

基于结构方程的学术期刊评价研究¹

俞立平 潘云涛 武夷山

中国科学技术信息研究所 北京 100038

摘要：本文利用中国科学技术信息研究所医学学术期刊数据，将指标分为影响力、时效性、期刊特征三个一级指标，它们之间存在两两相关关系，在此基础上建立结构方程进行模拟，计算出3个一级指标值并采用主观赋权法进行期刊的综合评价，取得了较好的结果。结构方程通过模拟期刊指标之间的错综复杂的关系，可以对一些隐含指标进行估计，还可以进行指标的筛选，为科技评价提供了一种新的思路。同时要注意的是，基础数据的完备性和模型建立的科学性对评价结果的影响很大。

关键词：学术期刊 评价 结构方程模型

1 引言

学术期刊是一个国家科技发展水平的重要窗口，是知识创新、科技成果转化成为生产力的重要桥梁，在推动社会科技进步方面发挥着不可替代的作用。期刊评价是文献计量学研究的重要组成部分，它通过对学术期刊的发展规律和增长趋势进行量化分析，揭示学科文献数量在期刊中的分布规律，为优化学术期刊的使用提供重要参考，同时可以提高学术期刊的内在质量，促进学术期刊的健康成长和发展。美国著名情报学家加菲尔德博士在20世纪60年代对期刊文献的引文进行了大规模统计分析，得到了大量被引用文献集中在少数期刊上，而少量被引用文献散布在大量期刊中的结论，这可以被认为是国外期刊评价理论的起源。

学术期刊评价是一项复杂的系统工程，牵涉到评价原则、指标选取、数据归一化、评价方法选择等诸多方面，国内外学者在该领域进行了大量广泛的研究。从评价方法的角度，Weiping Yue、Concepcion S. Wilson (2004)^[1]利用结构方程的原理建立了一个期刊影响力的分析框架，但没有进行实证。苏新宁 (2008)^[2]采用指标体系赋权进行中国人文社会科学期刊的评价。邱均平、张荣等 (2004)^[3]提出了期刊评价指标体系的三维层次结构图，并利用灰色关联法进行评价。庞景安、张玉华等 (2000)^[4]及李凯扬、贾玉萍 (2005)^[5]利用层次分析法对期刊进行评价。王小唯、杨波等 (2003)^[6]将期刊以往状态的评价结果作为各期刊基础条件的一种度量，再运用数据包络分析方法 (DEA) 测算出它们的二次相对评价价值。李修杰、陈景武 (2006)^[7]运用判别分析法建立的期刊评估指标体系。王玖、徐天和 (2003)^[8]运用秩和比法进行医学学术期刊学术质量综合评价。陈汉忠 (2004)^[9]应用主成分分析对学术期刊进行评价。凌春艳、莫琳 (2004)^[10]提出自然科学学术期刊质量指标体系的属性数学综合评价模型并进行了评价。李继晓、蔡成瑛 (2006)^[11]对各种核心期刊的评价方法进行了介绍和分析，认为选择核心期刊的评价方法并不是一件一劳永逸的事，只能通过不断实践、比较、分析，才能使核心期刊的评价日臻完善。从目前的科技评价现状看，广大学者采用各种不同的评价方法进行学术期刊评价，在指标选取、权重赋值、数据处理、评价结果的适用等方面取得不少进展，但至今难以产生较公认的评价结果。

对学术期刊进行评价离不开各种各样的主客观评价方法，目前国内外综合评价方法有数十种之多，根据权重确定方式结合评价原理，可以分为三大类：第一类是主观评价法，其基本原理是进行指标主观赋权，然后将数据标准化后加权汇总，如专家会议法、德尔菲法、层

¹国家十一五支撑计划项目 (2006BAH03B05)；国家自然科学基金资助 (70673019)

作者简介：俞立平 (1967-) 男，江苏泰州人，博士，中国科学技术信息研究所博士后，扬州职业大学教授，主要从事信息经济、科学计量领域的教学科研工作。Email: chinayangzhou@yahoo.com.cn

