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

Philipp Boeing
Goethe University, ZEW

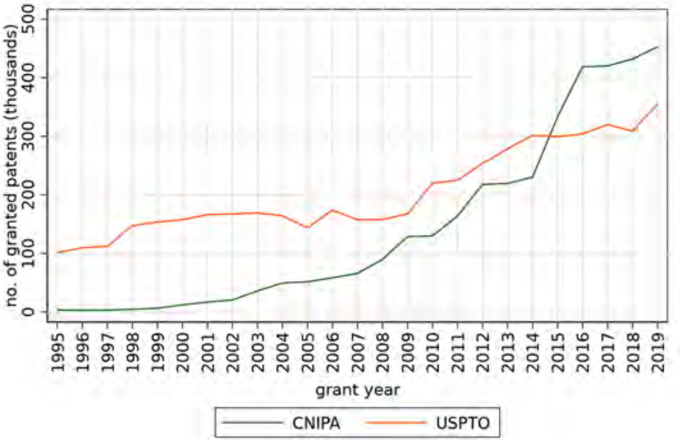
Loren Brandt
University of Toronto

Ruochen Dai
CUFE

Kevin Lim
University of Toronto

Bettina Peters
ZEW

August 2025



- There has been rapid growth in patenting activity in China over the last 20 years
  - faster growth than at the USPTO
  - despite slowdowns in Chinese GDP, productivity and export growth

- How has the quality of patenting in China changed over time and what sources of knowledge have been important for driving innovation?
- Three major challenges in answering these questions:
  - knowledge embodied in patents is codified almost entirely through text
  - there is little existing theory to guide measurement of patent quality
  - standard patent data provide little information about patentees
- We make progress by:
  - using a Large Language Model (LLM) to incorporate patent text data into a quantitative analysis of patenting activity
  - developing a new but simple model of innovation to motivate how patent quality should be measured
  - leveraging information from a comprehensive business registry in China to differentiate between patentee types

- We study patent data from CNIPA, 1985-2020, 11+ million invention patents

(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 103944225 B
(45)授权公告日 2017.04.26

(21)申请号 201410153009.6

(22)申请日 2014.04.16

(65)同一申请的已公布的文献号
申请公布号 CN 103944225 A

(43)申请公布日 2014.07.23

(73)专利权人 华为技术有限公司
地址 518129 广东省深圳市龙岗区坂田华为总部办公楼

(72)发明人 马向民 朱丁旺

(74)专利代理机构 广州三环专利代理有限公司 44202
代理人 郝传鑫 熊永强

(51)Int.Cl.
H02J 7/00(2006.01)
H01M 10/42(2006.01)
H01M 10/44(2006.01)
H01M 10/65(2014.01)
G01R 31/36(2006.01)

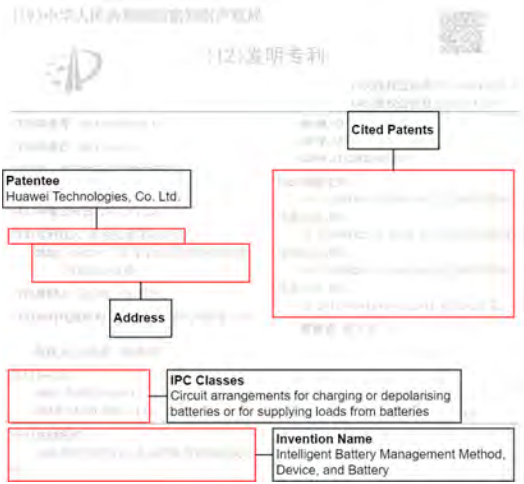
(56)对比文件
CN 103367823 A,2013.10.23,说明书具体实施方式,图1.
CN 103367823 A,2013.10.23,说明书具体实施方式,图1.
CN 103336245 A,2013.10.02,说明书具体实施方式,图1.
US 2011/0181245 A1,2011.07.28,全文.

审查员 曹卫琴

权利要求书3页 说明书8页 附图4页

(54)发明名称
电池智能管理方法、电池智能管理装置及电池

- We study patent data from CNIPA, 1985-2020, 11+ million invention patents



Patent text data

(57)摘要

本发明公开一种电池智能管理装置，包括识别单元、存储单元和控制单元；所述识别单元用于识别供电组件的规格；所述存储单元用于存储各种规格供电组件的管理模式；所述供电组件的管理模式与所述供电组件的规格一一对应，所述控制单元用于根据所述识别单元识别的所述供电组件的规格，从所述存储单元内提取与所述供电组件的规格相对应的管理模式，并采用所述管理模式对所述供电组件进行充电管理。本发明还提供一种电池及一种电池智能管理方法。本发明针对不同规格的供电组件，区别的管理供电组件的充放电，保持电池的最佳状态，延长电池的使用寿命，实现了电池智能化管理。

Summary

The invention discloses a battery intelligent management device. The battery intelligent management device comprises a recognition unit, a storage unit and a control unit, wherein the recognition unit is used for recognizing the specifications of power supply assemblies; the storage unit is used for storing the management modes of the various specifications of power supply assemblies; the management modes of the power supply assemblies correspond to the specifications of the power supply assemblies one to one; the control unit is used for extracting the management modes corresponding to the specifications of the power supply assemblies from the storage unit according to the specifications, recognized by the recognition unit, of the power supply assemblies, and the management modes are adopted for performing charging management on the power supply assemblies. The invention further provides a battery and a battery intelligent management method. According to the battery intelligent management method, the battery intelligent management device and the battery, charging and discharging of the power supply assemblies are managed differently, the best state of the battery is kept, the service life of the battery is prolonged, and the battery can be intelligently managed.

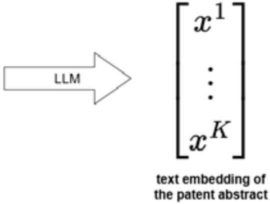
claims text

Measuring patent quality

- Key question (1): what is patent "quality" and how has this changed in China?
- A basic premise: a high quality patent is one that is important for innovation
- Some measures that have been explored in the literature:
  - number of forward citations received details
  - citation network centrality details
  - legal status changes (e.g., grants, unpaid renewal fees) details
  - legal status timing (e.g., time to grant) details
  - number and length of claims details
  - existence of overseas filings details
- None of these make direct use of the information content of patent text
  - this is a key source of codified knowledge
  - but has been traditionally difficult to incorporate into quantitative analysis
- We use **text embeddings** from large language models (LLMs) to make progress

(57)摘要

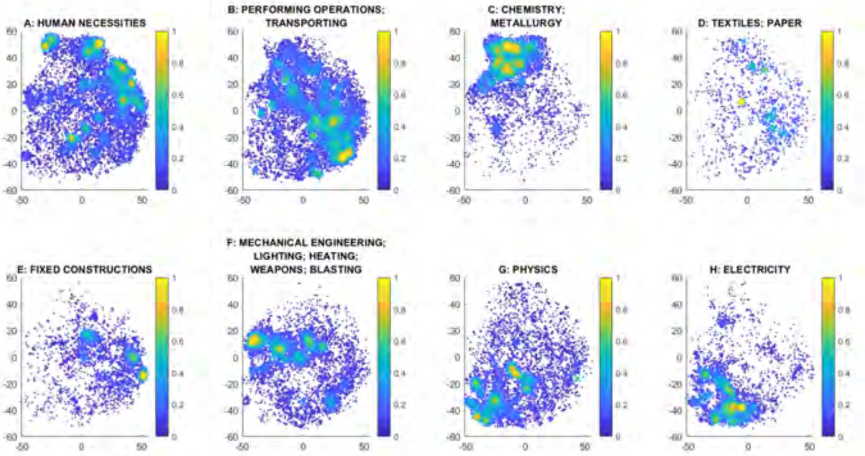
本发明公开一种电池智能管理装置,包括识别单元、存储单元和控制单元;所述识别单元用于识别供电组件的规格;所述存储单元用于存储各种规格供电组件的管理模式;所述供电组件的管理模式与所述供电组件的规格一一对应,所述控制单元用于根据所述识别单元识别的所述供电组件的规格,从所述存储单元内提取与所述供电组件的规格相对应的管理模式,并采用所述管理模式对所述供电组件进行充电管理。本发明还提供一种电池及一种电池智能管理方法。本发明针对不同规格的供电组件,区别的管理供电组件的充放电,保持电池的最佳状态,延长电池的使用寿命,实现了电池智能化管理。



$$\begin{bmatrix} x^1 \\ \vdots \\ x^K \end{bmatrix}$$

- A text embedding = representation of text in the form of a vector
- We generate text embeddings using the **Cohere multilingual model**:
  - Canadian startup based in Toronto
  - has quickly become a front-runner in the LLM space
  - model generates $K = 768$ vector representations of text
- With embeddings, we can compute formal measures of distance between patents
  - standard metric for text data: cosine similarity details

Text embeddings



- 2D projections of patent abstract text embeddings (UMAP)
- Embeddings represent patents in the same IPC section in the same "space"

A naive approach

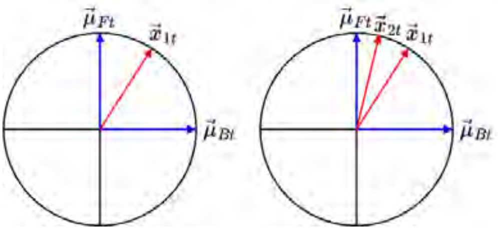
- How can we use embeddings to measure the **importance** of a patent?
- Borrowing from the literature, e.g., Kelly et al (2021), important patents should:
  - look like the future ("impact")
  - not look like the past ("novelty")
- A candidate measure of the importance of patent i would then be:

$$p_{it} = \underbrace{c\left(\vec{x}_{it}, \vec{\mu}_{Ft}\right)}_{F_{it}} - \underbrace{c\left(\vec{x}_{it}, \vec{\mu}_{Bt}\right)}_{B_{it}}$$

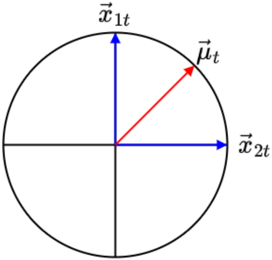
  - $\vec{\mu}_{Ft}$: average (normalized) embedding of future patents (**F**orward)
  - $\vec{\mu}_{Bt}$: average (normalized) embedding of past patents (**B**ackward)
  - F_{it}: cosine similarity of patent i to the average future patent
  - B_{it}: cosine similarity of patent i to the average past patent

Limitations of the naive approach

- There are two major limitations of this definition of importance
- There is no theoretical justification for why this particular transformation of patent embeddings should reflect the importance of a patent
  - for example, why not $p_{it} = \alpha F_{it} - \beta B_{it}$ for some constants $\alpha, \beta \neq 1$
- The measure does not take into account the existence of other patents



- To make progress, we develop a simple theory of how past and present knowledge affects innovation in the future
- Suppose that $\vec{x}_{it}$ is a random vector drawn from some distribution with mean $\vec{\mu}_t$



