

科技基本投入对论文产出的影响研究

——以中国国际论文为例

高 璐¹, 仝芳妍², 邓心安¹

(1. 中国农业大学 人文与发展学院, 北京 100094; 2. 中国科学院 科技政策与管理研究所, 北京 100080)

摘 要: 科技论文产出是衡量一个国家科技创新能力与科学影响力的重要指标。选取1987-2004年基础研究和应用研究的R&D经费、从事R&D活动的科学家与工程师人数以及国际论文数3个指标数据, 构建出科技基本投入与论文产出之间的关系模型, 首次对论文产出相对于科技投入的滞后效应进行了分析。

关键词: 研究与发展; 科学家与工程师; 科技论文; 相关与回归分析; 滞后分析

中图分类号: G311

文献标识码: A

文章编号: 1001-7348(2007)09-0025-04

科技论文是国家科技创新能力的重要体现, 能够从一个侧面反映该国科学的国际地位和影响力。研究科技基本投入与论文产出之间的关系, 对于合理评价科学研究水平、制定基础研究支持政策具有重要参考作用。

1 研究目的与意义

近年来中国政府加强了对科学研究的投入强度, 尤其是经费投入显著增加, 对论文数量增加和质量提高都产生了积极的影响。目前我国对科技论文的统计研究侧重于对其数量及变化趋势的常规分析, 包括论文数量动态分析、结构比例分析、论文应用情况分析等, 如中国科学技术信息研究所自1987年起开始对科技论文进行统计与分析研究工作, 其中包含了科技投入与产出状态分析; 其系列报告《中国科技论文统计与分析》从数量上分析了中国论文总体、各学科、各地区、各科研机构的变化。自20世纪60年代加菲尔德的《科学引文索引》(SCI)问世以来, 影响因素分析日益发展。它以美国科技信息研究所(Institute for Scientific Information, 简称ISI)发布的检索工具为基础, 运用科学计量学、文献计量学等方法, 对一个国家或地区及机构的论文产出进行影响因子与引用情况分析。运用文献计量学方法评价研究产出, 针对的产出形式主要是论文, 研究内容侧重于论文的分布、比重、引用情况^[1]。国内对科研绩效评价问题的争论颇多, 论文数量多了是否质量就低了。柳卸林、赵捷应用道格拉斯模型法, 对影响我国SCI论文数量与质量的关键因素进行了分析, 认为“科研经费是

决定论文数量和质量的关键因素^[2]”。

美国科技投入与产出关系研究^[3], 将投入形式分为两个主要方面: 一是从事科技活动的人力, 二是用于科技活动的资金, 包括R&D投入、科技和工程教育培训费用等; 产出形式中最主要的就是在国际知名度较高的刊物上发表论文数及其被引情况。国内外有关科技投入与产出关系的多元分析研究很多, 如李基磐将科技投入与产出当作一个多元复杂系统, 认为“投入和产出指标间的相关皆在1%水平呈显著^[4]”; 王贻志等从生产函数的角度研究了科技投入与产出的定量关系^[5]。但将论文从产出中抽提出来, 专门研究科技投入与论文产出关系的研究尚不多见, 其形式主要表现在以平均指标、相对指标为特征的常规描述分析, 如瑞士洛桑国际管理开发研究院(IMD)《国际竞争力报告》研究, 结果显示当年中国从事R&D活动的科学家与工程师平均千人发表的论文数仅有11篇, 在12个大国中排最后一位^[6]。

将科技投入与论文产出结合起来进行统计推断分析缺乏的原因: 一是由于基础研究的强探索性和产出不确定性; 二是由于科技投入不是当年就能见效, 与论文产出之间存在一个明显的滞后期。所谓“不确定性”及高风险的探索性, 都是针对具体的研究对象和微观的个体机构而言, 一旦将分析对象建立在国家或地区整体与宏观的统计分析基础上, 那么, 以多元分析为特征的定量分析便可能成为研究科技投入与论文产出关系的一个有效手段。

