

科学知识输入对发明价值的影响效应研究 ——基于反事实框架的因果推断

欧桂燕¹, 吴江^{1,2,3}, 岳名亮⁴, 王凯利⁵

(1. 武汉大学信息管理学院, 湖北武汉 430072; 2. 武汉大学电子商务研究与发展中心, 湖北武汉 430072;
3. 武汉数据智能研究院, 湖北武汉 430072; 4. 中国科学院武汉文献情报中心, 湖北武汉 430071;
5. 华中师范大学国家数字化学习工程技术研究中心, 湖北武汉 430079)

摘要: 科学知识被广泛视为创新过程中的“发明地图”, 其与发明价值之间的关系一直是科技政策与创新管理领域的核心议题。然而, 已有研究主要聚焦于二者的相关性分析, 研究结论不尽一致, 难以为科技决策提供确切的参考依据。本文基于反事实框架, 运用倾向得分匹配(PSM)与广义倾向得分匹配(GPSM)方法, 以美国专利商标局 USPTO 2001-2010 年授权的 1,685,970 件发明专利及其与科学论文构成的 6,473,214 个专利-论文引用对为研究样本, 构建专利科学引用 H 指数(PSH)作为量化科学知识输入强度的代理指标, 系统考察了科学知识输入及其强度对发明价值的因果效应。研究结果表明: (1) 科学知识输入对专利的前向引用、技术通用性和专利转让均具有显著的正向因果效应; (2) 科学知识输入强度与专利前向引用、专利转让次数均存在非线性关系, 随着 PSH 指数的递增, 专利的前向引用和转让次数表现出“先上升-后下降-再上升”的波动趋势, 而低强度与高强度区间的发明专利则都具有更高的技术影响力和市场交易潜力; (3) 高强度的科学知识输入是形成通用性技术的重要驱动力。本研究从因果推断视角验证了科学知识对发明价值的影响效应, 并结合 H 指数提出了评估科学知识输入强度的新方法, 为科技政策优化提供了重要的实证依据。

关键词: 科学知识; 发明价值; PSM 匹配; GPSM 匹配; 反事实框架

中图分类号: G203

文献标识码: A

创新驱动高质量发展的时代背景下, 提升发明创造的质量与价值已然成为科技创新领域重点关注的议题。Fleming 和 Sorenson 的开创性研究表明, 科学知识作为技术知识系统的重要输入, 能够充当发明者的“地图”, 指引其在复杂技术景观中进行有效搜索^[1]。这一理论贡献不仅深化了学界对科技互动的理解, 更激发了学者们从科学知识输入视角探究其对发明价值的影响机制。

基于这一研究视角, 学界开展了广泛的实证探索, 但实证结果却呈现出显著的差异。一方面, 以 Hohberger 为代表的研究表明, 基于科学的发明比非科学导向的发明在扩散速度上的更具优势, 特别是在生物技术等中等科学强度的行业中^[2]。另一方面, Sapsalis 等基于比利时大学的专利样本却未能证

实科学知识输入与发明价值的显著关联^[3]。更值得注意的是, 部分研究甚至发现了负向关系, 如曾德明等研究表明, 在生物医学工程行业, 科学知识获取广度反而抑制企业技术创新绩效^[4]。

这种实证证据的不一致性, 很大程度上源于未能充分考虑对科学知识采用过程的选择性偏差。知识重组理论指出, 技术发明是对现有知识元素搜索重组的结果。知识重组特征(如复杂性、新颖性、成熟度)与发明人特征(如发明经验、知识广度)都会影响科学知识输入和应用的结果^[5,6]。忽视这些因素导致的选择性偏差, 不仅会模糊科学知识对发明价值间的真实关系, 也会给政策制定与科技决策带来显著的不确定性。

鉴于传统回归分析难以有效处理上述样本选择

收稿日期: 2024-12-13; 修回日期: 2025-02-07

基金项目: 武汉大学区域国别研究课题; 国家自然科学基金专项项目(72442030); 国家自然科学基金重点项目(72232006)

作者简介: 欧桂燕(1994-), 女, 博士后。吴江(1978-), 男, 教授, 博士, 通讯作者, E-mail: jiangw@whu.edu.cn。岳名亮(1986-), 男, 研究员。王凯利(1992-), 女, 讲师。

偏差问题,有必要借助因果推断的方法准确评估科学知识输入对发明价值的净因果效应。因果效应评估的核心是基于反事实的比较:即在保持其他因素不变的情况下,比较技术发明过程中输入科学知识的结果(“实际发生的结果”)与不输入科学知识的结果(“反事实结果”)之间的差异。然而,科学知识的输入对发明价值的影响并非单一维度的问题,而是涉及两个重要方面:一是科学知识是否被输入;二是科学知识输入的强度(即输入的“量”)。前者关注科学知识输入的与否对发明价值的基本影响,后者则进一步探讨输入量的变化是否会引致发明价值的动态变化。

基于此,本研究从因果推断的视角,着重解决以下两个问题:

- (1) 科学知识输入与否对发明价值的影响效应如何?
- (2) 科学知识输入强度对发明价值的影响效应如何?

1 基于反事实框架的因果效应评估思路

1.1 反事实框架——潜在结果模型

早在 18 世纪初,Hume 就提出可以基于反事实思想解释事物间因果关系的形成与认知^[7]。这一思想,在 Rubin 的潜在结果模型中得到系统化发展,并成为因果推断研究中使用最广的“反事实框架”之一^[8]。该模型通过比较同一研究对象在接受与未接受处理两种状态下的结果差异来估计因果效应。由于历史的不可回溯性,这两种状态无法被同时观察,其中实际发生的状态即为“事实”,未发生的状态则为“反事实”。基于此,本研究构建的反事实分析框架如下:

(1) 基于科学知识输入这一处理变量(表示为 $D_i = \{0, 1\}$),将样本分为处理组(输入科学知识)与控制组(未输入科学知识)。具体而言, $D_i = 1$ 表示第 i 个技术发明输入了科学知识, $D_i = 0$ 表示第 i 个发明专利未输入科学知识。以发明价值为结果变量,用 Y_i 表示。对于任一发明 i ,存在两种潜在结果: $Y_i(1)$ 为处理条件下的发明价值, $Y_i(0)$ 为未处理条件下的发明价值。对于个体 i 而言,接受处理的因果效应定义为 $Y_i(1)$ 与 $Y_i(0)$ 之间的差值,公式表示如下:

$$\tau = Y_i(1) - Y_i(0) \quad (1)$$

(2) 由于同一个体的两种潜在结果无法被同时观察,个体因果效应难以精确估计。因此,本研究采用平均处理效应来估计它的期望值,公式表示如下:

$$E(\tau) = E[Y_i(1) | D_i = 1] - E[Y_i(0) | D_i = 0] = (E[Y_i(1) | D_i = 1] - E[Y_i(0) | D_i = 1]) + (E[Y_i(0) | D_i = 1] - E[Y_i(0) | D_i = 0]) = E[Y_i(1) - Y_i(0) | D_i = 1] + \varepsilon \quad (2)$$

$E[Y_i(1) - Y_i(0) | D_i = 1]$ 表示处理组的平均处理效应(简称 ATT), ε 称为选择性偏误。当 $\varepsilon = 0$ 情况下,处理组与控制组平均结果差异等同于 ATT,即实现了无偏估计。

