arXiv:2301.06374v1 [cs.DL] 16 Jan 2023

# 科学职业生涯中创新峰值动态的量化

**Mingtang Li**伦敦大学学院
计算机科学系 伦敦（英国）
mingtang.li.20@ucl.ac.uk

**Giacomo Livan**
伦敦大学学院计算机科学系 伦敦（英国）
g.livan@ucl.ac.uk

**Simone Righi**威尼斯卡福斯卡里大学经济学系 威尼斯（意大利）
simone.righi@unive.it

**摘要**

我们考察了在计算机科学和物理学领域拥有长期职业生涯的研究人员的创新表现。尽管这些学科在认识论上存在差异，我们始终发现，研究人员最具创新性的出版物出现得比预期更早——如果创新是在其职业生涯的出版物序列中随机分布的话；此外，研究人员还会经历一个峰值年份，在这一年里，他们发表的其他作品也比平均水平更具创新性。通过一系列线性模型，我们表明，若研究人员的峰值年份之前经历了较长一段生产率较低的时期，那么他们在峰值年份取得的创新成果会更高。这些发现与学术影响力的动态形成鲜明对比：当前盛行的科学评价范式通过奖励高生产率和渐进式、创新性较低的工作，激励研究人员追求学术影响力。

**关键词** 科学创新 · 科学生涯 · 科学影响力 · 创新

## 1 引言

科学专业的毕业生可能会选择投身学术职业生涯，希望为其研究领域带来创新。研究人员或许能凭借一篇出版物在其领域留下深远影响，但在当今的学术环境中，要建立长期的职业生涯，就必须证明自己有能力持续产出，并不断在知名的同行评审发表平台上发表作品。从这个角度看，要定量理解科学创新，归根结底需要识别一些规律，以揭示研究人员的创新能力如何随着其职业生涯的发展而变化，这种变化可由其出版物序列来刻画。创新能力大致保持不变，还是会在职业生涯的某个阶段达到峰值？如果会，哪些因素有助于科学生涯中创新能力较强的时期？

在当今竞争异常激烈的研究生态系统中，回答这些问题尤为重要。这个生态系统的特点是，出版物数量不断增加，彼此争夺关注 [1]。事实上，这样的环境会激励研究人员最大限度地提升其科学研究的影响力。一方面，科学影响力是一个多面向的概念，涵盖多个维度，其中包括——但不限于——科学出版物的合理性 [2, 3], 原创性 [4, 5], 科学价值 [6, 7, 8], 和社会价值 [9, 10, 11]。然而，另一方面，当前的学术评价实践大多将科学影响力具体化为文献计量影响力，即已发表科学研究从其他出版物获得的引文数量 [12, 13, 14]。反过来，这也导致基于引文的文献计量指标大量涌现，旨在量化论文或研究人员吸引引文的能力的不同方面 [15, 16, 17]。

大量文献表明，将科学影响力等同于文献计量影响力，会以至少两种对科学创新影响深远的方式左右研究人员的职业选择。首先，这会激励研究人员开展保守型研究，以便发表更“稳妥”的成果 [18, 19]。例如，众所周知，与旨在拓展成熟领域的渐进式研究相比，跨学科研究获得的引文相对较少 [20, 21]。这种研究方式或许有利于个别研究人员的职业发展，但从整体上看，它会严重抑制冒险 [22], ，而冒险对科学进步至关重要。其次，这会激励研究人员过度追求高生产率（即所谓的“不发表就出局”文化 [23]），导致科学研究失去可持续性 [24]。

文献计量影响力指标得到广泛应用，一个关键原因是它们易于使用和计算。而量化创新的度量指标则并非如此。创新研究历史悠久，例如经济学 [25, 26], 管理学 [27], 和人类学 [28], ，但创新指标的发展却是相对较新的工作。早期尝试量化科学出版物的创新性时，研究者试图判断论文参考文献列表的典型程度 [29]。然而，这类度量与跨学科性高度相关，而且往往低估界定明确的领域中的研究创新性 [30]。另一种称为颠覆性得分的度量方法 [31] 克服了这一问题，并已被证明能够可靠地区分创新性贡献与发展性贡献。其稳健性已通过科学出版物、专利和软件产品的数据得到验证 [31, 32]。

本文借助这一度量指标，旨在量化科学创新与影响力之间的张力。具体而言，我们提出假设：科学创新要求研究人员在较长时间内将努力投入特定项目或研究问题，也就是说，最终会导致较低的生产率。我们将通过追踪计算机科学和物理学领域一大批职业生涯漫长的研究人员的出版物序列，验证这一假设。首先，我们将证明，研究人员的职业生涯中存在一些特定阶段，他们在这些阶段的创新性持续高于职业生涯的其他时期。其次，我们将探讨研究人员在这些阶段到来之前的生产率与阶段内所达到的创新水平之间的关系。我们的结果将表明，研究人员会经历一个以创新性出版物为特征的“神奇的一年”，而在低生产率时期之后发表的此类出版物创新性更高。

## 2 结果

我们考察职业生涯始于1980年到2000年之间、且至少持续20年的计算机科学和物理学研究人员。在这些人中，我们仅保留发表论文至少10篇、且每5年至少发表一篇论文的研究人员（与 [33]一致）。因此，计算机科学和物理学研究人员分别有27,641名和34,527名（见方法）。

我们使用颠覆性得分 [32, 31],来量化研究人员的创新程度。该得分的特点是：如果后续发表的同一领域论文引用某篇论文的次数多于引用其参考文献中论文的次数，也就是说，该论文盖过了先前研究的关注度，那么这篇论文就被视为更具创新性（见方法）。我们计算研究人员群体中每篇论文的颠覆性得分，并将研究人员整个职业生涯所取得的创新描述为其出版物对应的颠覆性得分序列。

### 2.1 科学创新的非随机性

我们首先测量创新（以颠覆性得分量化）如何随着研究人员的职业生涯而变化。首先，我们找出研究人员发表其最具创新性论文的年份，即其颠覆性得分最高的出版物发表年份。下文将此年份称为“峰值年份”。借鉴Sinatra等人关于职业生涯中科学影响力随机性的分析 [34], ，我们在保留论文发表日期不变的同时，随机重新分配这些出版物对应的颠覆性得分，从而对研究人员的创新轨迹进行部分随机化。

我们分别根据原始数据和随机化数据，计算研究人员首次发表论文与其峰值年份之间相隔的年数。结果以直方图的形式呈现在图1中。在我们考察的两个学科中，原始数据和随机化数据所得的分布显著不同（$p < 0.01$ 计算机科学和物理学均如此，双侧柯尔莫哥洛夫-斯米尔诺夫检验）。我们还通过比较研究人员在峰值年份之前发表的论文数量分布，进一步验证了这一结果；这些分布也显著不同（两种情况下均为 $p < 0.01$，双侧柯尔莫哥洛夫-斯米尔诺夫检验），见附录 A 中的图4。

