

·推动高质量发展·

中国专利质量的测度、变化机制 与经济增长效应*

陈志远 章逸然 张 杰 刘泽轩

内容提要:构建有效测算中国专利质量的方法体系,既是该领域的前沿问题之一,也是事关中国全面落实创新驱动发展战略的基础性工作。本文利用1985—2021年中国发明专利文本信息,运用自然语言处理方法,在计算专利间文本相似度这一核心指标的基础上,构建了适用于中国专利数据特征的专利质量测算方法。各项可靠性检验验证了本文设计和测算的专利质量指标相较于传统专利质量指标的优势。基于这一指标,本文从专利申请人类别、地域分布、技术分类和战略性新兴产业发展等角度梳理和归纳了1991—2016年中国专利质量水平的变化趋势与驱动机制,揭示了中国专利质量全面提升的特征事实。在此基础上,利用包含专利质量和数量变化的GDP增长分解框架,本文发现样本期间专利质量对经济增长的平均贡献度约为3%,且其对经济增长的作用相较专利数量呈逐步增强态势。本研究既可为有效评估各类微观主体的专利真实市场价值和自主创新能力提供基础性的识别工具,也可为中国专利制度改革及各类科技创新激励政策的绩效评估提供参考依据。

关键词:发明专利 专利质量测度 文本相似度 经济增长效应

一、引言

加快建设科技强国、实现高水平科技自立自强,已成为中国式现代化的关键支撑和战略任务。“十五五”规划建议指出,要“统筹国家战略科技力量建设”,“提升国家创新体系整体效能”,将知识产权保护和运用作为激发社会创新活力的重要抓手。专利被普遍视作体现国家自主创新能力的基础性指标。自2011年起,中国成为全球专利申请数量排名第一的国家;2019年,中国的PCT国际专利申请量首次超过美国,位居全球第一;自2022年开始,中国正式成为世界上拥有有效专利数量最

* 陈志远,中国人民大学商学院,邮政编码:100872,电子信箱:chenzhiyuan@rmbc.ruc.edu.cn;章逸然,复旦大学中国社会主义市场经济研究中心,邮政编码:200433,电子信箱:yiran_zhang@fudan.edu.cn;张杰(通讯作者),中国人民大学中国经济改革与发展研究院,邮政编码:100872,电子信箱:zhangjie0402@ruc.edu.cn;刘泽轩,南京审计大学联合研究院,邮政编码:211815,电子信箱:zxliu@nau.edu.cn。本文研究得到国家自然科学基金青年项目(72103192、72203046、72103098)、中国人民大学重大规划项目(23XNLG03)的资助。作者感谢匿名审稿专家的宝贵建议,特别感谢廖漠雨和张蒙博为文章给予的有益建议以及车明佳提供的助研工作,感谢中国人民大学校级计算平台的支持。当然,文责自负。

多的国家。尽管如此,大量研究发现中国可能在一定范围内存在专利数量增长“虚假”问题甚至专利“泡沫”现象(Hu et al., 2017; 张杰和郑文平, 2018)。尤其在政府创新补贴政策的激励驱动下,各类申请主体表现出策略性创新而非实质性创新的倾向(黎文靖和郑曼妮, 2016; 张杰等, 2022)。其中,最被担忧和诟病的问题是中国各类发明主体申请专利质量相对低下,市场价值不足,难以转化为企业或产业发展的有效生产力和市场竞争力。因此,从专利质量视角界定中国专利的真实价值内涵及其在企业与产业层面的市场价值,构建兼具现实适配性与科学逻辑性的中国专利质量测算框架,是一项具有学术价值与现实政策意义的基础研究议题。

近年来,已有文献开始关注中国专利数量快速增长的现象,并探讨其背后的专利质量测算问题,但系统梳理相关研究后不难发现,既有测算方法仍存在一定的局限性。在较大程度上,这些测算方法尚未充分利用专利文本中所包含的丰富的、系统性的专利技术特征信息。正如 Bergeaud et al. (2025) 指出的,创新研究领域一直面临难以科学界定和精确测算创新活动的难题,机器学习技术的专利文本分析方法开拓了创新领域的研究工作,正在成为创新领域研究的前沿探索方向。在充分借鉴这些研究的基础上,本文利用 Kelly et al. (2021) 中 TF-BIDF 方法计算了专利文本向量相似度指标,并根据不同技术领域专利文本特点改进和拓展了 KPST 方法的专利质量指标,构建了适用于中国专利制度情境和符合中文专利文本特点的专利质量指标。此外,本文使用大语言模型标注文本训练的 Chinese-RoBERTa-wwm-ext 模型进行了稳健性检验。在充分检验专利质量指标有效性的基础上,本文梳理和归纳了中国专利质量变化趋势和驱动机制的特征事实,提供了中国专利质量提升的技术进步效应的微观证据,并评估了专利数量扩张和质量提升的宏观经济增长效应。

与既有文献相比,本文具有如下方面的边际贡献。第一,本研究提供了跨发明主体和全口径的发明专利质量长面板分析,具有基础性研究价值。既有研究多以高校、重点企业或短期样本为对象,如陈强远等(2024)聚焦顶层设计语境下的高新技术企业,孙雅慧等(2024)考察上市公司 2010—2022 年期间专利摘要,尚缺乏对中国境内所有授权发明主体的涉及多种专利申请类型的长期专利质量动态描述。本文构建了 1991—2016 年期间全口径中国发明专利质量面板数据,填补了跨主体、长时间维度上的空白,是后续微观主体与宏观经济联动研究的基础性工作。第二,结合专利中文文本特征的专利质量测度方法拓展。本文在借鉴 KPST 方法基础上,针对中文专利中常见的术语歧义、分词挑战及技术承接路径,提出“全分类前向+同分类后向”相似度混合指标,并引入基于 IPC 技术分组的加权文本相似度加以校正。该方法既兼顾跨领域技术溢出,又精准捕捉本领域创新深度,与陈强远等(2020, 2024)侧重专利基因词汇分析、孙雅慧等(2024)侧重渐进式创新指标的做法形成互补,显著提升了中文专利质量度量指标的适用性与解释力。第三,系统性揭示了中国发明专利质量演进的六大特征事实与驱动机制。本文从专利申请人国别、类型、地区、技术分类和战略性新兴产业发展等不同侧面总结提炼了 1991—2016 年期间的专

利质量变化特征,并分析了背后的驱动机制。第四,提供了专利质量促进技术进步的微观证据及其对经济增长的边际贡献与政策含义。本文发现在样本期内,平均而言,中国各省专利质量水平提高对经济增长的贡献度约为3%,且贡献度呈现上升的基本趋势,凸显了中国创新由“量”到“质”的发展过程。第五,本文开发的中国专利质量测算指标是有助于中国构建自主创新体系和促进科技创新能力持续提高的基础性研究工作。本研究不仅能为中国专利质量水平的前沿测算方法以及相应变化机制提供基础性的学术成果,也可为评估各类微观主体层面的专利真实市场价值和自主创新能力提供基础性的识别工具,更将为中国专利制度改革以及各类科技创新政府激励政策绩效评估提供有效的分析判断工具。

本文其余部分安排如下:第二部分系统梳理现有文献中专利质量的测度方法,并据此提出本文的专利质量测度策略;第三部分详细阐述中国专利质量的具体测算步骤,并开展指标有效性验证工作;第四部分深入分析中国专利质量的变化趋势与核心驱动机制;第五部分实证检验专利质量提升的技术进步效应,并测算其对经济增长的贡献度;第六部分为结论与政策建议。

二、现有专利质量测度方法比较与本文测度策略

(一)现有专利质量测度方法述评

既有文献中代表性的专利质量测度方法在数据要求、适用范围和应用效果上各具特色且存在显著局限。第一,基于专利引用数据的方法以专利前向被引数目为核心指标,其理论基础在于被引次数反映技术重要性及未来影响力(Trajtenberg, 1990; Hall et al., 2005),在以欧美国家为对象的研究中被广泛用于衡量知识转移与专利价值(Artz et al., 2010; Marco, 2007),然而该方法应用于中国情境时面临多重挑战:国家知识产权局不强制要求引用披露导致数据系统性缺失(Dang et al., 2015),发明人策略性隐瞒已知文献的逆向选择行为使得引用无法反映真实知识流向(Thompson, 2006),加之引用数量的整数特性造成质量测度不连续、专利授权时滞引发的数据截断问题,以及中国专利制度多次调整对引用行为的结构冲击,使得该指标难以进行跨期比较并可能严重失真(Lerner & Seru, 2017; 陈强远等, 2024)。第二,基于专利经济效应的方法主要分为两类:一是通过股价反映测度创新的事先价值(Pakes, 1985; Kogan et al., 2017),二是基于维护成本与收益权衡推断专利价值分布(Pakes, 1986; 朱雪忠等, 2009)。但前者仅适用于上市公司且要求符合有效资本市场假设,在发展中国家覆盖面极为有限,后者则因需要观察专利完整生命周期而存在数据截断偏误,难以应用于近期授权专利。第三,基于专利权范围的方法以知识宽度代理专利质量,常用主权项技术分类数量(Lerner, 1994)或文本字数、名词数目(Dang & Motohashi, 2015)测度,其理论逻辑在于保护范围越广则技术复杂度越高、模仿成本越高、垄断利润越丰厚。然而该指标具有显著的“欺骗性”,申请者可能通过宽泛权利要求和模糊措辞进行专利寻租,利用后续创新者的“侵权”风险索取使用费(Schwartz & Kesan, 2013),而中国知识产权保护制度不完善与政府补贴政

策的叠加效应进一步强化了企业的策略性申请动机,导致该指标难以真实反映创新的新颖程度。第四,基于专利技术词汇的方法通过识别专利文本中的新词测度创新程度(Balsmeier et al., 2018),陈强远等(2020, 2024)进一步借鉴基因编码逻辑构建包含时间位序新颖度和未来影响力频次的企业层面创新指标,但此类方法聚焦专利名称而非技术细节,难以完整刻画专利的真实技术特征,且缺乏对不同发明主体的系统性比较。第五,基于文本相似度的方法利用自然语言处理技术,通过计算专利文本与前后专利的相似度识别兼具创新性和影响力的专利(Kelly et al., 2021; 孙雅慧等, 2024)。其中,运用KPST方法对美国1840—2010年专利质量的测度显示,基于文本分析的指标相较于传统引用指标,能更好地捕捉分子生物学和基因技术等领域的重大技术突破,为识别突破性创新提供了新的方法路径。