1.2 倾向得分匹配

为了回答第一个研究问题,准确评估科学知识输入对发明价值的因果效应,本文采用倾向得分匹配法(Propensity Score Matching, PSM)消除选择性偏误。PSM 由 Rosenbaum 和 Rubin 提出^[8],其核心思想在于通过计算协变量的倾向值,构建与处理组特征相似的控制组,进而估计平均处理效应(ATT)。PSM 实施包含 4 个关键步骤:

(1) 选择协变量。纳入同时影响处理变量和结果变量的可观测因素以及仅影响结果变量的相关因素。

(2) 估计倾向得分。本研究采用 Logit 模型估计倾向得分 $p(X_i)$,即在给定协变量条件下个体 i 接受处理的条件概率:

$$p(X_i) = \Pr(D_i = 1 | X_i) = \frac{\exp(\beta X_i)}{1 + \exp(\beta X_i)} \quad (3)$$

(3) 选择匹配算法。本研究采用最近邻匹配算法。在控制组中识别倾向得分与处理组最为接近的样本。

(4) 平衡性检验和共同支撑假设。通过标准化偏差和核密度估计图等方法检验匹配后的平衡性,并验证共同支撑假设,以确保 ATT 估计的可靠性。

1.3 广义倾向得分匹配

为了回答第二个研究问题,本研究采用广义倾向得分匹配法(Generalized Propensity Score Matching, GPSM)进行分析。GPSM 突破了 PSM 对处理变量必须为二值的约束,能够估计连续处理变量的因果效应,进而估计不同强度下科学知识对发明价值影响的动态变化^[9]。本研究的 GPSM 实施包含三个步骤:

(1) 根据协变量 X_i 估算处理强度的条件概率密

度。采用专利科学引用 H 指数量化科学知识强度 K_i 。经标准化处理后,采用 Fractional Logit 模型估计,即:

$$E(K_i | X_i) = F(X_i \beta) = \frac{\exp(X_i \beta)}{1 + \exp(X_i \beta)}, \widehat{GPS}_i = [F(X_i \beta)]^{K_i} \cdot [1 - F(X_i \beta)]^{1-K_i} \quad (4)$$

$\widehat{GPS}_i$ 表示广义倾向得分值,它表示的是当处理变量 K_i 取值在 $[0,1]$ 范围内的概率。

(2) 在给定科学知识强度 K_i 以及倾向得分 $\widehat{GPS}_i$ 的情况下,构建结果变量 Y_i (即发明价值)的条件期望模型。本研究使用三阶多项式的模型,即:

$$E(Y_i | K_i, \widehat{GPS}_i) = \alpha_0 + \alpha_1 K_i + \alpha_2 K_i^2 + \alpha_3 K_i^3 + \alpha_4 \widehat{GPS}_i + \alpha_5 \widehat{GPS}_i^2 + \alpha_6 \widehat{GPS}_i^3 + \alpha_7 K_i \times \widehat{GPS}_i \quad (5)$$

(3) 估计剂量反应函数。将公式 7 的回归系数代入公式 8,计算不同科学知识强度水平下发明价值的平均剂量反应函数:

$$\widehat{E}[Y_i(k)] = \frac{1}{N} \sum_{i=1}^N \{ \widehat{\alpha}_0 + \widehat{\alpha}_1 k + \widehat{\alpha}_2 k^2 + \widehat{\alpha}_3 k^3 + \widehat{\alpha}_4 \widehat{r}(k, x) + \widehat{\alpha}_5 \widehat{r}(k, x)^2 + \widehat{\alpha}_6 \widehat{r}(k, x)^3 + \widehat{\alpha}_7 k \times \widehat{r}(k, x) \} \quad (6)$$

通过上述 3 个步骤,将科学知识强度的区间

$[0,1]$ 以 0.1 为步长划分为 10 个子区间,估计各区间因果效应系数,并通过剂量反应函数图可视化实证结果。

2 研究设计

2.1 数据与变量

为探究上述研究问题,本研究首先获取了美国专利商标局 (USPTO) 2001 - 2010 年间授权的 1,685,970 件发明专利数据。专利数据涵盖了专利标题、申请年份、发明人、申请人、引文、IPC 分类号及专利转让等基本信息。其次,本研究通过发明专利对科学论文的引用来量化科学知识输入的过程^[10]。数据源自于 Marx 和 Fuegi 于 2020 年开发的“Reliance on Science”专利 - 论文引用数据集^[11]。该数据集结合手动调整的启发式方法和 GROBID 机器学习包,实现了对专利文献中科学引文的全方位识别(包含首页引用和文本引用)。

基于上述,本文识别出 463,393 件(约占总样本的 27.5%)至少引用过 1 篇科学论文的发明专利。这些专利共引用了 1,407,439 篇科学论文,构成了 6,473,214 个专利 - 论文引用对。图 1 展示了本文的数据处理流程。

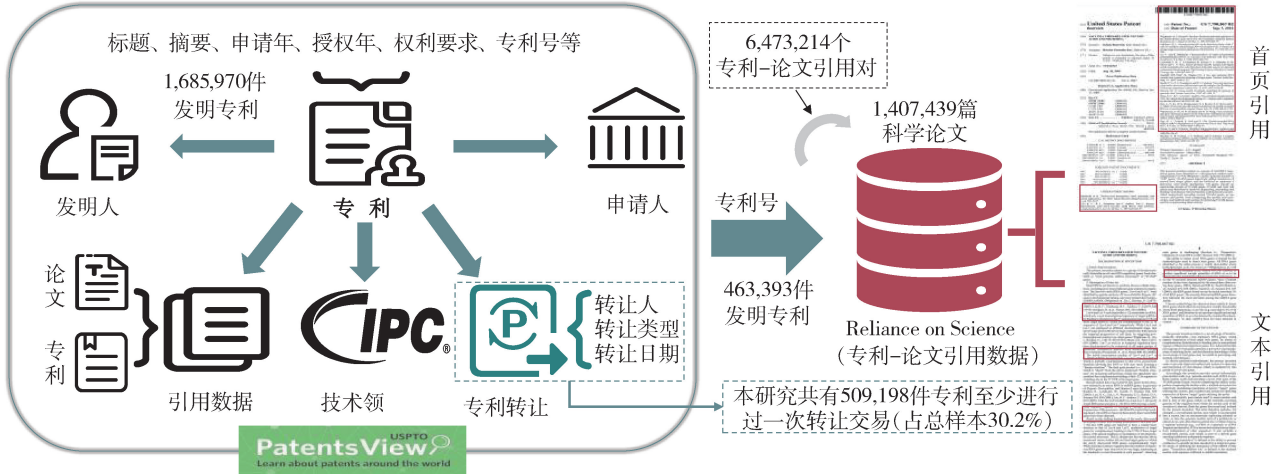


图1 数据处理流程图

Figure 1 Data processing flowchart

(1) 结果变量

发明价值涉及技术价值、商业价值等多个维度。但大多文献更关注科学知识对发明技术价值的影响,而忽略了科学知识对发明商业价值的作用。技

术发明的最终目标是实现市场的应用和检验,因此与高质量发展核心要义最为匹配的必然是专利发明的商业价值^[12]。基于此,本文旨在通过技术价值与商业价值的综合评估维度,系统地探索科学知识对