基于上述结果，我们可以得出结论：与峰值影响力不同，创新峰值并非完全随机地出现在研究人员的职业生涯中。另一个值得注意的现象是，平均而言，创新峰值出现的时间早于人们

根据随机化基准所预期的时间（$p < 0.01$ 对于计算机科学和物理学均如此，Mann-Whitney U 检验——MWU 检验）。

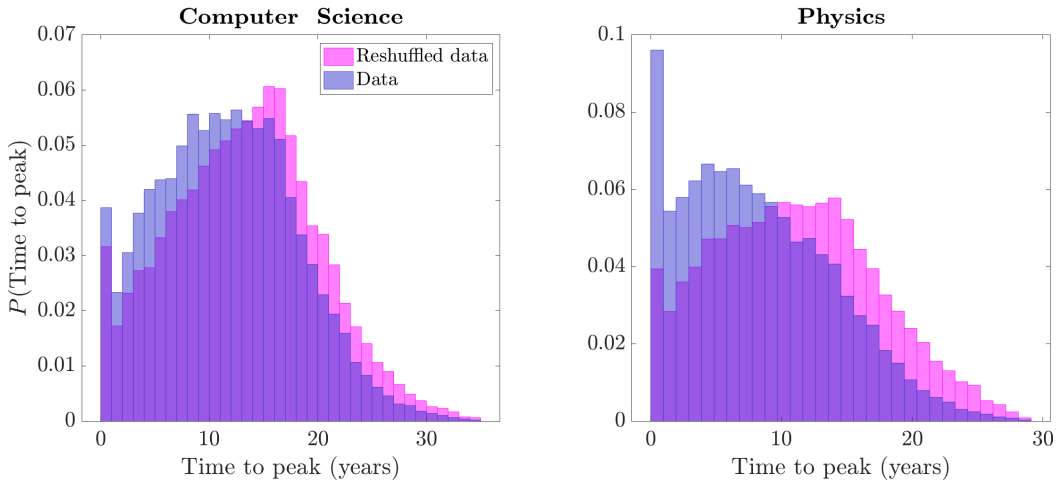


图1：根据原始数据（紫色）和随机化数据（粉色），展示计算机科学（左）和物理学（右）领域研究人员达到创新峰值所需时间（以年计）的归一化直方图。

### 2.2 表征创新峰值年份

我们现在着手确定研究人员的峰值年份是否呈现出独特的统计规律。具体而言，我们考察在峰值年份发表不止一篇论文的研究人员（即除其职业生涯中颠覆性得分最高的那篇论文外，至少还在峰值年份发表过一篇论文），并探究这些论文是否也具有更高的创新水平。我们假设，峰值年份通常以高度创新的研究为特征，而非仅仅是在较缺乏创新的论文中发表了一篇创新论文。

为了检验这一猜想，我们将研究人员的职业生涯划分为“峰值年份之前”（BPY）和“峰值年份之后”（APY）两个阶段（我们排除峰值年份恰好是其职业生涯第一年或最后一年的研究人员）。随后，我们计算样本中研究人员在这两个阶段的平均创新水平（BPY 阶段，计算机科学为 $N_{\mathrm{CS}} = 24{,}209$，物理学为 $N_{\mathrm{PHY}} = 31{,}139$；APY 阶段分别为 $N_{\mathrm{CS}} = 25{,}781$ 和 $N_{\mathrm{PHY}} = 33{,}465$）。接着，我们根据峰值年份发表的论文计算平均“峰值年份”（PY）创新水平，但不计入促成峰值的那篇论文（PY 阶段分别为 $N_{\mathrm{CS}} = 18{,}640$ 和 $N_{\mathrm{PHY}} = 26{,}543$）。

通过比较三组的创新水平，我们发现，峰值前年份和峰值后年份阶段的颠覆性得分分布与峰值年份的分布存在显著差异（两种情况下均为 $p < 0.01$，双侧柯尔莫戈洛夫-斯米尔诺夫检验）。更重要的是，我们发现，峰值年份的平均颠覆性得分高于另外两个阶段（两种情况下均为 $p < 0.01$，MWU 检验），见图2。

为减轻峰值前后阶段时长差异可能带来的偏差，我们将前述检验重复进行，并将峰值前年份和峰值后年份阶段设为 2 年（对于峰值前年份，$N_{\mathrm{CS}} = 20{,}980$ 和 $N_{\mathrm{PHY}} = 28{,}867$；对于峰值后年份，$N_{\mathrm{CS}} = 21{,}987$ 和 $N_{\mathrm{PHY}} = 30{,}792$），结果基本一致（计算机科学中峰值年份与峰值前年份之间的 MWU 检验结果为 $p = 0.021$ 的情况除外），见附录B中的图5。

### 2.3 峰值年份创新的决定因素

在确认存在“神奇的一年”——其特点是创新水平普遍较高——并确认它并非随机出现在研究人员的职业生涯中之后，我们现在要探究它与为达到这一峰值所付出的努力程度之间的关系。为此，我们将研究人员在峰值年份的整体创新水平量化为其在该年发表的论文的平均颠覆性得分，其中包括促成创新达到峰值的那篇论文。

我们拟合了一系列线性模型，以探究研究人员在峰值年份取得的创新水平，与一组旨在量化其为该年发表的科学研究投入的努力程度的变量之间的关系。我们旨在捕捉频繁发表的动机（即生产率）与产出创新性贡献的能力之间的权衡。我们预计，投入时间研究一篇论文越多，这篇论文

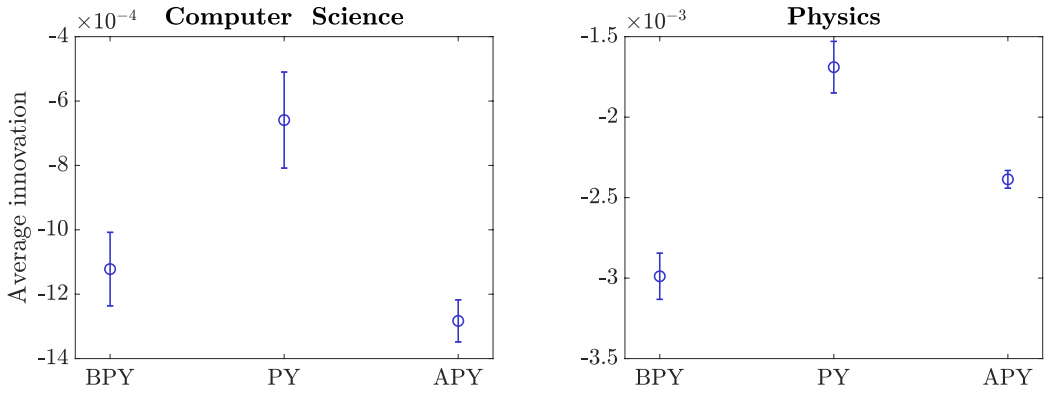


图2：研究人员在峰值年份之前发表的论文（峰值前年份，BPY）、峰值年份期间（不包括导致创新峰值的论文本身，PY）以及峰值年份之后（峰值后年份，APY）所取得的平均创新水平（颠覆性得分）。误差棒表示标准误。

就越有可能具有创新性。我们将努力程度量化为在某一感兴趣时期内，完成一篇已发表论文平均所花费的时间（以年计）。更具体地说，我们考察一个变量（下文称为“投入时间”）与研究人员在该时期内发表的论文数量（即其生产率）之比。前者定义为研究人员在感兴趣时期之前发表的最后一篇论文的年份，与该时期最后一年之间的年数。例如，某研究人员在 2005 年发表了 1 篇论文、2007 年发表了 3 篇、2008 年发表了 5 篇，那么其在 2007—2008 年的努力程度为 3/8。

我们还定义了相应的相对量，即将研究人员在某一感兴趣时期内的努力程度、投入时间和生产率分别除以其整个职业生涯中的相应量所得到的指标。在附录C的图6中，我们绘制了以峰值年份作为感兴趣时期时这些量的分布。下文仅呈现基于这些相对量得到的结果，因为基于相应绝对量得到的结果在定性上是等价的（见附录D表1—4）。

在三个不同的模型中，我们测量了研究人员在其峰值年份的创新与该年份相关的相对努力、相对生产率和相对投入时间之间的线性关系。在每个模型中，我们都控制了相同的其他因素。具体而言，我们纳入达到峰值年份所需的时间（以年为单位）以及与峰值年份对应的日历年份，以控制可能存在的时间效应。我们还纳入研究人员在峰值年份发表的论文中的合著者数量，以控制网络效应。最后，我们纳入峰值年份之前两年所发表论文的平均颠覆性得分，以控制创新能力中的个体特异性。作为稳健性检验，我们还将上述变量在峰值之前两年计算得到的值作为额外控制变量纳入模型。由这些模型得到的结果在定性上与本文呈现的主要结果一致（见附录D表5—8）。