现有研究表明,不同专利质量指标仅刻画质量的某一侧面,不存在“一刀切”的度量(Higham et al., 2021)。结合中国数据环境,本文将文本相似度指标定位为高覆盖、强可复现且与“知识距离—扩散”机制直接对应的核心度量,并遵循“数据可得性—理论一致性—方法创新”三重动机构建指标。在数据层面,中国专利引文由审查员补充导致“非内生性”,而文本数据覆盖广、可复现性强;在理论层面,文本相似度直接刻画知识距离(Jaffe, 1986),契合KPST方法的“新颖性—影响力”双维度框架。具体地,本文以同类先验技术的后向相似度刻画新颖度(相似度越低越新颖),以全域后续专利的前向相似度刻画影响力,以两者之比构造质量指数,并在技术领域—年份层面进行标准化。

为适配中文文本结构,本文在KPST方法框架上做出四项改进:一是对前向相似度采用所有技术分类、后向相似度采用相同IPC二分位技术分类的混合策略;二是剔除同一申请人产生的专利,避免申请人与自己历史产生专利的“自我相似”低估专利的新颖性以及未来产生专利的相似高估专利的影响力;三是引入IPC技术领域加权校正,修正部分领域词汇共现度高的偏误;四是对权利要求、摘要与说明书核心段落进行分词与术语化预处理,基于TF-BIDF生成表示,在技术领域—年份尺度做标准化与稳健化处理。这些改进使本方法成为面向中文语境的系统化拓展。

(二) 本文构建的中国专利质量测度方法

本文的专利质量测度方法主要包括两部分:专利文本相似度计算与专利质量指标构建。基于中国专利文本数据信息特征,本文拓展了KPST方法并设计适用于中国专利文本特征的专利质量测度方法。^①

1. 专利文本相似度计算

借鉴KPST方法的思路,本文采用词频—反向逆文档频率(TF-BIDF)构建专利文本向量。其定义为:

$$TFBIDF_{pw} = TF_{pw} \times BIDF_{pw} \quad (1)$$

^① 因篇幅所限,相关内容详见本刊网站登载的附录。

其中, $TF_{pw} = \frac{n_{pw}}{\sum_k n_{pk}}$ 表示词 w 在专利 p 的文本中出现的频率。 $BIDF_{pw} = \log\left(\frac{p\text{-之前专利数量}}{1 + p\text{-之前专利中含词}w\text{专利数量}}\right)$ 为反向逆文档频率。以 $TFBIDF_{pw}$ 构建专利文本词汇的高维矩阵 T 。矩阵 T 的第 p 行向量 $TFBIDF_p$ 代表专利 p 的词向量,其长度等于所有专利文本词汇并集的长度。对于该专利中存在的词汇 w ,其对应向量元素取值为 $TFBIDF_{pw}$,而该专利中不存在的词汇,其对应向量元素取值为零。最后,将词汇向量标准化为 $U_p = TFBIDF_p / \|TFBIDF_p\|$ 。则任意两个专利对 (p, q) 之间的相似度为两者的余弦相似度:

$$s_{p,q} = U_p \cdot U_q \quad (2)$$

其取值范围为 $[0, 1]$,取值越大代表专利 p 和专利 q 相似度越高。

2. 专利质量指标构建

基于专利相似度,定义新颖度(后向相似度)和影响力(前向相似度):

$$BS_p(\tau) = \sum_{q \in \mathcal{B}_p(\tau)} s_{p,q}, FS_p(\tau) = \sum_{q \in \mathcal{F}_p(\tau)} s_{p,q} \quad (3)$$

其中, $\mathcal{B}_p(\tau)$ 和 $\mathcal{F}_p(\tau)$ 分别为专利 p 申请时前后 τ 年内的专利申请集合。KPST 方法将专利质量定义为 $Q_p^{KPST}(\tau) = FS_p(\tau) / BS_p(\tau)$ 。该公式应用于中国存在一定局限,因此本文进行了四个方面的改进:第一,为了避免专利数量快速膨胀对专利文本相似度造成的影响,本文使用专利数量对专利的相似度进行标准化处理,得到专利的平均前向相似度和后向相似度,继而将专利质量定义为专利的平均前向相似度和平均后向相似度的比值。该步骤剔除了专利数量急剧增长带来的专利质量虚假膨胀因素,能够更加准确地刻画专利质量随时间的变化情况。第二,为了避免同一申请人持续累积创新对专利质量测算造成的影响,本文在计算前向和后向相似度时剔除了同一申请人的专利申请。第三,为了考虑专利之间跨技术领域比较和相同技术领域内比较的不同含义,本文分别考察了限定跨专利技术分类和同技术分类下的专利前向和后向相似度。第四,考虑到专利技术领域的不同,尤其是一些词汇概念相似度较高的技术领域的文本特征,本文使用专利文本的相对后向相似度进行标准化调整,以考虑不同技术领域专利文本的内在特征差异性。

具体而言,定义平均后向相似度和平均前向相似度分别为:

$$\overline{BS}_p(\tau) = \sum_{q \in \mathcal{B}_p(\tau) \setminus C_p(\tau)} \frac{s_{p,q}}{\sum_{k=t_p-\tau}^{t_p-1} N_k}, \overline{FS}_p(\tau) = \sum_{q \in \mathcal{F}_p(\tau) \setminus D_p(\tau)} \frac{s_{p,q}}{\sum_{k=t_p+1}^{t_p+\tau} N_k} \quad (4)$$

其中, $C_p(\tau)$ 和 $D_p(\tau)$ 表示根据相同申请人或 IPC 技术分类剔除的专利集合, N_k 表示获得授权的专利样本中 k 年申请的发明专利数量。引入不同技术领域调整系数:

$$\xi_i(T) = \frac{N_i}{N_i(T)} \frac{\sum_{p \in T(t)} \overline{BS}_p(\tau)}{\sum_{T(t) \in T} \sum_{q \in T(t)} \overline{BS}_q(\tau)} \quad (5)$$

其中, $T(t)$ 表示在 t 年申请的属于专利技术分类 T 的集合, $N_i(T)$ 表示技术分类 T 下 t 年申请的专利总数, T 表示所有专利技术领域划分的集合。在该调整系数的作用下, 技术概念文本内在相似度较高的技术领域(如化工、冶金)会自然具有较大的相对文本相似度。

本文最终构造的专利指数计算公式为:

$$Q_{p \in T}(\tau) = \xi_i(T) \times \frac{\overline{FS}_p(\tau)}{\overline{BS}_p(\tau)} \quad (6)$$

基准结果中, 本文设定 $\tau = 5$, 以匹配中国专利的短生命周期特征。

三、中国专利质量水平的测算及可靠性验证

(一) 数据处理与专利质量指标计算

本文使用的是国家知识产权局 1985—2021 年中国境内授权的发明专利数据, 包含申请号、授权号、年份、申请人类别及 IPC 分类等信息。从样本数据统计来看, 发明专利数量在 2000 年《中华人民共和国专利法》修订后呈爆炸式增长, 因此需剔除数量因素偏误。由于发明专利一般需要经过较长时间的实质审查, 数据截断导致 2016 年之后数据中被授权的发明专利申请量下降。

本文分析的专利文本信息主要包含专利的标题、摘要和主权项要求文本。为了便于运用独热法对专利文本词汇进行编码, 我们遵循如下操作步骤。首先, 专利文本编码与格式统一。该步骤主要包括去除文本中含有的数字和特殊字符, 并对文本进行分词处理。分词处理的主要目的包括两个方面: 一方面, 帮助降低词向量的维度, 从而减少计算所需的计算机内存并缩短计算时间; 另一方面, 去除无实际意义词汇对文本分析的干扰。本文使用了中文文本分析工具 Jieba 分词对文本进行分词操作, 将连续的文本序列切分成有意义的词语。分词处理要求去除专利文本中的停用词和低频词。本文使用的停用词表包含中文分析的常规停用词和作者根据阅读专利文本挑选的无实际意义的高频词。通过对词频分布的观察, 本文对词频低于 20 的词进行了剔除。^①其次, 考虑到发明专利授权时间相较于专利申请时间存在滞后性, 而专利公布的时间一般早于专利授权的时间, 因此本研究基于专利的申请年份对专利进行分类。在基准研究中, 对每个专利计算其与之后 5 年所有专利的相似度, 比如, 对 2000 年申请的专利, 我们计算其与 2001—2005 年间各年申请专利文本之间的余弦相似度。遍历所有专利文本数据便可得到数据中 1991—2016 年间任意专利文本与其前后 5 年专利文本的相似度。最后, 为减少计算的内存消耗, 提高计算效率, 本文针对性地设计了相似度的计算方法, 计算得到 1991—2016

^① 在本研究中, 经敏感性测试, 降低低频词的门槛值不会对计算结果产生实质影响。

年共计 351 万余条发明专利的质量指标。^①

(二) 专利质量指标构建策略与结果比较

基于前述方法,我们比较了三类计算策略:(1)全 IPC 分类相似度;(2)同 IPC 分类相似度;(3)前向全 IPC+后向同 IPC 组合。每类均对照是否剔除同一申请人申请的专利,并且所有类别均计算了是否采用技术领域调整系数的两种结果。^②实证发现:剔除同一申请人对结果影响较小,限定同 IPC 部类影响较大;前向相似度采用全 IPC 分类更能捕捉跨领域影响力,后向相似度限定同 IPC 分类可避免跨领域噪声,精准测度本领域知识承接。经检验,方法(3)在实证中综合表现最优,具有现实合理性。本文提出的技术领域调整系数有效修正了化学、冶金等概念组合创新领域的质量低估问题。^③综合考虑,基准指标采用前向全 IPC、后向同 IPC 且剔除同一申请人并经领域调整的方法。

(三) 专利相似度与质量指标分布特征

后向相似度和前向相似度指标均呈对数正态分布。同 IPC 分类下的后向相似度高于全 IPC 分类下的平均相似度,体现同领域文本更加相似的特征。专利平均后向相似度和前向相似度的线性拟合系数为 0.064(在 1% 的水平上显著),表明领域内技术承接性与跨领域影响力可从不同侧面反映专利质量。不同 IPC 部类的相似度分布存在显著差异,印证了技术领域异质性。^④

(四) 稳健性检验:调整中文预训练模型

专利文本具有法律性和高度技术性特征,术语含义依赖上下文内容。基准 TF-BIDF 方法侧重技术关键词,难以充分捕捉语境信息。为检验指标稳健性,本文采用 Chinese-RoBERTa-wwm-ext 预训练模型计算文本相似度。该模型通过引入完全屏蔽词汇(Whole Word Masking, WWM)策略,对中文词汇的全部组成字进行完整屏蔽,而非仅屏蔽单个字符。通过这种方式,模型能够更深入地理解中文词汇作为一个整体的语义内涵,有效处理多义性和上下文依赖问题,在中文语义相似度任务中表现优异。本文结合人工标注与大语言模型标注的专利样本集,对 Chinese-RoBERTa-wwm-ext 模型进行了微调。^⑤TF-BIDF 与 Chinese-RoBERTa-wwm-ext 方法计算的专利质量对数相关系数为 0.813,且在 1% 的水平上显著,表明本文指标对更加前沿的自然语言处理方法具有高度稳健性。^⑥