发明专利技术价值与商业价值的因果效应。

首先,技术价值是决定发明价值总量最为重要的下位价值,是其他价值存在的基础。①最常用于衡量一件发明技术价值的变量通常是焦点发明在后续发明中获得的引用次数,也称为“前向引用”。为了保证时间窗口的统一,本研究使用焦点发明自授权日起五年内获得的前向引用次数进行衡量^[10]。②其次,一项技术发明在不同领域应用的通用性也被认为是衡量技术价值的重要指标^[13]。拥有高技术通用性的发明专利(也被称为通用技术,即 GPT)由于可以与不同市场和领域的技术解决方案相结合并被采用来创建新的解决方案,通常被视为经济发展的“增长引擎”。然而,与前向引用衡量的技术价值相比,鲜少有研究分析科学知识在以技术通用性衡量的技术价值方面的作用。因此,本研究计算了 Trajtenberg 等提出的技术通用性指数^[14],以分析在技术发明过程中输入科学知识是否能够带来更通用的解决方案。技术通用性指数计算公式如下:

$$TechnologyGenerality = 1 - \sum_i^n \left(\frac{FC_{ip}}{FC_p} \right)^2 \quad (7)$$

FC_{ip} 表示焦点发明 p 在以前四位 IPC 分类号 (IPC4) 衡量的技术领域 i 获得的前向引用次数。 FC_p 则为焦点发明 p 的总前向引用次数。一项发明的通用性越高,通常意味着其具有更广泛的技术影响,因为其影响了多个技术领域的后续创新。

技术发明的最终目标是商业化,商业价值构成了技术发明价值的核心内容。专利转让是商业价值的代表性指标^[15]。一般而言,只有当一项发明专利具有潜在的经济价值时,才会有人从专利权人手中购买它,即发生专利转让的行为。本研究将专利转让次数作为商业价值的代理变量。

(2) 处理变量

根据前文提出的研究问题,本文的处理变量主要有两个:

①科学知识输入与否。实证研究表明,专利中的非专利文献引用不仅构成了衡量科学活动的重要且稳健的指标,更为探究发明过程中的科学知识提供了大规模分析基础^[4]。基于此,学界普遍将专利对科学文献的引用作为发明过程中科学知识输入的代理变量。本文的科学知识输入是一个二值变量,即当发明专利至少引用了一篇科学文献时,编码为 1,否则编码为 0。

②科学知识输入强度。学术界在评估专利输入

的科学知识强度时,形成了两种主流量化方法:一是采用专利引用科学文献的绝对频次^[16];二是考察科学文献引用在总引用中的相对比重^[4]。这两种方法都隐含着—个基本假定:即专利引用的科学文献越多,科学知识的输入强度越大。然而,这种仅关注引用数量的单维度评估方法存在明显局限,难以准确刻画专利采纳科学知识的实际情况。在科学研究中,不同科学文献的学术影响力存在显著差异,其对技术创新的贡献程度也不尽相同。例如,某项专利虽然引用了大量科学文献,但如果这些文献的学术影响力普遍较低,其获取的实质性科学知识可能有限;相反,另一项专利即使引用的文献数量较少,但若引用的都是领域内的突破性成果,科学知识的贡献可能更为突出。基于这一认识,本文借鉴科学计量学中 H 指数的测度思想,构建了专利科学引用 H 指数 (Patent Scientific Citation h-index, PSH),以评估专利中的科学知识输入强度。该指标通过将引用数量与引用质量有机结合,既考虑了专利引用科学文献的广度,又体现了被引文献的学术贡献,从而更准确地评估专利对科学知识的实际吸收程度。

具体而言,PSH 指数的测算遵循如下过程(如图 2 所示):识别焦点专利引用的全部科学文献,并且查找出科学文献的被引频次;按照被引科学文献的被引频次进行降序排序;确定 PSH 值,即排序第 h 位的科学文献的被引频次大于或等于 h 。

本文对 463,393 件发明专利的 PSH 指数进行了统计分析。数据显示,PSH 指数的整体分布范围为 0~250,其中 98% 的专利集中在 PSH 指数 ≤ 50 的区间。图 3 展示了 PSH 指数不超过 50 的发明专利分布情况。其中大约 3.5 万件专利的 PSH 指数为 0,表明这些专利虽然引用了科学文献,但是所引文献并未获得任何引用。PSH 指数为 1 的专利数量最多,达 13 万余件。

(3) 协变量

根据协变量选取原则,参考现有文献,本研究将以下可能的混淆因素作为协变量纳入模型:技术复杂性、技术新颖性、技术成熟度、技术累积速度、技术多样性、专利范围、PCT 专利、发明人科学发明经验、发明人技术知识广度、发明人数量、申请人数量、申请人科学发明经验、后向专利引用数量^[17,18]。同时控制专利申请人组织类型、时间、地理位置和技术领域等 4 组虚拟变量。本研究的所有变量名称、变量符号及变量含义见表 1。

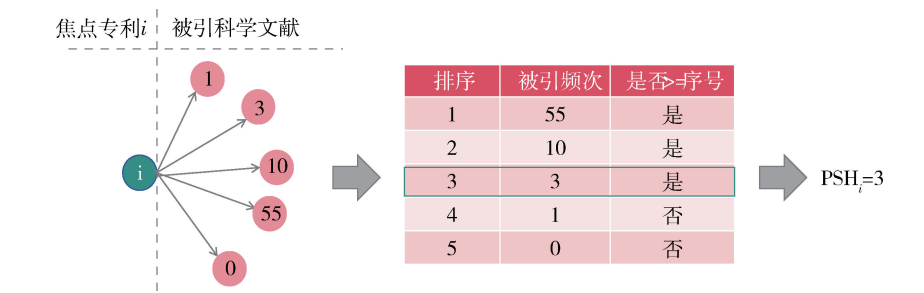


图 2 PSH 指数计算过程示意图
Figure 2 PSH index calculation process

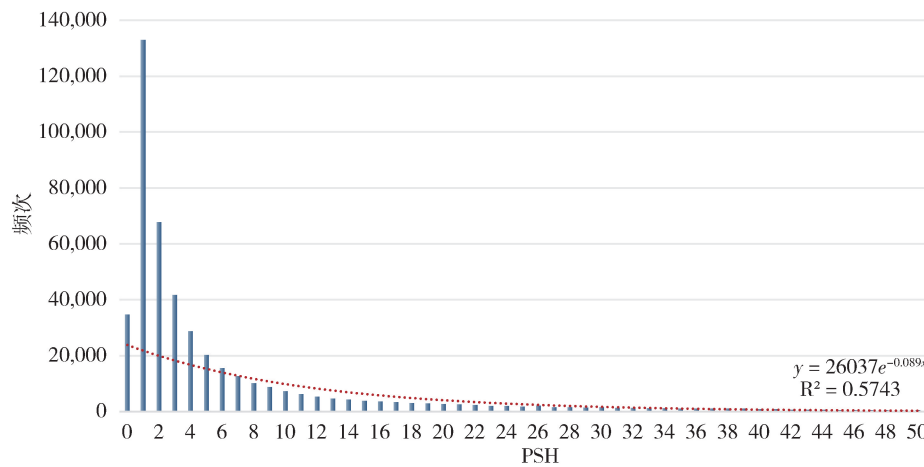


图 3 发明专利 PSH 指数分布 (≤50)
Figure 3 Patent PSH index distribution

2.2 描述统计与相关性分析

表 2 展示了本研究主要变量的描述性统计结果。其中,前向引用 (Citations_5) 的平均值为 6.902,最大值为 2,348。技术通用性 (Patent_Generality) 最小值为 0,最大值为 0.976。专利转让次数 (Num_Assignment) 最小值为 0,均值为 0.475。此外,科学知识输入 (Science_Input) 均值为 0.275,科学知识输入强度 (Science_Intensity) 的均值则为 7.004。

