Codifying Prior Art and Patenting: A Natural Experiment of Herbal Patent Prior Art Adoption at the USPTO and EPO

Prithwiraj (Raj) Choudhury and Tarun Khanna[1]
June 2014

The innovation literature in economics has long studied how changes in the patent system affect the level and direction of patenting. One of the propositions of this literature is that improving the patent examination process and the availability of relevant prior art will lead to better patenting outcomes. In this paper, we create a unique dataset of Chinese and Indian herbal patents filed on the USPTO and EPO from 1977 to 2010 and exploit a natural experiment where the USPTO and EPO adopted a codified database of traditional herbal prior art at different points in time. This database (the 'Traditional Knowledge Depository Library' or TKDL) was created by Indian state-owned R&D labs and provided the USPTO and EPO patent examiners with codified, searchable prior art on herbal formulations based on a translation of ancient Indian medicinal texts. We establish that the time lag of the USPTO adopting TKDL compared to the EPO was related to idiosyncratic differences in how the agreements were structured and negotiated, not differences in policy toward herbal patents at the USPTO and EPO. We find that the adoption of TKDL affects the level of herbal patent filing and grants. It also shifts the direction of patenting from pure herbal remedies that are similar to prior art available in the ancient texts to applications involving herbals and synthetics, which are more distant from the prior art and arguably less contestable. Further, we study the ethnic origins of the inventors of herbal patents filed on the USPTO. For this analysis, we use ethnic name matching for all patentees of herbal patents. We also exploit an exogenous reduction in employment (H1B) visa quotas and find that herbal patents filed by Western firms based in the U.S. are driven by scientists belonging to the Chinese and Indian diaspora.

[1] Harvard Business School; corresponding author: Prithwiraj (Raj) Choudhury—pchoudhury@hbs.edu.
The authors thank seminar participants at the Harvard Business School and International Economic Association conference for comments on a prior draft.

1. Introduction

The innovation literature in economics has a rich tradition of using patents as a measure of innovative activity. To quote Lanjouw and Schankerman (2001: 130), 'innovation is at the heart of productivity growth, and granting inventors property rights through the patent system is the most widespread policy to encourage innovation.'

Over the past several decades, researchers have studied the effect of modifying patent systems and changing patent laws on the *level* of innovations being patented (Sakakibara and Branstetter, 2001; Lerner, 2002; Branstetter *et al*., 2006; Lerner, 2009). There is also related enquiry on how changing the patent system affects the *direction* of innovative activity (Moser, 2003) and the *antecedents* of innovation (Moser, 2007). This paper adds to this stream of research in economics and asks the following question: *for a newly emerging technology, how does the codification of prior art affect the level of patenting and direction of innovation?* In other words, we study how better ex ante information provisioning affects patenting outcomes.

To explain why codifying prior art might have an effect on the level of patenting and the direction of innovation, at the onset we define 'codifying prior art.' As the innovation literature has documented, an invention might be patented if the claims contained in the patent meet the twin criteria of novelty and nonobviousness. To ascertain the validity of the claims contained in a patent, the claims are to be examined against prior art contained in prior patents and non-patent sources such as publications, books, databases, etc. The literature has also documented that prior art is mostly embedded in non-patent sources such as publications, books, and software for emerging technological areas such as biotechnology, software, and business method patents (Sampat, 2004; Hall and MacGarvie, 2010). Codifying prior art is a process of transcribing such prior art and making such information available to patent examiners, legal experts, and other

actors. In summary, codifying prior art allows patent examiners, legal experts, and other actors in the patenting system easier access to prior art previously contained in difficult to access and difficult to search non-patent sources.

We posit that codifying prior art should affect the *level* of patent filings and patent grants, as it increases the efficiency of prior art examination, reduces possible information asymmetry between patent examiners and inventors, and weeds out 'bad patents.' Codifying prior art into a searchable database increases the chances that patent examiners access appropriate prior art to examine the claims of a patent application. This prior art might have been earlier embedded in non-patent sources that might be difficult to access during the patent examination process. Anticipating this increased scrutiny of prior art and a higher likelihood of rejection, *ex ante* inventors and firms should be discouraged to file patents that violate prior art considerations. Additionally, codifying prior art should also impact the *direction* of innovation, as post codification, inventors should be encouraged to conduct additional R&D and file only patents that comprise novel claims.

A key actor in the system of patenting innovation is the patent and trademark office (PTO) that examines patent applications filed by inventors and either grants patents or rejects/modifies the claims of filed patents. The recent literature in economics and law has delved deep into the internal workings of patent offices around the world, especially studying the efficiency of the prior art examination process (Cockburn, Kortum, and Stern, 2003; Alcacer and Gittelman, 2004; Sampat, 2004; Lemley and Sampat, 2012; etc.). The literature has documented the inefficiencies of the patent examination process at the PTOs and has pointed out that these inefficiencies lead to the granting of 'bad patents' with significant legal and social costs. To quote Lemley (2001: 1-2), 'the PTO has come under attack of late for failing to do a serious job of

examining patents, thus allowing bad patents to slip through the system...several solutions have been proposed, but the common thread among them seems intuitively obvious: the PTO should do a more careful job of reviewing patent applications.' Sampat (2010) alludes to constraints around incentives, resources and capabilities of patent examiners in examining prior art. To quote the author, 'to assess whether a patent application meets these criteria, a patent examiner compares the claimed invention with previously disclosed information, or prior art, found in references to patents and printed publications. There is growing concern that resource-constrained examiners face difficulties in identifying relevant prior art, thereby contributing to the issuance of patents of questionable validity' (Sampat, 2010: 399).

In this paper, we create a unique dataset of Chinese and Indian herbal patents filed on the United States Patent and Trademark Office (USPTO) and the European Patent Office (EPO) from 1977 to 2010. We also exploit a natural experiment where the USPTO and EPO adopted a codified database of traditional herbal prior art at different points in time. This database (the 'Traditional Knowledge Depository Library' or 'TKDL') was created by Indian state-owned R&D labs and provided the EPO and USPTO patent examiners with codified, searchable prior art on herbal formulations based on a translation of ancient Indian medicinal texts. We establish that the time lag of the USPTO in adopting the TKDL agreement compared to the EPO was related to idiosyncratic differences in how the agreements were structured and negotiated, not differences in policy toward herbal patents at the EPO and USPTO. To confirm this, we conducted detailed interviews with two individuals who played critical roles in creating TKDL and rolling it out to patent offices around the world. We interviewed Dominic Keating, the director of the IP Attaché Program at the USPTO, who played a key role in negotiating the TKDL Access Agreement between the USPTO and the Indian public R&D labs. We also

interviewed Dr. V.K Gupta, the architect and director of TKDL. We followed up these interviews with qualitative analysis using secondary sources, and our interviews and analysis generated several insights that validate the natural experiment. As an example, one of the reasons the USPTO adopted the TKDL later than the EPO was due to the fact that the U.S. Access Agreement signing was postponed to take place during the state visit of the Indian prime minister to Washington D.C. While the USPTO Access Agreement was officially signed by the head of the Indian R&D labs during the prime minister's visit, the EPO Access Agreement signing was a much less publicized affair, with Dr. Gupta himself signing the agreement during one of his visits to the EPO.

We find that the adoption of TKDL affects the level of herbal patent filing and patent grants. It also shifts the direction of patenting from pure herbal remedies that are similar to prior art embedded in the ancient texts to applications involving herbals and synthetics, which are more distant from the prior art. Further, we study the ethnic origins of the inventors of herbal patents filed on the USPTO. For this analysis, we use ethnic name matching for all inventors of herbal patents. We also exploit an exogenous reduction in employment (H1B) visa quotas and find that herbal patents filed by Western firms based in the U.S. are driven by scientists belonging to the Chinese and Indian diaspora.

In addition to informing the literature on innovation and the literature on diaspora, our results contribute to the literature on patent litigation. Lemley (2001) estimates the cumulative legal costs of granting bad patents as around $2.4 billion. Sampat (2004) points out that new technological fields as well as old fields that are newly patent eligible, like software, might particularly benefit from better quality prior art. The author quotes former USPTO Commissioner Q. Todd Dickinson in suggesting that 'rapid progress in emerging technologies

continues to challenge the USPTO's ability to access the most current information that demonstrates the state of that art' (USPTO, 1999a: 3). One reason for this is that much of the prior art in such fields tends to be in the non-patent literature, which, as suggested earlier, tends to be more difficult to search (Merges, 1999). Indeed, the USPTO may not even have access to the relevant non-patent databases in some new fields.

The rest of the paper is structured as follows: Section 2 outlines the theory; Section 3 describes the empirical setting and the natural experiment while Section 4 outlines the empirical questions; Section 5 describes the data and variables; Section 6 presents the results and Section 7 concludes. References, tables and figures are at the end.

2. Theory: Codifying Prior Art and Effects on Levels and Direction of Patenting

2.1. Patent Examination Process

In order for an invention to be patentable, a patent claim must be shown to be both novel and nonobvious. As Sampat (2004) outlines, under the novelty bar, an invention cannot be patented if it was previously known or used. Under the nonobviousness bar, an invention cannot be patented if 'the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which the subject matter pertains' (35 USC ξ 103a).

As Alcacer and Gittelman (2004: 6) state, 'the core value of a patent is expressed in its claims...claims cover intellectual property that is not already anticipated by existing patents or public knowledge.' The authors also describe how, at a practical level, patents contain 'prior art,' comprising both prior patented and non-patented information such that the claims of the patent are viewed as novel and nonobvious. To quote the authors, 'claims and prior art operate together to describe the invention, show its novelty over existing knowledge, and delineate the scope and

the corresponding strength of the intellectual property covered by the claims' (Alcacer and Gittelman, 2004: 6-7).

Given the importance of recognizing prior art in framing the claims of a patent, it is intuitive that the efficiency of the patent-granting process hinges on the quality of prior art examination at patent offices around the world. Cockburn, Kortum, and Stern (2003) conducted a comprehensive study of the prior art examination process within one of the most important patent granting systems in the world—the USPTO. For an individual patent, the prior art search process typically begins with a review of prior U.S. patents in relevant technology classes and subclasses, either through computerized tools or by hand examination of hard copy stacks of issued patents. The search then proceeds to a word search of foreign patent documents, scientific and technical journals, and other databases. The USPTO's Scientific and Technical Information Center maintains an extensive collection of reference materials.

Patent examiners also rely, in part, on applicant disclosure of the prior art submitted with the patent application, as part of the information disclosure statements. As Cotropia, Lemley, and Sampat (2012) describe, in the United States, patent applicants have a 'duty of candor' to disclose any prior art material to patentability of an invention. However, Thomas (2001) suggests that applicants also face disincentives to search for relevant prior art. Applicants might be weary of searching for and discovering prior art in the fear that failure to disclose may later be deemed inequitable conduct in a court of law. This implies that patent examiners play a critical role in discovering prior art related to the claims of a patent.

There exists a recent stream of empirical papers that studies the relative share of examiners and applicants in citing prior patents. Alcacer and Gittelman (2004) use a dataset of citing-cited pairs of patents of 1,500 citing patents generated from a large, random sample of

patents issued over the period 2001-03, and they find that compared to the applicants, examiners are 87 percent more likely to cite patents that are in other countries than the inventors' own. Sampat (2004) uses a dataset of 502,687 utility patents issued by the USPTO from January 2001 to December 2003 and finds that at the patent level, 62 percent of the average share of citations to U.S. patents are inserted by examiners. In subsequent work, Alcacer, Gittelman, and Sampat (2009) report similar results. They examine prior art citations listed in all U.S. patents granted in 2001-03 and find that examiners add 63 percent of citations on the average patent and all citations on 40 percent of patents granted.[2] Sampat (2010) reports that there is self-sorting within technology fields and even within firms with regard to the share of applicant citations; applicants contribute more prior art for their more important inventions.[3] In a related paper, Cotropia, Lemley, and Sampat (2012) find that patent examiners did not use applicant-submitted prior art in the rejections that narrowed down the claims before the patents were issued. Instead, they relied almost exclusively on prior art that they found themselves. The authors also find that examiners do not even use applicants' late-submitted art (i.e., prior art submitted by applicants later in the process of examination) in making these rejection decisions. Cotropia *et al*. (2012) hint at patent examiner myopia in interpreting these results. In other words, patent examiners tend to focus on references that they themselves identify.