上述模型的结果如图3所示。在计算机科学和物理学领域，相对努力程度都会对峰值年份实现的创新水平产生积极影响，相对投入时间也有同样的作用。相比之下，相对生产率被发现会损害峰值年份创新。总体而言，我们将这些结果解读为一个明确的信号：若在峰值年份之前的一段时间内专注于最终将在峰值年份发表的工作，创新峰值就会提高；而在峰值年份前后减少论文写作数量，则会带来更高的峰值创新能力。

先验来看，我们不能排除这样一种可能：以颠覆性得分衡量的创新，提供的信息与论文影响力相似。为了检验这一假设，我们使用相同的一组自变量，并以影响力作为因变量，对上述线性模型进行校准。我们以峰值年份论文在发表后前5年内获得的累计引用次数来衡量后者。我们发现，此类模型的解释力相当低。不过，它们仍然为我们揭示努力程度相关变量与科学影响力之间的相关性提供了重要见解。事实上，我们发现，较高的相对生产率会提升影响力，而相对努力程度和相对投入时间则会产生相反的作用（见图3）。在考察相应的绝对量时，也有类似发现（见附录D表1—8）。

我们模型中的控制变量也有助于解释科学创新峰值的决定因素。事实上，我们发现，在近些年达到峰值的研究人员，其创新水平平均较低；这与多项研究的发现一致，即随着时间推移，科学的创新性有所下降 [35, 36, 37, 38]。在

控制了这一影响后，我们观察到创新峰值与达到峰值所需的时间之间存在正相关关系。

我们的结果还表明，在这两个学科中，研究人员此前发表的论文越具创新性，他们在峰值年份也越具创新性。最后，我们观察到，与更多合著者合作发表论文会对计算机科学中的峰值年份创新水平产生积极影响，但在物理学中则不会。

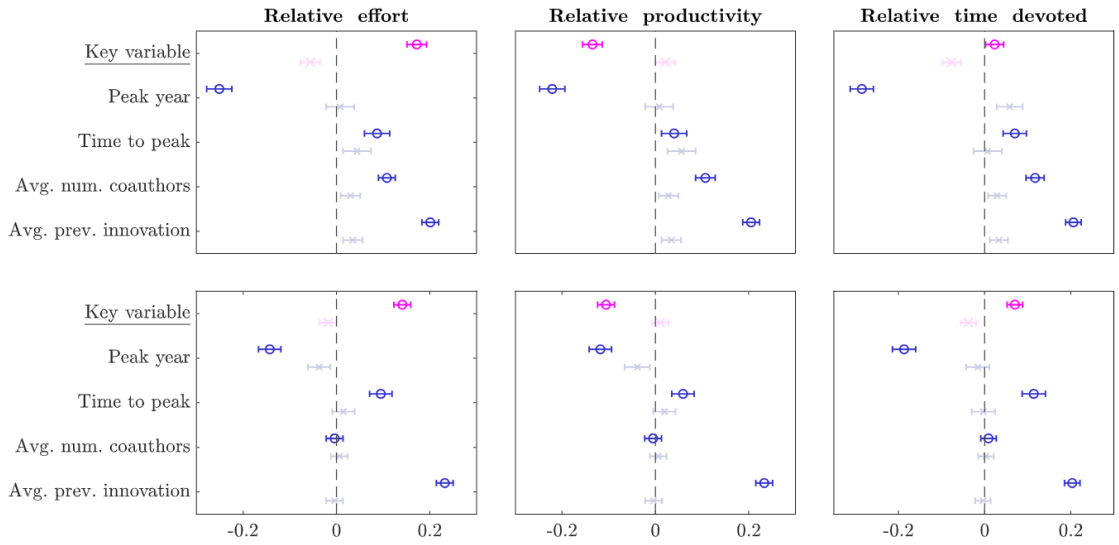


图3：峰值年份创新的线性回归模型系数（圆圈，亮色）和影响力的线性回归模型系数（叉号，暗色）。每列展示了采用本研究不同关键变量所得的结果，即相对努力程度（左）、相对生产率（中）和相对投入时间（右）。误差棒表示乘以三倍的标准误。上排（下排）模型对应计算机科学（物理学）。从左至右，计算机科学中创新模型的 $R^2$ 为 0。142、0。132 和 0。116，而影响力模型的 $R^2$ 为 0。009、0。007 和 0。011。在物理学中，创新模型的 $R^2$ 为 0。086、0。078 和 0。071，而影响力模型的 $R^2$ 为 0。001、0。001 和 0。002。详细的回归表请参见附录D表1–4。

## 3 讨论

我们从研究人员创新的角度考察了他们的职业生涯动态，并以颠覆性得分 [31, 32] 来定义创新。我们发现，创新峰值并非完全随机地出现在研究人员的职业生涯中，而且大多数研究人员达到峰值所需的时间因其学科而异。这与所谓的随机影响力规则 [34, 39]，形成对比，该规则认为，研究人员职业生涯中的每篇论文成为其影响力最大作品（即被引用次数最多的出版物）的可能性相同。具体而言，我们表明，科学家职业生涯中的创新峰值平均出现得比根据“随机创新规则”所预期的更早。所谓“随机创新规则”，是指一个零模型：将创新随机分配到研究人员职业生涯中的出版物序列上。

我们的发现与一些观察结果相符：年轻研究人员往往比资深研究人员更具创新性 [40,41]，例如，较具创新性的研究人员可能会被学术界淘汰 [42],，获得终身教职后创新激励减弱 [41],，或出现认知衰退 [43]。与此同时，我们的回归分析显示，在职业生涯后期出现的创新峰值往往更高。这表明，关于资深科学家往往较缺乏创新性的直觉需要加以限定。尽管研究人员在职业生涯后期达到创新峰值的情况比预期少见，但一旦出现，往往意味着更高的创新水平（在控制分析中纳入的协变量之后）。我们推测，这可能与科学领域创新整体衰退有关 [35, 36, 37, 38]——这一趋势在我们的模型中始终显现——而这或许为经验丰富的研究人员进行创新留下了更多空间。

关于职业生涯动态的研究表明，个人往往会经历一段段“高产期”，在此期间，他们的工作持续取得成功 [39, 44, 45]。我们的结果表明，科学创新也存在类似效应。事实上，我们发现，峰值年份的特点不仅是一篇极具创新性的论文，更是一系列

比平均水平更具创新性的论文。换言之，我们的研究发现，在峰值年份发表的论文会产生协同效应，也就是说，促成峰值的那篇论文并不会耗尽研究人员用于其他出版物的资源。

既然我们已经确定，生产率峰值并非随机出现，而是属于一段创新能力较高的特殊时期，那么探明峰值幅度的决定因素就十分重要。为此，我们建立了一系列回归模型，将创新峰值的幅度与多种指标进行回归分析，这些指标分别用于衡量个人在峰值年份论文中投入的努力程度，以及生产率。我们发现，在峰值年份的研究中投入更多时间和努力，与该时期所达到的创新水平呈正相关；而生产率指标则与之呈负相关。对此一种可能的解释是，同时撰写多篇论文可能会消耗研究人员的精力和想法。值得注意的是，将同一项研究的发现拆分成多篇论文发表，往往是一种有意采取的策略，通常称为“切香肠式发表” [46] ，目的是使预期引用量所体现的影响力最大化。

我们的模型充分揭示了创新与影响力之间的这种张力：这两个维度与上述努力程度和生产率指标呈现相反的相关性。当投入更多努力于科学研究时，创新便会蓬勃发展，也就是说，这种情况会出现在生产率较低的时期。相反，生产率低下不利于影响力。尽管对于后一项结果不应过度解读，因为我们的模型对影响力的解释变异较少，但这些发现有力地表明，科学创新与影响力遵循截然不同的模式，支持了 Aksnes等人 [5] 提出的观点：这两个概念确实应当加以区分。
研究人员经常仅根据文献计量影响力指标接受评估，例如在终身教职决策或晋升时 [47]。我们的分析表明，还应评估他们的创新能力，因为后者可能为其工作提供截然不同的视角。此外，我们的研究结果表明，职业生涯早期阶段的研究人员对科学领域的颠覆性创新至关重要，因此在科学政策制定中应强调他们的作用，例如在资源（如资助）分配和职业稳定性方面。

