

基于Nbreg模型的SCI论文被引频次 影响因素分析*

——中日比较研究的视角

张肃^{1,2} 王泽衡¹

(1. 长春理工大学经济管理学院, 长春 130022; 2. 长春理工大学区域创新与科技发展研究中心, 长春 130022)

摘要: 为更全面客观地认识论文被引频次、寻找提升中国SCI论文质量与影响力的途径, 本文以中国和日本的SCI论文数据为基础, 构建Nbreg负二项回归模型, 分析中日SCI论文被引频次的影响因素, 并从是否影响、影响方向与影响程度3个方面比较多维因素下中日SCI论文被引频次的影响异同点。研究表明, 与日本相比, 当参考文献数量越多, 作者数量越多, 合作方式为国内合作, 论文语种为母语, 文献类型为论文类、综述类时, 中国SCI论文被引频次提高的可能性越大。据此, 可通过建设更多高水平中文期刊、鼓励团队合作等方式来提升中国论文的被引频次, 进而增强我国科研成果的学术影响力。

关键词: Nbreg模型; SCI论文; 被引频次; 影响因素; 中日比较

中图分类号: G322

DOI: 10.3772/j.issn.1673-2286.2019.09.006

被引频次是SCI论文评价的重要定量指标之一^[1-3]。通常情况下, 一篇论文的被引频次越高, 可认为该论文的科研影响力越大, 学者可据此来评估SCI论文质量的优劣^[4-6]。那么, 在SCI论文质量的评估中, 哪些因素与论文被引频次有关、这些因素如何影响论文被引频次等问题, 已引起了学者的关注和探讨^[7-9]。

一些学者试图从不同方面对我国论文被引频次的影响因素进行研究。姜磊等^[10]从参考文献的数量与论文被引频次相关性的角度, 探讨了参考文献的数量与质量对论文被引频次的影响; 徐庆富等^[11]选取15种不同学科的代表性期刊收录的论文为样本, 对论文篇幅、参考文献数量等因素对被引频次的影响分别进行实证检验; 高丹等^[12]利用阐述重要性、文献综述、研究不足等语步, 计算语步的词表重合率与被引频次的关系来分析影响因素; 牟象禹等^[13]从论文特征、引文特征等方面选取指标, 利用这些指标与被引频次的相关性进行影响

因素的分析; 陈悦等^[14]从被引频次主要源于论文自身因素影响的角度, 分析了论文标题和摘要区别度、论文使用次数、热点持续能力等因素对被引频次的影响。当前研究对了解我国论文被引频次的影响因素情况起到积极作用, 但仍存在不足。首先, 孤立地分析我国论文的情况, 缺少与其他国家的比较, 使得到的研究结果缺少参照依据; 其次, 多从论文自身属性等单一维度选取影响因素, 缺少从来源出版物、作者等多角度选取因素, 使得对被引频次影响因素的认识单一, 研究范围过窄; 最后, 多数研究仅明确论文被引频次的影响因素, 而如何从这些影响因素的特征入手, 寻求提高论文被引频次的办法, 当前的研究鲜有阐述。

因此, 本文选择中国与日本为比较对象, 以Web of Science数据库中的SCI论文为数据源, 从来源出版物、作者信息、论文自身信息及基金资助4个方面选取影响论文被引频次的因素, 通过负二项回归模型的构建, 分

*本研究得到吉林省社会科学基金2018年度面上项目(编号: 2018B41)资助。

析中日SCI论文被引频次影响因素的差异,有助于学者更全面客观地认识我国SCI论文的被引频次及其影响因素情况,进而为寻找提升我国SCI论文质量、增强我国科研影响力等方面的途径提供参考依据。

1 研究设计

1.1 数据准备

本文数据来源于Web of Science中的SCI数据库,数据获取步骤为:首先,在SCI中分别检索中国和日本2006—2015年的论文属性数据,检索式分别为“地址=Peoples R China AND 出版年=2006—2015”“地址=Japan AND 出版年=2006—2015”;其次,根据检索结果,逐年下载中国和日本的论文属性数据。

对于下载后的论文属性数据,根据作者地址(C1)字段,构造合作方式字段,合作方式取值为国内合作、

国际合作以及独立作者。对于中国(日本)SCI论文,选取合作方式字段取值为独立作者以及国内合作的论文。需要特别说明的是:①由于中国国情的特殊性、港澳台地区与中国内地科研体制的差异性,本文所涉及的情况均不考虑港澳台地区,仅为中国内地SCI论文数据;②本文着力于分析不同因素对中国和日本的SCI论文分别产生的影响比较,为使得到的研究结论更具比较性、更有说服力,不考虑国际合作论文的情况;③考虑到被引频次指标的时滞性^[15],2015年以前的论文被引频次至今已趋向稳定,并且为进一步分析出版时间可能对论文被引频次产生的影响,本文选择2006—2015年的SCI论文数据。

经上述数据处理后,得到2006—2015年中国SCI论文有效数据1 038 508篇、日本SCI论文676 251篇,其被引频次分布情况如图1所示(由于篇幅原因只给出2015年情况)。可以看出,无论是中国还是日本,论文的被引频次均呈明显的偏态分布。

