INDIA POLICY FORUM 2025

Some Facts About Indian Innovation in the Twenty-First Century

Namrata Kala
MIT Sloan School of Management

Josh Lerner
Harvard University

Junxi Liu
University of Warwick

India Policy Forum
June 26–27, 2025

NATIONAL COUNCIL OF APPLIED ECONOMIC RESEARCH
NCAER, 11 Indraprastha Estate, New Delhi 110002
Tel: +91-11-2345 2698, +91-11-6120 2698
Email: events@ncaer.org | www.ncaer.org

Some Facts About Indian Innovation in the Twenty-First Century*

Namrata Kala
MIT Sloan School of Management

Josh Lerner
Harvard University

Junxi Liu
University of Warwick

Abstract

Innovation is a key driver of growth and development, and understanding its evolution in developing countries can provide important insights into the development process. Drawing on Kala, Lerner, and Liu (2025) which assemble novel data on Indian patents, this paper describes several key stylized facts of Indian innovation in the twenty-first century. We find considerable variation in patenting by geography, Indian vs. foreign inventors, ownership of inventing organization (public vs. private) and type of organization (academic institutions vs. firms). Given the rapid growth of patenting and venture capital in India and the intense interest in innovation in developing countries more generally, it is our hope that this work will stimulate research in this area by numerous economists and other innovation scholars.

Keywords:

JEL Classification:

*MIT Sloan School of Management; Harvard University; University of Warwick. Kala and Lerner are affiliates of the National Bureau of Economic Research. We thank MIT Sloan School of Management and the Harvard Business School Division of Research and Doctoral Programs for financial support. Brooke Jones, Zhaoying (Laura) Wang, and Austin Zhu provided very valuable research assistance in this effort. Lerner has received compensation for advising limited partners in venture funds, venture capital groups, and governments designing policies relevant to venture capital. All errors and omissions are our own.

Preliminary draft. Please do not circulate beyond the **NCAER India Policy Forum 2025**, for which this paper has been prepared.

Email: Namrata Kala: kala@mit.edu; Josh Lerner: josh@hbs.edu;
Junxi Liu: junxi.liu@warwick.ac.uk

1 Introduction

Innovative activity is key to the development process and is increasingly common in developing countries. In India for instance, R&D expenditures, have climbed twenty-five fold in rupee terms (though only six times over in U.S. dollars) between 1990-91 and 2020-21) (Department of Science and Technology, Ministry of Science and Technology, Government of India, 2023). Furthermore, the venture capital sector has been strongly linked to innovation in both developed (Akcigit et al., 2022; Kortum and Lerner, 2000) and developing (Lerner et al., 2024) economies. This sector has also grown sharply in India in recent decades. Investment climbed from a level of $67 million in 2005 (0.2% of the world total) to $14.5 billion in 2021 (5.2% of the world total).[1]

This rapid growth in innovation has been facilitated by internal and external factors. Domestic firms — such as Tata Motors, Sun Pharmaceuticals, and Dr. Reddy's — have been increasingly undertaking R&D. The government-owned National Investment and Infrastructure Fund and Indian high net worth families have similarly become important investors in local venture funds. The other driver has been the inflow of capital from global investors who recognize the deep pool of technical talent, large domestic market, and widespread adoption of information technology as critical strengths. This trend has accelerated as Western investors and corporations have increasingly look for alternative investing opportunities than those in China, and recent policies have tried to encurage innovative activity at the state and national level (Many of these policies are summarized in Department of Science and Technology, Ministry of Science and Technology, Government of India (2020)).

In this, paper we draw on an Indian patent dataset that we constructed in Kala, Lerner, and Liu (2025), to showcase some stylized factors about the levels and trends of innovation in between 2005 and 2024. We constructed this data using the methodology pursued with the U.S. patent data since the inception of the first patent project at National Bureau of Economic Research (Griliches, 1984). It emulates many of the key steps undertaken in the literature using the U.S. Patent and Trademark Office (USPTO) data, such as the compilation of citations (e.g., Jaffe and Trajtenberg (2002)), the calculation of measures of patent importance (Kelly et al., 2021), the analysis of patent text (Kalyani et al. (2025); Myers and Lanahan (2022)), the construction of international patent families (Putnam, 1997), the disambiguation of assignees (e.g., Jaffe and Trajtenberg (2002); Magerman et al. (2006), and the identification of venture-backed patents (Bernstein et al., 2016).

We examine patents granted between the effective date of the Patents (Amendment)

[1]These statistics are based on the authors' analysis of the PitchBook data.

Bill 2005, and the time that the data were downloaded at the end of 2024. Following the conventions of the economics of innovation literature (e.g., Jaffe and Trajtenberg (2002)), we focus exclusively on the most important patent class, utility patents. We do not consider design patents and India does not issue plant patents. Finally, we cover the lack of comprehensive availability of Indian patents in international data sources like WIPO in Kala, Lerner, and Liu (2025), which necessitate the construction of this dataset.

We document seven stylized facts about Indian innovation in recent decades. First, there is a large increase in patenting activity - total patents increased from just over 5,000 in 2005 to over 25,000 in 2018, a five-fold increase. Second, amongst the top twenty contributors to Indian patents during this time period, thirteen are firms, six are academic institutions, and one is non-academic government organization, with the Indian Institutes of Technology prominently represented in academic organizations. Third, the number of patent citations have been increasing over time but remain well below US levels and are driven largely by the extensive margin of patents citing any other patents. Fourth, geographically, innovative activity is dispersed but some states like Maharashtra have consistently been responsible for large amounts of patent grantees, with patenting quite concentrated in a few districts within these states. Fifth, firms (and in particular private firms rather than state-owned enterprises) are the largest source of innovative activity, followed by academic organizations. Sixth, innovation by sector is quite dispersed, with the top innovative sectors ranging from software, pharmaceuticals, telecommunications, and motor vehicles. Seventh, innovation has become more collaborative over time across individuals and organizations, and private-sector-academia collaborations are an important source of cross-organizational collaborations.

2 Background

The promulgation of the World Trade Organization's Trade-Related Aspects of Intellectual Property (TRIPS) agreement in the mid-1990s coincided with a period of economic liberalization in India. While India signed to the TRIPS agreement in 1994, it was not until 2005 that a fully compliant patent law was enacted.[2]

In particular, the Patent Law of 2005 – which came into effect at the beginning of that year after a protracted legislative process — modernized the Indian patent system. It extended full patent protection (including product awards) for pharmaceuticals and food,

[2]In April 1995, the Indian government agreed to receive patent applications for pharmaceutical and agricultural products and to store them for processing once the Indian patent law to make such subjects patentable took effect. See for details, https://www.wto.org/english/tratop_e/dispu_e/79r.pdf

as well as those involving chemical processes and extended patent life to 20 years from the application date, bringing India into compliance with the minimum standards established by the TRIPS agreement. It did retain some provisions from the earlier acts. In particular, the ability of the state to impose compulsory licensing for essential patents that were not being put into practice and the creation of a library to document the prior art associated with traditional Indian medicinal knowledge (and thus prevent its incorporation into patents in India or elsewhere, as had happened in the case of turmeric (Jayaraman, 1997)) reflected longstanding concerns about the potential abuse of the patent system.

3 Data Assembly

The main source of our data is IP India, the official website of the Indian Patent Office. We scrape every granted patent with a filing year from 2005 to 2022, along with all documents attached with each patent. Details about the dataset are documented in Appendix A.

We extract most information directly from the website data, including patent filing information (year of filing, year of grant, inventors, grantees) and patent content information (title, abstract, complete specification). We conduct disambiguation of patent grantees and link business grantees to the Prowess database (compiled by the Center for Monitoring the Indian Economy (CMIE)) to trace firms' financial performance. Finally, we also geolocate both grantees and inventors to the district-level.

We additionally leverage attached documents of each patent. First, where the complete specification from the website is incomplete, we directly extract the complete specification from the Form 2 document. Second, we extract backwards citation information of each patent from the First Examination Report (FER) document.

In total, we scrape a total of 295,743 granted patents from IP India for patents awarded between 2005 and 2022 and granted before 2024. (For the subsequent versions of analyses, we will extend the sample to include all patents awarded from the beginning of the modern Indian patent system in January 2005. We will also use the entire body of patent awards, whether domestic and foreign, to create textual measures of patent importance a la (Kelly et al., 2021).

A large number of patents recorded in the Indian Patent Office are patents originally granted in other countries and transferred to the Indian Patent Office through Patent Cooperation Treaty (PCT) or other arrangements. To focus on localized Indian innovation, we create a specific Indian patent dataset by including the following patents of either of the following types: (1) patents with at least one assignee whose address is in India; and (2) patents with at least one inventor whose address is in India. We identify 67,369 Indian

patents using this definition. We document the details on the process of identifying Indian patents in Appendix B. We plot the number of all patents and Indian patents over time by application time in Figure 1 and by date of grant in Figure 2.

While our scraping ends for the year of 2023, even if we have scraped all granted patents applied from 2005 and 2022, the time lag between application between the date of application and the date of grant means that we will have an incomplete coverage for the later years. In Figure 3, we document the very long wait time between application to grant: for most of the years, the median wait time for patents granted in that year is more than 5 years. Therefore, for analysis of time trends based on application date, we cut off our data at 2018 to avoid the distortion in trends in later periods.

3.1 Grantee and Inventor Information

In the Indian context, the grantee is the original organizational or individual owner of the patent rights according to the Indian Patent Act 1970, similar to the concept of "assignee" in the U.S. context. From the IP India website, we obtain the Grantee Name, Grantee Type, and Grantee Address information of the grantee(s) for the patent. Each patent can have multiple grantees.

On average, each patent has 1.24 grantees, though 87.7% patents only have one grantee. In comparison, patents granted between 2005 to 2022 in the U.S. on average has 1.04 assignees, with 96.43% patents only having one assignee.

An inventor is the individual(s) contributed to the claimed invention. From the IP India website, we obtain Inventor Name, Inventor Address, Inventor Country, and Inventor Nationality information of the inventor(s) for the patent. Each patent can have multiple inventors. On average, each patent has 3.18 inventors. In comparison, patents granted between 2005 to 2022 in the U.S. on average has 2.80 inventors.

3.1.1 Disambiguation of Grantees

As in many other nations, assignee names in India can be difficult to work with. In some cases, assignees can be subsidiaries of other firms. In others, names can be misspelled or abbreviated. We disambiguate the grantees to combine different names for the same entity due to typographical errors, alternative spellings, and different localities. We focus on the set of grantees that are organizations or enterprises (grantee type is not "Natural Person"). We document the detailed disambiguation procedures in Appendix C.

The raw Indian data contains 32,868 unique grantee names, of which 11,236 are organizations or enterprise grantees (i.e. not individual persons). After disambiguation, the

number of unique organization or enterprise grantees reduces from 11,236 to 7,129.

3.1.2 Identifying Grantee Types

While data from IP India do provide information on grantee types, the variable is very inconsistently defined. For instance, India Institute of Technology is sometimes classified as a "Academic Institution," but sometimes classified as "Other than Natural Person". Therefore, we conduct our own identification of grantee types by using keyword matching.

Specifically, for all grantees that are not "Natural Persons", we search for possible keywords that could identify them as academic institutions or government entities. We regard the remaining grantees as commercial firms (business) grantees. We describe the detailed process of matching in Appendix D.

Among the disambiguated grantees that are not natural persons, we are able to identify 1,497 academic institutions, 79 government organizations, and 5,553 businesses. We report the top patent grantees by number of patents and their type in Table 1. If considering the first grantee of each patent, among the 67,061 patents that have grantee information, 36,507 patents are granted to grantees which are identified to be businesses, 15,380 are granted to academic institutions, 13,797 are granted to individual (natural persons), and 1,375 are granted to government entities. The heavy representation of university assignees (22.9%) and individuals (20.6%) can be compared to the 4.5% and 1.0% seen in the U.S. (for patents awarded between 1985 and 2023) and the much more similar 17.6% and 14.0% in China (again for 1985 to 2023) (Lerner et al., 2025).

