

// Philipp Boeing (Goethe University Frankfurt and ZEW Mannheim)
Loren Brandt (University of Toronto)
Ruochoen Dai (Central University of Finance and Economics)
Kevin Lim (University of Toronto)
Bettina Peters (ZEW Mannheim and University of Luxembourg)

The Anatomy of Chinese Innovation: Insights on Patent Quality and Ownership

China's patenting activity has surged over the past two decades, yet questions remain about the quality and sources of innovation. We develop a new method to measure the importance of a patent for innovation, based on the use of a Large Language Model to process patent text data and a new theory of the innovation process. We apply this method to study the evolution of patenting in China from 1985 until recently, and also classify patent ownership using a comprehensive business registry. Our method and data yield several novel facts about Chinese patenting. Among these are that the patents which are important for innovation have become less important on average; that knowledge within China has become more important than knowledge outside of China for directing innovation in China; and that knowledge produced by Chinese entities has been more important than knowledge produced by foreign entities in China. These findings have implications for China's growth trajectory and reflect both global trends in the decline of innovativeness and potential effects of domestic policy.



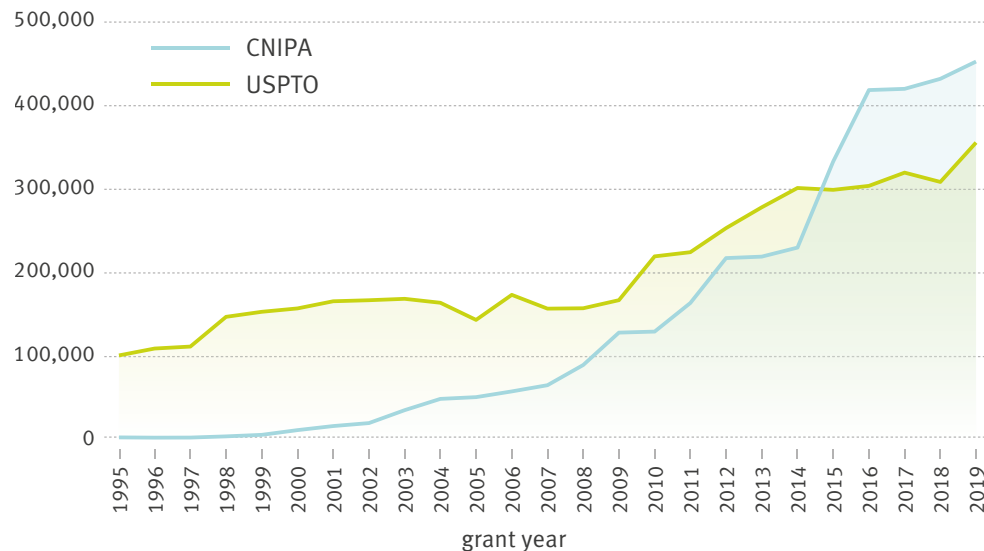
KEY MESSAGES

- Chinese patenting has become narrower and less innovative over time.
- The role of overseas knowledge has also declined sharply.
- These findings are salient in the context of a marked slowdown in economic growth in China and rising concerns of technological decoupling with the US.

Patents are an important indicator of technological change and innovation (Griliches, 1990). In China, patenting activity has accelerated substantially over the last two decades, far outpacing growth in countries like the US (see Figure 1). In [new research](#), we leverage frontier methods in machine learning, a statistical model of innovation, and unique data on patent ownership to investigate how the quality of Chinese patenting has changed over time and quantify which sources of knowledge have been important for driving innovation in China.

Our analysis leads to several novel insights. First, patents that are important for innovation in that they shift the knowledge frontier have become much less important on average, reflecting an overall decline in innovativeness and increase in crowdedness of patenting in China. Second, there is positive growth in the number of breakthrough patents in China (those with the greatest importance), but this growth has declined steadily over time. Third, knowledge within China has become more important than knowledge outside China for driving the direction of innovation in China. Finally, knowledge produced by Chinese entities within China has become more important than knowledge produced by foreign entities within China. These findings are salient in the context of a marked slowdown in economic growth in China and rising concerns of technological decoupling with the US (Han et al., 2024).