However, examiners account for a much smaller share of references to prior art embedded in the non-patent literature and in non-U.S. patents. Sampat (2004) finds that references to non-U.S. patents and to non-patent literature are, respectively, 27 and 26

[2] In a recent paper, Hegde and Sampat (2009) report that examiner citations to a patent are stronger predictors than applicant citations of a widely used measure of patent value—whether or not the patent will be renewed.

[3] This result builds on prior work on strategies employed by firms in acquiring patents, including work by Hall and Ziedonis (2001) and Cohen, Nelson, and Walsh (2000). The prior literature suggests that firms have stronger incentives to ensure that their patents are defensible against validity charges based on missed art in discrete product fields (such as pharmaceuticals) where patents are important for appropriating rents from R&D. In contrast, in complex product industries where many patents cover a given product, firms might have less incentives to conduct a thorough search of prior art.

percentage points less likely to be inserted by examiners rather than by applicants, compared to references to U.S. patents. Moreover, in 69 percent of the patents citing non-patent prior art, all the citations to the non-patent prior art came from applicants.

Starting with Cockburn, Kortum, and Stern (2003), the literature has also documented significant heterogeneity in the patent examination process both at the level of Art Units and at the level of individual examiners.[4] As an example, the authors document that patent examiners examining biotechnology patents rely extensively on non-patent prior art, such as scientific articles indexed by Medline. Lemley and Sampat (2012) show that more experienced examiners cite less prior art, are more likely to grant patents, and are more likely to grant patents without any rejections.

2.2. Constraints of the Patent Examination Process

The literature on prior art examination has outlined several constraints faced by patent examiners in the prior art examination process. Lemley and Sampat (2012) describe the time constraints faced by patent examiners in searching for prior art. They cite Allison and Lemley (2000) and Lemley (2001) to describe how the patent examination process for an individual patent might take three to four years, but the examiner spends an average of only 18 hours over those years working on any given application. Merrill *et al*. (2004) also document that examiner time constraints have increased over the years as the number of patent applications has outpaced the growth in the number of examiners.

Starting with Cockburn, Kortum, and Stern (2003) and Jaffe and Lerner (2004), the literature has also outlined how incentives of patent examiners might be aligned against conducting a comprehensive search for prior art. As Lemley and Sampat (2012) describe, the

[4] As Cockburn *et al*. (2003) document, patent applications are sorted and allocated into one of 235 Art Units, a group of examiners who examine closely related technologies and constitute an administrative unit.

promotion and bonus decisions of examiners is tied to the number of counts they accumulate. The counts are determined by the number of first office actions or by disposal of cases, which occurs when an application is either granted or abandoned.[5] The authors also state that this system of incentives does not reward searching for prior art.

The literature, starting with Lemley (2001), has also outlined the constraints of patent examiners in accessing relevant prior art embedded in non-patent sources, such as publications, books, and software. Lemley and Sampat (2012) also cite Thomas (2001) to argue that searching for prior art is more difficult for non-patent prior art, given that non-patent art is not readily accessible, is not conveniently identified and classified, and is not printed in a common format.

Hall and MacGarvie (2010) provide interesting evidence of this from the field of software patent examination. To quote the authors, 'when the USPTO began to handle large numbers of patents in this area in 1994-95, they lacked adequate databases with software prior art' (Hall and MacGarvie, 2010: 996). The authors also document an interesting verbal exchange between a patent examiner and the editor of the *Aharonian's Patent Newsletter* concerning constraints in accessing non-patent prior art.

The authors first quote the patent examiner as saying that 'the problem with obviousness is evidence…Also, a problem with ordering non-patent publications or translations of foreign documents is the time it takes to process these requests. An examiner cannot simply call a company who is making, or is believed to have made, a product which appears to infringe on a claim. At best, the examiner could ask a librarian at our library to call a company to request info,

[5] Cockburn *et al*. (2003) describe the process of issuing first office action letters as part of the patent examination process. After reviewing a patent application, examiners compose a 'first action' letter either accepting or rejecting their claims. Applicants then receive a fixed time to respond by amending the claims and/or by supplying additional evidence.

but again that takes time. With 10 hours to do a case, movement is paramount' (Hall and MacGarvie, 2010: 996).

The authors then highlight the issue of resource constraints in accessing non-patent prior art such as books and publications. They quote Aharonian as saying, 'additionally for subjects like software, the cost of purchasing copies of technical papers would exceed the application fee, so I doubt many examiners would get the authority to spend such sums. Since for most software patent applications, the most relevant prior art is non-patent materials, between the statistics I cited on citing prior non-patent prior art (an average of two out of about 30) plus the above and below comments, one could make a good case that it is impossible for the PTO to conduct adequate novelty analyses' (Hall and MacGarvie, 2010: 997).

2.3. Codifying Prior Art: Effect on Patent Filings and Patent Grants

Given these constraints of the patent examination process, codifying prior art embedded in books, publications, and other non-patent sources should impact the level of both patent filings and patent grants. Codifying prior art in a format easily accessible and searchable by patent examiners should *ex ante* increase the probability that patent examiners identify relevant prior art and should additionally reduce the time needed to search for prior art. This, in turn, should impact the level of patent filings, as patent examiners are now better equipped to examine prior art and weed out 'bad patents.' Anticipating this, *ex ante*, inventors and firms should be discouraged to file such patents in the first place and incur patent filing costs. Codifying prior art should also impact the level of patent grants, as examiners are now better equipped to reject bad patents in case they are filed.

2.4. Codifying Prior Art: Effect on Direction of Innovation

The broader literature on innovation in economics has also long studied the effect of changing patent laws and the overall patent system on the *level* and *direction* of innovation within a country, and the evidence is quite mixed.

Early researchers such as Nordhaus (1969) and Gilbert and Shapiro (1990) predicted that stronger patent laws raise the number of innovations within a country. However, several subsequent studies in this literature provide evidence to the contrary. In other words, the more recent literature has established that patent system reform has no effect on the *level* of innovation, especially by domestic inventors.

Lerner (2002) studies 177 events of patent reforms in 51 countries over a 150-year period and finds that residential patent filings did not react positively to domestic patent reform. In a related study that takes a historical perspective of patent system reform and innovation, Moser (2003) constructs a dataset of innovations from the catalogs of two nineteenth-century world fairs, and this dataset includes innovations that were not patented as well those that were. The dataset also includes innovations from countries both with and without patent laws. In summary, the author finds no evidence that patent laws increased the level of innovative activity. Sakakibara and Branstetter (2001) study the effect of expanding patent scope in Japan in 1988 and find no evidence of an increase in either R&D spending or innovative output that could plausibly be attributed to patent reform. Branstetter *et al*. (2006) examine how technology transfer within U.S. multinational firms changes in response to a series of patent system reforms undertaken by 16 countries from 1982 to 1999. They find that royalty payments for technology transferred to affiliates increase at the time of reforms, as do affiliate R&D expenditures and total levels of foreign patent applications. However, they also report that relative to the pre-reform

period, patenting grows for nonresidents after reform, but remains flat for domestic residents. This indicates that patent system reform had no reported effect on domestic innovation.

However, the recent empirical literature in economics on innovation has also looked at how patent system reform influences the *direction* of innovation within a country, and here the evidence is more positive. Moser (2003) finds strong evidence that reforming patent systems influenced the distribution of innovative activities across industries. Inventors in countries without patent laws focused their innovative activity in industries where secrecy was effective relative to patents. The author interprets her results as evidence that more effective patent laws might affect the *direction of innovative activity*, rather than the *level of patenting*. There is also a related literature in innovation that studies how changing the patent system affects the *antecedents* of innovation. Moser (2007) argues that the ability to keep innovations secret may be a key determinant of patenting. In other words, inventors' propensity to patent depends on the effectiveness of secrecy, hence patenting is likely to increase in response to fundamental advances in science and engineering. The author tests this proposition in the context of the chemical industry and the introduction of the periodic table in 1869. She finds that in 1851, none out of 32 chemical innovations in the U.S. had been patented; however by 1893, 16 percent of U.S. chemicals were protected by patents.

Our study adds to this stream of research in economics and asks how the codification of prior art affects the *direction* of innovation. In so doing, we study an important antecedent of innovation that has not been studied by prior researchers—codifying prior art. We posit that the codification of prior art into a codified, searchable database for patent examiners should shift the direction of patenting. Instead of inventors filing patents with claims that might be closer to prior art, post codification, inventors should file patents with claims that are more distant from prior

art. This should be particularly true for a newly emerging technology, where the prior art is much more likely to be embedded in non-patent sources such as publications, books, etc. The reasoning for this argument is the following: post codification of prior art, inventors are aware that patent examiners are likely to reject patents with claims that are similar to prior art. Knowing this, *ex ante*, they are less likely to submit such patents for examination. Instead, post codification, they are more likely to submit patents with claims distant from prior art. However, in framing this argument, we posit that prior art codification has an effect only on the direction of patenting; we do not wish to suggest any normative implications. As an example, we do not suggest that patents with claims more distant from prior art are 'more innovative' patents.

2.5. Diaspora and Innovation

We also study the ethnic origins of the inventors of herbal patents filed on the USPTO and build on the emerging literature on the role of diaspora in the propagation of intellectual property (IP) across global borders.

The early literature on diaspora and innovation had viewed the migration of the diaspora as 'brain drain' and viewed the effect as detrimental to those left behind in the home country (Grubel and Scott, 1966; Johnson, 1967; Bhagwati and Hamada, 1974).[6] More recent studies, however, have documented a more positive role played by the diaspora with regard to their country of origin (i.e., 'home country'). Remittances from migrants of developing countries totaled $351 billion and dwarfed other cross-capital flows, such as international lending, portfolio investments, and government-to-government aid (World Bank, 2011).[7] In addition to

[6] A number of studies documenting the magnitude of the problem of brain drain has shown high and increasing migration from poor-to-rich country for workers who received tertiary education in their home countries (Docquirer and Marfouk, 2005; Dumont and Lemaitre, 2005). Emigration of skilled labor has been found to be welfare reducing if the size of migrants is substantial (Bhagwati and Hamada, 1974).

[7] Part of the academic debate is now focused on how emerging market governments can tap into and purposefully channel these financial resources (Wilmaladharma *et al*., 2004). Vaaler (2011) theorizes and provides evidence that

remittances, the diaspora also affects the propagation of IP from the Western countries back to emerging markets. Kerr (2008) documents the role of the diaspora in international technology diffusion. Nanda and Khanna (2009) outline the role played by the diaspora in helping Indian entrepreneurs, especially individuals living outside the software hubs, secure access to business leads and finance. Saxenian (2005) uses the term 'brain circulation' to describe the process. As the author points out, the diaspora, aided by the lowered transaction costs associated with digitization, are transferring technical and institutional know-how between distant regional economies faster and more flexibly than most large corporations.

All of this previous research focuses on the effect of the diaspora from the point of view of their home country. However, in this paper, we explore whether the diaspora can act as an agent of knowledge transmission from the home country to the host country by transferring context-specific knowledge embedded in emerging markets to firms and institutions of the Western world. Here we are motivated by the prior literature on knowledge transfer and inventor mobility. The key insight is that mobility helps bridge the spatial and social distance between the distant inventors and leads to new knowledge creation by recombining knowledge that distant inventors might possess. There are several detailed insights from this literature. In introducing the concept of 'spatial myopia,' Levinthal and March (1993) outline how physical distance constrains the search for distant knowledge. As a solution, Rosenkopf and Almeida (2003) document that inventor mobility can serve as 'bridges to distant contexts.' Other papers that look at inventor mobility and knowledge creation include Jaffe, Trajtenberg, and Henderson (1993) and Saxenian (1994); an excellent summary of this literature is provided by Breschi and Lissoni (2001). We build on this literature and posit that given the tacitness and *ex ante* non-codifiability

---

emigrant remittances of money to developing countries substantially enhance home country venture funding, founding, and broader internationalization.

of knowledge of traditional herbal medication, scientists belonging to the diaspora can help transfer this knowledge from their countries of origin to the firms and research entities of the Western world. If this turns out to be true, in a way, the diaspora could be seen as actors involved in transferring contextual knowledge from their home countries to the host country.

