


## Modeling the dynamics of firms' technological impact

Shuqi Xu(徐舒琪), Manuel Sebastian Mariani, and Linyuan Lü(吕琳媛)
**Citation:**Chin. Phys. B, 2021, 30 (**12**): 120517. DOI: 10.1088/1674-1056/ac364c
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

**What follows is a list of articles you may be interested in**

---

## Stability analysis of hydro-turbine governing system based on machine learning

Yuansheng Chen(陈元盛) and Fei Tong(仝飞)
Chin. Phys. B, 2021, 30 (**12**): 120509. DOI: 10.1088/1674-1056/ac04a9

## Nonequilibrium free energy and information flow of a double quantum-dot system with Coulomb coupling

Zhiyuan Lin(林智远), Tong Fu(付彤), Juying Xiao(肖菊英), Shanhe Su(苏山河), Jincan Chen(陈金灿), and Yanchao Zhang(张艳超)
Chin. Phys. B, 2021, 30 (**8**): 080501. DOI: 10.1088/1674-1056/abe119

## Collective stochastic resonance behaviors of two coupled harmonic oscillators driven by dichotomous fluctuating frequency

Lei Jiang(姜磊), Li Lai(赖莉), Tao Yu(蔚涛), Maokang Luo(罗懋康)
Chin. Phys. B, 2021, 30 (**6**): 060502. DOI: 10.1088/1674-1056/abd9b0

## Stable water droplets on composite structures formed by embedded water into fully hydroxylated *β*-cristobalite silica

Hanqi Gong(龚菡琪), Chonghai Qi(齐崇海), Junwei Yang(杨俊伟), Jige Chen(陈济舸), Xiaoling Lei(雷晓玲), Liang Zhao(赵亮), and Chunlei Wang(王春雷)
Chin. Phys. B, 2021, 30 (**1**): 010503. DOI: 10.1088/1674-1056/abc2b7

## A new control method based on the lattice hydrodynamic model considering the double flux difference

Shunda Qin(秦顺达), Hongxia Ge(葛红霞), Rongjun Cheng(程荣军)
Chin. Phys. B, 2018, 27 (**5**): 050503. DOI: 10.1088/1674-1056/27/5/050503

---

SPECIAL TOPIC — Interdisciplinary physics: Complex network dynamics and emerging technologies

# Modeling the dynamics of firms' technological impact*

Shuqi Xu(徐舒琪)[1], Manuel Sebastian Mariani[1,2,†], and Linyuan Lü(吕琳媛)[1,3,‡]

[1] *Yangtze Delta Region Institute (Huzhou) & Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Huzhou* 313001, *China*
[2] *University Research Priority Program on Social Networks, University of Zurich, CH*-8050 *Zurich, Switzerland*
[3] *Beijing Computational Science Research Center, Beijing* 100193, *China*

(Received 31 July 2021; revised manuscript received 17 September 2021; accepted manuscript online 4 November 2021)

Recent studies in complexity science have uncovered temporal regularities in the dynamics of impact along scientific and other creative careers, but they did not extend the obtained insights to firms. In this paper, we show that firms' technological impact patterns cannot be captured by the state-of-the-art dynamical models for the evolution of scientists' research impact, such as the $Q$ model. Therefore, we propose a time-varying returns model which integrates the empirically-observed relation between patent order and technological impact into the $Q$ model. The proposed model can reproduce the timing pattern of firms' highest-impact patents accurately. Our results shed light on modeling the differences behind the impact dynamics of researchers and firms.

**Keywords:** firm technological impact, patent analysis, impact dynamics

**PACS:** 05.90.+m, 89.65.Gh, 07.05.Tp **DOI:** 10.1088/1674-1056/ac364c

## 1. Introduction

Complexity science methods have been recently applied to understand the dynamics of impact along creative careers.[1–7] In particular, researchers have unveiled remarkable regularities through the quantitative analysis of individuals' creative works (such as scientists' articles and directors' movies). These works found that the career trajectories of scientists,[1,8] artists,[8] directors,[6,8] and best-seller authors,[6,9] among others, are determined by a uniform pattern called random-impact rule: the most important works occur randomly within an individual's work sequence. Sinatra *et al.* emphasized that the ability of an individual to produce high-impact works is independent of his/her career stage,[1] based on which they developed the $Q$ model, where the impact of an individual's works is determined by a random factor, reflecting the potential of the idea, and a parameter $Q$, representing the individual's capability to improve the work. According to the model, high-impact discoveries require a combination of an excellent idea and high $Q$. These methods and findings are mainly focused on individuals' creative careers, and they have not been extended to model the impact of organizations, such as firms.

Existing evidence for firms suggests discrepancies among the impact dynamics of firms and individuals. Vast management literature acknowledged that firm age is an essential determinant of a firm's impact.[10,11] And it is commonly recognized that young firms tend to present higher capability of innovation.[12,13] As a consequence, firm age might be negatively related to the impact of its technology. Recently, Xu *et al.*[14] provided initial evidence for systematic differences between temporal patterns of the highest-impact works in scientists' careers and firms' lifecycles. However, this work focused on the empirical evidence, and it did not propose a dynamical model to describe firms' impact evolution. How to modify existing impact dynamical models to accurately reproduce the observed firms' impact patterns?

To answer this question, we analyze the patent licensing histories of worldwide firms from the U.S. Patent and Trademark Office (USPTO) and the European Patent Office (EPO). We propose a new dynamical model to capture the observed dynamic regularities. Specifically, by analyzing the evolution of average patent impact within firms' lifecycles, we observe a constantly decreasing trend, which indicates that firms tend to receive diminishing returns from their patents as they are issued more patents. We show that the $Q$ model is unable to capture firms' empirical impact patterns. Motivated by this finding, we generalize the $Q$ model by incorporating the property that patents' impact depends not only on the firm's capability and the potential of the patent idea, but also on the patent's position within the firm's lifecycle. Therefore, we propose a time-varying returns model to reproduce the observed patterns of the dynamics of firms' impact. We show that for both two

*Project supported by the National Natural Science Foundation of China (Grant Nos. 61673150 and 11622538). M.S.M. acknowledges financial support from the URPP Social Networks at the University of Zurich, and the UESTC professor research start-up (Grant No. ZYGX2018KYQD215). L.L. acknowledges the Science Strength Promotion Programme of UESTC, Chengdu.

†Corresponding author. E-mail: manuel.mariani@business.uzh.ch
‡Corresponding author. E-mail: linyuan.lv@uestc.edu.cn

© 2021 Chinese Physical Society and IOP Publishing Ltd 

analyzed datasets, the proposed model matches the empirical results more accurately than the state-of-art random-impact model and $Q$ model. Our study sheds light on modeling the patterns of firms' impact dynamics, which can provide useful insights for managerial practices.