(五) 专利质量指标的可靠性检验

1. 与已有指标的分布比较

本文将所构建的专利质量指标与文献中已有的专利质量指标进行了对比,包括专利被引用数、专利主权项要求数和专利主权项要求字数。^⑦传统指标在分布特征上存在明显缺陷:专利被引次数呈现显著的零膨胀与长尾分布,约 51% 的专利被引次数为零;主权项要求数高度离散并在整数处堆积;主权项字数虽近似钟形,但受撰写规范影响,变异有限。这些离散化与堆积特征导致排序粒度不足,易引发两类识别错误:第一,专利文书篇幅较长但技术含量有限的专利可能被误判为高质量

^{①②③④⑤⑥⑦} 因篇幅所限,相关内容详见本刊网站登载的附录。

(Type I 错误);第二,早期尚未获得引用或策略性压缩主权项数的真正高质量专利可能被低估(Type II 错误)。

相较而言,本文基于文本相似度构建的质量指标在不同IPC门类下均呈连续、单峰且左右展开的分布,跨领域形态一致性较好,能够提供更细粒度的排序信息。分位数比较分析进一步揭示:首先,在0—25分位的低质量区间,专利被引次数与主权项数的曲线近乎贴近横轴,显示出明显的变异性不足;而主权项字数曲线整体位于45度线之上,意味着相对于本文指标,其在低质量样本上给出了偏高的分位位置,存在“高估低质量”倾向的Type I 错误。其次,在25—75分位与75—99分位的中高质量区间,三类传统指标普遍呈现斜率偏小或上升受限的特征,即相对于本文指标存在分位提升迟滞、区分度不足现象,指向对高质量专利系统性低估的Type II 错误。这一模式与计数型指标受零膨胀与饱和上限约束、长度型指标受撰写规范约束的机制一致。在文本构建的专利质量指标分布两端,三种传统专利质量指标走势分化明显,相关性与一致程度较弱;而在中间区间则相对一致并保持平滑变化。具体来讲,在低质量区间,传统计数型指标变异性不足,而长度型指标存在高估倾向;在中高质量区间,传统指标普遍呈现区分度不足的特征。综上,本文基于文本相似度的连续型专利质量指标在跨技术领域保持稳定的分布形态与更高的分位区分度,能够在分布两端同时降低误判风险,更适用于对专利质量进行全分布刻画与排序。

2. 对于高、低质量专利的有效识别检验

(1) 基于专利存续时间的指标有效性检验

本部分以专利实际存续时间作为外部结果变量,检验文本相似度专利质量指标的有效性。中国专利制度设定了逐年递增的维持费用,使得专利存续时间能够反映权利人对专利经济价值的理性判断,是衡量专利内在质量的稳定代理变量。本文构建如下专利层面回归模型:

$$patlife_p = \alpha + \beta patquality_p + otherq_p \gamma' + assignee_j + appyear_p + \varepsilon_p \quad (7)$$

其中, $patlife_p$ 表示专利 p 在获得授权后专利生存时长的对数,以考虑专利寿命分布尾部的强右偏特征。 $patquality_p$ 表示本文构建的专利质量指标。 $otherq_p$ 包括主权项数、主权项字数和被引次数三类传统指标。^① $assignee_j$ 为专利授权人的固定效应, $appyear_p$ 表示专利 p 申请年份的固定效应, ε_p 为随机扰动项。如果 $patquality_p$ 的回归系数 β 显著为正,则说明质量越高的专利,其存续时间越长。

表1报告了全样本回归结果。本文测算的专利质量指标在逐步控制主权项数、主权项字数和被引次数后,系数始终稳定在0.1882—0.1945之间($p < 0.01$),表明其对专利寿命具有独立于传统指标的增量解释力。相比之下,主权项字数的系数在加入被引次数后由0.7993降至0.5676,显示传统指标间存在信息重叠。这一结果支持了文本指标作为专利质量度量的合理性。

^① 因篇幅所限,相关内容详见本刊网站登载的附录。

表 1 专利存续时长与各类专利质量指标相关性

变量	专利存续时长 (1)	专利存续时长 (2)	专利存续时长 (3)	专利存续时长 (4)
专利质量指标	0.1934*** (0.0089)	0.1945*** (0.0089)	0.1882*** (0.0089)	0.1930*** (0.0089)
主权项数		2.1197*** (0.0634)	2.1721*** (0.0635)	1.9586*** (0.0635)
主权项字数			0.7993*** (0.0473)	0.5676*** (0.0474)
被引次数				1.0653*** (0.0165)
观测值	3491515	3491514	3491487	3491487
Adj. R ²	0.6465	0.6467	0.6467	0.6471

注：所有回归中均包含专利授权人与申请年份固定效应，括号中为稳健标准误。***、**和*分别表示在1%、5%和10%的水平上显著。下同。

为检验文本指标在传统指标区分度不足区间的表现，本文进一步进行分组回归。^①在0—25分位低质量区间，文本指标系数显著为正，表明其能够识别被传统指标误判的优质专利，纠正 Type II 错误。在75—99分位高质量区间，文本指标仍具有增量区分力，尤其在主权项字数较长样本中，文本指标显著为正而主权项字数指标显著为负，有效抑制了将冗长专利误判为高质量的 Type I 错误。综上，文本指标在全分布范围内表现出稳健的预测力，能够同时降低两类识别错误，可作为中国情境下更为有效的专利质量度量指标。

(2) 专利申请“年末突击”的质量效应

本文利用中国“创新锦标赛式”考核背景作为准自然实验，进一步检验文本质量指标的外部有效性。在年度绩效考核约束下，各类创新主体在11月和12月呈现“年终冲刺式”专利申请行为，这一现象往往伴随专利质量的系统性下滑(陈强远等, 2024; Chen et al., 2025)。本文构造如下计量模型：

$$patquality_p = \delta patrush_p + apptype_p + assignee_j + \mu_{ct} + \varepsilon_p \quad (8)$$

其中， $patquality_p$ 为专利质量指标的对数。 $patrush_p$ 为年终冲刺哑变量，当专利申请月份为11月或12月时取值为1，其他月份取值为0。控制申请人类型固定效应 $apptype_p$ (个人、企业或科研机构)、授权人固定效应 $assignee_j$ 和城市一年份固定效应 μ_{ct} 。 ε_p 为随机扰动项。考虑同一城市内部发明专利申请决策的关联性，标准误在城市层面聚类。

表2报告了估计结果。列(1)显示年终冲刺的估计系数为-0.0072 ($p < 0.05$)，表明本文专利质量指标准确识别了制度性行为导致的质量下滑。相比之下，列(2)主权项字数和列(3)主权项数的系数分别为0.0043和0.0107，均具有统计意义上的显著性，表明冲刺期专利申请倾向于写得更长、列得更多，但并未转化为更高质量——这印证

^① 因篇幅所限，相关内容详见本刊网站登载的附录。

了专利文本长度型指标易产生 Type I 错误的判断。列(4)显示冲刺期申请专利被引次数显著更低,但在被引次数为零的样本中,列(5)专利年度冲刺对专利质量指标的回归系数为-0.0079($p < 0.01$),与全样本高度一致,显示其对传统指标无法识别的低质量专利仍具有甄别能力。上述结果表明,文本指标能够对制度性“负质量冲击”给出方向正确、幅度稳定的响应,在创新绩效考核评估、专利筛选预警等政策应用中具有更强的外部效度。

表2 专利申请“年末突击”的专利质量效应

变量	全样本				专利零被引样本
	专利质量 (1)	主权项字数 (2)	主权项数 (3)	被引次数 (4)	专利质量 (5)
专利年终冲刺	-0.0072** (0.0028)	0.0043* (0.0026)	0.0107*** (0.0026)	-0.0582*** (0.0024)	-0.0079*** (0.0030)
申请人类别固定效应	是	是	是	是	是
授权人固定效应	是	是	是	是	是
城市—年份固定效应	是	是	是	是	是
观测值	1984897	1984885	1984898	1984898	968765
Adj. R ²	0.4018	0.2515	0.4243	0.2507	0.3914

注:括号中为城市层面聚类的标准误。

3. 专利质量指标与企业各类经营表现

为验证专利质量指标的经济有效性,本文将专利数据与1998—2007年的中国工业企业数据匹配,考察专利质量与企业经营表现的关系。^①构建如下计量方程:

$$Y_{it} = \alpha Y_{it-1} + \beta patquality_{i(t-1)} + X_{it} \gamma' + \lambda_i + \lambda_t + \varepsilon_{it} \quad (9)$$

其中, Y_{it} 表示企业*i*在年份*t*的经营指标变量,包括企业对数形式的销售额、增加值、固定资产和员工人数。 $patquality_{i(t-1)}$ 为企业上期申请发明专利的平均质量。控制变量 X_{it} 主要包含发明专利数量、是否有发明专利申请。此外,控制企业固定效应 λ_i 和年份固定效应 λ_t , ε_{it} 为随机扰动项。此处为说明专利指标的解释预测能力而非追求关于因果关系的无偏估计,因此使用OLS估计量。

表3列(1)—(4)显示,企业平均专利质量对各类经营指标均有显著正向影响。将质量高于全样本95分位数的专利定义为“高质量专利”,质量低于5分位数的专利定义为“低质量专利”。^②列(5)—(8)结果表明:高质量专利数量与经营指标呈显著正相关,而低质量专利数量的系数均不显著,验证了本文指标的区分效度。本文进一步尝试同时纳入多种质量指标进行比较,^③发现本文指标在增加值和固定资

① 本文选用了争议最小也是使用最为广泛的1998—2007年中国工业企业数据库样本与专利数据进行匹配,并参考He et al.(2018)与寇宗来和刘学悦(2020)的做法,确保匹配效率。本文最终的估计样本包含1998—2007年工业企业申请并获得授权的发明专利95843件,占He et al.(2018)公布的工业企业合并发明专利申请总数的61%,体现出较高的一致性。

② 如无特殊说明,全文余下部分高、低质量专利划分均按照此标准。

③ 因篇幅所限,相关内容详见本刊网站登载的附录。

产模型中显著为正,专利存续时间表现稳定,而被引次数的系数为负且部分显著,主权项数和主权项字数的解释力有限。高质量专利数量在所有模型中均显著为正,低质量专利则不显著,说明本文质量指标对企业层面各类经营表现具有良好的解释力,优于被引次数等传统指标,展现出独立的信息增益与应用价值。