次分析法等等。第二类是客观评价法，包括两种，一种是采用客观赋权法确定指标权重，然后进行加权汇总，如熵权法、变异系数法、复相关系数法等；另一种是不需要赋权的系统方法，如主成分分析法、因子分析法、TOPSIS等。第三种是主客观相结合的赋权法，首先采用主观赋权方法确定权重，然后采用系统方法进行综合处理，如ELECTRE法、模糊综合评价法、PROMETHEE等等。指标体系综合评价方法存在的主要问题是，针对同一评价对象，选取相同的指标，采取同样的数据，不同评价方法得出的评价结果不一致，结果难以得到公认，必须采用全新的思路进行学术期刊评价的探索。

结构方程是一种模拟社会经济系统复杂关系的一种方法。学术期刊评价体系也是一个复杂系统，任何期刊都包括影响力、期刊特征、新颖度等关键要素，他们之间存在着互相依存的复杂关系，采用结构方程可以对学术期刊指标与要素间的复杂关系进行模拟，并对这些要素进行测度，同时还可以进行指标的筛选，这就为进一步评价打下了良好基础。本文以中国科学技术信息研究所的医学期刊评价为例，采用结构方程进行指标筛选和评价。

2 研究方法

2.1 结构方程模型 SEM

结构方程式模型（Structural Equation Modeling，SEM）习惯上也叫做因果关系模型、协方差结构模型，或者直接称为 LISREL 模型。它是一种建立、估计和检验因果关系模型的多元统计分析技术。它包含了回归分析（multiple regression）、因子分析（factor analysis）、路径分析（path analysis）和多元方差分析（multivariate analysis of variance）等一系列多元统计分析方法，是一种非常通用的、线性的、借助于理论进行假设检验的统计建模技术。该模型由 K.Joreskog 与其合作者在 70 年代提出并逐步改进和完善，到 90 年代初期开始得到了广泛的应用。随着 SEM 理论和分析软件的不断发展和完善，结构方程式模型不仅在市场研究中成为分析数据、检验理论的好工具，而且在心理学、社会学、计量经济学、管理学、行为科学和传播学等领域都得到了广泛的应用。

结构方程式模型本质上是利用联立方程求解。人们希望的是模型拟合的再生数据尽可能接近原始数据，如果真是这样的话，假设的因果关系结构与变量间的相互关联模式就是拟合的或是一致的。结构方程模型中方程分为结构模型（内部模型）与测量模型（外部模型）。结构模型为：

$$\eta = B\eta + \Gamma\xi + \zeta \quad (1)$$

结构模型主要测量各个隐变量之间的因果关系，其中 ξ 为模型中的外生隐变量， η 为模型中的内生隐变量， Γ 和 B 为结构系数矩阵，分别表示结构模型中外生隐变量 ξ 对内生隐变量 η 的影响以及 η 之间的互相影响， ζ 为模型中的残差。测量模型为：

$$X_{\eta} = \pi_{\eta}\eta + \varepsilon \quad X_{\xi} = \pi_{\xi}\xi + \delta \quad (2)$$

测量模型主要测量隐变量与显变量之间的对应关系，两个方程分别表示内、外生隐变量与其对应的显变量之间的关系， π_{η} 、 π_{ξ} 分别为各自的测量系数矩阵， ε 、 δ 分别为各自的残差矩阵。

一般采用路径图（path diagram）的形式表示结构方程式模型，这是最简单、最直观的描述模型的方法，研究人员可以借助路径图直接和明了地将变量之间的关系以图形的方式表现出来。流行的 AMOS7.0 软件都可以将路径图直接转化为建模的方程，并将分析结果直接标识在图中。