## 2. Method

### 2.1. Patent data of firms

Patent activity is usually considered as a rich source to reflect firms' impact in terms of technology.[15–17] In this study, we investigate two large patent datasets: (i) The record of the patents granted by the EPO to 163565 firms from 1978 to 2016, covering 772127 patents. Note that the original data involves all types of applicants which are not limited to firms. In the analysis, we identify firm applicants by specific name identifiers, see Appendix A for details. (ii) The patent data from the USPTO to 7440 firms from 1926 to 2017, amounting to 2458402 patents.[18,19] Only patents whose applicant is a listed firm on the US stock market are included in the data. For each patent, the two datasets both include patent number, issuance date, names and ids of applicant firms, and a list of patent references. Therefore, we can identify the applicant firms of each patent, and based on the reference information, we can calculate the citation count of each patent at any given time. We summarize the main properties of the two studied datasets in Table 1.

**Table 1.** Datasets description.

| Datasets | Patents | Firms | Average number of patents per firm | Analyzed firms | Time range |
|---|---|---|---|---|---|
| EPO | 772127 | 163565 | 5.0 | 3548 | 1978–2016 |
| USPTO | 2458402 | 7440 | 330.6 | 2819 | 1926–2017 |

EPO and USPTO are regarded as the most representative and highly reliable databases for patent analysis and have been used in many information science and cross-disciplinary studies.[20–23] Table 1 demonstrates that at the patent level, USPTO contains more patents than EPO. Kim *et al.*[20] argued that the United States is considered as one of the world's largest commercial and technological markets, and for this reason, international firms tend to file patents in the USPTO to claim their rights to protect the innovation in the US. As for the EPO data, it contains rich information on technological innovations mostly related to firms located in European countries. At the firm level, the USPTO data involves fewer firms than the EPO data, because all firms in the USPTO data are limited to listed firms, whereas there is no such restriction for firms in the EPO data.

We further show the cumulative distribution of the number of issued patents for firms in the two datasets in Fig. 1(a). We observe that firms' patenting activities is heavier in the USPTO than in the EPO. In the EPO data, 61.5% of firms have been issued only one patent, 93.9% of firms have been issued no more than 10 patents. For the USPTO data, there are only 13.44% of firms that have been issued one patent in total, 82.34% of firms have been issued no more than 100 patents. One of the main reasons for this difference is that the listed firms usually have an advantage over ordinary firms in terms of resource and strength (for example, in terms of finance and innovation), which allows them to issue more patents.

Table 2 reports the top five firms issued the largest number of patents in the two datasets. In the EPO data, the top firms are mostly from Japan, the United States, and European countries; as for the USPTO data, most of the top firms are from the United States and Japan.

In the following, we only consider patents issued up to 2006. In this way, we have at least a 10-year time window to observe their impact. Furthermore, the research subjects of this study are firms, we need to focus on firms with sufficiently active patenting activities. For this reason, we limit our firm-level analysis to firms that have been issued at least 20 patents for the EPO data, which refers to 3548 firms, and at least 15 patents for the USPTO data, which filters 2819 firms (see Table 1). Our conclusions are robust with respect to alternative choices of the filter standard of analyzed firms (see Appendix C for the results of filtering firms based on their number of years of patenting activities).

**Table 2.** Top firms by the number of issued patents. Time range refers to the first and last patent issuance year.

| EPO | Firm | $N$ | Country | Time range |
|---|---|---|---|---|
| 1 | CANON CO LTD | 6472 | JP | 1980–2007 |
| 2 | IBM CORP | 4864 | US | 1980–2001 |
| 3 | SIEMENS AG | 4653 | DE | 1980–1996 |
| 4 | GENERAL ELECT CO | 3813 | US | 1985–2016 |
| 5 | PHILIPS ELECTS NV | 3671 | NL | 1979–1995 |
| USPTO | Firm | $N$ | Country | Time range |
| 1 | INTERNATIONAL BUSINESS MACHS CORP | 123220 | US | 1926–2017 |
| 2 | GENERAL ELECTRIC CO | 75381 | US | 1926–2017 |
| 3 | HITACHI LIMITED | 52325 | JP | 1973–2012 |
| 4 | CANON INC | 41095 | JP | 1976–2017 |
| 5 | PANASONIC CORP | 37950 | JP | 1974–2012 |

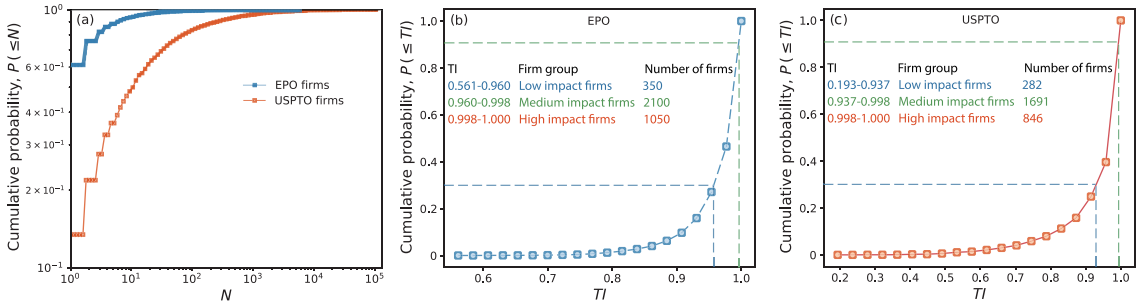



**Fig. 1.** Cumulative distribution of firms' patent number and technological impact. (a) Cumulative distribution of the number of issued patents for firms in the EPO and USPTO data. (b)–(c) Cumulative distribution of $TI$ for firms in EPO (b) and USPTO (c) data, as well as the impact thresholds that distinguish the three firm groups of impact.

### 2.2. Quantifying technological impact

**Patent's technological impact.** To evaluate the novelty of the innovation seeking patent protection, patent applicants are required to disclose the prior knowledge on which they relied,[24] represented by citations to prior patents. Citations are proposed by the patent applicant, and subsequently checked and corrected by the patent examiner.[25,26] Because of this, patents' citation count is the most used metric for patents' technological impact.[23,27] Despite its wide use, citation count suffers from a serious age bias by favoring older patents over more recent ones.[28–30] The main reason is that it takes time for a patent to accumulate citations, and the preferential attachment (also known as the "rich-get-richer" mechanism)[31] further amplifies discrepancies in the citation count of patents issued at different times.[32,33] The age bias makes it unreliable to directly compare the citation count of patents issued in different years.[34] To overcome this issue, following Ref. [14], we measure a given patent's technological impact in terms of its normalized citation count ($NC$), which corresponds to its ranking by citation count when compared against all patents issued in the same year. In this way, $NC \in [0,1)$ close to one (zero) corresponds to high (low) impact. Normalized citation count can effectively suppress the age bias held by citation count,[14] thus allowing us to compare on the same scale the impact of patents of different ages.