本研究还对各变量之间进行了 Spearman 相关性分析。结果显示,相关性系数的最大值为 0.655,并未超过 0.7 的临界值。除此之外,各变量的平均方差膨胀因子 VIF 值为 1.24,最大值为 2.18,未超过 10 的临界值,说明变量间并不存在多重共线性。由于篇幅的限制,在此不展示具体的相关性分析结果和 VIF 值。

3 实证结果

3.1 科学知识输入与否对发明价值的影响效应

根据 PSM 匹配方法步骤,本研究首先使用 Logit 模型估计倾向得分,继而采用 1:2 最近邻匹配算法将输入科学知识的处理组与未输入科学知识的控制组样本进行匹配。然后,进行平衡性检验以验证匹配结果的质量。如表 3 的所示,与匹配前相比,匹配后所有协变量的标准化偏差的绝对值均小于 10%,这意味着匹配后各变量在处理组与控制组之间是均衡的。除此之外,图 4 的核密度函数表明,匹配后两个组的分布更为接近,表明 PSM 匹配有效地消除了处理组和控制组之间的大部分差异。最后,PSM 还应满足共同支撑假设。图 5 显示了处理组与控制组的倾向得分分布。大多数观测值都在共同支撑范围内 (即处理组与控制组的倾向得分值存在重叠),在进行 PSM 匹配时仅丢失 609 个观察样本。

表 1 变量名称、变量符号及变量含义
Table 1 Variable names, variable symbols, and variable meanings

变量名称	变量符号	变量含义
结果变量		
前向引用	Citations_5	焦点专利自授权日起五年内获得的前向引用次数
技术通用性	Patent_Generality	根据 Trajtenberg 等提出的技术通用性指数计算 ^[14] ,取值范围为 0~1
专利转让次数	Num_Assignment	焦点专利被转让的次数
处理变量		
科学知识输入与否	Science_Input	虚拟变量,如果一项焦点专利引用过科学文献,则设置为 1,否则设置为 0
科学知识输入强度	Science_Intensity	焦点专利的 PSH 指数
协变量		
技术复杂性	Complexity	根据 Trajtenberg 等提出的原创性指数计算的结果 ^[14] ,取值范围为 0~1 ^① 。计算公式:
		$Originality_p = 1 - \sum_k^{N_p} \left(\frac{Nreferences_{pk}}{Nreferences_p} \right)^2$
技术新颖性	Novelty	根据 Squicciarini 等改进的激进性指数计算的结果 ^[19] ,取值范围为 0~1 ^② 。计算公式:
		$Radicalness_p = \sum_j^{n_p} \frac{CT_j}{n_p}; IPC_{pj} \neq IPC_p$
技术成熟度	Maturity	焦点专利与后向引用专利的平均时间间隔。
技术累积速度	Tech_Cum_Speed	焦点专利主 IPC4 分类号在五年内的平均发展变化程度 ^③ 。计算公式:
		$Tech_Cum_Speed_p = \sqrt[5]{\frac{n_p}{n_5}} * 100\%$
技术多样性	Tech_Diversity	焦点专利所属的 IPC 分类号数量
专利范围	Patent_Scope	焦点专利的权利要求数量
PCT 专利	PCT	虚拟变量。如果焦点发明为 PCT 申请专利则设置为 1,否则设置为 0
发明人科学发明经验	Inventor_Sci_Expe	焦点专利的发明人团队过去五年中已授权的专利中引用过科学文献的专利占其授权专利数量比值的最大值
发明人技术知识广度	Inventor_Tech_Bread	焦点专利的发明人团队过去五年授权专利所属不同 IPC4 分类号数量
发明人数量	Num_Inventors	焦点专利的发明人数量
申请人数量	Num_Applicants	焦点专利的专利权人数量
申请人科学发明经验	Num_Applicants_Sci	专利权人在焦点专利申请前就已授权的发明专利中引用过科学文献的发明专利数量
后向专利引用数量	Num_PC	焦点专利参考的专利文献数量
申请人组织类型	OrgTypes	虚拟变量。根据组织类型设置个人、公司、政府、大学、医院 5 个虚拟变量
授权时间	Year	虚拟变量,根据授权年份设置
地理位置	Country	虚拟变量,根据焦点专利第一个发明人所属国家设置
技术领域	IPC	虚拟变量,根据焦点专利主 IPC 分类号所属的“部”设置

① 公式中 k 表示某个技术领域, N_p 是专利 p 引用的所有专利参考文献中所属不同技术领域的总数。技术领域则使用 IPC4 表示。 $Nreferences_p$ 即专利 p 引用的专利参考文献数量。 $Nreferences_{pk}$ 则表征专利 p 引用的属于 k 技术领域的参考文献数量。

② 公式中 n_p 表示的是焦点专利 p 引用的所有专利文献中 IPC 分类号的总数。 CT_j 则代表被引用的专利 j 中与焦点专利 p 所属 IPC4 分类号不同的数量。

③ 公式中 n_p 是指焦点专利 p 的主 IPC4 分类号在其申请年的累积数。 n_5 则表明 IPC4 分类号在 5 年前的累积数。

表 2 主要变量的描述统计结果

Table 2 Descriptive statistics of main variables

变量名称	样本量	均值	标准差	最小值	最大值
Citations_5	1,685,970	6.902	16.355	0	2,348
Patent_Generality	1,496,992	0.530	0.281	0	0.976
Num_Assignment	1,685,970	0.475	0.899	0	21
Science_Input	1,685,970	0.275	0.446	0	1
Science_Intensity	463,393	7.004	11.822	0	250
Complexity	1,685,970	0.477	0.284	0	0.975
Novelty	1,632,772	0.474	0.341	0	1
Maturity	1,660,462	3,359.372	3,232.240	0	638,550
Tech_Cum_Speed	1,682,142	1.047	0.093	0.574	2.887
Tech_Diversity	1,685,970	1.174	0.526	1	23
Patent_Scope	1,685,970	17.763	14.170	1	887
PCT	1,685,970	0.100	0.300	0	1
Inventor_Sci_Expe	1,685,970	0.312	0.324	0	1
Inventor_Tech_Bread	1,685,970	5.734	8.076	0	148
Num_Inventors	1,685,970	2.510	1.804	0	76
Num_Applicants_Sci	1,530,794	2,188.774	4,688.280	0	45,899
Num_Applicants	1,530,794	1.034	0.213	1	14
Num_PC	1,685,970	22.541	46.978	0	2,399