结构方程式模型假定在一组隐变量中存在因果关系，这些隐变量可以分别用一组可观测

的变量表示。假设的模型通常包括某个基本线性回归模型和很多观测变量，而这个基本的线性回归模型应该是一组隐变量的结构关系模型。这一组隐变量分别是那些观测变量中某几个的线性组合。在技术上，通过验证观测变量之间的协方差，可以估计出这个基本线性回归模型的系数值，从而在统计上检验所假设的模型对所研究的过程是否合适，也就是检验观测变量的协方差矩阵与模型拟合后的引申协方差矩阵的拟合程度，如果证实所假设的模型合适，就可以得出结论：我们所假设的隐变量之间的关系是合理的。从一定意义上讲，结构方程模型是一种证实性（confirmatory）技术，而不是一种探索性（exploratory）技术。

2.2 学术期刊评价路径图

本文对中国科学技术信息研究所的医学类学术期刊进行评价，选取的指标有总被引频次、扩散因子、学科影响指标、学科扩散因子、影响因子、基金论文比、平均引文数、被引半衰期、即年指标、引用半衰期、平均作者数、地区分布数、海外论文比共 13 个指标，根据指标的内涵将其分为三大类，如表 1 所示。

表 1 指标及分类

一级指标	二级指标	变量名	一级指标	二级指标	变量名
	总被引频次	X1		平均作者数	Z1
	扩散因子	X2		地区分布数	Z2
影响力 X	学科影响指标	X3	期刊特征 Z	海外论文比	Z3
	学科扩散因子	X4		基金论文比	Z4
	影响因子	X5		平均引文数	Z5
	被引半衰期	Y1			
时效性 Y	即年指标	Y2			
	引用半衰期	Y3			

期刊影响力是与期刊被引用有关的指标，包括总被引频次、扩散因子、学科影响指标、学科扩散因子、影响因子 5 个指标。期刊的时效性是指期刊的生命力和活跃程度，是期刊与时间相关的指标，包括被引半衰期、即年指标、引用半衰期 3 个指标，用来反映期刊的新颖程度，被引半衰期和引用半衰期越短的期刊，其时效性越好，当年就被引较多的期刊，其时效性也越好。期刊特征是期刊自身的一些参数，与被引及时间因素无关，包括平均作者数、地区分布数、海外论文比、基金论文比、平均引文数 5 个指标，这些指标从各个不同侧面反映了期刊的特点。

影响力、时效性、期刊特征三个变量是隐变量，是不能直接度量的，他们之间是存在着两两相关关系。影响力和时效性之间存在着互为因果的关系，因为影响力较大的期刊，一般都注重其时效性，时效性好的期刊，容易受到欢迎，其影响力一般也较大，容易被引用。影响力和期刊特征之间也存在着互为因果关系，影响力大的期刊，其作者数量和地区更为发散，容易吸引基金论文，作者参考文献数量更多，同样，期刊特征指标越高的期刊，其影响力一般也越大。时效性和期刊特征之间也存在着互为因果的关系，时效性强的期刊，更容易吸引不同地区作者投稿，也容易吸引基金论文和参考文献较多的论文，期刊特征指标越高的期刊，一般质量也较高，其时效性也越好。