FIGURE 1: GRANTED PATENTS IN CHINA AND THE US



BASIC PATENT STATISTICS

We study an administrative database covering all invention patents applied for at the China National Intellectual Property Administration (CNIPA) from 1985 to 2019. This includes Patent Cooperation Treaty (PCT) invention patents, which are filed almost exclusively by overseas applicants. In our analysis, we utilize similar data on patents applied for at the United States Patent and Trademark Office (USPTO) over the same period.

CLASSIFYING PATENTEE TYPES

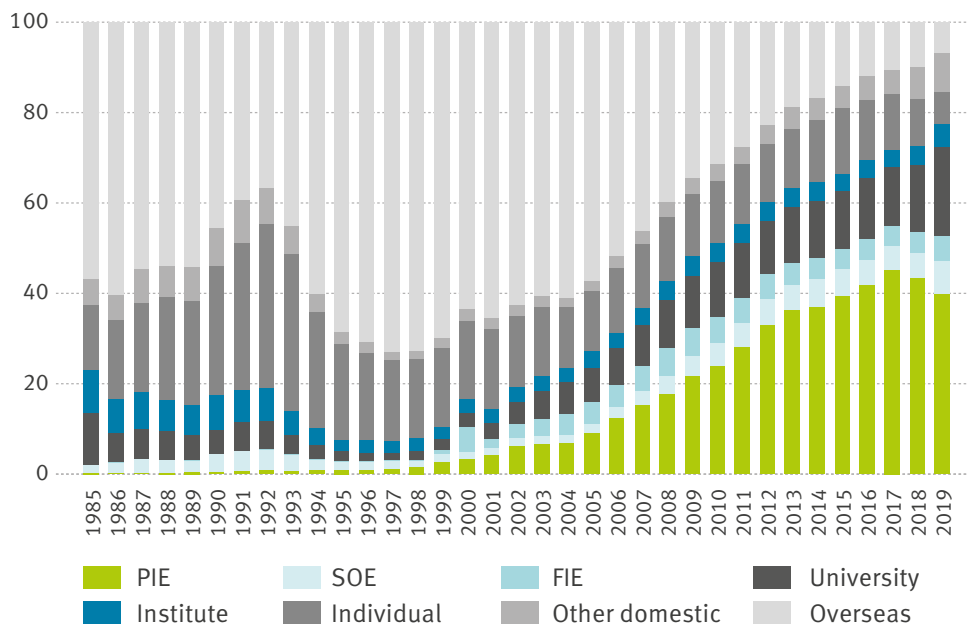
Patent application documents provide information about the name of the patent applicant(s) and the address of the main applicant. Little is known, however, about who is responsible for patenting activity. For example, is the applicant of a patent an enterprise or a research institute? If it is an enterprise, who owns the firm? Drawing on the intersection of three sources of information relating to patentee location, entity type, and ownership type, we define eight mutually exclusive

patentee types: (i) privately invested enterprises (PIEs); (ii) state-owned enterprises (SOEs); (iii) foreign-invested enterprises (FIEs); (iv) universities; (v) research institutes; (vi) individuals; (vii) other domestic patentees; and (viii) overseas patentees.

Figure 2 shows the share of patents applied for in each year by patentee type. Before the mid-2000s, overseas patentees account for more than half of all patent applications in almost every year. However, from the early 2000s onward, there is rapid growth in the patenting activity by PIEs, which quickly become the dominant patentee type in terms of patent shares. By the late 2010s, PIEs account for around 40% of patent applications, with the share of overseas patentees falling to only 10%. In contrast, the share of overseas patentees among granted USPTO utility patents increases from 45% to 55% during the same period. We also observe that the role of Chinese universities and SOE patentees increases slightly over time.

Private firms lead in patent applications

FIGURE 2: PATENT SHARES BY PATENTEE TYPE



PATENT TEXT EMBEDDINGS

The most basic source of information about the knowledge that a patent represents is provided by a set of technology class codes reported by each patent, referred to as the International Patent Classification (IPC) codes.¹ A patent can and often does report more than one IPC code, but the main IPC code of each patent is also observed. A patent's IPC codes are a coarse indicator of the knowledge that the patent embodies. Even at the finest level of disaggregation, it is difficult if not impossible to discern from the IPC codes alone what the patent is about. This information is contained in the textual descriptions of the patent – specifically, in the patent's abstract, which is a summary of the key elements of the patent, and claims, which are a detailed description of how the patent is claiming to be innovative. Hence, the text of a patent is a key source of codified knowledge, but this has traditionally been difficult to incorporate into quantitative analyses of patenting activity.

