patent filing as a proxy for patent quality. The idea is very simple: if a patent i possesses some new words that have never appeared in the past, we expect it to have higher quality, since there is something *new* in the patent. We analyze the *abstract* of each application and identify the new words. The new word is defined as the word and phrase which has never appeared in the abstract of previously filed patents.

New Word Counts. We construct the new words for China and US patents *separately*. Since the new word lists are constructed for China and US separately, a new world may appear as a new word both in China's list and US lists. The upper panel in Figure 9 plots the number of new words in Chinese and US granted patents separately, sorted by application year. The number of new words in US patents peaked around 2000, which is consistent with the findings in Balsmeier et al. (2018). For the number of new words of Chinese patents, we observe a growing trend since 1985 and this trend peaked around 2013. From 2013, the number of new words stopped increasing. [14]

In order to exclude the potential typo and man-made word just for being different from the previous patents, we define the *wasted word* as following: If a word '*xyz*' appeared less than three times in the word list, then the word '*xyz*' is said to be a *wasted word*. [15] The lower panel in Figure 9 shows the number of new words without the wasted words. Only 20% of the new words are left for both the Chinese and US patents. The number of the US patents peaked around 1998, which is similar as before. For the number of the Chinese patents, it peaked around 2008. A possible explanation is that after 2008, the patent stimulate programs in China produced lots of patents with new concepts but not influential at all. In the following analysis, we will rely on the word list without these wasted words.

[Figure 9 About Here]

Prospective and Retrospective Technology Gaps. Figure 9 may suggest that the Chinese patents somehow have better quality than US patents after around 2007. There are two possibilities to consider, however. First, the new words that first appeared in China were picked up later in US patents; second, the new words that first appeared in China stayed in Chinese patents only. These two possibilities have different implications regarding which country is leading the technological race. To distinguish these two possibilities, we combine separate new word list and identify the new word share by both countries. Now, we can ask whether the new words reported in Figure 9 would appear later in the other

[14] The detailed text data processing method is documented in Appendix B.2.

[15] The definition of the wasted word implies that all the new words in the last two years of the sample, 2018 and 2019, are will not shown up at all. We also present some robustness checks in Appendix B.3 to confirm the validity of the results.

country's patent filings. Figure 10 shows that indeed, most of new words in the separately constructed China and US lists are words that only appeared in one country. However, a larger share of the new words in the separately constructed China lists actually appeared in the US list first.

[Figure 10 About Here]

Besides the new word counts, we can infer the *technology gap* between China and US based on the time lag of occurrence of the same new words. In order to deal with the left- and right-censored problem of patent data and citation problem, we use the Kaplan-Meier style method to infer the technology gap between China and US. To simplify the graphical representation of or results, we divide the patents into 5 groups according to their application years.

In the first scenario, we construct the average time lag by sorting the new words according to their first appearance in the US patents. Under this definition, the time lag τ means: when the new knowledge appeared in US, the Chinese scientist takes τ years to absorb this idea. The technology gap in each year is defined as following, for the US patents applied in year s, $r_s(\tau)$ percent of the new words comes from China τ years ago.

$$p_s(\tau) = \int_{-\infty}^{\tau} \frac{\text{Number of Words from CN}_{st}}{\text{Number of New Words in US}_s} dt \tag{1}$$

Figure 11(a) plots the time gap between words' first appearance in China and US. A positive number indicates the absorbing time taken by Chinese inventors, and we call it the *prospective* technology gap.

The lines in the figure can be interpreted as the following: For the new words appeared in the US patents between 1996 and 2000, 0.06% of them had a 10-year or above time lag with the Chinese patents, and 0.12% of them had a 5-year or above time lag with the Chinese patents. For the new words appeared in the US patents between 2011 and 2016, 0.2% of them had a 10-year or above time lag with the Chinese patents, and 0.8% of them had a 5-year or above time lag with the Chinese patents. The interaction of those curves and $\tau = 0$ tells that, the share of new words in the US patents that origin from the Chinese patents was increasing since 1990. From this perspective, the technology gap between China and US was shortening over years.

In the second scenario, we show the technology gap between China and US from the perspective of the Chinese. The time gap t calculated with this method means: when the new knowledge appeared in China, the Chinese scientist were using the technology that was invented in US t years ago. The technology gap in each year is defined as following, for the Chinese patents applied in year s, $r_s(\tau)$ percent of the new words comes from US τ years ago.

$$r_s(\tau) = \int_{-\infty}^{\tau} \frac{\text{Number of Words from US}_{st}}{\text{Number of New Words in CN}_s} dt \tag{2}$$

Figure 11(b) plots the time gap between words' first appearance in China and US, where a positive number indicates the year lag of Chinese patent. We are going to show that for the frontier technology in China, how many years ago they have already appeared in the US, and we call it the *retrospective* technology gap.

The lines in this figure can be interpreted as following: For the new words appeared in the Chinese patents between 1996 and 2000, 10.50% of them had a 10-year or above time lag with the US patents and 14.26% of them had a 5-year or above time lag with the US patents. For the new words appeared in the Chinese patents between 2011 and 2015, 5.82% of them had a 10-year or above time lag with the US patents and 6.16% of them had a 5-year or above time lag with the US patents. The interaction of those curves and $\tau = 0$ tells that, in the Chinese patents, fewer and fewer of new words were using the existing technology the in US since 1990. From this perspective, the technology gap between China and US was narrowing over years.

[Figure 11 About Here]

4.3 Knowledge Flows as Measures of Patent Quality

In this part, we first aggregate the patent citation network among US and Chinese patents to country and patent-class level. The patent citation network then reflects the flows of knowledge across countries and across patent classes. Moreover, we construct the new word citation network across countries and across patent classes to measure the knowledge transfer.[16] We visualize the knowledge flows with the contour plots (Acemoglu, Akcigit and Kerr, 2016). Each row represents the knowledge-using (target) class and each column represents the knowledge-supplying (source) class. The darker the color, the more important the patent class is as the source of knowledge. The red dashed lines are used to indicate different patent classes.

We show that, first, in terms of both patent citations and new word citations, the

[16]We analysis the knowledge transfer from the aspects of patent citation network and the new word citation network.

Chinese patents rely heavily on the US patents. Second, the Chinese patents in the classes electricity (H) and physics (G) are becoming more influential since 2000, in terms of knowledge transfer from China to US. Third, some technology classes such as human necessities (A), fixed construction (E) and chemistry (C) consistently contribute to the knowledge transfer from China to US measured by new word citation, though the share of knowledge flows from China to US is still tiny.

4.3.1 Patent Citation Network

We first construct the cross-patent-class citation network of US and Chinese patents separately based on domestic patent citation. The results are shown in Figure 12.

There are two common features shared by Chinese and US patent citation network. First, we observe networks with strong diagonal vector. Not surprisingly, the most important knowledge source of each patent class is itself. Second, some patent classes, such as physics and electronics are important knowledge source of all other patent class. However, Chinese patent citation network is much more sparse than the US, which indicates that the cross-class knowledge flows among the Chinese patents is much weaker than that in US. The Chinese scientists cite fewer patents outside their own research fields and the knowledge produced by them has less impact on other technology classes.

[Figure 12 About Here]

We then pool the Chinese and US patents together and construct the cross-country and cross-patent-class citation network. The result is shown in Figure 13. The upper left quadrant presents CN-CN citations; the upper right quadrant presents CN-US citations, which means knowledge flows from US to China; the lower left quadrant presents US-CN citations, which represents knowledge flows from China to US; and finally the lower right quadrant presents the US-US citations.

The first salient pattern is the strong within-country and within-patent-class citations, which is represented by the dark diagonals in Figure 13. The second salient feature is that there is a strong diagonal in the upper right quadrant, suggesting that the Chinese patents learn a lot from the US patents, especially from the patents in the same technology class. To our surprise, we observe an almost empty lower left quadrant, suggesting that although the US patents do cite the Chinese patents, the share is quantitatively negligible.

[Figure 13 About Here]

From the data underlying Figure 13, we can construct the yearly knowledge flows within and across the two countries and the patent classes, and calculate the flow share for all combination. Specifically, there are four kinds of knowledge flows to China: within-patent-class knowledge flows from China to China, which is denoted by CN-CN-within; cross-patent-class knowledge flows from China to China, which is denoted by CN-CN-cross; within-patent-class knowledge flows from US to China, which is denoted by US-CN-within; and cross-patent-class knowledge flows from US to China, which is denoted by US-China-cross. Analogously for the knowledge flows to US, there are four types.

The composition of knowledge flows in each year is shown in Figure 14. It shows that within-country knowledge flow dominates the others, that the inventor cites domestic knowledge most. Second, there is substantial knowledge flow from US to China, but the flow is gradually decreasing. This feature is consistent with the first one that the number of Chinese patents is increasing and inventors tend to cite domestic knowledge more. Third, the knowledge flow from China to the US is tiny relative to that from the US to China, but it is emerging in recent years. We find that, since 2001, the top three contribute to knowledge transfer from China to US were electric communication technique (H04), basic electric elements (H01), and computing, calculating and counting (G06). In year 2015, the knowledge transfer from China to US accounted for only 0.8% of the total China-US knowledge transfer with almost equal share of within- and cross-patent-class transfer.

[Figure 14 About Here]

4.3.2 New Word Citation Network

Besides the patent citation network, we can also study the knowledge flows based on new word citation network.

A few words about the pros and cons of the patent citation network and the new word citation network are in order. In the application files, the applicant might strategically choose not to cite some patents in order to get the patent granted. This activity would potentially bias the outcome of patent citation network and mislead the analysis on the importance of patent classes. The new word citation network can better address this problem to a certain degree, since it can trace the original source of knowledge. Even if the patents with the original ideas are not shown in the reference list, their ideas would be reflected in the citing patents through wording. One may argue that the applicant might strategically create new words to make difference. This activity is less likely to bias the results, since such man-made new words are unlikely to be cited by later patents. This is the advantage of new word citation network. The construction of the new word citation network is similar to the patent citation network, but they differ in one important aspect. The new word citation focus on the knowledge flows from the patent where the word first appeared, but does not track the important later patents on the knowledge flow path.

We present the citation network of new words within China and US in Figure 15(a) and Figure 15(b), where the new word lists are constructed separately for US and China. The first common feature shared by Chinese and US new word citation network is the strong diagonal vector. Meanwhile, some patent classes supply knowledge to all other fields. Similarly, the new word citation network in US is denser than the Chinese one. The patent citation network in US is consistent with the new word citation network. The key supplying patent classes identified by patent citation network also apply for the new work citation network. But in China, the situation id different. The new word citation network in China and in US are quite alike, which share similar density.

[Figure 15 About Here]

Next, we pool Chinese and US patent data to construct the cross-country new word citation network and investigate the cross-country and cross-patent-class knowledge flows, using the pooled new word list. This new word citation network is shown in Figure 16. One distinct feature is that those highly-cited patent classes in Chinese network seem to originate from the US. We observe a dense upper right quadrant of the network but sparser upper left one. The dense upper right quadrant tells us that the highly cited new words in the Chinese patents originates from the US patents. Similar to the patent citation network, the new word citation network also exhibits an almost empty lower left quadrant, which indicates there is few knowledge flows from China to US. For the US-US knowledge flow, it is similar to the within US new word citation network. From the cross-country and cross-patent-class new word citation network, we can get an interesting conclusion. When a new word first appears in the US patents, the Chinese scientists in the same technology class will learn it and use the idea in their own innovation. Then this idea will spillover and innovators in other fields in China.

[Figure 16 About Here]

We construct the knowledge flow share across years using new word citations, and the result is reported in Figure 17. One distinct feature of the new word flow is that the cross-patent-class flow dominates the others, which is different from the patent citation flow. Another feature is that the change in share of new word flow is small. Contrary to the patent citation flow, we find that the new word flow from China to US was quite stable from 1985 to 2015. The top three technology classes that the Chinese patents contribute new knowledge were medical and veterinary science (A61), water supply and sewerage (E03) and organic chemistry (C07). This result is different from the findings in the patent citation network, where the technology class of electricity is becoming more influential. One possible explanation is that the novel idea in Chinese patents is not attractive to the US investors. However, the findings based on the US patents attract more attention from the US inventors.

[Figure 17 About Here]

The second and the third measurements are internally consistent that the the second measurement captures the extensive margin, that is how extensively the new words in China will enter the knowledge scope of the US patents and reversely. The knowledge flow measure captures the intensive margin, that how intensively the new words in the Chinese patents will be used in the US patents. From both the techno0logy gap and the knowledge flow, we show that the Chinese patents are having larger influence on the US patents extensively and intensively.

5 Conclusion

In this paper, we provide an anatomy of the patent quality in China from 1985 to 2020. We construct various quality measures to investigate the evolution of the quality of granted Chinese patents, relative to that of US patents, over the period of 1985 and 2020. The quality measures we study include standardized forward citation counts, new words, and knowledge flow as well as the originality and generality, centrality, and patent renewals. The average standardized forward citations of the Chinese patents is approaching the US level. Based on new words in patent filing text, we find that over the period of 1985 and 2020, the fraction of US new words which origin from China is increasing. We also examine the interactions between the US and Chinese patents in terms of cross-country knowledge transfer as proxied by patent citations and new word citations. We find that the knowledge flowed mostly from the US patents to Chinese patents, though in some technology classes, particularly electric communication technique (H04) and basic electric elements (H01), knowledge started to flow from China to the US from around 2001, though in 2017, the knowledge transfer from China to US accounted for only 1.1% of the total China-US knowledge transfers.