- $\vec{\mu}_t$ is the **state of knowledge**
  - the goal is to understand how this state changes over time and to quantify the importance of each patent for such changes

- To illustrate the key ideas, consider a simplified version of the model where the innovation process in a given IPC is:

$$\vec{\mu}_{t+1} = \rho_t \vec{\mu}_t + \sum_{i \in \Omega_t} p_{it} \vec{x}_{it} + \vec{\epsilon}_t$$

  - ρ_t: the **memory** of the innovation process
  - p_{it}: the **importance** of patent i for future innovation
  - Ω_t: set of patents applied for at time t
  - $\vec{\epsilon}_t$: random vector orthogonal to $\vec{\mu}_t$,$\{\vec{x}_{it}\}_{i \in \Omega_t}$
- Define $\vec{d}_{t+1} \equiv \vec{\mu}_{t+1} - \rho_t \vec{\mu}_t$ as the **direction of innovation**
- Given orthogonality of $\vec{\epsilon}_t$, patent importance $\{p_{it}\}_{i \in \Omega_t}$ can be estimated as OLS coefficients from a regression of $\vec{d}_{t+1}$ on $\{\vec{x}_{it}\}_{i \in \Omega_t}$
  - "observations" are the K dimensions of the embeddings
- Intuition: importance of i is high if $\vec{d}_{t+1}$ points in a more similar direction to $\vec{x}_{it}$

2D and 3D examples

- OLS estimate of the importance vector can also be written as:

$$\vec{p}_t = C_t^{-1}\left(\vec{F}_t - \rho_t \vec{B}_t\right)$$

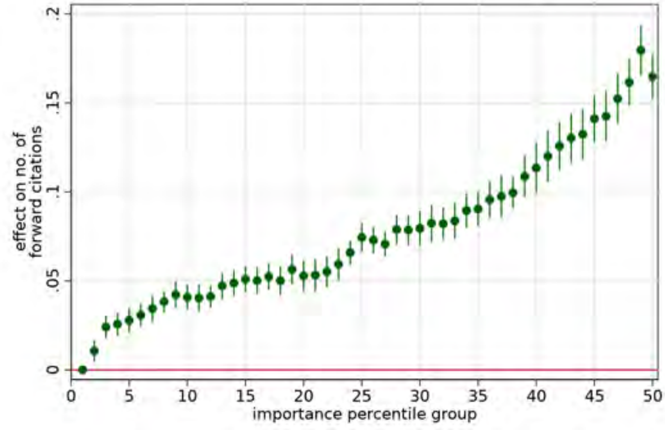
  - C_t: matrix of cosine similarities between $\{\vec{x}_{it}\}_{i\in\Omega_t}$
  - $\vec{F}_t$: vector of forward similarities between $\{\vec{x}_{it}\}_{i\in\Omega_t}$ and $\vec{\mu}_{t+1}$
  - $\vec{B}_t$: vector of backward similarities between $\{\vec{x}_{it}\}_{i\in\Omega_t}$ and $\vec{\mu}_t$
- This looks like the "naive" measure $F_{it} - B_{it}$ but with two key differences:
  - it adjusts for the influence of other patents on innovation (through C_t^{-1})
  - the relative weight on F ("impact") vs. B ("novelty") is pinned down by the memory of the innovation process (ρ_t)

Estimating patent importance

- In practice, we estimate p_{it} using a LASSO estimator instead of OLS
  - promotes sparsity in $\{p_{it}\}_{i=1}^{N_t}$ by adding $L1$ penalty to OLS MSE objective
  - deals with multicollinearity (e.g., when # patents > # embed elements)
  - can be interpreted as an "attention cost" for inventors
- Full empirical model for innovation in IPC g is:

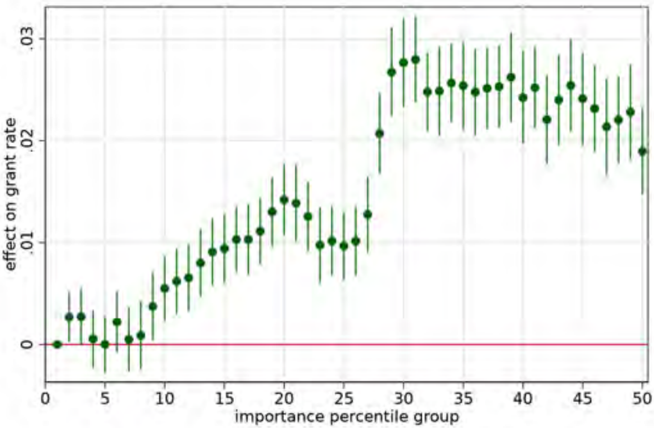
$$\vec{\mu}^g_{Ft} = \rho^g_t \vec{\mu}^g_{Bt} + \sum_{i \in \Omega^g_t} p^g_{it} \vec{x}^g_{it} + \underbrace{p^g_{US,t} \vec{\mu}^g_{US,Bt}}_{\text{foreign influence}} + \underbrace{\sum_{g' \in \Gamma_g} \gamma^{gg'}_t \vec{\mu}^{g'}_{Bt}}_{\text{IPC interactions}} + \vec{\epsilon}^g_t$$

  - $\vec{\mu}^g_{Ft}$: average embedding from $t+1$ to $t+3$
  - $\vec{\mu}^g_{Bt}$: average embedding from $t-2$ to t
  - $\vec{\mu}^g_{US,Bt}$: average embedding of USPTO patents from $t-2$ to t
  - $p^g_{US,t}$: importance of US patents for Chinese innovation
  - $\gamma^{gg'}_t$: importance of patents in IPC g' for innovation in g
  - Γ_g: set of related IPCs (based on secondary IPCs reported by patents in g)



- Plot shows coefficients from a regression of no. of forward citations (excluding self-citations) on percentile group of p_{it}
- Both variables are residualized by IPC-year effects
- Error bars show 95% confidence intervals

Internal validation: correlation with grant status



- Plot shows coefficients from a regression of grant status on the percentile of p_{it}
- Both variables are residualized by IPC-year effects
- Error bars show 95% confidence intervals

External validation: TFP and output regressions

- To assess whether patent importance is a "better" predictor of firm outcomes than other measures of patent quality, we merge patent data with NBS above-scale firm data, 1998-2007
- For each patent, we determine whether it is:
  - in the top p% of the importance distribution within its application year
  - in the top p% of the forward citation distribution within its application year
- For each firm f and year t, we then measure:
  - stock of top-p% important patents, $N_{ft}^{topimp,p}$
  - stock of top-p% cited patents, $N_{ft}^{topcit,p}$
  - stock of active patents, N_{ft}^{active}
- We then estimate the following regression via OLS:

$$y_{ft} = \beta^{topimp,p} N_{ft}^{topimp,p} + \beta^{topcit,p} N_{ft}^{topcit,p} + \beta^{active} N_{ft}^{active} + \gamma \delta_{ft} + \alpha_{h(f)t} + \epsilon_{ft}$$

  - y_{ft}: firm-year outcome of interest (log TFP or output)
  - δ_{ft}: dummy for whether firm has any patents
  - $\alpha_{h(f)t}$: 4-digit industry-year fixed effect

External validation: TFP regressions

outcome: log TFP ($\times$100); fixed effects: 4-digit industry x year					
$p =$	2	5	10	25	50
i. stock of top-p% important patents, $\beta^{topimp,p}$	0.40	0.31	0.17	0.05	0.03
	(3.87)	(5.18)	(5.39)	(4.68)	(4.16)
ii. stock of top-p% cited patents, $\beta^{topcit,p}$	0.14	0.08	0.05	0.02	-0.00
	(2.88)	(3.17)	(3.22)	(2.21)	(-0.24)
iii. stock of active patents, β^{active}	-0.03	-0.07	-0.09	-0.09	-0.09
	(-1.48)	(-2.91)	(-3.55)	(-3.46)	(-3.25)
iv. has patents, γ	3.66	3.63	3.63	3.66	3.68
	(22.33)	(22.14)	(22.15)	(22.35)	(22.50)
observations (m)	1.78	1.78	1.78	1.78	1.78
R^2	0.73	0.73	0.73	0.73	0.73
adjusted R^2	0.73	0.73	0.73	0.73	0.73

- Coefficients on top-important patents are always positive and significant, declining in p, larger than coefficients on top-cited and active patents, and decaying more slowly with p

with naive importance

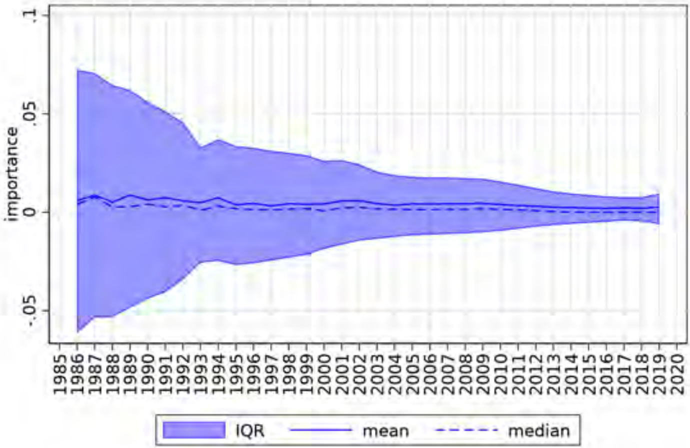
External validation: output regressions

outcome: log output ($\times 100$); fixed effects: industry-year					
$p =$	2	5	10	25	50
i. stock of top-p% important patents, $\beta^{topimp,p}$	5.42	5.11	2.69	0.76	0.29
	(15.02)	(23.95)	(24.27)	(18.72)	(11.18)
ii. stock of top-p% cited patents, $\beta^{topcit,p}$	0.59	0.47	0.34	-0.01	-0.13
	(3.43)	(5.15)	(5.79)	(-0.19)	(-4.14)
iii. stock of active patents, β^{active}	0.35	-0.28	-0.62	-0.32	-0.01
	(4.90)	(-3.36)	(-6.91)	(-3.32)	(-0.10)
iv. has patents, γ	110.64	110.09	110.17	110.64	111.01
	(191.48)	(190.30)	(190.53)	(191.57)	(192.37)
observations (m)	1.78	1.78	1.78	1.78	1.78
R^2	0.15	0.15	0.15	0.15	0.15
adjusted R^2	0.15	0.15	0.15	0.15	0.15

- Coefficients on top-important patents are always positive and significant, declining in p, larger than coefficients on top-cited and active patents, and decaying more slowly with p

with naive importance

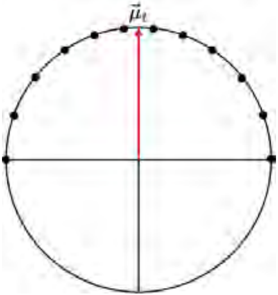
The distribution of patent importance



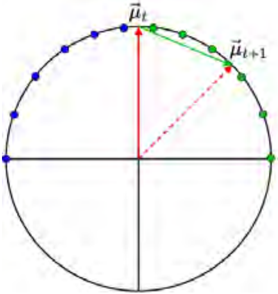
- Avg. importance is consistently around zero: it is a **relative** measure within year
- Dispersion of patent importance is falling over time