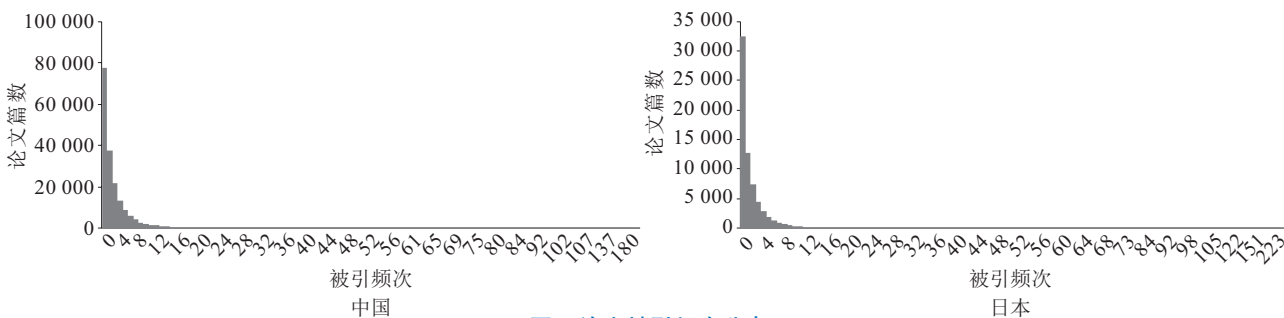


图1 论文被引频次分布

1.2 影响因素的选择

一篇SCI论文的完成,包含科研人员对论文的撰写、修改与发表等过程^[13-14]。其中,参与的作者、所需的资助基金、选择发表的期刊以及论文创作设计等方面均体现科研成果产生的不易。正因为如此,学者往往希望自己的成果能够被更多人关注并认可,而衡量被认可的方式之一,就是论文被引频次的高低。那么,什么因素影响了SCI论文被引频次,这些因素对SCI论文被引频次有怎样的影响等,将是本文的研究核心。

为分析不同相关因素对中日SCI论文被引频次的影响异同,需先构建影响因素的理论模型,以明确所涉及的因素。在阅读并分析现有文献的基础上^[16-18],结合相关的实践经验,本文选择从来源出版物、资助基金、作者信息及论文自身信息4个方面设定影响因素(见图2)。

(1) 论文自身信息类因素。有学者认为,可以通过

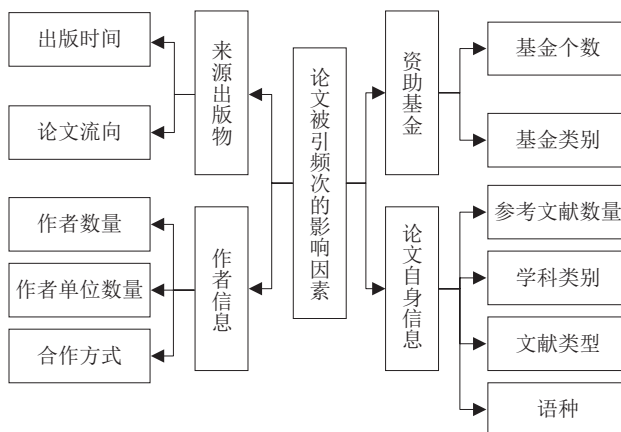


图2 影响因素的理论模型

控制参考文献规模、选择撰写论文学科等途径来人为地影响SCI论文的被引用情况。如徐庆富等^[11]发现参考文献数量在大多数期刊中与被引频次显著正相关;牟象禹等^[13]认为学科类别、文献类型等因素会影响论文

被引频次的特征。在此基础上结合当前学者的研究,本文设定论文自身信息类因素包括参考文献数量、学科类别、文献类型及语种4个因素。

(2) 资助基金类因素。很多学者认为有无基金会会对论文被引频次产生很大影响。如段宇锋等^[19]认为基金论文的被引频次和影响力普遍高于非基金论文;陈悦等^[14]认为有科研资助的论文,被引频次会更高。结合这些分析,并在此基础上,本文设定资助基金类因素包括基金个数与基金类别2个因素。

(3) 作者信息类因素。参与作者作为SCI论文产生的源动力,有学者认为其特征对论文的被引与否也会产生一定的影响。如王海涛等^[18]发现作者数量与论文被引频次有正向影响关系;段宇锋等^[19]认为作者数量对被引频次有较强的贡献度。基于已有研究的发现,结合论文创作中的实际考量,设定本文的作者信息类因素包括作者数量、作者单位数量以及合作方式3个因素。

(4) 来源出版物类因素。出版物是论文成果发表的重要载体,有学者认为出版物的状况也会影响人们是否引用期刊上的论文。如俞立平等^[20]认为被引频次受出版时间影响;余莉^[21]认为论文被引频次受文献流向的刊物影响。据此,本文设定来源出版物类因素包括出版时间与论文流向2个因素。

2 模型构建

2.1 模型变量的设定

2.1.1 变量的选择

根据图2所示,本文设定被引频次为因变量,各影响因素为自变量,其取值情况如表1所示。①本文使用的数据为2006—2015年中国和日本国内作者的论文属性数据,下载时间为2018年12月21—27日,考虑到被引频次本身的动态性,其数值确定为自发表之日起至数据被下载时止的被引频次。②学科类别是在SCI的176个学科条目的基础上概化形成的。③文献类型是在SCI的39种文献类型的基础上概化形成的。

2.1.2 变量相关性分析

在构建Nbreg模型前,需要验证各变量是否存在多重共线性,以防止相关性过高的变量干扰模型结果

表1 设定的变量

变量符号	变量名称	取值方法
Citation	被引频次	计数
A_1	参考文献数量	篇数值
A_2	文献类型	1=论文类; 2=书评类; 3=综述类; 4=概要; 0=其他
A_3	学科类别	0=工学; 1=理学; 2=农学; 3=社会科学; 4=生命科学; 5=医学
A_4	语种	1=英语; 2=母语; 0=其他
A_5	基金类别	1=国内基金; 2=国外基金; 3=国外和国内基金; 0=无
A_6	基金个数	计数
A_7	作者数量	计数
A_8	合作方式	0=独立作者; 1=国内合作
A_9	出版时间	年
A_{10}	作者单位数量	计数
A_{11}	论文流向	0=国内; 1=国外