3. Empirical Setting

3.1. Indian Traditional Medicinal Knowledge

India's diversified agro-climatic zones span from the Trans Himalayas to Coastal Andaman and Nicobar. India is one of the most bio-diverse countries of the world. India comprises 16 out of the 26 possible agro-climatic zones. These diversified agro-climatic zones have resulted in India accounting for 7 to 8 percent of the recorded species of the world, with only 2.4 percent of the world's land area. The Botanical Survey of India and the Zoological Survey of India have recorded more than 47,000 plant and 81,000 animal species. This biodiversity has led to a long-standing interest in traditional medication among ancient civilizations in India. This has been formalized in several traditional medicine systems, including the *Unani, Yoga, Ayurveda,* and *Siddha* systems. India officially recognizes a wide variety of plants and herbs for their medicinal values. As Dubey *et al*. (2004) note, more than 6,000 plants in India are used in traditional, folk, and herbal medicine. Three of the 10 most widely selling herbal medicines in the developed countries, namely preparations of Allium sativum, Aloe barbedensis, and Panax sp., have been long available in the traditional form in India. Starting in the 1990s, there has been an increase in the trend of Western firms filing patents related to medicinal herbs.

3.2. Western Entities Patenting Traditional Indian Medicine

In 2000, the state owned Council of Scientific & Industrial Research (CSIR) in India conducted a study and found that 80 percent of the 4,896 references of individual plant-based medicine patents in the USPTO that year related to just seven medicinal plants of Indian origin.[8] Three year later, there were almost 15,000 patent filings related to Indian herbs and medicinal plants across the USPTO, EPO, and other global patent locations. In the view of CSIR and other domestic Indian R&D entities, several of these patents encroach on prior art documented in ancient Indian medicinal texts. Several of the patents granted to Western entities fueled growth in the health foods and herbal medicine industries in the U.S. and other Western countries. Kamboj (2000) estimated the market size of health foods and herbal medicine in the U.S. to be around $4 billion in 1996.

The filing of Indian and Chinese herbal patents by Western entities is, arguably, 'IP stealing.' This is based on the provisions of the Trade Related Aspects of Intellectual Property (TRIPS) agreement signed in 1994. Article 27.1 of the TRIPS agreement laid forth three requirements for patentability: (1) the invention is new; (2) the invention involves an inventive step; and (3) the invention is capable of industrial application. Interpretation of these three requirements was left to domestic laws, which lead to different standards across the globe. Internationally, 'new' means never published, used, or known anywhere in the world (absolute novelty standards). Unlike other laws, 'new' in United States patent law means never published anywhere in the world or used/known in the United States one year before filing of the patent application (partial novelty standards).[9] Also, around the same time, many countries around the world joined a treaty called

[8] India's 42 state-owned national laboratories are organized under an autonomous umbrella organization, the CSIR. Collectively, it has about 12,500 scientific and technical employees.
[9] United States Code (USC) 35 § 102(b).

the Convention on Biological Diversity (CBD). Under this treaty, if a research team enters a CBD-contracting host country to gain access to biodiversity of the host country for their research, they have to sign an agreement with the host country detailing access terms, profit sharing, etc. The United States is the only country that has not signed the CBD, and it has arguably shielded its pharmaceutical company or researchers from any profit-sharing agreement.[10]

3.3. Litigation

The surge of Western patents based on Indian traditional knowledge evoked a strong reaction from the Indian scientific community. The CSIR was among the entities that initiated litigation against the Western firms securing such patents. In several instances, litigation was successful in revoking the patent. Two of the prominent examples of USPTO herbal patents being revoked include the cases of *turmeric* and *neem.*

Turmeric is one of the oldest spices used in Indian cooking for flavoring. It also has properties that make it an effective ingredient in medicines, cosmetics, and dyes. It has been used as medicine for healing wounds and rashes for centuries. In 1995, two expatriate Indians at the University of Mississippi Medical Center (Suman K. Das and Hari Har P. Cohly) were granted a U.S. patent (No. 5,401,504) on the use of turmeric to heal wounds. Soon after, the CSIR filed a lawsuit challenging the novelty of the patent. This claim was supported by the documentary evidence of traditional knowledge, including ancient Sanskrit texts and a paper published in 1953 in the *Journal of the Indian Medical Association*. This led to revocation of the patent in 1997.

[10] http://www.cbd.int/countries/.

The extract of neem (Azadirachta indica) has been used for hundreds of years by Indians against pests and fungal diseases that attack food crops. Its oil has been used to cure colds, the flu, malaria, skin diseases, etc. In 1994, the EPO granted a patent (EPO patent No. 436257) to a U.S.-based firm, W.R. Grace Company, and the U.S. Department of Agriculture for a method for controlling fungi on plants by the aid of hydrophobic neem oil. This evoked protests from the Indian farming community and led to a lawsuit filing in 1995. The Indian R&D entities submitted evidence that the fungicidal properties of extracts of neem seeds had been known and used for centuries in Indian agriculture to protect crops and, therefore, were non-patentable. The patent granted on neem was revoked by the EPO in May 2000.

3.4. The Traditional Knowledge Digital Library (TKDL)

Based on the mounting evidence of IP stealing or 'biopiracy' on the part of Europe and the United States regarding traditional medicines and recipes developed in India, particularly in the late 1990s, the Government of India moved to construct a digital library of traditional formulations. The most prominent examples of such 'bioprospecting' were the revocation of the turmeric and basmati patents granted by the USPTO in 1997, as well as the revocation of the neem patent granted by the EPO in 2000 (Hirwade, 2009). Consequently, in 1999, India's Department of Ayurveda, Yoga & Naturopathy, Unani, Siddha, and Homeopathy (AYUSH) established a task force for the creation of the TKDL. This task force consisted of Indian traditional medicine experts, patent examiners, IT practitioners, scientists, and technical officers. The task force found that internationally, more than 2,000 patents relating to traditional Indian medicine were being granted each year. Moreover there were severe constraints in examining the claims of these patents given the lack of formal academic publications describing herbal formulations. Section 102(a) of the U.S. Patent Act states that rejection of a patent requires

evidence in a printed publication describing the invention; mere proof of use in a foreign country is not good enough to reject the patent (Balasubramanian, 2012). With the task force's findings as the foundation, the TKDL project officially began in 2001 as a collaborative project between India's CSIR (of the Ministry of Science and Technology) and the Department of AYUSH (part of the Ministry of Health and Family Welfare).

The goal of TKDL was to codify knowledge available in the traditional Indian medicine literature on Ayurveda, Unani, and Siddha in a digitized format, translated into English, French, German, Japanese, and Spanish. After a period of five years, the project neared its completion and totaled more than 200,000 compositions, with approximately 82,900 Ayurveda formulations, 115,300 Unani formulations, and 12,950 Siddha formulations over the span of 34 million A4-size pages. In December 2005, these formulations were categorized through the creation of a new structured system developed for the TKDL, known as the Traditional Knowledge Resource Classification (TKRC). The TKRC consisted of approximately 25,000 subgroups and encompassed 'medicinal plants, minerals, animal resources, effects and diseases, methods of preparations, mode of administration, etc.

3.5. The Natural Experiment

We now outline a natural experiment where the introduction of TKDL at the USPTO and EPO was staggered based on idiosyncratic differences in how the Access Agreement for TKDL was structured and negotiated at the patent offices.

In order to prevent bioprospecting and intellectual property spin-offs resulting from the newfound availability of traditional Indian compositions due to the TKDL, the access and use of the TKDL is subject to a restrictive non-disclosure agreement, called the Access Agreement

(Oguamanam, 2008). After completion of a first edition of the TKDL categorized by the TKRC, the CSIR proceeded to send a formal Access Agreement to the EPO for approval in July 2006. After a number of meetings between the EPO and CSIR regarding clarifications to the agreement, as well as extensions and additions to the original library, the Access Agreement signing by the EPO took place in February 2009. *Ex post* analysis by the Government of India regarding pending patents at the time of the EPO agreement revealed that after sending relevant TKDL files, in two cases, the EPO set aside its intention to grant the patent, both within three weeks; in 59 additional cases, the applicant withdrew the patent application due to TKDL evidence; and in 18 other cases, the applicant modified the patent's claims as a consequence (Balasubramanian, 2012). Our detailed interviews and analysis of secondary data also reveal that the EPO was much quicker to react to the TKDL prior art compared to the USPTO.[11]

With requests for the Access Agreement expressed by Germany, Malaysia, and Thailand, the United States began taking steps to enjoin the agreement with the Indian government. After expressing an 'in principle' agreement in May, the USPTO formally signed the TKDL Access Agreement in November 2009 (Pappas and Byrne, 2009). Citing concerns of misappropriation of traditional knowledge by patent applicants in the West, as well as of removing knowledge from

[11] EPO application EP1520585 (concerning cancer treatment using Pistacia species) was one of the applications against which TKDL filed a third-party observation. In July 2009, the EPO asked the applicant to 'take position on TKDL-cited references.' In the same month, acting on the TKDL third-party submission, EPO 'set aside intention to grant patent' on an anti-vitiligo cream application (application EP1747786) and reopened the case for further examination. Almost immediately, a chain of application withdrawals began at the EPO. Unilever NV withdrew application EP1607006 for 'functional berry composition' in August 2009. Purimed Co. Ltd. (Seoul, Korea) withdrew application EP1781309 in October 2009. Similarly, applications EP2044850, EP1889638, EP1750809, EP1807098, and EP1967197 were all withdrawn in 2009 in the EPO. This series of withdrawals continued into 2010.

In contrast, there was a lag in the USPTO's reaction to the TKDL prior art. In June 2011, the Morinaga Milk Industry Co., Ltd. (based in Tokyo, Japan) withdrew 15 claims while retaining the application with only a single claim on their publication 20090093450, on the basis of the examiner's report. In November 2011, Laila Nutraceuticals (based in India) amended the claims of publication 20100203078 based on TKDL evidence. This was almost two years after the USPTO signed the TKDL Access Agreement.

the public domain, the USPTO urged other countries to follow suit and create facilitating mechanisms such as the TKDL. Consequently, Access Agreements were enacted with Canada and the United Kingdom in 2010 and with Australia and Japan in 2011.

In summary, the request for access to TKDL was made by the EPO in July 2005 and the same request was made by the USPTO in December 2006. The signing of the Access Agreements was concluded by the EPO in February 2009 and by the USPTO in November 2009. We conducted qualitative analyses and interviews to establish that the lag in the USPTO in requesting access and concluding the agreement was related to statutory reasons unrelated to USPTO/EPO policy toward herbal patents; in subsequent analyses, we exploit this time lag in identifying the effect of codifying prior art on the rate and nature of herbal patenting.

To seek out the roots of this almost one and a half year delay in request to access and nine-month delay in signing by the USPTO, we interviewed Dominic Keating, the director of the IP Attaché Program at the USPTO, who played a key role in negotiating the TKDL Access Agreement between the USPTO and CSIR. Keating noted that the primary cause for the EPO's earlier signing was that 'the EPO had different statutory requirements than the USPTO,' which led to a number of terms in the original Access Agreement being acceptable to the former but 'nonstarters' for the latter.

Keating went on to note three specific clauses that were problematic for the USPTO. First, there existed an indemnity clause in the original Agreement that would have required the USPTO to compensate for any lawsuits, which was quickly negotiated out of the Agreement. Second, initially only five users were allowed to access the TKDL database, which posed a problem because the USPTO did not know how many of its examiners, supervisors, and other employees

would need access at a particular moment. Consequently, in the final Agreement, the user access rights were altered to a maximum of 30 IP addresses at one time. Finally, the most problematic clause for the USPTO was that the office could not provide any aspects of the TKDL to the public. Since the biopiracy episodes that had triggered the TKDL effort, there had been widespread concern that providing TKDL to the public would allow for the exploitation of the knowledge to invent around the prior art. Unfortunately, this requirement did not fit with the U.S. patent system; under its statutory restrictions, the USPTO was mandated to provide documents detailing reasons for rejection to applicants. 'We can't just reject an application and give no reasons why it was rejected,' Keating noted. 'It doesn't work with the system. The USPTO has to provide all rejection documents to applicants, and this Agreement wouldn't let us do that.' As a result, several months were spent in negotiations on these points, during which time the 'USPTO's Access Agreement changed much more significantly than the EPO's.' Both parties were constrained by strictly statutory matters—the USPTO by its mandate to provide documentation for rejection and the CSIR by the necessity of preventing exploitation of the TKDL by the public—that were clearly independent of the innovation environment and economic conditions of the time. Ultimately, the Agreement was signed with the clause substantially altered to meet the USPTO's requirements, stating that 'the USPTO shall not disclose any information of the TKDL contents to third parties unless and only to the extent that it is necessary for patent search and examination...The USPTO, whenever required may give printouts from TKDL contents to patent applicants and/or their legal representatives only for the purpose of citations of prior art, and may publicly post the search results on the USPTO's Patent Application Information Retrieval system and on other search and examination results digital access systems' ('TKDL Access Agreement'). Thus, in this paper, we exploit these conditions as

a basis upon which to examine whether an improved *ex ante* information provision to patent examiners, in the form of the TKDL, results in changes in patenting behavior and the quality of patents issued.

