

# 发达国家科研投入效率初探 ——基于16个OECD国家纵列数据的统计学分析

余 昕<sup>1</sup>, 王 冬<sup>2</sup>, 韩 楠<sup>2</sup>, 王 欣<sup>3</sup>

(1.北京大学 人口研究所, 北京 100871; 2.北京大学 信息管理系, 北京 100871; 3.清华大学 公共管理学院, 北京 100084)

摘 要: 以16个OECD国家作为研究对象, 以SCI来源期刊论文量作为衡量一国科研产出的指标, 通过纵列数据(Panel Data)分析方法建立了科研投入-科研产出关系模型, 定量描述了发达国家科研产出与科研经费投入、科研人员数、时间等因素的关系。为了排除母语因素对SCI论文量可能产生的影响, 语言变量也被引入到模型中。同时, 还比较了不同统计模型估计出的不同结果, 验证了统计模型选取的重要性。

关键词: 科研投入; 效率; 回归; OECD国家; 纵列数据

中图分类号: G311

文献标识码: A

文章编号: 1001-7348(2007)08-0129-03

## 0 前言

科研产出究竟在多大程度上依赖于科研投入?本文以16个OECD国家近10年来被SCI收录的论文数量作为衡量科研产出的指标, 将科研经费、科研人员数量作为衡量国家科研投入的指标。试图利用纵列数据(Panel Data)建立一个科研投入-科研产出数学模型, 寻找科研投入与科研

产出之间的关系。

国外学者对科研投入效率的研究比较深入, 有学者利用数据包络分析模型(DEA: Data Envelopment Analysis)对国家科研活动进行了研究(Rousseau S. et al., 1997)<sup>[1]</sup>。其主要思想是将不同国家作为研究对象, GDP、科研人员与科研经费作为输入变量, 学术出版物与专利作为输出变量, 由此构建模型来分析评估国家科研投入获得产出的效

组数据,  $C_1 \sim C_{10}$  分别为 0.78, 0.85, 0.92, 0.88, 0.85, 0.69, 0.80, 0.79, 0.92, 0.91, 专家组评价H的学习力为0.82。将H的学习力三级指标得分输入已经得到的学习力评价模型, 输出结果为0.86。这样就获得了该成员的学习力结果。

经过比较, 基于神经网络的学习力评价模型计算的结果与专家组评价的结果基本一致, 相对误差仅为2.3%, 表明该评价模型具有较高的准确度; 同时, 该评价模型避免了专家组预测存在的人为偏好、个人情感等主观影响因素, 特别是在评价人数多、范围广时, 更能显示其优越性。

## 4 结论

学习力是核心竞争力, 是形成创造力的基础。对人们学习力的正确测量和评估, 有助于人们创造力的提升。本文引入具有高度非线性映射功能和很强的自学习能力的神经网络工具, 建立了适合于团队和个人的学习力评价模型, 解决了学习力难以定量分析的问题, 探索了学习力评价的技术途径, 对学习型组织建设以及团队或个人学习力的提升有着积极的意义。

## 参考文献:

- [1] [美] Stephen. Def. 学习力[M]. 常桦译. 延吉: 延边人民出版社, 2003.2-4.
- [2] 张声雄. 二十一世纪管理模式[M]. 上海: 上海科学技术文献出版社, 2000.10-11.
- [3] 黄键. 造就组织学习力[M]. 上海: 上海三联书店, 2003.10-11.
- [4] A.P.Carnevale. America And The New Economy[J]. Training And Development Journal, 1990, (11):31-52.
- [5] L.Thornburg. Accounting For Knowledge[J]. Hr Magazine, 1994, (8):51-56.
- [6] 许学国. 组织学习力提升策略研究[J]. 当代经济管理, 2005, (2):56-58.
- [8] D.Senge. The Learning Organization Made Plain And Simple[J]. Training And Development Journal, 1991, (8):37-44.
- [9] 夏敏. 高校教师组织学习力的提升[J]. 中国高等教育, 2005, (11):98-101.
- [10] 韩力群. 人工神经网络理论、设计及应用[M]. 北京: 化学工业出版社, 2002.121-125.