表3 专利质量指标与企业各类经营表现

变量	平均专利质量				区分高、低专利质量			
	销售额 (1)	增加值 (2)	固定资产 (3)	员工数 (4)	销售额 (5)	增加值 (6)	固定资产 (7)	员工数 (8)
平均专利质量	0.115*** (0.019)	0.181*** (0.031)	0.148*** (0.023)	0.093*** (0.015)				
高质量专利数					0.001*** (0.000)	0.002** (0.001)	0.001* (0.001)	0.001*** (0.001)
低质量专利数					0.000 (0.001)	-0.001 (0.003)	-0.001 (0.002)	0.001 (0.002)
总专利数	0.044*** (0.006)	0.081*** (0.009)	0.073*** (0.007)	0.036*** (0.005)	0.055*** (0.005)	0.099*** (0.009)	0.089*** (0.006)	0.044*** (0.005)
观测值	1029725	983983	1029725	1029725	1029725	983983	1029725	1029725
Adj. R ²	0.901	0.784	0.898	0.912	0.901	0.784	0.898	0.912

注:所有模型均控制了企业发明专利数量、是否有发明专利申请以及企业和年份固定效应,括号中为企业层面聚类的标准误。

四、中国专利质量变化趋势与驱动机制的特征事实

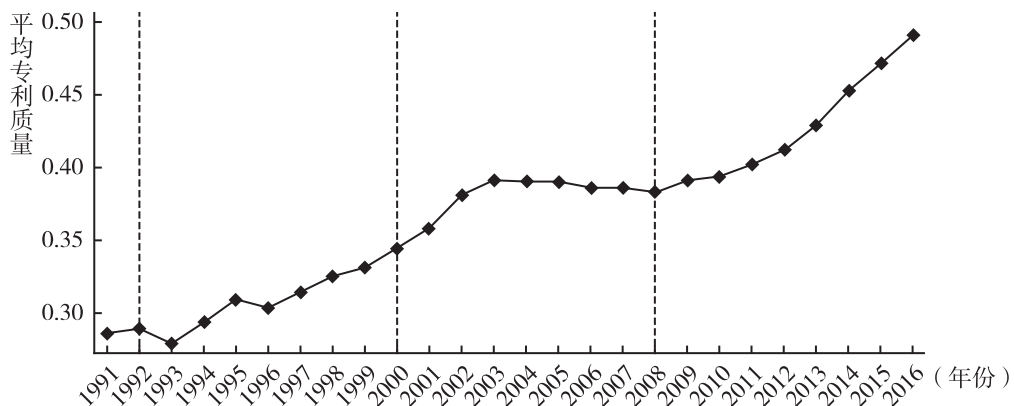
(一)整体变化趋势分析

传统专利质量指标(如被引次数、权利要求项数和字数)在跨时间比较中存在系统性偏误,如被引次数受“暴露期”和“引文通胀”影响,权利要求项数受专利审查收费政策和撰写规范影响,权利要求字数更多反映文件规范化趋势。这些外生因素导致传统指标呈现“时间漂移”的伪趋势。相比之下,本文指标通过控制专利类别、技术领域及时间窗口,在较大程度上规避了时序偏误,更可能反映专利质量的真实动态。需要明确的是,本文构建的专利质量指标至多只能反映中国技术进步的一个侧面,对于特征事实的一般性意义需要做谨慎的解读。

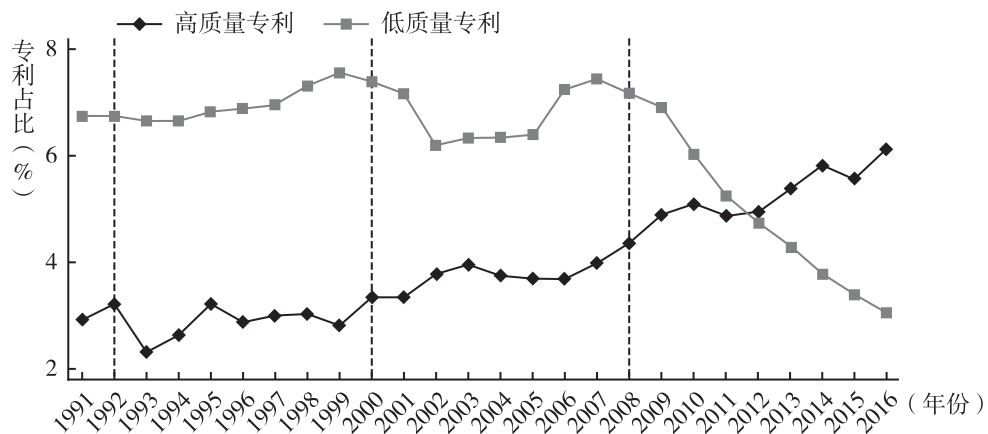
定义年份 t 申请的专利平均质量为 $\bar{Q}_t = 1/N_t \sum_{p \in I_t} Q_p$, 其中 N_t 表示 t 年专利数量, I_t 为该年申请并获得授权的发明专利集合, Q_p 表示专利 p 的专利质量。图 1(1) 显示, 1991—2016 年中国发明专利平均质量呈现 W 形的上升态势。图 1(1) 中竖直虚线标记了中国在 1992 年、2000 年和 2008 年三次修订《中华人民共和国专利法》。1991—2002 年中国平均专利质量出现了稳步的上升, 而在 2002 年之后出现了平均专利质量提升放缓乃至停滞的现象。这背后反映出地方政府偏向于专利数量增长的竞赛行为与严格的专利审查标准间存在内在不一致性, 与文献中指出的中国专利“泡沫”出现的时点一致 (Hu & Jefferson, 2009; 张杰和郑文平, 2018)。在 2008 年《中华人民共和国专利法》修订之后, 平均专利质量快速上升且至 2016 年增速未见

放缓,表明该轮制度改革对专利质量提升产生了持续促进效应。这为中国专利制度改革对专利质量水平提升的重要作用提供了初步证据。

图1(2)显示,高、低质量两类专利占比呈反向变动:低质量专利占比在2000年前持续上升,在2000—2008年呈U形波动,2008年后持续下降;高质量专利占比则稳步上升,2000年后增速加快,并于2012年超过低质量专利占比。这表明近年来专利审查和知识产权保护制度的完善有效遏制了低质量专利增长,专利质量提升主要由高质量专利相对数量增长所驱动。



(1) 平均专利质量



(2) 高、低质量专利占比

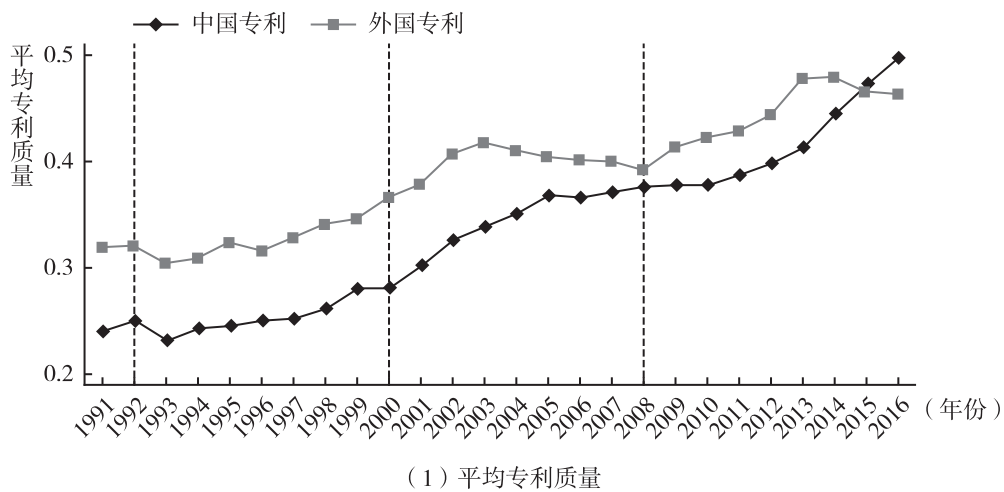
图1 中国境内发明专利质量变化趋势

(二)区分国内外主体专利质量

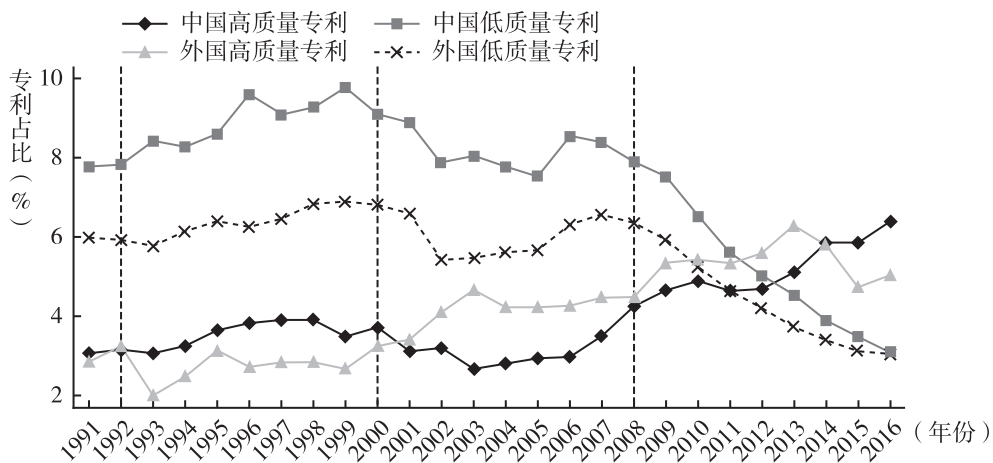
为分析中国境内专利质量的来源构成,本文进一步比较国内与国外申请人的平均专利质量变化特征。图2(1)显示,1991年以来国内主体平均专利质量呈显著上升趋势,增速快于国外主体,两者差距在2003年后逐步收敛,国内主体的平均专利质量于2015年实现反超。早期国内外专利质量差距的存在,一定程度上反映了国外发明者较高的技术知识存量 and 创新能力。然而,这一差距也可能源于国外主体的策略性申请行为。一方面,出于保护核心技术的考量,国外发明者倾向于不在中国布局最前沿创新;另一方面,其在华申请专利的动机更多在于通过专利制度限制本

土企业的技术模仿与赶超,以维护市场份额和商业利益。与此逻辑一致,在中国加入WTO之后,2003—2008年,国外主体在华授权平均专利质量呈现下降趋势;2008年后,国外主体专利质量开始回升并呈显著上升态势。这一转变反映了竞争性倒逼机制的作用:随着中国本土创新能力持续增强、国内主体平均专利质量稳步提升,国外申请人面临提高专利质量的竞争压力。图2(2)显示,国内主体低质量专利占比较高,但下降速度更快,样本期末与国外主体呈现收敛趋势;两类主体的高质量专利占比均逐年上升,且出现高低轮替现象。这表明中外专利质量差距的缩小主要源于国内低质量专利的相对减少。基于上述观察,我们总结出如下特征事实。