There are several interesting avenues for future research. The results in this paper suggest that the improvement of Chinese patents in the last twenty years is not solely in quantity, but also to some extent in the quality. What are the driving forces of the increase in the quality of Chinese patents? Are they driven by government policies that subsidize R&D? Or are they driven by China's private sector that is increasing integrated to the international market? We will pursue these questions in future research.

References

Abrams, David, Ufuk Akcigit, and Jillian Grennan. 2019. "Patent Value and Citations: Creative Destruction or Strategic Disruption?"

Acemoglu, Daron, Ufuk Akcigit, and William R Kerr. 2016. "Innovation network." *Proceedings of the National Academy of Sciences*, 113(41): 11483–11488.

Akcigit, Ufuk, William R Kerr, and Tom Nicholas. 2013. "The mechanics of endogenous innovation and growth: evidence from historical US patents." *Unpublished*, 106(5): 67.

Balsmeier, Benjamin, Mohamad Assaf, Tyler Chesebro, Gabe Fierro, Kevin Johnson, Scott Johnson, Guan-Cheng Li, Sonja Lück, Doug O'Reagan, Bill Yeh, et al. 2018. "Machine learning and natural language processing on the patent corpus: Data, tools, and new measures." *Journal of Economics & Management Strategy*, 27(3): 535–553.

Deng, Yi. 2007. "Private value of European patents." *European Economic Review*, 51(7): 1785–1812.

Hagedoorn, John, and Myriam Cloodt. 2003. "Measuring innovative performance: is there an advantage in using multiple indicators?" *Research Policy*, 32(8): 1365–1379.

Hall, Bronwyn H, Adam B Jaffe, and Manuel Trajtenberg. 2001. "The NBER patent citation data file: Lessons, insights and methodological tools."

Hall, Bronwyn H, Adam Jaffe, and Manuel Trajtenberg. 2005. "Market value and patent citations." *RAND Journal of economics*, 16–38.

Hu, Albert Guangzhou, and Gary H. Jefferson. 2009. "A great wall of patents: What is behind China's recent patent explosion?" *Journal of Development Economics*, 90(1): 57–68.

Hu, Albert G. Z., Peng Zhang, and Lijing Zhao. 2017. "China As Number One? Evidence from China's most recent patenting surge." *Journal of Development Economics*, 124: 107–119.

Jaffe, Adam B, and Manuel Trajtenberg. 1999. "International knowledge flows: evidence from patent citations." *Economics of Innovation and New Technology*, 8(1-2): 105–136.

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy. 2020. "Measuring technological innovation over the long run." National Bureau of Economic Research.

Kleinberg, Jon M. 1999. "Authoritative sources in a hyperlinked environment." *Journal of the ACM (JACM)*, 46(5): 604–632.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman. 2017. "Technological innovation, resource allocation, and growth." *The Quarterly Journal of Economics*, 132(2): 665–712.

Lanjouw, Jean O, Ariel Pakes, and Jonathan Putnam. 1998. "How to count patents and value intellectual property: The uses of patent renewal and application data." *The journal of industrial economics*, 46(4): 405–432.

Lerner, Josh, and Amit Seru. 2017. "The use and misuse of patent data: Issues for corporate finance and beyond."

Liang, Mark. 2011. "Chinese patent quality: Running the numbers and possible remedies." *J. Marshall Rev. Intell. Prop. L.*, 11: i.

Liang, Yousha, Kang Shi, Hanyi Tao, and Juanyi Xu. 2019. "Knowledge Absorbing by Exporting: Evidence from Patent Citations in China." *Working Paper*.

Lin, Jia, Ho-Mou Wu, and Howei Wu. 2020. "Has China enhanced her Innovation Capability? A Multi-dimensional examination of the quality of China's new patents." *Working Paper*.

Martinez, Catalina. 2010. "Insight into different types of patent families."

Pakes, Ariel. 1986. "Patents as Options: Some Estimates of the Value of Holding European Patent." *Econometrica*, 54(4): 755–784.

Prud'homme, Dan, and Taolue Zhang. 2017. "Evaluation of China's intellectual property regime for innovation: Summary Report." *Summary Report for the World Bank*.

WIPO. 2019. *World intellectual property indicators 2019.* WIPO publication.

Zuniga, P, D Guellec, H Dernis, M Khan, T Okazaki, and C Webb. 2009. "OECD patent statistics manual." *Francia: OECD Publications*.

List of Figures

Figure 1: Time Line of the Evolution of the Chinese Patent System

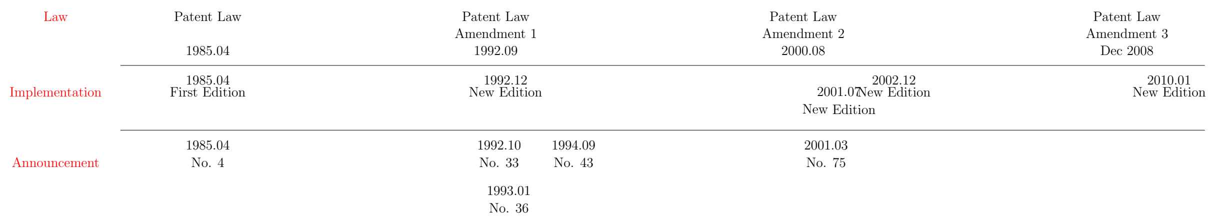


Figure 2: Average Patent Examination Time: China and US

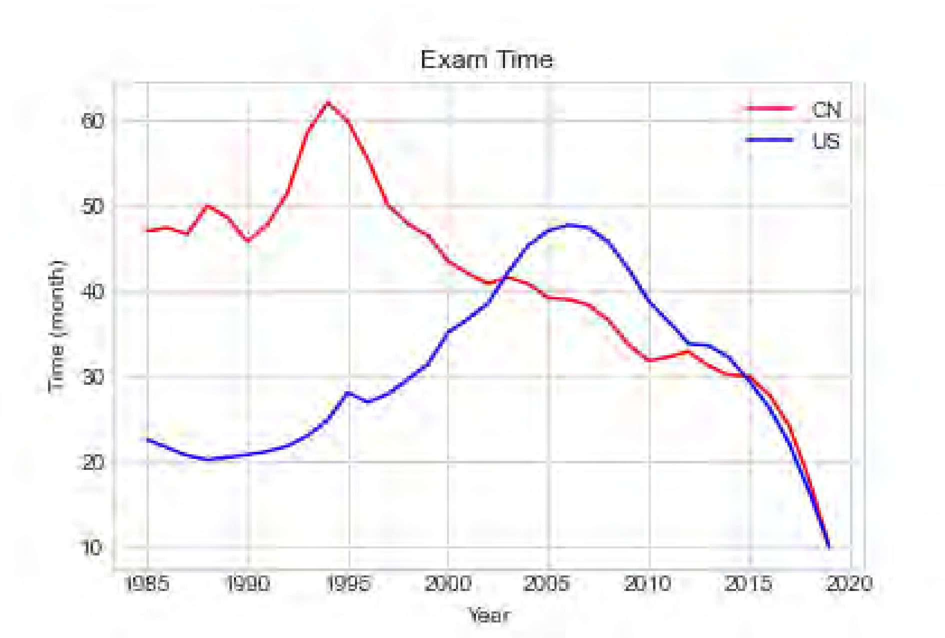
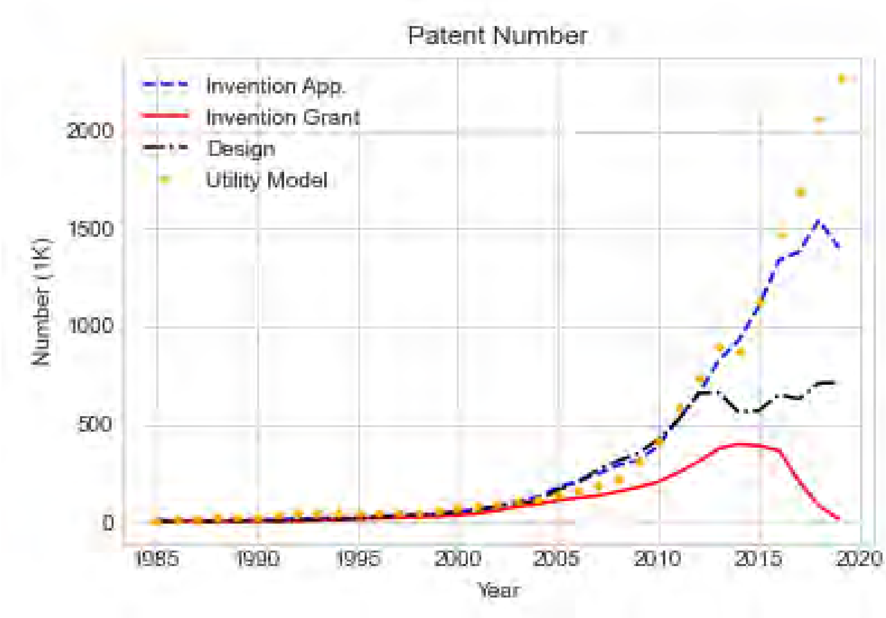
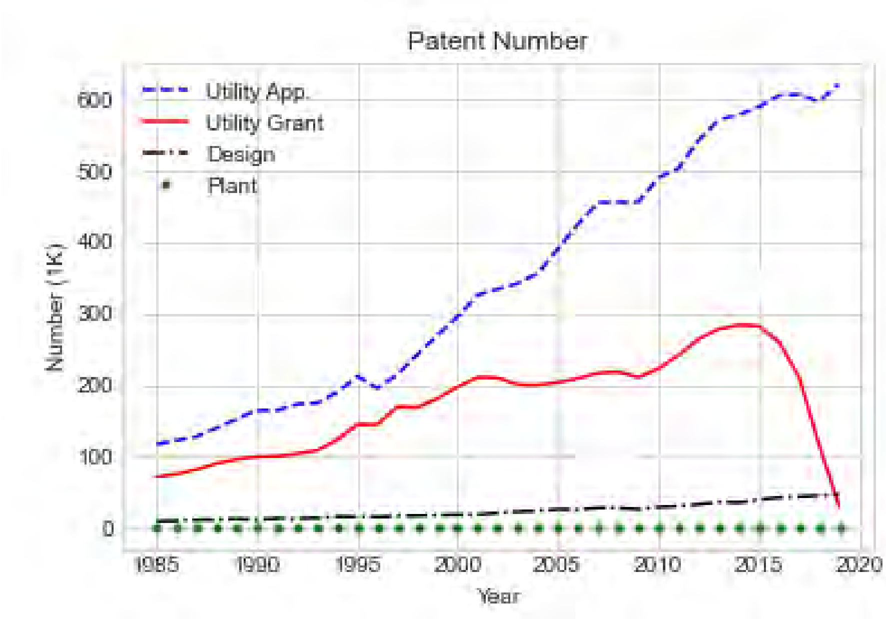