(责任编辑: 高建平)

收稿日期: 2006-09-10

作者简介: 余昕(1985-), 男, 汉族, 陕西西安人, 北京大学人口研究所硕士研究生, 研究方向为应用统计学、定量方法; 王冬(1982-), 男, 汉族, 湖北襄樊人, 北京大学信息管理系硕士研究生, 研究方向为文献计量学、信息资源管理; 韩楠(1983-), 男, 汉族, 陕西咸阳人, 北京大学信息管理系硕士研究生, 研究方向为信息政策与法规、传播学; 王欣(1984-), 女, 汉族, 湖北武汉人, 清华大学公共管理学院硕士研究生, 研究方向为科技政策与创新、公共政策与管理。

率。还有学者研究了经济发展水平和文化因素对科研产出的影响(Inonu E, 2003)<sup>[2]</sup>。此研究利用“每百万人口的学术出版物数量”与“人均GDP(依照购买力平价)”两项指标对国家进行分类,描述了经济发展水平与科研产出的大致关系。国内学者也对科研投入的效率进行了大量的研究,例如有学者利用我国30个省市的纵列数据(Panel Data),对我国科技投入的技术进步效应进行了实证分析(陈利华等, 2005)<sup>[3]</sup>,结果显示科技投入对技术进步的作用是明显的,由于各区域自身条件的不同,科技投入所产生的技术进步效应是不一样的。

然而,我国学者主要将研究视野集中于国内,缺乏国际间科研投入效率的比较与综合分析,也缺乏对外国纵列数据的模型化研究。本文收集了16个OECD国家的纵列数据,估算了发达国家的科研投入效率,为我国科研投入效率的提高提供了一个参照系,具有一定的政策含义。

## 1 模型建立

### 1.1 数据来源

本项研究所需的数据主要包括两方面:学术指标及经济指标。学术指标主要来自于ISI的Web of Knowledge检索系统。经济指标主要来自于世界三大经济组织和联合国教科文组织的官方网站。通过多方搜集,我们获得了1993~2002年间16个OECD国家的SCI来源期刊论文量、名义GDP、通货膨胀率、购买力平价指数、科研资金投入量、科研人员总量和劳动力总量。

### 1.2 变量选取

SCI收录了世界范围内绝大部分高水平的学术出版物,其权威性举世公认。因此,SCI来源期刊论文量可以作为衡量一国科研实力的重要指标。根据Rousseau.S等人的研究<sup>[4]</sup>,GDP、科研资金投入量、科研人员数量都是一国科研产出的重要解释变量。为了排除人口总量对科研人员数量的影响,我们选用“每千劳动力中科研人员的数量”作为衡量科研人员数量的指标。为了排除通货膨胀和货币购买力对GDP和科研资金投入量的影响,我们所使用的数据均

表1 主要变量摘要统计表

变量	观测量	均值	标准差	最小值	最大值
SCI 论文量 a	160	42365.18	71666.94	1352	327258
GDP b	160	1282317	2124508	45872.95	1.13e+07
科研资金 b	160	28994.38	54353.95	366.14	246186.9
科研人员 b	160	5.645312	2.063399	1.9	10.2
GDP的对数 b	160	13.08446	1.444845	10.73363	16.24212
科研资金的对数 d	160	8.928851	1.736262	5.903016	12.41385
SCI 论文量的对数 d	160	9.731304	1.383573	7.20934	12.6985

a: 某国某一年内在SCI收录期刊上发表的论文总量。

b: 以1995年为基准年计算,扣除了通货膨胀和购买力平价因素,单位为百万美元。

c: 某国某一年每一千劳动力中科研人员的数量。

d: 均取自然对数。

数据来源:世界三大经济组织官方网站,ISI的Web of Knowledge检索系统。

经过了通货膨胀加权处理和购买力平价加权处理。此外,SCI收录的期刊大多是英文期刊,这也许会影响一些非英语国家的论文发表量。因此,笔者将语言因素也引入到模型中,以验证其是否会对论文发表量产生影响。