特征事实1:样本期间内,中国发明主体平均专利质量水平整体低于外国发明主体,但专利平均质量提升更快,主要原因是中国发明主体低质量专利占比的快速下降。



(1) 平均专利质量



(2) 高、低质量专利占比

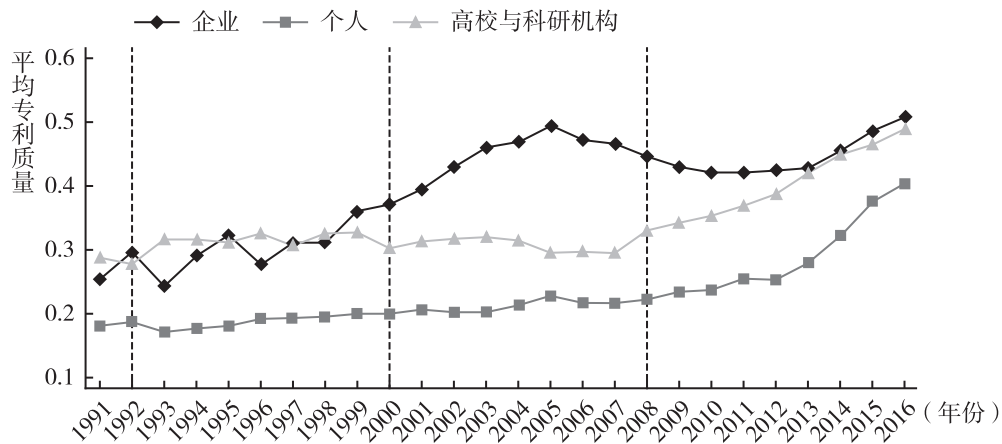
图2 中国境内授权的中国与外国专利质量变化趋势

(三)区分申请人类型专利质量

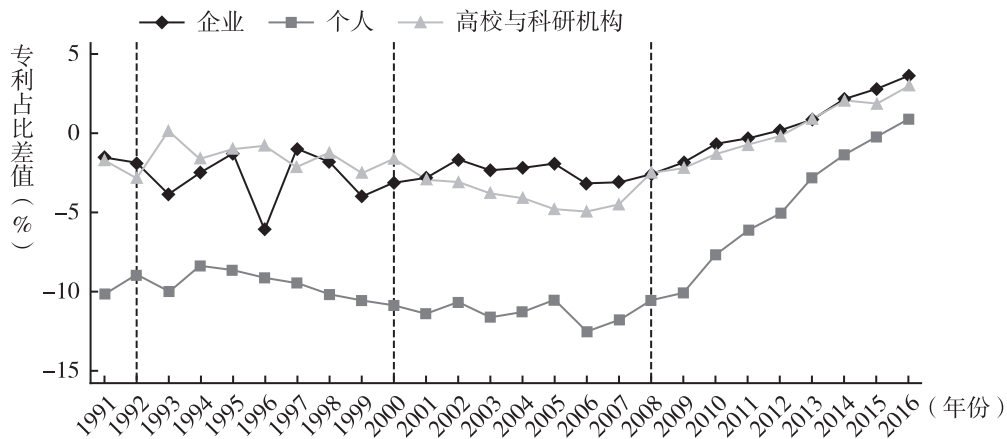
本部分将中国境内专利申请人划分为企业、个人以及高校与科研机构三类,分

别考察其专利质量的变化特征。图3(1)展示了三类申请主体的平均专利质量水平。1991—2016年三类主体的平均专利质量整体呈上升态势,但水平与增速存在明显差异:企业申请人平均专利质量最高,高校与科研机构次之,个人申请人最低。从变化趋势看,个人申请人及高校与科研机构的平均专利质量在2008年之前提升幅度较小,此后才呈现显著上升趋势;企业申请人的平均专利质量则呈波动式上升,1991—2005年稳步提升,之后经历U形调整,拐点出现在2011年。受此影响,高校与科研机构专利质量与企业专利质量的差距在2005年后逐步缩小,2012年后两者平均专利质量趋于接近。图3(2)展示了三类申请人高质量专利与低质量专利占比的差值。三类主体均呈现先降后升的趋势,拐点出现在2006年前后。相比较而言,企业申请人贡献了较多的高质量专利,高校与科研机构次之,而个人申请人贡献了较多的低质量专利。基于以上分析,本文得到如下特征事实。

特征事实2:样本期间内,不同申请主体的平均专利质量呈现企业>高校与科研机构>个人的基本格局,三类主体的平均专利质量均有所提升。其中,个人申请人贡献了较多的低质量专利,企业及高校与科研机构贡献了较多的高质量专利。



(1) 平均专利质量

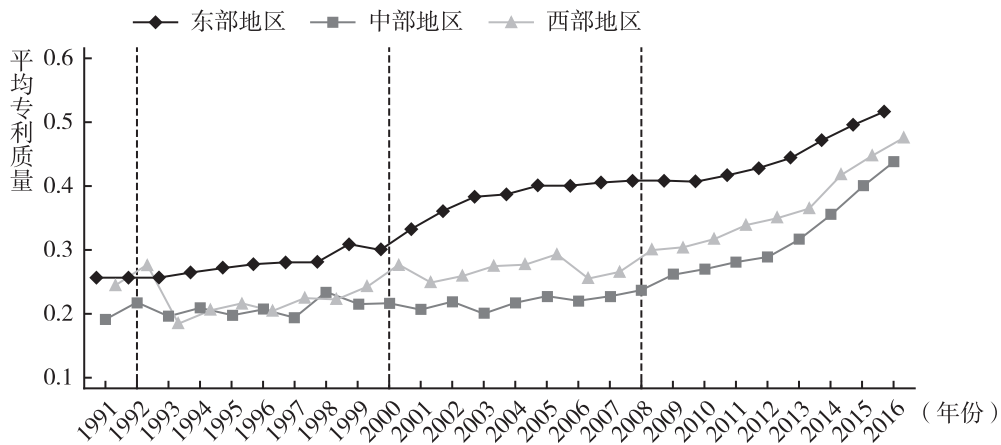


(2) 高质量与低质量专利占比差值

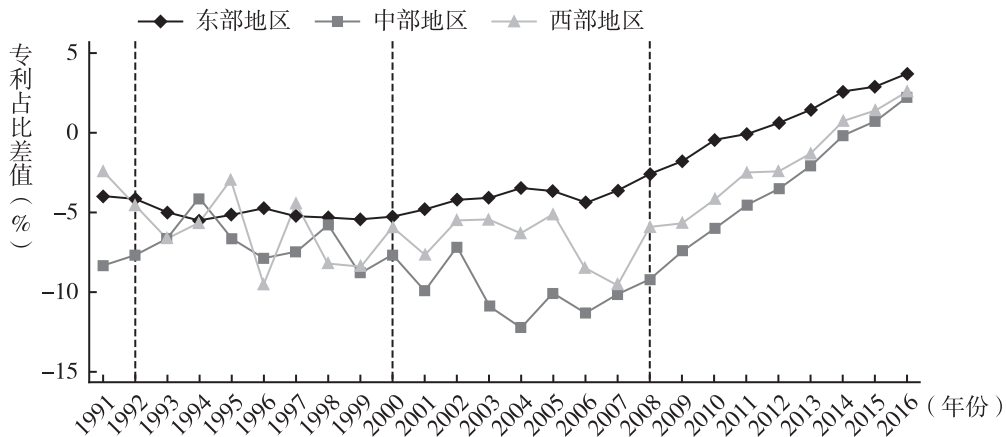
图3 中国不同专利申请主体的专利质量变化趋势

(四)区分地区专利质量

为探究中国发明专利质量的区域分布特征,本文按照专利申请人所在地将样本划分为东、中、西部三大地区。^①图4(1)显示,三大地区的平均专利质量均呈逐步上升态势,但水平存在差异:东部地区平均专利质量最高,西部地区次之,中部地区最低。2008年后,中、西部地区平均专利质量增速加快,呈现向东部地区追赶的趋同态势。这一趋同效应可能源于创新知识的空间溢出,也可能与向中、西部倾斜的创新支持政策有关。值得注意的是,西部地区平均专利质量高于中部地区,这一格局与区域产业结构密切相关。西部地区(如四川、陕西、重庆)聚集了航空航天、新材料、电子信息等国家级科研平台和大型军工集团,创新活动更为集中,形成了



(1) 平均专利质量



(2) 高质量与低质量专利占比差值

图4 中国不同地区专利质量变化趋势

^① 东、中、西部定义如下:东部为北京、天津、河北、辽宁、上海、江苏、浙江、福建、山东、广东、海南,中部为山西、内蒙古、吉林、黑龙江、安徽、江西、河南、湖北、湖南、广西,西部为四川、贵州、云南、西藏、重庆、陕西、甘肃、青海、宁夏、新疆。由于研究时间跨度较大,为保持年代一致性,主要依据20世纪90年代广泛采用的“七五”计划中的划分方式。考虑到对东、中、西部省份数量的平均以及东部省份普遍较高的经济发展水平,本文将广西移至中部地区统计,该处理不影响本文主要结论。

若干高价值专利集群。相比之下,中部地区以传统制造和原材料产业为主,2010年前后受专利数量考核与财政补贴政策驱动,策略性和低质量申请比例较高,拉低了整体平均水平。进一步分析表明,中部地区的低均值主要源于大量低质量专利的稀释效应,而其高质量专利的绝对数量并不逊于西部。图4(2)展示了三大地区高质量专利与低质量专利占比的差值变化。2008年之前,东部地区该差值较为稳定,而中、西部地区呈波动式下降,可能与地区间围绕专利数量的“创新锦标赛式”竞争有关。2008年后,各地区高质量专利占比相对于低质量专利占比均显著提升。根据以上分析,本文总结出如下特征事实。

特征事实3:样本期间内,平均专利质量呈现东部>西部>中部的区域格局,三大地区平均专利质量均稳步上升。东部地区贡献了较多的高质量专利,中部和西部地区贡献了较多的低质量专利。