研究科研经费和人力资源投入对论文产出的影响, 对

收稿日期: 2007-08-31

作者简介: 高璐(1983-), 女, 汉族, 黑龙江省大庆市人, 中国农业大学人文与发展学院硕士研究生; 研究方向为科技政策, 生物经济; 仝芳妍(1983-), 女, 汉族, 云南昆明市人, 中国科学院科技政策与管理研究所研究生, 研究方向为科技政策; 邓心安(1963-), 男, 汉族, 中国农业大学人文与发展学院副教授, 研究方向为生物经济和科技发展战略与政策。

论文产出的滞后期进行分析,对于把握科技投入与论文产出关系变化的规律性,提高论文的国际影响力、加速论文的应用等都具有重要意义。

2 研究方法 with 数据处理

2.1 指标选取

将中国论文的数量与质量放在世界范围内进行分析,应选取国际上通用的文献检索方法,因而本文国际论文数指的是3个系统(SCI、EI、ISTP)收录的各年度中国大陆论文总数。

科研经费和人力资源是科技基本投入的两个代表性指标,二者分别以研究与发展(R&D)经费、科学家和工程师(S&E)人数作为衡量指标。

R&D活动是科技创新之源。R&D经费投入对我国科技论文产出有重要影响,一个国家科技水平的高低,在一定程度上取决于其对科技的投入的多少^[4]。由于研究对象是科技论文,相对于R&D活动中的试验发展而言,基础研究与应用研究对论文产出有着比较直接的贡献,而前者产出效果更直接地表现在专利和经济效益方面,因而选取R&D经费中基础研究和应用基础研究支出作为科研经费投入指标。

高素质的科技人员是科学研究的中坚力量。选用从事R&D活动的S&E人数作为人力资源投入指标,因为这部分人员代表着科技创新的实力,对创造性的科技成果有着直接的贡献。

综上,本文选取1987-2004年的中国大陆国际论文数、基础研究和应用基础研究的R&D经费(以下简称BAR&D)、从事R&D活动的S&E人数3个指标进行3个时间序列分析。

2.2 数据预处理及散点图

按照商品零售价格总指数,将研究年份的BAR&D经费投入统一折算为用2000年的固定价格计算的绝对数。然后对数据进行无量纲化处理,将1987年的论文数、BAR&D经费及S&E人数作为基数标准值,令其各等于100;由此推算出1988-2004年相应指标的无量纲化值。从事R&D活动的S&E人数按全时当量折算数。(表1)

将BAR&D经费、S&E人数分别作为自变量X1, X2, 国际论文数作为因变量Y, 作变量的分布散点图(图1), 可以直观地判断变量间的关系。

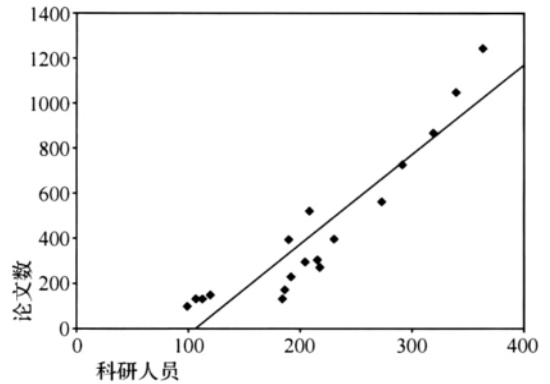
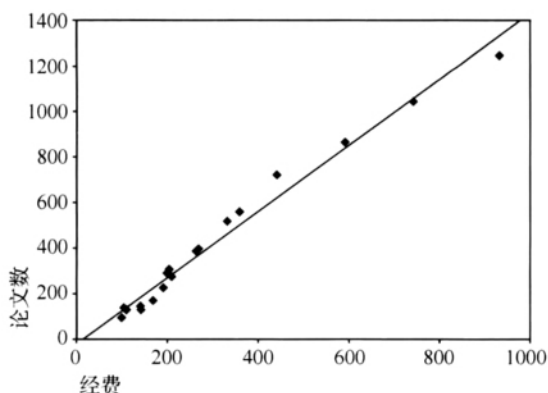


图1 BAR&D经费、S&E人数对论文数影响散点图

由图1可见: 这些点大致分布在一条直线的上下; 论文数与BAR&D经费、S&E人数之间大致呈线性关系。

表1 中国国际论文数、BAR&D经费及S&E人数(1987-2004年)