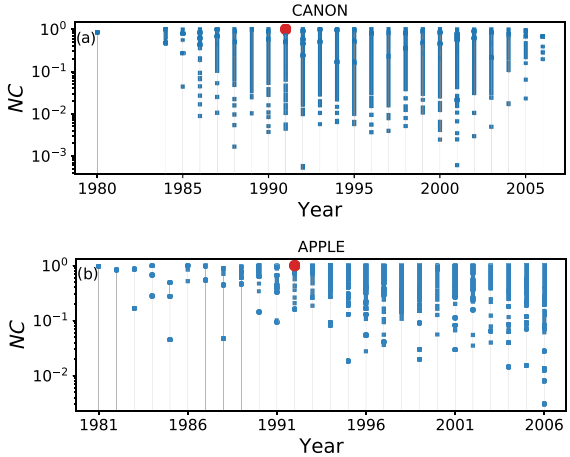



**Fig. 2.** An illustration of two firms' lifecycles. Each dot represents a patent, and the $x$-axis refers to the issuance year of patents, $y$-axis refers to the technological impact ($NC$) of each patent, red dots denote breakthrough patents (patents with the highest $NC$). (a) Firm CANON's lifecycle from the EPO data. (b) Firm APPLE's lifecycle from the USPTO data.

**Firm's technological impact and classification.** Motivated by previous works,[14,35,36] a firm's lifecycle is defined as a time-ordered sequence of its issued patents (see Fig. 2). And we define a given firm's technological impact ($TI$) as the impact of its highest-impact patent, which we refer to as the firm's "breakthrough". Based on the technological impact of all analyzed firms, we classify firms into three groups: we consider the top 10% firms as high-impact firms, the bottom 30% firms as low-impact firms, and the remaining 60% firms as medium-impact firms, see Figs. 1(b) and 1(c) for the classification details.

### 2.3. Modeling impact dynamics

To describe the temporal dynamics of the technological impact in firms' patenting lifecycles, we resort to three models to reproduce patents' impact.

**State-of-the-art models and their shortcomings.** We consider two models from previous works:[1] the random-impact model and the $Q$-model. The random-impact model assumes that the impact of each of a firm's issued patents is a random variable. Specifically, the model keeps a given firm's total patent number unchanged, and it extracts the impact of each of the firm's patents from the impact values of patents issued in that issuance year (without replacement). According to the random-impact model, the only difference between two firms is their patent productivity: a productive firm is more likely to have a breakthrough patent with high technological impact merely because of a larger number of attempts with the same likelihood to have high impact.

The $Q$ model assumes that each firm has an capability value $Q$, which is unique to the firm and stable throughout the firm's lifecycle. $Q$ captures the capability of firms to improve ($Q > 1$) or reduce ($Q < 1$) the impact of a patent. Specifically, in the model, a firm $\alpha$ randomly chooses a patent idea $i$ with potential impact $r_i$, and it modulates the idea according to its $Q$, resulting in a patent of impact,

$$R_{i\alpha} = r_i Q_\alpha, \tag{1}$$

where $r_i$ is a patent-dependent random factor which is lognormal distributed with mean 1, identical for all firms.

The above two models were both motivated by the empirical findings on scientists' careers,[1] in which the highest-impact works are randomly distributed among individuals' careers. For firms, the validity of the two models has not been examined.

**Time-varying returns model.** In the $Q$ model, a firm's early and late patents have the same probability to be the breakthrough patents, and the potential impact ($r$) of a patent idea itself and the firm's capability ($Q$) are the only two factors that determine the impact of a patent, independently of the patent's order within the applicant firm's lifecycle. However, in management literature, many studies[12,13,37] have also uncovered the role of age played in firms' technological quality. These works argued that as firms become more experienced, the cost and benefits brought by innovation change, which suggests that firms may receive time-varying returns from their patents depending on the stage in their lifecycles. Inspired by these findings, based on the $Q$ model, we propose a time-varying returns model which assumes that a firm will enhance or diminish the potential impact of a patent based on its inherent ability combined with its age, quantified by its number of previously issued patents. In formulas, the impact of patent $i$ from firm $\alpha$ is modeled as

$$R_{i\alpha} = r_i Q_\alpha \Phi\left(\frac{n_i}{N_\alpha}\right), \tag{2}$$

where $\Phi$ is a time-varying function representing the influence of firm age on patent impact, $n_i \in \{1,\ldots,N_\alpha\}$ is the position of patent $i$ within the time-ordered sequence of firm $\alpha$'s patents, and $N_\alpha$ is the total patent number issued to firm $\alpha$. In the next section, we will show that a monotonously decreasing function exhibits an accurate fit to the empirical data, which we refer to as a diminishing returns effect.

## 3. Results

### 3.1. Productivity and impact dynamics in firms' patenting lifecycles

To understand the patenting behaviors of firms, we firstly analyze the productivity (i.e. number of issued patents) and impact dynamics for different groups of firms in the two datasets. Figures 3(a) and 3(b) show that $N(\tau)$, the total number of patents that firms are issued to up to age $\tau$ after their first patent issuance, is well matched by a sigmoidal function on a logarithmic scale for all the three groups of firms, i.e., $N(\tau) \sim \exp(\frac{1}{1+e^{-b(\log\tau - a)}})$. This is markedly distinct from the scaling patterns of scientists' productivity,[38] which reveals that firms generally experience periods of emergence, growth, and maturity in terms of their patenting activities. The results also indicate that for firms in the USPTO data (Fig. 3(b)), the duration of the growth stage is longer on average, which confirms the distinction between firms in the EPO and USPTO datasets (see Subsection 2.1). Comparing parameter $b$ of the