的准确度。根据相关性分析可知,变量的相关系数小于0.5,则可认为两变量间相关性不明显,相互独立^[18-19]。因此,本文认为如果两变量间相关系数小于0.5,则变量可以引入模型中进行运算,相关性分析工具采用SPSS 22.0。

经计算可得,各变量间的相关系数均小于0.5,变量间的相关性不明显,各变量相互独立,可以将这些变量引入后面的模型中进一步计算。

2.2 模型假设的提出

根据影响因素的选择与变量的具体设定,结合本文的研究核心,并参照过往研究^[13-15],分别从来源出版物、作者信息、资助基金以及论文自身信息4个方面,针对中日SCI论文被引频次影响因素的差异提出本文研究情境下的模型假设(见表2),以待后续的验证。

2.3 Nbreg模型的构建

由于被引频次呈偏态分布,传统的多元线性回归模型并不适合^[22]。同时,被引频次属于离散型变量,取值是典型的计数型,因此考虑利用计数型回归模型进行分析。对于计数型回归模型,首先考虑泊松回归,但此模型的内在假定是计数变量取值的条件均值等于条件方差,即等离散^[13],而经计算本文因变量的方差明显大于期望,即存在过离散的情况,因此泊松回归模型也不

表2 模型的研究假设

类别	影响因素	研究假设
来源出版物	出版时间	H1.1: 出版时间对SCI论文被引频次有负向作用, 并且对中国SCI论文被引频次的影响程度高于日本
	论文流向	H1.2: 论文流向国内对SCI论文被引频次有正向作用, 并且对中国的影响程度高于日本
资助基金	基金个数	H2.1: 基金个数对SCI论文被引频次有正向作用, 并且对中国的影响程度高于日本
	基金类别	H2.2_1: 国内基金对SCI论文被引频次有正向作用, 并且对中国的影响程度高于日本
		H2.2_2: 国外基金对SCI论文被引频次有正向作用, 并且对中国的影响程度高于日本
		H2.2_3: 国内和国外基金对SCI论文被引频次有正向作用, 并且对中国的影响程度高于日本
作者信息	作者数量	H3.1: 作者数量对SCI论文被引频次有正向影响, 并且对中国的影响程度高于日本
	作者单位数量	H3.2: 作者单位数量对SCI论文被引频次有正向影响, 并且对中国的影响程度高于日本
	合作方式	H3.3: 合作方式为国内合作对SCI论文被引频次呈正向影响, 并且对中国的影响程度高于日本
论文自身信息	参考文献数量	H4.1: 参考文献数量对SCI论文被引频次有正向影响, 并且对中国的影响程度高于日本
	学科类别	H4.2_1: 学科类别为理学时对SCI论文被引频次有正向影响, 并且对中国的影响程度高于日本
		H4.2_2: 学科类别为工学时对SCI论文被引频次有正向影响, 并且对中国的影响程度高于日本
		H4.2_3: 学科类别为社会科学时对SCI论文被引频次有正向影响, 并且对中国的影响程度高于日本
		H4.2_4: 学科类别为农学时对SCI论文被引频次有正向影响, 并且对中国的影响程度高于日本
		H4.2_5: 学科类别为生命科学时对SCI论文被引频次有正向影响, 并且对中国的影响程度高于日本
	语种	H4.3_1: 语种为母语时对SCI论文被引频次有正向影响, 并且对中国的影响程度高于日本
		H4.3_2: 语种为英语时对SCI论文被引频次有正向影响, 并且对中国的影响程度高于日本
	文献类型	H4.4_1: 文献类型为论文类时对SCI论文被引频次有正向影响, 并且对中国的影响程度高于日本
		H4.4_2: 文献类型为书评类时对SCI论文被引频次有正向影响, 并且对中国的影响程度高于日本
H4.4_3: 文献类型为综述类时对SCI论文被引频次有正向影响, 并且对中国的影响程度高于日本		
H4.4_4: 文献类型为概要类时对SCI论文被引频次有正向影响, 并且对中国的影响程度高于日本		

适用。再继续评估广义泊松回归、零膨胀泊松回归、零膨胀负二项回归、负二项回归等模型的适用性^[18], 最终选定负二项回归模型, 因文章篇幅限制, 模型选取和确定过程在此不做过多陈述。

负二项回归的概率密度函数为: $P(Y=y|\theta, J) = C_{y+J-1}^{J-1} \theta^J (1-\theta)^y$, 其中 $y=0, 1, 2, \dots, 0<\theta<1$ 。模型计算分析工具采用Stata12.0, 分类变量的参照基准为每组“最后一个”, 取置信度为95%。

3 实证分析

3.1 模型运行结果分析

3.1.1 中国SCI论文被引频次影响因素分析

对中国的论文数据进行负二项回归, 并按照论文自身信息、资助基金、作者信息及来源出版物的顺序将这4类因素逐步引入到模型中计算, 最终分别得到模型Nbreg1—4的回归结果(见表3)。根据结果可分析得出影响中国SCI论文被引频次的因素情况如下。

(1) 各因素是否影响中国SCI论文的被引频次。模型Nbreg4中, 引入模型中的全部因素在5%的水平下均通过显著性检验, 说明论文参考文献数量、作者数量等引入模型中的全部因素均对中国SCI论文的被引频次有影响。