USPTO results

- Why is the importance distribution consistently centered around zero?
- p_{it} is essentially measuring the extent to which patent i is "pointing" in the same direction as the direction of innovation
- Suppose that the distribution of $\vec{x}_{it}$ is symmetric around $\vec{\mu}_t$

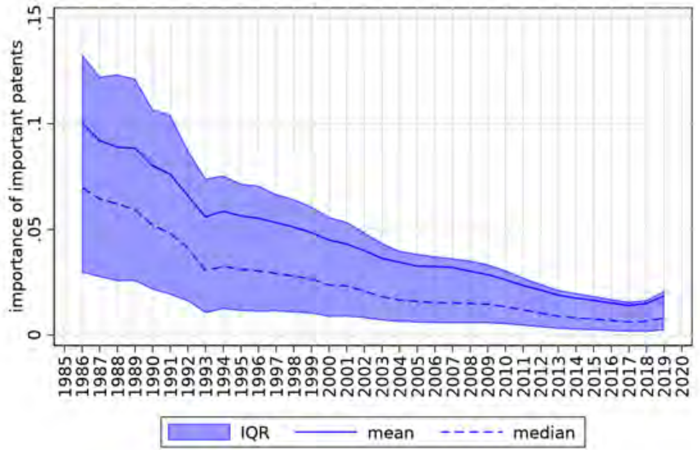


- When $\vec{\mu}_t$ moves in any direction, it must move:
  - in the same direction as half of the patents
  - in the opposite direction as half of the patents



- Hence, importance tends to be positive/negative for half of all patents
  - average patent importance is not a meaningful statistic
- Instead, we should be looking at:
  - the average importance of important patents ($p_{it} > 0$)
  - the relative importance of different types of patents

Importance of important patents



- Define an important patent to be one with $p_{it} > 0$
- The average importance of important patents has declined over time - why?

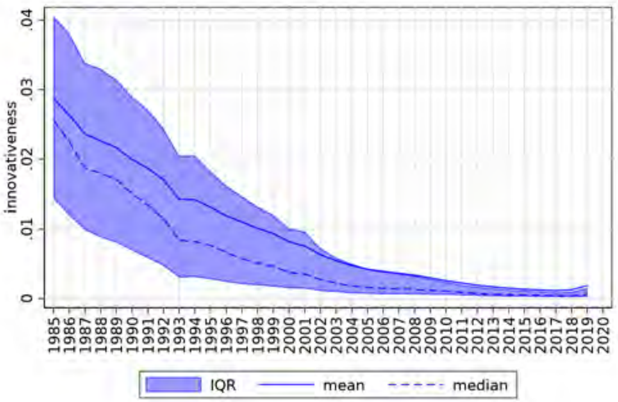
USPTO results

Explanation 1: patenting in China has become less innovative

- Define **innovativeness** of an IPC at time t as:

$$I_t = \frac{1}{2}\left[1 - c\left(\vec{\mu}_{Ft}, \vec{\mu}_{Bt}\right)\right] \in [0, 1]$$

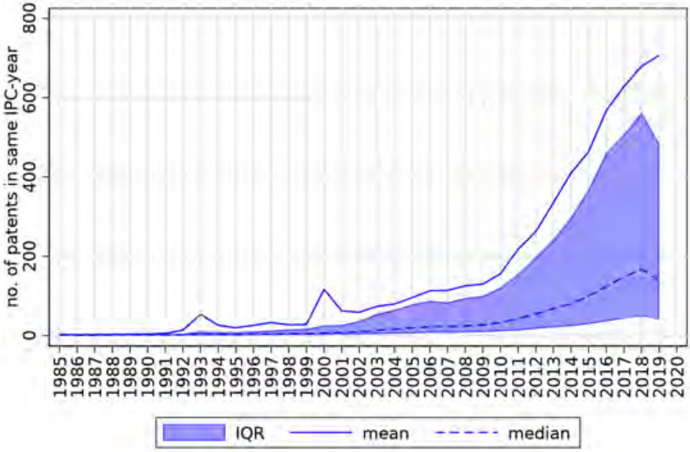
- When I_t is large, the future is less similar to the past



intuition USPTO results

Explanation 2: patenting in China has become more crowded

- Define **crowdedness** as no. of patents N_t in an IPC at time t



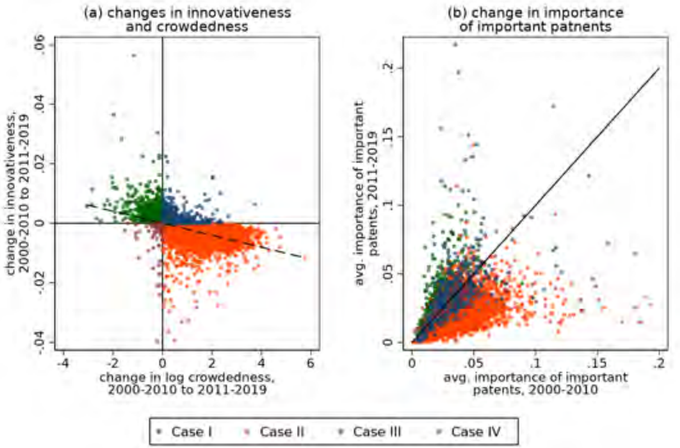
intuition USPTO results

Regressions of patent importance on innovativeness and crowdedness

	(1)	(2)	(3)	(4)
observation =	patent-year	patent-year	IPC-year	IPC-year
i. innovativeness	0.40	0.51	0.45	0.52
	(448.79)	(847.01)	(130.30)	(182.98)
ii. crowdedness	-0.13	-0.25	-0.10	-0.22
	(-91.45)	(-568.15)	(-40.85)	(-125.19)
IPC fixed effects	yes	no	yes	no
year fixed effects	yes	yes	yes	yes
observations (m)	6.01	6.01	0.42	0.42
R^2	0.36	0.28	0.39	0.24
adjusted R^2	0.36	0.28	0.33	0.24
within R^2	0.05	0.25	0.06	0.22

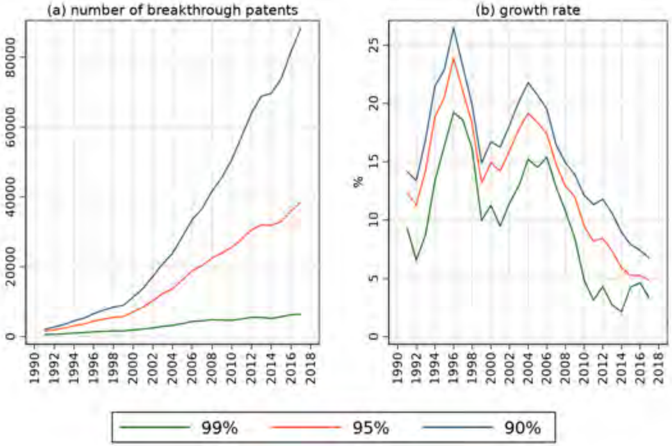
- (1), (2): regress p_{it} on I_t and $\log N_t$ for the sample of important patents
- (3), (4): regress average p_{it} of important patents on I_t and $\log N_t$ by IPC-year
- Table reports standardized beta coefficients and t-statistics in parentheses

Changes in innovativeness, crowdedness, and importance



- Each point in the scatter plot is an IPC, falling into four cases
- Most IPCs are in Case II or III: negative correlation between changes over time in innovativeness and crowdedness
- Some IPCs are in Case I: innovativeness is increasing despite an increase in crowdedness

Breakthrough patents

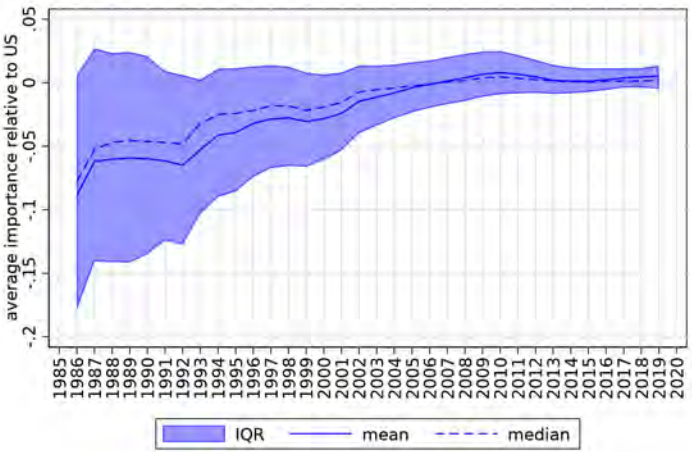


- Avg. importance of important patents declines but no. of patents grows rapidly
- Define a breakthrough patent as one with p_{it} above x-percentile of distribution
- No. of breakthrough patents still grows although growth rate is declining

- Key question (2): what sources of knowledge have been important for innovation?
  - knowledge inside China (CNIPA) vs. outside China (USPTO)
  - knowledge produced by different patentee types within China
- We use three sources of information to differentiate between patentees
  - location based on the address of the main patent applicant
  - entity type based on a keyword search of applicant names
  - ownership based on registered capital in business registry
- Using this, we define 8 **patentee types**:

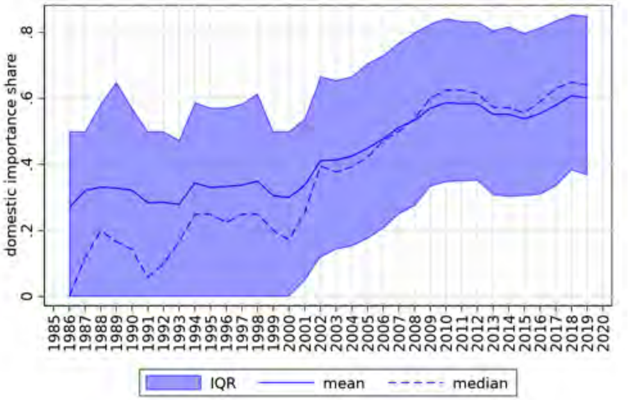
	location	entity type	ownership
private-invested enterprise (PIE)	domestic	enterprise	private
state-owned enterprise (SOE)	domestic	enterprise	state
foreign-invested enterprise (FIE)	domestic	enterprise	foreign
university	domestic	university	any
institute	domestic	institute	any
individual	domestic	individual	any
other domestic	domestic	other	any
overseas	overseas	any	any

