

基于 PCA - SFA 的医院科研活动效率评价*

钟 华 李海存 安新颖

(中国医学科学院医学信息研究所 北京 100020)

[摘要] 首先在分析医院科研活动效率影响因素的基础上, 从投入和产出两个层面建立医院科研活动评价指标体系; 然后通过主成分分析法对产出指标开展降维, 进而运用随机前沿分析开展效率评价; 最后运用实际案例验证医院科研活动效率模型的准确性和有效性。

[关键词] 医院科研活动效率; 主成分分析; 随机前沿分析

[中图分类号] R - 056 **[文献标识码]** A **[DOI]** 10. 3969/j. issn. 1673 - 6036. 2016. 10. 012

Efficiency Evaluation of Hospital Scientific Research Activities Based on PCA - SFA ZHONG Hua, LI Hai - cun, AN Xin - ying, Institute of Medical Information, Chinese Academy of Medical Sciences, Beijing 100020, China

[Abstract] At first, based on the analysis of factors influencing the efficiency of hospital scientific research activities, the paper sets up the evaluation index system for hospital scientific research activities from aspects of inputs and outputs. Then, by the Principal Component Analysis (PCA), it reduces the dimensionality of the output indexes and conducts efficiency evaluation by Stochastic Frontier Analysis (SFA). At last, it verifies the accuracy and validity of the efficiency model for hospital scientific research activities through practical cases.

[Keywords] Efficiency of hospital scientific research activities; Principal Component Analysis (PCA); Stochastic frontier analysis (SFA)

1 引言

“医疗、教学、科研”是医院的基本职责, 被

称作医院发展的“三驾马车”, 而科研能力是体现一所医院医疗业务水平和发展潜力的重要指标。医院科学研究的重要任务是从基础理论研究、试验研究、临床实践应用等方面入手, 着力解决重大疾病的防患问题, 维护国民健康素质, 进而提高人民的生活质量, 而科研活动效率是科学研究活动成绩与效果好坏的关键。因此, 医院科研效率的高低关系到医院的发展大局。同时, 政府是医院科研资金的主要资助来源, 开展医院科研活动效率评估, 不仅可以为政府对科研投入资金在不同主体间的分配提供参考, 而且也可为医学研究优先领域部署以及科研管理模式的改进提供依据。

[收稿日期] 2016 - 02 - 25

[作者简介] 钟华, 助理研究员, 发表论文多篇; 通讯作者: 安新颖。

[基金项目] 北京协和医学院青年教师培养项目(项目编号: 2014zlgc0753); 协和青年基金资助项目“基于 TOPSIS - Benchmarking 模型的医学学科科技影响力评价研究”(项目编号: 3332015084); 中国医学科学院医学信息研究所中央公益性基本科研业务费课题“中国医院科技影响力评价中的关键问题研究”(项目编号: 15R0111)。

2 随机前沿分析方法在医院效率评价领域的应用

2.1 进行综合效率分析的主要方法

从 20 世纪 80 年代后期开始国内外陆续出现大量关于医院效率评价的研究,从现有资料来看,针对医院多种产出与多种投入之间的关系进行综合效率分析的方法主要有随机前沿分析、数据包络分析、模糊评价、灰色关联分析等,其中,随机前沿分析(Stochastic Frontier Analysis, SFA)是近年科技评价领域发展起来的一种新的分析方法,能较好的测算多投入和多产出之间的关系,已在多个领域的效率评价中应用。

2.2 随机前沿分析方法

随机前沿分析由 Aigner 等^[1]、Battese 等^[2]于 1977 年分别提出,已经广泛地应用于国外医学效率评价领域,如 Wagstaff (1989) 利用随机前沿分析方法研究了 49 家西班牙医院的效率^[3]。Linna (1998) 基于随机前沿方法分析在医院效率评价中数组数据模型和横断面数据模型的特点和优势,还对医院改革与芬兰医院效率变化的关系进行了研究^[4]。Gerdtham (1999) 利用随机前沿模型的产出函数对瑞典 1991 年实施的补偿制度改革对医院效率的影响及其程度进行检验^[5]。国内也有学者也采用随机前沿分析方法对医院评价开展研究,如吴明等(2000) 利用随机前沿成本函数评价了威海市县级以上医院的经济效率^[6];姚红等(2003) 建立了上海市医院成本效率的随机前沿成本模型^[7];王伟成等(2005) 运用随机前沿成本模型对全国东部、中部和西部 3 个地区 60 家中医院的效率进行了评价^[8];刘启贵等(2005) 应用随机前沿分析法研究低效率值间的关系,分析了铁道部所属 101 家二级以上医院的成本状况^[9]。