(2) 各因素对中国SCI论文被引频次的影响方向。负二项回归中, 可利用直接系数或边际效应的正负值来判断影响方向, 在模型Nbreg1—4中, 随着被引入模型中的影响因素不断增加, 影响方向始终保持前后一致。因此, 本文以模型Nbreg4的结果为准, 分析各因素对中国SCI论文被引频次的影响方向。①影响方向为正向的因素有参考文献数量、论文类、书评类、综述类、概要、工学、英语、母语、作者数量、合作方式, 说明这些因素每增加一个单位, 中国论文的被引频次会随之增加; ②影响方向为负向的因素有理学、农学、社会科学、生命科学、国内基金、国外基金、国内和国外基金、基金个数、论文流向、出版时间、作者单位数量, 说明这些因素每增加一个单位, 中国论文的被引频次会随之减少。

(3) 各因素对中国SCI论文被引频次的影响程度。

表3 中国论文数据回归结果

类别	Nbreg1		Nbreg2		Nbreg3		Nbreg4	
	直接系数	边际效应	直接系数	边际效应	直接系数	边际效应	直接系数	边际效应
参考文献数量	0.023***	0.148	0.024***	0.156	0.023***	0.146	0.030***	0.157
论文类	0.987***	4.090	0.937***	3.940	0.987***	4.090	0.978***	3.994
书评类	0.055***	0.593	0.041***	0.401	0.043***	0.451	0.048***	0.516
综述类	0.684***	3.426	0.768***	3.538	0.739***	3.521	0.743***	3.595
概要	0.465***	1.602	0.475***	1.766	0.488***	1.883	0.461***	1.596
工学	0.267***	1.787	0.237***	1.530	0.238***	1.511	0.131***	0.698
理学	-0.037***	-0.250	-0.027***	-0.173	-0.007***	-0.045	-0.089***	-0.449
农学	-0.507***	-2.634	-0.455***	-2.329	-0.429***	-2.184	-0.314***	-1.427
社会科学	-0.044***	-0.293	-0.016***	-0.100	-0.030***	-0.187	-0.154***	-0.772
生命科学	-0.093***	-0.596	-0.088***	-0.544	-0.162***	-0.954	-0.201***	-0.993
英语	0.617***	3.156	0.737***	3.471	0.639***	3.070	0.690***	2.751
母语	0.439***	2.408	0.363***	1.976	0.479***	2.449	0.189***	1.086
国内基金	-	-	-0.434***	-2.485	-0.421***	-2.378	-0.422***	-2.379
国外基金	-	-	-0.592***	-3.020	-0.552***	-2.804	-0.532***	-2.574
国内和国外基金	-	-	-0.557***	-3.162	-0.437***	-2.493	-0.421***	-2.314
基金个数	-	-	-0.014***	-0.090	-0.021***	-0.133	-0.025***	-0.139
作者数量	-	-	-	-	0.037***	0.228	0.076***	0.399
合作方式	-	-	-	-	0.612***	3.044	0.521***	2.265
出版时间	-	-	-	-	-	-	-0.247***	-1.305
作者单位数量	-	-	-	-	-0.058***	-0.364	-0.042***	-0.224
论文流向	-	-	-	-	-	-	-0.799***	-3.619

注：***表示在5%的水平下通过显著性检验

Nbreg回归模型中,可用边际效应值来确定各因素的影响程度。在模型Nbreg1—4中,各因素的平均边际效应值前后变化幅度不大,因此,本文以模型Nbreg4中的边际效应值为准来分析各因素的影响程度。对照表3中模型Nbreg4的结果,根据边际效应值的相对大小对各因素的影响程度降序排列,结果依次为论文类、论文流向、综述类、英语、国外基金、国内基金、国内和国外基金、合作方式、概要、农学、出版时间、母语、生命科学、社会科学、工学、书评类、理学、作者数量、作者单位数量、参考文献数量、基金个数。由此可见,当文献类型为论文类时对中国SCI论文被引频次的影响程度最大,基金个数对中国SCI论文被引频次的影响程度最小。

3.1.2 日本SCI论文被引频次影响因素分析

对日本的论文数据进行负二项回归,并按照论文自身信息、资助基金、作者信息及来源出版物的顺序将这4类因素逐步引入到模型中计算,最终分别得到模型

Nbreg1—4的回归结果。

根据表4中的结果,可分析得到影响日本SCI论文被引频次的因素情况如下。

(1)各因素是否影响日本SCI论文的被引频次。模型Nbreg4中,引入模型中的全部因素在5%的水平下均通过显著性检验,说明参考文献数量、作者数量等因素均对日本论文的被引频次有影响。

(2)各因素对日本论文被引频次的影响方向。模型Nbreg1—4中,随着被引入模型中的影响因素不断增加,影响方向始终保持前后一致。因此,本文以模型Nbreg4的结果为准分析各因素对日本SCI论文被引频次的影响方向。①属于正向影响因素的有论文类、书评类、综述类、概要、工学、理学、农学、社会科学、生命科学、英语、母语、合作方式、作者单位数量、论文流向;②属于负向影响因素的有国内基金、国外基金、国内和国外基金、基金个数、作者数量、出版时间。