Figure 3: Total Number of Patent Fillings and Grants: China and US

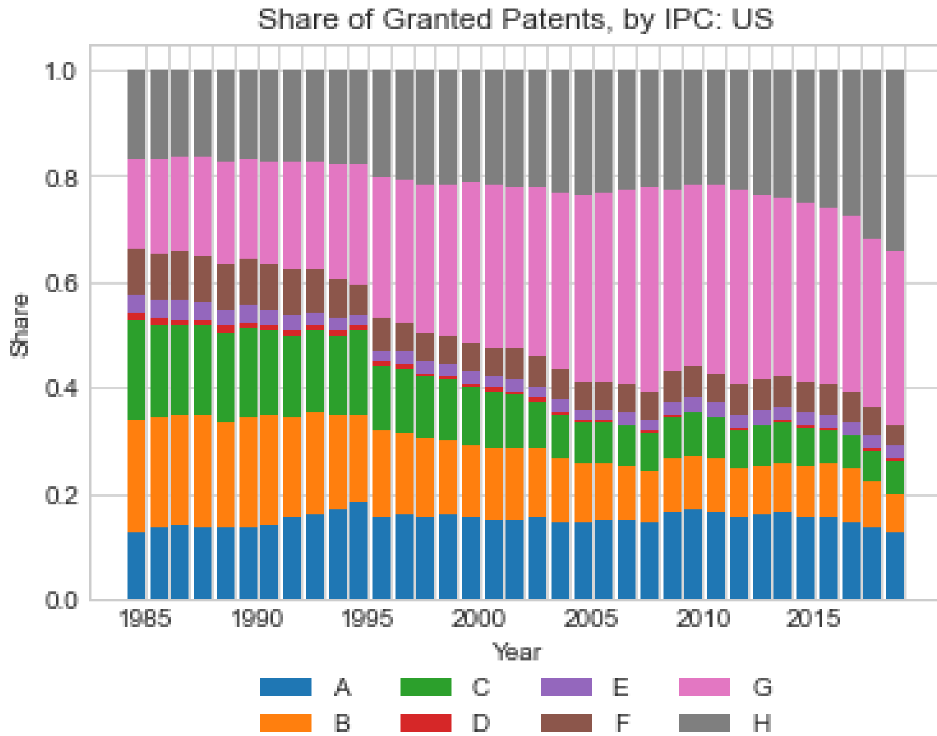


(a) China

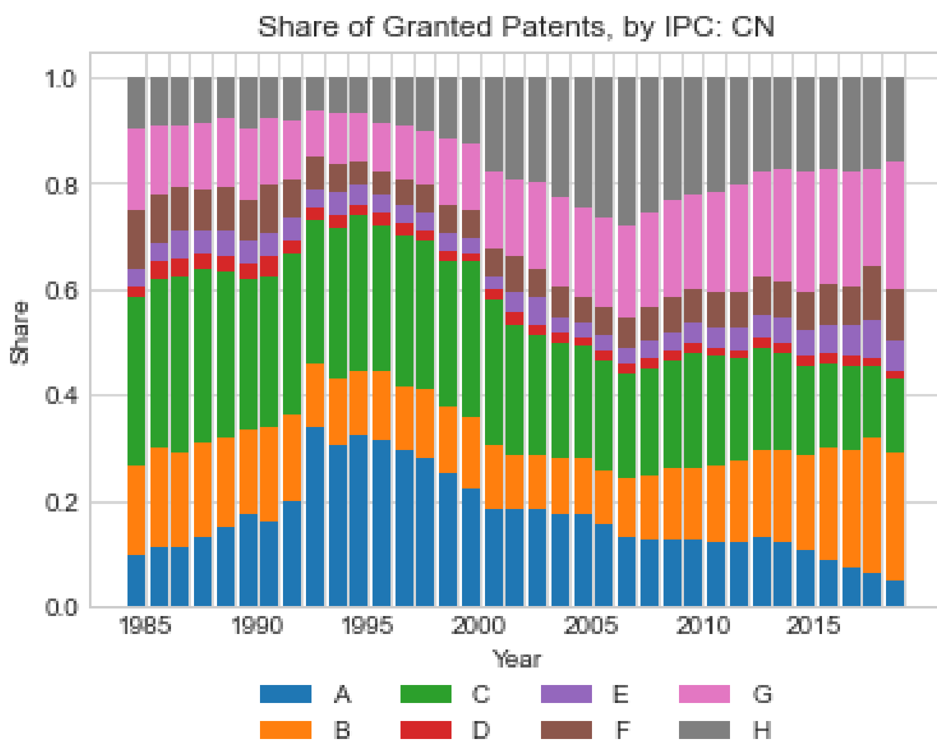


(b) US

Figure 4: Shares of Granted Patents, by IPC

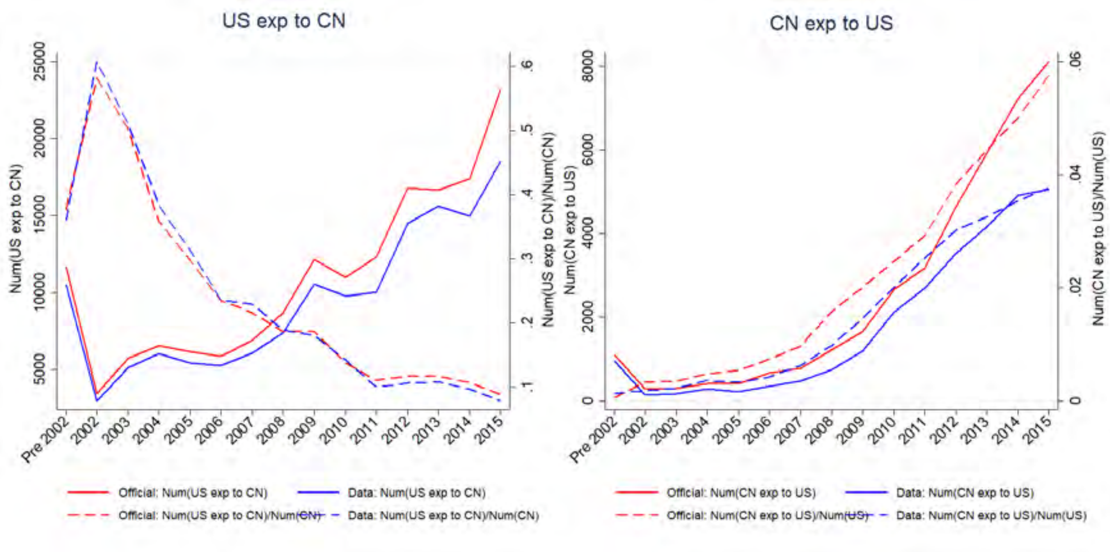


(a) US



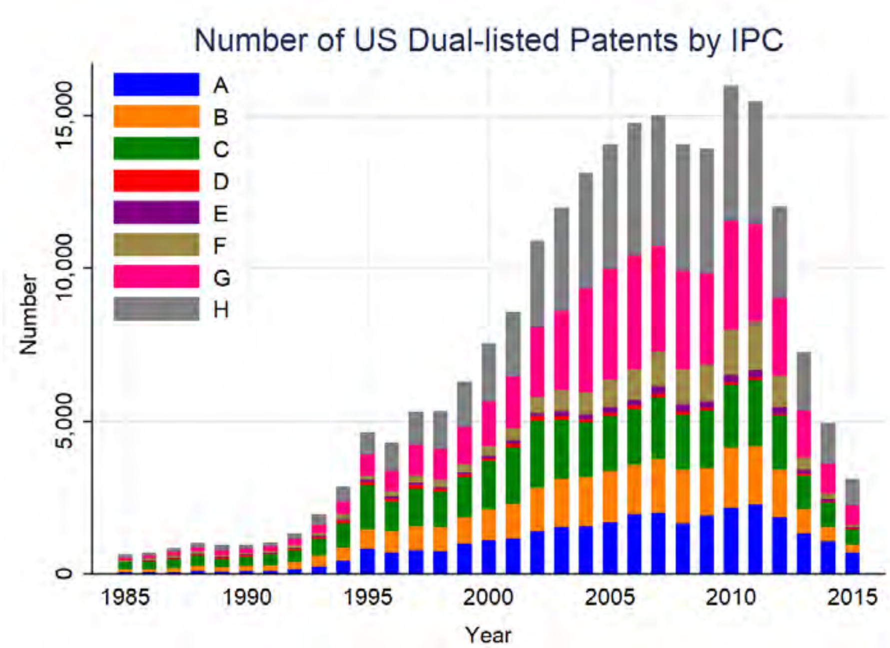
(b) China

Figure 5: Comparison with Official Statistics

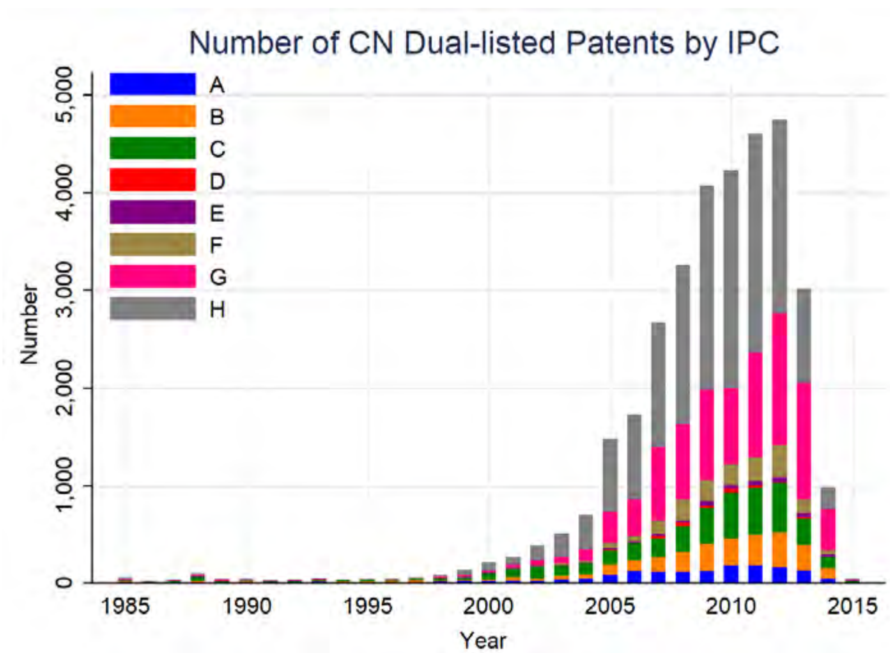


Note: (1) The red line indicates official data, and the blue line represents our data. We compare our data with the official data in two dimensions. First is the number of the patent exported to US or CN; Second is the ratio of exported patents number over domestic patents number. (2) The left graph documents the matching conditions of US patents exported to CN. (3) The right graph documents the matching conditions of CN patents exported to US The official statistics source from yearbook of U.S. patent,https://www.uspto.gov/web/offices/ac/ido/oeip/taf/reports.htm. (4) Pre 2002 is the total accumulated number before 2002.