Importance of Chinese vs. US patents



- Gap between Chinese and US importance has been steadily narrowing

Domestic importance share

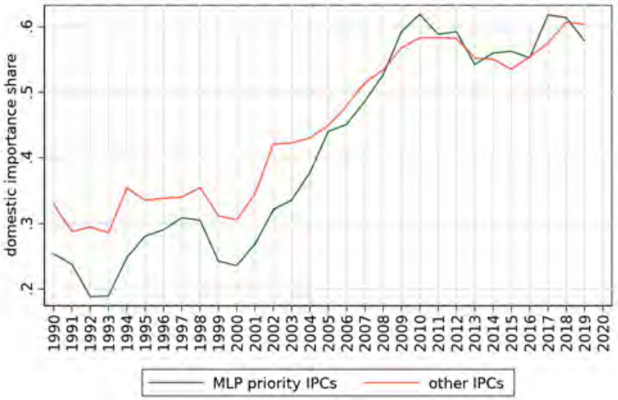


- Domestic importance share: share of Chinese patents in an IPC with $p^g_{it} > p^g_{US,t}$
- Constant from 1985-2000, increasing from 2000-2010, constant after 2010

Domestic importance share

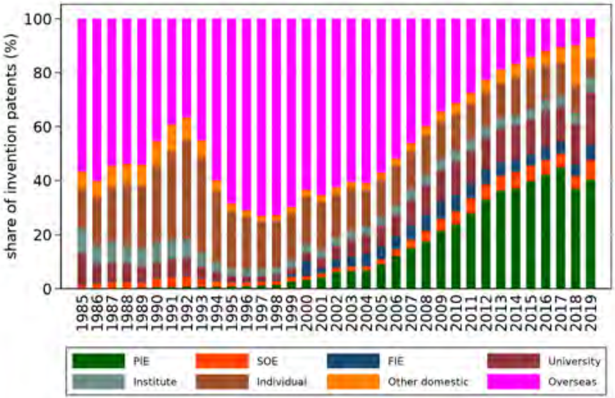
- As a preliminary examination of the role of policy, we consider the **Medium- to Long-Term Plan for S&T Development (2006-2020) (MLP)**
  - lists 62 "priority technologies" for economic development, national security
- To link MLP priority technologies to IPCs, we do the following:
  - generate embeddings of MLP priority text descriptions
  - compute average patent embedding in each IPC from 2000-2010
  - compute cosine similarity for each MLP priority x IPC pair
  - define an IPC to be an MLP priority if average cosine similarity of top three scores is above the 90^{th} percentile of the distribution across all patents
- Findings are robust to:
  - using top score or average of top 5 cosine similarity scores
  - defining threshold at the 75^{th} or 95^{th} percentiles

Domestic importance shares for MLP priority and non-priority technologies



- MLP priorities initially have lower domestic importance share than non-priorities
- Gap begins to close around 2004, catch-up complete around 2010

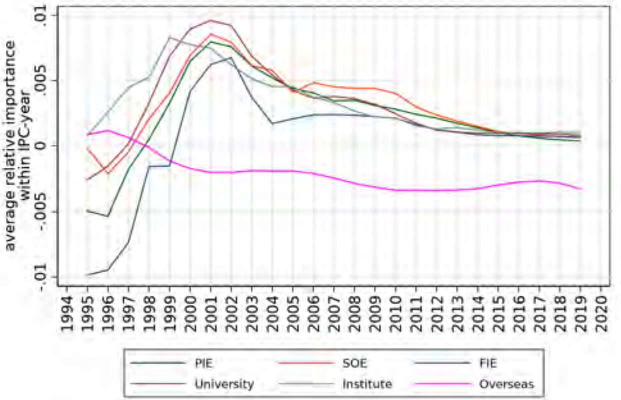
Share of patenting activity by patentee type



- Rapid patent growth largely accounted for by PIEs
- Role of Chinese universities also increases over time
- Role of overseas patentees declines substantially
- In contrast, share of overseas patents at USPTO increases from 45% to 55%

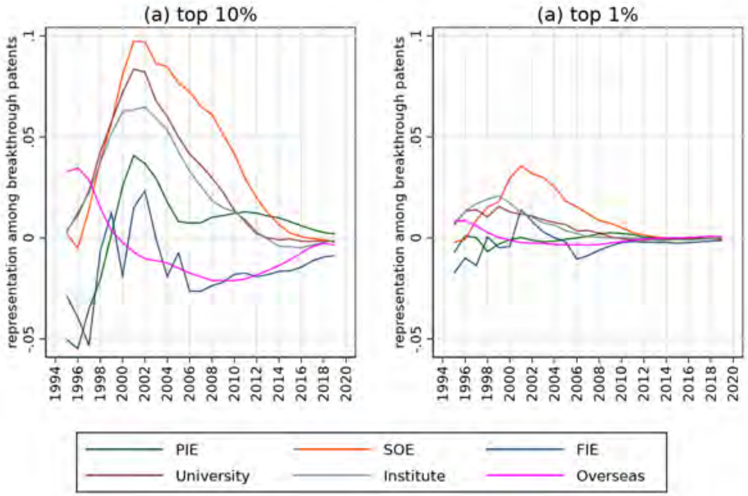
by IPC section USPTO results

Relative patent importance across patentee types

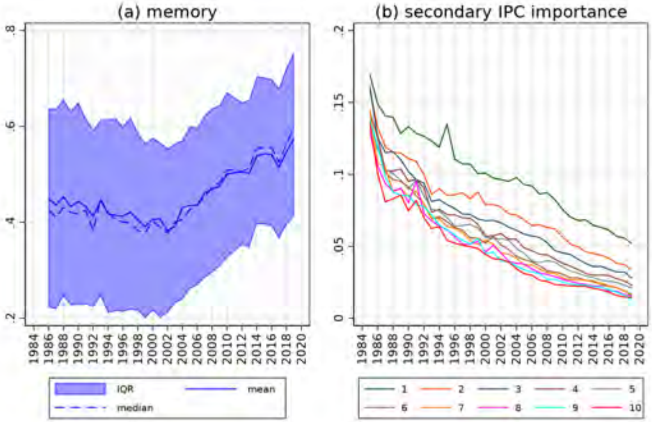


- Overseas patentees have consistently been less important than domestic, even though they account for more than 50% of patents up to mid-2000s
- SOEs and universities/institutes were more important among domestic patentees, but gap disappears by 2003 for universities/institutes and 2013 for SOEs

decomposition | citation-based measure | USPTO results



- Representation = share of patents by each patentee type that are breakthrough minus overall breakthrough patent share



- Secondary IPCs ranked by frequency of being listed as a secondary IPC
- Patenting is becoming "narrower": within-IPC knowledge increasingly more important than across-IPC knowledge

USPTO results

Conclusion

- We use frontier methods (LLM embeddings) and a new theory of innovation to quantify the importance of patents for the direction of innovation in China
- Our measure of patent importance is:
  - positively correlated with citations, grant status
  - a better predictor of firm TFP/output than citations, grant status
- Four key takeaways:
  1. patenting in China has become less innovative and more crowded, with the average importance of important patents declining over time
  2. the stock of breakthrough patents continues to grow but at a declining rate
  3. knowledge inside China has become more important for Chinese innovation than knowledge outside China, especially throughout the 2000s and in technologies targeted by policy
  4. overseas patents are increasingly less important than domestic patents, in stark contrast to patenting in the US
- **Ongoing work**: role of business groups; impact of state policy on innovation; role of key state labs in directing innovation; US-China comparisons

权 利 要 求 书

1.一种电池智能管理方法,其特征在于,所述电池智能管理方法包括:

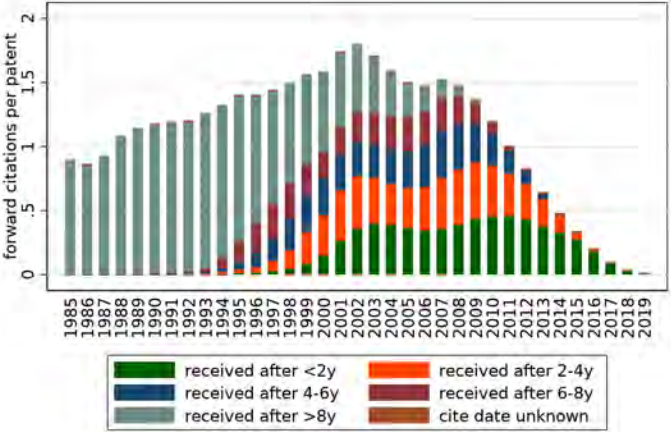
电池智能管理装置侦测预设时间内电网断开累积时间和预设时间内电池工作温度超过预设温度的高温累积时间,并根据侦测结果识别所述电池的供电组件的规格;

Claims

1. A smart battery management method, characterized by the following:

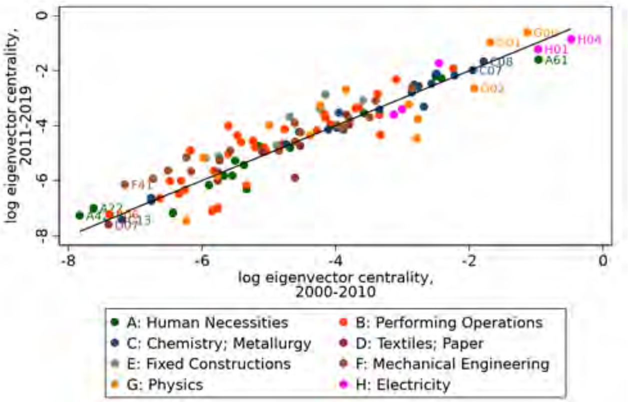
The battery intelligent management device detects the accumulated time when the power grid is disconnected within a preset time and the high-temperature accumulated time when the battery's operating temperature exceeds the preset temperature. It identifies the specification of the power supply assembly of the battery based on the detection results.

back



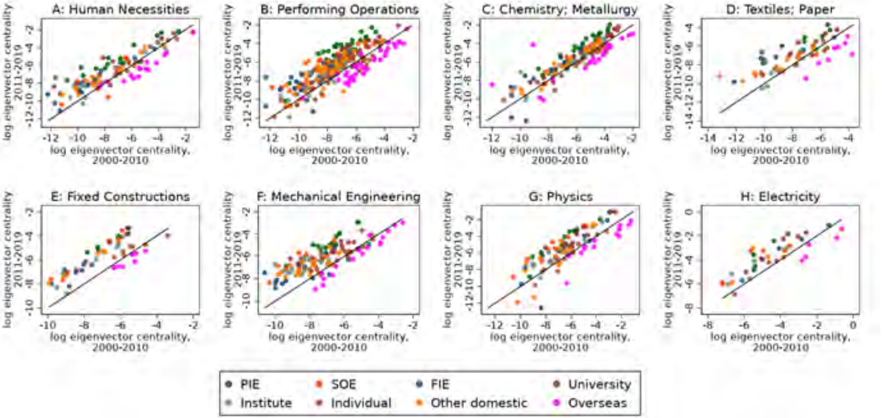
- Most citations are received within 7 years after application
- Citation measures are limited by truncation before 2000 and after 2010

back



- Node = IPC3, link = share of backward citations
- High centralites: G01 - Measuring; G06 - Computing; H01 - Electric Elements; H04 - Electric Communication Techniques; A61 - Medical/Veterinary Science
- Low centralities: A22 - Butchering ; A42 - Headwear; C13 - Sugar; D07 - Ropes