(3)各因素对日本论文被引频次的影响程度。在模型Nbreg1—4中,各因素的平均边际效应的数值前后

表4 日本论文数据回归结果

类别	Nbreg1		Nbreg2		Nbreg3		Nbreg4	
	直接系数	边际效应	直接系数	边际效应	直接系数	边际效应	直接系数	边际效应
参考文献数量	0.002***	0.024	0.003***	0.027	0.003***	0.026	0.002***	0.019
论文类	0.394***	3.925	0.392***	3.908	0.395***	3.967	0.350***	3.329
书评类	0.698***	6.337	0.719***	6.497	0.715***	6.486	0.693***	6.329
综述类	0.163***	1.466	0.129***	1.117	0.150***	1.301	0.197***	1.852
概要	0.759***	7.076	0.765***	7.286	0.779***	7.311	0.703***	6.405
工学	0.368***	3.885	0.365***	3.802	0.369***	3.837	0.348***	3.258
理学	0.294***	3.264	0.285***	3.097	0.282***	3.065	0.284***	2.795
农学	0.074***	0.733	0.091***	0.899	0.082***	0.811	0.196***	1.849
社会科学	0.492***	5.607	0.485***	5.435	0.481***	5.377	0.423***	4.182
生命科学	0.254***	2.510	0.246***	2.393	0.242***	2.353	0.217***	1.903
英语	0.097***	0.890	0.156***	1.373	0.151***	1.328	0.413***	2.919
母语	0.336***	2.757	0.298***	2.450	0.299***	2.438	0.346***	0.404
国内基金	-	-	-0.262***	-0.369	-0.257***	-0.383	-0.299***	-4.216
国外基金	-	-	-0.550***	-2.505	-0.558***	-2.482	-0.566***	-1.901
国内和国外基金	-	-	-0.039***	-4.506	-0.040***	-4.564	-0.445***	-3.271
基金个数	-	-	-0.013***	-0.121	-0.014***	-0.135	-0.003***	-0.030
作者数量	-	-	-	-	-0.003***	-0.030	-0.004***	-0.032
合作方式	-	-	-	-	0.065***	0.602	0.070***	0.578
出版时间	-	-	-	-	-	-	-0.208***	-1.781
作者单位数量	-	-	-	-	0.016***	0.152	0.033***	0.284
论文流向	-	-	-	-	-	-	0.198***	1.598

注：***表示在5%的水平下通过显著性检验

变化幅度不大，因此，本文以模型Nbreg4中的边际效应值为准来分析各因素的影响程度。对照表4中模型Nbreg4的结果，根据边际效应值的相对大小对各因素的影响程度降序排列，结果依次为概要、书评类、国内基金、社会科学、论文类、国内和国外基金、工学、英语、理学、生命科学、国外基金、综述类、农学、出版时间、论文流向、合作方式、母语、作者单位数量、作者数量、基金个数、参考文献数量。由此可见，当文献类型为概要时对日本SCI论文被引频次的影响程度最大，参考文献数量对日本SCI论文被引频次的影响程度最小。

3.2 中日SCI论文被引频次影响因素的比较分析

3.2.1 影响与否的比较

结合表3、表4中模型Nbreg4的数值，并综合前文的分析，为清晰起见，将中国和日本论文被引频次影响

因素的比较结果绘制成图（见图3）。

由图3所示可知，除参照组不予考虑外，文中列出的因素均对中国和日本的SCI论文被引频次有影响，这一点中国和日本情况相同。

3.2.2 影响方向的比较

（1）对中日SCI论文被引频次影响方向相同的因素。①参考文献数量，文献类型为论文类、书评类、综述类、概要，学科类别为工学，语种为英语、母语，合作方式为国内合作等因素对中国和日本的SCI论文被引频次影响均为正向；②国内基金、国外基金、国内和国外基金、基金个数、出版时间等对中国和日本的SCI论文被引频次影响均为负向。

（2）对中日SCI论文被引频次的影响方向不同的因素。①学科类别为农学、生命科学、社会科学、理学，作者单位数量，论文流向国内这些因素对中国SCI论文被引频次的影响为负向，而对日本SCI论文被引频次的影