Figure 6: Dual-listed Patents Number

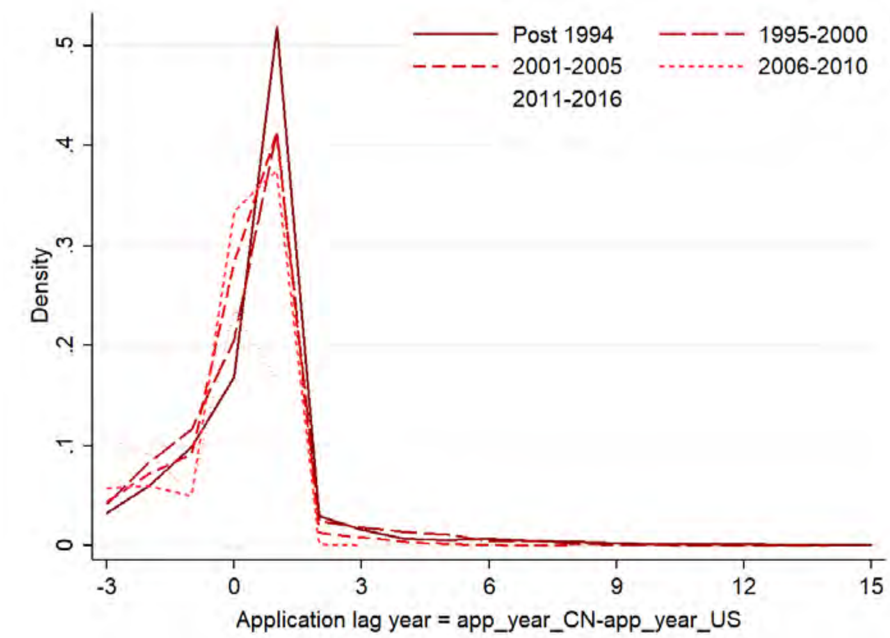


(a) US Patents exported to China

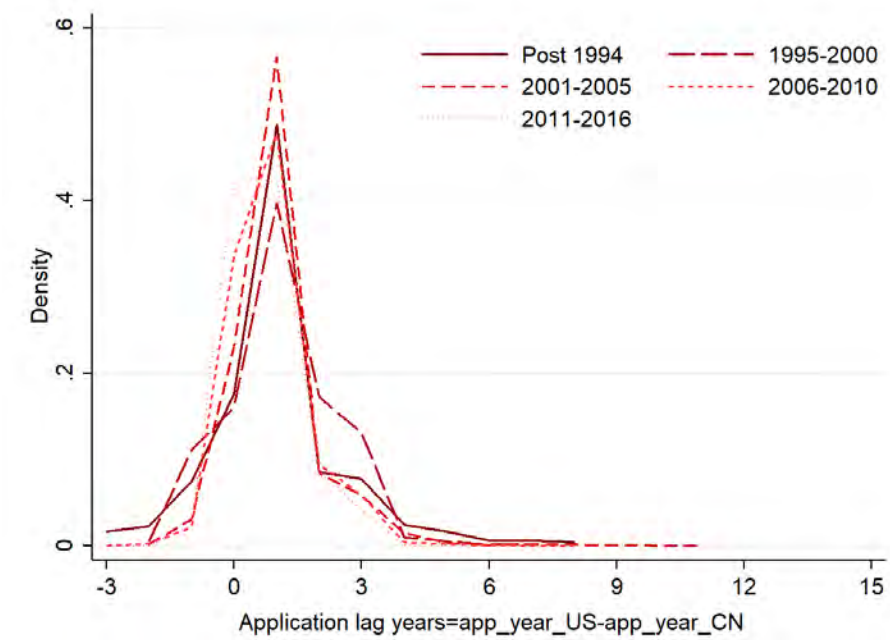


(b) China's Patents exported to US

Figure 7: Application Lag of Multinational Patents

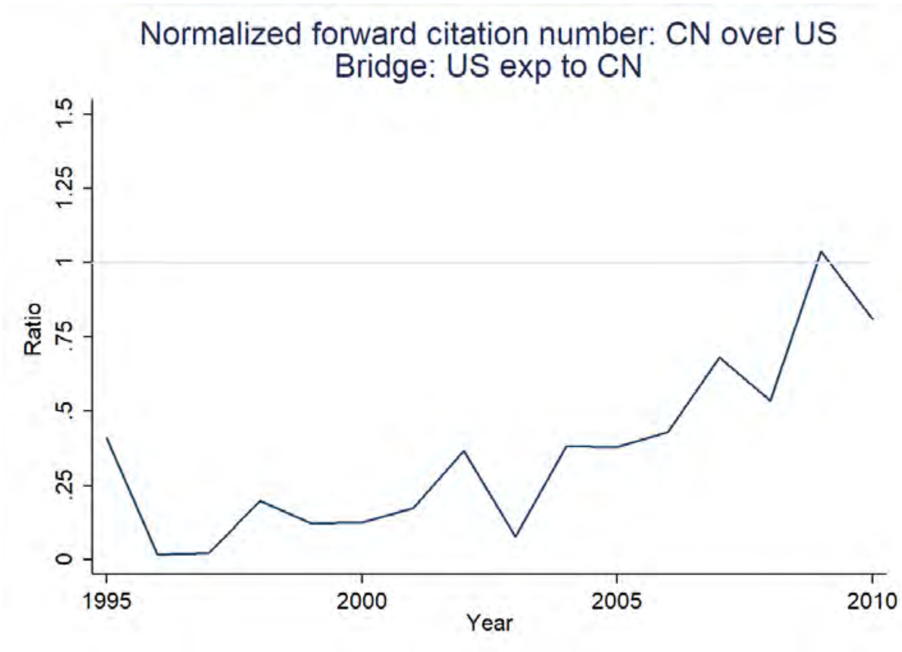


(a) Application Lag of US Patents Exp to China

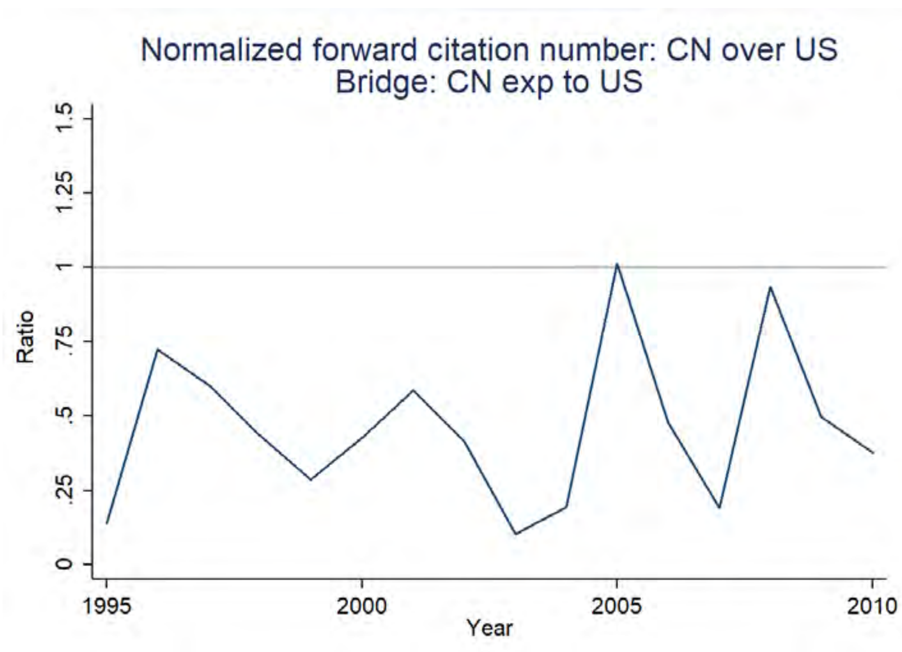


(b) Application Lag of China's Patents Exp to US

Figure 8: Standardized Forward Citation Gap between China and US

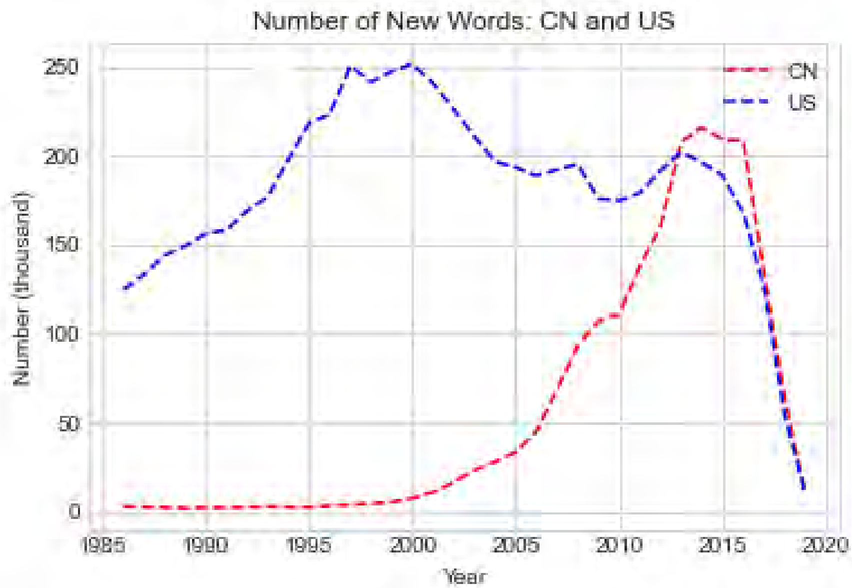


(a) US exp to CN as Bridge

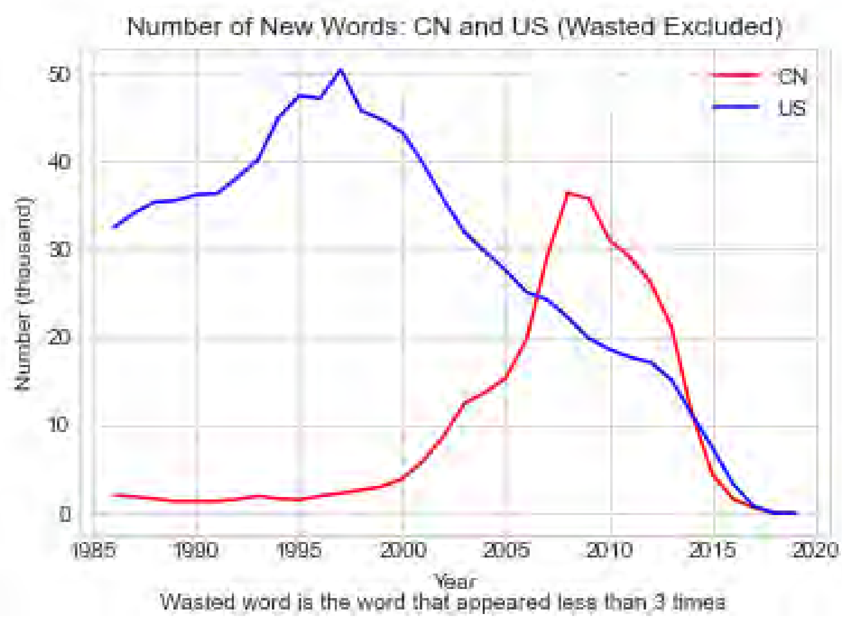


(b) CN exp to US as Bridge

Figure 9: Number of New Words: Separate Lists for China and US

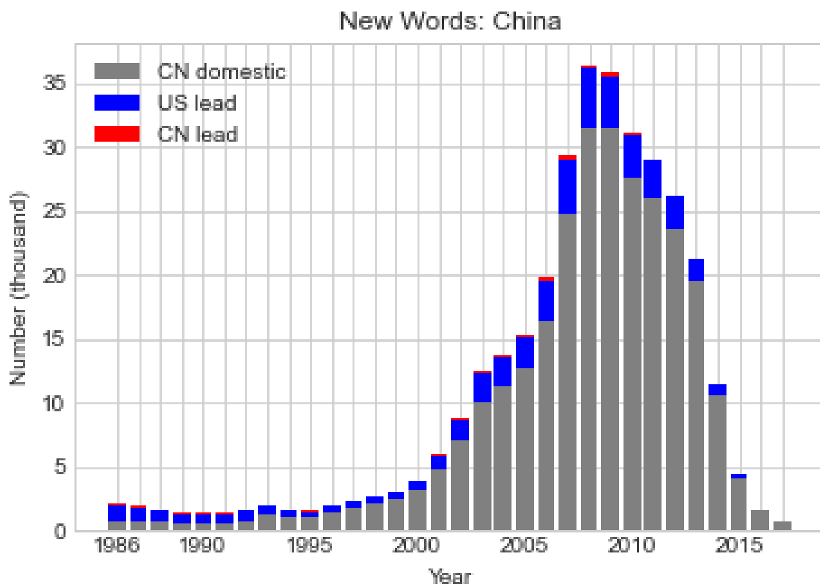


(a) All

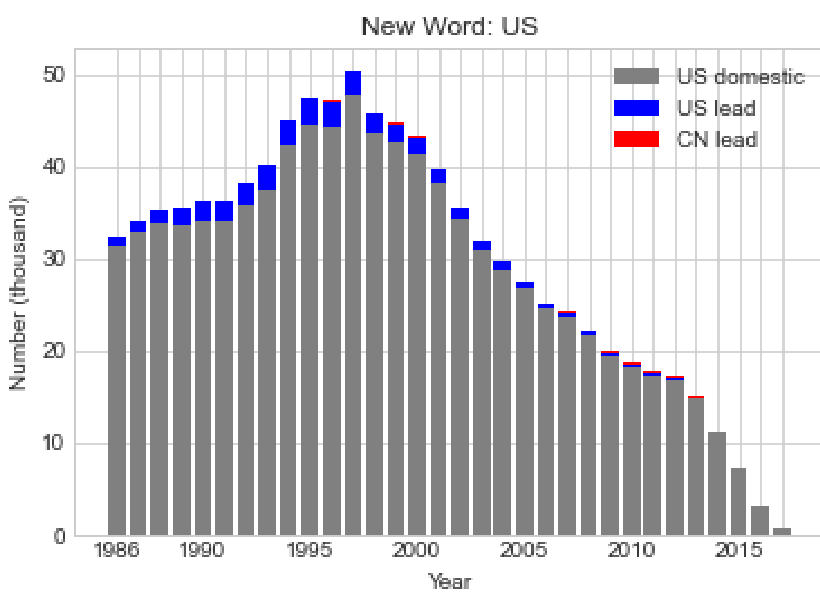


(b) Non-wasted

Figure 10: Decomposing the Country-Specific New Words

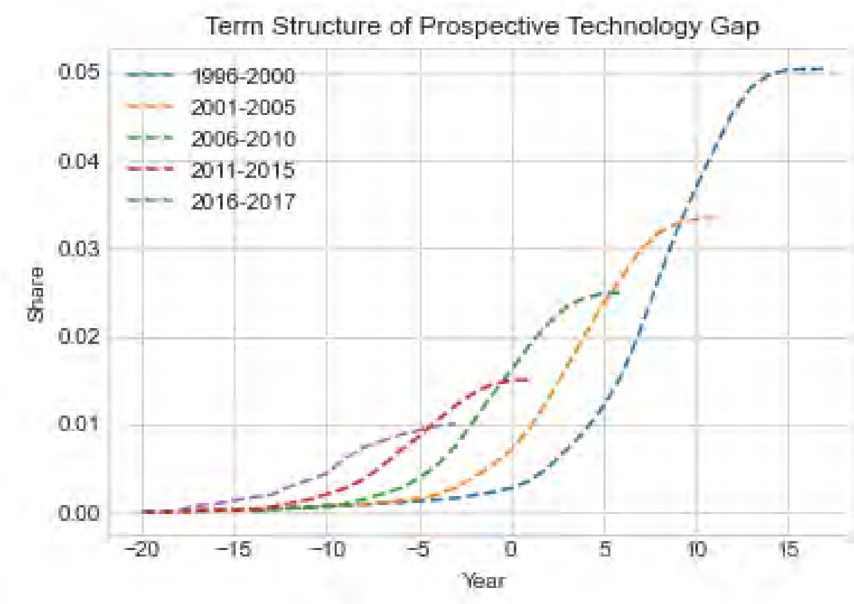


(a) China

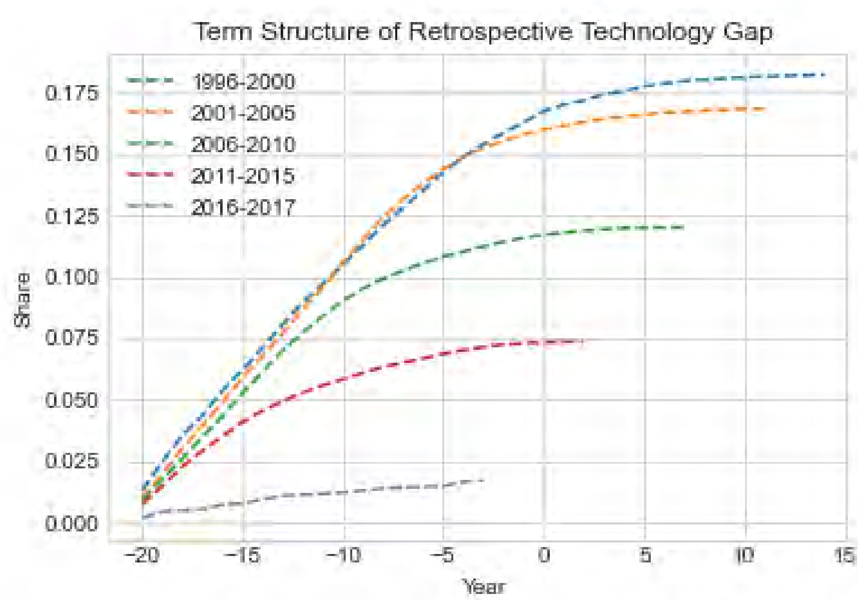


(b) US

Figure 11: Prospective and Retrospective Technology Gaps Between China and US Using New Words