总体而言,本研究使用的匹配算法实现了处理 的平均处理效应的分析。
组与控制组之间特征的良好匹配,可以进行下一步

表 3 匹配变量平衡性检验结果

Table 3 Results of balance test for matching variables

变量名称	样本	均值		标准偏差(%)
		处理组	控制组	
Complexity	匹配前	0.510	0.454	20.1
	匹配后	0.510	0.506	1.3
Novelty	匹配前	0.496	0.463	9.7
	匹配后	0.496	0.498	-0.7
LnMaturity	匹配前	7.574	7.778	-25.6
	匹配后	7.574	7.551	2.9
Tech_Cum_Speed	匹配前	1.058	1.048	9.7
	匹配后	1.058	1.063	-5.6
LnTech_Diversity	匹配前	0.145	0.084	20.3
	匹配后	0.145	0.137	2.5
LnPatent_Scope	匹配前	2.820	2.610	27.5
	匹配后	2.820	2.842	-3.0
PCT	匹配前	0.095	0.095	-0.2
	匹配后	0.095	0.098	-1.1

续表 3

变量名称	样本	均值		标准偏差(%)
		处理组	控制组	
Inventor_Sci_Expe	匹配前	0.514	0.250	85.0
	匹配后	0.514	0.509	1.4
LnInventor_Tech_Bread	匹配前	1.628	1.495	14.8
	匹配后	1.628	1.591	4.1
LnNum_Inventors	匹配前	0.893	0.710	29.4
	匹配后	0.893	0.869	3.9
LnNum_Applicants_Sci	匹配前	5.238	4.789	14.8
	匹配后	5.237	5.275	-1.3
LnNum_Applicants	匹配前	0.025	0.020	4.3
	匹配后	0.025	0.023	1.9
LnNum_PC	匹配前	2.898	2.517	37.3
	匹配后	2.898	2.859	3.7

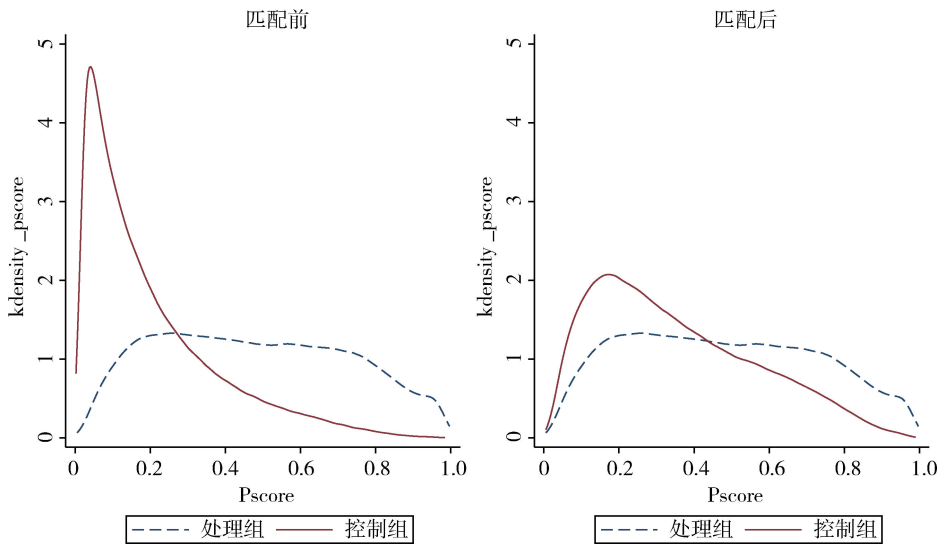


图 4 匹配前后的核密度函数图对比

Figure 4 Comparison of kernel density function before and after matching

在完成样本匹配后,本研究通过考察处理组与控制组在技术价值指标(前向引用次数、技术通用性)和商业价值指标(专利转让次数)维度的平均处理效应(ATT),考察科学知识输入对发明价值的因果效应。结果如表 4 所示。

由表 4 可知,在技术价值维度,前向引用(Citation_5)和技术通用性(Patent_Generality)匹配后的 ATT 值分别为 0.542 和 0.006,且均在 1% 的水平上显著。这表明科学知识输入能够显著提升发明专利技术影响力和技术影响力的范围。具体而言,与控

制组相比,具有科学知识输入的专利平均多获得 0.542 次引用,技术通用性则平均高出 0.006。在商业价值维度,专利转让次数(Num_Assignment)匹配后的 ATT 为 0.025,同样在 1% 水平上显著,表明科学知识输入显著促进专利的市场转化。值得注意的是,匹配前三个变量的 ATT 值均高于匹配后的 ATT 值。这说明若不控制样本的自选择性偏差,会高估科学知识输入对发明价值实际的因果效应。

总而言之,建立在科学知识基础上的专利在技术价值和商业价值方面均表现出明显的优势。基于

科学的发明专利更可能成为后续技术创新的基础，被商业化交易的潜力也更大。并产生更广泛的技术影响。与此同时，专利在后续

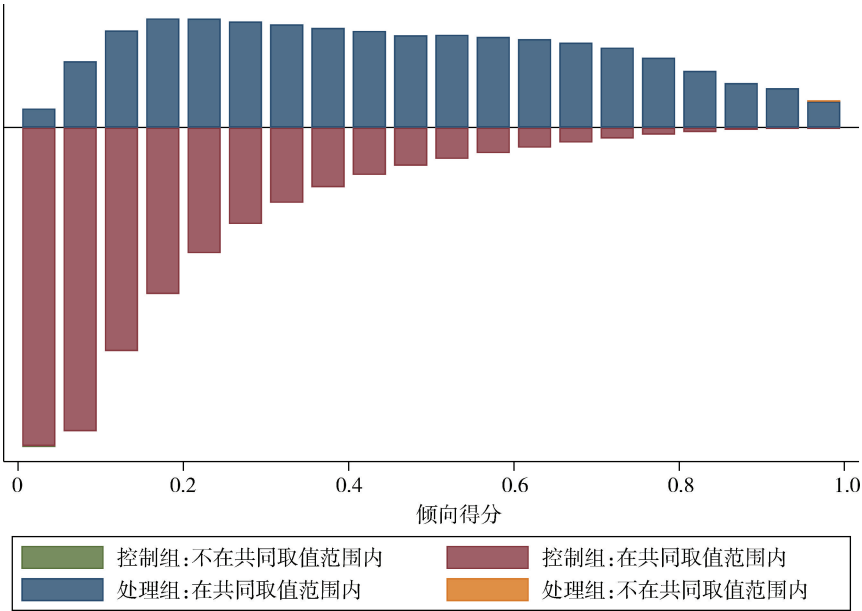


图 5 倾向得分共同取值范围
Figure 5 Common range of propensity score

表 4 平均处理效应 (ATT) 结果
Table 4 Average treatment effect (ATT) results

变量名称	样本	处理组	控制组	ATT	T 值
Citation_5	匹配前	10.765	6.927	3.839	112.29***
	匹配后	10.765	10.223	0.542	7.77***
Patent_Generality	匹配前	0.559	0.516	0.043	78.6***
	匹配后	0.559	0.553	0.006	6.27***
Is_Assignment	匹配前	0.378	0.297	0.080	88.62***
	匹配后	0.378	0.353	0.025	14.62***