### 1.3 主要变量摘要统计

### 1.4 模型设定

根据1.1节的分析,我们选取“科研资金投入量(加权值)”、“每千劳动力中科研人员的数量”以及“母语是否为英语”作为“SCI论文量”的解释变量。由于“GDP”与“科研资金投入量(加权值)”高度相关(相关系数为0.82),如果同时加入模型会造成“共线性”的问题,严重影响参数的估计,所以我们剔除了“GDP”变量,设定如下形式的模型:

$$\ln sci_{it} = \beta_0 + \beta_1 \ln expend_{it} + \beta_2 person_{it} + v_i + eng + u_{it}$$

模型各变量的含义如下:  $\ln sci_{it}$  为某国某一年内在SCI期刊上发表的论文总量的自然对数。  $\ln expend_{it}$  为某国某一年科研资金投入量的自然对数,科研资金投入量以“1995年百万美元”为单位,扣除了通货膨胀和购买力平价因素。  $person_{it}$  为某国某一年每一千劳动力中科研人员的数量。  $eng$  (二值变量)表示该国是否属于英语国家,“是”为1,“否”则为0。  $u_{it}$  为第*i*个国家第*t*期的误差项。

## 2 模型估计

### 2.1 估计方法

由于本研究使用的数据是纵列数据(Panel Data),所以需要决定是采用Random Effect模型还是采用Fixed Effect模型。总的来说,如果个体效应与其它解释变量不相关,用Random Effect模型较好;反之,只能使用Fixed Effect模型。在本研究中,Hausman检验拒绝了Random Effect模型( $\chi^2=185.71, Prob > \chi^2 = 0.0000$ ),所以我们决定选取Fixed Effect模型。为此,需要使用带虚拟变量的最小二乘法。

### 2.2 估计结果与解释

由于个体效应与时间效应不是本文研究的重点,下面只列出几个核心解释变量的估计结果。

表2 主要变量估计结果(Fixed Effect Model)

变量	估计值	标准差	z	P> z	95%置信区间
lexpend	.5772991	.0499342	11.56	0.000	.4794298 .6751684
person	.0372001	.0124657	2.98	0.003	.0127673 .0616329
eng	.2475164	.23113	1.07	0.284	-.20549 .7005229
_cons	4.20347	.4079656	10.30	0.000	3.403872 5.003068
R <sup>2</sup> (overall)=0.8934 N=160 X <sup>2</sup> =283.39 Prob>X <sup>2</sup> =0.0000					

注:本表由Stata8处理生成,为了排除序列相关和异方差性对回归的影响,笔者使用了xtregar命令。

根据以上结果看出,本模型的总体拟合系数R<sup>2</sup>(overall)=0.8934,说明一国一年SCI期刊论文量的变异有89.34%可以被此模型解释。通过 $\chi^2$ 检验说明本模型的所有变量是联合显著的( $\chi^2$ 值为283.39, Prob>  $\chi^2 = 0.0000$ )。

通过z值可以看出变量lexpend和person的系数非常显著地异于零。这说明国家对科研活动投入资金,培养科研人

员对科研产出的增加、科研能力的提高有着至关重要的作用。而变量eng的z值仅为1.07, P值高达0.284, 说明eng对lsci的影响不显著。虽然SCI收录的期刊绝大部分是英文期刊, 可一个国家的母语是否为英语对该国SCI期刊论文量的影响不大, 这也从一个侧面反映出英语在全世界的普及程度。

具体说来, lexpend的系数估计值为0.577, 说明根据此模型预测, 如果一个国家的科研资金投入增长1%, 该国SCI期刊论文量在其它条件相同时会增长0.577%。person的系数估计值为0.037, 说明一个国家每千名劳动力中科研人员的数量增加1人, 该国SCI期刊论文量在其它条件相同时会增长3.7%。以上两个数据再次清楚地表明科研资金投入量和科研人员数量在科学技术发展中所起的重要作用。我国经济虽已取得了飞速的增长, 国民受教育水平也大幅提高, 但我国的高级科研人才仍然比较匮乏, 科研教育投入仍然较低, 这些都是制约我国科技进步的重要瓶颈。