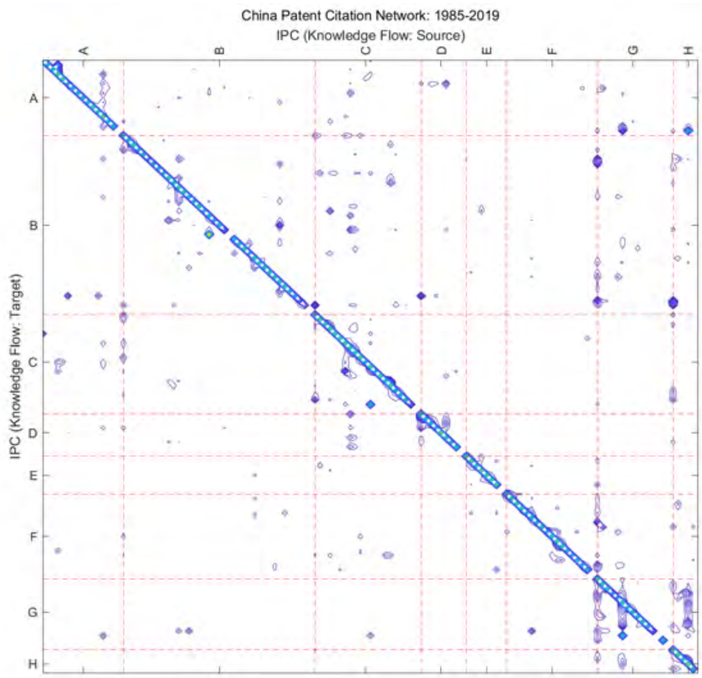


(a) Prospective Technology Gap

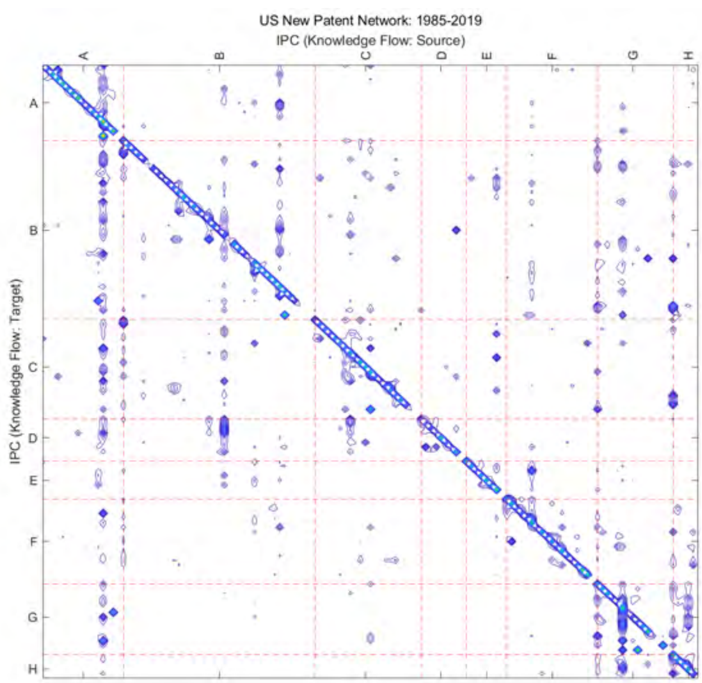


(b) Retrospective Technology Gap

Figure 12: Cross IPC Patent Citation Network in China and US: 1985-2019



(a) China



(b) US

Figure 13: China-US and Cross IPC Patent Citation Network: 1985-2019

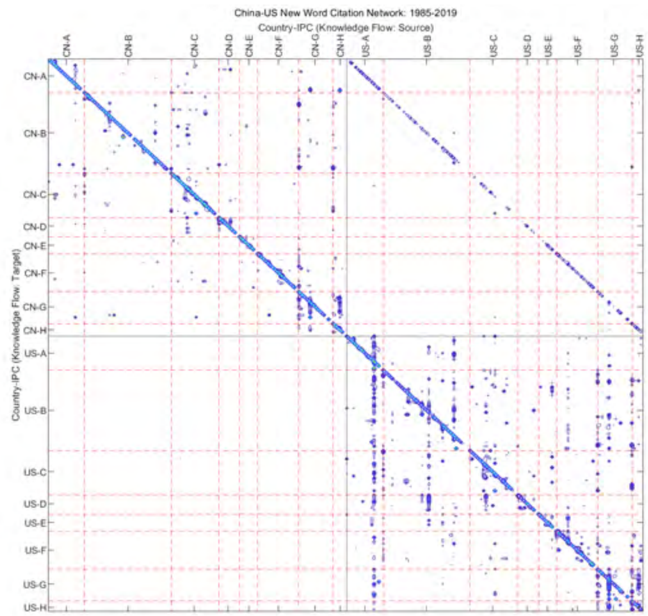


Figure 14: China-US Knowledge Flow Share: 1985-2019

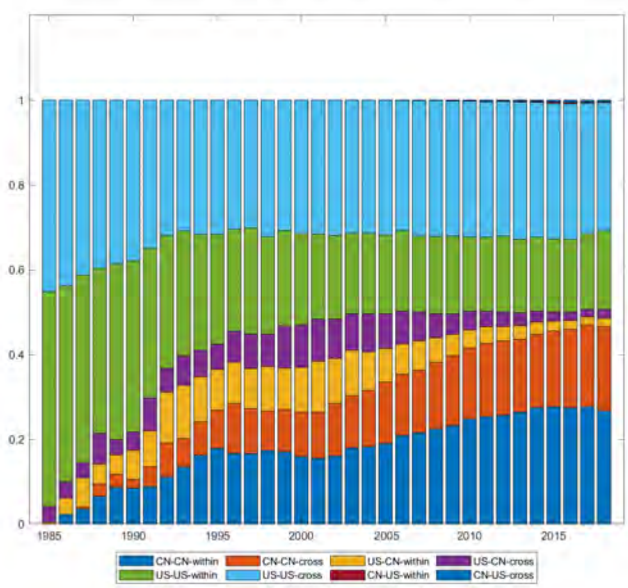
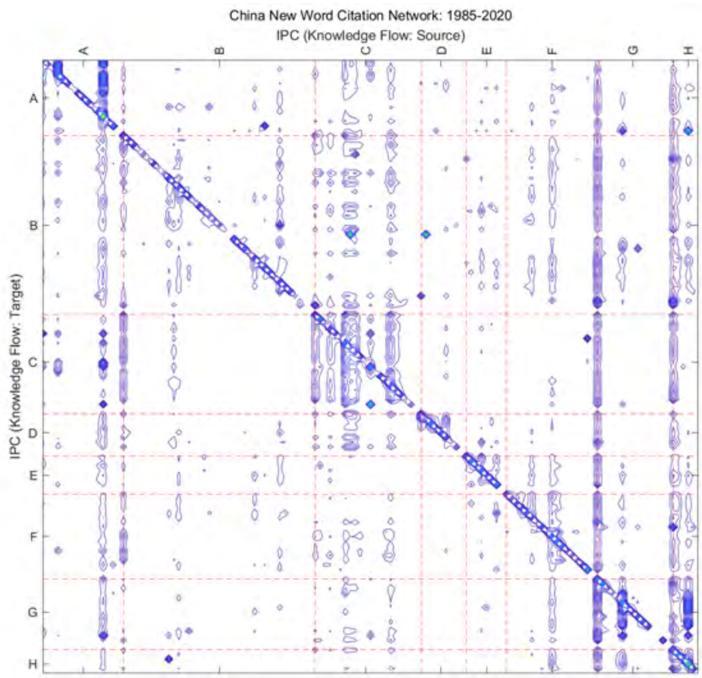
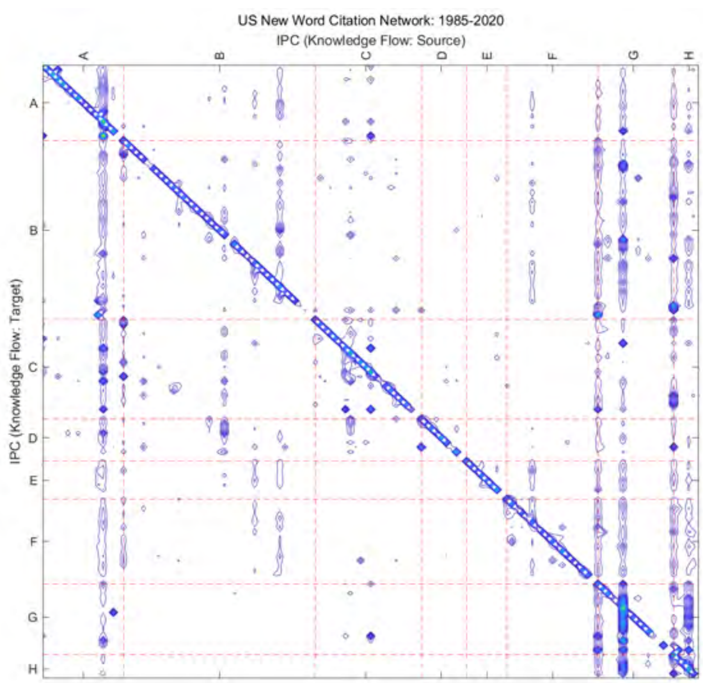


Figure 15: Cross IPC New Word Citation Network in China and US: 1985-2019



(a) China



(b) US

Figure 16: China-US Cross IPC New Word Citation Network: 1985-2019

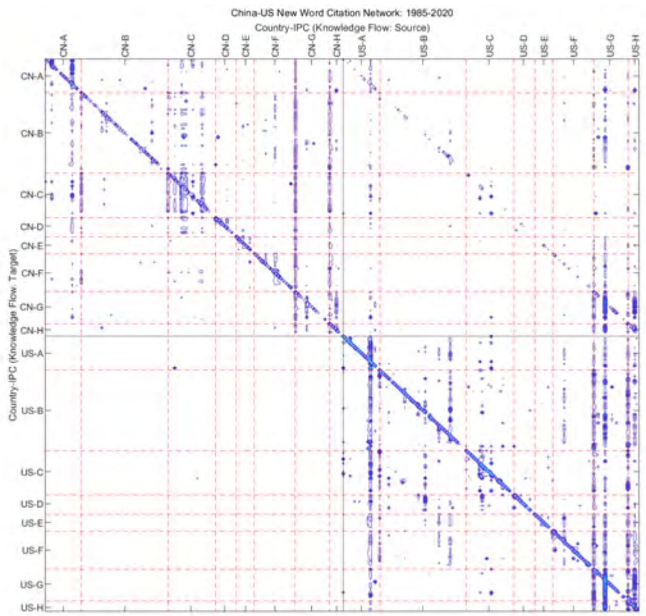
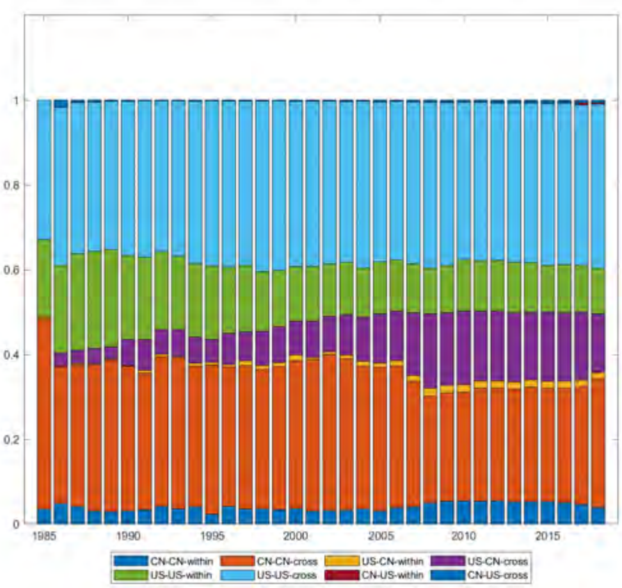