3.2 科学知识输入强度对发明价值的影响效应

本小节使用 GPSM 进一步探究科学知识输入强度对发明价值的影响效应。遵循 GPSM 分析步骤，本研究首先对处理变量“科学知识输入强度”进行标准化处理后，并使用 Fractional Logit 模型估计其条件概率密度。在此基础上，根据给定的科学知识输入强度以及 GPS，分别估计表征技术价值与商业价值指标的条件期望。具体而言，将前向引用、技术通用性以及专利转让次数作为被解释变量，将科学知识输入强度和 GPS 作为解释变量代入公式 5。结果如表 5 所示，GPS 估计系数均与零有显著差异，表明协变量设定适当，未引入估计偏差。

接着，在 GPSM 的第三个步骤中，本研究以 0.1

为步长，将科学知识强度划分为 10 个子区间。将回归系数代入公式 6，计算各区间对应的科学知识输入强度与发明价值的平均剂量反应，进而构建剂量反应函数图，如图 6 所示。

图 6(a) 和 (c) 显示，科学知识输入强度与前向引用、专利转让均呈现相似的非线性关系：在科学知识输入强度较低区间 (0 ~ 0.2)，随着知识强度的提升，这两个变量均呈现显著的上升趋势；在中等强度区间 (0.3 ~ 0.7)，两个变量均处于下降趋势；在高强度区间 (> 0.8)，两个变量则又呈现显著上升趋势。这种“先升 - 后降 - 再升”的非线性特征表明，科学知识输入强度对专利价值的影响存在明显的阶段性。低科学知识输入强度与高科学知识输入强度

段,专利表现出更高的技术影响力,同时也展现出更大的市场交易潜力。

表 5 给定科学知识输入强度和 GPS 情况下的结果变量条件分布参数估计

Table 5 Conditional distribution parameter estimation of result variables given scientific knowledge input intensity and GPS conditions

变量名称	Citation_5	Patent_Generality	Num_Assignment
K	28.768 *** (1.237)	0.064 *** (0015)	0.242 *** (0.054)
K2	-86.369 *** (4.192)	-0.148 *** (0.053)	-1.436 *** (0.183)
K3	64.880 *** (3.382)	0.179 *** (0.043)	1.110 *** (0.148)
GPS	74.467 *** (2.339)	1.668 *** (0.029)	3.235 *** (0.102)
GPS2	-338.529 *** (12.129)	-6.684 *** (0.154)	-17.528 *** (0.529)
GPS3	384.357 *** (16.043)	7.672 *** (0.206)	21.470 *** (0.700)
K * GPS	-17.147 *** (2.952)	-0.123 *** (0.038)	0.607 *** (0.129)
_cons	5.505 *** (0.106)	0.473 *** (0.001)	0.474 *** (0.005)

图 6(b)展示了科学知识输入强度与技术通用性的关系。可以发现,科学知识输入强度处于 0 ~ 0.6 区间时,技术通用性呈现缓慢增长趋势;当该强

度超过 0.6 的阈值后,技术通用性表现出急剧上升趋势。也就是说,建立在高水平科学知识基础上的专利更可能带来通用的解决方案。

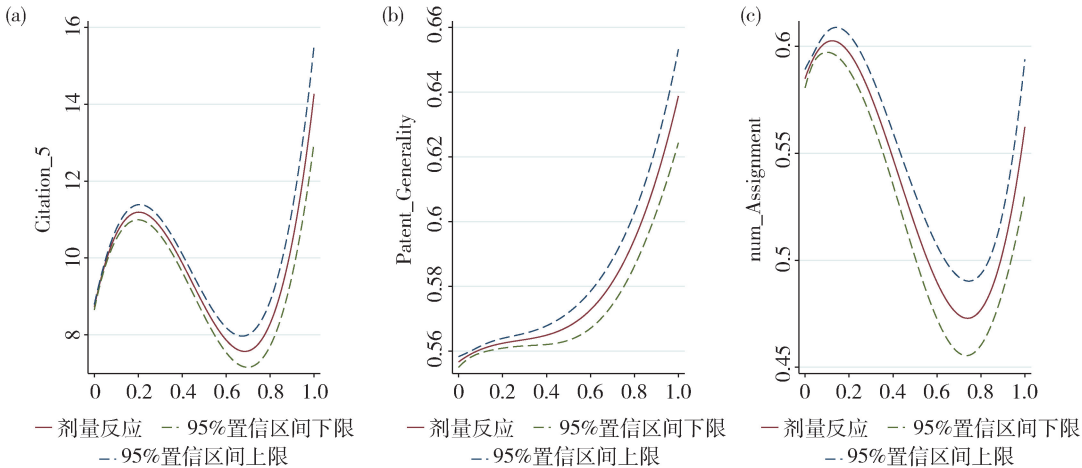


图 6 3 个结果变量的剂量反应函数图

Figure 6 Dose - response function plots a for three outcome variables

3.3 稳健性检验

为了验证实证结果的稳健性,本文借鉴 Benja-min 等人做法^[47],采用分层随机抽样方法从原始数据集抽取 30% 的子样本,并基于该子样本重复进行

PSM 和 GPSM 的估计。如表 6 所示,子样本 PSM 估计结果的符号与显著性水平与全样本保持一致;同时,图 7 所呈现的剂量反应函数图的趋势也与主分析结果一致。由此可见,本研究的分析结果是可靠

和稳健的。

表 6 减少样本后的 PSM 稳健性检验结果
Table 6 Results of PSM robustness test after sample reduction

变量名称	样本	处理组	控制组	ATT	T 值
Citation_5	匹配前	10. 710	6. 934	3. 775	60. 14 ***
	匹配后	10. 710	10. 280	0. 430	3. 38 ***
Patent_Generality	匹配前	0. 557	0. 516	0. 041	41. 16 ***
	匹配后	0. 557	0. 550	0. 007	3. 85 ***
Num_Assignment	匹配前	0. 613	0. 465	0. 148	45. 42 ***
	匹配后	0. 613	0. 568	0. 045	7. 21 ***

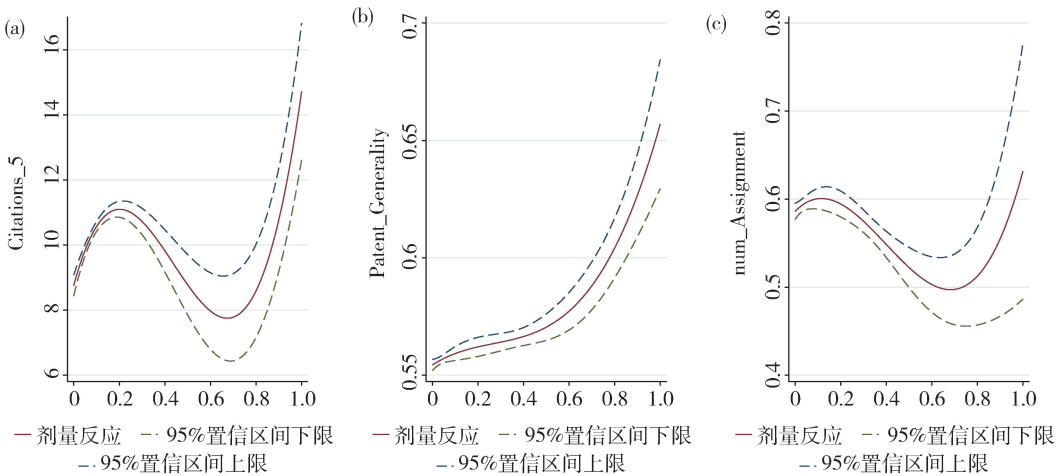


图 7 3 个结果变量的剂量反应函数图(30% 样本)
Figure 7 Dose – response function plots for three outcome variables (30% sample)