¹ For example, IPC code "A01B 1/00" indicates that a patent is related to human necessities ("A"), specifically agriculture and forestry ("01"), more specifically soil working ("B"), and even more specifically hand tools ("1/00").

We utilize an industry leading Large Language Model (LLM) to generate text embeddings for the abstract and claims of every patent in the CNIPA database.² Each embedding is a vector that represents the meaning of the text, where the mapping from text to vector is determined by the LLM based on vast quantities of training data, e.g., all text that is publicly available on the internet. The model represents each text with a 768-element vector, thereby reducing the extremely high-dimensional text data to a smaller but still high-dimensional object. A widely used measure of “distance” is cosine similarity, which lies in the interval $[-1, 1]$. When the content of two patents is more similar, their vectors will be pointing in the same direction and their cosine similarity closer to 1.

Leveraging patent text with a Large Language Model

MEASURING PATENT IMPORTANCE

How can patent text embeddings be used to determine which patents are important for innovation? A common conception in the literature is that a patent that is important for innovation should have two features: it should have an influence on future patents (“impact”) and differ from past patents (“novelty”). For example, forward citations are by far the most widely used measure of impact, based on the assumption that if one patent cites another, then the cited patent must have had some influence on the citing patent. Similarly, others have proposed that a lack of backward citations indicates that a patent is dissimilar to the past and hence is indicative of novelty, e.g., the so-called “radicalness” measure used by Ahuja and Lampert (2001) and Banerjee and Cole (2011). More recently, Kelly et al. (2021) have also proposed that the differential similarity of a patent’s keywords to the future relative to the past is indicative of the patent’s importance for innovation. Measures of forward and backward cosine similarity based on patent embedding can be used in an analogous way.

There are two major limitations to measuring patent importance in this way, however. First, there is no theoretical underpinning for why any one particular transformation of the information contained in the embeddings is the best measure of importance. For example, why should forward and backward similarity be equally weighted? Second, this approach evaluates the importance of a patent without regard for the existence of other contemporaneous patents that might also be important for innovation. Arguably, the set of competing ideas needs to be considered jointly to determine which ideas win out in shaping the future. Yet, without additional structure, it is not exactly clear how one should do this.

Limitations of existing measures

To address these issues, we model innovation within a given technological field as a dynamic process with two key features. First, patent embeddings are random vectors drawn from a time-varying mean, which we refer to as the “state of knowledge”. This measures the location in the embedding space where innovation in China is currently centered around. Second, changes in the state of knowledge – which we refer to as the “direction of innovation” – depend on knowledge embodied in current and past patents. By examining the empirical relationship between the direction of innovation and existing patent embeddings, we can estimate the importance of each patent for shifting the knowledge frontier, while effectively controlling for the existence of other sources of knowledge that may also matter for the frontier. Intuitively, patents with embeddings that are aligned with the direction of innovation receive positive importance scores, while those with embeddings that diverge from it receive negative scores.³

Introducing a theoretically grounded measure

² The model we use is the embed-multilingual-v2.0 model from Cohere, a Canadian technology company that specializes in natural language processing (NLP) and LLMs.

³ In our paper, we provide validation that text embeddings can be used to capture meaningful variation across patents and to measure patent importance. First, we show that patents in the same IPC section have embeddings that are more like each other. Second, our measure of patent importance is positively correlated with other traditional measures of patent quality, including the number of forward citations and grant status. And third, our measure of patent importance is a better predictor of firm TFP and output than traditional measures of patent quality.

Finally, innovation does not occur in isolation. Patents often cite patents from other technology fields. Similarly, Chinese patents frequently draw upon foreign knowledge, most notably, patents in the US. To capture these interactions, our model also incorporates cross-field and international influences, recognizing that technological progress is shaped by a complex web of interdependent ideas.⁴ By integrating these factors into the model, our approach provides a more holistic and theoretically grounded measure of patent importance in driving innovation.