3.1.3 Geolocation of grantees and inventors

In order to understand the spatial distribution of Indian patents, we parsed postcodes, districts, and states from grantee and inventor addresses. Indian postcodes were provided in about 90% of the addresses, which allowed us to extract corresponding district and state information at a higher level. For the remaining 10% of addresses without postcodes, we were able to extract the district (and, thus, the state) information using a combination of natural language processing detection and manual checks. Overall, we extracted district and state information for 98% of inventor and grantee addresses. Appendix E provides further detail on our process for extracting location information.

Our data show the wide variation in the location of inventors and grantees across India, illustrated in Table 2. About 22% of the inventors are located in Maharashtra (which includes Mumbai and is the third most populous state in India), 16% in Tamil Nadu (including Chennai), and 15% in Karnataka (which encompasses Bengaluru, "the Silicon

Valley of India"). Grantees show a similar distribution, with about 25% of addresses in Maharashtra, 14% in Tamil Nadu, and 13% in Delhi. The remaining addresses for inventors and grantees are spread across 37 other states.

3.1.4 Identifying Inventor's Organization

While grantee information documents the property of the patent, it is often the case that multiple inventors collaborate on a single patent. To identify inventor organizations from patent filings, we analyze the free-form addresses provided for each inventor, as these data often reveal the organization of the inventor.

We utilize a large language model (LLM) to parse these unstructured addresses. The model extracts and standardizes organization names when present. Furthermore, for cases where the address does not reveal an organization, but is indeed a well-known organization address, the LLM can infer the organization correctly based on public knowledge it has. We document how we extract, post-process, and validate the results in Appendix F.

3.1.5 Linking Grantees to Prowess

To investigate firms' performance and innovation, we link business grantees (as opposed to natural persons, academic institutions, or government entities grantees) to firms in Prowess data, one of the largest database of financial performance of Indian companies. We describe the detailed process of matching in Appendix G.

There are 5,553 unique firm grantees and we are able to match 1,239 of them with Prowess data. Since we have manually checked all grantees with more than 10 patents, this set of matched firms should cover the most important firms and represent a large portion of the total patent grants: among the 36,507 patents whose first grantee is a Business, our matched grantees are granted a total of 28,704 patents as the first grantee, representing 78.6% of the total.

3.2 Patent Information

3.2.1 Patent Full Text: Complete Specification

The textual information of the patents will be used to perform various analyses. We obtain the complete specifications (i.e., the full text description) for most of patents, and conduct quality checks for complete specifications. Some of the complete specifications downloaded from the website are missing, corrupted, or refer to attached PDF files. For these cases, we go to the attached document for the complete specification and parse

from the PDF directly. We also check each parsed complete specification: whether they have a minimal length and whether they contain keywords that are common in most complete specifications. We document details on our parsing and quality check procedures in Appendix H.

We sought to extract complete specifications for the full set of granted patents in India (295,743 patents) and succeed in extracting 97.8% of the patents (289,832 patents).

3.2.2 Backwards Citations

To assess the relative novelty of each patent, we extracted backward citations — the patent literature referenced — from the First Examination Report (FER) of each patent. Citations may serve as a unsatisfactory role of indicating quality for patents in India, as in many other developing countries, because of the scarcity of citation activities. This scarcity of citations by domestic applicants seems to be a characteristic of emerging market patent offices: for a discussion of citation practice by Chiense inventors, see Lerner et al. (2025). Despite the paucity of these cite, we extract backwards citations to patent literature for all patents.

Among the entire set of 295,743 granted patents, 244,431 of them have FER attached. Patents without FER files (51,312 patents) are predominantly in the early period - 49,516 of them are in the years between 2005 and 2010. We extract citation records for 233,319 patents, representing 95.5% of all patents with FER reports. Unsuccessful cases are mostly due to the patent not citing any patent, with a small number of FER reports having PDF quality issues. Appendix I provides details on our extraction process.

On average, each patent cites 2.28 patents, where non-Indian patents make 2.29 citations and Indian patents make 2.24 citations[3].

4 Stylized Facts About Indian Innovation

4.1 Large Increase in Patenting Activity

One of the most striking facts about Indian patenting activity has been its fast growth over time - total patents increased from just over 6,000 in 2005 to over 25,000 in 2018, a nearly five-fold increase. This is largely driven by foreign patents, with the share of Indian patents increasing from less than 800 in 2005 to almost 7,000 in 2018.

[3]We assume patents without FER files or FER files having PDF quality issues to have cited zero patents here. If conditional on patents with any citation, each patent cites 2.90 patents, where non-Indian patents make 2.95 citations and Indian patents make 2.74 citations

4.2 Top Contributors to Innovation by Patent Numbers and Their Composition

Who are the organizations most responsible for this large amount of innovation in recent decades? Table 1 shows the identity, type, and number of patents granted to the top twenty organizations by number of patents (we focus on the first listed grantee for this exercise). Thirteen of the them are firms, six are academic institutions, and one is a government (non-academic) organization. The Council of Scientific and Industrial Research (CSIR) received the largest number of granted patents in this period, at 2154, representing 3.2% of total Indian patents, followed by Bharat Heavy Electricals at 1,779 patents. Large firms that are part of conglomerates are in this list of top innovators, with three of the twenty being Tata firms, and two Samsung firms. Moreover, prominent Indian academic institutions also form part of this list, with four Indian Institutes of Technology (IITs) in this list, as well as the Indian Insitute of Science.

4.3 Stability of Geographical Distribution of Innovation

How geographically concentrated is Indian innovation? Geocoding grantee's and inventors' addresses allows us to analyze the levels and trends in the geography of innovation. First, Figure 4 shows state-level as well as district-level maps of where grantees and inventors are located, by quintile of the occurrences of grantee/inventor appearances in patents. Grantees are usually the organization to which the patent is granted, and inventors are the individuals involved in developing the patent. In terms of aggregate innovation (number of patents), several Southern states are in the top quintile, including Karnataka and Tamil Nadu, though some Central and Northern states are also in the top quintile (Uttar Pradesh and Gujarat). Panels c) and d) show district-level maps, and shows that within states, innovative activity is highly concentrated - for instance, in Maharashtra, only a few districts are responsible for greater innovation like Mumbai and Pune.

How stable is this spatial distribution of innovative activity in recent years? Figure 5 present trends for the number of appearances for the top five states for grantees and inventors, respectively. It shows that the relative ranking of these five states (in order of patenting activity, Maharashtra, Tamil Nadu, Delhi, Karnataka, and Uttar Pradesh) is quite stable for grantees with Tamil Nadu being an exception, going from rank four to two between 2015 and 2018. For inventors, there are more changes in ranks, but Maharashtra is consistently the highest across time. Overall, the relative contribution of these two states seems relatively stable in recent years.

Finally, Figure 6 and Figure 7 presents results for the level and growth in patent

applications for granted patents for both inventors and grantees. In terms of levels, from 2005 to 2018, there is general stability in leading states like Mahaharashtra, Karnataka, Gujarat and Uttar Pradesh, both in absolute number of patents and density. In terms of growth, Andhra Pradesh and Chattisgarh are amongst the top quintile states for patenting growth. Some other states have a high growth rate of grantees, such as Madhya Pradesh, but not of inventors. Overall, these figures indicate that where innovative activity is already high, and where it is growing fast, are different. Furthermore, a few districts drive a large share of innovative activity.

4.4 Business Innovation's Large and Increasing Share of Innovation

Figure 8 shows trends in the composition of which types of organizations are conducting innovative activity, namely academia, firms, government organizations (other than academic organizations and state-owned enterprises), and individual inventors not affiliated with any organization (labeled "natural persons") in the data. The figure shows that across the 2005-2018 time period, firms have been by some distance the largest driver of patenting activity, followed by academic institutions (this considers the first grantee in the patents). Moreover, the growth rate for patenting activity for firms has been higher over time, and in 2018, the number of patents issued to firms is over three times that issued to academia. Third, while patenting activity by firms, academia, and individuals have increased over time, those by other government institutions has been relatively more stable over time. Overall, the trends point to robust and growth patenting activity in both the private sector and academia.

With firms, we can also see the relative contribution of patenting activity between state-owned enterprises such as public sector undertakings vs. other firms (which are predominantly private firms). Figure 9 presents trends in innovation by these two types of firms. The gap between private sector firms and state-owned enterprises has increased over time, with the growth rate for the latter relatively stable. For the former, the number of patent awards grows relatively faster especially after 2009.

4.5 Top Contributors to Innovation by Sector and Trends Over Time

Which sectors in the Indian economy are drivers of innovation in the private sector? To understand this, we look to the five-digit National Industrial Classification (NIC) codes, in the merged patents-Prowess data.[4] We focus on the top ten sectors (by first grantee) by total number of granted patents. Figure 10 presents trends in innovation

[4]This is the finest level at which NIC codes are available in the Prowess database.

over time for these sectors as well as which sectors these are. First, the list of top ten sectors by number of patents is diverse, and shows that innovative activity is occurring in a wide swathe of sectors ranging from pharmaceuticals, motor vehicles, energy, and telecommunications. Second, it shows while that growth has been rapid in several of these sectors, it is quite heterogeneous, with some sectors like software growing faster than others like manufacturing of motor vehicles. Third, some sectors have started with a lower relative ranking but climbed rapidly, such as telecommunications (the NIC code "manufacture of pagers, cellular phones and other mobile communication equipment"), which goes from ranked tenth in 2014 to second in 2018 in terms of number of granted patents.

4.6 The Evolving Nature of Innovation Collaboration Over Time

How have collaboration networks changed in recent decades? We explore this in several ways. First, Figure 11 depicts the share of patents where inventors have identical addresses for patents with more than one inventors. The blue line shows this for patents granted to organizations (which are likely higher-value) while the orange dotted line shows this for patents granted to individuals. In both cases, we see that in recent years (after 2012), there has been a decrease in non-collaborative collaboration i.e. a decrease in the share of patents with identical inventor addresses. The magnitude here is economically significant, with such patents (with identical inventor addresses) comprising about 57% of patents in 2013, but about 44% in 2018 for patents granted to organizations.

Figure 12 shows trends in more detailed types of collaborations over time. The share of patents comprising collaboration among inventors from the same organization has declined over time, especially after 2013. Interestingly, the share of patents from a single inventor has also declined significantly and continuously over time, from over 20% in 2005 to under 15% in 2018. Both these patterns indicate an increase in collaborative innovation in recent years, which is shown with increases in the share of patents attributable to multiple organizations and multiple individuals.

To better understand the types of cross-organizational collaboration, we focus on the top twenty inventing organizations, and plot their networks. Figure 13 shows this network, with the width of the link representing the strength of the collaboration (the number of patents) they have together. The figure highlights several facts about the network of collaboration for these high-innovation organizations. First, while the majority of these organizations, thirteen of twenty, have collaboration networks with other high-innovation organizations, a significant fraction do not, including large private firms such as Maruti

Suzuki, Unilever, and Tata Motors. Second, cross-organization innovation frequently takes the form of private firms collaborating with academia, such as Tata Steel's collaboration with IIT Kharagpur, Mahindra and Mahindra's collaboration with IIT Madras, or Tata Consultancy Services collaborating with the Indian Institute of Science, as well as with several IITs including Kanpur, Bombay, and Kharagpur. Third, this type of academic-private firm collaboration is much more common than private firms collaborating with each other. Overall, the figure highlights the significance of collaborative innovation in India, and shows the importance of frontier academic organizations' participation with the private sector in recent innovation.

4.7 Increasing Trends in Patent Citations

While number of patents is one measure of innovation, a long literature on innovation shows the importance of patent citations as measures of quality. We will focus on this in future work, but for now, note the general increase in patent citations over time depicted in Figure 14. The average number of citations (shown in Panel a) is negligible in 2005 both for Indian and non-Indian patents, but rises rapidly especially after 2007 and stabilizes at about three for non-Indian patents by 2012 and about 2.5 for Indian patents around the same time. Moreover, Panel b shows that the average number of citations *conditional on citing any patent* has remained stable over time, at about 2.7 for Indian patents and 3 for non-Indian patents. However, note that a large number of patents in the early period does not include the FER (First Examination Report) files on IP India Website, making it impossible to determine whether they indeed cited or not.