4 结论与展望

本研究基于因果推断的反事实框架,利用倾向得分匹配(PSM)与广义倾向得分匹配(GPSM)方法分别验证了科学知识输入与否以及科学知识输入强度对发明专利技术价值与商业价值的因果效应。研究发现:

- (1)科学知识输入对发明专利的技术价值以及商业价值均具有正向的因果效应。
- 首先,本研究从因果推断的角度为科学知识对前向引用的影响提供了额外的证据。研究发现,相比未涉及任何科学知识组件的专利,输入科学知识的专利更有可能成为后续创新的基础,尽管两者之间的差异很小。
- 其次,不同于既有研究对 GPT 影响因素的探

讨,本研究从科学知识输入的角度考察了其在 GPT 形成中的作用。研究发现,科学知识输入显著提升了发明专利的技术通用性水平。也就是说,科学知识的输入,有助于形成具有更广泛影响力的 GPT 发明专利。先前的研究表明^[21],开发 GPT 的可能性是在科学与技术层面发生的知识重组机制的一个复杂而微妙的函数。并且,对科学正确和全面的理解能够赋予科学家有效搜索和重组来自更遥远技术领域知识的能力,从而开发出更通用的技术。这也许能够解释为什么科学知识能够成为 GPT 的驱动因素之一。

第三,本研究的发现似乎挑战了资源基础理论中关于私人价值的创造需要依托不可复制性和不易转移性的特殊资源的观点。虽然科学知识的公共物品特性表面上与私人价值创造的逻辑相悖,但本研究的实证结果表明,与未输入科学知识的发明相比,

输入科学知识的发明被转让的平均次数更高。也就是说,建立在科学基础上的发明更有可能获得私人价值。对此,Arora 等提供了一个可能的解释^[22]:科学影响技术市场的主要渠道并非是降低了转让成本,而是因为成熟的科学更有可能为市场参与者提供共同语言或者背景知识。

(2)本研究将 H 指数引入,建立了兼顾数量与质量维度的专利科学引用的 PSH 指数,以此评估科学知识输入强度。基于 GPSM 方法,本研究揭示了科学知识输入强度与发明价值之间的非线性关系。

①随着 PSH 指数的递增,发明专利的前向引用和转让次数呈现“先上升-后下降-再上升”的非线性变化模式。建立在低强度与高强度科学知识基础上的发明专利表现出更高技术影响和市场交易潜力。在低 PSH 区间,适度科学知识的输入能够提升发明的技术基础,提升其创新性和应用潜力。这种程度的科学知识输入既能保持技术的先进性,又不会显著偏离市场主流认知,因此能够获得更高的技术影响力(即前向引用次数的增加)和市场接受度(即转让次数的上升)。当进入中等 PSH 区间时,科学知识投入的边际效应开始转负。正如 Ardito 和 Svensson 的研究所强调的,过度依赖科学知识可能使技术方案偏离用户的认知范畴和使用习惯,从而导致潜在使用者因其过于陌生而产生认知距离,进而降低技术的影响力和市场接受度^[23]。在高 PSH 区间,尽管技术方案可能依然具有较高的认知门槛,但丰富和高质量的科学知识积累可能能够突破关键技术的瓶颈,带来突破性创新^[24],使其最终能够获得较高的技术影响力和市场价值。

②就技术通用性而言,研究发现,在低 PSH 指数区间,专利的技术通用性随着 PSH 指数的增加而缓慢提升;而当 PSH 指数超过一定阈值时,技术通用性显著提高。也就是说,丰富和高质量的科学知识输入对通用技术的形成具有显著的影响。知识整合是开发 GPT 的基础^[25]。较高 PSH 指数的专利往往包含着更多基础性的科学原理,这些原理的普适性特征有助于技术在不同应用场景中的延展与扩散^[26]。

本文存在一定的局限性。首先,实证分析的数据来源仅限于 USPTO 授权的发明专利。鉴于各国专利制度的差异性,未来研究应将分析范围扩展至其他主要专利局,尤其是欧洲专利局(EPO)、日本专利局(JPO)、韩国专利局(KIPO)和中国专利局

(CNIPA)等世界主要专利机构,以验证研究结论的普适性。其次,尽管本文在测度科学知识输入时综合考虑了数量与质量维度,但发明人引用科学文献的实际动机和使用程度仍难以准确识别。未来研究可从文献内容特征、引用频率分布以及引用上下文等多维度深入分析专利对科学知识的实质性采纳程度。最后,本文基于全领域专利数据的分析可能掩盖了不同技术领域对科学知识依赖程度的异质性,后续研究应针对细分技术领域展开对比分析,以揭示科学知识输入效应的领域特异性。

参考文献:

- [1] Fleming L, Sorenson O. Science as a map in technological search[J]. Strategic Management Journal, 2004, 25(8-9): 909-928.
- [2] Hohberger J. Diffusion of science-based inventions[J]. Technological Forecasting and Social Change, 2016, 104: 66-77.
- [3] Sapsalis E, van Pottelsberghe de la Potterie B. The institutional sources of knowledge and the value of academic patents[J]. Econ. Innov. New Techn., 2007, 16(2): 139-157.
- [4] 曾德明,赵文静,文金艳. 外部科学知识获取与企业技术创新——桥梁科学家的调节作用[J]. 中国科技论坛, 2020(5): 109-117. Zeng D M, Zhao W J, Wen J Y. External scientific knowledge acquisition and enterprise technological innovation ——The regulatory role of bridging scientists[J]. Forum on Science and Technology in China, 2020(5): 109-117.
- [5] Barbieri N, Marzucchi A, Rizzo U. Knowledge sources and impacts on subsequent inventions: Do green technologies differ from non-green ones? [J]. Research Policy, 2020, 49(2): 103901.
- [6] 荣雪云,杨中楷,徐鑫. 发明者特征,技术新颖性和发明质量关系研究[J]. 科学学与科学技术管理, 2020, 41(10): 93. Rong X Y, Yang Z K, Xu X. The relationship research on inventor features, technological novelty and invention quality[J]. Science of Science and Management of S. & T, 2020, 41(10): 93.
- [7] Hume D. An Enquiry Concerning Human Understanding and Selections from a Treatise of Human Nature: With Hume's Autobiography and a Letter from Adam Smith [M]. Open Court Publishing Company, 1907.
- [8] Rosenbaum P R, Rubin D B. The central role of the propensity score in observational studies for causal effects[J]. Biometrika, 1983, 70(1): 41-55.