此外，总被引频次和影响因子存在着相关关系，学科影响指标与学科扩散因子也存在着相关关系，据此分析，期刊各显变量与隐变量之间关系的路径图如图 1 所示。

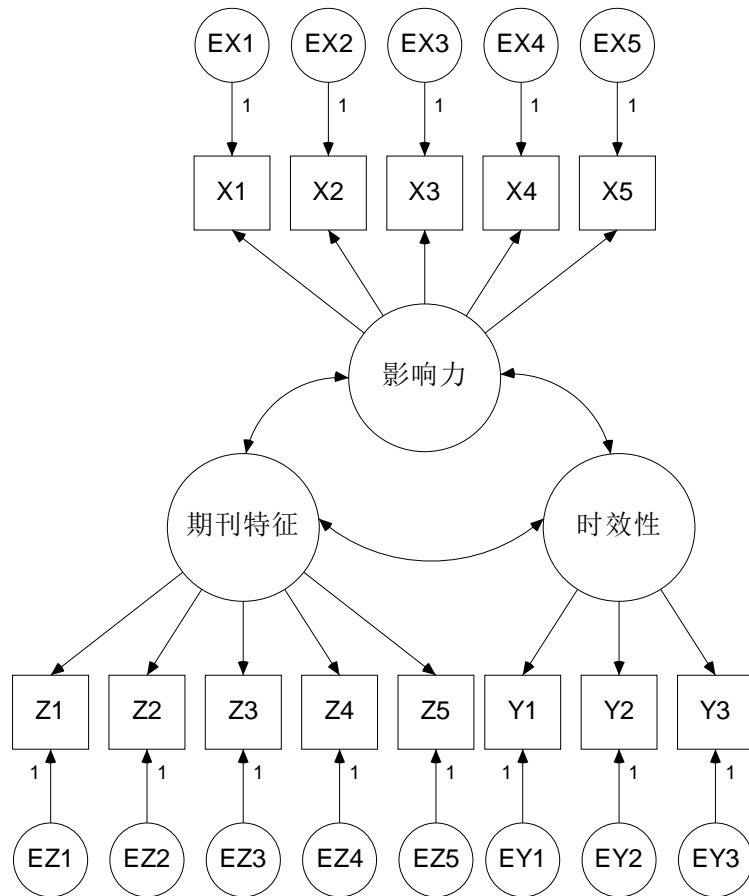


图 1 学术期刊评价路径图

3 变量和数据

本文数据来自于中国科学技术信息研究所 CSTPC 数据库，以医学类期刊为例进行分析。中国科学技术信息研究所从 1987 年开始对中国科技人员在国内外发表论文数量和被引情况进行统计分析，并利用统计数据建立了中国科技论文与引文数据库，同时出版《中国学术期刊引证报告》。本文数据是 2006 年的医学学术期刊数据，共 518 种医学期刊。

表 2 变量描述统计量

变量名	内容	均值	最大值	最小值	标准差
X1	总被引频次	794.02	5805.00	17.00	811.65
X2	扩散因子	31.95	78.43	3.67	15.74
X3	学科影响指标	0.54	1.00	0.02	0.26
X4	学科扩散因子	5.14	31.20	0.17	3.81
X5	影响因子	0.44	1.86	0.06	0.28
Y1	被引半衰期	4.27	10.00	1.75	1.13
Y2	即年指标	0.05	0.33	0.00	0.04
Y3	引用半衰期	5.63	8.52	2.78	0.86
Z1	平均作者数	4.04	8.16	2.04	0.87
Z2	地区分布数	22.61	31.00	2.00	5.81
Z3	海外论文比	0.02	0.78	0.00	0.05
Z4	基金论文比	0.26	0.97	0.01	0.19
Z5	平均引文数	9.25	34.72	3.12	3.88

数据来源：中国科学技术信息研究所 CSTPC
数据库 2006 年数据

由于要对期刊进行评价，所以必需对数据进行标准化处理，每项指标最大值设为 100，然后按比例进行调整。此外，被引半衰期和引用半衰期是两个反向指标，必须进行适当处理，方法是用 100 减去其标准化后的结果，这样就变成了正向指标。

4 结构方程评价

4.1 结构方程的估计

在模型中对应于每一个隐变量有几个测量指标，规定其中的一个系数等于 1，其效果相当于规定隐变量的度量单位分别与对应的测量指标相同。每个观测变量都存在测量误差（EX1-EX5, EY1-EY3, EZ1-EZ5）。利用 AMOS7.0 软件采用极大似然估计对结构方程进行估计，发现扩散因子 X2、引用半衰期 Y3 的系数为负，并且统计检验是显著的，不符合常识，因此删除这两个指标，重新进行模拟，发现地区分布数 Z2 的系数为负，并且没有通过统计检验，将其删除后新模型共 10 个内生变量，13 个外生变量，模型是过度识别的，求解后所有变量都通过了统计检验，结果如图 2、表 3 所示。