Whether patent citations are as likely to measure quality in Indian patents as in US patents is not clear, as citation behavior may vary across contexts. Suggestive evidence shows that there is very little activity in the field of citing Indian patents - the top three countries of patents cited by Indian patents are U.S. (46.48%), China (14.10%), and European Patent Office (6.13%), and on average, non-Indian patents cite 0.003 patents filed in the Indian patent office, whereas Indian patents cite 0.048 patents filed in the Indian patent office. By comparison, data from PatentsView for U.S. granted patents applied between 2005 to 2018, on average U.S. patents cite 19.19 U.S. granted patents and 6.73 foreign patents. China is intermediate between the two extremes: patents ten or more years of age have on average 4.5 citations by other Chinese awards, while those granted between 2013 and 2017 have 1.3 citations (Lerner et al., 2025).

5 Conclusion

In this paper, we draw on a newly constructed dataset covering all granted patents in India from 2005 until the present, to uncover key stylized factors about innovation in India. The data show that innovative activity is growing rapidly over time, largely driven by the private sector and academia, and distributed in diverse sectors in the economy. Furthermore, private-sector and academic collaborations are an important source of patenting activity for the top organizations by number of patents, and innovation is becoming more collaborative over time.

In future work, we will test how recent policy changes have changed the rate and type of innovation over time, as well as build newer measures of innovation quality. Given the relative importance of the Indian economy, we also hope this work will be useful for academics and professionals in better understanding innovation landscape in India, and the developing world broadly.

References

Akcigit, Ufuk, Emin Dinlersoz, Jeremy Greenwood, and Veronika Penciakova, "Synergizing Ventures," *Journal of Economic Dynamics and Control*, 2022, *143*, 104427.

Bernstein, Shai, Xavier Giroud, and Richard R. Townsend, "The Impact of Venture Capital Monitoring," *Journal of Finance*, 2016, *71* (4), 1591–1622.

Department of Science and Technology, Ministry of Science and Technology, Government of India, *Science, Technology, and Innovation Policy*, Government of India, 2020. https://dst.gov.in/sites/default/files/STIP_Doc_1.4_Dec2020.pdf.

_ , *Research and Development Statistics at a Glance*, Government of India, 2023. url{https://dst.gov.in/sites/default/files/Updated

Griliches, Zvi, *R&D, Patents, and Productivity*, University of Chicago Press, 1984.

Jaffe, Adam B. and Manuel Trajtenberg, *Patents, Citations, and Innovations: A Window on the Knowledge Economy*, MIT Press, 2002.

Jayaraman, K.S., "US Patent Office Withdraws Patent on Indian Herb," *Nature*, 1997, *389* (6646), 6.

Kalyani, Aakash, Nicholas Bloom, Marcella Carvalho, Tarek A. Hassan, Josh Lerner, and Ahmed Tahoun, "Diffusion of New Technologies," *Quarterly Journal of Economics*, 2025, *140*, forthcoming.

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy, "Measuring Technological Innovation over the Long Run," *American Economic Review: Insights*, 2021, *3* (3), 303–320.

Kortum, Samuel and Josh Lerner, "Assessing the Contribution of Venture Capital to Innovation," *RAND Journal of Economics*, 2000, *31* (4), 674–692.

Lerner, Josh, Junxi Liu, Jacob Moscona, and David Yang, "Appropriate Entrepreneurship? The Rise of China and the Developing World," Working Paper 32193, National Bureau of Economic Research 2024.

_ **, Namrata Narain, Dimitris Papanikolaou, and Amit Seru**, "Creating the Chinese Patent Dataset," Unpublished Working Paper, Harvard University 2025.

Magerman, Tom, Bart Van Looy, and Xiaoyan Song, "Data Production Methods for Harmonized Patent Statistics: Patentee Name Harmonization," SSRN Scholarly Paper 944470 2006.

Myers, Kyle R. and Lauren Lanahan, "Estimating Spillovers from Publicly Funded R&D: Evidence from the US Department of Energy," *American Economic Review*, 2022, *112* (7), 2393–2423.

Putnam, Jonathan, "The Value of International Patent Rights," Unpublished Working Paper, Yale University 1997.

Tables and Figures

Table 1: Top Patent Grantees in India

Rank	Organization	Type	#Patents
1	Council of Scientific and Industrial Research	Academic Institutions	2154
2	Bharat Heavy Electricals	Business	1779
3	TVS Motor	Business	1679
4	Tata Service	Business	1635
5	Samsung Electronics	Business	1186
6	Mahindra & Mahindra	Business	1100
7	Samsung R&D Institute India Bangalore	Business	1065
8	Tata Steel	Business	1032
9	Tata Motors	Business	1029
10	Indian Institute of Technology Bombay	Academic Institutions	944
11	Indian Institute of Technology Madras	Academic Institutions	925
12	Schneider Electric India	Business	822
13	Wipro	Business	721
14	Director General of Defense R&D Organisation	Government	649
15	Maruti Suzuki India	Business	624
16	Indian Institute of Technology Delhi	Academic Institutions	535
17	Hindustan Unilever	Business	517
18	Indian Institute of Technology Kanpur	Academic Institutions	484
19	Indian Institute of Science	Academic Institutions	436
20	Reliance Industrial	Business	429

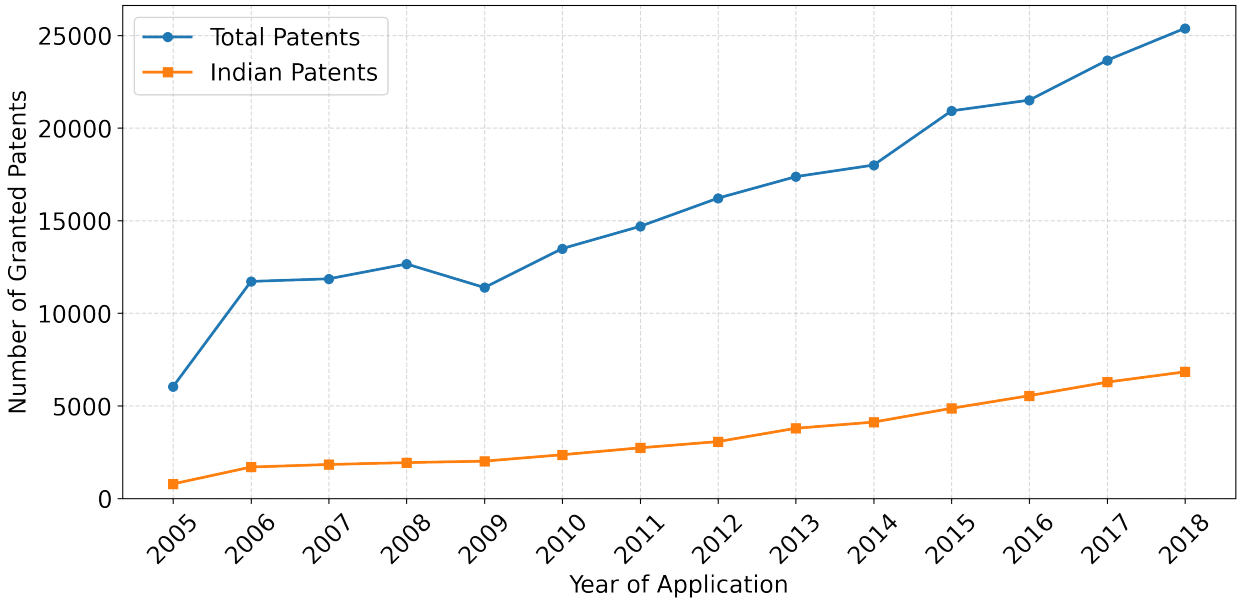
Notes: This table reports the top patent grantees for Indian patents. A patent is defined as an Indian patent if either at least one of the inventors or at least one of the grantees has an address in India. Number of patents calculates the total number of patents that a given grantee appears in the list of grantees, not necessarily the first grantee.

Table 2: Inventor and Grantees by State

State	Inventors	Inventors (Pct)	Grantees	Grantees (Pct)	Patents (Pct)	Population
Maharashtra	40,672	22%	16,490	25%	22%	112,374,333
Karnataka	28,341	15%	6,982	11%	15%	61,095,297
Tamil Nadu	30,517	16%	9,509	14%	14%	72,147,030
Uttar Pradesh	11,125	6%	3,413	5%	6%	199,812,341
Telangana	11,080	6%	2,690	4%	6%	35,003,674
Delhi	8,168	4%	8,632	13%	5%	16,787,941
Gujarat	8,203	4%	2,839	4%	4%	60,439,692
Haryana	9,268	5%	1,781	3%	4%	25,351,462
West Bengal	6,241	3%	3,408	5%	4%	91,276,115
Kerala	4,094	2%	1,005	2%	2%	33,406,061
Jharkhand	6,036	3%	1,379	2%	2%	32,988,134
Punjab	4,239	2%	1,232	2%	2%	27,743,338
Andhra Pradesh	2,592	1%	999	2%	2%	84,580,777
Rajasthan	2,349	1%	852	1%	2%	68,548,437
Madhya Pradesh	2,606	1%	725	1%	2%	72,626,809
Uttarakhand	3,546	2%	771	1%	2%	10,086,292
Odisha	1,936	1%	625	1%	1%	41,974,218
Bihar	962	1%	347	1%	1%	104,099,452
Assam	1,501	1%	612	1%	1%	31,205,576
Chandigarh	1,114	1%	540	1%	1%	1,055,450
Himachal Pradesh	1,149	1%	384	1%	1%	6,864,602
Chhattisgarh	527	0%	265	0%	0%	25,545,198
Puducherry	489	0%	208	0%	0%	1,247,953
Jammu & Kashmir	636	0%	168	0%	0%	12,541,302
Goa	399	0%	110	0%	0%	1,458,545
Tripura	61	0%	30	0%	0%	3,673,917
Manipur	48	0%	39	0%	0%	2,855,794
Meghalaya	84	0%	37	0%	0%	2,966,889
Nagaland	57	0%	33	0%	0%	1,978,502
Dadra & Nagar Haveli & Daman & Diu	37	0%	13	0%	0%	586,956
Mizoram	38	0%	30	0%	0%	1,097,206
Sikkim	29	0%	22	0%	0%	610,577
Arunachal Pradesh	17	0%	10	0%	0%	1,383,727
Andaman & Nicobar Islands	24	0%	8	0%	0%	380,581
Total	188,185	100%	66,188	100%	100%	1,245,794,178

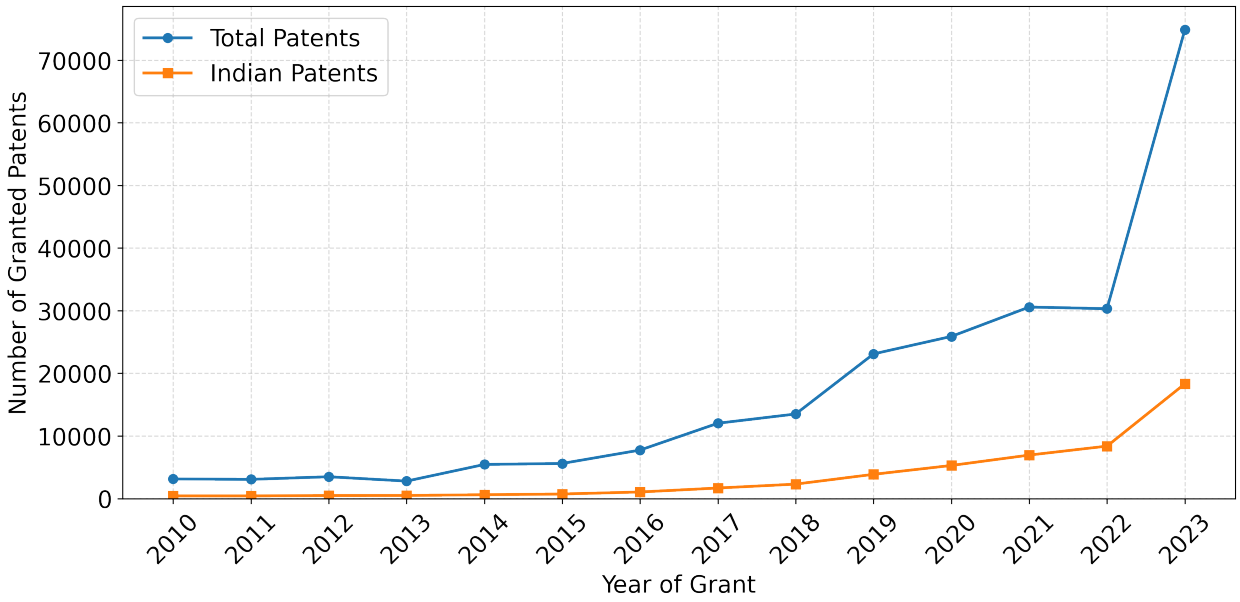
Notes: Population information provided by 2011 Census https://censusindia.gov.in/census.website/data/census-tables. For Telangana, population retrieved from state website https://www.telangana.gov.in/About/State-Profile/.