(五)区分技术领域专利质量

按照IPC部类划分标准,本文将中国发明专利按技术领域划分为8个大类,考察各领域专利质量的变化特征。图5(1)展示了不同技术领域的平均专利质量变化趋势。除化学、冶金领域外,其他技术领域平均专利质量均呈上升态势,反映出中国在多数技术领域的创新质量得到有效提升。具体而言,电学和物理领域平均专利质量较高,且增幅最大;人类生活需要领域专利质量稳步上升,2003年后位居各领域第三;化学与冶金领域经历一段时间的质量下滑后趋于平稳,但仍处于较高水平;其他技术领域则呈现较为平缓的增长或稳定态势。图5(2)展示了各技术领域高质量专利与低质量专利占比的差值变化。作业与运输、纺织与造纸、固定建筑物、机械工程等领域贡献了较多的低质量专利,而电学、物理、人类生活需要等领域贡献了较多的高质量专利,这与平均专利质量的变化趋势基本一致。根据以上分析,本文归纳出如下特征事实。

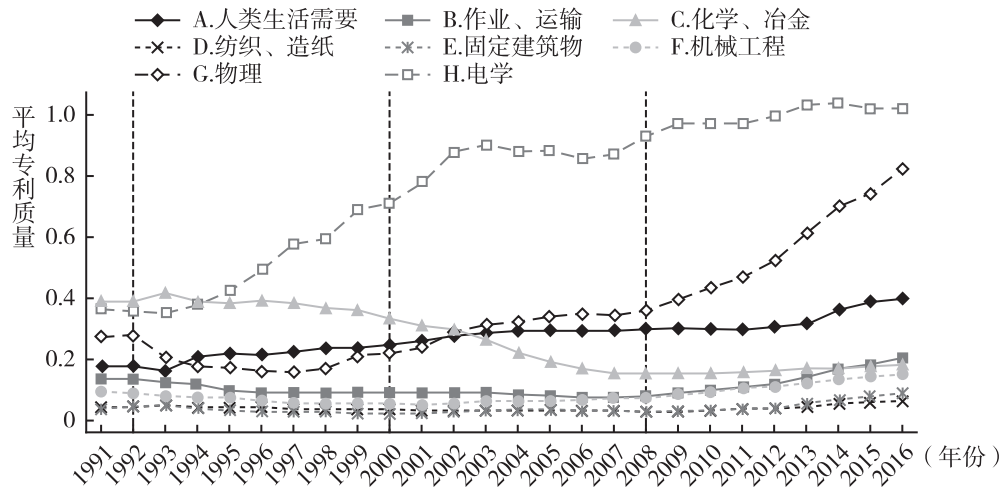
特征事实4:样本期间内,电学、物理和人类生活需要领域的专利平均质量较高,增长趋势最为明显,贡献了较多的高质量专利。

(六)专利质量变化机制分解

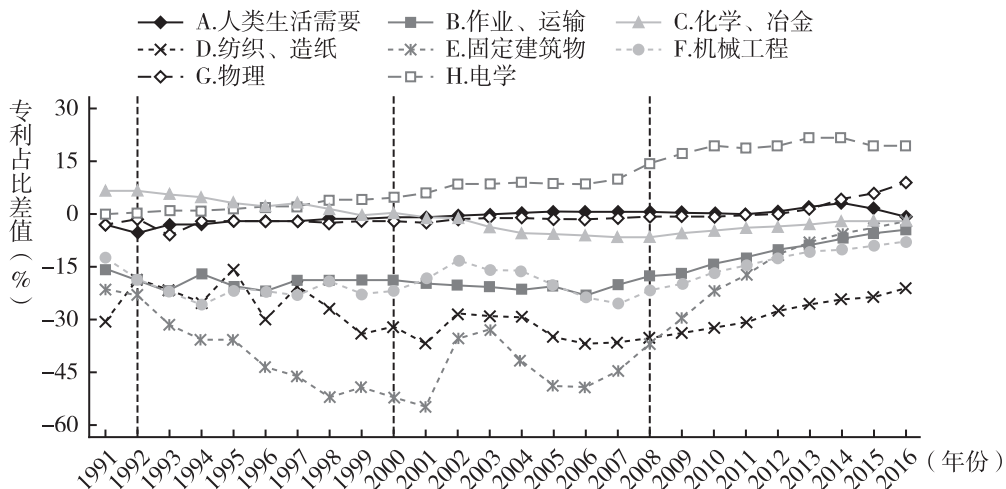
为深入理解专利质量变化的驱动机制,本文借鉴Olley & Pakes(1996)的分解思路,对专利质量变化进行均值—协方差分解。 $\bar{Q}_{g,t}$ 表示分组 g 专利在年份 t 的平均质量, \bar{Q}_t 表示年份 t 的所有专利平均质量,则有:

$$\bar{Q}_t = \underbrace{\sum_{g \in G} \frac{N_{g,t}}{N_t} (\bar{Q}_{g,t} - \bar{Q}_t)}_{\text{质量与数量协方差}} + \underbrace{\bar{Q}_t}_{\text{质量均值}} \quad (10)$$

其中, $N_{g,t}$ 为分组 g 在年份 t 的专利数量, N_t 表示年份 t 的专利总量。 G 表示分组指标集合,如按申请人类型分组,则 $G = \{\text{企业, 个人, 高校与科研机构}\}$ 。专利质量均值 $\bar{Q}_t = \frac{1}{\|G\|} \sum_{g \in G} \bar{Q}_{g,t}$ 为各分组专利质量均值, $\|G\|$ 表示组数。该分解将整体专利质量变动拆解为两部分:均值项反映各类主体质量水平的普遍性提升,协方差项反映



(1) 平均专利质量



(2) 高质量与低质量专利占比差值

图5 中国不同技术领域专利质量变化趋势

高质量专利向特定分组集聚所带来的结构性提升。

表4报告了按申请人类型、地区和IPC技术分类三个维度的分解结果。结果显示:第一,按申请人类型和地区分组时,均值项对质量增长的贡献显著高于协方差项,约为后者的10倍,表明各类主体和地区间的质量提升较为普遍,未出现严重的质量集聚效应。第二,按IPC技术分类时,均值项贡献约为协方差项的3倍,显示专利质量提升存在一定的技术领域集中性,高质量专利更倾向于集聚在特定技术门类。综合以上分析,本文归纳得到如下特征事实。

特征事实5:历次专利制度改革对中国整体专利质量提升具有显著推动作用。不同申请人类型和地区间的质量提升主要源于各自质量水平的普遍提升;在技术领域维度,专利质量提升既体现为整体均值提升,也反映出高质量专利向特定领域的集聚趋势。

表4 专利质量变化的均值—协方差分解

年份区间	按申请人类型		按地区		按IPC技术分类	
	协方差	质量均值	协方差	质量均值	协方差	质量均值
1991—1996	-0.0165	0.2558	0.0138	0.2301	0.0603	0.1836
1997—2001	-0.0154	0.2878	0.0231	0.2528	0.0677	0.2082
2002—2006	0.0222	0.3286	0.0587	0.2915	0.1074	0.2428
2007—2011	0.0435	0.3355	0.055	0.3229	0.1239	0.2541
2012—2016	0.0349	0.4092	0.032	0.4135	0.1153	0.3302

注：年份区间内专利质量和数量的协方差以及专利质量均值为区间内各年均值。

(七) 战略性新兴产业专利质量

中国战略性新兴产业概念于2008年国际金融危机后正式提出，同年《中华人民共和国专利法》完成重要修订。本文选取2008年后的专利数据，考察战略性新兴产业的专利质量变化。依据IncoPat数据库的产业分类，在本文使用的专利数据中，中国战略性新兴产业的专利产出数量占比超过70%，体现出战略性新兴产业在推动和引领技术创新方面的核心作用。^①

2012年《“十二五”国家战略性新兴产业发展规划》(以下简称《规划》)明确提出的战略目标包括：“掌握一批具有主导地位的关键核心技术，建成一批具有国际先进水平的创新平台，发明专利质量数量和技术标准水平大幅提升。”由此可见，该《规划》强调专利产出“质”“量”并重的发展导向。为理解这一《规划》对战略性新兴产业专利质量的作用，本文以美国、德国、日本、法国和韩国的战略性新兴产业为对照组，构建双向固定效应模型。表5报告了回归结果：该《规划》对发明专利质量的提升效应为8.2%($p < 0.01$)，政策对专利数量的提升效应为48.4%($p < 0.01$)。事件研究法显示专利质量满足事前平行趋势，但专利数量不符合事前平行趋势；使用合成双重差分方法(Arkhangelsky et al., 2021)调整后，专利数量回归的估计结果为50.3%($p < 0.01$)，与调整前接近。^②据此，本文总结如下特征事实。

特征事实6：《“十二五”国家战略性新兴产业发展规划》不仅显著提升了相关产业的发明专利申请数量，也有效提高了专利的平均质量水平，表明政策在促进创新能力提升方面发挥了重要作用。

表5 《“十二五”国家战略性新兴产业规划》对专利质量与数量的影响

变量	双向固定效应		合成双重差分
	专利质量	专利数量	专利数量
	(1)	(2)	(3)
《“十二五”国家战略性新兴产业发展规划》	0.082*** (0.017)	0.484*** (0.038)	0.503*** (0.025)

①② 因篇幅所限，相关内容详见本刊网站登载的附录。

续表 5

变量	双向固定效应		合成双重差分
	专利质量	专利数量	专利数量
	(1)	(2)	(3)
IPC—行业固定效应	是	是	是
年份固定效应	是	是	是
观测值	15327	15327	15327

注:括号中为二分位战略性新兴产业层面得到的聚类标准误,合成双重差分模型标准误由 500 次 Bootstrap 得到。

五、专利质量的技术进步效应及其对经济增长的贡献

本部分首先提供专利质量的技术进步效应的微观证据,然后在此基础上构建包含专利质量和数量的增长分解框架,从“质”与“量”两个维度评估创新对中国经济增长的贡献。

(一)专利质量的技术进步效应:微观证据

大量的专利技术内容是流程创新,一般表现为生产流程优化和效率提升,这会提高企业的生产率。为了验证专利质量和企业生产率变动的关系,本文考虑了不同方法测算的生产率指标作为被解释变量。本文将专利质量数据与 1998—2007 年中国工业企业数据库相匹配,对专利质量和企业生产率变动的关系进行回归分析,首先尝试使用人均销售额对数和人均增加值对数测度企业劳动生产率。为进一步考察专利质量对企业全要素生产率的影响,本文还借鉴 Chen et al. (2021) 的研究,分行业估计了包含专利质量的内生生产率演化方程,得到企业全要素生产率。^①

表 6 报告了估计结果。列(1)(2)显示,平均专利质量与劳动生产率(人均销售额、人均增加值)呈显著正相关。列(3)—(5)考察全要素生产率:控制企业固定效应后,平均专利质量的生产率促进效应不再显著,但高质量专利数量对全要素生产率有显著正向作用,低质量专利未体现出显著的生产率促进效应。值得注意的是,专利数量在所有模型中均呈显著正向作用,表明企业生产效率变化受到创新“量”与“质”的共同驱动。