We also interviewed Dr. V.K. Gupta, director of TKDL and senior advisor to the Indian state-owned labs that created TKDL. Dr. Gupta was, arguably, the most important architect of the TKDL database and had oversight of the process from the inception of the database until the signing of the Access Agreements with the USPTO and EPO. Dr. Gupta confirmed several insights from the prior interview with Keating, most importantly that the USPTO delay in signing the agreement could be attributed to its concern about not being able to make the prior art known to the public. However, to quote Dr. Gupta, 'the EPO wanted access to TKDL as early as 2005, as our efforts had become quite popular and quite well known by then at the EPO.' Dr. Gupta also outlined another reason why the signing of the USPTO Access Agreement was delayed—the formal signing was scheduled to match the schedule of the visiting Indian Prime Minister Dr. Manmohan Singh, who visited the United States in November 2009. The USPTO Access Agreement was signed by Dr. Samir Brahmachari, head of the Indian public R&D labs, in the presence of the Indian prime minister. In contrast, the EPO Access Agreement signing was a much less publicized event and Dr. Gupta signed the agreement in one of his visits to the EPO.

Table 1 outlines the milestones of TKDL adoption by the USPTO and EPO, and Table 2 outlines USPTO-EPO differences in the Access Agreement clauses.

4. Empirical Questions

In this paper we study the following three empirical questions:

4.1. Effect of TKDL on Levels of Herbal Patent Filing and Grants

Our first empirical question relates to whether or not the *level* of herbal patenting on the USPTO and EPO is affected by the TKDL project, which was formally codified in 2003 and formally introduced to the EPO in 2006. As Table 1 indicates, the first version of the compact disc (CD) was launched in October 2003, and the request for access was made by the EPO in July 2005. In December 2006, the USPTO made an official request to access the TKDL database.

To conduct this analysis, we perform a unit root test developed by Perron and Vogelsang (1992), treating all herbal patent applications on the USPTO as time-series data. This test allows for identification of structural changes in the time series and provides information on the time and the direction of the change. We use an innovational outlier model, which assumes that the impact of measures to counter IP theft is gradual rather than instantaneous (Perron, 1990).[12]

4.2. Effect of TKDL on Direction of Herbal Patents

Our second empirical question relates to whether or not there was a change in the *direction* of herbal patents in the post-TKDL years. We code herbal patents as comprising two types: (1) pure herbal patents that are exclusively based on herbal formulations; and (2) herbal patents combined with synthetic compounds. We then use a Logit regression to study whether the mix of herbal

[12] This is based on Clemente, Montanes, and Reyes' work (1998) based on finding multiple breaks in time-series data. The authors proposed tests based on both the 'additive outliers' (AO) model, which captures a sudden change in a series, or the 'innovational outliers' (IO) model, allowing for a gradual shift in the mean of the series. This taxonomy of structural breaks follows from Perron and Vogelsang's work (1992).

patents exhibits any change in the years following the TKDL database being created and being adopted by the patent offices.

4.3. Ethnicity of Herbal Patentees and Role of Diaspora

Our third and final empirical question is focused on identifying whether herbal patents filed on the USPTO by Western firms and research entities are driven by scientists of Indian and Chinese origin. To conduct this analysis, in the first step, we generate a 'control' patent for each herbal patent filed on the USPTO. We use the methodology first used by Jaffe *et al*. (1993) and several other researchers to generate control patents for each herbal patent. A control patent is selected for each herbal patent by matching the herbal patent on application year and same three-digit U.S. technology classification. Having generated the set of patents with the same application year and same original three-digit U.S. classification as the herbal patent, we identify the patent in the set that has the closest grant date to the herbal patent. In the next step, we code the ethnicities of patentees on each herbal patent and each control patent using a methodology described in the next section.

In the final step, we use an exogenous shift in H1B visa quotas for identification. The H1B visa is a temporary immigration category visa that allows U.S. employers to seek short-term help from skilled foreigners in 'specialty occupations.' This visa category applies to Indian and Chinese workers, but not to workers from a large number of Western European countries. We exploit an exogenous change in the total quota of H1B visas issued to identify whether scientists of Indian ethnic origin are driving herbal patenting. The total cap on H1B visas was exogenously reduced in 2004 from a total of 195,000 to 65,000. This, arguably, constrained the supply of Indian scientists migrating from India who could work at U.S. firms and research entities. We use this exogenous shock to conduct a difference-in-differences test using herbal and control

patents. If scientists migrating from India are disproportionately driving herbal patent filing, when compared to similar control patents, we should observe this effect to diminish post-2004. We also conduct a placebo outcome test using European ethnicities, where we expect to get no effect.

5. Data Collection and Coding of Variables

This section outlines how we created the unique dataset of Chinese and Indian herbal patents filed on the EPO and USPTO and how we coded the variables. Given that there is no easily identifiable way of extracting Chinese and Indian herbal patents from any of the EPO/USPTO-based datasets, we had to create our own dataset.

Our data sources comprise two widely used patent databases—Thomson Innovation and LexisNexis TotalPatents. We searched through every EPO/USPTO patent filed from 1977 to 2011 to construct our database, and we used both 'keyword search' and 'patent classification search' to identify the patents of interest.

5.1. Keyword Search

In the first step, we used keyword search in the title and abstracts of every USPTO patent to identify herbal patents of interest. There are 291 herbs documented in the Indian TKDL.[13] However, for computational ease, we started with keywords from the U.S. National Center for Complementary and Alternative Medicine (NCAM). In this database, 45 herbs are documented, along with their traditional and common use. In the first step, the keywords (herb name + traditional use/common use) were searched in the title/abstract/claim sections of every USPTO patent.

[13] The Traditional Knowledge Digital Library (TKDL) Web site is available at www.tkdl.res.in.

5.2. Classification Search

We then used two relevant patent classification systems—the International Patent Classification (IPC) and the U.S. Patent Classification (USPC) to search for herbal patents. The most relevant IPC was A61K36+ (with 207 subgroups) that was introduced in 2002 by a Committee of Experts at IPC Union for coding traditional medicine formulations.

Table 3 lists the specific IPC classifications relating to herbal medicine that were used in our classification based search. We also based our analysis on the USPTO classification system focused on herbal medicines. This was first outlined by Dominic Keating, first secretary for intellectual property at the U.S. Embassy, New Delhi/USPTO and patent attorney at USPTO, in a presentation made to the World Intellectual Property Organization (WIPO).[14] He revealed that major U.S. classifications 424/725 (having 55 subgroups) and 514/783 relate to herbal medicines. After mapping these U.S. classifications with IPC, we identified the relevant IPC classes. The same result was achieved using information from Georg Schiwy-Rausch's presentation at the EPO on Traditional Knowledge.[15] The corresponding U.S. Patent Classifications are outlined in Table 4.

We also used additional databases to augment our search and used the Traditional Chinese Medicine (TCM) database to extract more patents. All patents having priority as the United States were extracted from the TCM database.[16] This gave us 400 Chinese patents with U.S. priority. Once these 400 priority numbers were inserted in the Thomson Innovation database, 703 U.S. patent records relating to herbal medicines were extracted. Ninety-five percent of these

[14] Dominic Keating, *Defensive Protection of Traditional Knowledge at the United States Patent & Trademark Office* (available at www.wipo.int/edocs/mdocs/tk/en/.../wipo_tkdl_del_11_ref_t7_2.pdf).
[15] Georg Schiwy-Rausch, information manager, data acquisition at the EPO, has the presentation on *Traditional Knowledge at the EPO: Present & Future* (13-15 December 2006) available at http://pame.european-patent-office.org/pubs/hararepdf/tk_aripo_present_future.pdf.
[16] *http://chmp.cnipr.cn/englishversion/...advance.asp*.

patents records were already present in our initial sample of patent records, thus validating our earlier search. In the last step, we read every patent record to filter out irrelevant patents. We read the title and abstract of each of the patent records to complete the extraction process.

5.3. Coding the Independent Variables

After creating the database of herbal patents, we coded the independent variables. The first independent variable relates to the type of assignee for each herbal patent. We coded patents based on five possible assignee types: (1) individual inventors; (2) Fortune 1000 companies; (3) U.S. university, research organization, or government organization; (4) foreign university, research organization, or government organization; and (5) others. The following steps were used to categorize the assignee type:

1. *Fortune 1000 companies:* In the first step, a partial string matching of each word in the assignee name was carried out with words in the names of each Fortune 1000 company. In the next step, exact matches were created based on a manual parsing of the matches from the previous step.
2. *Individuals:* All patent records with no assignee were labeled as '*individual*.'
3. *University/research organization:* We carried out partial string matching of the assignee name with any of the following keywords: 'council,' 'board,' 'college,' 'center,' 'centre,' 'university,' 'research,' 'organization,' 'school,' '*laboratoire*,' and '*institut*.' After carefully examining the filtered results, university or research organizations were labeled as a U.S. or foreign university/research entity depending on the country mentioned under the assignee column.
4. *Others:* All other patents were labeled as having assignee type equal to 'others.' This mostly includes firms outside the Fortune 1000 list.

This analysis was conducted only for USPTO herbal patents. In the final dataset, all patents filed by individuals were removed, and this shrunk the database from 5,472 patents to 2,965 patents. We then categorized these 2,965 patents based on whether or not they were 'pure herbs' or whether or not they were a 'mixture of herbs and other synthetic compounds/drugs.' To do this categorization, we used the Derwent classification for each patent. After analyzing various mixes of Derwent classes in herbal medicine patent records, we concluded that herbal medicine patent records containing Derwent classes B05, B06, or B07 comprised a mix of herbal medicines and other synthetic compounds/drugs. In the absence of any of these classes, the composition is purely made up of herbal medicines.[17]

Next, we coded the ethnicity of herbal patent inventors. While the actual data on the ethnicity of the inventor cannot be collected, one can determine probable ethnicities of the inventor through his/her last name (Kerr and Lincoln, 2009). Bill Kerr helped us conduct this analysis, and we used two commercial ethnic names databases as well as name-matching algorithms customized for the USPTO data. When multiple inventors exist on a patent, we make individual ethnicity assignment for each inventor. Kerr (2008) provides further details on the matching process, lists frequent ethnic names, and provides multiple descriptive statistics and quality assurance exercises.

[17] In Derwent classification, the B class refers to 'pharmaceuticals.' Subclass B05 refers to 'other organics,' B06 to 'inorganics,' and B07 to 'general.' B04 refers to 'natural products and polymers,' which is also the most popular class in the herbal medicine patent group. B05, B06, and B07 are the only three classes in B (pharmaceuticals) that contain synthetic Western drugs. Thus, a presence of these three classes signifies a combination of synthetic compounds/drugs with herb. Fifty random abstracts of patent records having any of these three classes and 50 random abstracts of patent records with absence of all of these three classes were studied and this result was independently verified.

6. Results

6.1. Summary Trends

Figures 1-4 outline several summary trends of herbal patents. Figure 1 shows the trend of all herbal patents filed by and granted to firms, universities, and research entities on the USPTO and EPO. For our statistical analyses, we exclude herbal patents filed by individuals. Figure 2 outlines the mix of herbal patents on the USPTO in terms of pure herbal patents and patents that represent a combination of herbal and synthetic formulations. Figure 3 displays a broad classification of usage of herbal patents based on the Derwent class analysis. Figure 4 reveals the trend of herbal patenting by assignee type for USPTO patents. Summary statistics for USPTO patents are reported in Table 5.

6.2. Empirical Question 1: How Herbal Patenting Reacts to Litigation and TKDL

The breakpoints unit root test developed by Perron and Vogelsang (1992) was implemented using the clemio2 routine in STATA (Baum, 2005).

For patent filings at the EPO, the test detected two structural changes in years 1994 and 2003. There was an increase in herbal patent filing in 1994 and a decrease in herbal patent filing in 2003. For the USPTO, the test detected two structural changes in years 1993 and 1999. Both years marked an increase in herbal patent filing. The results indicate that the flattening out of herbal patent filing in 2003 is specific to the EPO and not the USPTO. To recap, the USPTO had a close to one and half years lag in requesting access to the TKDL Access Agreement versus the EPO. (Figure 5)

For patent grants, we conducted the analysis for the period post 1995. For the EPO, the test indicated breakpoints in patent filing years 1998 (negative breakpoint) and 2001 (negative

breakpoint). The variable 'year' here indicates 'year of filing the patent'; there is an estimated four or five year lag between a filing a patent and the patent getting granting. This indicates that at the EPO, there was a decline in patent grants around both 2003 and 2006. For the USPTO, the test indicated breakpoints for patent filing years 1998 (positive breakpoint) and 2001 (negative breakpoint). Given the four to five year lag between year of patent filing and year of patent grant, this indicates that at the USPTO, there was an increase in patent grants in 2003 and a decline in patent grants around 2006. (Figure 6)

In summary, the EPO exhibits a relatively greater and quicker decline on both patent filings and on patent grants, compared to the USPTO.