Figure 1: Number of Patent Awards in India, by Application Year



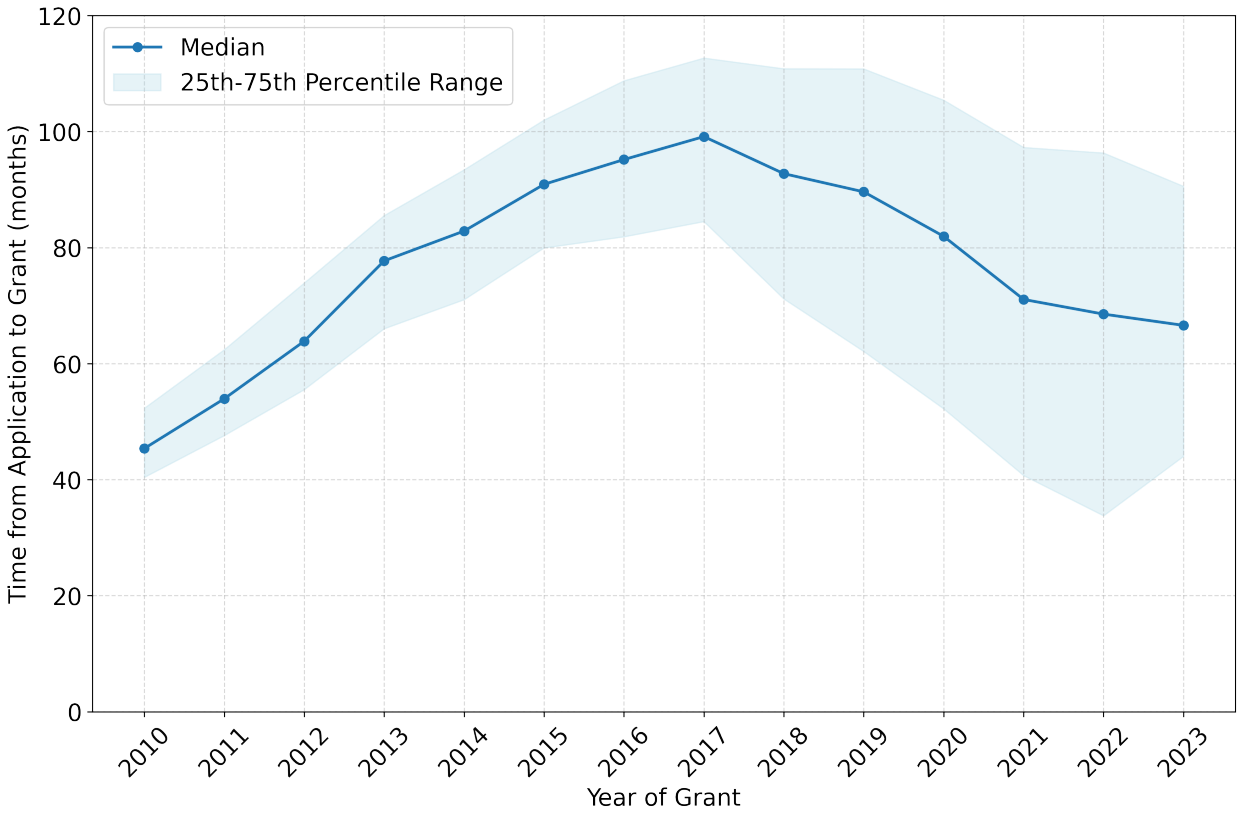
Note: This figure plots the number of patent awards by patents' application year. Indian patents are defined as either (1) patents with at least one assignee whose address is in India; or (2) patents with at least one inventor whose address is in India. We omit the years of 2018 to 2023 in this figure, since only a fraction of patents applied in these years have been examined - the median wait time between application to grant in India is over five years.

Figure 2: Number of Patent Awards in India, by Grant Year



Note: This figure plots the number of patent awards by the year of grant. Indian patents are defined as either (1) patents with at least one assignee whose address is in India; or (2) patents with at least one inventor whose address is in India. We omit the years from 2005 to 2010 since we have only scraped patents applied from 2005, and the time between application to grant means that we do not have complete coverage for the early years yet.

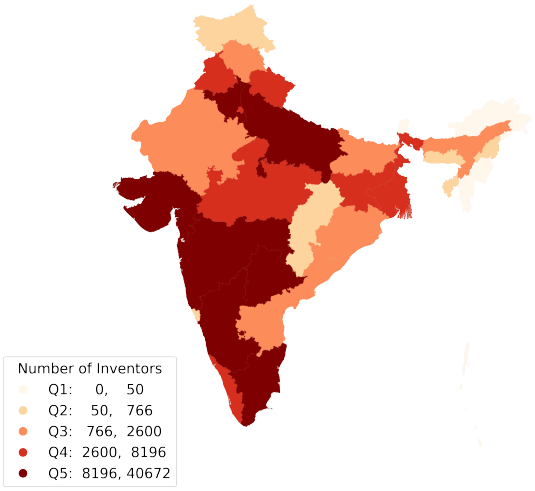
Figure 3: Time from Application to Granted Patents, by Grant Year



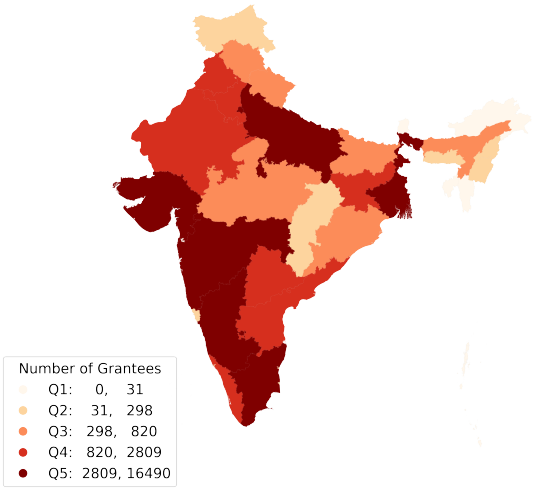
Note: This figure plots the median and inter-quartile range of time (in months) from patent application date to patent grant date for Indian patents, grouped by year of grant. Indian patents are defined as either (1) patents with at least one assignee whose address is in India; or (2) patents with at least one inventor whose address is in India.

Figure 4: Spatial Distribution of Innovation in India

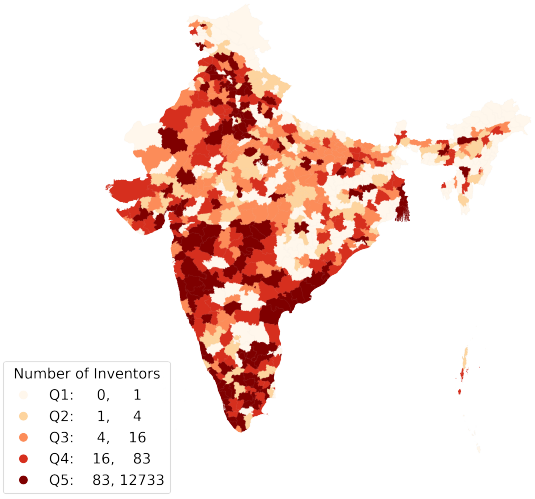
(a) Inventors by State



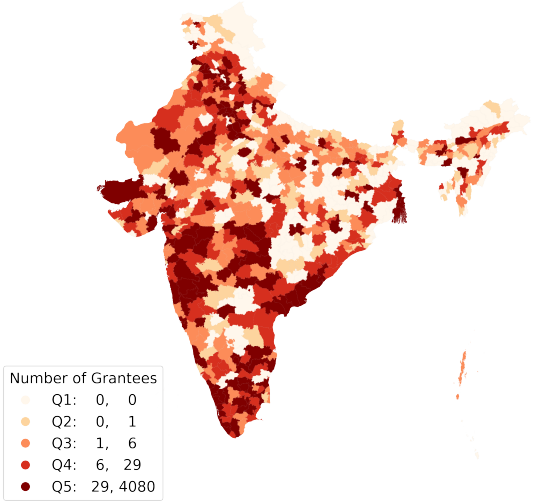
(b) Grantees by State



(c) Inventors by District



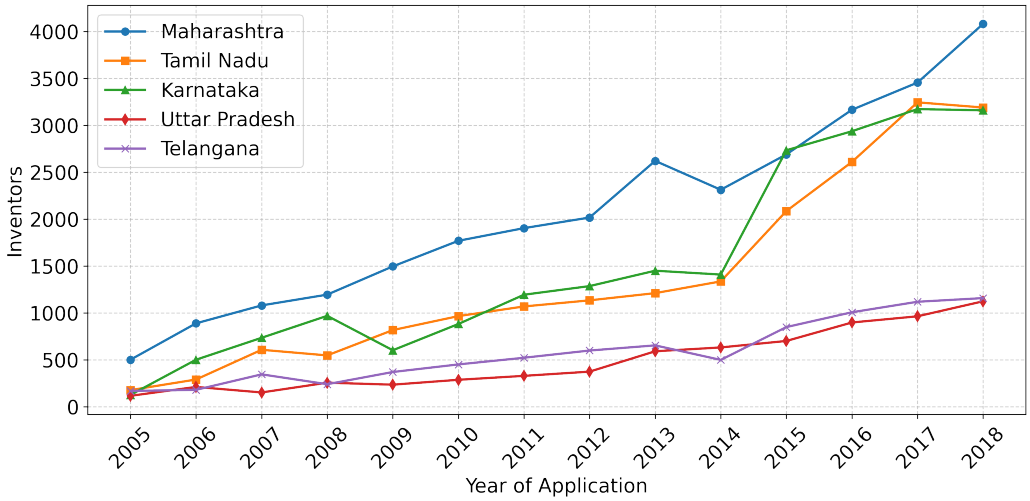
(d) Grantees by District



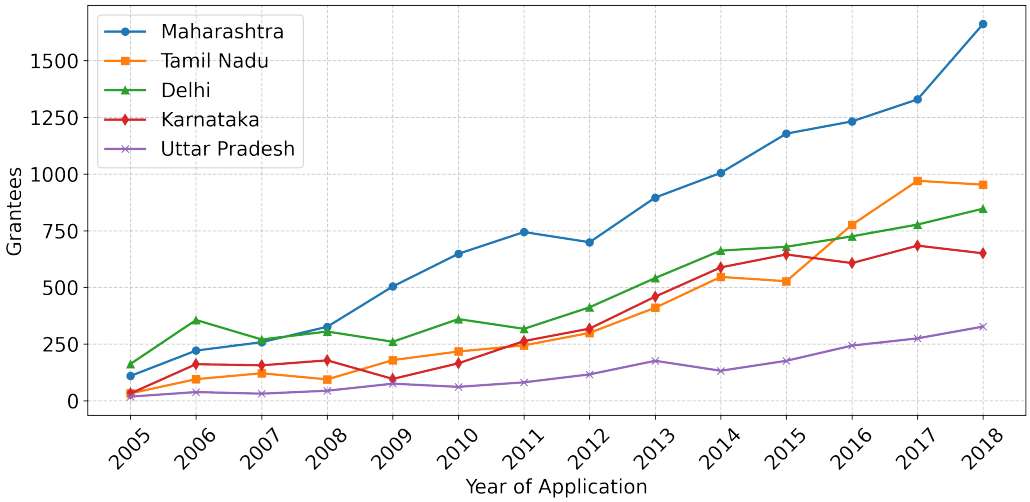
Note: This figure plots the distribution of inventors and grantees by state and district in quintiles for Indian patents from 2005 to 2022. The legend denotes the range of inventors and grantees in each quintile. Each observation is an inventor/grantee-patent pair. Indian patents are defined as either (1) patents with at least one assignee whose address is in India; or (2) patents with at least one inventor whose address is in India.