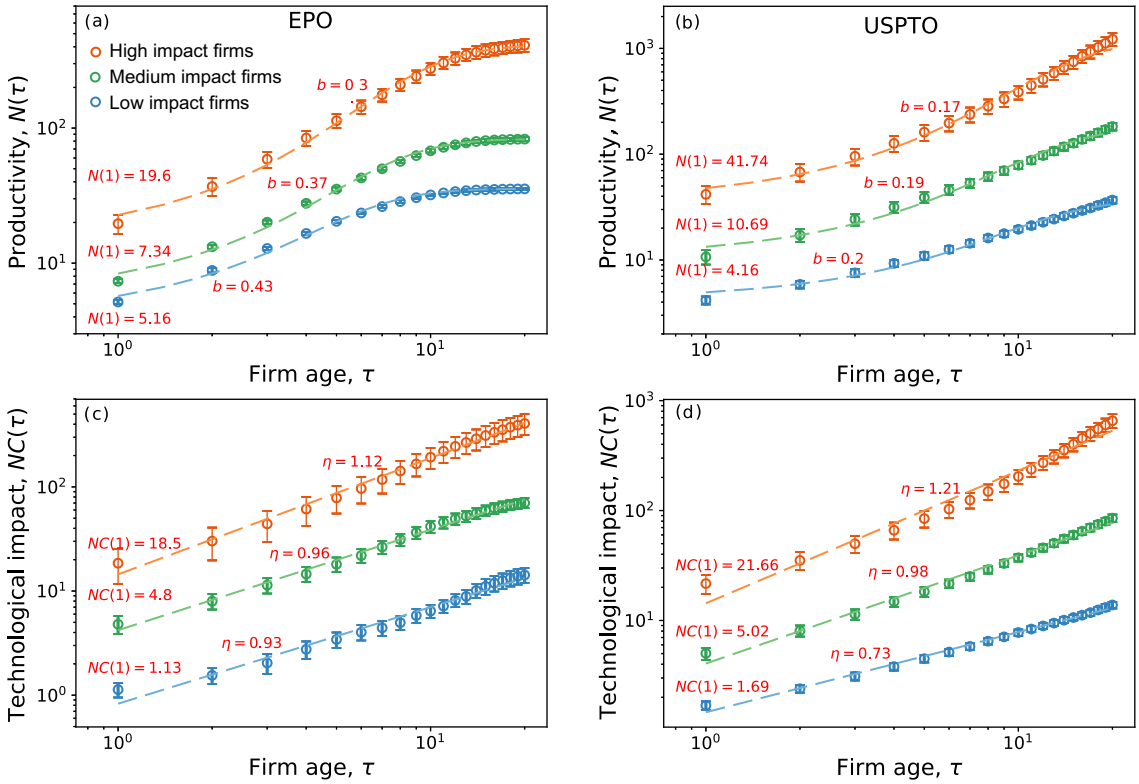



**Fig. 3.** Temporal dynamics of patent productivity and technological impact during firms' lifecycles. The panels show the temporal trends of the total number of issued patents (a), (b) and cumulative technological impact (c), (d) for the three different groups of firms. Only firms with at least 20-year lifetime are involved in this analysis. Error bars refer to the standard error of the mean. The results show that there are sustained differences across the three impact groups of firms, and the growing trends of productivity and impact are well-described by sigmoidal and power-law functions, respectively, for both datasets. The fitting parameters are shown along the curves.

fitting curves, we find that high-impact firms tend to have smaller $b$ than other firms in both datasets, this reflects that the development of the former kind of firms is more rapid. Figures 3(c) and 3(d) demonstrate that $NC(\tau)$, the cumulative technological impact of firms' patents that were issued within firms' $\tau$-year lifecycles, can be described by a power-law pattern ($NC(\tau) \sim \tau^{\eta}$). And the increase is much faster for high-impact firms, for which the exponent $\eta$ is larger than that for medium and low-impact firms. Comparing firms in EPO (Fig. 3(c)) and USPTO (Fig. 3(d)) data, the results also indicate that the latter show a stronger growth trend, reflected by a larger power exponent. We further show the distribution of $b$ and $\eta$ in Appendix B, Fig. B1, which also indicates the striking gaps between the three groups of firms.

### 3.2. Timing pattern of the firms' breakthrough patents

We focus here on the recent finding uncovered by the recent study[14] that breakthrough patents (highest-impact patents) tend to occur early within a firm's patenting lifecycle. This finding is revealed by the ratio $N^*/N$ between the relative position of a firm' breakthrough patent, $N^*$, and the total number of patents issued to the firm, $N$; a low (high) value of this ratio indicates that the breakthrough tends to be among the early (late) patents by the firm.[14] The results on the EPO and USPTO data are shown in Fig. 4. We find that the probability that a firm is issued its breakthrough patent is indeed not constant over time. A firm's technological breakthrough is more likely to be among the firm's earliest patents, regardless of firm's group of impact, represented by an early peak of $P(N^*/N)$. This is true for both the EPO and USPTO datasets; the same is not true for reshuffled lifecycles constructed from the original data by randomizing the temporal order of each firm's patents (see the light lines in Fig. 4). These findings are consistent with the previous findings by Xu *et al.*.[14] In addition, Fig. 4 reports that the peak for high-impact firms is slightly higher than that for the medium and low-impact firms. While through statistical tests, we find only the difference between the high-impact and low-impact firms in the USPTO data is significant, which demonstrates that the timing pattern of top-impact firms' breakthrough patents is similar to that of ordinary firms' breakthroughs. So in the following, we will construct a unified model for all analyzed firms. In Table 3, we further provide specific impact values ($TI$) and relative positions of breakthrough patents ($N^*/N$) for a few high-impact firms that are involved in both the EPO and USPTO data. For some of them, such as MOTOROLA MOBILITY and INTERNATIONAL BUSINESS MACHS, the breakthrough patents in the two datasets exhibit similar temporal positions. While there also exist firms that demonstrate gaps in terms of the timing of their breakthrough patents in the EPO and USPTO data, such as EASTMAN KODAK and EXXON MOBIL. This difference confirms firms' heterogeneity in terms of patenting activities in different patent offices, under which the uniform early timing has more credibility.

The observed early peak of $P(N^*/N)$ can find explanations from previous management literature which claimed that young firms tend to present a higher probability of innovation than mature ones.[12,13] The underlying reason could be that as firms age, benefits from technological progress diminish,[13] and therefore, patents from old firms are more likely to be the extension and improvement of their established technologies,[39] which tend to hold relatively less impact compared with those from new entrants. The result proves that patent impact is correlated with the patent's temporal order among the applicant firm's lifecycle, and provides initial evidence in favor of a time-varying mechanism, i.e., firms will obtain varying returns from patents at different stages in their lifecycles. This is radically different from the random-impact phenomenon in scientific and artistic careers, for which the return is stable.

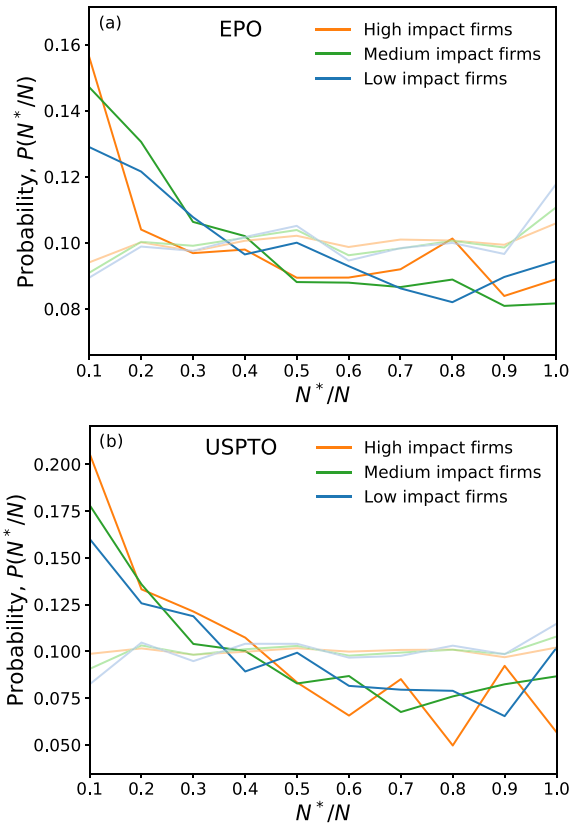



**Fig. 4.** Distribution of the temporal order of firms' highest-impact patents. We show the probability distribution of the relative order ($N^*/N$) of firms' technological breakthroughs, and the same distribution for randomized patenting lifecycles (denoted by the light lines) for high-, medium-, and low-impact firms, for the EPO and USPTO data, respectively. The results represent the average of 200 times statistics (to take the "paralleling" breakthroughs with the same impact value into consideration) and randomized firms' lifecycles. The early-stage peak of the distributions for all three groups of firms indicates that firms are more likely to achieve their highest-impact patent at the beginning of their patenting activity, contradicting the random-impact rule.