年份	国际论文数(篇)	无量纲化论文数	BAR&D(当年价, 亿元)	无量纲化BAR&D	S&E人数(万人)	无量纲化S&E人数
	Y_i			X_{1i}		X_{2i}
1987	8932	100	22.63	100	25.45	100
1988	11744	131.48	29.75	110.89	27.18	106.79
1989	12102	135.49	32.73	103.48	28.80	113.16
1990	13183	147.59	44.90	139.05	30.41	119.48
1991	11556	129.37	46.28	139.28	47.14	185.22
1992	15466	173.15	59.13	168.71	47.19	185.42
1993	20277	227.01	75.79	190.97	48.92	192.22
1994	24584	275.23	100.51	208.11	55.20	216.89
1995	26305	294.50	110.08	198.44	52.20	205.10
1996	27569	308.65	119.36	202.86	54.80	215.32
1997	35311	395.33	159.93	269.53	58.87	231.31
1998	35003	391.88	153.54	265.69	48.55	190.76
1999	46188	517.10	185.41	330.95	53.11	208.68
2000	49678	556.18	198.67	359.94	69.51	273.12
2001	64526	722.41	240.40	439.06	74.27	291.82
2002	77395	866.49	320.48	593.07	81.05	318.46
2003	93352	1045.14	399.06	739.26	86.21	338.74
2004	111356	1246.70	517.7	932.34	92.63	363.96

注:(1)资料来源:根据中国科技信息研究所1989-2004年《中国科技论文统计与分析年度研究报告》(科学技术文献出版社)和国家统计局/科学技术部历年《中国科技统计年鉴》(中国统计出版社)整理。

(2)历年BAR&D经费在进行无量纲化之前,按照2000年=100的定基价格指数,将其按当年价格计算数分别统一折算成2000年固定价格计算数。

(3)本统计不包括港、澳、台的统计数据。

3 相关与回归分析

3.1 相关分析

由散点图可直观得出,论文数与BAR&D经费、S&E

人数之间具有相关关系, 为此可作进一步的相关分析。应用 SPSS 软件作二元等距变量相关分析, 得出相关系数的矩阵(表 2)。

表 2 相关系数矩阵

	论文数	BAR&D 经费	S&E 人数
论文数	1	0.991	0.928
BAR&D 经费	0.991	1	0.909
S&E 人数	0.928	0.909	1

由表 2 容易看出, 论文数与 BAR&D 经费、S&E 人数之间均呈现高度线性相关; 相对于 S&E 人数而言, BAR&D 经费与论文数的相关程度更加明显。经 s 检验, BAR&D 经费、S&E 人数与论文数均显著相关, 可以进行下一步回归分析。

3.2 回归分析

回归分析侧重于考察变量之间的数量变化规律, 通过一定的数学表达式来描述变量之间的关系, 进而确定一个或几个变量的变化对另一个特定变量的影响程度。

假设 Y 与 X_1 , X_2 符合二元线性回归模型:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

式中: β_0 是回归常数, β_1, β_2 为回归系数, ε 是随机误差。

运用 SPSS 软件, 得出回归方程系数、常用统计量以及方差分析表。关于论文数、BAR&D 经费、S&E 人数的回归方程为:

$$Y = -95.277 + 1.237X_1 + 0.67X_2 \quad (1)$$

表 3 回归统计量

相关系数	判定系数	调整的判定系数	标准差	观测值个数
0.993	0.986	0.984	43.545	18

表 4 方差分析表

	df	SS	MS	F	Sg
回归	2	1937926	968963.096	511.011	0.000
残差	16	28442.56	1896.171		
总和	18	1966396			

由表 3 可见, 判定系数接近于 1, 表明线性回归效果显著, 回归方程与样本观测值的拟合优度高; 自变量与因变量之间存在显著的线性关系, 自变量的变化能够反映因变量的线性变化, 回归方程显著。