Figure 5: Number of Patent Inventors and Grantees by State

(a) Inventors



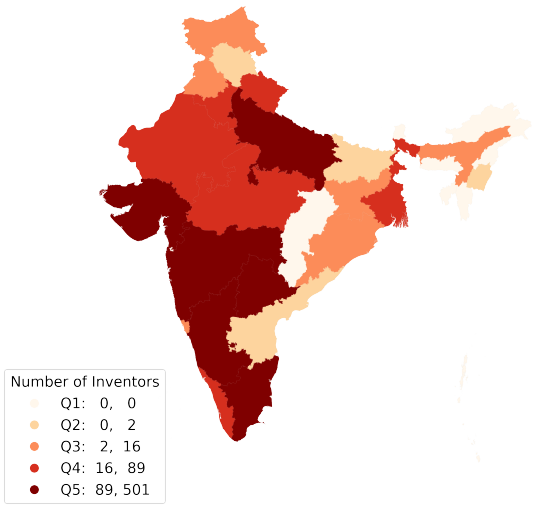
(b) Grantees



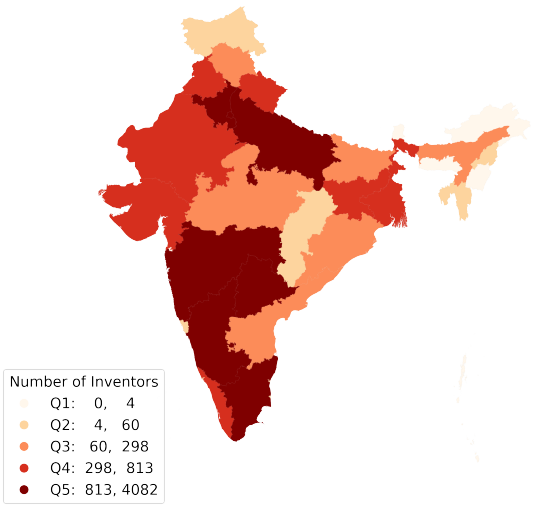
Note: This figure plots the number of patent inventors and grantees of Indian patents by state in the top 5 states between 2005-2018. The x-axis represents the application filing year. The legend presents the states in descending order of total inventors. Each observation is an inventor/grantee-patent pair. Indian patents are defined as either (1) patents with at least one assignee whose address is in India; or (2) patents with at least one inventor whose address is in India. We omit the years of 2018 to 2023 in this figure, since only a fraction of patents applied in these years have been examined - the median wait time between application to grant in India is over 5 years.

Figure 6: Spatial Distribution of Innovation in India, 2005 and 2018

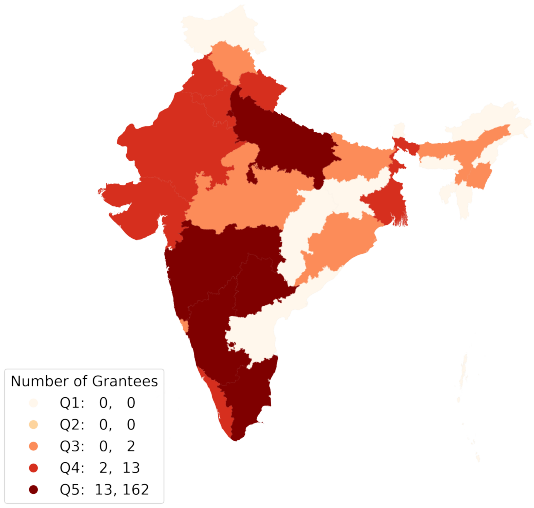
(a) Density of Inventors by State (2005)



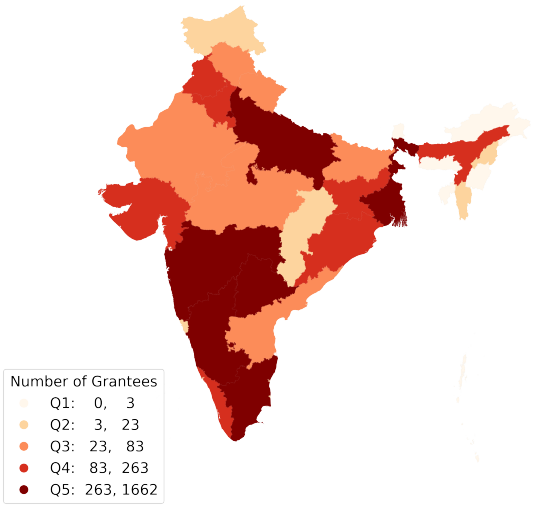
(b) Density of Inventors by State (2018)



(c) Density of Grantees by State (2005)



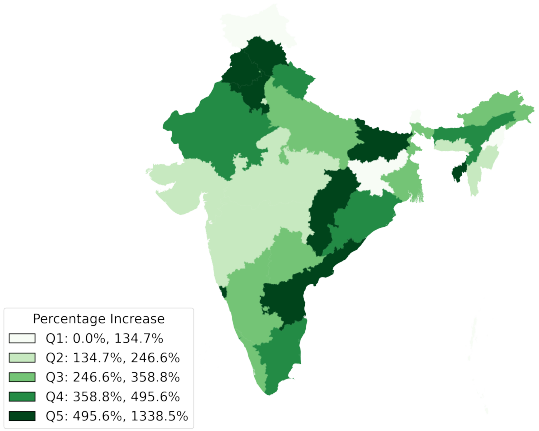
(d) Density of Grantees by State (2018)



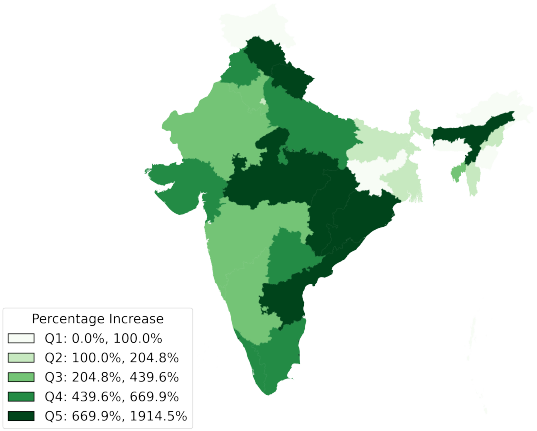
Note: This figure plots the distribution of inventors and grantees by state in application filing years 2005 and 2018 for Indian patents. The legend denotes the range of inventors and grantees in each quintile. Each observation is an inventor/grantee-patent pair. Indian patents are defined as either (1) patents with at least one assignee whose address is in India; or (2) patents with at least one inventor whose address is in India. We omit the years of 2018 to 2023 in this figure, since only a fraction of patents applied in these years have been examined - the median wait time between application to grant in India is over 5 years.