## 4 方法

### 4.1 数据

我们分别从 AMiner引文网络数据集（第12版）和 Web of Science 数据库收集计算机科学与物理学领域的出版物和引文数据。AMiner 数据集收录了来自 DBLP、ACM、MAG 以及计算机科学领域其他主要来源的 20 世纪 60 年代至 2020 年的论文 [48], , 共包含 4,894,081 篇论文及其间的 45,564,149 条引文。A Miner 数据集已被应用于多项文献计量学研究 [49, 50, 51]。
对于物理学论文，我们从 Web of Science（WOS）数据库中提取数据。我们提取了特定研究人员群体的出版物（见下文）及其出版物的引文网络。总体而言，我们收集了 1985 年至 2020 年间共 1,619,039 篇论文及其相互之间的 12,621,175 条引文。重要的是，WOS 不维护唯一的作者标识符。因此，为了将作者与其出版物关联起来，我们采用了 Caron 和 van Eck 提出的一种先进方法来消解作者姓名歧义 [52]。具体而言，该方法根据一系列属性计算作者两两之间的相似度得分，这些属性包括 ORCID标识符、姓名、所属机构、电子邮件、合著者、资助编号、学科类别、期刊、自引、书目耦合和共被引。两位作者之间的相似度得分越高，他们越有可能被视为同一人。近期一项研究表明，该方法优于其他无监督消歧方法 [53],，其精确率和召回率均超过 90%。

在我们的分析中，我们仅计算2016年之前发表的论文的颠覆性分数，从而确保样本中的论文至少有5年的时间积累引文。我们按照 [33], 的要求设定筛选标准，仅选择职业生涯较长的研究人员。入选的研究人员在1980年至2000年间开始职业生涯，职业生涯至少持续20年，发表至少10篇论文，并且每五年至少发表一篇论文。按照这些筛选标准，计算机科学和物理学领域分别有27,641名和34,527名研究人员符合条件。

### 4.2 颠覆性得分

为回答我们的研究问题，我们采用颠覆性得分来量化数据集中每篇论文的创新水平。该指标能够有效区分创新性出版物与发展性出版物，并且已通过科学论文、专利和软件产品的数据验证其稳健性 [31, 32]。这一指标的核心思想是，高度创新的论文会将注意力从同一领域的先前研究上转移开；也就是说，后续出版物往往会比引用其参考文献中的文献更多地引用该论文。

更具体地说，考虑一个围绕焦点论文、其参考文献（先前论文）和后续论文构建的引文网络。后续论文可分为三类：仅引用焦点论文的论文、同时引用焦点论文和先前论文的论文，以及仅引用先前论文的论文。假设每组中的论文数量分别为 $n_i$、 $n_j$ 和 $n_k$。那么，颠覆性得分的计算公式为

$$D = \frac{n_i - n_j}{n_i + n_j + n_k} \tag{1}$$

其中， $n_i - n_j$ 衡量焦点论文在多大程度上将注意力从先前论文转移开来，而$n_i + n_j + n_k$ 表示网络中的后续论文总数。

根据上述定义，颠覆性得分的范围为 -1 到 1。正分表明，与其参考文献相比，焦点论文吸引了后续论文更多的关注，这意味着焦点论文更具创新性。如果一篇焦点论文足够创新，那么其颠覆性 $D$ 应接近 1。同样，负分意味着焦点论文很可能是一项发展性研究成果。得分越接近 -1，论文的开发性就越强。因此，颠覆性得分不仅使我们能够量化每篇论文的创新水平，还能比较不同论文的创新水平。

## 致谢

G.L. 感谢 EPSRC 数字经济早期职业研究员奖学金（项目编号：EP/N006062/1）提供的支持。我们感谢 Ye Sun 协助处理 WOS 数据中的作者消歧。

## 参考文献

[1] Lutz Bornmann 和 Rüdiger Mutz。现代科学的增长率：基于出版物数量和被引参考文献数量的文献计量分析。《信息科学与技术协会期刊》，66(11):2215–2222，2015年。[2] Daniele Fanelli 和 Tom Tregenza。科学家中有多少人伪造和篡改研究？对调查数据进行的系统综述与荟萃分析。*PLoS ONE*，4:5738，2009年。[3] F. C. Fang、R. G. Steen 和 A. Casadevall。不端行为是撤稿的科学出版物的大多数原因。《美国国家科学院院刊》，109，2012年。[4] Michael Polanyi、John Ziman 和 Steve Fuller。科学共和国：其政治与经济理论。来源：*Minerva*，38:1–32，2000年。[5] Dag W. Aksnes、Liv Langfeldt 和 Paul Wouters。引文、引文指标与研究质量：基本概念和理论概述。*SAGE Open*，9，2019年。[6] B R Martin。多项指标在基础研究评估中的应用。36:343–362，1996年。[7] H. F. Moed。研究评价中的引文分析。Springer，多德雷赫特，2010年。[8] Ludo Waltman、Nees Jan Van Eck 和 Paul Wouters。统计出版物和引文：越多就一定越好吗？*Journal ofInformetrics*，7:635–641，2013年。[9] Michele La mont。教授如何思考：学术评判的奇妙世界内幕。哈佛大学出版社，马萨诸塞州剑桥，2009年。[10] Ben R Martin。研究卓越框架与“影响力议程”：我们是否正在创造一个弗兰肯斯坦怪物？《研究评价》，20:247–254，2011年。[11] Lutz Bornmann。衡量研究的社会影响。《欧洲分子生物学组织报告》，13:673–676，2012年。[12] Henk Moed、Renger De Bruin 和 TH Van Leeuwen。用于评估国家科研绩效的新型文献计量工具：数据库描述、指标概述及初步应用。《科学计量学》，33(3):381–422，1995年。[13] Filippo Radicchi、Santo Fortunato 和 Claudio Castellano。引文分布的普适性：迈向科学影响力的客观衡量。*PNAS*，105:17268–17272，2008年。[14] Ole Ellegaard、Johan A Wallin 和 Sdu Dk。学术产出的文献计量分析：影响力有多大？《科学计量学》，105:1809–1831，2015年。[15] J. E. Hirsch。量化个人科学研究产出的指数。《美国国家科学院院刊》，102:16569–16572，2005年。