3 研究方法和模型

3.1 概述

本文将医院科研过程中科研成果的产出过程

视为生产过程,采用随机前沿生产函数方法进行估计。由于随机前沿生产函数的应变量和解释变量分别对应医院的产出指标和投入指标,因此对指标的确定即是对函数变量的确定。随机前沿生产函数要求产出 Y 为单一变量,而实际情况往往与此不符;同时由于医院科研活动效率是涉及多个因素的多指标问题,会导致分析的复杂性和指标间的多重相关性等;而且,医院科研活动效率评价涉及的各个指标间常还经常存在一定的相关性,这也使评价数据在一定程度上反映的信息有所重叠。针对此种情况,本文采用主成分分析(Principal Component Analysis, PCA)方法将投入和产出指标进行多指标综合,对原始的评价指标体系进行化简和降维处理,用少数指标代替原始多指标。

3.2 PCA 的基本原理和模型

主成分分析是处理降维的一种统计方法,它借助于正交变换,将其分量相关的原随机向量转化成其分量不相关的新随机向量,这在代数上表现为将原随机向量的协方差阵变换成对角形阵,在几何上表现为将原坐标系变换成新的正交坐标系,使之指向样本点散布最开的 P 个正交方向,然后对多维变量系统进行降维处理,再通过构造适当的价值函数,进一步把低维系统转化成一维系统^[10]。

3.3 SFA 的基本原理和模型

随机前沿分析方法自 20 世纪 70 年代末提出后被广泛用于评价技术效率。SFA 有多个分析模型,本文借鉴 Battese 等的模型^[11],采用对数型柯布-道格拉斯生产函数分析医院科研活动效率。SFA 模型如下:

$$\ln(y_i) = \beta_0 + \beta_1 \ln(K_{it}) + \beta_2 \ln(L_{it}) + \dots + \varepsilon_{it} \quad (1)$$

式中: y_i 为医院科研活动的产出指标; k_{it} 、 l_{it} 等为投入指标值, F 标 i 为医院代号; t 为时间; β_0

为截距项； β_1 、 β_2 等为待估计参数。

$$\varepsilon_{it} = v_{it} - u_{it} \tag{2}$$

式中： ε_{it} 为误差项，由相互独立的两部分组成； v_{it} 服从 $N(0, \sigma^2 v)$ ； u_{it} 为非负数，反映在第 t 时期应仅仅影响第 i 个医院的随机因素， u_{it} 服从正半部的正态分布 $N(u, \sigma^2 u)$ 。

$$TE_{it} = \exp(-u_{it}) \tag{3}$$

式中： TE_{it} 为样本中第 i 个医院在 t 时期的技术效率水平。当 $u_{it} = 0$ 时， $TE_{it} = 1$ ，即处于技术效率状态；当 $u_{it} > 0$ 时， $0 < TE_{it} < 1$ ，即处于非技术效率状态。

$$\gamma = \frac{\sigma_u^2}{\sigma_v^2 + \sigma_u^2} \tag{4}$$

式中： γ 为待估参数，可以利用 γ 判断是否适用于 SFA 分析模型。在 $r = 0$ 时，误差项 $\varepsilon_{it} = v_{it}$ ，此时所有待检测的生产点都在前沿曲线上，可以直接使用最小二乘法 (Method of Ordinary Least Squares, OLS)，不需要进行 SFA 模型分析^[12]。

4 实证分析

4.1 样本和投入产出变量的选择

根据国家卫生和计划生育委员会提供的数据^[13]，实证分析选取北京和上海 20 家三甲综合医院作为评价对象，抽取 2011 年的数据。在投入产出变量的选择方面，经相关研究归纳后发现，目前科技投入产出研究中最常用的指标基本包括人员、项目、论文等方面。在本研究中，投入指标选择人员投入和项目投入两个方面的指标，既考虑质量因素也考虑数量因素；产出指标选择主要直接产出的论文指标，包括 SCIE 收录论文数量、中文核心期刊论文数量、SCIE 总被引频次、中文核心期刊论文总被引频次，见表 1。评价数据来源于中国医学科学院医学信息研究所科技评价基础数据库。国内论文以中国医学科学院医学信息研究所 SinoMed 数据库收录的国内出版的科技期刊为统计源，国际论文以 SCIE 数据库收录的文章为统计源。