6.3. Empirical Question 2: How Mix of Herbal Patents Reacts to TKDL

Results reported in Tables 6 and 7 indicate that with the introduction of TKDL in 2003 and with TKDL being adopted by the EPO in 2006, there is a shift in the mix of herbal patenting. From pure herbals, the mix shifts toward herbals in combination with synthetics. Compared to the base year of 2003, in 2005 the odds ratio that patents are herbs mixed with synthetics is greater than 1 and the result is statistically significant. For year 2006, the odds of a patent being a mix of herbal and synthetic formulations is 96 percent higher compared to the base year of 2003.

6.4. Empirical Question 3: Ethnicity of Inventors for Herbal Patents

Results reported in Tables 8 and 9 indicate that herbal patents are disproportionately being driven by scientists of Indian origin. As the top panel of Table 9 indicates, the odds of a patentee being of Indian origin are 86 percent higher if the patent is based on herbal medicines compared to a similar control patent. We also find that this effect weakens for herbal patents post the reduction in the quota for H1B visas. This indicates that herbal patent filing is possibly driven by Indian

migrant scientists. The bottom panel of Table 9 indicates that we do not get any statistical significance for using a placebo outcome, i.e., scientists of European origin. Finally, Table 10 indicates that a diaspora Chinese scientist is likely to file a Chinese herbal patent.

7. Discussion

In summary, we have three major findings. We find that the codification of prior art affected the trend of herbal patent filing and herbal patent grants. We also find that starting in 2006, the mix of herbal patents shifts toward patents that are a combination of herbal and synthetic formulations. Finally, we find that herbal patent filing in the U.S. is driven by scientists of Indian ethnic origin—most likely by first-generation migrants.

Our results make a contribution to the research by economists and legal scholars on how to make patent examination more efficient. Prior literature has studied several solutions to increase the efficiency of the patent examination process at the PTO. Lemley (2001) discusses the implications of increasing the time that patent examiners spend in searching prior art from the average of 18 hours to 36 hours. Farrell and Merges (2004) discuss solutions related to administrative reform at the PTO, including improving incentives for patent examiners. The key insight here is that the patent examination process should not only reward examiners who grant patents, but also reward examiners who reject patents based on a careful examination of prior art. The issue of patent examiner incentives also has been studied extensively by Cockburn, Kortum, and Stern (2003). Cockburn *et al*. (2003) also discuss implications of processes such as patent examiner supervision, selection of examiners, and training. Yet another solution is 'sorting.' As Sampat (2004) points out, patent examiners should conduct a more thorough prior art search for patent applications with the greatest potential social cost, i.e., those that would impinge on many future inventive efforts. An example of this is provided by Bessen and Meurer (2004): the case

of the USPTO's 'second pair of eyes' policy meant to improve the patent examination process for business method patents belonging to patent class 705. However, an important white space in this literature is that prior research has not adequately studied the implications of codifying and improving the quality of prior art provided to patent examiners.

Our results contribute to this rich literature by studying the effect of prior art codification on the level and direction of innovation. We could find prior studies that analyze the effect of codifying prior art on either the levels or direction of innovation. Prior researchers have extensively studied the effect of other forms of patent system reform on innovation, but not the effects of prior art codification. As an example, Branstetter *et al*. (2006) coded patent system reforms across 16 countries from 1982 to 1999 along the following five dimensions: (1) an expansion in the range of goods eligible for patent protection; (2) an expansion in the effective scope of patent protection; (3) an increase in the length of patent protection; (4) an improvement in the enforcement of patent rights; and (5) an improvement in the administration of the patent system.[18] As mentioned earlier, Sakakibara and Branstetter (2001) studied the effect of expanding patent scope in Japan; wherein the Japanese patent system moved from a single-claim system to a multi-claim system.[19]

Our findings have several implications for patenting innovation in both new technological fields and older technological fields that are newly patent eligible, like software. Sampat (2004) summarizes the constraints in searching for prior art in such areas. He quotes

[18] The authors find strong similarity in the 16 reforms, with 15 out of 16 exhibiting expansion of patent rights along at least four of the five dimensions listed above.

[19] As Sakakibara and Branstetter (2001) document, until1976, Japanese patent law allowed only one independent, single claim to be included in an invention. A 1976 amendment to the patent law allowed the inclusion of multiple dependent claims, which defined the technical ways to implement an independent claim, in the same patent application. In contrast, the 1988 reforms significantly expanded the extent to which multiple claims could be included in one patent. Patent applicants could now define the coverage of an invention with multiple claims, and those claims could be either independent of or dependent on other claims. In addition, the new law expanded the extent to which related inventions could be included in a single patent.

Merges (1999) in stating that prior art in such fields tends to be in the non-patent literature, which is more difficult to search. Patent examiners at the USPTO, EPO, and other patent offices might not have access to the relevant non-patent databases in such fields. Sampat also cites Thomas (2001) and other authors to outline the learning curve in searching for prior art in new technological fields and cites Popp and Johnson (2001) and Thomas (2001) in describing the difficulties in recruiting and retaining patent examiners in new technological areas. In his own work, Sampat (2004) analyzes prior art searching in an emerging technology field—nanotechnology. From 2001 to 2003, the USPTO granted 3,748 patents in this field. To compare prior art searches in nanotechnology with those in other technological fields, the author constructs a comparison sample with the same primary patent examiners as those on the nanotechnology patents and finds that any given reference in a nanotechnology patent is approximately 5 percentage points less likely to be inserted by an examiner than in other fields. He also finds that in this emerging technology, 30 percent of the references are to non-patent prior art, almost 10 percentage points greater than the control set.

Our findings also have several implications for the literature on patent litigation. Codifying prior art from books, publications, and other non-patent sources and enabling patent examiners access to easily searchable codified prior art might lead to lower grants of legally contestable patents and might reduce litigation and social costs. The literature in economics and law has documented both the growing preponderance and the legal and social costs of granting 'bad patents,' i.e., patents that do not hold ground upon later scrutiny of the twin premises of novelty and nonobviousness. As Lemley and Sampat (2009) state, 75 percent of all patent applications lead to a patent being granted. Allison and Lemley (1998) report that nearly half the granted patents that are later litigated turn out to be invalid. As Lemley and Sampat (2012) state,

the PTO seems positioned to narrow claims in patent applications, but generally not to reject applications. The literature has also outlined several facts related to patent litigation. Lemley (2001) documents that of the roughly 2 million patents in force as of 2001, around 1,600 patent lawsuits involving about 2,000 patents were filed every year. Though the number of patents that were litigated is relatively small compared to the number of patents granted, the costs of litigation were very high. Lemley (2001) also states that the cost of patent litigation to each side is around $799,000 through end of discovery and $1.5 million through trial and appeal. In subsequent work, Cotropia, Lemley, and Sampat (2012) cite the AIPLA survey (2009) to state that bad patents impose substantial attorney fees on defendants, a median of $5 million per case. The literature has also documented the social costs of granting bad patents (Farrell and Merges, 2004). Cotropia, Lemley, and Sampat (2012) cite Chien (2009) and Lemley and Shapiro (2007) to document that bad patents might lead smaller firms to drop their products rather than defend the legality of the contested patents and might lead to inefficient outcomes in licensing technologies. Related to this, Lemley (2001) outlines the *in terrorem* effects of bad patents, where potential competitors or follow-on innovators in a field might be deterred from entering the field by the existence of bad patents owned by their competitors and the prohibitive costs of litigating these patents. In prior related work, Lerner (1995) found that small biotechnology firms avoided R&D in fields where the threat of litigation from larger firms was high. In subsequent work, Lanjouw and Lerner (2001) show that preliminary injunctions in patent cases tend to be used by larger firms hoping to impose financial distress on smaller rivals.

Finally, our results have implications for the literature in economics that has characterized Western research entities as the innovative 'North' and has labeled emerging market entities as the imitating 'South.' The underlying premise here is that most of the patented

products or processes that are consumed in the South are developed in the North. While stronger patent protection in the South would protect the North against imitation in their export market, the South would have to pay higher prices for those products. As a result, it is in the interest of the South to choose weak patent protection to facilitate imitation, thus benefitting domestic consumers in the South through lower prices (Chin and Grossman, 1988). In addition to considering consumer surplus in the South, Grossman and Lai (2004) also consider the incentives and benefits of local innovation.[20] However, the example of herbal patents is one where Western entities patented herbal formulations from China and India, where prior art was public knowledge for decades, if not centuries.

[20] In their model, the South tries to balance between two opposing priorities. On the one hand, there is the sum of extra deadweight loss that results from extending patents for domestic firms and the loss in consumer surplus that results from monopolistic pricing by foreign firms. On the other hand, by better protecting IP in the South, there is the benefit of greater innovation by firms in both the countries. They also find that the incentives to strengthen IP protection increase as the relative capacity for research grows in the South.

References

Alcacer, Juan, and Michelle Gittelman. 'Patent citations as a measure of knowledge flows: The influence of examiner citations.' The Review of Economics and Statistics 88.4 (2006): 774-779.

Alcacer, Juan, Michelle Gittelman, and Bhaven Sampat. 'Applicant and examiner citations in US patents: An overview and analysis.' Research Policy38.2 (2009): 415-427.

Allison, John R. and Mark A. Lemley (1998), 'Empirical Evidence on the Validity of Litigated Patents.' American Intellectual Property Law Association (AIPLA) Quarterly Journal 26: 185-268.

Balasubramanian, Namita (2012), 'Extent of Use of TKDL in Patent Offices: Trends and Concerns.' Working Paper Series, SSRN.

Baum C. (2001) 'Stata: The language of choice for time series analysis', The Stata Journal.

Bessen, J., and Meurer M., J. (2004), 'Lessons for patent policy from empirical research on Patent litigation', Boston University Working Paper.

Bhagwati, Jagdish N., and Hamada, Koichi (1974). 'The Brain Drain International Integration of Markets for Professionals and Unemployment: A Theoretical Analysis,' Journal of Development Economics 1: 19-24.

Branstetter L., Fisman R., and Foley, F. (2006), 'Do stronger intellectual property rights increase international technology transfer? Empirical evidence from U.S. firm-level data.' The Quarterly Journal of Economics.

Cockburn, I., S. Kortum, and S. Stern. 'Are all patent Examiners Equal? Examiners, Patent Characteristics and Litigation Outcomes in Cohen, W. and Merrill, S.(eds.) Patents in the Knowledge $ Based Economy.' (2003).

Cohen, Wesley M., Richard R. Nelson, and John P. Walsh. *Protecting their intellectual assets: Appropriability conditions and why US manufacturing firms patent (or not)*. No. w7552. National Bureau of Economic Research, 2000.

Cotropia, Christopher A., Mark A. Lemley, and Bhaven Sampat. 'Do applicant patent citations matter?.' *Research Policy* 42.4 (2013): 844-854.

Farrell, Joseph and Robert P. Merges (2004), 'Incentives to Challenge and Defend Patents: Why Litigation Won't Reliably Fix Patent Office Errors and Why Administrative Patent Review Might Help.' Berkeley Technology Law Journal 19: 943-970.

Gilbert, Richard, and Carl Shapiro. 'Optimal patent length and breadth.' *The RAND Journal of Economics* (1990): 106-112.

Grossman, G., M., and Lai, E. (2004), 'International Protection of Intellectual Property', The American Economic Review 94: 1635-1653.

Hall, Bronwyn H., and Megan MacGarvie. 'The private value of software patents.' *Research Policy* 39.7 (2010): 994-1009.

Hall, Bronwyn H., and Rosemarie Ham Ziedonis. 'The patent paradox revisited: an empirical study of patenting in the US semiconductor industry, 1979-1995.'*RAND Journal of Economics* (2001): 101-128.

Hegde, Deepak, and Bhaven Sampat (2009), 'Examiner Citations, Applicant Citations, and the Private Value of Patents.' Economics Letters 105: 287-289.

Jaffe, Adam B. and Josh Lerner (2004), Innovation and Its Discontents: How Our Broken Patent System is Endangering Innovation and Progress, and What to Do About It. Princeton: Princeton University Press.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson (1993), 'Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations.' The Quarterly Journal of Economics 108: 577-598.