[16] Filippo Radicchi 和 Claudio Castellano。大型数据集中的个体学者文献计量指标分析。《科学计量学》，97:627–637，2013年。[17] R Klavans、JP Ioannidis 和 K W Boyack。多项引文指标及其在各科学学科中的综合指标。PLoS Biol，14，2016年。[18] Santo Fortunato、Carl T. Bergstrom、Katy Börner、James A. Evans、Dirk Helbing、Staša Milojevi´c、Alexander M. Petersen、Filippo Radicchi、Roberta Sinatra、Brian Uzzi、Alessandro Vespignani、Ludo Waltman、Dashun Wang 和 Albert László Barabási。科学之科学。Science，359，2018年。[19] Giacomo Livan。不要追随领头者：排名表现如何削弱精英主义。R. Soc. open sci.，6，2019年。[20] Jonathan M Levitt 和 Mike Thelwall。多学科研究是否获得更多引文？一项宏观层面的研究。Journal of the American Society for Information Science and Technology，59(12):1973–1984，2008年。[21] Ye Sun、Giacomo Livan、Athen Ma 和 Vito Latora。跨学科研究人员获得更好的长期资助绩效。Communications Physics，4(1):1–7，2021年。[22] Andrey Rzhetsky、Jacob G. Foster、Ian T. Foster 和 James A. Evans。选择实验以加速集体发现。Proceedings of the National Academy of Sciences，112:14569–14574，2015年。[23] Andrew Plume 博士和 Daphne van Weijen 博士。发表或灭亡？分数作者的兴起？Research trends，1(38):5，2014年。[24] Niklas Reisz、Vito DP Servedio、Vittorio Loreto、William Schueller、Márcia R Ferreira 和 Stefan Thurner。科学研究的可持续性丧失。New Journal of Physics，24(5):053041，2022年。[25] Joseph Schumpeter。经济发展理论。哈佛大学出版社，马萨诸塞州剑桥，1934年。[26] Joseph A Schumpeter 等。商业周期。Mcgraw-hill，纽约，1939年。[27] Mark Dodgson、David M. Gann 和 Nelson Phillips。《牛津创新管理手册》。Oxford University Press，牛津，2014年。[28] Homer Garner Barnett。创新：文化变迁的基础。1953年。[29] Brian Uzzi、Satyam Mukherjee、Michael Stringer 和 Ben Jones。非典型组合与科学影响力。Science，342:468–472，2013年。[30] Magda Fontana、Martina Iori、Fabio Montobbio 和 Roberta Sinatra。新颖且非典型的组合：对新颖性和跨学科性的评估。Research Policy，49，2020年。[31] Lingfei Wu、Dashun Wang 和 James A. Evans。大型团队推动科学与技术的发展，小型团队则颠覆科学与技术。Nature，566:378–382，2019年。[32] Russell J. Funk 和 Jason Owen-Smith。技术变革的动态网络测度。Management Science，63:791–817，2017年。[33] Weihua Li、Tomaso Aste、Fabio Caccioli 和 Giacomo Livan。与顶尖科学家早期合作，预示着学术职业生涯的成功。Nature Communications，10，2019年。[34] Roberta Sinatra、Dashun Wang、Pierre Deville、Chaoming Song 和 Albert László Barabási。量化个体科学影响力的演变。Science，354，2016年。[35] Benjamin F. Jones。知识负担与“文艺复兴人物之死”：创新是否越来越难？Review of Economic Studies，76:283–317，2009年。[36] Mark Buchanan。创新放缓。Nature Physics，11，2015年。[37] Hyejin Youn、Deborah Strumsky、Luis M. A. Bettencourt 和 José Lobo。作为组合过程的发明：来自美国专利的证据。Journal of The Royal Society Interface，12，2015年。[38] E. Richard Gold。创新帝国的衰落及其通过开放科学实现复兴的可能性。Research Policy，50，2021年。[39] Lu Liu、Yang Wang、Roberta Sinatra、C. Lee Giles、Chaoming Song 和 Dashun Wang。艺术、文化与科学职业生涯中的高产期。Nature，559:396–399，2018年。[40] Mikko Packalen 和 Jay Bhattacharya。年龄与新想法的尝试。Journal of human capital，13(2):341–373，2019年。[41] Arthur M Diamond。年龄与计量史学的接受。The Journal of Economic History，40(4):838–841，1980年。

[42] Seongsu Kim. 研究生产力对大学教授提前退休的影响。*IndustrialRelations: A Journal of Economy and Society*，42(1):106–125，2003。[43] Arne Dietrich 和 Narayanan Srinivasan。发动革命的最佳年龄。*TheJournal of Creative Behavior*，41(1):54–74，2007。[44] Oliver E Williams、Lucas Lacasa 和 Vito Latora。量化并预测演艺界的成功。*Nature Communications*，10(1):1–8，2019。[45] Lu Liu、Nima Dehmamy、Jillian Chown、C Lee Giles 和 Dashun Wang。理解艺术、文化和科学职业生涯中高产期的出现。*Nature Communications*，12，2021。[46] Debra Jackson、Garry Walter、John Daly 和 Michelle Cleary。单项研究的多项产出：研究发现拆分的可接受方式与“香肠切片”式拆分。*Journal of Clinical Nursing*，第1–2页，2014。[47] David Moher、Florian Naudet、Ioana A Cristea、Frank Miedema、John PA Ioannidis 和 Steven N Goodman。评估科学家的聘用、晋升和终身教职。*PLOS Biology*，16(3):e2004089，2018。[48] Jie Tang、Jing Zhang、Limin Yao、Juanzi Li、Li Zhang 和 Zhong Su。ArnetMiner：学术社交网络的提取与挖掘。载于*KDD'08*，第990–998页，2008。[49] Tasleem Arif、Rashid Ali 和 M. Asger。使用向量空间模型和混合相似度度量进行作者姓名消歧。载于*2014*年第七届当代计算国际会议（*IC3*），第135–140页，2014。[50] An Zeng、Zhesi Shen、Jianlin Zhou、Ying Fan、Zengru Di、Yougui Wang、H Eugene Stanley 和 Shlomo Havlin。科学家转换研究主题的趋势不断增强。*Nature Communications*，10(1):1–11，2019。[51] Akash Anil 和 Sanasam Ranbir Singh。类别不平衡对异构网络嵌入的影响：一项实证研究。*Journal of Informetrics*，14:101009，2020。[52] Emiel Caron 和 Nees Jan van Eck。使用基于规则的评分和聚类进行大规模作者姓名消歧。载于第*19*届科学与技术指标国际会议论文集，第79–86页。莱顿大学科学技术研究中心，莱顿，2014。[53] Alexander Tekles 和 Lutz Bornmann。文献计量数据的作者姓名消歧：多种无监督方法的比较。*Quantitative Science Studies*，1(4):1510–1528，2020。

# 附录

### 附录 A. 科学创新的非随机性：更多细节

我们通过比较研究人员依据原始出版物序列和重新排列后的序列达到创新峰值所需时间的分布，证明科学创新并非随机。此处，我们进一步比较研究人员在峰值年份之前发表的论文数量分布，以验证这一结果。分布比较见图4。可以看出，在两个学科中，原始数据与随机化数据所得的分布均存在显著差异（两种情况下均为 $p < 0.01$，双侧柯尔莫戈罗夫-斯米尔诺夫检验）。

### 附录B. 表征峰值年份创新：更多细节

本节中，我们展示一幅图，比较“峰值前年份”（BPY）、“峰值年份之后”（APY）和“峰值年份”（PY）各阶段的创新水平（以颠覆性得分衡量）。值得注意的是，此处 BPY 和 APY 阶段的创新水平是根据峰值年份前后各 2 年的时段计算的，而正文中 BPY 和 APY 组的创新水平则是根据完整的出版物历史计算的。这样，我们可以减轻因科学生涯长度或阶段不同而可能产生的偏差。在这项分析中，计算机科学和物理学领域的 BPY 样本数分别为 $N_{CS} = 20$，980 和 $N_{PHY} = 28$，867，PY 样本数分别为 $N_{CS} = 18$，640 和 $N_{PHY} = 26$，543，APY 样本数分别为 $N_{CS} = 21$，987 和 $N_{PHY} = 30$，792。结果绘于图 5。我们发现，峰值年份的平均颠覆性得分高于其他两个阶段（$p < 0.01$ 在所有成对曼–惠特尼 U 检验和双侧柯尔莫戈洛夫-斯米尔诺夫检验中均如此，但$p = 0.021$ 计算机科学领域 PY 与 BPY 之间的曼–惠特尼 U 检验除外），这进一步支持了科学生涯中存在一个“神奇的一年”。

### 附录C. 与努力程度相关变量的分布图

为了量化峰值年份创新的决定因素，我们建立了一系列线性模型，以衡量研究人员在其峰值年份的创新与该年份的相对努力、相对生产率和相对时间投入之间的关系。这里，我们展示了计算机科学和物理学中这些量的分布，并在图6中以归一化直方图呈现。如图所示，相对努力和相对生产率的对数分布在不同学科间非常相似，而相对投入时间的分布则不规则得多，并呈现出一些差异。