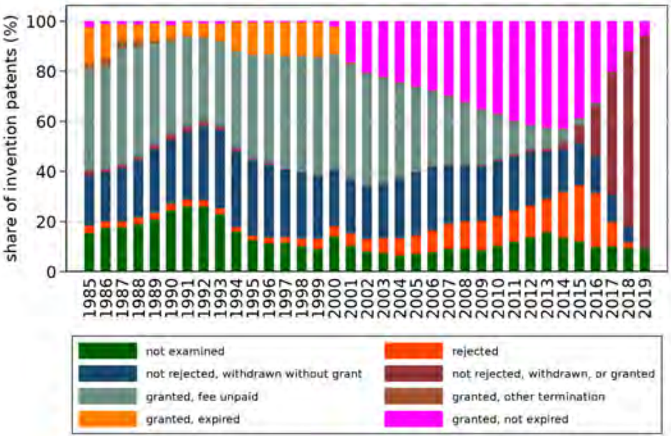
heterogeneity back



- Node = IPC3 x patentee type, link = share of backward citations
- Overseas patents are the most central initially, but centrality declines
- PIE and University patents become the most central over time

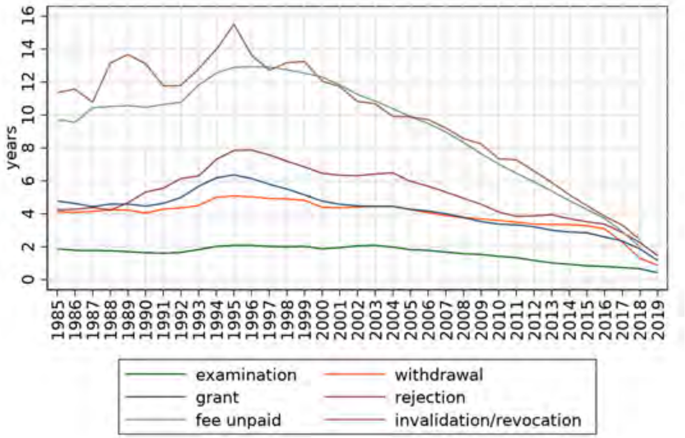
back

1. The quality of patenting: legal status



- Grant rates have been declining since 2002
- Most patents are examined but rejection rate increases over time
- Many patents lapse due to unpaid fees but share declines over time

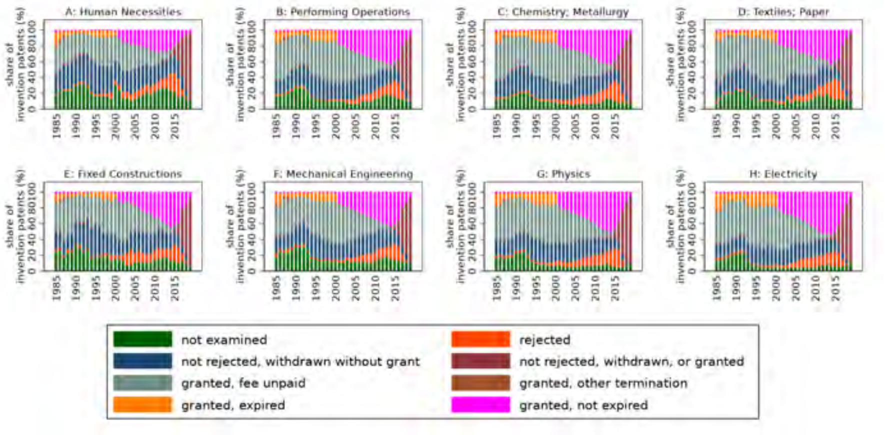
heterogeneity by IPC1 | heterogeneity by patentee type | back



- Average grant lags have been declining since the mid 1990s

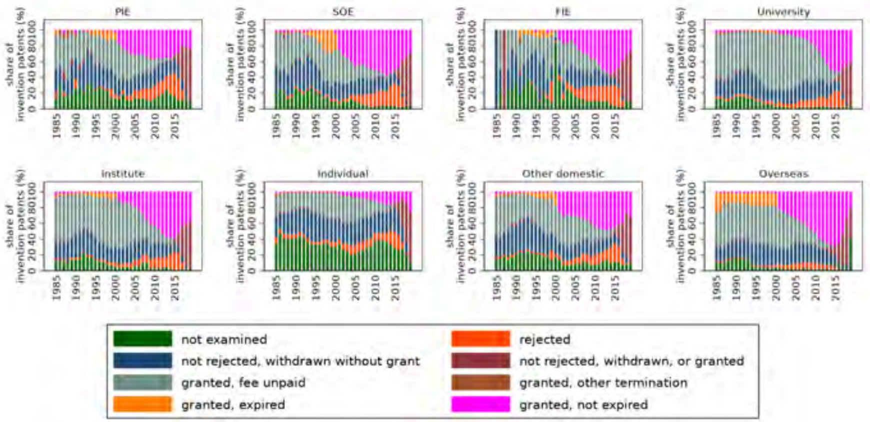
heterogeneity by IPC1 heterogeneity by patentee type back

1. The quality of patenting: legal status



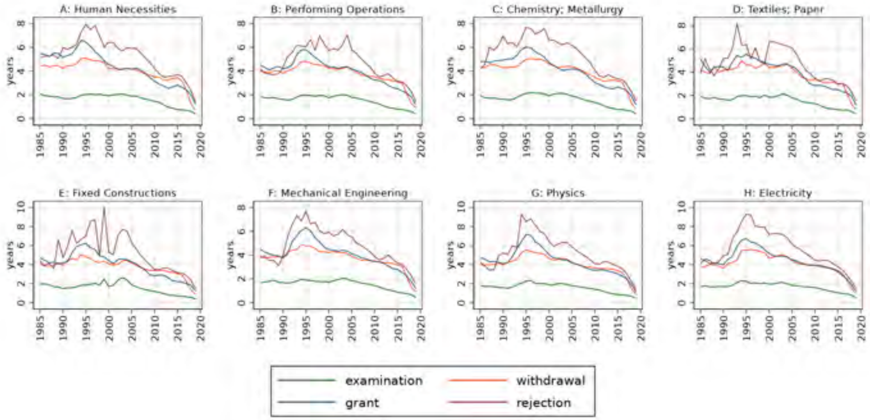
back

1. The quality of patenting: legal status



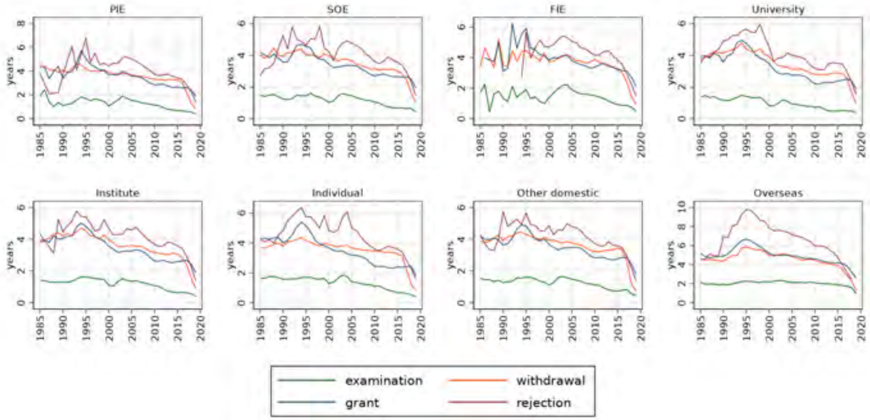
back

1. The quality of patenting: legal status lags

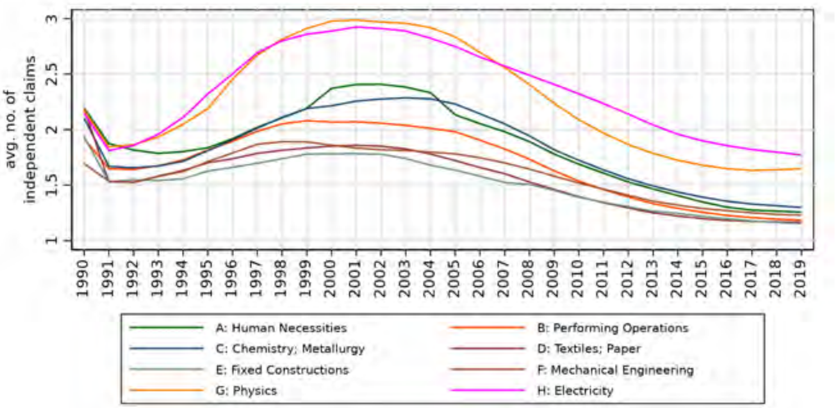


back

1. The quality of patenting: legal status lags

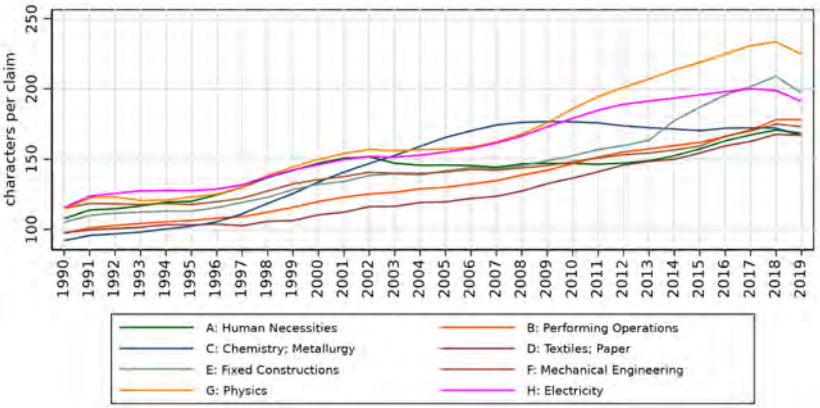


back



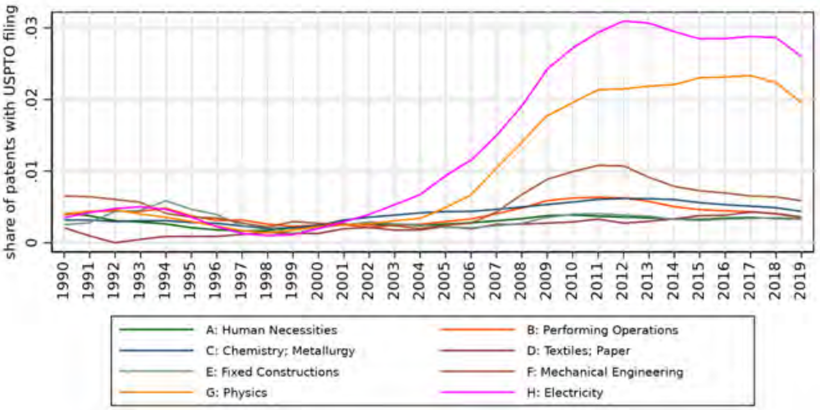
- Average no. of independent claims increasing from 1990-2000, declining after

back



- Average claim length has been increasing over time

back



- Increase in USPTO filing rates in early 2000s, but mainly in Physics, Electricity

back

Cosine similarity

- A patent i applied for at time t is now represented by its embedding, which is a K-dimensional unit row vector $\vec{x}_{it}$:

$$\vec{x}_{it} = \begin{bmatrix} x_{it}^1 & \dots & x_{it}^K \end{bmatrix}$$

- We use **cosine similarity** as the (inverse) distance metric:

$$c_{ijt} \equiv c\left(\vec{x}_{it}, \vec{x}_{jt}\right) = \vec{x}_{it} \cdot \vec{x}_{jt} \in [-1, 1]$$