Figure 7: Spatial Growth in Patent Applications from 2005 to 2018

(a) Growth of Inventors by State

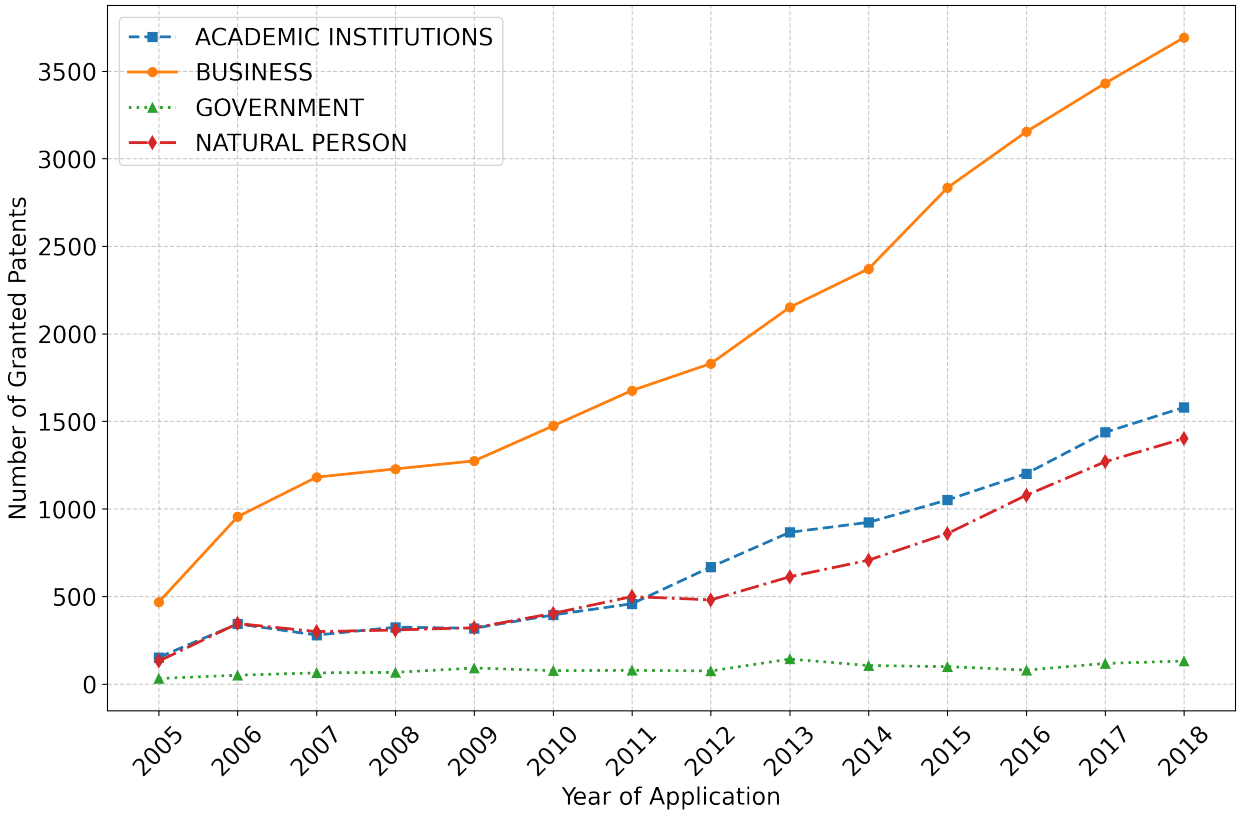


(b) Growth of Grantees by State



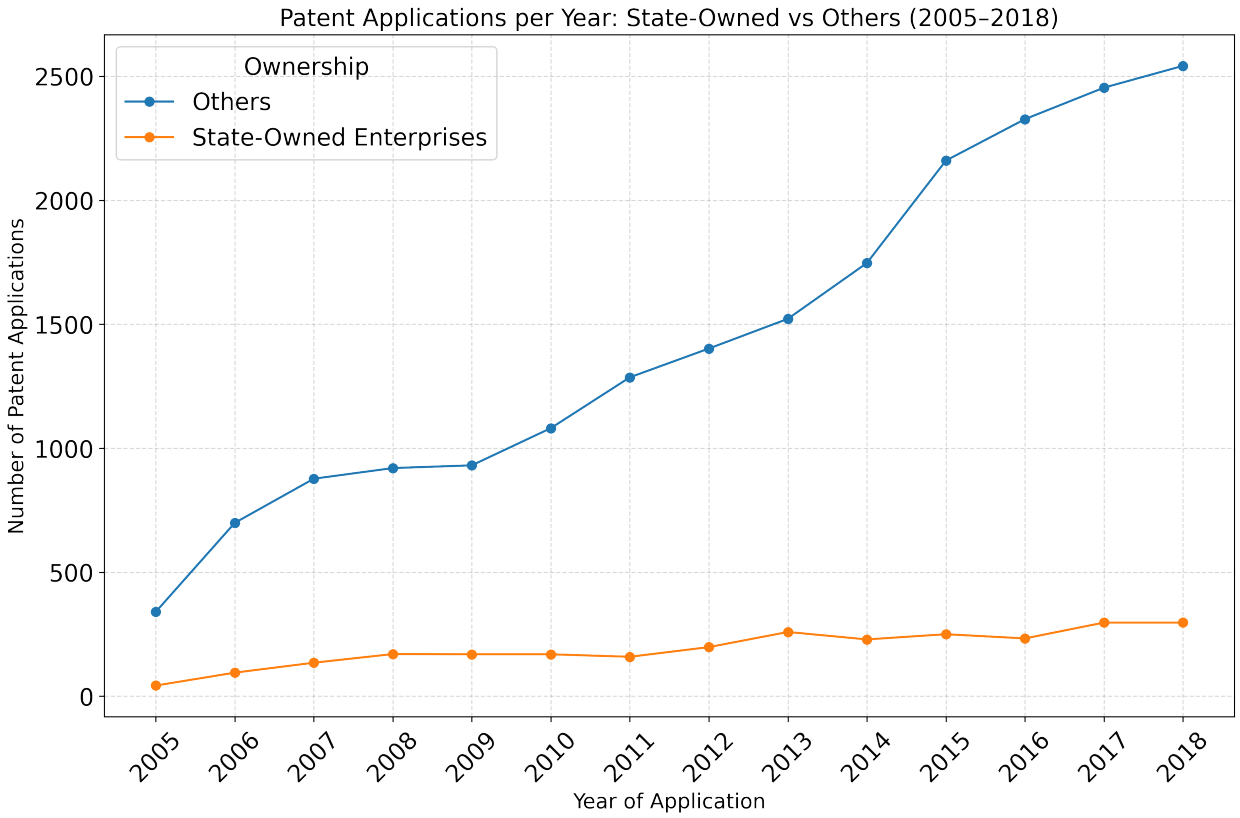
Note: This figure plots the quintiles distribution in the growth rate of inventors and grantees by state for Indian patents. We calculate the growth rate from the average number of patents in the first five years of our data (2005-2009) to the average in the last five years (2014-2018). States are classified into quintiles based on growth, and states with an average of 0 in the first five years are treated as 1 in the growth calculation. The legend denotes the range of growth in each quintile. Each observation is an inventor/grantee-patent pair. Indian patents are defined as either (1) patents with at least one assignee whose address is in India; or (2) patents with at least one inventor whose address is in India. We omit the years of 2018 to 2023 in this figure, since only a fraction of patents applied in these years have been examined - the median wait time between application to grant in India is over 5 years.

Figure 8: Number of Indian Patent Awards, by First Grantee Type and Application Year



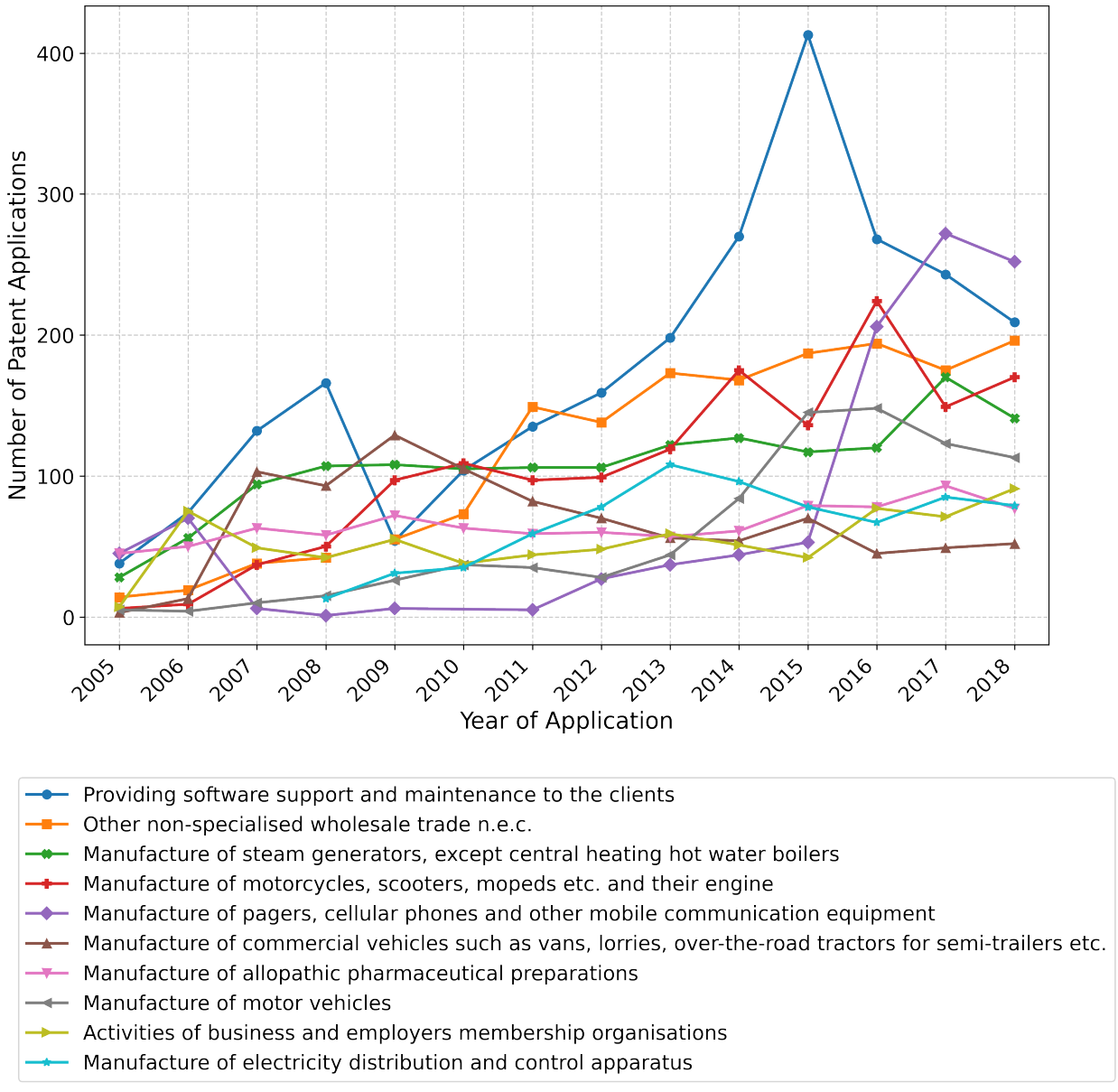
Note: This figure plots the number of Indian patent awards grouped by the type of each patent's first grantee and by the year of application. We describe how we classify grantees into the four groups in Appendix D. Indian patents are defined as either (1) patents with at least one assignee whose address is in India; or (2) patents with at least one inventor whose address is in India. We omit the years of 2018 to 2023 in this figure, since only a fraction of patents applied in these years have been examined - the median wait time between application to grant in India is over 5 years.

Figure 9: Number of Indian Patent Awards, by First Grantee's Prowess Ownership



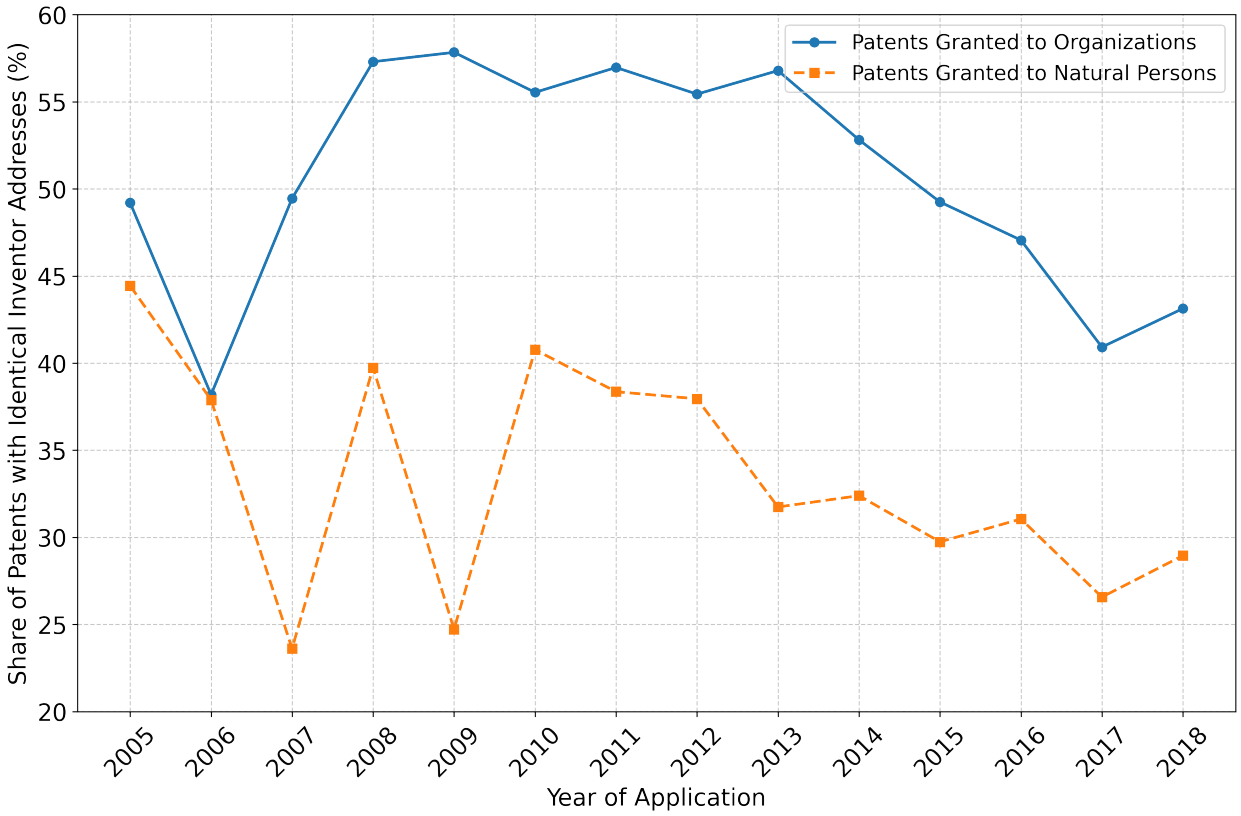
Note: This figure plots the number of Indian patent awards grouped by the ownership of each patent's first grantee, if the first grantee is matched to Prowess data. Indian patents are defined as either (1) patents with at least one assignee whose address is in India; or (2) patents with at least one inventor whose address is in India. We define State-Owned Enterprises if the "owner_gp_name" variable in the Prowess data contains the word "Govt" or "State". We describe how we match patent grantees to Prowess company names in Appendix G. We omit the years of 2018 to 2023 in this figure, since only a fraction of patents applied in these years have been examined - the median wait time between application to grant in India is over 5 years.