ANALYZING PATENT IMPORTANCE

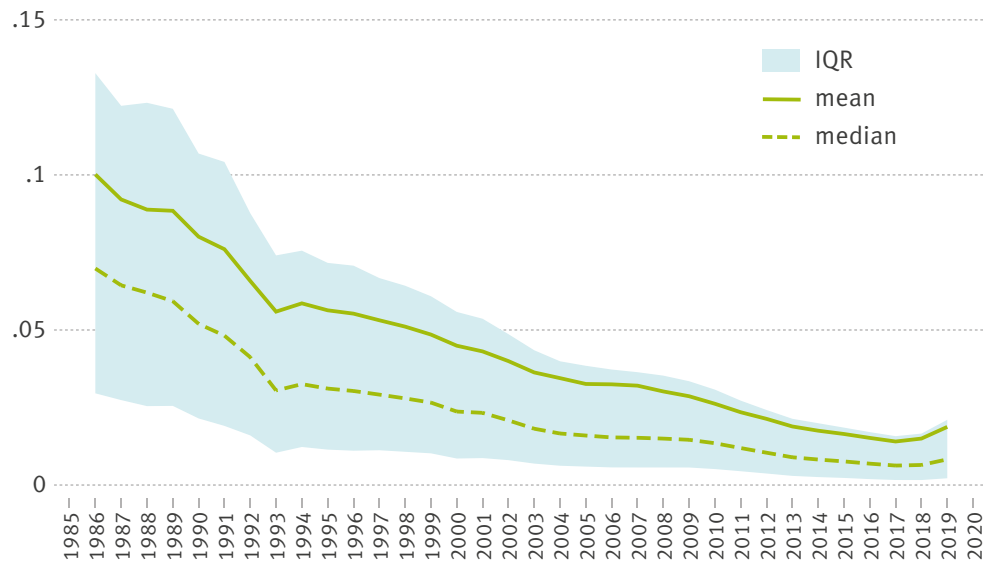
Figure 3 shows how the distribution of our estimated patent importance measure changes over time for the set of important patents (defined as those with a positive importance score).⁵ We see that the average importance of important patents is steadily declining over time. In other words, the influence of important patents on future innovation is falling. We propose two explanations for this trend. First, the innovativeness of patenting – defined as the rate of change in the state of knowledge – has fallen over time. Second, the crowdedness of patenting – defined as the number of patent applications in a given IPC-year – has increased over time.

**Diminishing influence
of important patents
on future innovation**

⁴ To draw comparisons with patenting activity outside of China, we also generate embeddings for all granted utility patents at the USPTO.

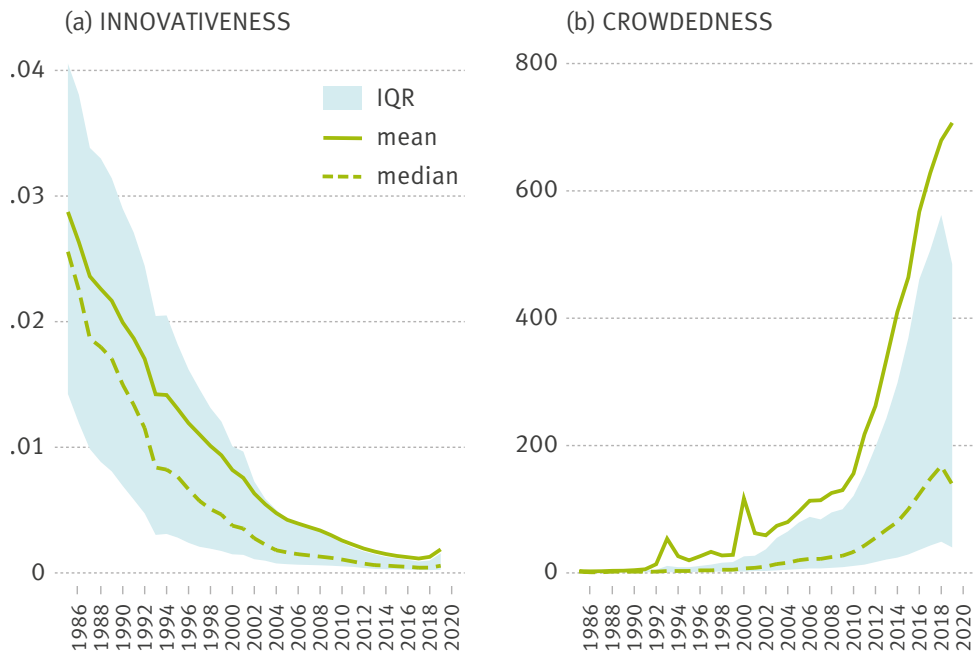
⁵ As we explain in detail in the paper, patent importance is best interpreted as a relative measure within an IPC-year and hence we focus on the importance of important patents instead of average patent importance.