- This measures the extent to which vectors are "pointing" in the same direction
  - in 2D, equivalent to the cosine of the angle between embeddings
  - invariant to the magnitude (Euclidean norm) of each embedding, which for text embeddings reflects the length of each text

example back

Cosine similarity

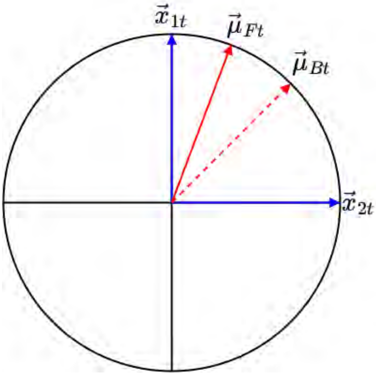
- Why are embeddings better than traditional measures of "meaning" (e.g., frequencies of key phrases)?
- As a thought experiment, suppose that all patents in a given year are represented by a single key phrase and consider two cases:

year	case 1	case 2
t-2	horse carriage	automobile
t-1	horse carriage	automobile
t	electric car	electric car
t+1	electric car	electric car
t+2	electric car	electric car

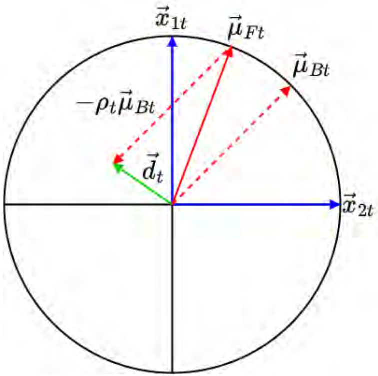
- In both cases, there is zero overlap in key phrases for patents before t and after t
  - traditional measures would view both cases as being equally innovative
- Yet clearly case 1 constitutes a "greater" innovation than case 2
- With the Cohere multilingual model:
  - $c\,(\text{“horse carriage”}, \text{“electric car”}) = 0.90$
  - $c\,(\text{“automobile”}, \text{“electric car”}) = 0.93$
- For reference, standard deviation of cosine similarity across patents is ≈ 0.01

back

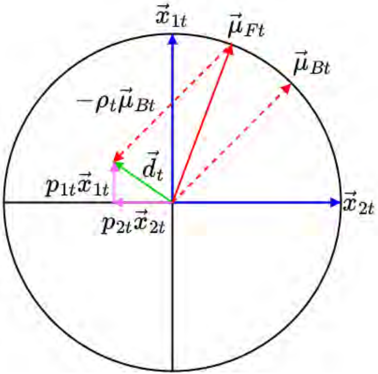
- Example in two dimensions with $||\epsilon_t|| = 0$:



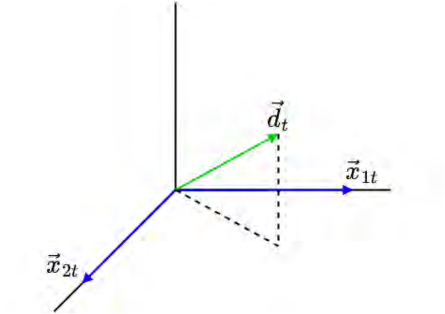
- Example in two dimensions with $||\epsilon_t|| = 0$:



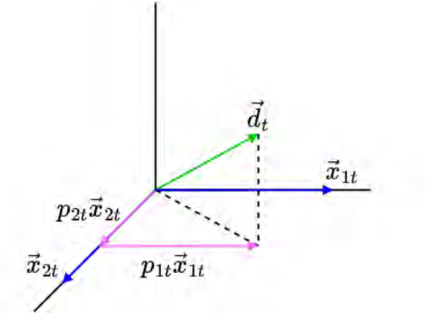
- Example in two dimensions with $||\epsilon_t|| = 0$:



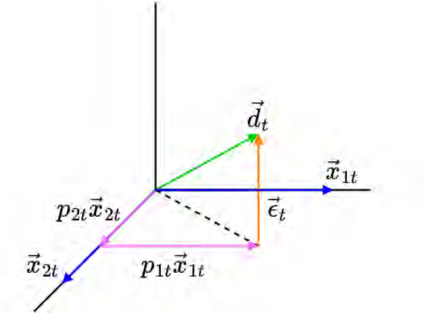
- Example in three dimensions with $||\epsilon_t|| > 0$:



- Example in three dimensions with $||\epsilon_t|| > 0$:



- Example in three dimensions with $||\epsilon_t|| > 0$:



back

External validation: TFP regressions

outcome: log TFP ($\times$100); fixed effects: industry-year year					
$p =$	2	5	10	25	50
i. stock of top-p% important patents, $\beta^{topimp,p}$	0.27	0.26	0.16	0.06	0.08
	(2.55)	(3.96)	(4.58)	(3.88)	(4.32)
ii. stock of top-p% important patents (naive), $\tilde{\beta}^{topimp,p}$	0.62	0.15	0.03	-0.01	-0.05
	(3.94)	(2.51)	(0.86)	(-0.60)	(-2.95)
iii. stock of top-p% cited patents, $\beta^{topcit,p}$	-0.27	-0.08	0.02	0.03	0.03
	(-2.35)	(-1.14)	(0.47)	(1.43)	(2.13)
iv. stock of active patents, β^{active}	-0.05	-0.08	-0.10	-0.09	-0.08
	(-2.56)	(-3.33)	(-3.64)	(-3.25)	(-2.89)
v. has patents, γ	3.65	3.64	3.63	3.64	3.60
	(22.31)	(22.17)	(22.07)	(22.06)	(21.85)
observations (m)	1.78	1.78	1.78	1.78	1.78
R^2	0.74	0.74	0.74	0.74	0.74
adjusted R^2	0.74	0.74	0.74	0.74	0.73

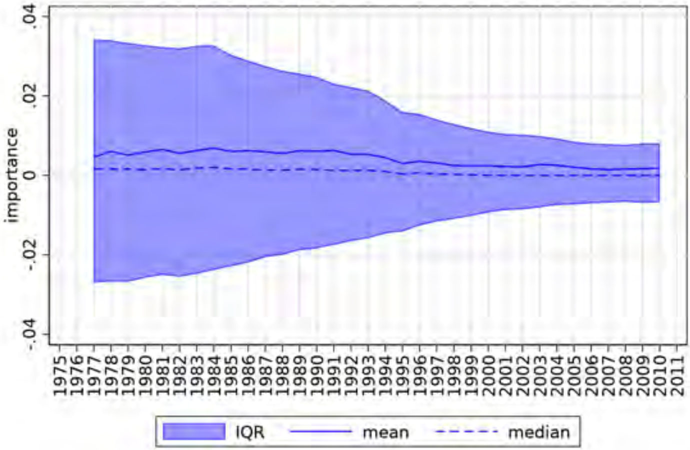
back

External validation: output regressions

outcome: log output ($\times 100$); fixed effects: industry-year					
$p =$	2	5	10	25	50
i. stock of top-p% important patents, $\beta^{topimp,p}$	5.31	5.94	3.72	1.80	2.41
	(14.07)	(26.05)	(30.78)	(33.21)	(35.64)
ii. stock of top-p% important patents (naive), $\tilde{\beta}^{topimp,p}$	0.59	-2.22	-2.34	-1.73	-2.17
	(1.07)	(-10.29)	(-21.26)	(-28.90)	(-33.95)
iii. stock of top-p% cited patents, $\beta^{topcit,p}$	0.20	2.81	3.24	2.11	1.17
	(0.49)	(11.47)	(21.83)	(26.46)	(23.55)
iv. stock of active patents, β^{active}	0.32	-0.12	0.04	0.29	0.37
	(4.37)	(-1.39)	(0.44)	(2.98)	(3.76)
v. has patents, γ	110.61	109.72	109.02	108.65	108.39
	(191.40)	(189.39)	(187.82)	(186.92)	(186.29)
observations (m)	1.78	1.78	1.78	1.78	1.78
R^2	0.15	0.15	0.15	0.15	0.15
adjusted R^2	0.15	0.15	0.15	0.15	0.15

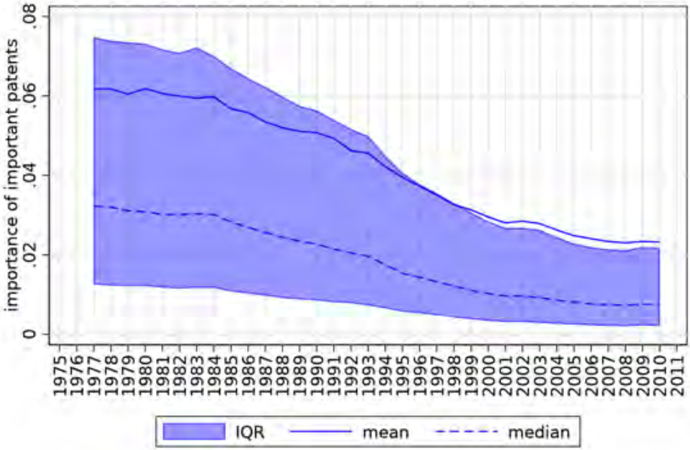
back

The distribution of patent importance - granted USPTO patents



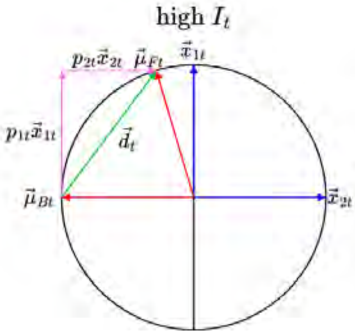
back

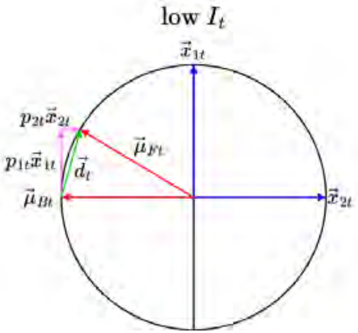
Importance of important patents - granted USPTO patents



back

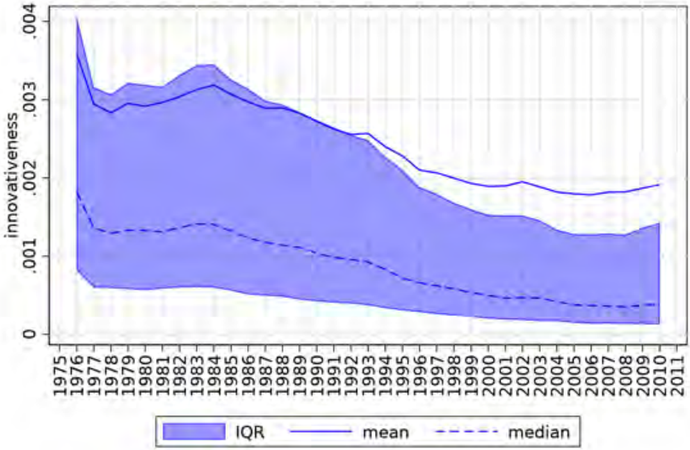
- Compare the following two cases:





