

# 科学家跨国流动促进了知识扩散吗？

——基于青年千人的实证分析

刘玮辰 郭俊华 史冬波

**【摘要】**科学家的跨国流动,有助于加强国际间科技交流与合作,促进知识跨国扩散。考察科学家的跨国流动对知识扩散的外部效应,对于认识引进海外人才的实际成效具有重要意义。以青年千人计划入选者为例,通过粗糙完全匹配方法构建实验组和对照组,采用双重差分模型分析青千回国前后国内科学家对其论文引用的变化。这些入选者回国后,国内科学家对其论文引用显著增加,并且该效应既存在于青千回国后发表的论文,也存在于回国前发表的论文。本文认为鼓励科学家的合理有序流动,对于提升我国基础研究水平和源头创新能力至关重要。揭示了人才流动背后隐性知识的正向扩散效应,为人才流动和知识扩散领域的研究提供参考借鉴。

**【关键词】**知识扩散;科学家跨国流动;青年千人计划;双重差分模型;粗糙完全匹配

**【作者简介】**刘玮辰,博士研究生,上海交通大学国际与公共事务学院,研究方向:科技政策与创新管理, vickylu26@sjtu.edu.cn;郭俊华,博士,教授,上海交通大学国际与公共事务学院,研究方向:公共部门绩效管理、创新管理、公共政策分析, guojunhua@sjtu.edu.cn;史冬波(通讯作者),博士,上海交通大学国际与公共事务学院,特别副研究员,研究方向:创新经济学、战略管理, shidongbo@sjtu.edu.cn(上海 200030)。

**【原文出处】**《图书情报知识》(武汉),2020.2.32~41

## 1 引言

近年来,全球范围内的科学家流动日益频繁。科学家的跨国流动推动了依附于其身上的知识,特别是隐性知识,在世界范围内的传播与扩散<sup>[1]</sup>,促进了技术创新<sup>[2]</sup>。在此背景下,各国家与地区纷纷根据自身优势,推出人才政策,以期通过吸引人才流入促进区域经济发展。一直以来,美国都是科学家流动的主要流入国。然而,随着美国公共部门研发经费投入的减少和移民政策的收紧,越来越多的非美籍科学家开始离开美国,选择回到其祖国工作<sup>[3]</sup>。学术界开始越来越关注,离开学术中心的科学家是否能够促进科学知识从学术中心向非中心国家的扩散,并促进非学术中心国家的科学进步。

特别地,随着经济的发展和引才力度的加大,海外青年科学家回流已成为时下中国最重要的人口学

现象之一。从数量上看,青年科学家是海归科学家群体的主体;从知识扩散的角度看,青年科学家处于学术研究最活跃的阶段,掌握着关键的知识和技术,是知识扩散的重要载体。因此,研究海归青年科学家对知识扩散的影响,兼具重要的现实和理论意义。当前,关于海归科学家的研究主要使用描述性分析与案例研究等方法对科学家流动的特征与规律进行总结,鲜有实证研究关注跨国流动对知识扩散的因果效应。

本文以青年千人计划获得者(以下简称“青千”)为例,回答以下研究问题:青千回国能否促进国外知识在国内学术界的扩散?这一作用是否存在学科差异?作用背后的机制是什么?笔者收集了2011-2013年四批入选青年千人计划的科学家的个人简历,并从Web of Science数据库中获取科学家SCI论文及引

文数据。为了克服选择性偏误,本研究运用粗糙完全匹配方法(CEM)构造青千科学家的对照组,并应用双重差分模型(DID)研究回国对青千知识扩散的影响。

本研究首次证实了中国青年科学家跨国流动对知识扩散的正向效应,在理论上具有先导性。跨国流动带来的知识扩散效应有利于缩小发展中国家与世界知识中心的距离,以及促进知识在全球范围内实现共享和传播,从而促进和保持国家和区域内的经济、科技与创新活力。同时,本研究在方法上也具有一定的示范意义:①运用严谨的政策评估方法,使用CEM+DID方法克服了选择性偏误估计政策的效果,为拓展公共政策研究的实证经验提供了依据;②本文使用论文引用数据测度知识扩散<sup>[4]</sup>,且首次在国内研究中使用消歧后的科学家论文发表数据为基础数据进行实证分析,为后续研究提供了方法论上的参考。