Figure 17: China-US Knowledge Flow(Word) Share: 1985-2019



List of Tables

Table 1: Cost of Applying and Holding Invention Patent

	1985.04	1992.10	1993.01	1994.09	2001.03
Announcement	No. 4	No. 33	No. 36	No. 43	No. 75
Application Fee		300	340	490	950
Application Maintenance Fee		200	200	300	300
Examination Fee		800	800	1200	2500
Re-Examination Fee		400	400	600	1000
Claim Fee (per claim)			50	80	80
Maintenance Fee (per year)					
1st - 3rd			400	600	900
4th - 6th			600	900	1200
7th - 9th		800	800	2000	2000
10th - 12th		1500	1500	2000	4000
13th - 15th		3000	3000	4000	6000
16th - 20th			6000	8000	8000
Additional Claim Fee (from 11th claim)		20	20	30	150

Table 2: Comparison of Patenting Cost

	China	US
Term of Patent	20 Years (from filing date)	20 Years (from filing date)
Application Fee	950 (135)	300
Examination Fee	2500 (357)	760
Claim Fee (per claim)	80 (11)	
Each claim in excess of 10	150 (21)	
Each independent claim in excess of 3		460
Each claim in excess of 20		100
Patent Maintenance Fees (per year)		
1st - 3rd	900 (128)	
4th - 6th	1200 (171)	
7th - 9th	2000 (285)	
10th - 12th	4000 (571)	
13th - 15th	6000 (857)	
16th - 20th	8000 (1142)	
Patent Maintenance Fees		
Due at 3.5 years		1600
Due at 7.5 years		3600
Due at 11.5 years		7400

Table 3: International Patent Classification

Class	Contents
A	human necessities
B	performing operations
	transporting
C	chemistry
	metallurgy
D	textiles
	paper
E	fixed constructions
F	mechanical engineering
	lighting
	heating
	weapons
	blasting engines or pumps
G	physics
H	electricity

Table 4: OST Technology Classification

sector_en	field_en
Electrical engineering	Electrical machinery, apparatus, energy
Electrical engineering	Audio-visual technology
Electrical engineering	Telecommunications
Electrical engineering	Digital communication
Electrical engineering	Basic communication processes
Electrical engineering	Computer technology
Electrical engineering	IT methods for management
Electrical engineering	Semiconductors
Instruments	Optics
Instruments	Measurement
Instruments	Analysis of biological materials
Instruments	Control
Instruments	Medical technology
Chemistry	Organic fine chemistry
Chemistry	Biotechnology
Chemistry	Pharmaceuticals
Chemistry	Macromolecular chemistry, polymers
Chemistry	Food chemistry
Chemistry	Basic materials chemistry
Chemistry	Materials, metallurgy
Chemistry	Surface technology, coating
Chemistry	Micro-structural and nano-technology
Chemistry	Chemical engineering
Chemistry	Environmental technology
Mechanical engineering	Handling
Mechanical engineering	Machine tools
Mechanical engineering	Engines, pumps, turbines
Mechanical engineering	Textile and paper machines
Mechanical engineering	Other special machines
Mechanical engineering	Thermal processes and apparatus
Mechanical engineering	Mechanical elements
Mechanical engineering	Transport
Other fields	Furniture, games
Other fields	Other consumer goods
Other fields	Civil engineering

Table 5: Patent Number: Home Country and Exported Country

Patent Type	Patent Num in Home Country	Patent Number in Exported Country				
	Patent Num	Mean	Std	Median	75p	99p
US exp CN	222,908	1.61	2.18	1	1	11
CN exp US	33,837	1.25	2.06	1	1	5

Appendix

A Other Measures in the Literature

A.1 Patent Renewal as a Measure of Patent Quality

First, we use patent renewal data to assess the quality of patents (Pakes, 1986; Lanjouw, Pakes and Putnam, 1998; Deng, 2007). Patent holders are more likely to renew their patents if they perceive that their patents have higher value. Since patent values are likely to monotonically increasing in patent quality, patent renewal decisions can be used to proxy th equality of patent, thus the patent renewal behavior gives us a window to investigate the quality of patents. We show the Kaplan–Meier curve of patents in different time periods in China and US. To simplify the graphical representation of or results, we divide the patents into 6 groups according to their application years.

Figure 1(a) shows the survival rate of the patents in China. Since the term of patent has been changed in the 1992 amendment of China's Patent Law, we do not compare the 1986-1990 group with the others. The 1991-1995 and 1996-2000 group showed similar patterns in terms of the short-run (5-year) survival rate, which were 98.22% and 99.51% respectively. There was also no big difference in the medium-run (10-year) survival rate, which were 73.14% and 74.48% respectively. For the long-run (20-year) survival rate, the 1991-1995 group was 21%, and the 1996-2000 group was higher at 33.6%. However, while there was a slight drop in the short-run survival rate for the 2001-2005 group, the medium-run survival rate decreased to 64.11%. The situation improved a little for the 2006-2010 group, that its medium-run survival rate was 77.89%. Because of the right-censored problem, we could get the short-run survival rate for the 2011-2015 group, which was 96.15% and it was better than the 2006-2010 group (88.68%).

For the survival rate of the US patents, it has changed a little since 1996, as shown in Figure 1(b).[17] This short-run survival rate of patents after 1996 was almost 1, because of the maintenance fee schedule in US. The medium-run survival rate was on average 81.54% for the groups after 1996. The long-run survival rate achieved 49.56% for the 1996-2000 group and was expected to be higher for the 2006-2010 group.

One distinct difference between the Chinese and US patents is that the medium-run and long-run survival rate of US patents is substantially higher (more than 20%), despite

[17] The stage-style pattern of the 1986-1990 and 1991-1995 group is resulted from term of patent in US before 1995.

the differences in the maintenance fees in the two countries.

A.2 Generality and Originality as Measures of Patent Quality

Second, we use the originality and generality scores proposed in (Hall, Jaffe and Trajtenberg, 2001) to measure the novelty of a patent. If a patent has higher *originality* score, it means that the patent cites patents from a broader set of technology classes, which indicates that it combines broader set of knowledges and thus more original; if a patent has higher *generality* score, it means that this patent has been cited by patents in a broader set of technology classes, which indicates that the focal patent has more impacts on broader knowledge. When computing the originality and generality scores, the patent classes are based on 3-digit IPC we discussed in Section 3.

Formally, patent i's *generality score* is defined as:

$$Generality_i = 1 - \sum_{j=1}^{m_i} \left(\frac{M_{ij}}{M_i} \right)^2 \tag{3}$$

where j is the index of patent classes, and m_i the number of patent classes of patents that cited the focal patent i; M_i is the total number of patents citing patent i, and M_{ij} is the number of patents in class j citing patent i. A higher value of generality score means that the patents citing patent i are less concentrated in a particular IPC class.

The originality score of patent i is defined as:

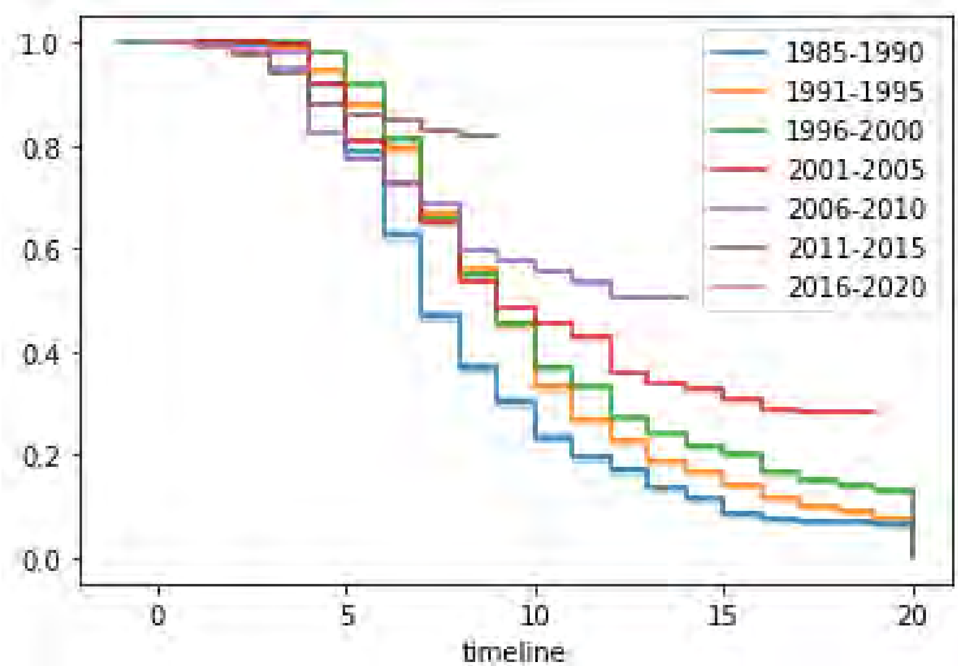
$$Originality_i = 1 - \sum_{j=1}^{n_i} \left(\frac{N_{ij}}{N_i} \right)^2 \tag{4}$$

where j is the index of patent classes, and n_i the number of patent classes of patents that are cited by the focal patent i; N_i is the total number of patents cited by patent i, and N_{ij} is the number of patents in class j cited by patent i. A higher value of originality score indicates that patent i cites more diverse set of patents.

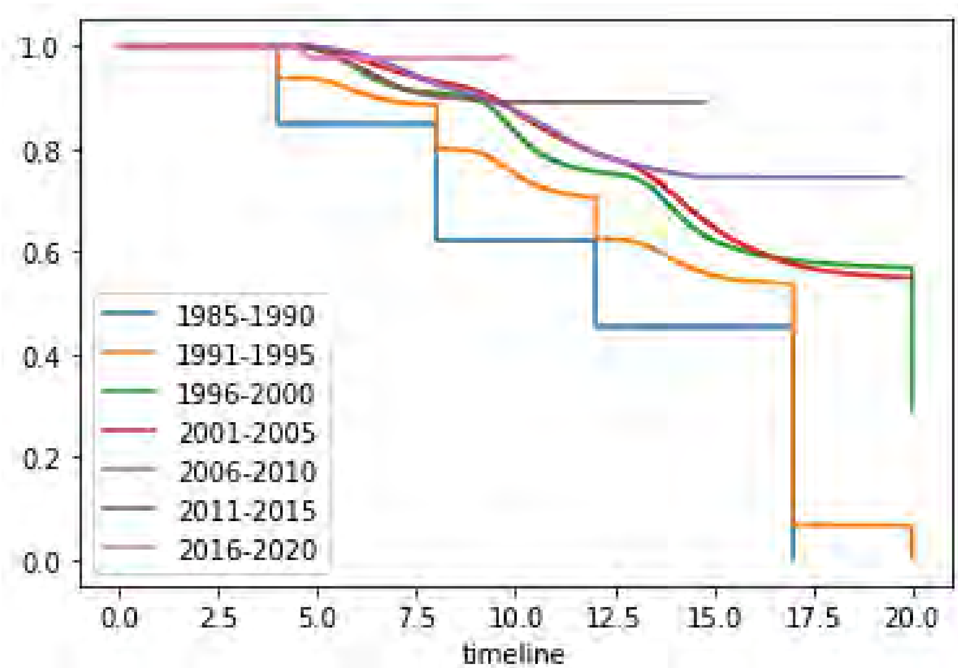
Figure A2 shows that the originality score of the Chinese patents was around zero before 1990, indicating that there was only within-patent-class knowledge transfer at that time in China. The originality score of the Chinese patents has risen steadily since then and in 2011 it reached 40% of the level of the US patents. After 2011, the originally score of Chinese patents grew more rapidly, and in 2015 it reached 80% of the US level.

Figure A2 also shows that the generality score of Chinese patents was only 20% of the

Figure A1: Kaplan-Meier Survival Curves of Patents: China and US



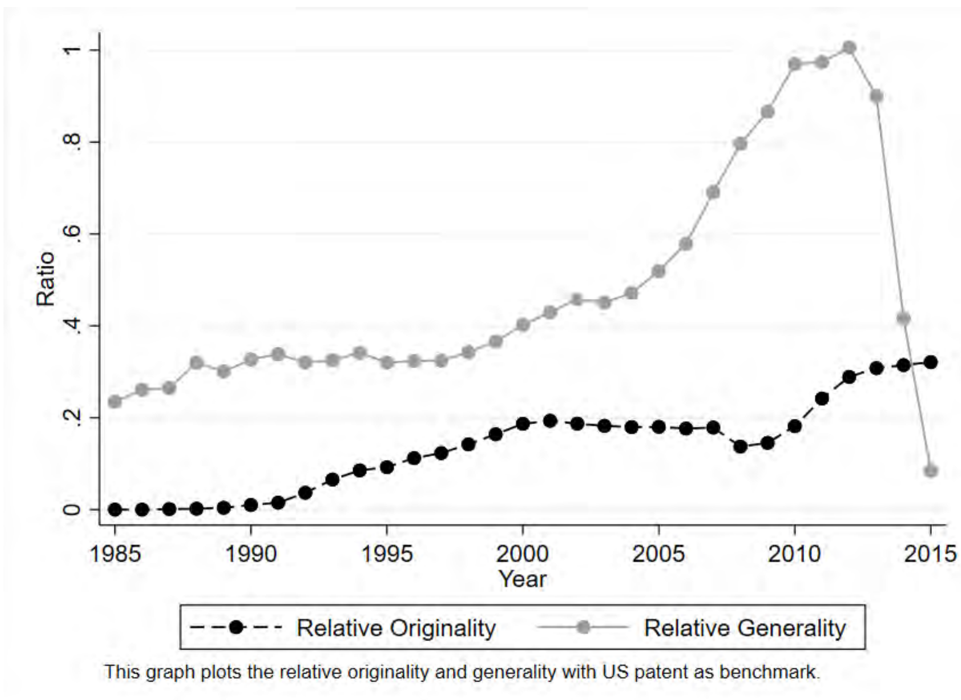
(a) China



(b) US

US patents in 1985. By 2000, the generality score of Chinese patents reached 50% of the level of US patents; and it grew rapidly after 2005, and in year 2012 there was almost no difference between the Chinese and the US patents in this measure. In the last three years, there was a large drop in the generality score of the Chinese patents relative to the US. Since both data suffers from the right-censored problem, but a possible explanation for the Chinese patents to be more impacted by the right-censored problem is that it takes the Chinese patents longer time than the US patents to be cited by other patents, i.e., the right-censored problems in forward citation faced by the Chinese patents are more severe than the US patents.