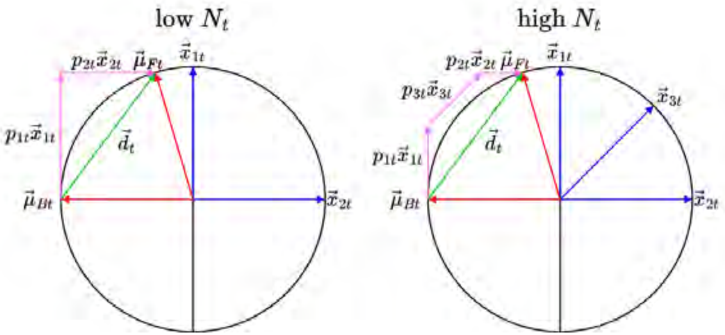
back

Innovativeness - granted USPTO patents



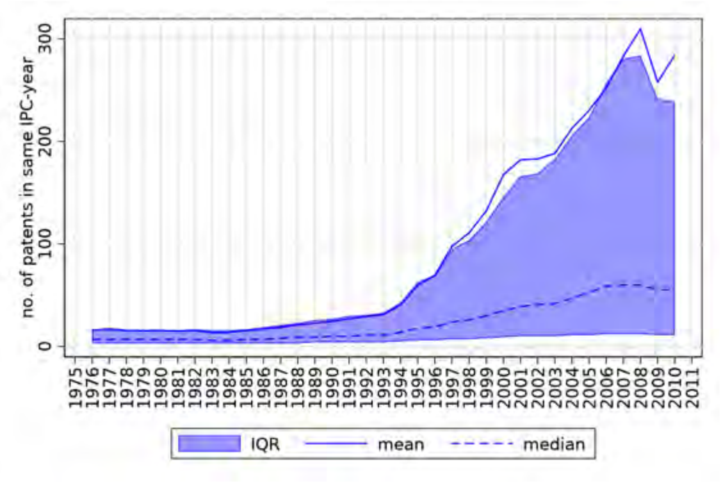
back

- Compare the following two cases:



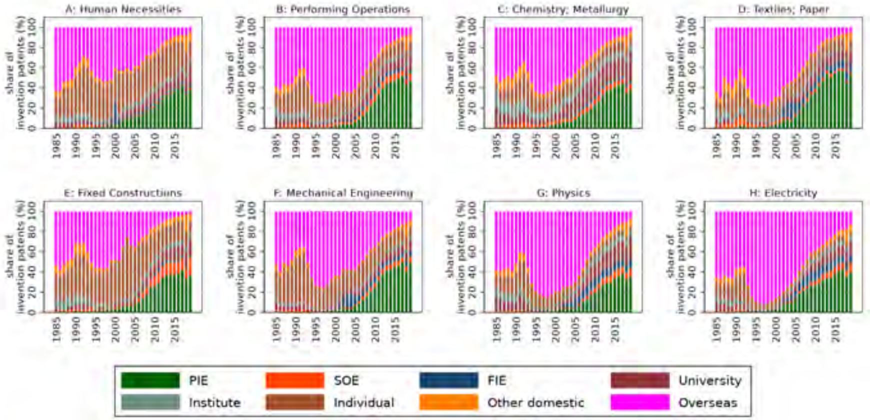
back

Crowdedness - granted USPTO patents

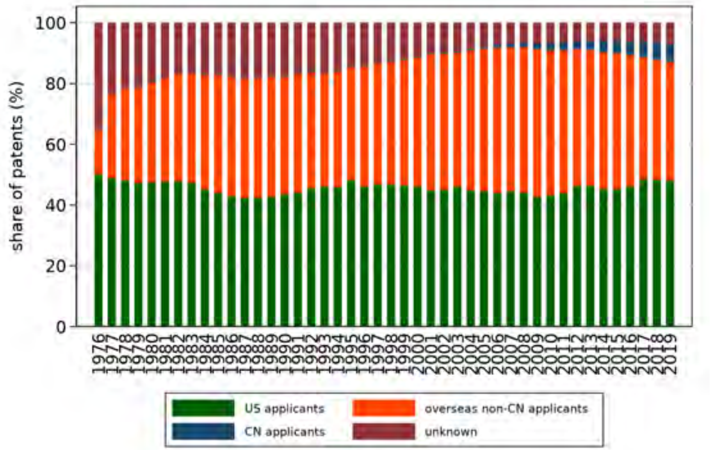


back

Share of patenting activity across patentee types - by IPC section



back


share of patents (%)
0
20
40
60
80
100
US applicants
overseas non-CN applicants
CN applicants
unknown

back

Decomposing patent importance by patentee type

- We can further decompose patent importance by patentee type pair:

$$p_{it}^{g} \approx \sum_{r} s_t^{gr} p_{it}^{gr}$$

  - p_{it}^{gr}: importance of patent i for patents in IPC g owned by patentee type r
  - s_t^{gr}: share of patents in IPC g owned by patentee type r

Decomposing patent importance by patentee type

- We can further decompose patent importance by patentee type pair:

$$p_{it}^{g} \approx \sum_{r} s_{t}^{gr} p_{it}^{gr}$$

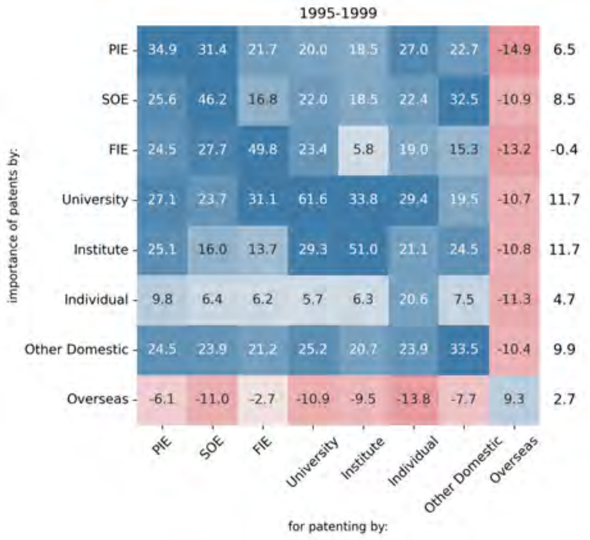
  - p_{it}^{gr}: importance of patent i for patents in IPC g owned by patentee type r
  - s_{t}^{gr}: share of patents in IPC g owned by patentee type r
- Decomposition is exact under OLS given linearity
  - hence approximation arises from our use of the LASSO estimator

- We can further decompose patent importance by patentee type pair:

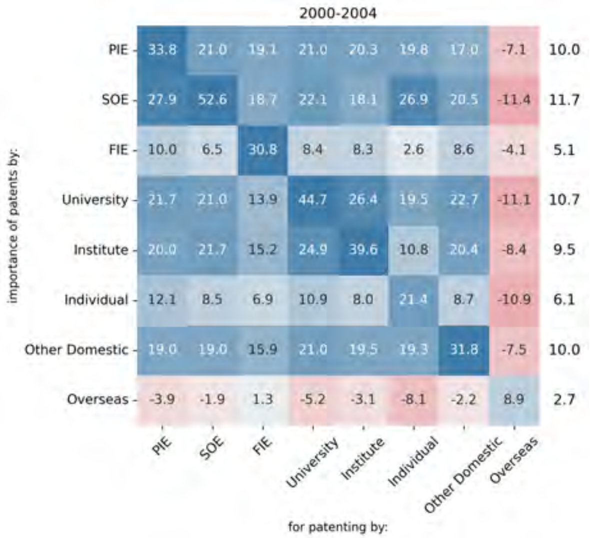
$$p_{it}^{g} \approx \sum_{r} s_{t}^{gr} p_{it}^{gr}$$

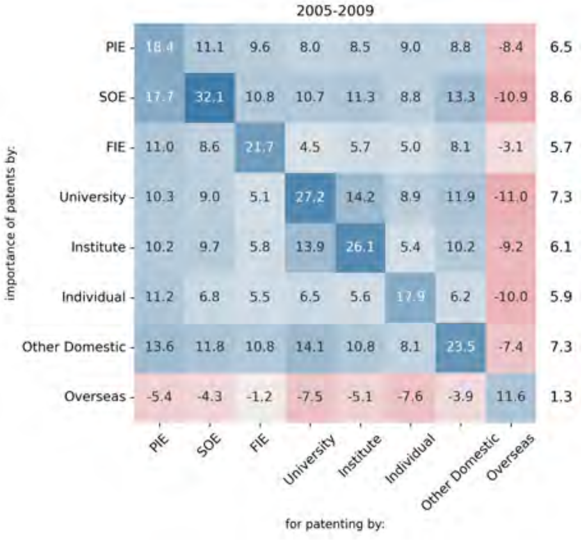
  - p_{it}^{gr}: importance of patent i for patents in IPC g owned by patentee type r
  - s_{t}^{gr}: share of patents in IPC g owned by patentee type r
- Decomposition is exact under OLS given linearity
  - hence approximation arises from our use of the LASSO estimator
- To measure importance of patents by patentee type r' for patenting by patentee type r, compute average of p_{it}^{gr} over patents i owned by patentee type r'

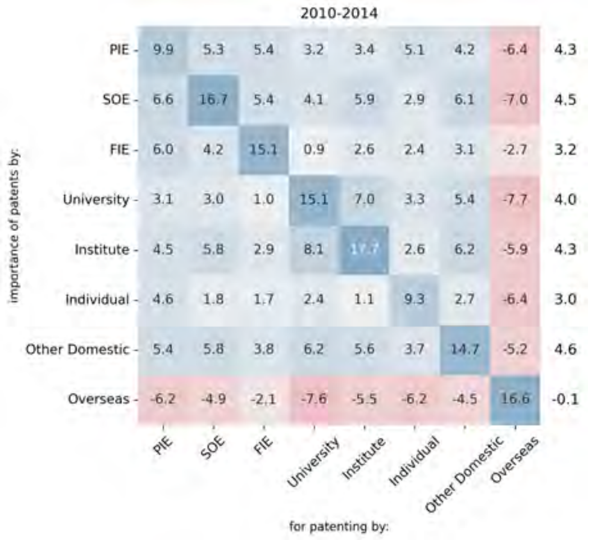
Decomposing patent importance by patentee type