Kanwar S, and Evenson R. (2009), 'On the strength of intellectual property protection that nations provide.' Journal of Development Economics.

Kerr, W., R. (2008), 'Ethnic scientific communities and international technology diffusion', The Review of Economics and Statistics.

Kerr W and Lincoln W. (2010), 'The Supply side of innovation: H-1B visa reforms and U.S. Ethnic invention.' Journal of Labor Economics.

Lanjouw J. (1998), 'The introduction of pharmaceutical product patents in India: Heartless exploitation of the poor and suffering?' National Bureau of Economic Research.

Lanjow J and Schankerman M. (2001), 'Characteristics of Patent Litigation: A Window on Competition.' The RAND Journal of Economics.

Lemley, Mark A. (2001), 'Rational Ignorance at the Patent Office.' Northwestern University Law Review 95: 1-34.

Lemley, Mark A., and Bhaven Sampat. 'Examiner characteristics and patent office outcomes.' *Review of Economics and Statistics* 94.3 (2012): 817-827.

Lemley, Mark A., Doug Lichtman, and Bhaven N. Sampat (2005), 'What to Do about Bad Patents?' Regulation 28: 10-13.

Lerner J. (2002), 'Patent protection and innovation over 150 years.' National Bureau of Economic Research.

Lerner, Josh. 'The empirical impact of intellectual property rights on innovation: Puzzles and clues.' *The American Economic Review* (2009): 343-348.

Merges, R.P. 1999. As Many as Six Impossible Patents Before Breakfast: Property Rights for Business Concepts and Patent System Reform. *Berkeley Technology Law Journal* 14.

Merrill, S.A., R.C. Levin & M.B. Myers (eds) 2004. *A Patent System for the 21st Century*. Washington, DC: The National Academies Press.

Moser, Petra. *How do patent laws influence innovation? Evidence from nineteenth-century world fairs*. No. w9909. National Bureau of Economic Research, 2003.

Moser, Petra. *Why Don't Inventors Patent?*. No. w13294. National Bureau of Economic Research, 2007.

Nordhaus, William D. 'The optimum life of a patent: reply.' *The American economic review* (1972): 428-431.

Oguamanam, Chidi (2008), 'Patents and Traditional Medicine: Digital Capture, Creative Legal Interventions, and the Dialectics of Knowledge Transformation.' Indiana Journal of Global Legal Studies 15: 489-528.

Perron P. (1990), 'Testing for a unit root in a time series with a changing mean.' Journal of business and economic statistics.

Rosenkopf L, and Almeida P. (2003), 'Overcoming local search through alliances and mobility.' Management Science 49: 751-766.

Sakakibara M, and Branstetter L. (1999), 'Do stronger patents induce more innovation? Evidence from the 1998 Japanese patent law reforms.' National Bureau of Economic Research.

Sampat, Bhaven N. *Examining patent examination: an analysis of examiner and applicant generated prior art*. Diss. University of Michigan, 2004.

Sampat, Bhaven N. 'When do applicants search for prior art?.' *Journal of Law and Economics* 53.2 (2010): 399-416.

Saxenian A. (2005), 'From brain drain to brain circulation: Transnational communities and regional upgrading in India and China.' Studies in Comparative International Development.

Thomas, J.R. 2001. Collusion and Collective Action in the Patent System: A Proposal for Patent Bounties. *University of Illinois Law Review* , 316-322.

Table 1. Milestones for the TKDL project

No.	Milestone	Timeline
1.	Recognition of need of creation of Traditional Knowledge database to protect Indian Traditional Knowledge, in particular, Ayurveda	June 1999
2.	Approach paper on setting up TKDL	October 1999
3.	Submission of approach paper to Standing Committee on Information Technology (SCIT) and World Intellectual Property Organization (WIPO).	December 1999
4.	Setting up interdisciplinary (inter-ministerial task force on TKDL)	January 2000
5.	Presentation of TKDL concept and vision at International Forum	May 2000
6.	Approval of TKDL by Cabinet Committee of Economic Affairs	January 2001
7.	Presentation on Traditional Knowledge Resource Classification at International Patent Classification (IPC) Union for getting established WIPO-TK Task Force consisting of USPTO, EPO, JPO, China, and India	February 2001
8.	TKDL software, specifications, and design	July 2001
9.	Internationally recognized specifications and standards for setting up of TK databases and registries based on TKDL specifications	November 2002
10.	Completion of data abstraction work on 36,000 Ayurveda formulations for creating TKDL in five languages	March 2003
11.	Release of CD containing sample of 500 formulations	October 2003
12.	Meeting of Ministers of Health and Family Welfare on providing access to TKDL database to European Patent Office	August 2004
13.	Request for access to TKDL by EPO	July 2005
14.	Approval on access to TKDL database by international patent offices (IPO) by Cabinet Committee of Economic Affairs	June 2006
15.	Access Agreement sent to EPO	July 2006
16.	Request by USPTO for access to TKDL database	December 2006
17.	EPO signing of Access Agreement	February 2009
18.	USPTO signing of Access Agreement (coinciding with state visit of Indian Prime Minister Dr. Manmohan Singh to Washington D.C.)	November 2009

Source: Documents from the Council for Scientific and Industrial Research (CSIR), India and interviews conducted by researchers

Table 2. Comparison of the final TKDL 'Access Agreements' signed by the USPTO and EPO with the CSIR

USPTO	EPO
No general restrictions on purpose of use of TKDL	General restriction on use of TKDL 'for the purposes of European patent grant procedures'
Explicit access permissions to both USPTO examiners/staff and 'contractors engaged in search of Patent Cooperation Treaty (PCT) applications'	Use limited to EPO only
Information use allowed for both patent examination and other 'internal purposes such as statistical and technical analysis, training, developing classification schedules, definitions and planning, etc.'	Information use permitted only for 'the purposes of the European patent grant procedure in all its phases'
TKDL information transfer is permitted to third parties for patent search and examination, to patent applicants/legal representatives for prior art purposes, and to the public via the USPTO's Patent Application Information Retrieval System	Sharing of TKDL information is limited only for patent grant procedures and to patent applicants, although form of information provision is not limited
User access restrictions to 30 IP addresses simultaneously	Options for user ID-based access restrictions or IP address-based access

Source: Based on analyses of researchers of the Traditional Knowledge Digital Library (TKDL) Access Agreements concluded by the Council of Scientific and Industrial Research with the USPTO and EPO and interviews.

Table 3. IPC classification related to herbal medication

IPC	Description	Number of subgroups
A61K35/00	Medicinal preparations containing material or reaction products thereof with undetermined constitution	0
A61K36+	Medicinal preparations of undetermined constitution containing material from algae, lichens, fungi or plants, or derivatives thereof, e.g., traditional herbal medicines	207
A61K125	Containing or obtained from roots, bulbs, tubers, corms, or rhizomes	0
A61K127	Containing or obtained from leaves	0
A61K129	Containing or obtained from bark	0
A61K131	Containing or obtained from seeds, nuts, fruits, or grains	0
A61K133	Containing or obtained from flowers or blossoms	0
A61K135	Containing or obtained from stems, stalks, branches, twigs, or shoots	0
A01N25+	Biocides, pest repellents or attractants, or plant growth regulators, characterized by their forms, by their non-active ingredients, or by their methods of application (fungicidal, bactericidal, insecticidal, disinfecting, or antiseptic paper D21H); substances for reducing the noxious effect of the active ingredients to organisms other than pests	17
A01N65+	Biocides, pest repellents or attractants, or plant growth regulators containing material from algae, lichens, bryophyta, multicellular fungi or plants, or extracts thereof	24
Total		**258**

Source: Interviews and analysis of the following secondary data sources – (i) Dominic Keating, Defensive Protection of Traditional Knowledge at the United States Patent & Trademark Office (available at www.wipo.int/edocs/mdocs/tk/en/.../wipo_tkdl_del_11_ref_t7_2.pdf) and (ii) Georg Schiwy-Rausch, information manager, data acquisition at the EPO, has the presentation on Traditional Knowledge at the EPO: Present & Future (13-15 December 2006) available at http://pame.european-patent-office.org/pubs/hararepdf/tk_aripo_present_future.pdf.

Table 4. Matching U.S. Patent Classifications (USPC) with the IPC classification

USPC	Description	Number of subgroups	List
424/725	PLANT MATERIAL OR PLANT EXTRACT OF UNDETERMINED CONSTITUTION AS ACTIVE INGREDIENT (E.G., HERBAL REMEDY, HERBAL EXTRACT, POWDER, OIL, ETC.)	55	725.1, 726-779
514/783	DESIGNATED ORGANIC NON-ACTIVE INGREDIENT CONTAINING OTHER THAN HYDROCARBON: Plant extract or plant material of undetermined constitution	0	
Total		**57**	

Source: Interviews and analysis of the following secondary data sources – (i) Dominic Keating, Defensive Protection of Traditional Knowledge at the United States Patent & Trademark Office (available at www.wipo.int/edocs/mdocs/tk/en/.../wipo_tkdl_del_11_ref_t7_2.pdf) and (ii) Georg Schiwy-Rausch, information manager, data acquisition at the EPO, has the presentation on Traditional Knowledge at the EPO: Present & Future (13-15 December 2006) available at http://pame.european-patent-office.org/pubs/hararepdf/tk_aripo_present_future.pdf.

Table 5. Summary statistics of characteristics of USPTO patents (1977-2010)

Variable	Mean	Std. dev.	Min.	Max.
% Synthetic	0.24	0.11	0.09	0.41
% Fortune 1000	0.06	0.04	0	0.14
% U.S. Univ.	0.06	0.03	0	0.11
% Foreign Univ.	0.09	0.04	0.01	0.17
% Others	0.80	0.07	0.67	0.94

Table 6. Percentage of USPTO patents based on herbs and synthetics pre-TKDL and post-TKDL (pre-TKDL is 1990-2003 and post-TKDL is 2003-2011)

% Synthetic	Mean	Std. dev.	Min.	Max.
Pre_TKDL	0.22	0.64	0.11	0.31
Post_TKDL	0.28	0.09	0.09	0.41

% Synthetic	Mean	S.E.	t-test	[95% Conf. interval]	
Diff (Pre_TKDL-Post_TKDL)	-0.06	0.03	-2.07	-0.13	0.00

Notes: This analysis is based on the sample of all herbal patents granted on the USPTO between 1977 and 2010

Table 7. Impact of TKDL on the mix of herbal patents (pure herbals vs. herbs mixed with synthetic compounds)

Dependent variable: *is_synthetic*

		1	2
Application year			
	2004	-0.15 (0.38)	-0.14 (0.38)
	2005	0.54 (0.40)	0.52 (0.40)
	2006	0.67* (0.35)	0.71* (0.35)
	2007	0.89** (0.34)	0.87** (0.34)
	2008	0.94** (0.43)	0.95** (0.43)
	2009	0.41 (0.39)	0.39 (0.39)
	2010	1.25*** (0.42)	1.20*** (0.42)
Control variables			
	Is_Fortune1000		13.5*** (1.04)
	Is_U.S._univ		-28.00*** (1.57)
	Is_foreign_univ		-13.52*** (1.06)
	N	2,299	2,299

Notes:

- This analysis is based on the sample of all herbal patents granted on the USPTO between 1977 and 2010
- The dependent variable for this analysis is a dummy variable representing whether the patent is purely based on herbal medicine or is based on herbal medicines mixed with synthetic compound. The variable *is_synthetic* takes values equal to 1 if the patent document (title and/or claims) indicated the presence of a synthetic compound. An example of a pure herbal patent is USPTO patent US20010012524A1, where the claim indicates the following – (this)"Invention relates to an anti-aging agent, the anti-aging agent is a solvent extract of Uncaria Gambir Roxburgh". An example of a mixed patent is USPTO patent US20030113388A1 which is a "Composition for Skin Treatment comprising turmeric extracts [herb], hydroxy acids, and alpha-1-antitrypsin [synthetic]"
- The coding of the dependent variable was done manually by the researchers by studying the patent title and claims.
- The main independent variables in this analysis comprise year dummies representing the year of patent application. The main control relates to assignee type.
- We run a Logit regression with robust standard errors.
- *p < 0.1; **p < 0.05; ***p < 0.01.