### 3.3. Dynamical model of firms' technological impact

To further understand and model the dynamics of firms' impact, we then analyze the average impact of patents issued to firms in different phases of lifecycles. In Fig. 5, we show patents' average impact as a function of their temporal order

**Table 3.** The technological impact and timing to achieve the breakthrough patents of several high-impact firms. Here we select firms that have records in both the EPO and USPTO data.

| | EPO | | USPTO | |
|---|---|---|---|---|
| Applicant firm | $TI$ | $N^*/N$ | $TI$ | $N^*/N$ |
| EASTMAN KODAK CO | 1.000 | 0.012 | 1.000 | 0.368 |
| GENENTECH INC | 1.000 | 0.232 | 1.000 | 0.045 |
| CANON INC | 1.000 | 0.166 | 1.000 | 0.078 |
| MOTOROLA MOBILITY INC | 1.000 | 0.032 | 1.000 | 0.090 |
| INTERNATIONAL BUSINESS MACHS CORP | 1.000 | 0.715 | 1.000 | 0.701 |
| FUJI PHOTO FILM CO LTD | 1.000 | 0.215 | 1.000 | 0.088 |
| EXXON MOBIL CORP | 1.000 | 0.026 | 1.000 | 0.393 |
| HITACHI LIMITED | 0.999 | 0.045 | 1.000 | 0.257 |
| TEXAS INSTRUMENTS INC | 0.999 | 0.550 | 1.000 | 0.222 |
| PFIZER INC | 0.999 | 0.250 | 1.000 | 0.459 |

within firms' lifecycles. We find an approximately linear decreasing trend, which indicates that firms tend to obtain diminishing returns from patents as they are issued more patents. Based on this finding, in the time-varying returns model shown in Eq. (2), we assume a linearly-decreasing function to model the phenomenon of the diminishing returns

$$\Phi\left(\frac{n_i}{N_\alpha}\right) = \beta - \gamma\frac{n_i}{N_\alpha}, \tag{3}$$

where $\beta$ and $\gamma$ are parameters to be estimated based on the empirical data. When $\beta = 1, \gamma = 0$ (no-varying scenario), the time-varying returns model reduces to the $Q$-model.

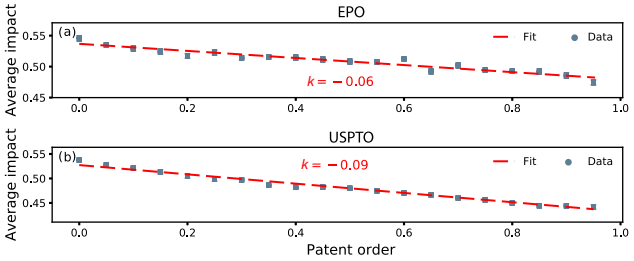



**Fig. 5.** Average impact of patents as a function of their temporal order within the applicant firms' lifecycle. We divide each firm's issued patents into 20 equally-sized groups according to the patents' issuance date. By aggregating all the analyzed firms, each dot represents the average impact of firms' patents that belong to the corresponding group. For example, the first dot from the left represents the average impact of firms' earliest 5% patents. Error bars refer to the standard errors of the mean. The results show that for both the two datasets, there is a linearly decreasing trend of patent impact as the patent order increases (via linear fit), which suggests a monotonously decreasing function for the time-varying returns model.

Assuming the proposed model, we firstly sum over the impact values of a firm's all patents, and empirically estimate each firm's $Q$ based on Eq. (2) through the approximate formula

$$Q_\alpha = \frac{I_\alpha}{\langle r\rangle \sum_{i=1}^{N_\alpha}(\beta - \gamma\frac{i}{N_\alpha})},$$
$$I_\alpha = \sum_{j\in\mathcal{P}_\alpha} NC_j, \tag{4}$$

where $I_\alpha$ is the total patent impact of firm $\alpha$, $\mathcal{P}_\alpha$ denotes the set of patents issued to firm $\alpha$. Subsequently, we generate each patent's simulated impact according to Eq. (2), and compare the simulated temporal pattern of firms' breakthrough patents (i.e., the probability distribution of $N^*/N$) with the corresponding pattern in the empirical data.

To calibrate the model parameters to the data, we apply a maximum likelihood approach. Specifically, we consider the Jensen Shannon divergence ($D_{\rm JS}$)[40] to measure the statistical "distance" between the observed temporal distribution, $P_{\rm data}(N^*/N)$, and the model-generated distribution, $P_{\rm model}(N^*/N)$, which is the symmetric version of the Kullback–Leibler divergence.[1,41] It is defined as

$$D_{\rm JS}(P_{\rm data}, P_{\rm model}) = \frac{1}{2}(D_{\rm KL}(P_{\rm data}, M)) + \frac{1}{2}(D_{\rm KL}(P_{\rm model}, M)),$$
$$D_{\rm KL}(P, Q) = \sum_x P(x)\log\left(\frac{P(x)}{Q(x)}\right), \tag{5}$$

where $M = (P_{\rm data} + P_{\rm model})/2$ is the mixed distribution. Then the likelihood function is defined as the fraction of model-generated distributions that are sufficiently close to the empirical one among 100 times simulations, i.e., the fraction of the simulated distributions such that $D_{\rm JS} < 0.002$ for the EPO data and $D_{\rm JS} < 0.005$ for the USPTO data, in which the thresholds are chosen based on the $D_{\rm JS}$ results in each dataset. The threshold is a little larger concerning the USPTO data than the EPO data considering the $D_{\rm JS}$ results are overall larger for the former. The value of the likelihood function ranges from 0 (no simulation fulfills the threshold) to 1 (all of the simulations fulfill the threshold). We choose the optimal parameter pair $(\hat{\beta}, \hat{\gamma})$ that maximizes the likelihood to generate a model whose temporal pattern is sufficiently close to the empirical one. The results are shown in Figs. 6(a) and 6(b), in which the color of each box refers to the value of the likelihood function. Figures 6(a) and 6(b) demonstrate that the optimal combination of parameters (denoted by the box with green contour) is $\hat{\beta} = 0.9$, $\hat{\gamma} = 0.06$ for the EPO data, and $\hat{\beta} = 0.9$, $\hat{\gamma} = 0.08$ for the USPTO data, which are similar to the slope of the fitted line in Fig. 5. We also observe that the $Q$ model corresponding to the box with pink contour achieves 0 in terms of the likelihood on both datasets, namely, none of the 100 times simulations fulfills the related thresholds. This reveals that to accurately model firms' impact dynamics, the role of the diminishing returns effect cannot be ignored.