- [9] Hirano K, Imbens G W. The propensity score with continuous treatments[J]. *Applied Bayesian Modeling and Causal Inference from Incomplete – data Perspectives*, 2004, 226164.
- [10] Wang F. Does the recombination of distant scientific knowledge generate valuable inventions? An analysis of pharmaceutical patents [J]. *Technovation*, 2024, 130: 102947.
- [11] Marx M, Fuegi A. Reliance on science: Worldwide front - page patent citations to scientific articles[J]. *Strategic Management Journal*, 2020, 41 (9): 1572 – 1594.
- [12] 杨思思,戴磊,郝屹. 专利经济价值度通用评估方法研究[J]. *情报学报*, 2018, 37(1): 52 – 60. Yang S S, Dai L, Hao Y. Study on the common evaluation methodology of patented economic value[J]. *Journal of the China Society for Scientific and Technical Information*, 2018, 37(1): 52 – 60.
- [13] Ardito L, Natalicchio A, Appio F P, et al. The role of scientific knowledge within inventing teams and the moderating effects of team internationalization and team experience: Empirical tests into the aerospace sector[J]. *Journal of Business Research*, 2021, 128: 701 – 710.
- [14] Trajtenberg M, Henderson R, Jaffe A. University versus corporate patents: A window on the basicness of invention[J]. *Economics of Innovation and New Technology*, 1997, 5(1): 19 – 50.
- [15] Liu W, Tao Y, Bi K. Capturing information on global knowledge flows from patent transfers: An empirical study using USPTO patents [J]. *Research Policy*, 2022, 51(5): 104509.
- [16] Petruzzelli A M, Rotolo D, Albino V. Determinants of patent citations in biotechnology: An analysis of patent influence across the industrial and organizational boundaries[J]. *Technological Forecasting and Social Change*, 2015, 91: 208 – 221.
- [17] 于英杰,曾德明,赵胜超,等. 跨领域科学搜索、网络位置与企业发明质量[J]. *科学学研究*, 2023, 41(1): 123 – 133. Yu Y J, Zeng D M, Zhao S C, et al. Cross – field scientific search, network location and firm invention quality [J]. *Studies in Science of Science*, 2023, 41(1): 123 – 133.
- [18] Zhu K, Malhotra S, Li Y. Technological diversity of patent applications and decision pendency[J]. *Research Policy*, 2022, 51(1): 104364.
- [19] Squicciarini M, Dernis H, Criscuolo C. Measuring patent quality: Indicators of technological and economic value[R]. *OECD Science, Technology and Industry Working Papers*, 2013.
- [20] Benjamin D J, Berger J O, Johannesson M, et al. Re-define statistical significance[J]. *Nature Human Behaviour*, 2018, 2(1): 6 – 10.
- [21] Gruber M, Harhoff D, Hoisl K. Knowledge recombination across technological boundaries: Scientists vs. engineers [J]. *Management Science*, 2013, 59(4): 837 – 851.
- [22] Arora A, Belenzon S, Suh J. Science and the market for technology[J]. *Management Science*, 2022, 68(10): 7176 – 7201.
- [23] Arora A, Belenzon S, Pataconi A. The decline of science in corporate R&D[J]. *Strategic Management Journal*, 2018, 39(1): 3 – 32.
- [24] Arts S, Fleming L. Paradise of novelty—or loss of human capital? Exploring new fields and inventive output [J]. *Organization Science*, 2018, 29(6): 1074 – 1092.
- [25] Ardito L, Petruzzelli A M, Albino V. Investigating the antecedents of general purpose technologies: A patent perspective in the green energy field[J]. *Journal of Engineering and Technology Management*, 2016, 39: 81 – 100.
- [26] Bresnahan T F, Trajtenberg M. General purpose technologies ‘Engines of growth’? [J]. *Journal of Econometrics*, 1995, 65(1): 83 – 108.

Study on the impact of scientific knowledge input on the value of inventions

——A causal inference based on a counterfactual framework

OU Gui – yan¹, WU Jiang^{1,2,3}, YUE Ming – liang⁴, WANG Kai – li⁵

- (1. School of Information Management, Wuhan University, Wuhan 430072, China; 2. Center for E – commerce Research and Development, Wuhan University, Wuhan 430072, China; 3. Wuhan Insititue of Data Intelligence, Wuhan 430072, China; 4. Wuhan Library, Chinese Academy of Sciences, Wuhan 430071, China; National Engineering Research Center for E – Learning, Central China Normal University, Wuhan 430079, China)

Abstract: The relationship between scientific knowledge—a critical input to technological systems—and invention value remains a central concern in science policy and innovation management. While existing studies have primarily focused on correlational analyses, yielding inconsistent findings, the integration of scientific knowledge into technological systems is not random but intimately linked to knowledge recombination patterns and invention team characteristics. The inconsistency in research findings likely stems from endogeneity issues caused by selection bias, which traditional regression analyses fail to address. Causal inference methods are thus necessary to accurately estimate the net causal effect of scientific knowledge on invention value, thereby providing more reliable evidence for science and technology policymaking.

Using a counterfactual causal framework, we employed Propensity Score Matching (PSM) and Generalized Propensity Score Matching (GPSM) to estimate the causal effects of both the presence and intensity of scientific knowledge inputs on patents' technological and commercial value. Existing measures of scientific knowledge input intensity focus solely on citation counts, overlooking variations in scientific literature impact. We address this limitation by developing the Patent Scientific Citation h -index (PSH), inspired by the h -index methodology, which integrates both quantitative and qualitative dimensions of scientific citations to more accurately assess patents' scientific knowledge absorption. We measured patents' technological value using two indicators: forward citation counts and technological generality. Commercial value was quantified through the frequency of patent transfers. Analyzing 1,685,970 utility patents granted by the USPTO (2001–2010) and their 6,473,214–citation links to 1,407,439 scientific papers, we established the following findings:

(1) Our findings demonstrate positive causal effects of scientific knowledge input on both the technological and commercial value of patents. Patents incorporating scientific knowledge receive significantly more forward citations than those without such inputs. This indicates that scientific knowledge integration enhances a patent's capacity to serve as a foundation for subsequent innovations. Moreover, scientific knowledge input significantly enhances patents' technological generality. Specifically, we observe that scientific knowledge incorporation enables the development of more influential general-purpose technologies. The theoretical mechanism underlying this relationship lies in how comprehensive scientific understanding enhances researchers' cognitive capabilities to effectively identify, access, and recombine knowledge from distant technological domains. Finally, patents incorporating scientific knowledge demonstrate significantly higher transfer frequencies compared to those without such knowledge inputs. This pattern suggests that science-based inventions exhibit enhanced potential for private value realization, primarily because established scientific frameworks provide a shared cognitive platform and common language among market participants.

(2) A nonlinear relationship exists between scientific knowledge input intensity and inventive value. Empirical analysis reveals a fluctuating "rise–decline–rise" pattern in patents' forward citations and transfer frequencies as the PSH index increases. Patents with either low or high PSH intensities demonstrate superior technological impact and market transaction potential. In low-PSH domains, moderate scientific knowledge integration maintains technological advancement while aligning with market cognition. Mid-range PSH shows diminishing returns due to potential divergence from user expectations. High-PSH patents leverage rich scientific knowledge to overcome technical barriers, yielding enhanced technological impact and market value. Regarding technological generality, our findings reveal that patents characterized by higher PSH indices embody more fundamental scientific principles. The universal applicability of these underlying principles facilitates technology extension and diffusion across diverse application contexts, resulting in enhanced levels of technological generality.

This study makes several significant contributions to the innovation policy literature. First, we employ a causal inference framework to establish the causal relationship between scientific knowledge input and inventive value. Second, we advance methodological development by introducing a novel approach that incorporates the H -index to quantify scientific knowledge input intensity. These empirically grounded findings provide robust evidence for optimizing science and technology policies, bridging the gap between theoretical frameworks and practical policy implementation.

Key words: scientific knowledge; invention's value; propensity score matching (PSM); generalized propensity score matching (GPSM); counterfactual framework.