Table 8. Summary statistics of percentage of patents filed by Indians and Europeans in the herbal patent sample and randomly generated control patent sample

Is_Indian	Mean	Std. dev.	[95% Conf. interval)	
Main sample	0.15	0.36	0.14	0.17
Control sample	0.11	0.31	0.10	0.13

Is_Indian	Mean	Std. err.	t-test	[95% Conf. interval]	
Diff	0.04	0.01	3.95	0.02	0.06

Is_European	Mean	Std. dev.	[95% Conf. interval)	
Main sample	0.20	0.40	0.18	0.21
Control sample	0.24	0.43	0.22	0.26

Is_European	Mean	Std. err.	t-test	[95% Conf. interval]	
Diff	-0.05	0.01	-3.38	-0.07	-0.02

Notes:

- This analysis is based on the sample of all herbal patents granted on the USPTO between 1977 and 2010
- Following methodology used by Jaffe et al. (1993), for every herbal patent, we randomly selected a 'control patent' by matching the herbal and control patents on the application year and same three-digit U.S. technology classification. Having generated the set of patents with the same application year and same original three-digit U.S. classification as the herbal patent, we identify the patent in the set that has the closest grant date to the herbal patent.
- We then employ methodology from Kerr (2008) and help from Bill Kerr to code the ethnicities of inventors on both the herbal and control patents

Table 9. Regression results – Ethnicities of inventors in the herbal patent sample and control patent sample

Top panel dependent variable: *is_Indian*

	1	**2**	**3**
Is_herbal	0.61**	0.64**	1.09***
	(0.45)	(0.45)	(0.49)
Post_2004		1.64**	2.11***
		(0.78)	(0.79)
Is_herbal*post2004			-1.2***
			(0.42)
N	872	872	872

*p < 0.1; **p < 0.05; ***p < 0.01.

Bottom panel dependent variable: *is_European*

	1	**2**	**3**
Is_herbal	-0.55**	-0.55**	-0.32
	(0.25)	(0.25)	(.29)
Post_2004		1.7	1.98
		(1.56e7)	(1.28e7)
Is_herbal*post2004			-0.60
			(0.38)
N	1,005	1,005	1,005

Notes:

- This analysis is based on the sample of all herbal patents granted on the USPTO between 1977 and 2010. Following methodology used by Jaffe et al. (1993), for every herbal patent, we randomly selected a 'control patent' by matching the herbal and control patents on the application year and same three-digit U.S. technology classification.
- The dependent variable for this analysis in the top panel is a dummy variable *is_Indian*, representing whether at least one of the individuals filing the patent is of Indian ethnicity. The variable takes a value equal to 1 if at least one of the applicants belongs to Indian ethnicity and 0 otherwise.
- The dependent variable for this analysis in the bottom panel is a dummy variable *is_European*, representing whether at least one of the individuals filing the patent is of European ethnicity. The variable takes value equal to 1 if at least one of the applicants belongs to a European ethnicity and 0 otherwise.
- The main independent variables are the following: (1) *is_herbal*: a dichotomous variable representing whether the patent is based on herbal medicine or not. It includes both pure herbal patents and patents with herbal mixed with synthetic. The variable takes a value equal to 1 if the patent is based on herbal medication and 0 if the variable is a control patent; (2) *post_2004*: a dichotomous variable representing whether the patent was applied after year 2004 or not. We use this variable to represent the exogenous event that led to a decrease in the overall H1B visas available for Indians. The variable takes a value equal to 1 if the patent application year is after 2004.
- The top panel represents the actual test where we are analyzing whether Indians are driving the filing of herbal patents; the bottom panel represents a placebo outcome test with European ethnicities.
- We employ robust standard errors and control for any assignee-specific characteristics by clustering the fixed effects for each assignee. *p < 0.1; **p < 0.05; ***p < 0.01.

Table 10. Regression results – Presence of diaspora inventors in the herbal patent sample and control patent sample

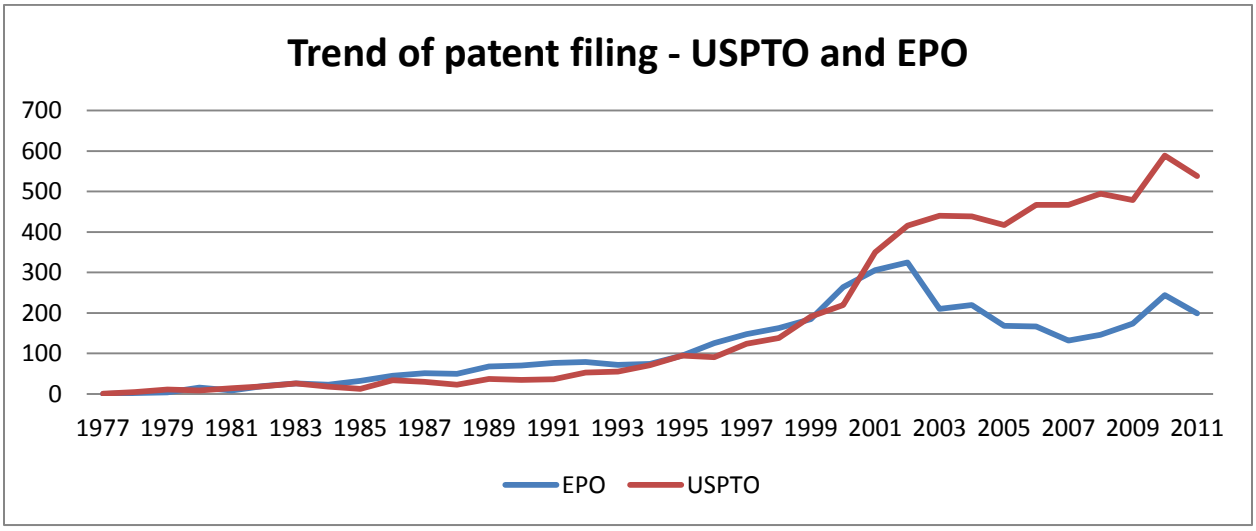
Chinese herbal patents and Chinese diaspora scientists

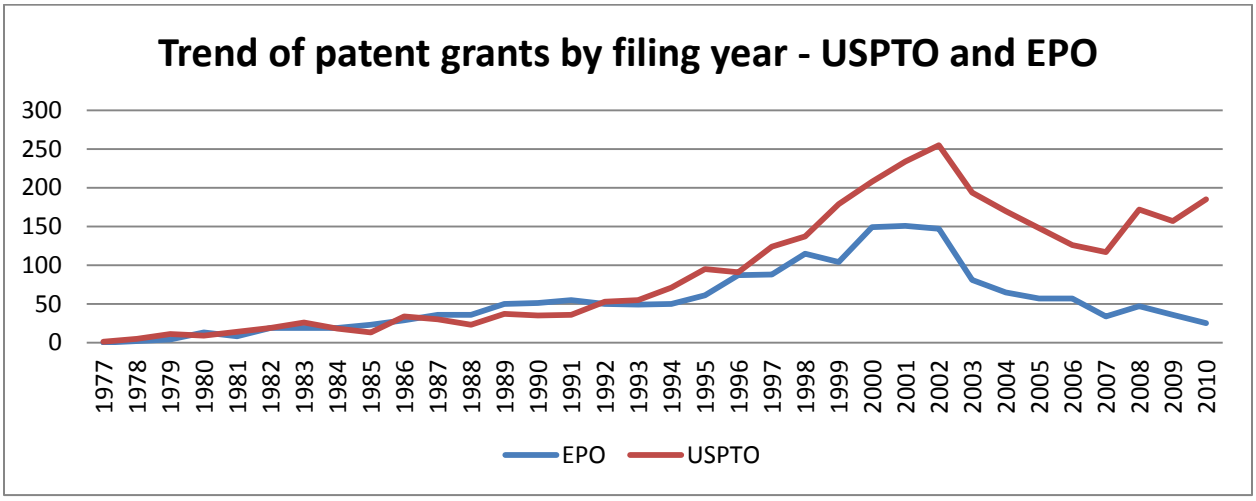
	Inventor list includes ethnic Chinese		Inventor list includes diaspora Chinese	
Is Chinese herbal patent	0.41*** (0.11)	0.62*** (0.12)	0.25** (0.12)	0.44*** (0.13)
Is Indian herbal patent	-	-1.16*** (0.19)	-	-1.01*** (0.19)
N	2,298	2,298	2,298	2,298

Indian herbal patents and Indian diaspora scientists

	Inventor list includes ethnic Indian		Inventor list includes diaspora Indian	
Is Indian herbal patent	1.28*** (0.13)	1.38*** (0.13)	0.09 (0.19)	0.24 (0.20)
Is Chinese herbal patent	-	-0.33** (0.13)	-	-0.59*** (0.20)
N	2,298	2,298	2,298	2,298

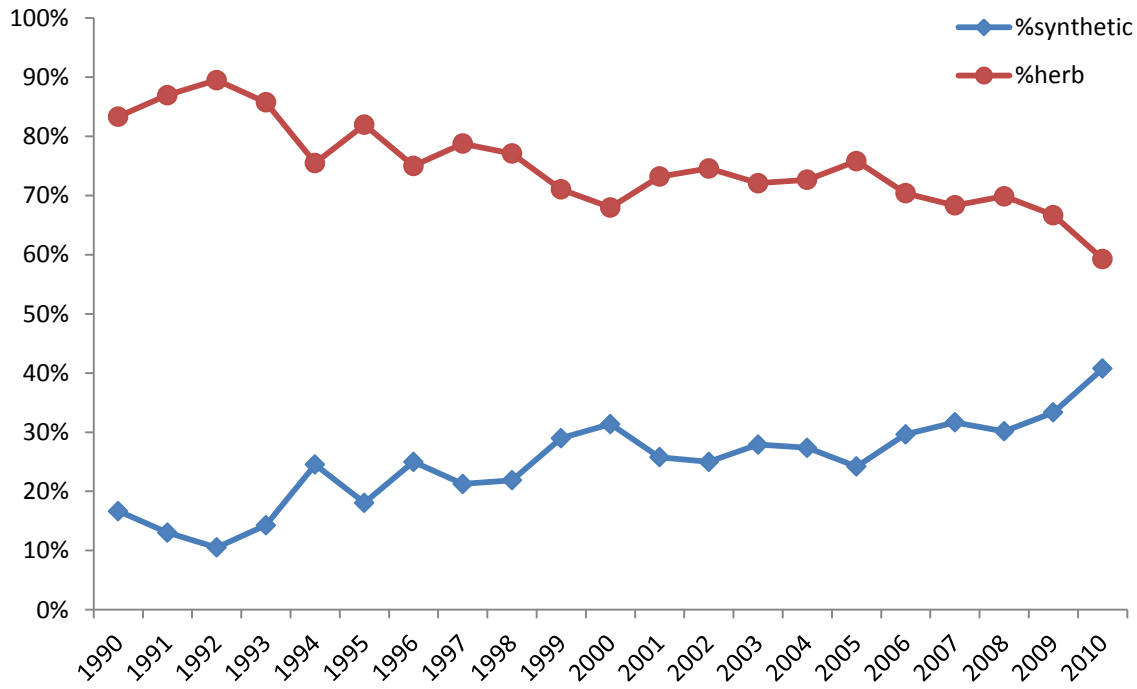
Figure 1. Trend of herbal patent filing and herbal patent grants at EPO and USPTO (1977-2011)





Notes: We used two widely used patent databases, Thomson Innovation and LexisNexis TotalPatents to create this unique herbal patent dataset. We searched through every EPO/USPTO patent filed from 1977 to 2011 to construct our database and used both 'keyword search' and 'patent classification search' to identify the herbal patents.

Figure 2. Mix of patents based on herbal and patents based on herbal mixed with synthetic (analyses only for USPTO)



Notes: This analysis is based on the sample of all herbal patents granted on the USPTO between 1977 and 2010. The top line represents the share of pure herbal patents, while the bottom line represents the share of herbals in conjunction with synthetic formulations in the total mix of herbal patents.