表 3 结构方程求解结果

变量	估计值	标准差	临界比	概率
X1<---影响力	1			
X3<---影响力	0.588	0.093	6.32	***
X4<---影响力	0.415	0.057	7.346	***
X5<---影响力	1.023	0.052	19.635	***
Z1<---期刊特征	0.62	0.066	9.381	***
Z3<---期刊特征	0.466	0.041	11.279	***
Z4<---期刊特征	1.326	0.126	10.511	***
Z5<---期刊特征	1			
Y1<---时效性	0.281	0.064	4.39	***
Y2<---时效性	1			

结构方程的卡方检验值为 420.06, GFI=0.868, IFI=0.801, CFI=0.800, RMSEA=0.159, 拟合程度良好，说明本模型总体上是可行的。之所以没有达到理想的拟合水平，一是文献计量学指标本身的缺陷所致，如影响因子存在自引、反引现象，部分作者为了增加发表的可能性人为增加基金项目，一些论文作者署名比较随意等等。二是样本量虽然足够，但一些指标并不完全服从正态分布，此外可能还有统计误差、指标缺失等问题。

扩散因子是统计源期刊内某期刊每被引 100 次所涉及的期刊数，医学期刊由于专业性较强，其他学科引用较少，同时由于医学期刊分类很细，除了一些综合性医学期刊扩散因子较高外，各子学科期刊扩散因子一般也不高，因此，用扩散因子作为衡量医学期刊影响力的指标是不合适的。

引用半衰期是用来反映作者应用文献的新颖程度的，医学期刊中各子学科学术水平发展速度并不相同，一些发展较快的子学科引用半衰期可能较短，但是一些发展较慢的子学科（成熟学科）引用半衰期可能相对较长，同样不能用来作为衡量期刊时效性的指标。

地区分布数是期刊刊载论文所涉及的地区数，医学水平在我国发展并不均衡，北京、上海及沿海经济发达地区医学相对较发达，这些地区发表的论文相对较多，而一些经济欠发达地区医学水平相对薄弱。不同医学子学科水平的地区差距更大，因此用地区分布数来衡量期