表 1 投入产出指标

1 级指标	2 级指标	指标说明
科研投入指标	科研人员总数	具备一定科学理论知识并从事科学研究工作的人员总数
	国家自然科学基金项目数量	国家自然科学基金项目数量
科研产出指标	SCIE 收录论文数量	以第一作者或通讯作者发表的 SCI 论文数量
	中文核心期刊论文数量	以第一作者发表的中文核心期刊论文数量
	SCIE 论文总被引频次	以第一作者或通讯作者发表的 SCI 论文的总被引频次
	中文核心期刊论文总被引频次	以第一作者发表的中文核心期刊论文的总被引频次

4.2 数据预处理

对原始数据的描述统计，见表 2。运用 SPSS 18.0 软件和主成分分析法，根据主成分特征值大于 1 的原则，对产出部分的指标数据进行主成分提取，提取结果，见表 3。共提取出一个主成分，对总方差的贡献率为 76.42%，说明该主成分可以反映产出指标原始数据中 76.42% 的信息。

表 2 指标变量描述统计

项目	最小值	最大值	均值	标准差
科研人员总数	1 105	3 203	1 906.85	525.523
自科基金数	6	70	25.90	15.286
SCI 论文数	34	294	147.55	75.051
中文核心论文数	289	1 192	572.65	199.917
SCI 总被引次数	74	3 117	1 104.20	698.147
中文核心论文被引次数	122	484	265.80	96.297

表 3 产出指标的主成分提取结果

成分	初始特征值			提取平方和载入		
	合计	方差百分比 (%)	累积百分比 (%)	合计	方差百分比 (%)	累积百分比 (%)
1	3.057	76.421	76.421	3.057	76.421	76.421
2	0.752	18.799	95.220	-	-	-
3	0.151	3.771	98.991	-	-	-
4	0.040	1.009	100.000	-	-	-

通过主成分载荷矩阵中的数据及主成分对应的特征值，可以计算得到主成分中每个指标所对的系数，由此可以得到产出综合指标值 (Y) 与各产出指标间的关系，见表 4。

$$Y = 0.5022O_1 + 0.4936O_2 + 0.4993O_3 + 0.5056O_4$$

表 4 产出主成分载荷矩阵及特征向量

指标	成分 1
O_1 : SCIE 收录论文数量	0.878
O_2 : 中文核心期刊论文量	0.863
O_3 : SCIE 论文总被引频次	0.873
O_4 : 中文核心期刊论文总被引频次	0.884

4.3 分析结果

根据以上数据和模型，采用 Frontier (Version4.1) 软件进行效率估算。Frontier 软件可对模型各参数和总的效率值进行估算，是进行 SFA 模型数据处理的常用软件。从表 5 可见， $\gamma = 0.965$ ，且在 10% 显著水平下具有统计显著性，说明式 (1) 中的误差项复合结构显著，因此使用 SFA 模型对医院科研效率进行测算是必要的。极大似然估计为 7.9669554，单边误差检验值为 1.0855728。

表 5 SFA 模型参数分析结果

待估参数	相关系数、参数	标准差	t-检验值
β_0	-3.4655	2.1130	-1.6400
β_1	1.3044	0.3184	4.0966
β_2	0.2350	0.0968	2.4268
δ^2	0.0804	0.0467	1.7219
γ	0.9651	0.1232	7.8352

说明投入因子对产出的影响是正向的，其中， $\beta_1 = 1.3044$ 说明研发人员投入每增加 1%，会带来创新产出增加 1.3044%； $\beta_2 = 0.2350$ ，说明研发项目投入每增加 1%，会带来医院科研产出增加 0.2350%。科研人员的投入产出弹性要高于科研项目的投入产出弹性，这一方面说明医院科研活动的产出的增加主要是靠科研人员投入拉动的，另一方面说明科研基金项目的投入强度还需进一步强化，因此，要提高医院科研活动效率，应重视对高素质科研人员以及科研项目的投入强度。20 家三甲医院科研活动效率值，见表 6。可看出在参加测评的 20 家三甲医院中，医院 4 的科研活动效率最高，为 97.1%；医院 1 的科研活动效率最低，为 59.22%。

表 6 医院科研活动效率分析值

医院	值	医院	值
1	0.5922	11	0.8771
2	0.9481	12	0.7564
3	0.7196	13	0.5959
4	0.9710	14	0.8207
5	0.7129	15	0.8459
6	0.9108	16	0.7892
7	0.6480	17	0.7916
8	0.8068	18	0.9315
9	0.8820	19	0.6567
10	0.9149	20	0.9579