FIGURE 3: THE IMPORTANCE OF IMPORTANT PATENTS



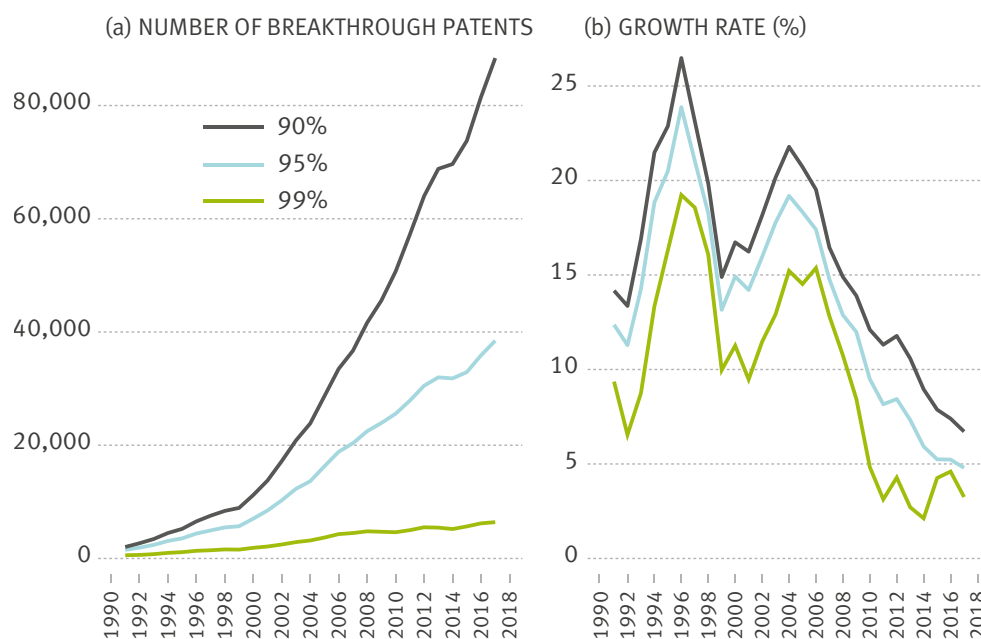
Panel (a) of Figure 4 shows how the distribution of innovativeness changes over time. Evidently, the innovativeness of Chinese patenting has declined steadily, indicating that the state of knowledge has been changing at a slower pace. Patent importance scores are essentially weights assigned to the patent embeddings to “explain” the direction of innovation. Hence, when innovativeness is low, there is less that needs to be explained in the first place and the magnitudes of patent importance scores tend to be small. In panel (b) of Figure 4, we show the distribution for the number of patents applied for in each IPC-year. Clearly, the crowdedness of Chinese patenting has increased substantially. For example, the average patent in 2005 has around 100 other patents in the same IPC-year, whereas this increases by about seven-fold by 2019. The crowdedness of innovation matters for the distribution of patent importance because when there are many candidate patents that can explain the direction of innovation, each individual patent is less likely to matter a lot.

**Chinese patenting:
less innovative,
more crowded**

FIGURE 4: PATENT SHARES BY PATENTEE TYPE

Note that the decline in the average importance of important patents occurs in parallel with rapid growth in the number of Chinese patents overall (Figure 1). Hence, the stock of highly important patents may well be increasing. To investigate, we define a breakthrough patent as one that has an importance score above the k th percentile of the importance distribution over the entire sample of patents (so that the threshold does not vary across years). Figure 5 shows the number of breakthrough patents in each year for the 99th, 95th and 90th percentile and the associated annual growth rates. Indeed, we find that the number of breakthrough patents is increasing over time. After growing in the late 1990s and the early 2000s at 15–20% per year, however, growth rates fall off significantly beginning in the mid-2000s. By the end of the period, growth rates are a third to a quarter of their earlier levels.