F 检验: 查 F 分布表, 得 $F_{\alpha}(2, 18-2-1) = 3.68 (\alpha=0.05)$, 表 4 中的 F 值远远大于 F_{α} , 说明在 0.05 的显著性水平下, Y 对 X_1, X_2 有显著的线性关系, 即回归方程(1)是显著的, 亦即“自变量全体对 Y 产生影响”结论错误的概率不超过 5%。方程(1)可用于进一步的预测、控制及比较研究。

4 论文产出的滞后分析

科技投入对论文产出影响存在一定滞后效应, 即当年投入并不一定对当年论文产出产生明显影响。因此, 研究科技投入对论文产出的影响, 要进一步分析论文发表时间

的滞后性。利用回归模型作滞后分析, 可以揭示出论文产出相对于科技投入的滞后期。其基本思路是:

分别用时间序列第 i 年的论文数作为 Y_i 值, 第 i-n 年的 BAR&D 经费和 S&E 人数分别作为相应 X_1 和 X_2 值 ($1 < i < 17$; $n=1, 2, 3, 4, \dots$; $i > n$; i, n 均为整数), 从而生成新的时间序列, 以此进行回归分析; 然后进行检验, 若发现滞后 n 年的回归方程拟合程度最高, 则说明在 BAR&D 经费、S&E 人数投入后的第 n 年对论文产出影响最为显著。

滞后 1 年的回归方程为:

$$Y = -126.782 + 1.651X_1 + 0.627X_2 \quad (2)$$

$R = 0.987$ $R^2 = 0.975$ 调整后的 $R^2 = 0.971$ F 检验值为 $F = 273.185$

滞后 2 年:

$$Y = -155.508 + 2.377X_1 + 0.267X_2 \quad (3)$$

$R = 0.988$ $R_2 = 0.977$ 调整后的 $R^2 = 0.973$ F 检验值为 $F = 274.050$

滞后 3 年:

$$Y = -186.493 + 3.644X_1 - 0.587X_2 \quad (4)$$

$R = 0.985$ $R^2 = 0.970$ 调整后的 $R^2 = 0.965$ F 检验值为 $F = 194.527$

滞后 4 年:

$$Y = -212.383 + 4.669X_1 - 1.138X_2 \quad (5)$$

$R = 0.981$ $R^2 = 0.962$ 调整后的 $R^2 = 0.955$ F 检验值为 $F = 138.991$

由判定系数 R^2 和 F 值检验可见, 回归方程(2)~(5)的显著水平均较高, 回归效果明显。其中, 滞后期为 2 年的回归方程的判定系数 R^2 和 F 检验值, 双双大于其它滞后期对应值; 其次为滞后 1 年。说明在滞后期为 2 年时, 回归方程的拟合优度最高。由此可得出初步结论: 在基础研究和应用研究 R&D 经费和从事 R&D 活动的科学家与工程师投入两年后, 对科技论文数的影响最大; 滞后 1 年的影响次之。

5 结论与讨论

关于科技基本投入与论文产出的关系。中国国际论文数与 BAR&D 经费、S&E 人数投入之间具有很强的线性关系, 利用 1987-2004 年时间序列, 可以构建起科技基本投入与论文产出的关系模型。经检验认为, 运用回归方程, 不仅可以对未来论文数进行预测分析, 而且可以根据科技发展规划目标, 对科技经费和人力资源投入进行控制分析。回归系数的大小能够描述变量的重要程度。由回归方程(1)可见, BAR&D 经费投入与论文产出的关系更为密切, 也就是说, 通过加大 BAR&D 经费投入的力度, 能够较快地提高国际论文的数量。

关于论文产出的滞后分析。回归方程的滞后分析表明, 中国国际论文产出的滞后期平均为 2 年。也就是说, 从总体上讲, 中国国际论文数产出相对于 BAR&D 经费投入和 S&E 人员投入的滞后期为 2 年。滞后 3 年及以上的 X_2

回归系数出现负值的现象表明,相对于在特定的 BAR&D 经费(X_1)投入保持不变的情况下, S&E 人数(X_2)增加,对论文产出(Y)反倒可能产出“负面”影响,即投入的人数越多,论文数量反而下降。