Figure 3. Derwent classification for herbal patents (analyses only for USPTO)

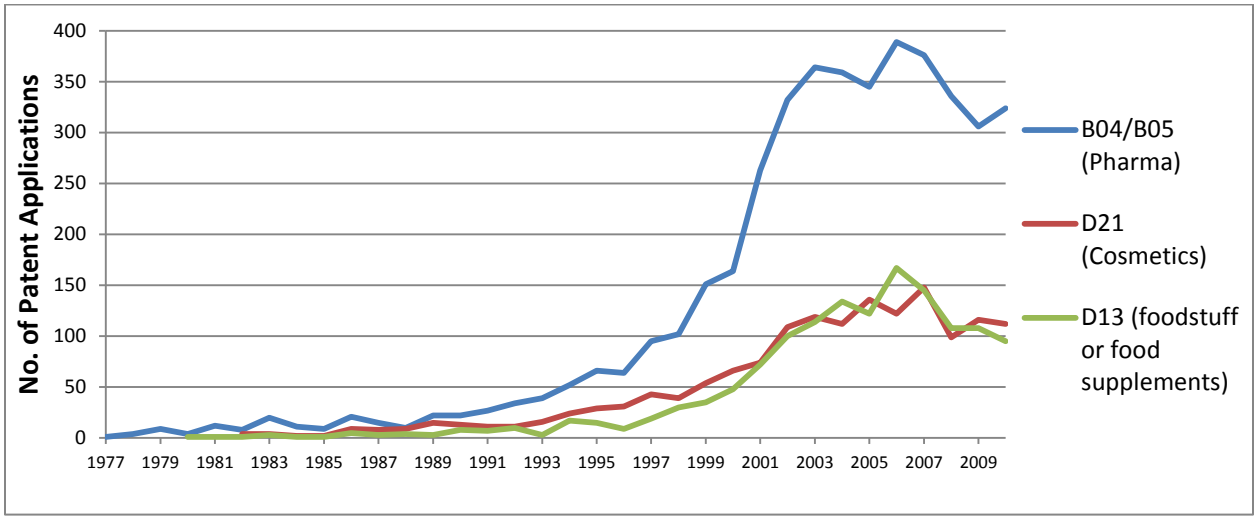
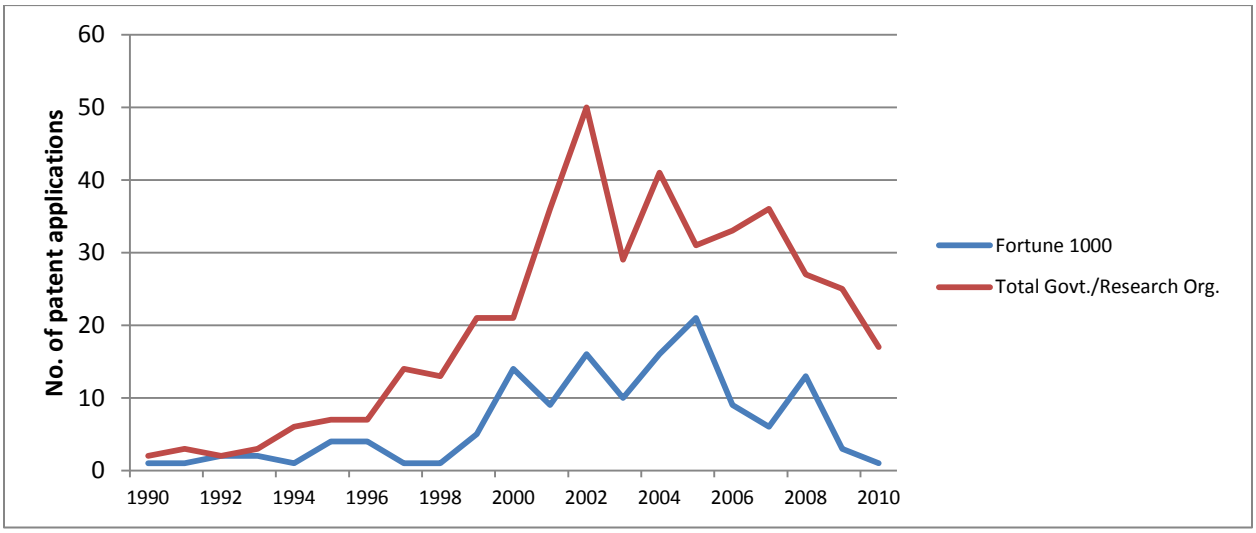
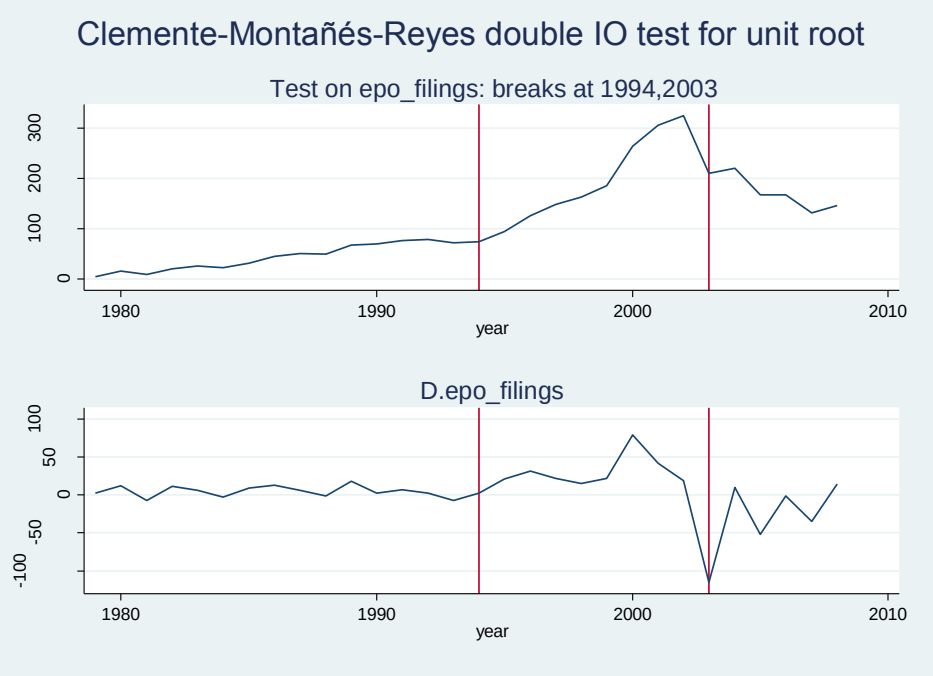


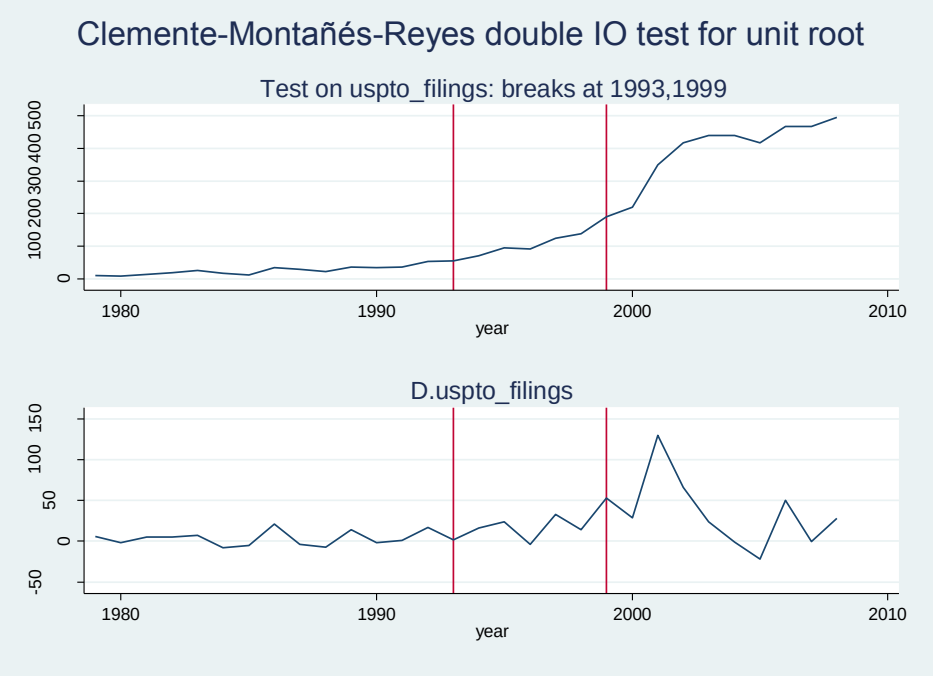
Figure 4. Herbal patent filing by assignee type (analyses only for USPTO)



Notes: This analysis is based on the sample of all herbal patents granted on the USPTO between 1977 and 2010

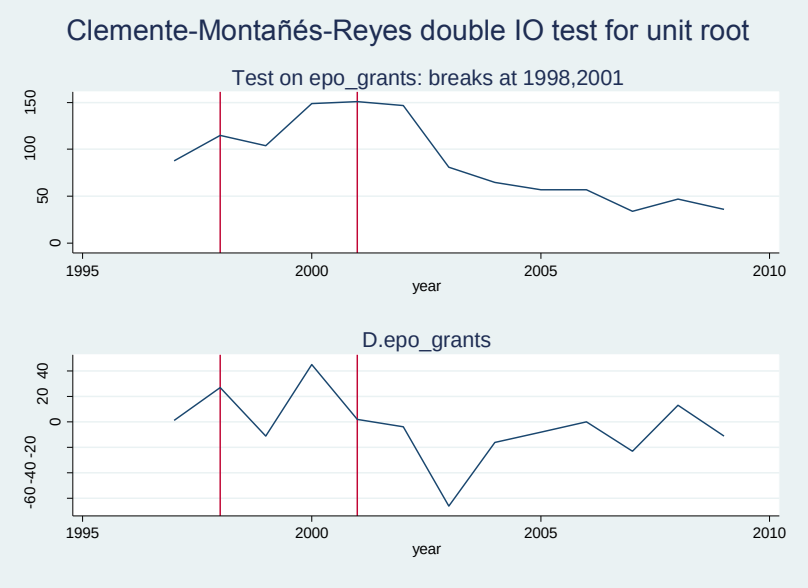
Figure 5. Results of breakpoints analysis for USPTO and EPO patent filings

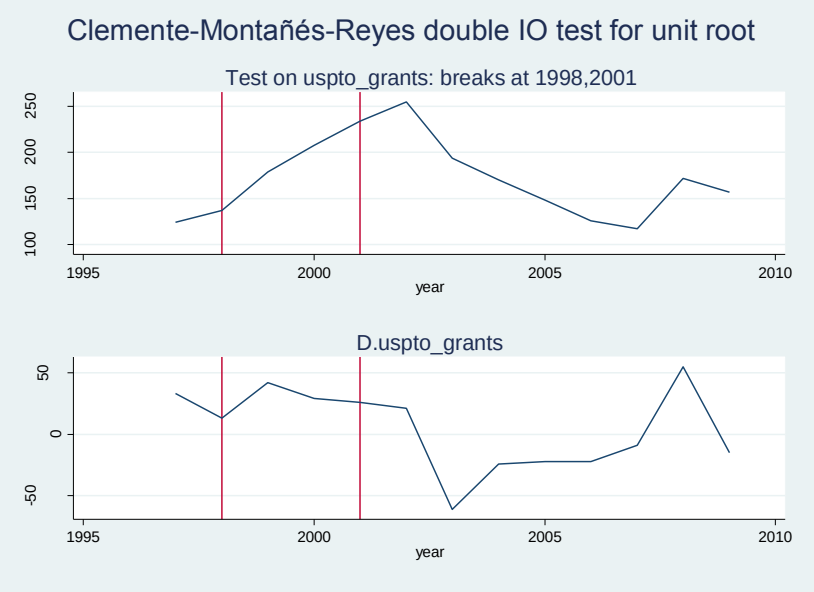




Notes: The breakpoints unit root test developed by Perron and Vogelsang (1992) was implemented using the clemio2 routine in STATA (Baum, 2005). For patent filings at the EPO, the test detected two structural changes in years 1994 and 2003. There was an increase in herbal patent filing in 1994 and a decrease in herbal patent filing in 2003. For the USPTO, the test detected two structural changes in years 1993 and 1999. Both years marked an increase in herbal patent filing.

Figure 6. Results of breakpoints analysis for USPTO and EPO patent grants (based on year of filing)





Notes: The breakpoints unit root test developed by Perron and Vogelsang (1992) was implemented using the clemio2 routine in STATA (Baum, 2005). For patent grants, we conducted the analysis for the period post 1995. For the EPO, the test indicated breakpoints in patent filing years 1998 (negative breakpoint) and 2001 (negative breakpoint). The variable 'year' here indicates 'year of filing the patent'; there is an estimated four or five year lag between a filing a patent and the patent getting granting. This indicates that at the EPO, there was a decline in patent grants around both 2003 and 2006. For the USPTO, the test indicated breakpoints for patent filing years 1998 (positive breakpoint) and 2001 (negative breakpoint). Given the four to five year lag between year of patent filing and year of patent grant, this indicates that at the USPTO, there was an increase in patent grants in 2003 and a decline in patent grants around 2006.