Breakthrough patents rising, but growth is slowing

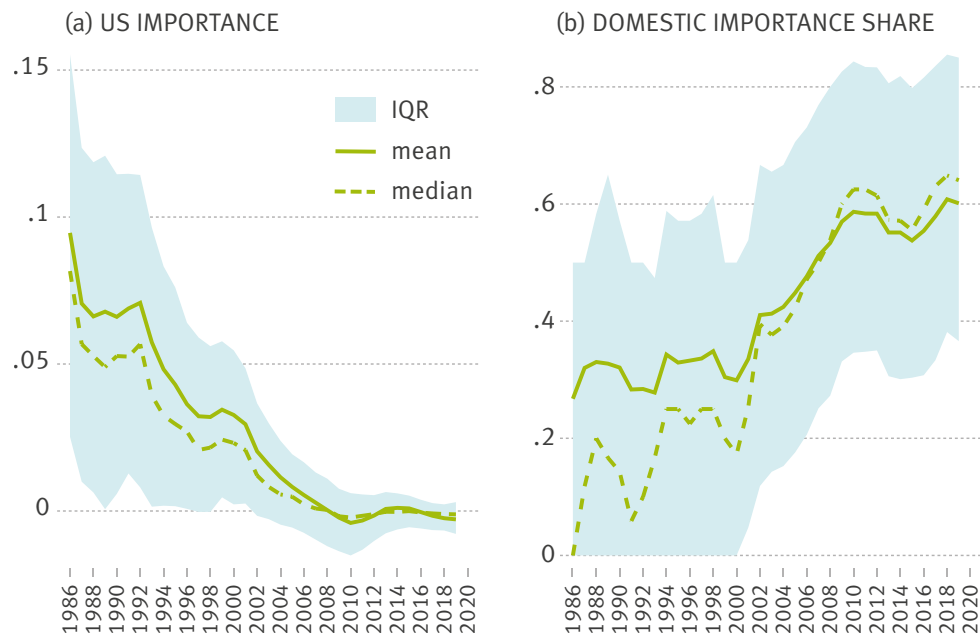
FIGURE 5: NUMBER AND GROWTH OF BREAKTHROUGH PATENTS

SOURCES OF KNOWLEDGE – CHINESE VERSUS US KNOWLEDGE

Which sources of knowledge are the most important for shaping the direction of innovation in China? We consider two margins: the relative importance of knowledge embodied in patents filed in China versus outside of China; and sources of knowledge within China. Panel (a) of Figure 6 shows how the distribution of US importance varies over time. Much like the distribution of importance for important Chinese patents (see Figure 3), we observe that the importance of US knowledge is steadily declining. This decline may be explained in part by the same forces as discussed above – i.e., less innovative and more crowded patenting in China – but some of the decline may also reflect changes in the relative importance of knowledge embedded in Chinese versus US patents.