Figure 10: Number of Indian Patent Awards, by First Grantee's Prowess NIC Product Code



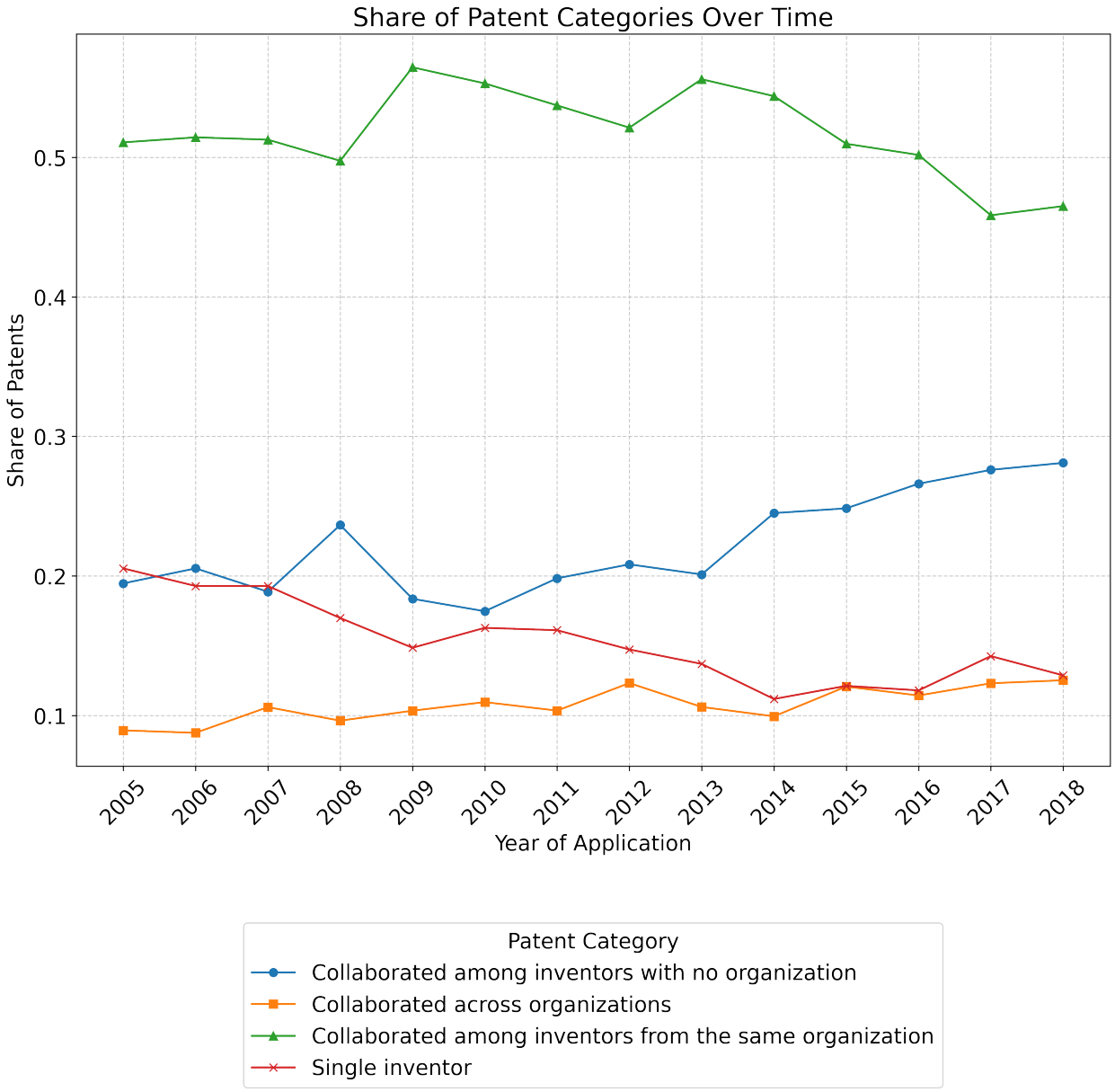
Note: This figure plots the number of Indian patent awards grouped by the NIC (National Industrial Classification) of each patent's first grantee, if the first grantee is matched to Prowess data. Indian patents are defined as either (1) patents with at least one assignee whose address is in India; or (2) patents with at least one inventor whose address is in India. We use Prowess' "nic_name" variable and list the top 10 groups. We describe how we match patent grantees to Prowess company names in Appendix G. We omit the years of 2018 to 2023 in this figure, since only a fraction of patents applied in these years have been examined - the median wait time between application to grant in India is over 5 years.

Figure 11: Share of Patents with Identical Inventor Addresses



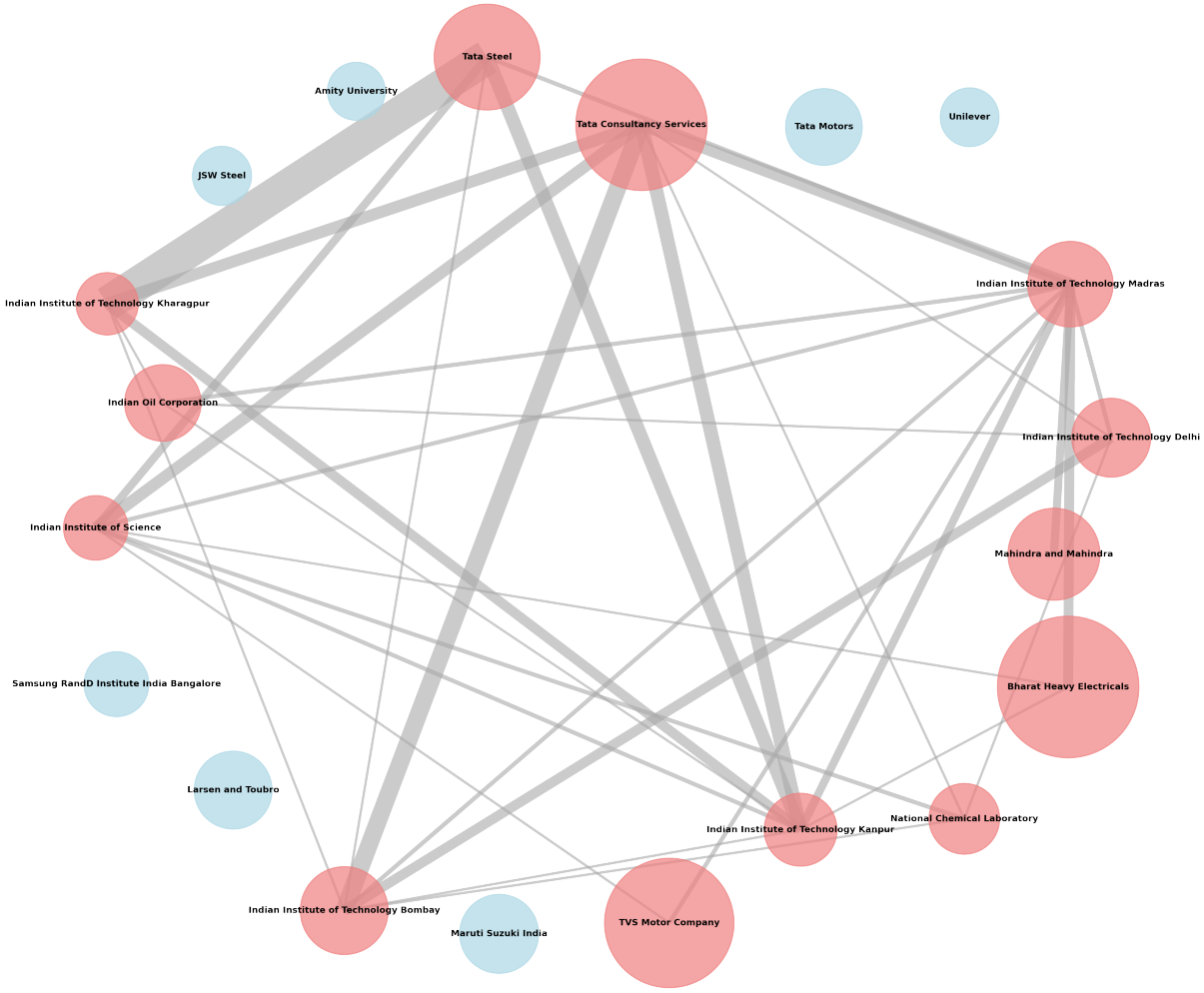
Note: This figure plots the share of patents with identical inventor addresses, grouped by inventor collaboration types and year of application. Indian patents are defined as either (1) patents with at least one assignee whose address is in India; or (2) patents with at least one inventor whose address is in India. We omit the years of 2018 to 2023 in this figure, since only a fraction of patents applied in these years have been examined - the median wait time between application to grant in India is over 5 years.

Figure 12: Collaboration Types for Patents Granted to Organizations



Note: This figure plots the share of patents granted to organizations (i.e. non-natural persons) grouped by inventor collaboration types and year of application. We describe how we infer inventors' organization in Appendix F. Patents are classified into "Collaborated across organizations" if at least two of the inventors are from different organizations, or at least one of the inventors are from a certain organization and others are without organization. Indian patents are defined as either (1) patents with at least one assignee whose address is in India; or (2) patents with at least one inventor whose address is in India. We omit the years of 2018 to 2023 in this figure, since only a fraction of patents applied in these years have been examined - the median wait time between application to grant in India is over 5 years.

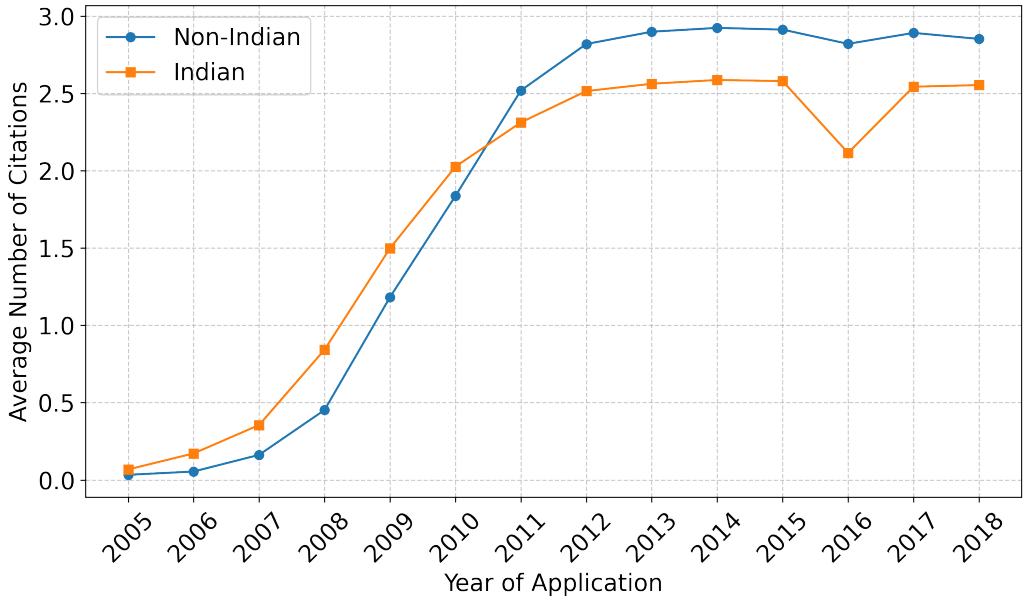
Figure 13: Collaboration Network



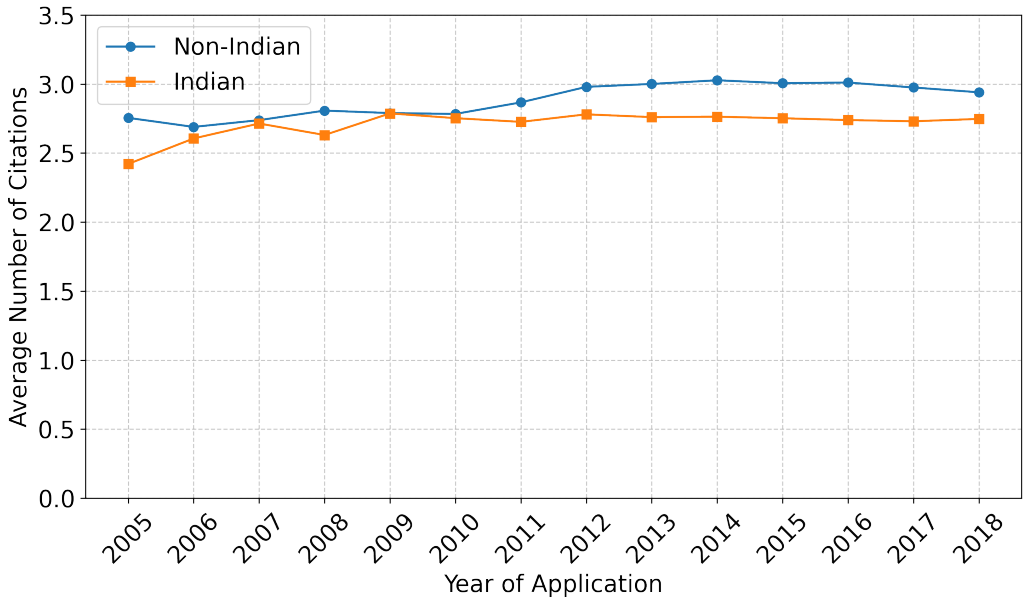
Note: This figure plots the collaboration network for the top 20 inventor organizations. The size of the node represents the total number of patents involving inventors from this organization, and the width of the link represents the number of collaborations between inventors from the two organizations. We document how we extract inventor organization information from inventor address in Appendix F. Blue nodes represents singleton organizations while red nodes represents organizations with collaborations. A collaboration is defined as at least one inventor from each organization appear on the inventors list for a given patent. Indian patents are defined as either (1) patents with at least one assignee whose address is in India; or (2) patents with at least one inventor whose address is in India.