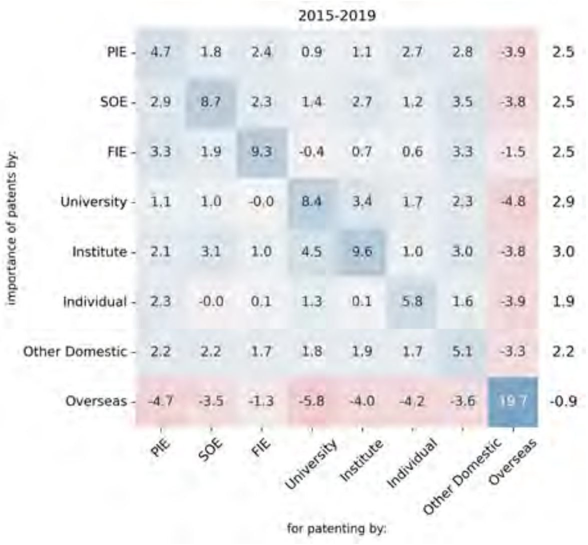
Decomposing patent importance by patentee type




2005-2009
importance of patents by:
PIE 18.4 11.1 9.6 8.0 8.5 9.0 8.8 -8.4 6.5
SOE 17.7 32.1 10.8 10.7 11.3 8.8 13.3 -10.9 8.6
FIE 11.0 8.6 21.7 4.5 5.7 5.0 8.1 -3.1 5.7
University 10.3 9.0 5.1 27.2 14.2 8.9 11.9 -11.0 7.3
Institute 10.2 9.7 5.8 13.9 26.1 5.4 10.2 -9.2 6.1
Individual 11.2 6.8 5.5 6.5 5.6 17.9 6.2 -10.0 5.9
Other Domestic 13.6 11.8 10.8 14.1 10.8 8.1 23.5 -7.4 7.3
Overseas -5.4 -4.3 -1.2 -7.5 -5.1 -7.6 -3.9 11.6 1.3
PIE
SOE
FIE
University
Institute
Individual
Other Domestic
Overseas
for patenting by:

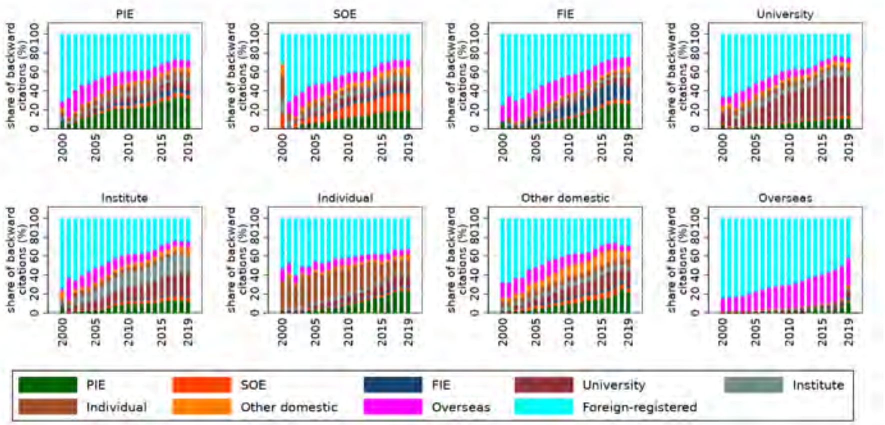

2010-2014
importance of patents by:
PIE 9.9 5.3 5.4 3.2 3.4 5.1 4.2 -6.4 4.3
SOE 6.6 16.7 5.4 4.1 5.9 2.9 6.1 -7.0 4.5
FIE 6.0 4.2 15.1 0.9 2.6 2.4 3.1 -2.7 3.2
University 3.1 3.0 1.0 15.1 7.0 3.3 5.4 -7.7 4.0
Institute 4.5 5.8 2.9 8.1 17.7 2.6 6.2 -5.9 4.3
Individual 4.6 1.8 1.7 2.4 1.1 9.3 2.7 -6.4 3.0
Other Domestic 5.4 5.8 3.8 6.2 5.6 3.7 14.7 -5.2 4.6
Overseas -6.2 -4.9 -2.1 -7.6 -5.5 -6.2 -4.5 16.6 -0.1
PIE
SOE
FIE
University
Institute
Individual
Other Domestic
Overseas
for patenting by:

Decomposing patent importance by patentee type



back

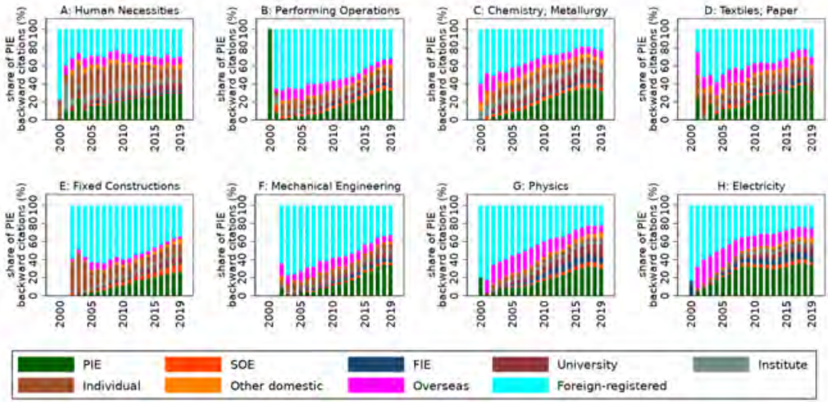
Citation-based measure of dependence



- All domestic patentees reduce dependence on foreign-registered patents
- Overseas patentees still rely heavily on overseas and foreign-registered patents

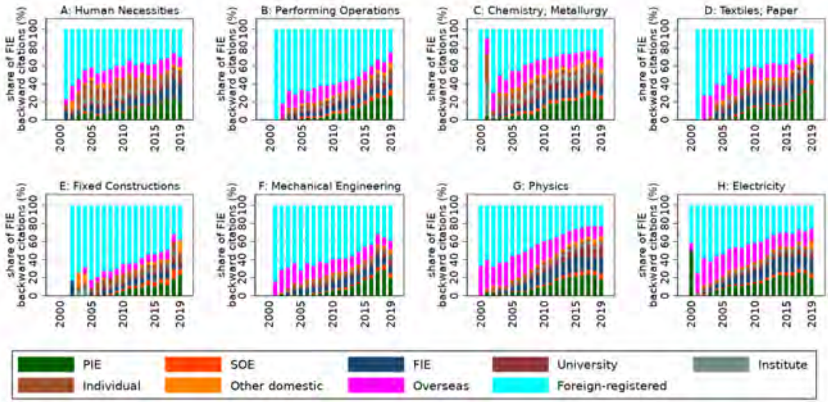
back | PIE citation heterogeneity | FIE citation heterogeneity | overseas citation heterogeneity

Citation-based measure of dependence - PIE patentees



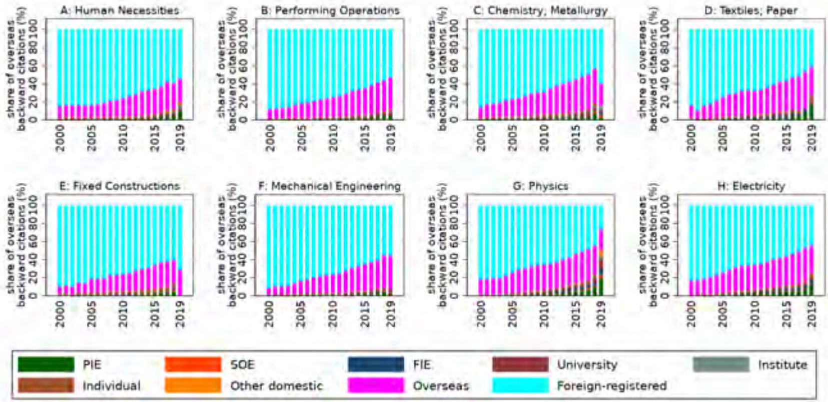
back

Citation-based measure of dependence - FIE patentees



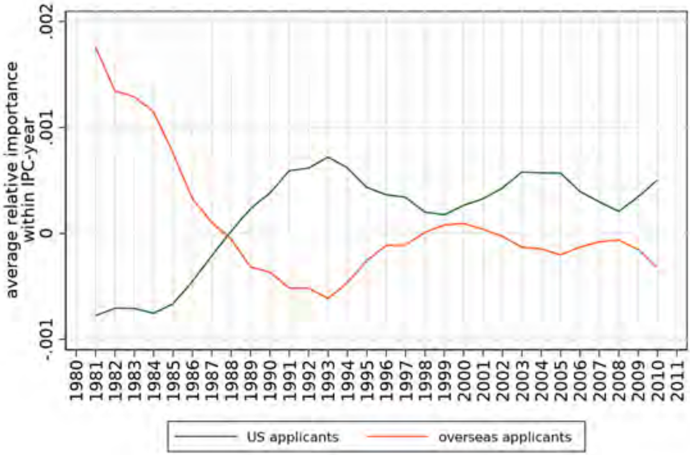
back

Citation-based measure of dependence - overseas patentees

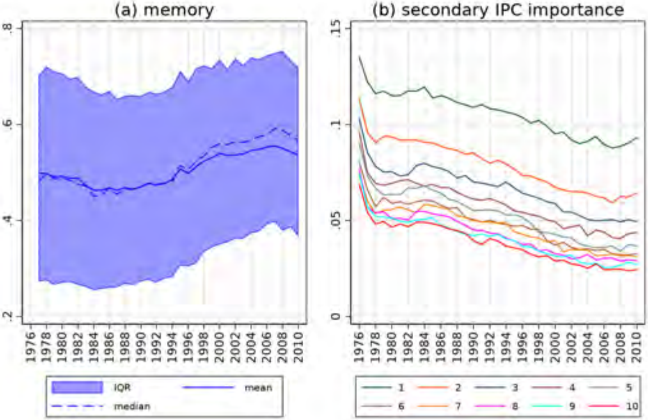


back

Relative patent importance by patentee type - granted USPTO patents



back


(a) memory
.2
.4
.6
.8
1976 1978 1980 1982 1984 1986 1988 1990 1992 1994 1996 1998 2000 2002 2004 2006 2008 2010
IQR
mean
median
(b) secondary IPC importance
0
.05
.1
.15
1976 1978 1980 1982 1984 1986 1988 1990 1992 1994 1996 1998 2000 2002 2004 2006 2008 2010
1 2 3 4 5
6 7 8 9 10

back