**Declining importance
of US patents**

FIGURE 6: US IMPORTANCE AND DOMESTIC IMPORTANCE SHARE



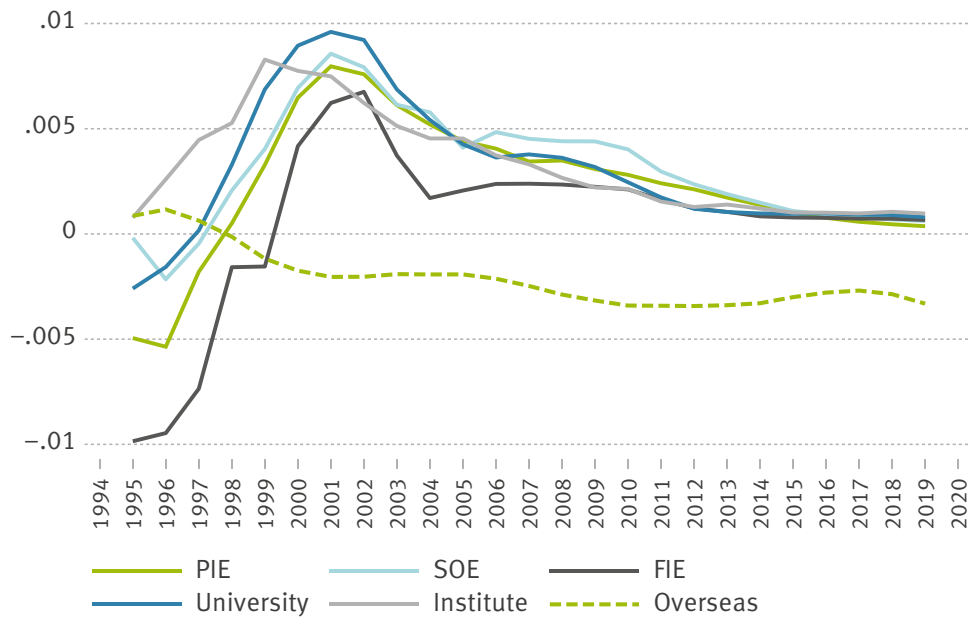
To investigate the latter, we measure in each IPC-year the share of Chinese patents with an importance score greater than the importance of US patents. Henceforth, we refer to this measure as the domestic importance share. Panel (b) of Figure 6 plots the distribution of the domestic importance share over time. We observe that the average domestic importance share is roughly constant from 1985 to 2000 at 30%. Throughout the 2000s, however, there is a sustained increase in the measure, so that by 2010, the average share is just under 60%. From 2010 onward, the domestic importance share levels off and again remains roughly constant at around 60%. Thus, the 2000s were a key period during which innovation in China became increasingly directed by knowledge within China as opposed to knowledge outside.

**60% of Chinese
patents exceed US
patents in importance**

SOURCES OF KNOWLEDGE WITHIN CHINA

To compare the importance of patents across patentee types, we first compute the importance of each patent parsed of IPC-year effects. We then average this measure over all patents applied for by a given patentee type each year. This measures the relative importance of patents by each patentee type within an IPC-year. Figure 7 shows how this relative measure changes over time.⁶

⁶ For visual clarity, we omit the "Individual" and "Other Domestic" patentee types and plot five-year rolling averages of the relative importance scores.

FIGURE 7: RELATIVE PATENT IMPORTANCE BY PATENTEE TYPE

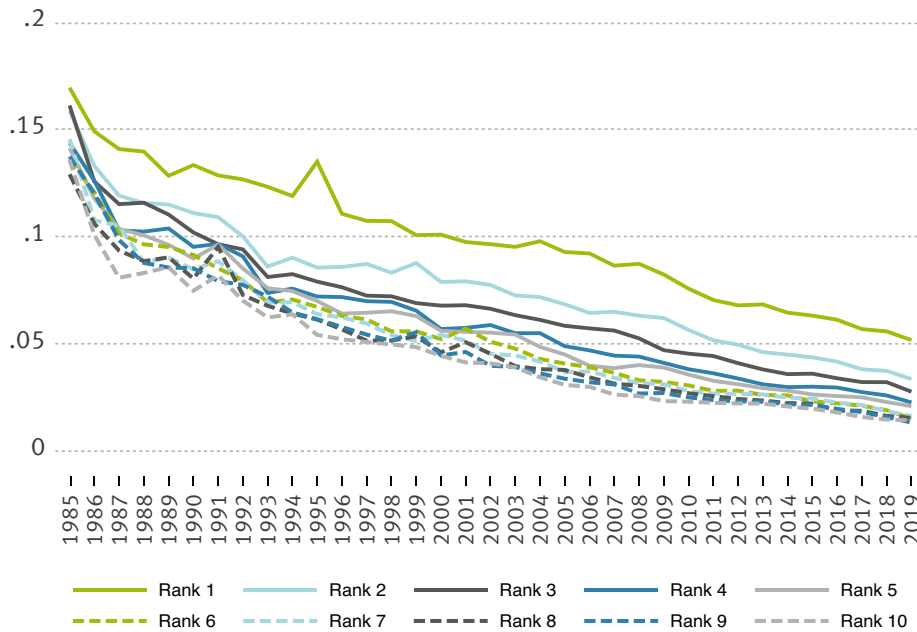
We highlight two key observations. First, before 2000, patents filed by Chinese universities and research institutes tended to be the most important. By 2005, however, SOE patents assume the top position in terms of relative importance. In almost all years, SOE patents are relatively more important than private sector patents, while PIE patents are also typically more important than FIE patents. Second, in the mid-1990s, overseas patents had slightly higher than average importance. However, the relative importance of these patents declines steadily, so that from 2000 onward, overseas patents are less important than patents applied for by enterprises, universities, and research institutes in China.