## 2 文献综述

科学知识,尤其是隐性知识的传播与扩散从来就不是凭空发生的,通过对以往研究的追溯,发现大量关于科学知识生产的文献出现在经济学、社会学领域<sup>[5-7]</sup>。与此同时,内生增长理论的兴起,以及其对于溢出的强调使得研究的重点聚焦于知识如何在个人、机构和区域之间流动<sup>[8,9]</sup>。由于知识传播于无形,且很少留下痕迹,因此测量知识扩散的难度极大<sup>[10]</sup>。尽管如此,仍有不少学者尝试量化知识扩散的方向和速度。经济学和科学计量领域的学者尝试使用“论文—论文”的引用关系测量学术界的知识扩散过程<sup>[11]</sup>,使用“专利—专利”的引用关系<sup>[12]</sup>或者“专利—论文”的引文数据<sup>[13,14]</sup>刻画知识从学术界扩散到业界的过程。

研究科学家流动对知识扩散的影响,需要克服内生性问题<sup>[15-18]</sup>。内生性的源头来自科学家的流动行为并非随机发生的。例如,学术影响力大、科研能力强的科学家更倾向于流动,无论流动与否,他们的论文都更有可能被流入地的科学家引用。因此,直接对比流动的科学家和未流动的科学家的论文被引次数,有可能会高估流动对知识扩散的作用。Azoulay等<sup>[19]</sup>观测9,483位精英科学家的职业流动,在“科学家—文章(专利)”的组合层面进行匹配,发现科学家的文章被流动前机构的引用并不受科学家离开的影响。

在这一研究思路的启发下,学者们开始将理论视野拓展到科学家的跨国流动对知识扩散的影响。Kahn和MacGarvie<sup>[20]</sup>使用249名富布莱特学者计划的人选人的论文数据,通过匹配控制组,发现富布莱特学者回国对其知识在国内扩散有显著的正向影响,并且该正向作用主要有利于科学环境相对薄弱的国家。与Azoulay等人的研究不同的是,该研究在科学家层面而不是论文层面进行配对,以此解决内生性问题,这也为本文的研究提供了参考。Ganguli<sup>[21]</sup>聚焦移民至美国的前苏联科学家群体,同样采用论文层面配对的方法,证明了移民科学家促进了前苏联的科学知识在美国的传播,并且在俄罗斯科学家移居美国后,苏联时代论文的引用量相较于对照组论文显著增加。事实上,跨国流动的科学家是十分重要的隐性知识的载体,作为“科学中间人”,他们通过跨国流动整合不同国家的信息与资源,不仅助力于实现科学界基础研究领域的知识创新,更促进了业界开发新产品、引进新技术的技术创新<sup>[22]</sup>。

当前跨国知识流动问题已经成为国内创新经济学与图情领域的研究热点。刘云和程旖婕使用全球52个国家和地区的引文数据,分别从个体和国家两个层面研究了国际知识流动的影响因素和内在机制<sup>[23]</sup>。李平等<sup>[24]</sup>使用CSCD数据库的引文数据,发现由国际引文带来的技术知识扩散对中国技术创新能力的正效应。林波等<sup>[25]</sup>使用专利的引用指标,以中美两国碳纳米管技术领域为例,构建了国家专利技术的知识流动模型。后续学者在测度海归学者的知识扩散方面进行了探索。孟晋宇等<sup>[26]</sup>以赴美留学后回国并继续从事科研工作的海归学者为研究对象,从论文数量、论文质量、论文贡献度和知识转移意愿四个维度构建了知识转移的绩效测度算式,证实了海归学者的知识转移是其实现知识创新的关键途径。特别地,作者考虑了重名问题,使用了科学家的姓名变形及履历的检索方法,并辅以人工核对的方式以最大程度保证结果的准确性。

通过回溯人才流动与知识扩散的经典文献,笔者发现当前面临三个重要的挑战:①宏观层面分析较多,聚焦于流动的科学家个体层面的研究不足;②当前研究中观测到的正效应未考虑到内生性的潜