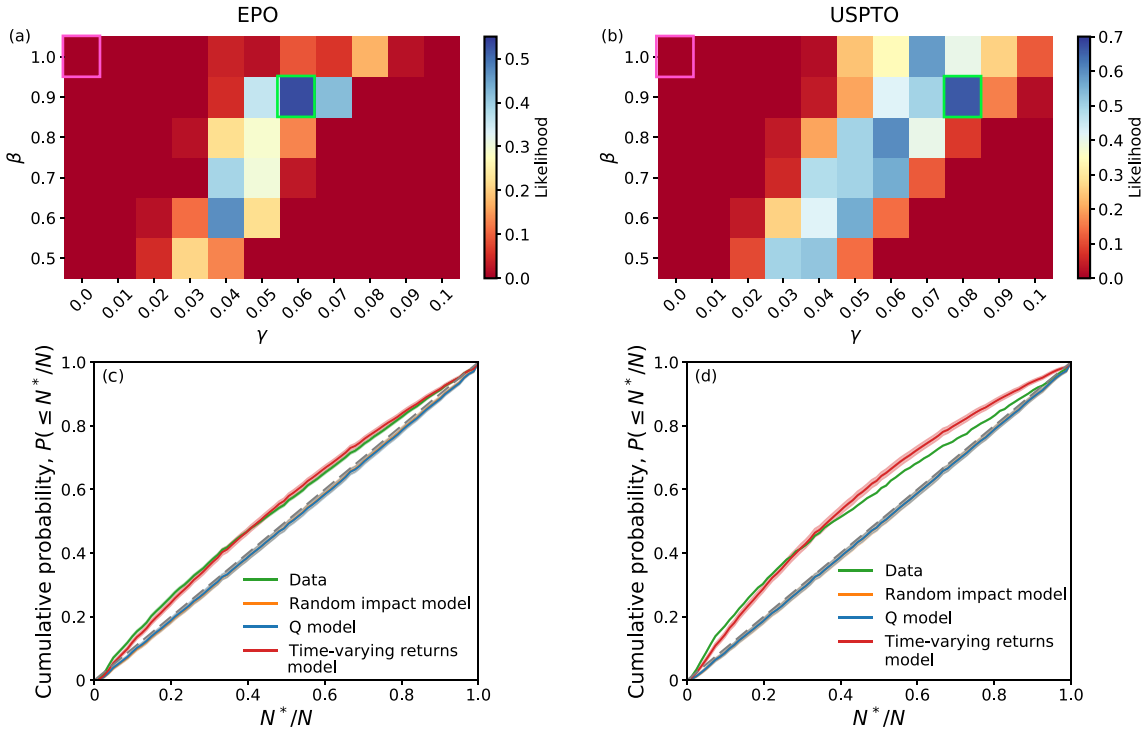



**Fig. 6.** Comparison of the empirical and model-based temporal patterns of firms' breakthrough patents. (a), (b) The likelihood that the timing distribution of breakthrough patents ($P(N^*/N)$) generated by the time-varying returns model is sufficiently close to the empirical one (see Subsection 3.3 for details) among 100 times simulations for different parameter combinations. The parameter pairs that lead to the maximum likelihood are $\hat{\beta} = 0.9$, $\hat{\gamma} = 0.06$ for the EPO data (a), and $\hat{\beta} = 0.9$, $\hat{\gamma} = 0.08$ for the USPTO data (b), which are corresponding to the boxes with green contour. The $Q$ model (corresponding to the box denoted by the pink contour) achieves the lowest likelihood. (c), (d) A comparison of the empirical and model-based distributions $P(\leq N^*/N)$ of the breakthrough patents' temporal positions of firms. We consider two state-of-art models and the proposed time-varying returns model, which uses the optimal parameter pairs as mentioned above. The results indicate that the empirical results in the two datasets are well reproduced by the time-varying returns model.

Figures 6(c) and 6(d) show the fitting performance of the proposed model with the optimal parameters against the random-impact model and the $Q$ model for the EPO and USPTO datasets. The result reveals that the proposed time-varying returns model can match the empirical curve much more accurately. While the random-impact model and $Q$ model assume that a patent's impact does not depend on its temporal position among its applicant firm's patents, thus fail to generate the concave timing distribution of breakthrough patents observed in the empirical data. The results further confirm the discrepancy of the timing pattern of the highest-impact works for firms and individuals (such as scientists and artists[1,6,8]). Besides, we also notice that although our proposed model outperforms the random-impact model and $Q$ model by a wide margin, it cannot fit the empirical results perfectly, especially for the USPTO data. This could be due to the heterogeneity in firms' impact dynamics and model parameters, which will be validated in future works.

## 4. Conclusion and discussion

In this work, we modeled the evolution mechanism of firms' technological impact within lifecycles. The USPTO and EPO patenting history data of applicant firms on a long timescale enable us to quantitatively trace and analyze the dynamic patterns of diverse firms worldwide. By treating each firm as a time-ordered sequence of its issued patents, we quantified the technological impact of their patents through a time-normalized citation measure, based on which we studied the dynamics of firms' patent productivity, technological impact, and the timing pattern of the highest-impact patents. There are three main takeaways from this work. Firstly, we provided evidence that firms tend to have their technological breakthroughs at the early stage within their lifecycles, which contradicts the random-impact rule that governing individual careers, and confirms previous findings.[14] Secondly, we found that the decrease of firms' average patent impact along with the patenting lifecycles is accurately matched by a linear function. Thirdly, most importantly, inspired by these findings, we proposed a time-varying returns model, which is proved to capture the temporal distributions of firms' technological breakthroughs more accurately than two state-of-art dynamical models.

A potential limitation of our approach is that our analysis assumes that all firms follow a uniform pattern. Yet different firms might have different trajectories, and the "average" behavior might not hold for all of them. Thus, our results could be not informative for specific types of firms. In future research, we will explore multiple patterns held by distinct types and industries of firms. Identifying drivers of the temporal position of firms' technological breakthroughs is also worth further research: The proposed model assumes that the time-varying function $\Phi$ is given, which leads to the impor-

tant problem of identifying its determinants at the firm level. Besides, incorporating our insights into real-world investment and policy decisions would need careful integration with a variety of signals beyond patent analysis. Addressing these questions will help us identify the best path for firm development and further catch a glimpse of the secret of business success.

To conclude, this paper enriches the understanding of firm impact dynamics. It provides the first step towards new models to describe the observed, firm-specific empirical patterns. The results can offer insights into technological innovation management for firms.

## Acknowledgment

We thank Luciano Pietronero, Emanuele Pugliese, and Andrea Zaccaria for providing us with the EPO patent data and enlightening discussions.

## Appendix A: Filtering firm applicants in the EPO data

As the research target in this paper is firm and not all applicants involved in the EPO data are firms, we remove some of the patent data that can be sure not assigned to firms. In detail, we identify four kinds of applicants based on their names and the information of their patents' inventors according to the following rules. (i) **Firm applicants**. If an applicant has the firm-related strings (refer to the first row in Table A1) in its name, this applicant is identified as a firm. (ii) **Individual applicants**. For each patent issued to a given applicant, if all its inventors also appear in the assignee field, and the applicant's name does not contain any firm-related string, this applicant is identified as an individual applicant.[42] (iii) **Educational applicants**. If the applicant does not have any firm-related string in its name but has the education-related strings (refer to the second row in Table A1), it is considered as an educational institution. (iv) **Other applicants**. If none of the above conditions is satisfied, the applicant belongs to this set. To exclude the patent data that is irrelevant to firms, we remove the patents issued to individual applicants and educational applicants, and leave the remaining data for the study. In total, we exclude 15.5% applicants and 4.3% patents, note that few of such applicants have larger than 20 patents, so the filtering has little influence on our analysis.