由于许多重要创新表现为产品更新而非流程改进,仅以生产率变动检验专利质量指标可能不够全面。本文进一步匹配 2008—2014 年全国企业创新调查数据库,考察专利质量与新产品产值的关系。^②列(6)—(7)显示,专利质量对新产品产值有显著正向影响;在控制专利平均质量后,高、低质量专利数量及专利总数的系数均不显著,凸显了本文专利质量指标在解释产品创新方面的价值。^③

^{①②} 因篇幅所限,相关内容详见本刊网站登载的附录。

^③ 本文没有对专利基于产品创新和流程创新进行区分,而是使用企业层面的专利质量指标综合反映企业的产品创新和流程创新水平。因此,无法单独反映流程创新和产品创新分别对生产效率和新产品生产的影响。

表6 中国工业企业发明专利质量的技术进步效应

变量	生产率					新产品	
	人均 销售额	人均 增加值	全要素 生产率	全要素 生产率	全要素 生产率	新产品 产值	新产品 产值
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
平均专利质量	0.0447** (0.0194)	0.0750** (0.0303)	0.0621*** (0.0234)	-0.0018 (0.0268)	-0.0033 (0.0269)	0.0189* (0.0111)	0.0191* (0.0111)
高质量专利 数量					0.0025*** (0.0008)		-0.0003 (0.0006)
低质量专利 数量					-0.0021 (0.0038)		0.0021 (0.0025)
专利数量	0.0275*** (0.0058)	0.0402*** (0.0093)	0.0260*** (0.0057)	0.0142* (0.0076)	0.0139* (0.0082)	0.0001 (0.0001)	0.0001 (0.0001)
企业固定效应	是	是	否	是	是	是	是
观测值	1029725	983983	700226	631818	631818	124512	124512
Adj. R ²	0.8280	0.6686	0.2371	0.4015	0.4015	0.815	0.815

注：所有回归均控制了滞后期被解释变量、是否有发明专利申请以及年份固定效应，其中全要素生产率估计方程控制了生产率的1—3阶滞后项。括号中为企业层面聚类的标准误。

(二) 专利质量与数量的经济增长效应

为识别中国专利质量变化对经济增长的贡献，本文在Cobb-Douglas生产函数框架下，将专利数量与质量同时引入全要素生产率(TFP)之中。基于Griliches(1979, 1998)、Doreszelski & Jaumandreu(2013)、Kelly et al.(2021)等研究，本文假设专利数量和平均专利质量都对全要素生产率产生影响，推导得到包含专利数量扩张与质量提升的扩展型Cobb-Douglas生产函数：

$$Y_{jt} = N_{jt}^{\beta} \bar{Q}_{jt}^{\gamma} K_{jt}^{\alpha_k} L_{jt}^{\alpha_l} e^{v_t + \varepsilon_{jt}} \quad (11)$$

其中， N_{jt} 为省份的发明专利数量， \bar{Q}_{jt} 为平均专利质量，参数 β 与 γ 分别衡量专利数量扩张与专利质量提升对经济增长的边际贡献， v_t 和 ε_{jt} 分别为外生的时间趋势和随机扰动项。对数化后得到估计所需的结构方程：^①

$$y_{jt} = \alpha_k k_{jt} + \alpha_l l_{jt} + \beta n_{jt} + \gamma \bar{q}_{jt} + v_t + \varepsilon_{jt} \quad (12)$$

估计结果显示，专利数量系数 β 估计值为0.0935，本文专利质量指标系数 γ 为0.117，且都在1%的水平上显著。将被引次数与权利要求字数作为控制变量后，系数估计变化较小，说明本文专利质量指标对于宏观产出具有显著且稳定的增量解释力。

基于专利数量与质量产出弹性系数的估计结果，本文按照“贡献=系数×年度指标变化量”的口径，分省测算了1992—2016年期间资本、劳动与TFP对经济增长的

① 因篇幅所限，相关内容详见本刊网站登载的附录。

贡献,并进一步分解了专利数量与专利质量对TFP提升的贡献,结果如表7所示。平均来看,资本投入贡献为75.7%,劳动贡献为3.9%,与既有文献对中国GDP增长的分解结论基本一致(Wei et al., 2017)。专利质量贡献约为3%,专利数量贡献为21%,约为平均专利质量贡献的7倍。尽管专利质量贡献的量级小于专利数量的贡献,但其相对专利数量的贡献在近年来呈现显著的上升趋势。这一结构与中国专利体系由“数量扩张优先”(Hu et al., 2017; 龙小宁和王俊, 2015)逐步转向“提质增效”的演进相一致。总体而言,专利质量的持续提升及其所反映的创新能力提升,对经济增长的拉动效应正在增强,但其绝对贡献仍偏低,表明科技成果转化能力仍有提升空间。

表7 专利质量与数量变化对中国经济增长贡献分解

年份区间	TFP 增长贡献			资本贡献	劳动贡献
	专利质量贡献	专利数量贡献	其他贡献		
1992—1996	0.021	0.275	-0.12	0.795	0.046
1997—2001	0.023	0.190	0.004	0.782	0.001
2002—2006	0.006	0.192	0.051	0.660	0.091
2007—2011	0.013	0.169	0.073	0.731	0.013
2012—2016	0.082	0.222	-0.166	0.819	0.043
平均	0.029	0.210	-0.031	0.757	0.039

六、结论与政策含义

专利作为中国自主创新活动与高水平科技自立自强的重要载体,已逐步演变为影响创新驱动发展战略能否有效落地的关键变量。其中,专利质量的提升日益成为各级政府政策制定者与国内外学界高度关注的前沿议题。相应地,对专利质量进行科学测算并揭示其形成与演化的驱动机制,已成为当前中国创新研究中最重要、最为紧迫的基础性问题之一。

基于此,本文依托中国发明专利文本数据库,在运用自然语言处理技术测算专利文本相似度的基础上,探索并构建了契合中文专利文本特征的专利质量测度方法。为检验该指标体系的有效性、可靠性与适用性,本文开展了一系列稳健性与外部一致性检验。结果表明,本文测算的专利质量指标在解释专利存续时间以及识别中国专利申请“年末突击”的质量效应方面表现高度稳健,在规避专利质量识别的两类典型错误上整体优于传统指标,对于企业经营绩效具有较强的预测与解释能力。

在此基础上,本文利用所构建的专利质量指标刻画并解释中国发明专利质量的变化特征及背后的驱动机制,并在国别、申请人类型、地区与技术类别维度上归纳提炼出若干特征事实。本文进一步提供了专利质量促进技术进步的微观

证据,并基于同时纳入专利质量与数量的扩展型生产函数,构建了包含专利质量的经济增长分解框架,估算了1992—2016年中国专利质量与数量变化对经济增长的贡献。总体而言,本文开发的测算策略与指标属于支撑中国自主创新体系建设与科技创新能力持续提升的基础性研究成果:既可为学界探索适合中国专利的质量测算策略提供方法工具,也可为专利价值与自主创新能力评估提供识别手段,并为理解中国专利制度改革与各类创新激励政策的效果评估提供基础变量,进而对各级政府落实创新驱动发展战略与高水平科技自立自强任务具有参考意义。

结合本文揭示的中国发明专利质量变化趋势、驱动机制及其对经济增长的贡献特征,本文的政策含义主要体现在以下三个方面。

首先,经验结果表明,中国历次专利制度的调整与改革对专利质量提升具有实质性的正向激励作用。尤其是2008年新一轮专利制度改革之后,中国专利质量摆脱了此前呈现的倒U形波动特征,进入更为持久的上升通道,专利发展体系开始显现“由量变到质变”的关键转换。由此引申的政策含义在于:在中国经济全面转向高质量发展、创新驱动发展更加依赖自主创新体系、加快推进自主可控的全球产业链供应链布局的战略阶段,中国亟须进一步优化专利制度与政策工具,从而激励各类微观主体面向原始创新、关键核心技术、颠覆性技术与关键共性技术开展高强度研发投入,并促进高质量专利产出。

其次,应高度重视专利质量结构性调整中“以质换量”的战略性转型。研究发现,尽管各区域专利平均质量总体上升,但高质量占比提升、低质量占比下降的结构优化在不同区域板块呈现显著差异,因此政策重心应由“数量扩张”转向“质量牵引”,构建差异化的区域创新激励机制:一是改革财政激励的方向和强度。在东部高技术密集区域适度降低对专利活动的数量型激励,避免过度补贴对质量提升的扭曲;在中、西部地区维持必要扶持力度,同时将补贴与高价值专利培育、实施转化、关键技术突破挂钩。二是对弱势行业和薄弱环节给予实质性扶持。研究发现,作业运输、纺织造纸、固定建筑物、机械工程等领域的低质量专利较高,可通过关键共性技术平台联合攻关,为中小企业提供专利导航支持,对产业链关键点开展专利组合培育,并配套高价值专利保险与质押融资增信。三是治理年末突击申请导致的质量下滑。在年度绩效考核约束下,各类创新主体呈现“年终冲刺式”专利申请行为,就此,应改革创新绩效考核机制,降低短期数量指标权重,强化质量导向的中长期评价。四是总结推广通过规划引导、集群培育提升创新质量的经验。西部地区依托航空航天、新材料、电子信息等国家级科研平台和大型军工集团,形成若干高价值专利集群,专利质量稳步提升,为其他区域提供了可借鉴的实践范式。

最后,应按申请人类型实施更精细的分类施策,构建“企业主导、个人活跃、高校院所支撑”的高质量创新生态。研究揭示,不同申请主体的专利质量呈现企业>高校与科研机构>个人的基本格局。对企业而言,其专利质量最高且贡献了较多的

高质量专利,体现了我国产业实践先行的创新特征。未来应进一步强化企业科技创新主体地位,围绕“强链补链延链”建立关键技术清单与专利路线图,支持龙头企业牵头组建创新联合体,完善研发费用加计扣除、高价值专利培育与产业化后补助等激励机制,健全职务发明权益分配与转化收益分享制度,并通过政府采购与示范工程加速落地验证。对个人申请者而言,尽管其专利质量偏低,但个体创新者在创新生态中仍具有活力与潜力。应提升创新服务可及性,提供更普惠的知识产权公共服务,鼓励合规使用AI技术辅助开展检索比对、权利要求优化与风险提示,并依托众创空间、开源硬件平台及专利池降低产业化门槛,激发个体创新者的质量意识与转化能力。对高校与科研机构而言,研究揭示了高校与科研机构在国家战略科技力量体系中发挥的作用尚不充分。建议科研院所深入推进评价改革,降低发明专利数量在成果评价中的权重,强化高价值专利与转化绩效;构建以质量与应用价值为核心的科研激励与评价体系,完善收益分配与晋升激励联动机制,并与企业共建联合实验室、概念验证中心和中试平台,提升可转化、可规模化的高质量创新供给能力。