Figure 14: Trend in Citations

(a) Average Number of Citations Made, All Indian Patents



(b) Average Number of Citations Made, Conditional on Cited Any Patents



Note: This figure plots the average number of citations made for all Indian patents and Indian patents conditional on having cited any patent. We describe how we extract backwards citation information from patents' associated files in Appendix I. Patents filed in the early years tend not to include the file documenting citation information. Indian patents are defined as either (1) patents with at least one assignee whose address is in India; or (2) patents with at least one inventor whose address is in India. We omit the years of 2018 to 2023 in this figure, since only a fraction of patents applied in these years have been examined - the median wait time between application to grant in India is over 5 years.

Appendix

A Data Source

We search in IP India's portal at https://iprsearch.ipindia.gov.in/publicsearch. Each patent has three main sources of information from the IP India website: Details, Register, and Status. An example is as follows:

Application Number	Title	Application Date	Status		
202041004552	GLASS SHUTTER FRAME ASSEMBLY WITH SEAMLESS FRAME INTEGRATION DESIGN	01/02/2020	Granted	E-Register	Application Status

We obtain various information from their respective websites, and the variable of interest has a prefix of "Details/", "Register/", or "Status/". Details page refers to the patent detail information page when clicking on the application number. Register page refers to the E-Register providing information on register and grantees. Status page refers to the Application Status page. We also download all attached documents for each patent.

To determine the application year of each patent, we primarily use the variable 'Details/Application Filing Date' from the web-scraped data. If this field is missing or contains erroneous values (e.g. 1900), we use information from application ID: either the first four digit (where the application ID is all digits) or the last four digit (where the application ID follows the format of DD-OFFICE-YYYY) represents the year that the application is filed.

B Identifying Indian Patents

To capture the localized Indian innovation, we include the following patents in our Indian Patents Dataset: (1) patents with at least one grantee whose address is in India; or (2) patents with at least one inventor whose address is in India. We identify 67,369 Indian patents in our dataset from 295,743 patents we scraped.

The India Patent Office records the nationality and country of inventors, so the determination of an Indian inventor is more straightforwards. The listed "inventor country" information is associated with the address of the inventor. We consider a patent to have an Indian inventor as long as one of the inventors' "inventor country" is India. In total, there are 64,070 unique patents with an Indian inventor.

Grantees only have a "grantee address" field to document address but not following any specific format. To determine whether a grantee address is Indian, we used a combination of mapping address strings to India directly (through the named country or Indian states), filtering out addresses clearly based in another country, and name entity recognition (NER) via a Python package. First, we utilize the locationtagger (NER) Python package to broadly extract country and related entities to location keywords in address string. Then, we determine an address is Indian if it contains a standalone "India" or any states of India (53,930 addresses). We determine an address is not Indian if it ends with any other country names, or states/provinces of U.S. and China (142,751 addresses). Among the remaining 128,819 addresses, 15,060 mapped to India via locationtagger, and 113,759 do not. We

consider the latter as not Indian. Overall, we determined that 56,577 unique patents with an Indian grantee.

Combining the two sets (Indian inventors and Indian grantees), we obtain our Indian Patent Dataset, consisting of 67,369 patents.

C Disambiguation

We closely follow PatentsView's disambiguation process on U.S. patents' assignees. Details can be found on PatentsView's website at https://patentsview.org/disambiguation. Our final pipeline proceeds in two main phases. In Phase One we closely follows PatentsView's methodology in disambiguation. We first create "mentions" of grantees and aggregate processed mentions into what PatentsView call "canopies" using the first four character of the name to reduce the total number of comparisons needed. Next, we apply PatentsView's Hierarchical Agglomerative Clustering (HAC) clustering model with TF-IDF and location features through the Grinch library. Records are clustered based on whether they exceeds the threshold that PatentsView's code use (0.85) in terms of cosine similarity of these engineered features.

Phase Two re-checks each cluster with similarity metrics after removing both "Stopwords" and "Industry" stopwords. In addition to PatentsView's clustering model, we introduced an additional verification layer based on four distinct similarity metrics—BERT cosine similarity, TF-IDF cosine similarity, Jaro-Winkler distance, and Levenshtein distance. If at least two of these metrics register above a 0.9 threshold, we accept that cluster.

We applied this approach to an Indian patents dataset of 11,236 raw Indian organization and enterprises grantee names. After cleaning and disambiguation, we ended with 7,130 distinct organization and enterprises entities.

D Identifying Grantee Types

In the web-scraped data from IP India each grantee is indeed associated with a field called "Grantee Type". This field takes one of the following values: Natural Person, Other Than Natural Person, Educational Institute, Small Entity, and Startup. However, upon further investigation, this field is defined very inconsistently.

To systematically determine each grantee's type, we conduct keyword matching process for all grantees that are not Natural Person. We start off with a probable set of keywords that identifies an educational institute, for instance "university", "indian institute of", and check the remaining grantees sorted by their total amount of patents granted to add more keywords to the list.

Among the disambiguated grantees, we are able to identify 1,497 academic institutions, 79 government organizations, and 5,553 commercial firms.

E Geolocation

E.1 Postcode Extraction

Indian postcodes are formatted in six digits, typically at the end of the address. They are either given as a full string (e.g., 123456) or delimited with a space in the middle (e.g., 123 456). We devise a regex command to extract postcodes based on these patterns from the Indian grantee and inventor addresses.

Out of 68,990 Indian grantees, we extracted postcodes from 60,050 (87%), leaving 8,940 (13%) without a postcode. Out of 196,064 Indian inventors, we extracted postcodes from 179,985 (92%), leaving 16,079 (8.2%) without a postcode. Those that did not have a postcode extracted were simply missing a postcode from their address. For example, "Perundurai, Erode, Tamilnadu, India" and "Anusandhan Bhawan, 2 Rafi Marg, New Delhi, Delhi" are addresses that are missing postcode but contain district and state information. For these addresses, we proceeded with district extraction, outlined in the next section.

E.2 District Extraction

Of the 13% grantees and 8% inventors who did not have a postcode extracted, we attempted to extract district information. First, we take the ADM2 Indian district shapefile as an official list of Indian districts and their associated states. Then, we import and clean the inventor and grantee addresses by removing line breaks and changing the address to lowercase. Next, we match the list of districts to the addresses by searching for the full district name in the string. The district name must be on its own. Then, we re-merge the ADM2 list to obtain the state corresponding to the matching district, if any.

We repeat these steps for the ADM3 Indian subdistrict shapefiles, as we noted some addresses are missing the district but contain the subdistrict name. The final dataset coalesces the ADM2-matched district and state with the ADM3-matched subdistrict's corresponding district and state. So if a district was matched, we use that district and state. Otherwise, we take the ADM3-matched district and state.

F Identifying Inventor Organizations

Since inventor address field is free-form texts that do not follow any templates or patterns, we leverage OpenAI's latest o3 model to extract the most likely organization from each patent inventor's address and return well-formatted full name of the organization. For cases which do not indicate any organizations, we ask the LLM to return "NA". In addition, if an address does not include specific organization name, we ask the LLM to infer based on its knowledge where possible.

We carry out basic disambiguation to the returned organization names in similar procedures as we do in the grantee disambiguation process, including removing suffixes like "limited", replacing "&" with "and". In the end, we obtain 6,775 unique inventor organizations. We conduct a random check on 100 results and only find 1 error among them, yielding a success rate of 99%.

G Matching with Prowess

Prowess is the industry-leading firm-level data on firms registered in India. We intend to link patent grantees to this rich dataset to obtain more information on firms.

Since Prowess data only includes commercial firms, we only match on the set of grantees that are identified as "Business" in AppendixD. We began matching grantee names to Prowess by applying the same systematic cleaning and disambiguation procedures that we originally used for patent grantees to the Prowess firm names. For any patent grantees who remained unmatched after these steps, we sorted them by their total number of granted patents and manually attempted to locate the best match within Prowess. We manually check all grantees with more than 10 patents granted by searching through all possible variation of the firm name.

In the end, we are able to match 1,239 firm grantees to the Prowess data, which are granted a total of 29,387 patents, representing 78.5% of the total patents.

H Parsing Complete Specification

"Complete specifications" are the main descriptive body of patents. They usually include the background, summary, and detailed description of the embodiments of the patents.

We initially use the complete specification text gathering from the IP India website. After parsing, we conduct several quality checks. We consider the provided complete specification to be invalid if the value is n/a or if the length of the text is under 1000 characters (most common cases are text stating "the complete specification is attached"). This round of processing generates 249,144 success cases. We also manually check a random sample of 50 results and all of them are parsed correctly.

To resolve the remaining patents, we directly use the attached documents of each patents to parse the complete specifications and claims using OCR package pytesseract (due to cases of non-digitized complete specifications). We also apply the same length and keywords validation as in the first round. This step generates 37,373 success cases. We manually check a random sample of 50 results and all of them are parsed correctly. For a small number of cases, the complete specification is actually in a file with a wrong name (most commonly, in the "Claims" file), or an unusual name. Therefore, for the remainder of patents, we further expand our search to all files whose names contain "claim", "form 2", or "desc" for that application. This step generates 3,315 success cases. We also manually check a random sample of 50 results and all of them are parsed correctly.

In total, we are able to identify 289,832 patents' complete specification information, representing 97.8% of all patents.

I Parsing Backwards Citations

First Examination Report (FER) of each patent documents the literature that this patent cites, which is issued by the Patent Office after review and outlines any potential issues or objections they have regarding the application's patentability based on their initial

examination. We read each patent's FER file and parse information on citations to patent literature. The extraction process relies on Python package regex to identify two specific patterns within the FER files: (1) the marker indicating the start of the patent citation section (2) the publication date format. Cited patent IDs are located between these two patterns. After rounds of experiments, we also include more variants of such patterns in our regular expression matching. We conduct backwards citations extraction on our entire set of patents.

To extract cited patent ID, we employ several cleaning steps. First, we remove formatting characters like "D1:", "D1-"; second, we extract the leading two characters of the cited patent ID, which is usually the patent office identifier of the cited patent (e.g. US); third, for the set of patents where the two leading characters are "WO" (international patent office), we try to match the cited patent ID against US patent IDs to see if there are any potential matches, and for cases there is a square brackets at the end we identify that country to be the patent's country of origin; fourth, for the set of Indian patents, we conduct another round of cleaning including removing suffixes and prefixes, to try to match them to our Indian patent dataset. In the end, we are able to identify patenting country for 99.10% cited patents.

As Figure 14 demonstrates, the average citations grouped by application year increases over time, but this is mostly due to the fact that patents in early periods tend not to include patents. We do not observe the number of citations increase over time if conditional on citing any patents.

We further extract each cited patent ID and infer the country of each cited patents (see Appendix I for further details). The top three countries of patents that Indian patents cite are U.S. (46.48%), China (14.10%), and European Patent Office (6.13%), whereas the top three countries that non-Indian patents cite are U.S. (43.79%), Japan (12.68%), and European Patent Office (9.51%). As expected, citations made to patents filed in the Indian patent office are very scarce, although we do see Indian patents tend to cite more: on average, non-Indian patents cite 0.003 patents filed in the Indian patent office, whereas Indian patents cite 0.048 patents filed in the Indian patent office.

NATIONAL COUNCIL OF APPLIED ECONOMIC RESEARCH

11 Indraprastha Estate, New Delhi 110002. INDIA.
Tel: +91-11-2345 2698, 6120 2698 www.ncaer.org

NCAER | Quality. Relevance. Impact.