**Table A1.** The identification strings in applicants' names that we apply to filter applicants in the EPO data.

| Identification targets | Strings in applicants' names |
|---|---|
| Firm-related name | corporation, cooperativa, limitee, limitata, corp, co, ltd, inc, gmbh, kgaa, se, sro, ag, llc, llp, lp, kg, spa, firma, srl, sa, cie , bv, nv, sdn, holding, bhd, plc, as, ab, sl, mbh, sarl, sprl, snc, sas, aps, ltda, kk, group, sagl, asa |
| Education-related name | university, universitaet, universitat, universita, universidad, universite, univ, college, institute, institutes, institut, instituto, institutet, inst, research, recherche, recherches |

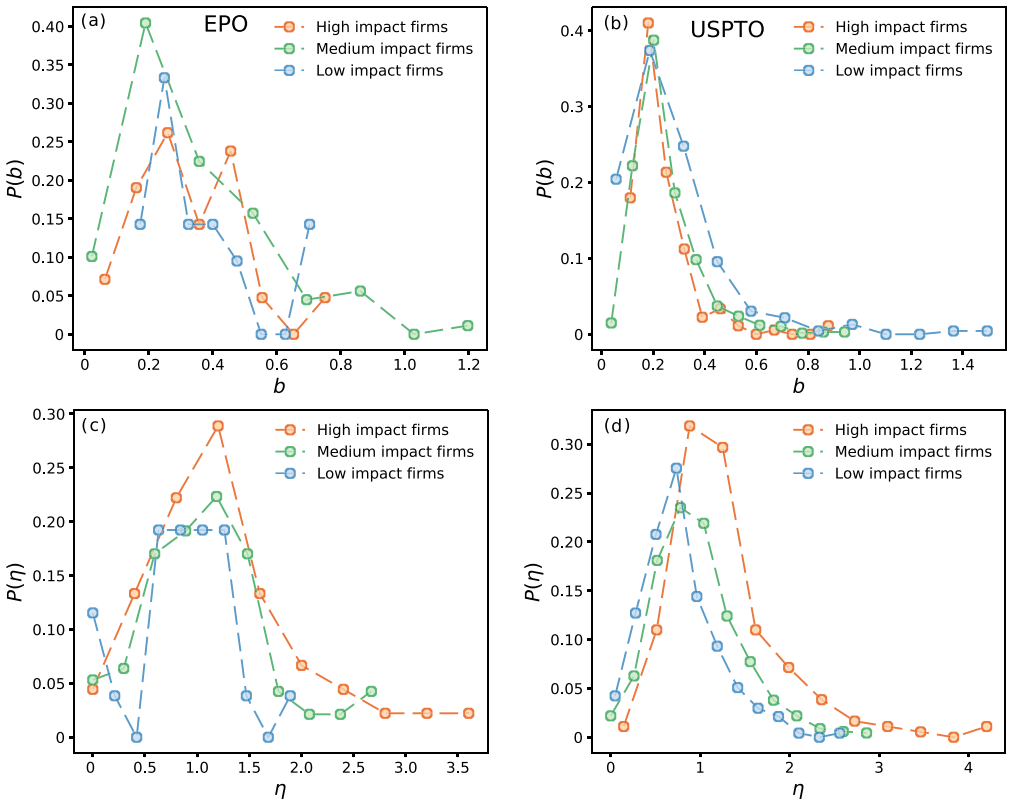



**Fig. B1.** Distribution of the parameter values ($b$ and $\eta$) in the fitting of productivity and impact dynamics (see Fig. 3) for firms in the (a), (c) EPO and (b), (d) USPTO data.

As for the USPTO data, in view of Kogan *et al.*[18,19] have made the extraction based on the Center for Research in Security Prices (CRSP) database, we do not implement any applicant filter on this data.

## Appendix B: Distribution of the fitting parameters in Fig. 3

Based on Fig. 3, we further show the distribution of the fitting parameters of the productivity growth and impact growth for the three groups of firms in Fig. B1. The results reveal that in terms of productivity, the development of high-impact firms ($\langle b_{\text{high}} \rangle = 0.33$ in EPO, $\langle b_{\text{high}} \rangle = 0.24$ in USPTO) is faster than that of low-impact firms ($\langle b_{\text{low}} \rangle = 0.36$ in EPO, $\langle b_{\text{low}} \rangle = 0.31$ in USPTO). In terms of technological impact, for low-impact firms, the growth is steady, with $\langle \eta_{\text{low}} \rangle = 0.88$ in EPO, and $\langle \eta_{\text{low}} \rangle = 0.82$ in USPTO. The increase of other firms is much faster, for whom the average $\eta$ is bigger, see Fig. B1.

## Appendix C: Robustness check of the main results

In the main text, we consider firms that have been issued at least a given number of patents. To guarantee the robustness of our main results, we further add analysis under a different filter criterion for firms in both datasets, i.e., the analyzed firms need to have long-lasting active patenting activities.[1,5,14] Specifically, for EPO data, we only consider firms that have at least 10-year patenting activity and have been issued to at least one patent every two years, which ends up with 1495 firms; for USPTO data, firms need to have at least 15-year patenting activity and have been issued to at least one patent every three years, involving 1007 firms. The main results under the new filter criterion are shown in Fig. C1, which demonstrate the robustness of our results.