参考文献

- 陈强远、林思彤、张醒,2020:《中国技术创新激励政策:激励了数量还是质量》,《中国工业经济》第4期。
- 陈强远、赵浩云、林思彤、申宇,2024:《中国高质量技术创新:情境叙事与测度体系》,《管理世界》第5期。
- 寇宗来、刘学悦,2020:《中国企业的专利行为:特征事实以及来自创新政策的影响》,《经济研究》第3期。
- 黎文靖、郑曼妮,2016:《实质性创新还是策略性创新?——宏观产业政策对微观企业创新的影响》,《经济研究》第4期。
- 龙小宁、王俊,2015:《中国专利激增的动因及其质量效应》,《世界经济》第6期。
- 孙雅慧、时省、彭飞、吴华清,2024:《研发补贴与渐进式创新锁定:基于机器学习的专利文本分析》,《经济研究》第11期。
- 张杰、陈容、郑姣姣,2022:《策略性创新抑或真实性创新——来自中国企业设立研发机构的证据》,《经济管理》第3期。
- 张杰、郑文平,2018:《创新追赶战略抑制了中国专利质量么?》,《经济研究》第5期。
- 朱雪忠、乔永忠、万小丽,2009:《基于维持时间的发明专利质量实证研究——以中国国家知识产权局1994年授权的发明专利为例》,《管理世界》第1期。
- Arkhangelsky, D., S. Athey, D. A. Hirshberg, G. W. Imbens, and S. Wager, 2021, “Synthetic Difference-in-differences”, *American Economic Review*, 111(12), 4088—4118.
- Artz, K. W., P. M. Norman, D. E. Hatfield, and L. B. Cardinal, 2010, “A Longitudinal Study of the Impact of R&D, Patents, and Product Innovation on Firm Performance”, *Journal of Product Innovation Management*, 27(5), 725—740.
- Balsmeier, B., M. Assaf, T. Chesebro, G. Fierro, K. Johnson, S. Johnson, G. C. Li, S. Lück, D. O’Reagan, and B. Yeh, 2018, “Machine Learning and Natural Language Processing on the Patent Corpus: Data, Tools, and New Measures”, *Journal of Economics and Management Strategy*, 27(3), 535—553.
- Bergeaud, A., A. B. Jaffe, and D. Papanikolaou, 2025, “Natural Language Processing and Innovation Research”, National Bureau of Economic Research.
- Chen, K. P., X. Dai, K. Li, and P. Lin, 2025, “Year-end Rush and Career Tournament: Theory and Evidence from Patent Applications in China”, MPRA Paper, No.124898.
- Chen, Z., J. Zhang, and Y. Zi, 2021, “A Cost-benefit Analysis of R&D and Patents: Firm-level Evidence from

China”, *European Economic Review*, 133, 103633.

Dang, J., and K. Motohashi, 2015, “Patent Statistics: A Good Indicator for Innovation in China? Patent Subsidy Program Impacts on Patent Quality”, *China Economic Review*, 35, 137—155.

Doraszelski, U., and J. Jaumandreu, 2013, “R&D and Productivity: Estimating Endogenous Productivity”, *Review of Economic Studies*, 80(4), 1338—1383.

Griliches, Z., 1979, “Issues in Assessing the Contribution of Research and Development to Productivity Growth”, *Bell Journal of Economics*, 10(1), 92—116.

Griliches, Z., 1990, “Patent Statistics as Economic Indicators: A Survey”, *Journal of Economic Literature*, 28(4), 1661—1707.

Hall, B. H., A. Jaffe, and M. Trajtenberg, 2005, “Market Value and Patent Citations”, *RAND Journal of Economics*, 36(4), 16—38.

He, Z. L., T. W. Tong, Y. Zhang, and W. He, 2018, “A Database Linking Chinese Patents to China’s Census Firms”, *Scientific Data*, 5(1), 1—16.

Higham, K., G. de Rassenfosse, and A. B. Jaffe, 2021, “Patent Quality: Towards a Systematic Framework for Analysis and Measurement”, *Research Policy*, 50(4), 104215.

Hu, A. G., and G. H. Jefferson, 2009, “A Great Wall of Patents: What Is Behind China’s Recent Patent Explosion?”, *Journal of Development Economics*, 90(1), 57—68.

Hu, A. G. Z., P. Zhang, and L. Zhao, 2017, “China as Number One? Evidence from China’s Most Recent Patenting Surge”, *Journal of Development Economics*, 124, 107—119.

Jaffe, A. B., 1986, “Technological Opportunity and Spillovers of R&D: Evidence from Firms’ Patents, Profits, and Market Value”, *American Economic Review*, 76(5), 984—1001.

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.

Kogan, L., D. Papanikolaou, A. Seru, and N. Stoffman, 2017, “Technological Innovation, Resource Allocation, and Growth”, *Quarterly Journal of Economics*, 132(2), 665—712.

Lerner, J., 1994, “The Importance of Patent Scope: An Empirical Analysis”, *RAND Journal of Economics*, 25(2), 319—333.

Lerner, J., and A. Seru, 2017, “The Use and Misuse of Patent Data: Issues for Corporate Finance and Beyond”, NBER Working Paper, No.24053.

Marco, A. C., 2007, “The Dynamics of Patent Citations”, *Economics Letters*, 94(2), 290—296.

Olley, G. S., and A. Pakes, 1996, “The Dynamics of Productivity in the Telecommunications Equipment Industry”, *Econometrica*, 64(6), 1263—1297.

Pakes, A., 1985, “On Patents, R&D, and the Stock Market Rate of Return”, *Journal of Political Economy*, 93(2), 390—409.

Pakes, A., 1986, “Patents as Options: Some Estimates of the Value of Holding European Patent Stocks”, *Econometrica*, 54(4), 755—784.

Schwartz, D. L., and J. P. Kesan, 2013, “Analyzing the Role of Non-practicing Entities in the Patent System”, *Cornell Law Review*, 99(2), 425—456.

Thompson, P., 2006, “Patent Citations and the Geography of Knowledge Spillovers: Evidence from Inventor- and Examiner-Added Citations”, *Review of Economics and Statistics*, 88(2), 383—388.

Trajtenberg, M., 1990, “A Penny for Your Quotes: Patent Citations and the Value of Innovations”, *RAND Journal of Economics*, 21(1), 172—187.

Wei, S. J., Z. Xie, and X. Zhang, 2017, “From ‘Made in China’ to ‘Innovated in China’: Necessity, Prospect, and Challenges”, *Journal of Economic Perspectives*, 31(1), 49—70.

Measurement, Mechanism of Change and Economic Growth Effect of China's Patent Quality

CHEN Zhiyuan^a, ZHANG Yiran^b, ZHANG Jie^c and LIU Zexuan^d

(a: Business School, Renmin University of China;

b: China Center for Economic Studies, Fudan University;

c: Institute of China's Economic Reform and Development, Renmin University of China;

d: Joint Research Institute, Nanjing Audit University)

Summary: Constructing an effective measurement system for China's patent quality is critical for implementing China's innovation-driven development strategy. Since 2011, China has been the world's largest patent applicant, yet concerns persist about patent bubbles and strategic rather than substantive innovation, particularly driven by government subsidies. This raises fundamental questions about scientifically measuring patent quality and accurately assessing the innovation capabilities of Chinese economic entities.

Existing methods face significant limitations in China's context. Citation-based methods suffer from systematic deficiencies as citation disclosure is not mandatory. Economic effect approaches are limited to listed companies with data truncation. Text similarity-based methods offer unique advantages: comprehensive coverage, strong reproducibility, and alignment with knowledge distance theory.

This study analyzes China's invention patents granted during 1985-2021, employing natural language processing to calculate text similarity as the core indicator. Building upon Kelly et al. (2021), we develop a measurement approach adapted to China's patent system through four improvements: hybrid similarity strategies, excluding same-applicant patents, International Patent Classification (IPC) technology field weighting, and processing text through segmentation with Term Frequency-Backward Inverse Document Frequency (TF-BIDF) representations and field-year standardization. Multiple reliability tests validate our indicators' superiority.

We systematically document six stylized facts regarding China's patent quality evolution from 1991 to 2016. First, foreign applicants consistently exceed Chinese applicants in quality, though gaps have narrowed significantly. Second, among Chinese entities, research institutions maintain the highest quality, followed by universities and enterprises, with individual inventors exhibiting the lowest level, though all categories show improvement trends. Third, regional disparities are pronounced: eastern regions significantly outperform central and western regions, yet all demonstrate quality enhancement. Fourth, electrical engineering and instruments exhibit the highest quality, while chemistry and metallurgy remain at a lower level, though quality-quantity correlations have strengthened. Fifth, post-2008 strategic emerging industries show varied quality: new-generation information technology, new energy vehicles, and renewable energy rank higher. Sixth, the 12th Five-Year Plan for Strategic Emerging Industries (2012) significantly improved both patent quantity (by approximately 50%) and average quality (by approximately 8%), demonstrating effective policy impact on innovation capability enhancement.

We provide micro-level evidence for technological progress effects. Regression analyses controlling for enterprise fixed effects reveal that average patent quality positively correlates with labor productivity; high-quality patent counts significantly promote total factor productivity; low-quality patents, however, show no significant effects. Furthermore, matching with the National Innovation Survey database (2008-2014) demonstrates that patent quality significantly influences new product output value.

Our growth decomposition framework, incorporating both quality and quantity, assesses the contribution of innovation to economic growth. During 1992-2016, quality contributed approximately 3% to economic growth with an increasing trend, while quantity contributed approximately 21% with a declining trend, demonstrating China's transition from "quantity expansion" to "quality improvement". Although the absolute contribution of quality remains smaller, its sustained growth underscores the positive role in promoting continuous patent quality enhancement.

This study makes several contributions. First, it provides a comprehensive long-panel analysis of invention patent quality, filling gaps in subject and temporal dimensions. Second, it develops measurement methods adapted to China's patent characteristics, proposing hybrid similarity indicators and IPC-weighted adjustments to enhance their applicability. Third, it systematically reveals stylized facts and mechanisms of China's patent quality evolution, providing novel micro-evidence on how quality fosters technological progress and its marginal contributions to growth. Fourth, the developed indicators serve as essential tools for assessing the innovation capabilities of micro-entities and provide analytical instruments for understanding patent system reforms and evaluating innovation policies, offering guidance for implementing innovation-driven strategies.

Keywords: Invention Patents; Patent Quality Measurement; Text Similarity; Economic Growth Effect

JEL Classification: O31, O34, O38

(责任编辑: 溯 明)(校对: 王红梅)