**Overseas patents
less important than
Chinese patents**

THE IMPORTANCE OF SECONDARY IPCS

Finally, in Figure 8 we examine how the importance of interactions across IPCs has evolved over time. Secondary IPCs are ordered based on their rank in terms of how frequently they are reported by patents in a given technology class. The average importance of secondary IPCs trends steadily downward throughout our sample period. This is suggestive that patenting in China is becoming “narrower”, where knowledge generated within an IPC is more important for innovation in the IPC as opposed to knowledge generated in other related IPCs.

**Patenting in China is
becoming narrower**

FIGURE 8: SECONDARY IPC IMPORTANCE

POLICY IMPLICATIONS AND FUTURE RESEARCH

Our analysis reveals that Chinese patenting has become narrower and less innovative over time. The role of overseas knowledge has also declined sharply. The reasons for this shift are less clear. Although some of the decline in innovativeness broadly reflects global trends (Kelly et al., 2021; Kalyani 2024), domestic policy may also figure importantly. Sorting out these influences, as well as their effect on China's growth trajectory are important topics for future research.

REFERENCES

- Ahuja, G., and C. Lampert (2001), Entrepreneurship in the Large Corporation: A Longitudinal Study of How Established Firms Create Breakthrough Inventions, *Strategic Management Journal* 22, 521–543.
- Banerjee, P. M., and B. M. Cole (2011), Globally Radical Technologies and Locally Radical Technologies: The Role of Audiences in the Construction of Innovative Impact in Biotechnology, *IEEE Transactions on Engineering Management* 58(2), 262–274.
- Boeing, P., L. Brandt, R. Dai, K. Lim, and B. Peters (2025), The Anatomy of Chinese Innovation: Insights on Patent Quality and Ownership, ZEW Discussion Paper Nr. 24-016 | 06/2025, Mannheim.
- Griliches, Z. (1990), Patent Statistics as Economic Indicators: A Survey, *Journal of Economic Literature* 28(4), 1661–1707.
- Han, P., W. Jiang, and D. Mei (2024), Mapping US–China Technology Decoupling: Policies, Innovation, and Firm Performance, *Management Science* 70(12), 8386–8413.
- Kalyani, A. (2024), The Creativity Decline: Evidence from US Patents. Working Paper.
- Kelly, B., D. Papanikolaou, A. Seru, and M. Taddy (2021). Measuring Technological Innovation Over the Long Run, *American Economic Review: Insights* 3(3), 303–320.



Imprint

Authors: Philipp Boeing (Goethe University Frankfurt and ZEW Mannheim), Loren Brandt (University of Toronto), Ruochen Dai (Central University of Finance and Economics), Kevin Lim (University of Toronto), Bettina Peters (ZEW Mannheim and University of Luxembourg)

Acknowledgments: The project on which this study is based was funded by the German Federal Ministry of Education and Research under grant number 01DO21006A. The responsibility for the content of this study lies with the authors.

This policy brief is also available on [VoxChina](#).

Publisher: ZEW – Leibniz Centre for European Economic Research
L 7, 1 · 68161 Mannheim · Germany · info@zew.de · www.zew.de/en · x.com/zew_en

President: Prof. Achim Wambach, PhD · Managing Director: Claudia von Schuttenbach

Editorial responsibility: Pascal Ausäderer · kommunikation@zew.de

Quotes from the text: Sections of the text may be quoted in the original language without explicit permission provided that the source is acknowledged.

© ZEW – Leibniz-Zentrum für Europäische Wirtschaftsforschung GmbH Mannheim

ZEW

Leibniz
Leibniz
Gemeinschaft