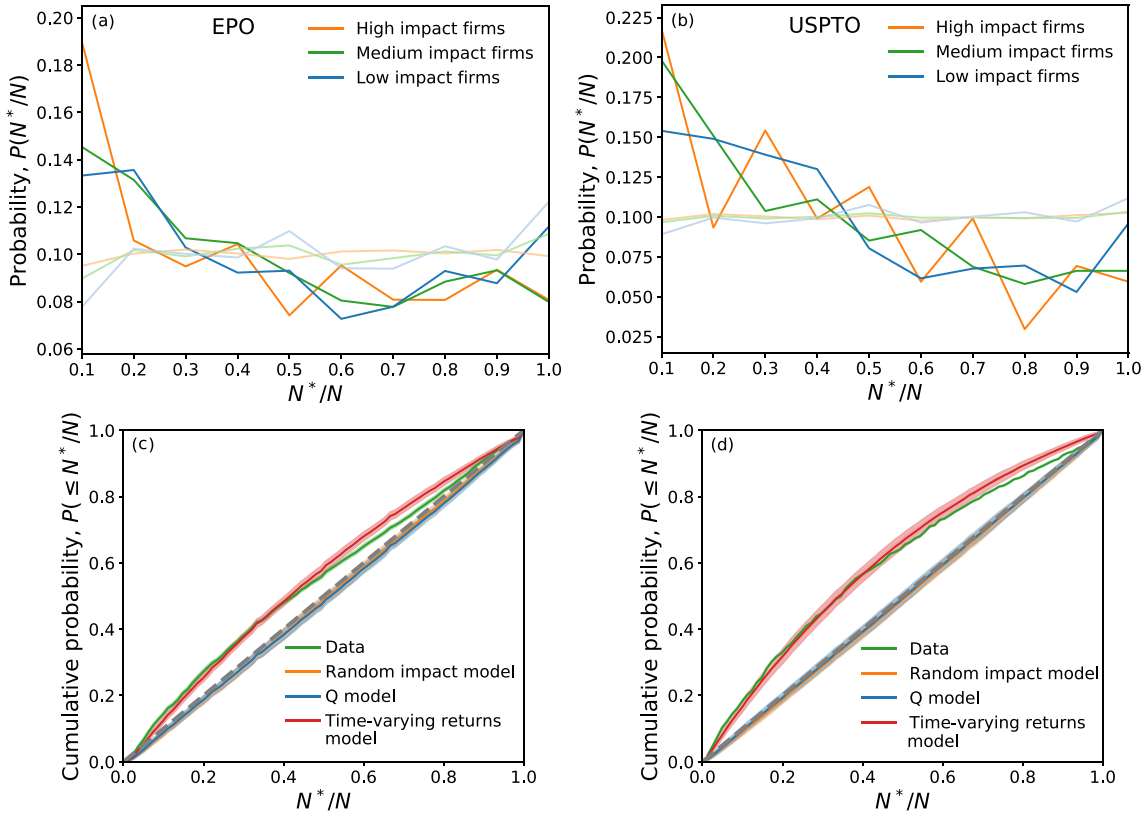



**Fig. C1.** Robustness check of the main results, in which we restrict firms' patenting activities to be long-lasting and active. The results are consistent with the conclusions in the main text. (a), (b) Distribution of the temporal order of firms' highest-impact patents, which is similar to Fig. 4. (c), (d) Comparison of the empirical and model-based temporal patterns of firms' breakthrough patents, which is similar to Fig. 6. The optimal parameters for the EPO data are $\hat{\beta} = 0.8$, $\hat{\gamma} = 0.06$, and for the USPTO data, they are $\hat{\beta} = 0.9$, $\hat{\gamma} = 0.08$.

## References


[1] Sinatra R, Wang D, Deville P, Song C and Barabási A L 2016 *Science* **354** aaf5239

[2] Bol T, De Vaan M and De Rijt A V 2018 *Proc. Natl. Acad. Sci. USA* **115** 4887

[3] Yucesoy B, Wang X, Huang J and Barabási A L 2018 *EPJ Data Sci.* **7** 7

[4] Williams O E, Lacasa L and Latora V 2019 *Nat. Commun.* **10** 2256

[5] Li W, Aste T, Caccioli F and Livan G 2019 *Nat. Commun.* **10** 1

[6] Janosov M, Battiston F and Sinatra R 2020 *EPJ Data Sci.* **9** 9

[7] Siudem G, Żogała-Siudem B, Cena A and Gagolewski M 2020 *Proc. Natl. Acad. Sci. USA* **117** 13896

[8] Liu L, Wang Y, Sinatra R, Giles C L, Song C and Wang D 2018 *Nature* **559** 396

[9] Yucesoy B and Barabási A L 2016 *EPJ Data Sci.* **5** 17

[10] Rossi M 2016 *Corp. Ownersh. Control.* **13** 217

[11] Coad A, Holm J R, Krafft J and Quatraro F 2018 *J. Evol. Econ.* **28** 1

[12] Huergo E and Jaumandreu J 2004 *Small Bus. Econ.* **22** 193

[13] Balasubramanian N and Lee J 2008 *Ind. Corp. Chang.* **17** 1019

[14] Xu S, Mariani M S, Lü L, Napolitano L, Pugliese E and Zaccaria A 2021 arXiv:2108.00200 [physics.soc–ph]
[15] Bloom N and Van Reenen J 2002 *Econ. J.* **112** 97
[16] Fallatah M I 2018 *J. Knowl. Manag.* **22** 678
[17] Pugliese E, Napolitano L, Zaccaria A and Pietronero L 2019 *PLoS One* **14** e0223403
[18] Kogan L, Papanikolaou D, Seru A and Stoffman N 2017 *Q. J. Econ.* **132** 665
[19] Stoffman N, Woeppel M and Yavuz M D 2020 *Kelley Sch. Bus. Res. Pap.* **19-5**
[20] Kim J and Lee S 2015 *Technol. Forecast. Soc. Chang.* **92** 332
[21] Strumsky D and Lobo J 2015 *Res. Policy* **44** 1445
[22] Acemoglu D, Akcigit U and Kerr W R 2016 *Proc. Natl. Acad. Sci. USA* **113** 11483
[23] Jaffe A B and De Rassenfosse G 2019 *Research Handbook on the Economics of Intellectual Property Law* (Cheltenham: Edward Elgar Publishing)
[24] Squicciarini M, Dernis H and Criscuolo C 2013 *OECD Science, Technology and Industry Working Papers* (Paris: OECD Publishing)
[25] Alcácer J, Gittelman M and Sampat B 2009 *Res. Policy* **38** 415
[26] Criscuolo P and Verspagen B 2008 *Res. Policy* **37** 1892
[27] Waltman L 2016 *J. Informetrics* **10** 365
[28] Mariani M S, Medo M and Zhang Y C 2015 *Sci. Reports* **5** 1
[29] Dunaiski M, Geldenhuys J and Visser W 2019 *J. Informetrics* **13** 270
[30] Xu S, Mariani M S, Lü L and Medo M 2020 *J. Informetrics* **14** 101005
[31] Barabási A L and Albert R 1999 *Science* **286** 509
[32] Valverde S, Solé R V, Bedau M A and Packard N 2007 *Phys. Rev. E* **76** 056118
[33] Higham K W, Governale M, Jaffe A and Zülicke U 2017 *Phys. Rev. E* **95** 042309
[34] Mariani M S, Medo M and Lafond F 2019 *Technol. Forecast. Soc. Chang.* **146** 644
[35] Ahuja G and Morris Lampert C 2001 *Strateg. Manag. J.* **22** 521
[36] Dunlap-Hinkler D, Kotabe M and Mudambi R 2010 *Strateg. Entrepreneurship J.* **4** 106
[37] Coad A, Segarra A and Teruel M 2016 *Res. Policy* **45** 387
[38] Petersen A M, Riccaboni M, Stanley H E and Pammolli F 2012 *Proc. Natl. Acad. Sci. USA* **109** 5213
[39] Sørensen J B and Stuart T E 2000 *Adm. Sci. Q.* **45** 81
[40] Lin J 1991 *IEEE Transactions on Inf. Theory* **37** 145
[41] Kullback S and Leibler R A 1951 *Ann. Math. Stat.* **22** 79
[42] He Z L, Tong T W, Zhang Y and He W 2018 *Sci. Data* **5** 180042