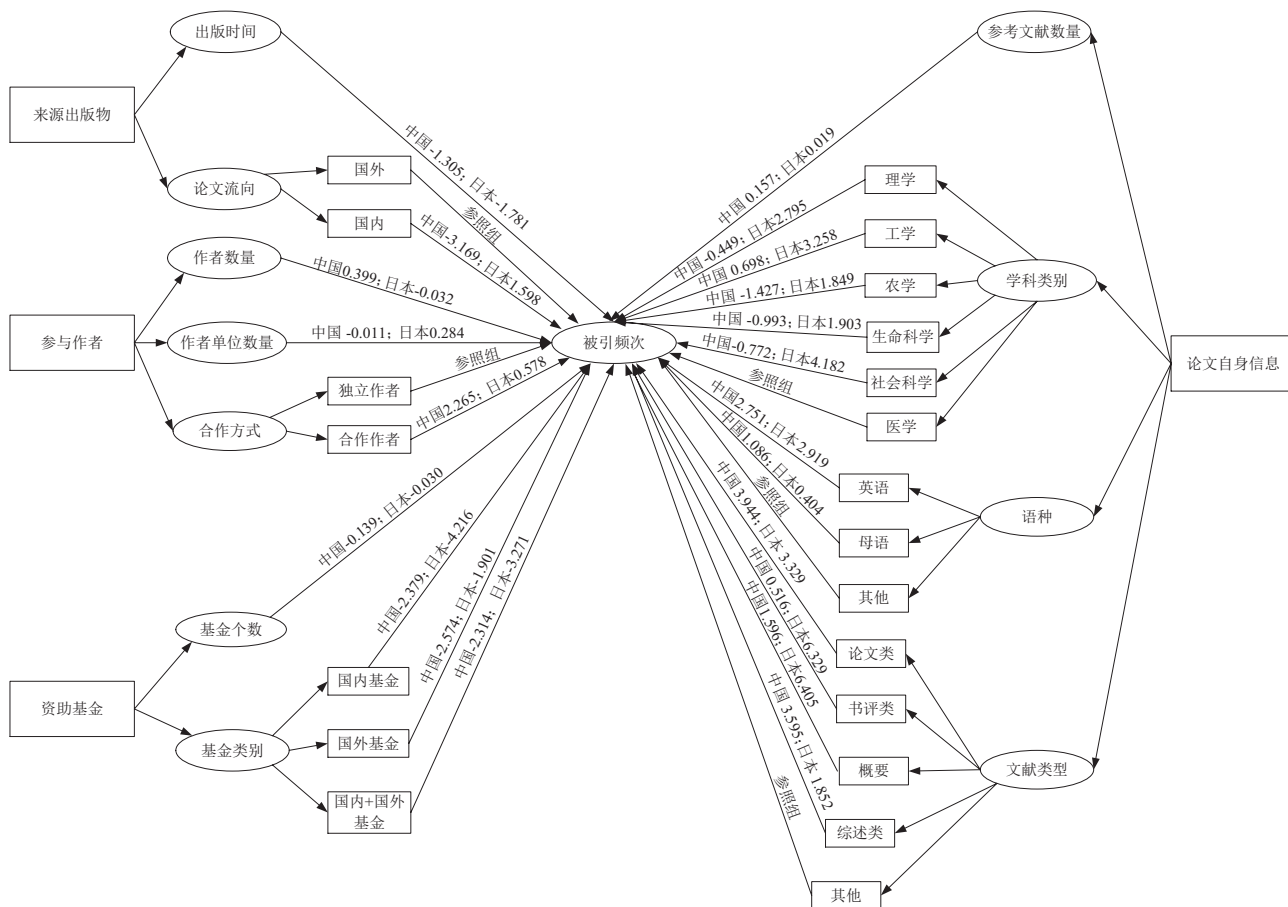


图3 中日论文被引频次的影响因素对比

响为正向；②作者数量对中国SCI论文被引频次的影响为正向，而对日本SCI论文被引频次的影响为负向。

3.2.3 影响程度的比较

(1) 论文自身信息类因素的比较。①参考文献数量的比较。参考文献数量每增加一篇，中国论文的被引频次平均增加0.157次，日本论文的被引频次平均增加0.019次，参考文献数量对SCI论文被引频次呈正向影响，并且对中国的影响程度高于日本。因此，模型研究假设H4.1获得支持。②文献类型的比较。由图3所示，相对于文献类型为其他时，每增加一篇文献类型为论文类、综述类、概要类、书评类的论文，中国和日本的SCI论文被引频次均增加，但论文类、综述类的论文对中国被引频次的影响程度高于日本，概要类、书评类的论文对中国的影响程度低于日本。因此，模型假设H4.4_1和H4.4_3获得支持，H4.4_2和H4.4_4获得拒绝。③学科类别的比较。相对于学科类别为医学时，每增加一篇

学科类型为工学、理学、社会科学、生命科学的论文，日本SCI论文的被引频次下降，并且除工学外的中国论文被引频次也下降，但对中国论文被引频次的影响程度均低于日本。因此，模型假设H4.2_1—H4.2_5均未获得支持。④语种的比较。相对于语种为其他时，英文论文每增加一篇，中国论文的被引频次增加2.751次，日本论文被引频次增加2.919次；语种为母语的论文每增加一篇，中国论文的被引频次增加1.086次，日本论文的被引频次增加0.404次。语种对SCI论文被引频次的影响均为正向，但语种为英语时对中国的影响低于日本，语种为母语时对中国的影响高于日本。因此，模型假设H4.3_1获得支持，H4.3_2被拒绝。

综上所述，相较于日本，参考文献数量越多，文献类型为论文类、综述类，语种为母语时均可使中国SCI论文被引频次增加幅度更高，说明这些因素是促进中国SCI论文被引频次提升的有利因素。这也证实了抛开论文本身内容的质量，通过主观控制引用前人的研究成果数量，有针对性地选择论文的文献类型、语种等方

式,可以影响中国SCI论文被引频次。

(2) 资助基金类因素影响程度的比较。①基金类型的比较。国内基金、国外基金、国内和国外基金每增加一个,中国和日本的SCI论文被引频次均下降,但基金类别为国外基金时对中国的影响程度高于日本,基金类别为国内基金、国内和国外基金时对中国的影响低于日本。因此,模型假设H2.2_1—H2.2_3均未获得支持。②基金个数的比较。基金个数每增加一个,中国论文的被引频次下降0.139次,日本论文的被引频次下降0.030次,对中国论文的影响更大。因此,模型假设H2.1未获得支持。

综上所述,与日本相似,基金类型和基金个数均没有对中国SCI论文被引频次产生提升作用,说明有无基金、基金类型并不是使中国SCI论文被引频次增加有利因素,也验证了论文创作的初衷不是为了吸引各级基金的资助,而是为了获得更多有意义的科研成果,通过吸引基金资助不能提高中国SCI论文质量。

(3) 作者信息类因素影响程度的比较。①作者数量的比较。作者数量每增加一个,中国论文的被引频次增加0.399次,日本论文的被引频次下降0.032次,对中国论文的影响更大。因此,模型假设H3.1获得支持。②合作方式的比较。合作方式为国内合作时,中国论文的被引频次增加2.265次,日本论文的被引频次增加0.578次,对中国论文的影响更大。因此,模型假设H3.3获得支持。③作者单位数量的比较。作者单位数量每增加一个,中国论文的被引频次下降0.011次,日本论文的被引频次增加0.284次,对日本论文的影响更大。因此,模型假设H3.2未获得支持。

综上所述,相较于日本,作者数量、合作方式为国内合作对中国SCI论文被引频次产生明显提升作用,而作者单位数量对中国并没有提升作用。说明作者数量、合作方式是中国SCI论文被引频次增加有利因素。这也证实了团队合作对于中国SCI论文被引频次的重要性,尤其是相较于跨单位的团队合作,与团队的学者因交流合作的便利性会更有利于高水平研究成果的出现。

(4) 来源出版物类因素影响程度的比较。①出版时间的比较。出版时间每靠近一年,中国论文的被引频次平均下降1.305次,日本论文的被引频次平均下降1.781次,对日本的影响更大。因此,模型假设H1.1未获得支持。②论文流向的比较。论文流向为国内时,每增加一篇论文,中国论文的被引频次下降3.619次,日本论文的被引频次增加1.598次。因此,模型假设H1.2未获