### 2.3 有关其它统计模型的讨论

如果不使用纵列数据模型, 而使用时间序列模型, 将会得出错误的结论。下面列出几个核心解释变量的估计结果。

表3 部分变量估计结果(时间序列数据模型)

变量	估计值	标准差	z	P> z	95%置信区间	
lexpend	.7728809	.0253424	30.50	0.000	.7226779	.8230839
person	-.0329463	.0130193	-2.53	0.015	-.0584641	.0074285
eng	.0314552	.1080827	0.29	0.772	-.1826559	.2455662
_cons	3.04382	.1686308	18.05	0.000	2.709764	3.377876
R <sup>2</sup> =0.9160 N=160 F=283.39 Prob>F=0.0000						

注: 本表由Stata8处理生成。使用时间序列数据模型。

尽管使用时间序列模型所得的R<sup>2</sup>=0.916, 高于纵列数据模型的R<sup>2</sup>, 可这种方法得出的结论显然是与事实是相悖的, person的系数为负, 说明科研人员数量越多科研产出越少。

由此可见, 统计模型的选取很关键, 使用错误的统计模型必然会导致错误的结论。

## 3 结论与总结

(1) 科研资金投入与科研人数的增加对科研能力的提

高有显著影响。预测一个国家科研资金投入增长1%, 该国SCI期刊论文量在其它条件相同时会增长0.577%; 一个国家每千名劳动力中科研人员的数量增加一人, 该国SCI期刊论文量在其它条件相同时会增长3.7%。

(2) 母语是否为英语不是影响一个国家科研产出的重要因素。

(3) 统计模型的选取很关键, 使用错误的统计模型必然会导致错误的结论。

## 4 问题及讨论

(1) 由于本研究所需的数据非常多, 经常有一些国家某些年份的科研经费投入、科研人数、通货膨胀率、汇率、购买力平价指数等数据缺失, 使得样本量大为减少, 在一定程度上影响了模型的估计。为了解决这一问题, 笔者尽可能多方收集数据, 尽最大可能补充数据, 但最终只得到了16个国家10年间的的历史数据。

(2) 变量遗漏的问题。例如每个国家对论文发表的激励机制可能存在不同, SCI在不同国家的地位也不尽相同, 这些都会影响某国的SCI期刊论文量。使用纵列数据模型可以在一定程度上解决这些问题, 这也是笔者选用纵列数据的重要原因。

参考文献:

- [1] Rousseau S, Rousseau R. Data Envelopment Analysis as a Tool for Constructing Scientometric [J]. *Scientometrics*, 1997, 40 (1): 45- 56
- [2] Inonu E. The Influence of Cultural Factors on Scientific Production[J]. *Scientometrics*, 2003, 56 (1): 137- 146.
- [3] 陈利华, 杨宏进. 我国科技投入的技术进步效应——基于30个省市跨省数据的实证分析 [J]. *科学与科学技术管理*, 2005(7): 55- 59
- [4] Rousseau S, Rousseau R. The Scientific Wealth of European Nations: Taking Effectiveness Into Account[J], *Scientometrics*, 1998, 42 (1): 75- 87.

(责任编辑: 焱 焱)

# Efficiency of Scientific Input in Developed Countries

Abstract: Based on panel data and regression models, this paper investigates the relationship of scientific input and scientific output of 16 OECD countries. The authors use the account of SCI papers as the indicator of scientific output, scientific expenditure and scientific personnel as the indicator of the scientific input. To eliminate the possible effect of native language on SCI papers, The authors also introduce language into the regression models. Furthermore, to show the importance of model selecting, The authors compare the parameters estimated from different models.

Key Words: scientific input; efficiency; regression; OECD countries; panel data