Figure A2: Generality and Originality Measures of Patent Quality: China vs. US



A.3 Betweenness, Authorities and Hubs Centrality as Measures of Patent Quality

In constructing the originality and generality measures on patent quality, only patents that cite or are cited by the focal patent, i.e., only patents that are linked to the focal patent by one layer, are used. Now, we extend the one-layer linkage of patents into multiple layers and construct measures of patent quality using the whole knowledge transfer *network* of patents.

In the patent citation network, the patents are connected by *citations*. Each patent is treated as a *node* in the network. The *edges* between any two nodes are defined by the forward and backward citations between them. For example, if Patent A cites Patent B, we will say that Patent Bis Patent A's *backward citation*, and Patent A is Patent B's *forward citation*. In this relationship, knowledge flows from Patent B to Patent A, and thus we define that Patent B is the *source*, and Patent A the *target*, of the knowledge transfer. The direction of edge points at the target of the knowledge transfer. We pool the Chinese and US patents, over 7 million patents in total in our analysis, to construct the patent citation network, and construct *authorities* centrality, *hubs* centrality and *betweenness* centrality measures to quantify the quality of patents.

Let A be the adjacency matrix resulting from the patent citation network. If patent i cites patent j, the cell a_{ij} in adjacency matrix A is one; otherwise, it is zero. Following the definition in Kleinberg (1999), each patent i in the patent citation network is assigned two scores: *authority score*, x_i and *hub score*, y_i. Let x and y denotes the vector of *authority score* vector and *hub score* vector respectively. The *authority score* and *hub score* are calculated as follows:

$$\lambda x = xA^T A \tag{5}$$

$$\lambda y = yAA^T \tag{6}$$

where λ is the (common) eigenvalue of the matrix $A^T A$ and its transpose AA^T; x is the eigenvector of authority matrix $C = A^T A$, y is the eigenvector of the hub matrix $H = AA^T$.

It can be show that the *authority score* or *authorities centrality* of patent i, x_i is related to the hub score vector y as follows:

$$x_i = \alpha \sum_k a_{ki}\, y_k \tag{7}$$

where α is a constant, and a_{ki} is the element of adjacency matrix A. A patent i's authorities centrality is proportional to the sum of hub scores of patents it cites. Intuitively, a patent i has a higher *authorities centrality* if it cites patents with high hub scores.

Similarly, the *hub score* or *hubs centrality* of a patent i, denoted by y_i, is related to the authority score vector x as follows:

$$y_i = \beta \sum_k a_{ik}\, x_k \tag{8}$$

where β is a constant, and a_{ki} is the element of adjacency matrix A. A patent's hub centrality is proportional to the sum of authorities centrality of the patents it is cited by. A patent has higher hub centrality if it is cited by patents with high authority scores.

Intuitively, if a patent has authority score, it indicates this patent receives knowledge transfer from other good (highly cited) patents. In words, this patent is a good "learner". If a patent has high hub score, it means this patent has been cited by other good patents (patents that cite highly cited patents). In words, this patent is a good "teacher".

Besides the *authorities centrality*, we also use *betweenness centrality* of a patent to measure its quality. In the patent citation network, for each pair of connected patents (i, j), there exists at least one *shortest* path between them, such that the number of patents linking i and j is minimized. The *betweenness centrality* of patent k defined as the number of the shortest paths that pass through patent k. It is based on the idea that a patent is central if more patents are connected via this patent to other patents. Intuitively, *betweenness centrality* measures how often a patent appears on the shortest paths linking other patents, where it serves as a bridge of past and future innovations. Formally, the *betweenness centrality* of a patent k is computed as follows:

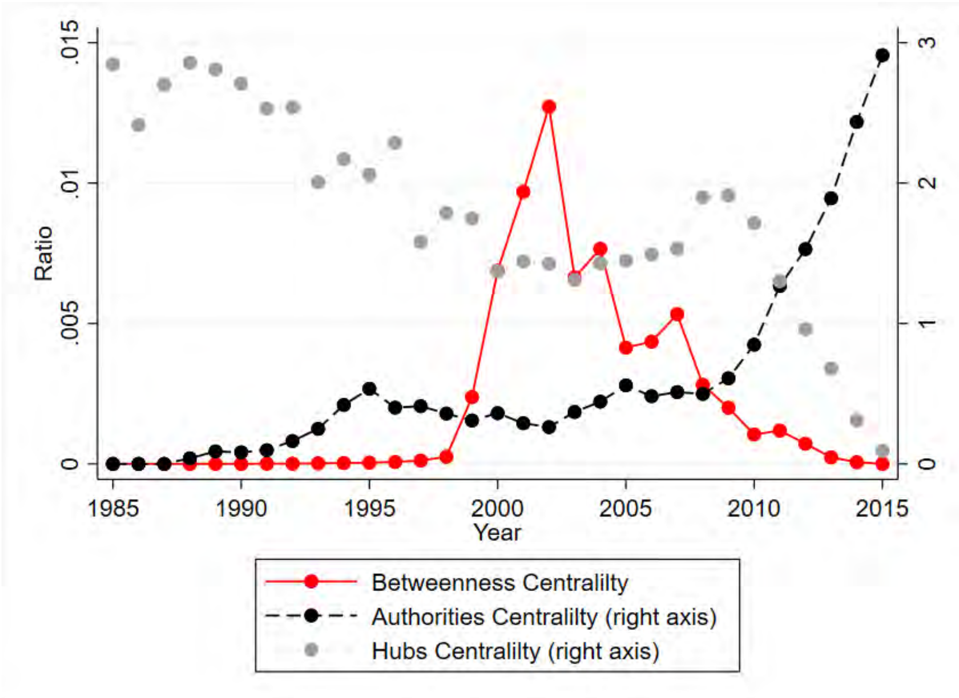
$$Betweenness\ Centrality_k = \sum_{i<j} \frac{g_{ij}(k)}{g_{ij}} \tag{9}$$

where g_{ij} is the number of shortest paths from patent i to patent j, and $g_{ij}(k)$ is the number of shortest paths from patent i to patent j which pass through patent k. A patent with a higher *betweenness centrality* indicates that it has more interaction than other patents in terms of passing knowledge in the patent citation network. Similarly, the larger the *betweenness centrality*, the more important this patent is in the patent citation network.

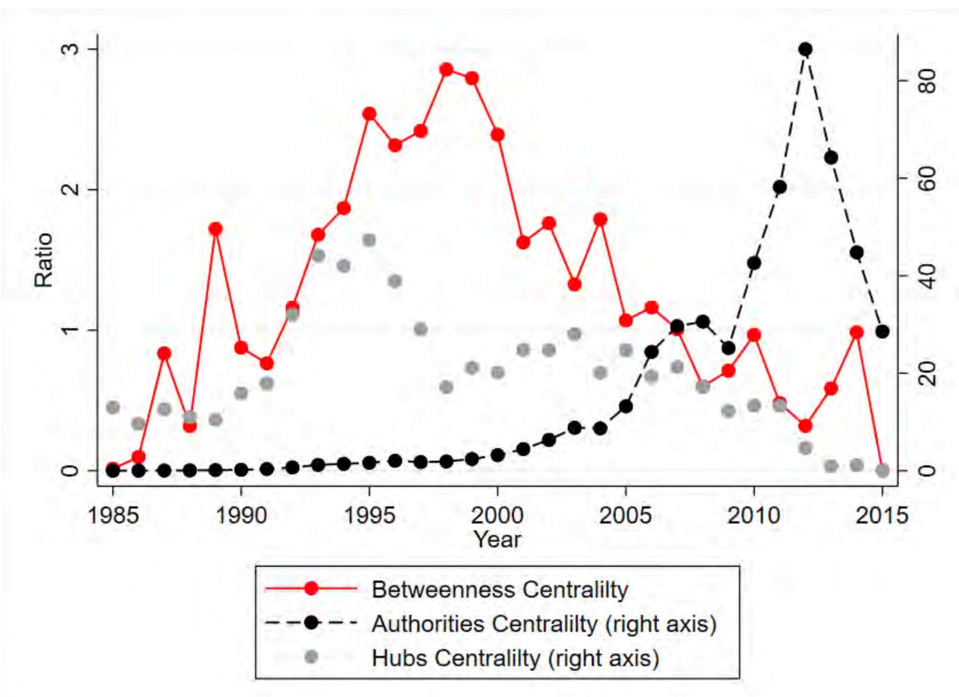
Figure A3 plots the betweenness centrality, hubs centrality and authorities centrality of Chinese and US patents separately. The upper panel shows the three centrality measures of the Chinese patents. The authorities centrality of Chinese patents slowly increased from 1985, and the growth sped up after 2010. It suggests that the Chinese inventors are becoming good "learners", which are connected to good hubs. On the other side, the hub centrality of Chinese patents was close to zero. The peak of betweenness centrality of the Chinese patents appeared around 2001, which could be the case that it is around the middle point in the periods covered by our data. The lower panel shows the centrality measurement on the US patents. In terms of level difference, they are much higher than the Chinese patents. The authorities centrality went up since 1985 and peaked in year

2012. The hubs centrality first went up and peaked in 1995, and then it continuously went down. The betweenness centrality showed a flat pattern between 1985 and 1998, and then it gradually fell.

Figure A3: Authorities, Hubs and Betweenness Centrality Measures of Patent Quality: China and US



(a) China



(b) US

B Data and Measurement

B.1 Assignee's Type and Location

In Patentview, the patents' assignee have been divided into 9 types, which are shown in Table below. However, this classification ignores the university and research institutions. So We make a classification according to NBER patent data clean code [18].

Table A1: Assignee Type

Patentview			NBER		
Type	Obs	Percent	Type	Obs	Percent
US Company or Corporation	2,951,997	49.45%	Firm	5,394,085	90.35%
Foreign Company or Corporation	2,897,671	48.54%	Government	77,847	1.30%
US Federal Government	45,232	0.76%	Institution	85,694	1.44%
US State Government	33	0.00%	Univeristy	170,022	2.97%
US County Government	14	0.00%	Individual	60,522	1.01%
Foreign Government	13,391	0.22%	Missing	181,819	3.05%
US Individual	32,104	0.54%	Total	5,969,989	100%
Foreign Individual	25,222	0.42%			
Unassigned	2	0.00%			
Missing	4,323	0.07%			
Total	5,969,989	100%			

From Table A1, the shortcoming of the assignee's type classification in patent view is that it lacks the identification of university and research institutions, which are include in the type of company, while the types in NBER lacks the identification of the nationality of the assignee. Accordingly, We need the assignee's location information.

[18] First, We standardize the assignee's name except for the individuals. Then We divide the assignee into four types based on the key word of the assignee's name. All the reference code can be found in https://sites.google.com/ site/patentdataproject/Home/posts/ namestandardizationroutinesuploaded .

Table A2: Assignee's Location

Country	Patent Num	Percent
USA	3,032,312	50.79%
Japan	1,169,335	19.59%
Germany	264,046	4.42%
UK	154,575	2.59%
South Korea	154,477	2.59%
France	154,025	2.58%
Missing	130,872	2.19%
Taiwan, China	115,846	1.94%
Canada	92,660	1.55%
China	88,738	1.49%
Others	613,103	10.27%

Then We redefine the assignee's type as following steps, where the original assignee's type in Patentview database is the foundation:

- Step 1. We correct the assignee's type which should be the university or research institutions but the other types in Patentview's classification. By the NBER assignee's type identification method, the keywords in assignee's name, like "UNIV", "RESEARCH COUNCIL", "ACAD", "COLLEGE" and etc., give clues to identify the types of university or institutions. As for the nationality of the university and institution, We use the location to identify it.

- Step 2. Apparent inconsistency exists between location and assignee's type in Patentview database. (1) The assignee type indicates that it is domestic (US company or corporation, government and etc.), however, the location of assignee is not "US" (e.g. Signetics Corporation is a firm starting up in US, but has plant in South Korea, which was merged by Philips in 1975; assignee type: US Company or Corporation; assignee location: South Korea). (2) The assignee type indicates that it is foreign, but the location of assignee is "US" (e.g. Robert Bosch Corporation is a multinational firm of which head locates in Germany. It has plant in USA; Assignee type: Foreign Company or Corporation; assignee location: United State). The numbers of patent of those kind of inconsistency can be found in Table A3.