5 结语

在科技评价领域，近年来 SFA 方法得到了深入发展，其不仅可以测算样本总体及其个体中的技术效率水平状态，而且还能分析影响技术效率的因素。但由于医院科研活动的复杂性，精确的产出投

由此可知， $\beta_0, \beta_1, \beta_2$ 均通过了显著性检验，

入指标很难界定,目前的实证研究只能采取可获指标进行分析。本文研究通过影响医院科研活动的因素确定科研效率的指标体系,运用主成分分析法对研究数据进行初步处理,并用随机前沿分析方法进行医院科研活动效率评价,通过 20 家三甲医院数据进行实证分析,应用结果说明了模型的可行性,并得出目前科研项目投入仍然是我国医院科研效率提高的一个关键因素,为进一步合理规划医院科研资源提供依据。

参考文献

- Aigner D J, Lovell C A K, Schmidt P. Formulation and Estimation of Stochastic Frontier Production Function Models [J]. *Journal of Econometrics*, 1977, 6 (1): 21 - 37.
- Battese G E, Corra G S. Estimation of a Production Frontier Model with Application to the Pastoral Zone of Eastern Australia [J]. *Australian Journal of Agricultural Economics*, 1977, 21 (3): 169 - 179.
- Rosko M D, Mutter R L. Stochastic Frontier Analysis of Hospital Inefficiency - A Review of Empirical Issues and an Assessment of Robustness [J]. *Medical Care Research and Review*, 2008, 65 (2): 131 - 166.
- Linna M. Measuring Hospital Cost Efficiency with Panel Data Models [J]. *Econometrics and Economics*, 1998, 7 (5): 415 - 427.
- Gerdtham U G. Internal Markets and Health Care Efficiency: A Multiple - Outputstochastic Frontier Analysis [J]. *Health Economics*, 1999, 8 (2): 151 - 164.
- 吴明, 李曼春, 侯建林, 等. 随机前沿成本函数方法在医院经济效率评价中的应用 [J]. *中华医院管理杂志*, 2000, 16 (8): 507 - 509.
- 姚红, 胡善联, 曹建文. 上海市 45 家医院供给的技术效率评价 [J]. *中国医院管理*, 2003, 5 (5): 9 - 11.
- 王伟成, 曾武. 随机前沿成本模型在中医院技术效率评价中的应用 [J]. *中华医院管理杂志*, 2005, 21 (5): 333 - 336.
- 刘启贵, 宋桂荣, 禹海波, 等. 随机前沿方法在评价医院效率中的应用 [J]. *中国卫生统计*, 2005, 22 (5): 303 - 305.
- 李清华, 郭耀煌. 主成分分析用于多指标评价的方法研究——主成分评价 [J]. *管理工程学报*, 2002, 16 (1): 39 - 43, 3.
- Battese G E, Coell T J. A Model of Technical Inefficiency Effects in Stochastic Frontier Production for Panel Data [J]. *Empirical Economics*, 1995, (20): 40 - 43.
- 王建宏, 陆志峰, 高岩波, 等. 基于 PCA - DEA 和 PCA - SFA 的大型医院绩效评价 [J]. *数学的实践与认识*, 2011, 41 (15): 36 - 44.
- 国家卫生和计划生育委员会. 医院名录查询 [EB/OL]. [2016 - 03 - 06]. <http://www.nhfp.gov.cn/>.

(上接第 56 页)

- 邱均平. 信息计量学 [M]. 武汉: 武汉大学出版社, 2007.
- 袁晓园, 华薇娜. 基于共词分析的我国医学信息学国际发文热点研究 [J]. *医学信息学杂志*, 2014, 35 (4): 8 - 14
- 赵健. 生物医药研发外包初探 [J]. *经营管理者*, 2014, (1): 113 - 114.
- 张翔, 王璨. 国内科技企业公司治理研究的主题分布及热点——基于共词分析法 [J]. *沈阳大学学报*, 2014, 16 (1): 14 - 19.
- 薛薇. 基于 SPSS 的数据分析 [M]. 北京: 中国人民大学出版社, 2006.
- 宋茂海, 李东方. 基于共词分析的国内生物信息学热点领域研究 [J]. *生物信息学*, 2014, 12 (1): 46 - 51.
- 刘跃. 我国 CRO 公司与国际 CRO 公司的管理对比分析 [J]. *中国卫生产业*, 2015, (23): 193 - 195.