刊的期刊特征也是不合适的。

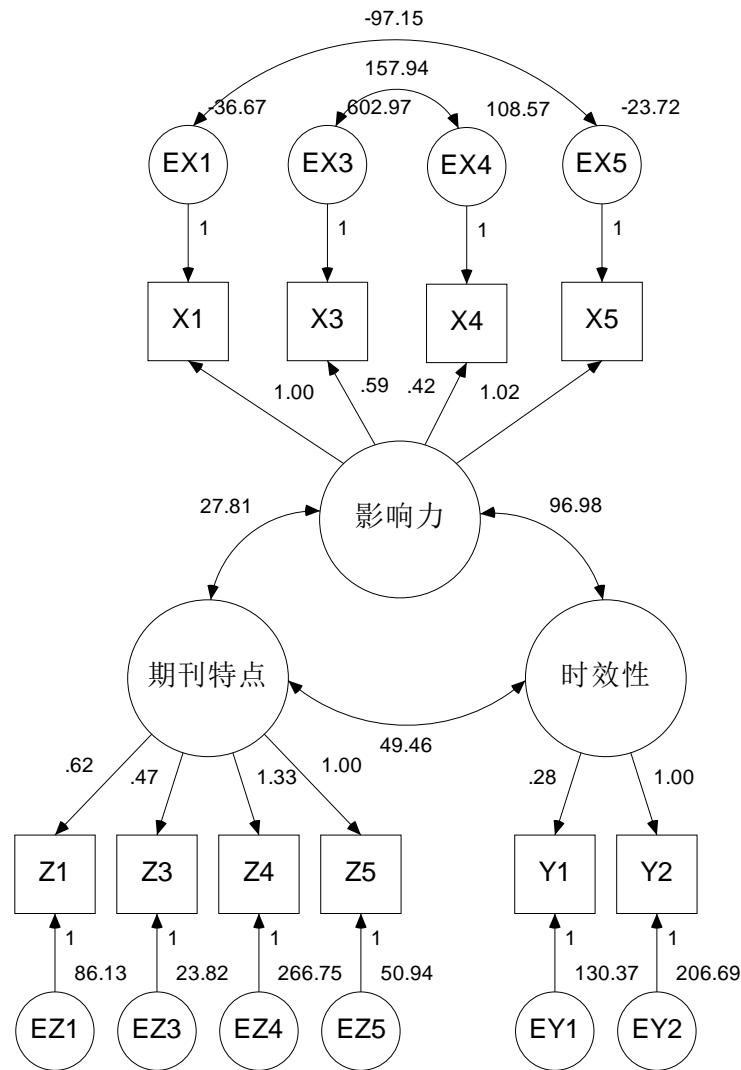


图 2 结构方程模拟结果

4.2 医学期刊的评价

将表 3 中的路径系数分别进行标准化处理，得到期刊影响力、时效性、期刊特征指数三个一级指标的计算公式：

影响力=0.33×总被引频次+0.19×学科影响指标+0.14×学科扩散因子+0.34×影响因子

时效性=0.22×被引半衰期+0.78×即年指标

期刊特征指数=0.18×平均作者数+0.14×海外论文比+0.39×基金论文比+0.29×平均引文数

接下来的影响力、时效性、期刊特征三个一级指标的权重是无法根据结构方程进行计算的，本文采取专家打分法确定，通过专家会议法设定影响力权重为 0.5，时效性和期刊特征权重各为 0.25，对医学期刊进行估计，表 4 是前 20 种医学期刊的评价结果。

表 4 前 20 种期刊评价结果

期刊名	影响力	时效性	期刊特征	总评分	排序
中华护理杂志	94.89	52.92	19.06	65.44	1
中华儿科杂志	73.44	52.30	30.92	57.53	2

WORLD JOURNAL OF GASTROENTEROLOGY	56.43	37.70	72.19	55.69	3
中华内分泌代谢杂志	45.26	78.91	41.67	52.78	4
中华心血管病杂志	61.07	49.81	37.43	52.34	5
中华结核和呼吸杂志	62.84	50.65	31.75	52.02	6
中华医学杂志	65.45	35.20	39.03	51.28	7
中华流行病学杂志	61.24	36.46	45.40	51.09	8
中国中西医结合杂志	57.93	44.06	43.77	50.92	9
ACTA PHARMACOLOGICA SINICA	42.97	39.65	76.12	50.43	10
中华医院管理杂志	47.11	85.46	19.80	49.87	11
中华骨科杂志	59.19	49.94	30.02	49.58	12
中国组织工程研究与临床康复	68.49	20.70	34.11	47.95	13
药学报	51.05	42.04	46.27	47.60	14
中华肿瘤杂志	64.31	24.70	36.86	47.54	15
中华外科杂志	60.98	37.84	29.77	47.39	16
中国危重病急救医学	46.96	46.51	46.98	46.85	17
中华内科杂志	59.89	37.48	29.92	46.80	18
中华放射学杂志	64.47	29.93	28.22	46.77	19
癌症	52.04	38.39	43.80	46.57	20

5 结论与讨论

5.1 用结构方程可以进行辅助评价

与其他评价方法相比，结构方程模型是一种革命性的方法，它是在模拟评价对象自身内部各要素复杂关系的基础上进行的评价，可以计算模拟出若干不可观测的要素，而这些要素中往往有许多就是一级指标，再根据这些指标值直接进行对比或赋权后进行加权汇总评价都可以。