得支持。

综上所述,出版时间、论文流向为国内均未对中国SCI论文被引频次产生提升作用,说明出版时间、论文流向并不是中国SCI论文被引频次增加有利因素。其中,出版时间的这一表现符合被引频次的时滞性^[15],因为论文需要在发表一段时间后,其被引频次才会逐步增加直至趋于稳定;而论文流向为国内并没有对中国SCI论文被引频次产生提升作用,这也从侧面说明,高水平的论文并不倾向发表在SCI国内期刊上,这一方面是因为中国缺乏有影响力的期刊,另一方面与当前国内科研评价体系过度追捧国际期刊有关。

4 结论与建议

通过本文的研究,从是否影响、影响方向与影响程度方面得到了中日SCI论文被引频次影响因素的异同点,可知相较于日本,当中国SCI论文具备以下情况时,更有利于提高论文的被引频次:①参考文献数量越多;②文献类型为论文类、综述类时;③语种为母语;④作者数量越多;⑤合作方式为国内合作。

基于此,为有效提升中国SCI论文被引频次,增强中国学术论文质量与影响力,本文提出以下建议。

(1) 鼓励学者进行团队合作,引导学者更多地参与高水平合作项目。与日本相比,作者数量越多、合作方式为国内合作时,更有利于提高我国SCI论文的被引频次,说明团队协作对高水平、高质量的科研成果的产生有积极的意义。因此,可以通过为科研团队提供优质的合作平台、对团队成果给予高额奖励等有效措施和办法吸引更多的科研人才加入到团队合作中来。

(2) 加强优质中文期刊的国际化建设,提高我国中文期刊的国际影响力,增强我国期刊在SCI等平台中的国际话语权。与日本相比,论文语种是母语时更有利于我国SCI论文被引频次的提高,可以从期刊建设方面寻找提升策略。首先,当前存在我国期刊影响力不足、国际化不够等问题,从而导致中文期刊在SCI等国际平台上话语权偏弱,因此,加强我国期刊,尤其是中文期刊的国际化建设刻不容缓;其次,英语是国际上的通用语言,打造中文国际期刊并不意味着对英语期刊的忽视,在中文期刊能够产生足够的国际影响力前,可从建设具有国际影响力的中英文期刊开始。

(3) 合理利用前人的研究成果,有效规划论文的参考文献规模。本文的结果证明了参考文献数量增多

会有利于我国SCI论文被引频次的提高,但这并不意味着参考文献数量越多越好。如果学者忽视了论文自身创作的需要,人为地增加论文参考文献数量,反倒不利于论文质量的提高,那么,论文被引频次的增加也是没有意义的。所以,合理控制参考文献数量的同时,从参考文献的质量、内容等方面进行筛选,将论文创作需求放在首位,才更有利于高水平论文的出现。

(4) 加强引导或鼓励学者更多地进行论文类或综述类的SCI论文创作。与日本相比,论文类或综述类的论文数量增多,有利于我国SCI论文被引频次的提高。因此,可以通过采取相应的奖励措施或激励办法,有意识地引导或吸引学者更多地进行这些类型论文的创作。

通过本文的研究,得出了中日SCI论文被引频次影响因素的差异,可以为学者更全面客观地认识中国论文被引频次影响因素,进而为寻找提升中国SCI论文质量的途径等方面提供参考依据。本文的研究仍存在局限:仅从SCI数据库分析中日论文被引频次影响因素的异同点,没有考虑其他权威论文数据库中中国和日本的情况;仅从论文自身信息、作者信息、来源出版物及基金资助方面来设定相关因素,没有考虑宏观环境因素与论文被引频次的关系;仅分析不同因素下的中日论文被引频次的影响异同点,没有分析这些因素对论文被引频次作用的内在机制等,这些都是需要未来进一步解决的问题。