由滞后分析可见,中国论文产出的滞后期处于较短的范围内。滞后期产生的原因主要是由研发时间和论文评审时间所造成。考虑到从投稿到刊物发表一般有半年到 1 年的时间,从“滞后 1 年影响次之”等结果分析,无论是 BAR&D 经费投入还是 S&E 人员投入,在第 2 年和第 1 年即当年对论文产出的影响最为显著,从侧面说明,我国论文的“短平快”项目现象比较突出。如果对其它国家尤其是科技发达国家的论文产出滞后期及科研体制进行对比分析,那么可以更全面地揭示出科技基本投入与论文产出的关系,进一步解释我国从事 R&D 活动人员平均论文产出偏低等现象。

本文仅仅考虑到 BAR&D 经费和 S&E 人数两个自变量同时滞后 n 年的情况,其实还应存在两个自变量分别滞后不同时间的情况,即有关自变量与因变量滞后的配伍问题。若考虑“配伍”问题,必然涉及 BAR&D 经费和 S&E 人员投入分别对论文产出的贡献率。考虑到资料收集的难易和翔实程度以及计算工作量,本文对两个自变量分别滞后对论文产出影响的“配伍”问题未予探讨。

关于其它因素量化问题。研究科技投入对论文产出的关系,除考虑科技经费和人力两个基本投入外,是否存在其它重要因素?论文产出的增加,除了存在经费和人力投入两大因素影响之外,ISI 等收录刊物数量及登载量的增加也是影响因素之一。如何量化其它因素的影响,是下一步研究应当考虑的方面。由刊物数量及登载量增加而导致的论文数增加,一定程度上也是要通过经费和人力投入的作用才能实现,因而,这些“人为”因素的暂未考虑对现有模型分析并不构成决定性影响,这些因素之间可能存在多重共线性问题。

关于论文数量与质量的关系问题。尽管中国论文数量上升很快,但在各个学科的论文篇均被引频次等方面均低于国际平均水平,论文质量有待于进一步提高。目前国内对基于科技论文的评价系统争议颇多。虽然“论文的数量与科学成果的价值和意义没有必然的逻辑关系^[7]”,但是基础研究对长远研究与开发的后劲、技术的引进消化吸收、人才培养等方面都有间接影响,况且高水平的论文质量也是建立在足够的论文数量基础之上的。重视国际论文与“片面追求 SCI”不是一个概念。

本文基本没有涉及论文质量方面的问题,具有一定局限性。但这并不意味着对质量的忽视,相反通过数量与质量的互补与对比分析,更有利于发现问题,制定更适于基础科学发展的政策,扩大我国论文的国际影响力和应用潜力。

参考文献:

- [1] 琳达·巴特勒. 用文献计量学方法评价研究产出[A]. 中国科学技术指标研究会. 科技指标及其要政策中的应用[M]. 北京: 科学技术文献出版社, 2002.41- 51.
- [2] 柳卸林, 赵捷. 影响中国 SCI 数量与质量的因素分析[J]. 中国软科学, 2003, (10): 87- 94.
- [3] 午许. 美国对科技投入产出评价方法的研究[J]. 全球科技经济瞭望, 1996, (3).
- [4] 李基磐. 科技投入——产出相关研究及管理效益分析[J]. 海洋水产研究, 1996, (1): (9) - 106.
- [5] 王贻志, 陈晓声, 茅国平, 阮大成. 科技投入与产出的计量研究[J]. 数量经济技术经济研究, 2002, (7).
- [6] 中国科学院可持续发展研究组. 2002 中国可持续发展战略报告[M]. 北京: 科学出版社, 2002.134- 135.
- [7] 郑金武. 潘际奎院士: 科研评价指挥棒应指向解决国家重大科学问题[N]. 科学时报, 2005.11.17.

(责任编辑: 董小玉)

Impact of Sci- tech Basic Inputs on the Output of Papers

Abstract: The output of international sci- tech papers is an important standard which can reflect the sci- tech innovation and the influence of science. Through expenditure of R&D for basic research and applied research, and the personnel of scientists and engineers engaged in R&D, the authors use correlation and regression analysis to get the regression equations, and analyse the lagging effects of paper outputs for the first time.

Key Words: R&D; scientists and engineers; sci- tech paper; correlation and regression analysis; analysis of paper output lagging