### 附录D. 关于科学创新和影响力的其他回归分析

在正文中，关于峰值年份论文的创新和影响力的回归结果，是根据与峰值年份数据相关的变量得出的。作为稳健性检验，我们使用在峰值年份及其前两个年份阶段计算的变量，针对相同的被解释变量建立线性模型。下文展示所有回归模型的结果，见表1—8。从创新模型可以看出，在两个学科中，相对努力和相对时间投入均对峰值年份创新水平产生正向作用，而相对生产率则对峰值年份创新产生负向影响。然而，在影响力回归中，同一组变量产生了相反的作用，这反映了科学创新与影响力机制的差异。考察相应的绝对量时，也能得出类似结果。所有这些结果都支持我们关于峰值年份创新决定因素的论点。

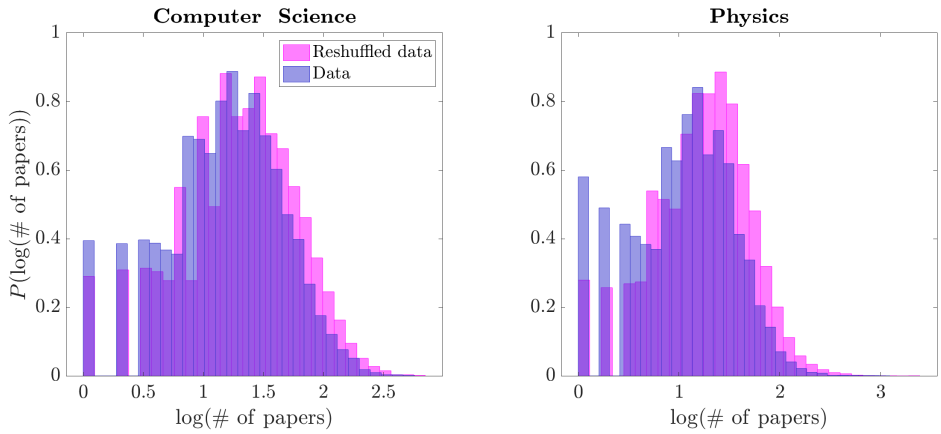


图4：研究人员在计算机科学（左）和物理学（右）领域的峰值年份之前发表的论文数量的归一化直方图，数据经过对数变换，展示原始数据（紫色）和随机化数据（粉色）。

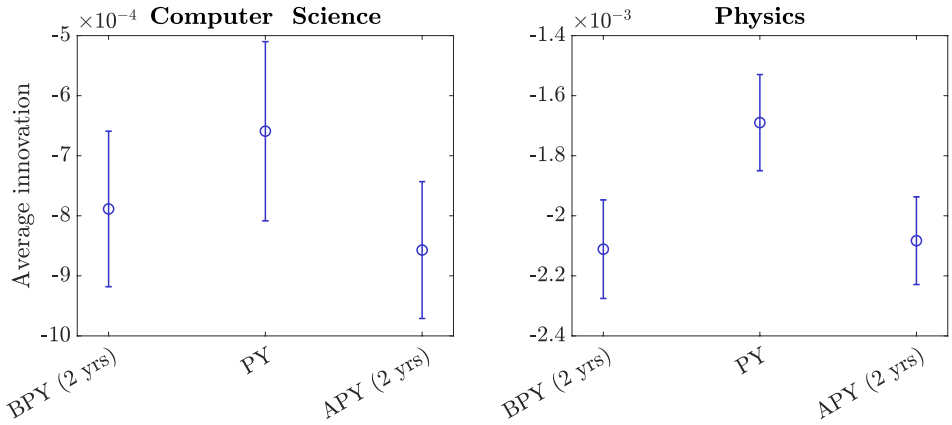


图5：研究人员在峰值年份前2年（峰值前年份，BPY）、峰值年份（不包括促成峰值的论文，峰值年份，PY）以及峰值年份后2年（峰值后年份，APY）发表的论文所达到的平均创新水平。误差棒表示标准误。在这两个学科中，峰值年份发表的论文的平均创新水平均高于其他阶段发表的论文。

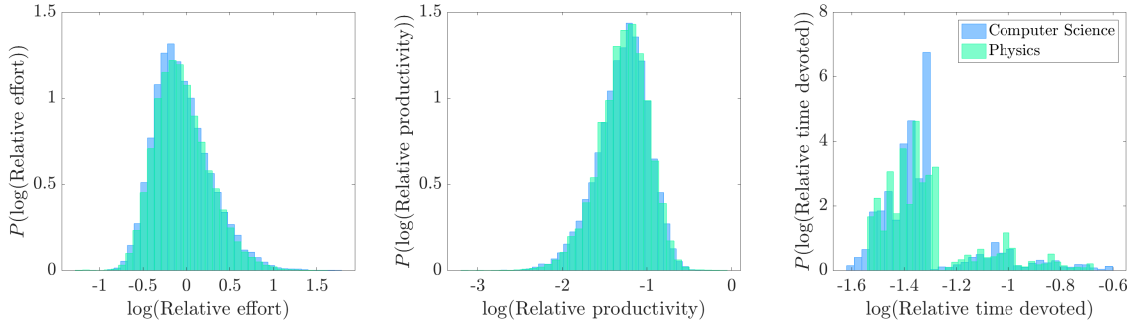


图6：我们样本中的研究人员在峰值年份相关的相对努力程度（左）、相对生产率（中）和相对投入时间（右）的分布直方图，经归一化处理。

表1：使用峰值年份计算的努力相关指标，对计算机科学领域峰值年份创新水平进行回归的结果。

| 自变量 | 计算机科学：使用峰值年份数据建立的峰值年份创新水平模型 | | | | | |
|---|---|---|---|---|---|---|
| | 模型1 | 模型2 | 模型3 | 模型4 | 模型5 | 模型6 |
| 相对努力 | 0.172*** | | | | | |
| | (0.007) | | | | | |
| 努力程度 | | 0.137*** | | | | |
| | | (0.007) | | | | |
| 相对生产率 | | | -0.135*** | | | |
| | | | (0.007) | | | |
| 生产率 | | | | -0.121*** | | |
| | | | | (0.007) | | |
| 相对投入时间 | | | | | 0.023*** | |
| | | | | | (0.007) | |
| 投入时间 | | | | | | 0.021*** |
| | | | | | | (0.007) |
| 平均合著者人数 | 0.108*** | 0.099*** | 0.107*** | 0.103*** | 0.117*** | 0.117*** |
| | (0.006) | (0.007) | (0.007) | (0.007) | (0.007) | (0.007) |
| 平均先前创新 | 0.201*** | 0.209*** | 0.205*** | 0.209*** | 0.206*** | 0.206*** |
| | (0.006) | (0.006) | (0.006) | (0.006) | (0.007) | (0.007) |
| 峰值年份 | -0.251*** | -0.256*** | -0.221*** | -0.256*** | -0.285*** | -0.272*** |
| | (0.009) | (0.009) | (0.009) | (0.009) | (0.010) | (0.009) |
| 达峰时间 | 0.087*** | 0.067*** | 0.040*** | 0.070*** | 0.070*** | 0.057*** |
| | (0.009) | (0.009) | (0.009) | (0.009) | (0.010) | (0.009) |
| $N$ | 20,980 | 20,980 | 20,980 | 20,980 | 20,980 | 20,980 |
| $R^2$ | 0.142 | 0.133 | 0.132 | 0.129 | 0.116 | 0.116 |