Table A3: Inconsistency between Assignee Type and Location

	Location≠US	Location=US	Location=Missing
Non-Domestic Assignee	2,554,571	305,301	76,412
	(42.79%)	(5.11%)	(1.28%)
Domestic Assignee	251,526	2,723,405	54,451
	(4.21%)	(45.62%)	(0.91%)
Missing type assignee	708	3,606	9
	(0.01%)	(0.06%)	(0%)

Actually, the former(red in Table) is probably the patent invented by American multinational firm's oversea subsidiary, and the later is possibly the patent invented by the Foreign multinational firm's American subsidiary. In these cases, the Patentview classification of assignee type is reasonable.

However, one caveat is that for the red part, there is another possibility that the firm is a FDI firm for USA. E.g. HISENSE USA CORPORATION has registered 255 patents in USPTO, it is identified as a domestic firm in Patentview database, but the location of assignee is "CN". At this circumstance, the location is more reliable to tell whether the firm is domestic or not. But Luckily, there are only two Chinese firm correlated with this errors. The other is Huaya Microelectronics, Ltd. with 22 registered patents.

Furtherly, we can restrict American domestic patents to the patents of which the assignee is domestic and the location is USA, and take this as a robust check.

B.2 Text Data Processing Method

The Chinese patent data contains the filling made to CNIPA from both Chinese applicants and foreigners. In this paper, we only select the fillings made by Chinese, which refers to the Chinese citizens, and the organizations, research institutes and firms registered in mainland China.

Since we focus on the new knowledge, we need to find a proper way to define the new knowledge embodied in the patents. In Section 4.2, we use the new words that appeared in the abstract of patent fillings to represent the new knowledge in the patents. For example, if a word *xyz* has never appeared in the abstract of patent fillings before, then we define this word *xyz* as the new word. We conduct the text analysis of all the patent fillings in the Chinese and US patent data. Since CNIPA provides the English translation of the abstract of each patents, we can directly do the comparison between Chinese patents and their US counterparts.

In order to identify the new words in the patent abstract, we need to tokenize the sentences to words and phrases.

Phrases. In this paper, we use N-gram method to get the phrases. We first calculate the number of all the n-gram of each patent, that is, the combination of the consecutive n words appearing in the text. n is set as 2 and 3 separately. Next, we filter the n-grams by dropping the phrases containing the number, date, symbol, punctuation, wh-word, and step words. Finally, after the case normalization and lemmatization of the noun., we keep the phrases which appear at least three times in a patent.

Words. After we get the phrases of 2-gram and 3-gram of each patent, we collect the word of each patent. In this stage, we use the text with the 2- and 3-grams removed. Again, we tokenize the text and do the case normalization. The number, date, symbol, punctuation, wh-word, and step word are removed and we lemmatize the noun. and verb., eg. plural to singular. We also tag the words (classify words by part-of-speech) and drop words with bad tag. For the tag of words, we use the definition from The Penn Treebank Project. The bad tags are listed as following: adjective, including comparative and superlative; adverb, including comparative and superlative; verb, including base form, past tense, gerund or present participle, past participle, non-3rd person singular present, and 3rd person singular present; wh-determiner; wh-pronoun; Possessive wh-pronoun; wh-adverb.

After those procedures, we generate a word/phrase list of each patent. Based on these word/phrase list, we construct the yearly new word list of the Chinese and US patents separately. We also construct the cross-country citation network based on the new word backward citation.

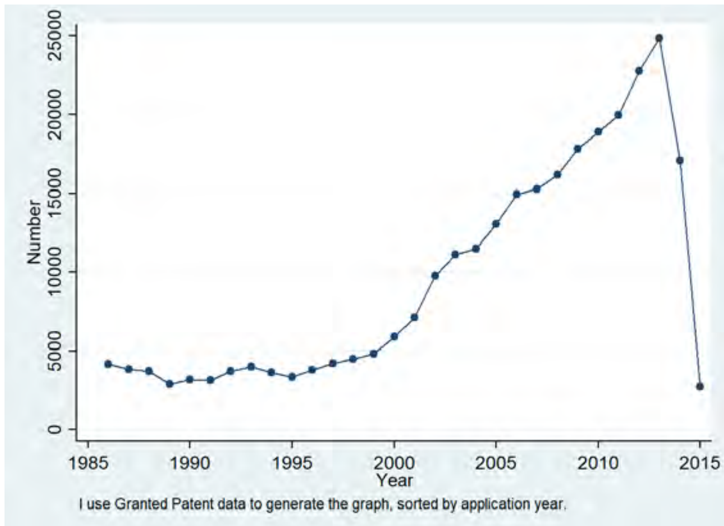
B.3 Robustness Check in Text Analysis

It is natural to ask whether the English translation of Chinese patent abstract is reliable and what if its quality if low? In order to address this concern, we conduct text analysis based on Chinese abstract and again use the above procedure to process Chinese text.

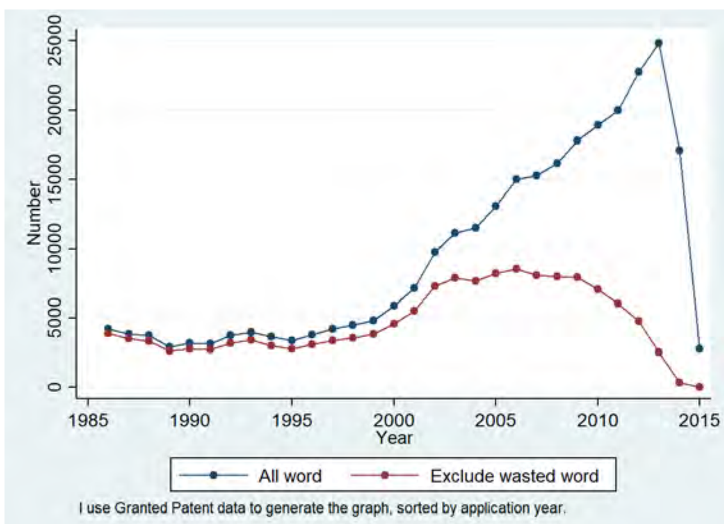
The upper panel in Figure A4 shows the results. It plots the number of new words in granted Chinese invention patent with Chinese text, sorted by patent application year. The similar pattern assures us that language will not mislead the analysis.

The lower panel in Figure A4 compares the number of new word between all word and excluding wasted word for Chinese granted invention patent (in Chinese text). The number of non-wasted new words was increasing and the trend became flat later. Similarly, since 2008, the number began to decrease.

Figure A4: Number of New Word in Chinese Patent



(a) Number of New Word in Chinese Patent(Chinese Text)



(b) Number of New and Non-wasted Word in Chinese Patent(Chinese Text)

B.4 Standardized Forward Citation Number Calculation Method

The followings are calculation steps in details:

- Step 1 Within home country (applied in T year): We define forward citation number gap between non-exported patents and exported patents as:

$$Gap_1 = \frac{FWCT(exported, T)}{FWCT(non-exported, T)} \tag{10}$$

$FWCT(exported, T)$ is the forward citation number of exported patents applied in T year in home country; $FWCT(non-exported, T)$ is the forward citation number of non-exported patents applied in T year in home country. One thing to be noted is that all the comparisons between different types of patents' forward citation are taken within every technology class, and we take the average of gaps across every technology class. What's more, to guarantee the gap of forward citation number comparable across the year, we restrict the forward citation to the forward citation received within 4 years after the patent being granted.

- Step 2 Within foreigner country: We taking the bridge that US patents exported to China as an example. In this example, China is the foreigner country for US exported patent. One problem is that how we deal with the application lag problem. The patent owned by US assignee applied in T-year in US may be applied in China several years later. Here, we first calculate the forward citation gap between the US exported patents which are T-year applied in US and t-year applied in China, and CN t-year applied domestic patents. Then, we take the average of this gap across t years weighted by the share of T-year applied in US patent exported to China in t year.

$$Gap_2 = \sum_t \frac{Fwct(P_{us,T;cn,t})}{Fwct(P_{cn,t})} \cdot \frac{Num(P_{us,T;cn,t})}{\sum_t Num(P_{us,T;cn,t})} \tag{11}$$

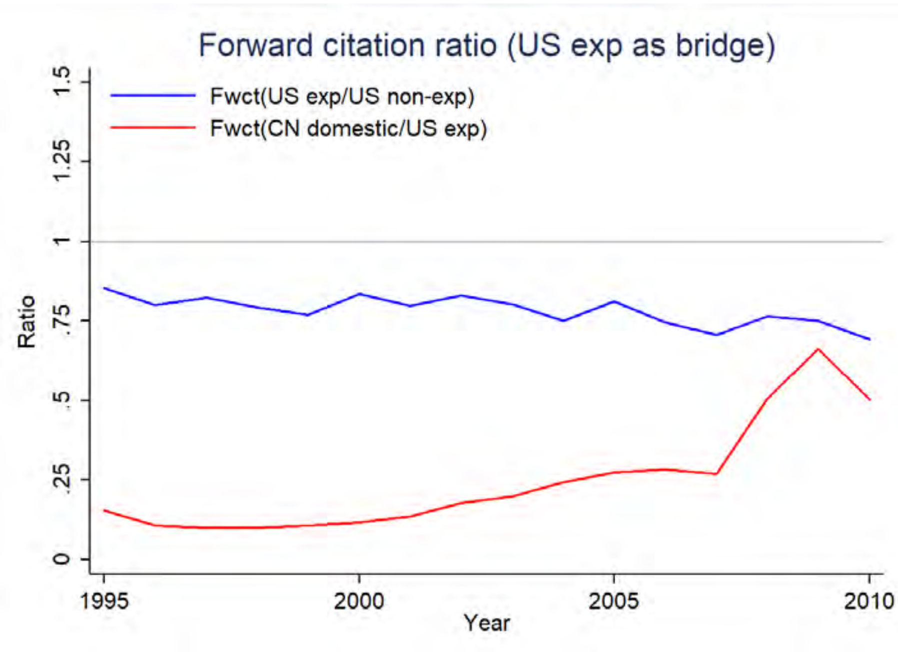
- Step 3 Normalize bridge patent's forward citation number=1; and take it as its' quality measurement

  - E.g. taking the bridge that US patents exported to China patents as an example

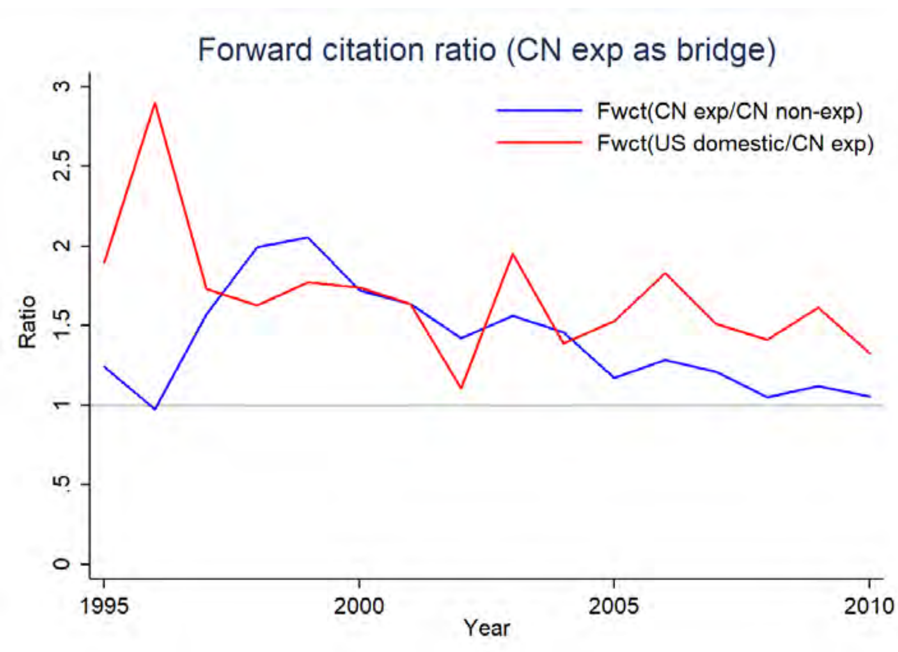
  - Quality of US non-exported patents= $Q(US\ non-exported, T) = \frac{1}{Gap1}$

- Quality of (US domestic patents = non-exported+exported) = $Q(USdomestic, T) = \frac{Num(USexported)}{Num(USdomestic)} \cdot 1 + \frac{Num(USexported)}{Num(USdomestic)} \cdot Q(US\ non-exported, T)$
- Quality of CN domestic patents, $Q(CNdomestic, T) = \frac{1}{Gap2}$
- Quality gap between patents invented by U.S. and China is $Qgap_{us,cn,T} = \frac{Q(USdomestic,T)}{Q(CNdomestic,T)}$

Figure A5: Standardized Forward Citation Gap Decomposition



(a) US exp to CN



(b) CN exp to US