5.2 结构方程可以进行指标筛选

结构方程的建立、调整是一个动态过程，在这个过程中如果发现问题，必须进行调整，指标筛选也是调整的重要内容之一。在科技评价中，若指标过多，数据获得就相对困难，评价成本增高；指标过少，又不能提供全部信息。在一些科技评价中，动辄几十个指标，很难用客观方法对指标进行筛选，而结构方程提供了一种科学的指标筛选方法。需要说明的是，采用结构方程进行指标筛选时，一定要剔除符号不对但是通过了统计检验的指标，并对该指标进行深入分析，而对于一些没有通过统计检验但符号正确的指标，建议根据具体情况进行取舍。

5.3 基础数据的完备性对评价结果影响很大

评价离不开指标的选取，如果数据不准确或错误，当然得不到满意的结果，所谓“输入是垃圾、输出也是垃圾”。在科技评价中，有时有些数据的获取是十分困难的，人们也许只能找到有限的一些指标，这些指标并不能反映科技评价中某个要素的全貌，在这样的情况下，无论采用什么评价方法都得不到满意的评价结果，当然采用结构方程评价也不例外。

5.4 模型建立偏差将影响结果

结构方程本质上是一种主客观相结合的评价方法，因为各变量之间的关系是人们主观定义的，比如期刊影响力、时效性、期刊特征等指标的选择，影响力、时效性、期刊特征指标的关系等，都体现了人的主观能动性。也正因为如此，如果结构方程各要素之间的关系设置不当，比如由于某种原因人们对一些要素之间的固有关系没有充分认识或者是错误认识，都容易导致计算偏差，得到有偏的结果。

5.5 评价方法选择与结果修正

迄今为止，已经有几十种多属性评价方法，在评价方法的选择中，除了考虑评价目的、评价方法原理、区分度、灵敏度等因素外，还有一个很重要的原因就是必须符合客观实际，比如若存在某种专家一致公认的最好的期刊，但是在用某种评价方法进行评价时，该期刊没有排在第一，此时要分析原因，比如指标选取是否得当？数据处理是否正确？评价方法选择是否存在问题？等等，在此基础上进行调整。

总之作为一种新的尝试，结构方程模型为学术期刊评价提供了一种新的思路，通过增加指标个数和期刊数量，优化结构方程模型可以提高评价的精度科学性。

参考文献：

- [1] Weiping Yue, Concepcion S. Wilson. Measuring the citation impact of research journals in clinical neurology: a structural equation modeling analysis[J]. *Scientometrics*, 2004 (3): 317-334
- [2] 苏新宁. 构建人文社会科学学术期刊评价体系[J]. *东岳论丛*, 2008 (1): 35-42
- [3] 邱均平、张荣等. 期刊评价指标体系及定量方法研究[J]. *现代图书情报技术*, 2004 (7): 23-26
- [4] 庞景安、张玉华等. 中国学术期刊综合评价指标体系的研究[J]. *中国学术期刊研究*, 2000(11): 217-219
- [5] 李凯扬、贾玉萍. 基于 AHP 的期刊全文数据库的模糊综合评价[J]. *情报科学*, 2005 (11): 1688-1703
- [6] 王小唯、杨波等. 学术期刊质量评估的二次相对评价方法[J]. *编辑学报*, 2003 (6): 231-232
- [7] 李修杰、陈景武. 运用判别分析法建立的期刊评估指标体系[J]. *江西图书馆学刊*, 2006 (3): 48-50
- [8] 王玖、徐天和. 秩和比法在医学学术期刊学术质量综合评价中的应用[J]. *数理医药学杂志* 2003 (3): 266-267
- [9] 陈汉忠. 主成分分析在学术期刊评价中的应用[J]. *中国学术期刊研究*, 2004 (6): 658-660
- [10] 凌春艳、莫琳. 自然科学学术期刊质量指标体系的属性数学综合评价模型[J]. *数学的实践与认识*, 2004 (5) :1-7
- [11] 李继晓、蔡成瑛. 对各种核心期刊评价方法的分析[J]. *中国学术期刊研究*, 2006 (2): 253-256