表2：使用峰值年份计算的努力相关指标对计算机科学领域峰值年份影响力的回归结果。

| 自变量 | 计算机科学：使用峰值年份数据预测峰值年份影响力的模型 | | | | | |
|---|---|---|---|---|---|---|
| | 模型1 | 模型2 | 模型3 | 模型4 | 模型5 | 模型6 |
| 相对努力 | -0.056*** | | | | | |
| | (0.007) | | | | | |
| 努力程度 | | -0.182*** | | | | |
| | | (0.007) | | | | |
| 相对生产率 | | | 0.021*** | | | |
| | | | (0.007) | | | |
| 生产率 | | | | 0.370*** | | |
| | | | | (0.007) | | |
| 相对投入时间 | | | | | -0.077*** | |
| | | | | | (0.008) | |
| 投入时间 | | | | | | -0.056*** |
| | | | | | | (0.007) |
| 平均合著者人数 | 0.030*** | 0.051*** | 0.028*** | 0.072*** | 0.029*** | 0.028*** |
| | (0.007) | (0.007) | (0.007) | (0.007) | (0.007) | (0.007) |
| 平均先前创新 | 0.035*** | 0.029*** | 0.034*** | 0.023*** | 0.033*** | 0.034*** |
| | (0.007) | (0.007) | (0.007) | (0.006) | (0.007) | (0.007) |
| 峰值年份 | 0.008 | -0.007 | 0.008 | -0.037*** | 0.058*** | 0.014 |
| | (0.010) | (0.010) | (0.010) | (0.009) | (0.011) | (0.010) |
| 达峰时间 | 0.044*** | 0.039*** | 0.056*** | 0.012 | 0.007 | 0.052*** |
| | (0.010) | (0.009) | (0.010) | (0.009) | (0.011) | (0.010) |
| $N$ | 20,980 | 20,980 | 20,980 | 20,980 | 20,980 | 20,980 |
| $R^2$ | 0.009 | 0.038 | 0.007 | 0.134 | 0.011 | 0.009 |

表3：使用峰值年份计算的与努力相关的指标，对物理学峰值年份创新水平进行回归的结果。

| 自变量 | 物理学：使用峰值年份数据对峰值年份创新水平建模 | | | | | |
|---|---|---|---|---|---|---|
| | 模型1 | 模型2 | 模型3 | 模型4 | 模型5 | 模型6 |
| 相对努力 | 0.141*** | | | | | |
| | (0.006) | | | | | |
| 努力程度 | | 0.205*** | | | | |
| | | (0.006) | | | | |
| 相对生产率 | | | -0.106*** | | | |
| | | | (0.006) | | | |
| 生产率 | | | | -0.097*** | | |
| | | | | (0.006) | | |
| 相对投入时间 | | | | | 0.070*** | |
| | | | | | (0.006) | |
| 投入时间 | | | | | | 0.059*** |
| | | | | | | (0.006) |
| 平均合著者人数 | -0.004 | 0.007 | -0.005 | 0.005 | 0.009 | 0.008 |
| | (0.006) | (0.006) | (0.006) | (0.006) | (0.006) | (0.006) |
| 平均先前创新 | 0.232*** | 0.229*** | 0.233*** | 0.233*** | 0.233*** | 0.233*** |
| | (0.006) | (0.006) | (0.006) | (0.006) | (0.006) | (0.006) |
| 峰值年份 | -0.143*** | -0.152*** | -0.118*** | -0.149*** | -0.187*** | -0.150*** |
| | (0.008) | (0.008) | (0.008) | (0.008) | (0.009) | (0.008) |
| 达峰时间 | 0.095*** | 0.098*** | 0.059*** | 0.089*** | 0.114*** | 0.080*** |
| | (0.008) | (0.008) | (0.008) | (0.008) | (0.009) | (0.008) |
| $N$ | 28,867 | 28,867 | 28,867 | 28,867 | 28,867 | 28,867 |
| $R^2$ | 0.086 | 0.108 | 0.078 | 0.076 | 0.071 | 0.070 |

表4：使用峰值年份计算的与努力相关的指标，对物理学峰值年份影响力进行的回归结果。

| 自变量 | 物理学：使用峰值年份数据分析峰值年份影响力的模型 | | | | | |
|---|---|---|---|---|---|---|
| | 模型1 | 模型2 | 模型3 | 模型4 | 模型5 | 模型6 |
| 相对努力 | -0.018*** | | | | | |
| | (0.006) | | | | | |
| 努力程度 | | -0.074*** | | | | |
| | | (0.006) | | | | |
| 相对生产率 | | | 0.010 | | | |
| | | | (0.006) | | | |
| 生产率 | | | | 0.157*** | | |
| | | | | (0.006) | | |
| 相对投入时间 | | | | | -0.038*** | |
| | | | | | (0.006) | |
| 投入时间 | | | | | | -0.029*** |
| | | | | | | (0.006) |
| 平均合著者人数 | 0.006 | 0.005 | 0.006 | 0.005 | 0.003 | 0.004 |
| | (0.006) | (0.006) | (0.006) | (0.006) | (0.006) | (0.006) |
| 平均先前创新 | -0.004 | -0.003 | -0.004 | -0.004 | -0.004 | -0.004 |
| | (0.006) | (0.006) | (0.006) | (0.006) | (0.006) | (0.006) |
| 峰值年份 | -0.037*** | -0.035*** | -0.039*** | -0.036*** | -0.016* | -0.036*** |
| | (0.008) | (0.008) | (0.009) | (0.008) | (0.009) | (0.008) |
| 达峰时间 | 0.015* | 0.010 | 0.019*** | -0.003 | -0.003 | 0.016* |
| | (0.008) | (0.008) | (0.008) | (0.008) | (0.009) | (0.008) |
| $N$ | 28,867 | 28,867 | 28,867 | 28,867 | 28,867 | 28,867 |
| $R^2$ | 0.001 | 0.006 | 0.001 | 0.025 | 0.002 | 0.002 |

表5：利用峰值年份及其前两年的努力程度相关指标，得出的计算机科学领域创新峰值水平回归结果。

| 自变量 | 计算机科学：基于全部数据对峰值年份创新进行建模 | | | | | |
|---|---|---|---|---|---|---|
| | 模型1 | 模型2 | 模型3 | 模型4 | 模型5 | 模型6 |
| **峰值年份变量** | | | | | | |
| 相对努力 | 0.175*** (0.007) | | | | | |
| 努力程度 | | 0.150*** (0.007) | | | | |
| 相对生产率 | | | -0.140*** (0.007) | | | |
| 生产率 | | | | -0.199*** (0.011) | | |
| 相对投入时间 | | | | | 0.026*** (0.008) | |
| 投入时间 | | | | | | 0.027*** (0.007) |
| 平均合著者人数 | 0.091*** (0.008) | 0.086*** (0.008) | 0.086*** (0.008) | 0.085*** (0.008) | 0.104*** (0.008) | 0.104*** (0.008) |
| **峰值变量前两年** | | | | | | |
| 相对努力 | -0.010 (0.007) | | | | | |
| 努力程度 | | -0.033*** (0.007) | | | | |
| 相对生产率 | | | 0.016** (0.007) | | | |
| 生产率 | | | | 0.099*** (0.011) | | |
| 相对投入时间 | | | | | 0.016** (0.007) | |
| 投入时间 | | | | | | 0.021*** (0.007) |
| 平均合著者人数 | 0.047*** (0.008) | 0.040*** (0.008) | 0.055*** (0.008) | 0.045*** (0.008) | 0.040*** (0.008) | 0.040*** (0.008) |
| **控制变量** | | | | | | |
| 平均先前创新 | 0.189*** (0.007) | 0.198*** (0.007) | 0.192*** (0.007) | 0.197*** (0.007) | 0.193*** (0.007) | 0.193*** (0.007) |
| 峰值年份 | -0.254*** (0.009) | -0.261*** (0.009) | -0.230*** (0.010) | -0.261*** (0.009) | -0.296*** (0.011) | -0.273*** (0.009) |
| 达峰时间 | 0.086*** (0.009) | 0.065*** (0.009) | 0.044*** (0.009) | 0.065*** (0.009) | 0.084*** (0.011) | 0.062*** (0.009) |
| $N$ | 20,081 | 20,081 | 20,081 | 20,081 | 20,081 | 20,081 |
| $R^2$ | 0.140 | 0.132 | 0.129 | 0.130 | 0.113 | 0.114 |