参考文献

- [1] 鲁玉妙,鞠建伟. SCI 论文学术质量指标与评价模型研究 [J]. 科研管理研究, 2012 (23): 71-74, 79.
- [2] ANDRADE H B, MARTIN E D. Dimensions of scientific collaboration and its contribution to the academic research groups' scientific quality [J]. Research Evaluation, 2009, 18 (4): 301-311.
- [3] 由庆斌,韦博,汤珊红. 基于补充计量学的论文影响力评价模型构建 [J]. 图书情报工作, 2014, 58 (22): 5-11.
- [4] 董政娥,陈惠兰. 基于ESI和SCIE论文来源期刊分区的东华大学学科发展分析 [J]. 东华大学学报(自然科学版), 2012, 38 (1): 107-112.
- [5] DELBARI S A, IMMING S, AZIZ Y A, et al. Measuring the influence and impact of competitiveness research: a Web of Science approach [J]. Scientometrics, 2015, 105: 773-788.
- [6] 俞征鹿,马峥,郭红,等. 中国科技计划项目论文产出和影响力分析 [J]. 情报杂志, 2013, 32 (4): 1-5, 15.
- [7] 程郁缀,刘曙光. 论文被引频次、期刊质量与期刊影响力 [J]. 陕西师范大学学报(哲学社会科学版), 2010, 39 (5): 64-69.
- [8] 胡德华,常小婉. 开放存取期刊论文被引频次和影响力的评价研究 [J]. 图书情报工作, 2008, 52 (2): 61-64.
- [9] 李秋实,刘红玉. 基于文献计量的期刊分区与论文学术评价量化实证研究 [J]. 图书馆工作与研究, 2015 (4): 60-66.
- [10] 姜磊,林德明. 参考文献对论文被引频次的影响研究 [J]. 科研管理, 2015, 36 (1): 121-126.
- [11] 徐庆富,康旭东,张春博. 多期刊比较视角下的论文被引频次若干影响因素研究 [J]. 情报杂志, 2018, 37 (2): 147-153.
- [12] 高丹,李秀霞,周娜. 基于改进CARS模型的学术文献被引频次影响因素分析 [J]. 数字图书馆论坛, 2018 (10): 18-24.
- [13] 牟象禹,龚凯乐,谢娟,等. 论文被引频次的影响因素研究——以国内图书情报领域为例 [J]. 图书情报知识, 2018 (4): 43-52.
- [14] 陈悦,宋超,周京生,等. 文献计量学视角下的论文被引频次影响因素研究——兼评使用与被引之间关系 [J]. 情报杂志, 2019, 38 (4): 96-104.
- [15] 刘雪立,魏雅慧,盛丽娜,等. 科研期刊总被引频次和影响因子构成中的自引率比较——兼谈影响因子的人为操纵倾向 [J]. 编辑学报, 2017, 29 (6): 602-606.
- [16] 邱长波,余莉. 基于数据挖掘的中国SCI论文在线首发论文下载次数典型特征研究 [J]. 图书情报工作, 2011, 55 (10): 83-87.
- [17] 朱雯,陈荣,刘颖. 期刊下载频次和被引频次的相关性研究——复合H指数视角 [J]. 数字图书馆论坛, 2018 (10): 25-31.
- [18] 王海涛,谭宗颖,陈挺. 论文被引频次影响因素研究——兼论被引频次评估科研质量的合理性 [J]. 科学学研究, 2016, 34 (2): 171-177.
- [19] 段宇锋,刘俊茹,步坤. 期刊特征与被引频次的关系研究 [J]. 情报资料工作, 2018 (2): 13-17.
- [20] 俞立平,王艳. 学术期刊被引频次的时间分布规律研究——以图书馆、情报、文献学期刊为例 [J]. 情报科学, 2016, 34 (5): 158-162.
- [21] 余莉. 基于回归树的我国SSCI文献被引频次的特征分析 [J]. 现代经济信息, 2017 (14): 351-353, 355.
- [22] 舒杏,王佳,胡锡琴. 中国企业对“一带一路”国家出口频率研究——基于Nbreg计数模型 [J]. 国际贸易问题, 2016 (5): 82-93.

作者简介

张肃, 女, 1972年生, 博士, 教授, 研究方向: 科技创新、信息计量。

王泽衡, 女, 1988年生, 博士, 讲师, 通信作者, 研究方向: 科学计量学、数据分析, E-mail: wangzeheng@live.cn。

Analysis on Factors Affecting Citation Frequency of SCI Articles Based on Negative Binomial Regression Model:
From the Perspective of Comparison Between China and Japan

ZHANG Su^{1,2} WANG ZeHeng¹

(1. School of Economics and Management, Changchun University of Science and Technology, Changchun 130022, China; 2. Science and Technology Innovation and Regional Development Research Center, Changchun University of Science & Technology, Changchun 130022, China)

Abstract: In order to understand the citation frequency of SCI articles comprehensively and seek ways to improve their quality, this paper chooses SCI articles of China and Japan respectively as the research foundation, and a negative binomial regression model is constructed to analyze the influencing factors of the citation frequency in the two countries, then compares the similarities and differences of factors between China and Japan from the three aspects of influence, influence direction and influence degree. The results show that, compared with Japan, when the number of references is more, or the number of authors is more, or the way of cooperation is domestic cooperation, or the language of the paper is mother tongue, or the type of literature is paper/review, the more likely the Chinese SCI papers are cited. Therefore, we can improve the citation frequency of Chinese papers by building more high-level Chinese journals, encouraging team cooperation, etc., so as to enhance the academic influence of scientific research achievements in China.

Keywords: Negative Binomial Regression Model; SCI Articles; Citation Frequency; Impact Factors; Comparison of China and Japan

(收稿日期: 2019-07-24)

■ 书 讯 ■

《汉语主题词表》

《汉语主题词表》自1980年问世以后, 经1991年进行自然科学版修订, 在我国图书情报界发挥了应有作用, 曾经获得国家科学技术进步二等奖。为适应网络环境下知识组织与数据处理的需要, 由中国科学技术信息研究所主持, 并联合全国图书情报界相关机构, 自2009年开始进行重新编制工作, 拟分为工程技术卷、自然科学卷、生命科学卷、社会科学卷四大部分逐步完成。目前工程技术卷和自然科学卷已出版。

《汉语主题词表(工程技术卷)》共收录优选词19.6万条, 非优选词16.4万条, 等同率0.84, 在体系结构、词汇术语、词间关系等方面进行了改进创新。《汉语主题词表(自然科学卷)》共收录专业术语12.4万条, 包含数学、物理学、化学、天文学、测绘学、地球物理学、大气科学、地质学、海洋学、自然地理学等学科领域, 收词系统、完整, 语义关系丰富、严谨, 每条词汇都有相应的学科分类号表现其专业属性, 并与同义英文术语对应。同时, 建立《汉语主题词表》网络服务系统, 提供术语查询、文本主题分析、知识树辅助构建等服务。《汉语主题词表》可用于汉语文本分词、主题标引、语义关联、学科分类、知识导航和数据挖掘, 是文本信息处理及检索系统开发人员不可或缺的工具。

《汉语主题词表(工程技术卷)》已于2014年由科学技术文献出版社出版, 分为13个分册, 总定价3 880元。

《汉语主题词表(自然科学卷)》已于2018年5月由科学技术文献出版社出版, 分为5个分册, 总定价1 247元。两卷均可分册购买。