表6：使用峰值年份及其前两年计算的与努力相关的指标，得出的计算机科学领域峰值年份影响回归结果。

| 自变量 | 计算机科学：基于全部数据的峰值年份影响模型 | | | | | |
|---|---|---|---|---|---|---|
| | 模型1 | 模型2 | 模型3 | 模型4 | 模型5 | 模型6 |
| **峰值年份变量** | | | | | | |
| 相对努力 | -0.057*** (0.008) | | | | | |
| 努力程度 | | -0.154*** (0.008) | | | | |
| 相对生产率 | | | 0.026*** (0.008) | | | |
| 生产率 | | | | 0.292*** (0.011) | | |
| 相对投入时间 | | | | | -0.092*** (0.008) | |
| 投入时间 | | | | | | -0.074*** (0.007) |
| 平均合著者人数 | 0.030*** (0.008) | 0.041*** (0.008) | 0.030*** (0.008) | 0.055*** (0.008) | 0.027*** (0.008) | 0.026*** (0.008) |
| **峰值变量前两年** | | | | | | |
| 相对努力 | -0.002 (0.008) | | | | | |
| 努力程度 | | -0.076*** (0.008) | | | | |
| 相对生产率 | | | -0.021*** (0.008) | | | |
| 生产率 | | | | 0.098*** (0.011) | | |
| 相对投入时间 | | | | | -0.073*** (0.008) | |
| 投入时间 | | | | | | -0.062*** (0.007) |
| 平均合著者人数 | 0.0001 (0.008) | 0.023*** (0.008) | -0.003 (0.008) | 0.031*** (0.008) | 0.010 (0.008) | 0.009 (0.008) |
| **其他变量** | | | | | | |
| 平均先前创新 | 0.035*** (0.007) | 0.027*** (0.007) | 0.034*** (0.007) | 0.020*** (0.007) | 0.033*** (0.007) | 0.034*** (0.007) |
| 峰值年份 | 0.005 (0.010) | -0.013 (0.009) | 0.011 (0.010) | -0.042*** (0.009) | 0.095*** (0.011) | 0.008 (0.009) |
| 达峰时间 | 0.039*** (0.010) | 0.032*** (0.009) | 0.050*** (0.009) | 0.007 (0.009) | -0.044*** (0.012) | 0.043*** (0.009) |
| $N$ | 20,081 | 20,081 | 20,081 | 20,081 | 20,081 | 20,081 |
| $R^2$ | 0.008 | 0.042 | 0.006 | 0.137 | 0.015 | 0.012 |

表7：利用峰值年份及其前两年的与努力相关的计算指标，对物理学领域峰值年份的创新水平进行回归分析所得的结果。

| 自变量 | 物理学：基于全部数据的峰值年份创新模型 | | | | | |
|---|---|---|---|---|---|---|
| | 模型1 | 模型2 | 模型3 | 模型4 | 模型5 | 模型6 |
| **峰值年份变量** | | | | | | |
| 相对努力 | 0.154*** (0.006) | | | | | |
| 努力程度 | | 0.203*** (0.006) | | | | |
| 相对生产率 | | | -0.105*** (0.006) | | | |
| 生产率 | | | | -0.109*** (0.009) | | |
| 相对投入时间 | | | | | 0.086*** (0.007) | |
| 投入时间 | | | | | | 0.080*** (0.007) |
| 平均合著者人数 | -0.017** (0.007) | 0.001 (0.007) | -0.017** (0.007) | -0.004 (0.007) | 0.006 (0.007) | 0.006 (0.007) |
| **峰值变量前两年** | | | | | | |
| 相对努力 | -0.027*** (0.007) | | | | | |
| 努力程度 | | 0.007 (0.006) | | | | |
| 相对生产率 | | | 0.004 (0.006) | | | |
| 生产率 | | | | 0.018 (0.010) | | |
| 相对投入时间 | | | | | 0.062*** (0.007) | |
| 投入时间 | | | | | | 0.056*** (0.007) |
| 平均合著者人数 | 0.017** (0.007) | 0.010 (0.007) | 0.019** (0.007) | 0.012 (0.007) | 0.003 (0.007) | 0.003 (0.007) |
| **其他变量** | | | | | | |
| 平均先前创新 | 0.248*** (0.006) | 0.245*** (0.006) | 0.250*** (0.006) | 0.250*** (0.006) | 0.247*** (0.006) | 0.248*** (0.006) |
| 峰值年份 | -0.137*** (0.008) | -0.146*** (0.008) | -0.114*** (0.009) | -0.141*** (0.008) | -0.214*** (0.010) | -0.140*** (0.008) |
| 达峰时间 | 0.091*** (0.008) | 0.094*** (0.008) | 0.062*** (0.008) | 0.088*** (0.008) | 0.151*** (0.010) | 0.084*** (0.008) |
| $N$ | 25,222 | 25,222 | 25,222 | 25,222 | 25,222 | 25,222 |
| $R^2$ | 0.095 | 0.116 | 0.084 | 0.083 | 0.081 | 0.080 |

表8：使用峰值年份及其前两年计算的与努力相关的变量，对物理学领域峰值年份影响力进行回归的结果。

| 自变量 | 物理学：基于全部数据的峰值年份影响力模型 | | | | | |
|---|---|---|---|---|---|---|
| | 模型1 | 模型2 | 模型3 | 模型4 | 模型5 | 模型6 |
| **峰值年份变量** | | | | | | |
| 相对努力 | -0.034*** | | | | | |
| | (0.007) | | | | | |
| 努力程度 | | -0.080*** | | | | |
| | | (0.007) | | | | |
| 相对生产率 | | | 0.021*** | | | |
| | | | (0.007) | | | |
| 生产率 | | | | 0.124*** | | |
| | | | | (0.010) | | |
| 相对投入时间 | | | | | -0.057*** | |
| | | | | | (0.007) | |
| 投入时间 | | | | | | -0.046*** |
| | | | | | | (0.007) |
| 平均合著者人数 | 0.014 | 0.004 | 0.014 | -0.001 | 0.005 | 0.006 |
| | (0.007) | (0.007) | (0.007) | (0.007) | (0.007) | (0.007) |
| **峰值变量前两年** | | | | | | |
| 相对努力 | 0.010 | | | | | |
| | (0.007) | | | | | |
| 努力程度 | | -0.038*** | | | | |
| | | (0.007) | | | | |
| 相对生产率 | | | -0.013 | | | |
| | | | (0.007) | | | |
| 生产率 | | | | 0.123*** | | |
| | | | | (0.010) | | |
| 相对投入时间 | | | | | -0.034*** | |
| | | | | | (0.007) | |
| 投入时间 | | | | | | -0.026*** |
| | | | | | | (0.007) |
| 平均合著者人数 | 0.0004 | 0.006 | -0.002 | 0.011 | 0.006 | 0.006 |
| | (0.007) | (0.007) | (0.008) | (0.007) | (0.007) | (0.007) |
| **其他变量** | | | | | | |
| 平均先前创新 | -0.007 | -0.004 | -0.007 | -0.006 | -0.006 | -0.007 |
| | (0.006) | (0.006) | (0.006) | (0.006) | (0.006) | (0.006) |
| 峰值年份 | -0.038*** | -0.035*** | -0.039*** | -0.038*** | 0.008 | -0.037*** |
| | (0.009) | (0.009) | (0.009) | (0.008) | (0.010) | (0.009) |
| 达峰时间 | 0.014 | 0.007 | 0.019** | -0.013 | -0.028*** | 0.014 |
| | (0.009) | (0.008) | (0.009) | (0.008) | (0.010) | (0.008) |
| $N$ | 25,222 | 25,222 | 25,222 | 25,222 | 25,222 | 25,222 |
| $R^2$ | 0.002 | 0.011 | 0.001 | 0.054 | 0.004 | 0.003 |