在影响;③较少论文关注了中国科学家姓名歧义问题,尽管有学者尝试解决作者的重名问题,然而普遍还是以科学家履历筛选为主,辅以手工核对,然而这一做法不仅耗时耗力,且不具备可复现性。基于此,本文通过在科学家个体层面进行匹配以控制科学家的科研能力差异克服内生性问题,并开发基于科学家简历和论文引用关系的消歧算法。

### 3 制度背景

自改革开放以来,我国初步形成了具有中国特色的人才政策体系框架,主要包括三个阶段:1978—1992年以人才解放和科技体制改革为主要方向的改革开放初期、1992—2005年以科教兴国和人才强国为方针的战略实施阶段、2006年至今以海外引才为重点的改革成型期。我国人才政策体系的发展过程呈现出以下趋势:①对人才的要求越来越严格,愈发注重质量而非数量;②在中央主导下,各级地方政府从不同层面推出了一系列配套的引才计划,形成了多层次全覆盖的人才体系;③引才的重点从前期主要专注于吸引拥有海外留学及工作经验的人才归国转向同时吸引外籍人才,充分促进了人才在全球范围内的合理流动。

近年来,中国在引进科研人才方面投入巨大,各大部委相继推出一系列的人才项目,为我国的科技实力的迅速提升奠定了人才基础。其中,青年千人计划作为面向海外高层次青年人才的领军计划,为我国海外人才引进做出了突出贡献。2010年,中央人才协调小组批准通过了《青年海外高层次人才引进工作细则》,青年千人计划正式启动,计划从2011年起5年内引进2,000名左右海外青年人才归国工作,主要待遇为中央财政给予每位入选者一次性50万元的生活补贴,三年100—300万元的科研经费补助。青千的入选条件是:主要面向自然科学或工程技术领域,年龄不超过40岁,在海内外知名高校获得博士学位并有3年以上的海外科研工作经验。青年千人计划实施以来,已超额完成5年引进2,000人的目标,并且2017年起至今青年千人计划依然持续实施,已逐渐成为海外引才的新常态。青千计划还带动了地方政府、高校一系列配套计划的实施,成功吸引了大批海外青年人才归国,极大调动了青年科技

人才的科研热情,青千逐渐成为了我国科技创新人才队伍中的中坚力量。

由于目前青年千人作为我国最重要的针对青年科学家的引才计划,其政策代表性、引进人才的质量以及全职回国的要求为本研究提供了优质的观测样本条件;同时,政策带来的天然冲击为研究跨国流动提供了理想的实证场景。因此,本文选取青年千人计划入选者为研究样本以探析青年科学家的跨国流动对知识扩散的影响。

## 4 数据与方法

### 4.1 数据

本文的数据准备过程分为两个部分:收集科学家数据和筛选其论文数据。在第一部分,首先以2011—2013年四批青年千人获得者共721位青年科学家为研究对象,通过千人计划官网收集入选名单,并通过大学官方网站、LinkedIn等网站手工收集721位科学家的简历,整理青年科学家简历数据库,包括科学家姓名、年龄、性别、本科至博士的毕业院校、毕业时间、专业、博士毕业后到目前的所有工作或学习机构以及在该机构的起止年份等数据;之后排除了目前并不在学术界工作以及并未全职回国的科学家,共获得599位观测科学家样本,其中394位科学家获得海外博士学位;然后采用粗糙完全匹配方法<sup>①</sup>,基于ProQuest博硕士学位论文数据库,在个体层面为所有博士毕业于海外高校的科学家匹配了毕业于同一博士院校、同一专业、毕业年份临近但并未回国且仍在学术界工作的华人科学家作为对照组样本。最终,笔者为207位实验组科学家匹配了541位对照组科学家。

在筛选科学家论文时,本文采用监督学习的方法,思路如下:

(1)根据科学家姓名的三种写法<sup>②</sup>,从WoS SCI数据库选取其截止到2017年底的全部SCI论文作为备选集 $P_0$ ;

(2)从 $P_0$ 中选择论文的发文单位(Affiliation)与科学家简历中的工作机构匹配,且论文发表年份与其工作年份相差不超过 $n_i$ 年的论文构成 $P_1$ ;

(3)从 $P_0$ 与 $P_1$ 的差集中,选择与 $P_1$ 论文存在引用关系的论文加入 $P_1$ ,不断迭代直到 $P_1$ 不再变化;

(4)从 $P_1$ 中选择论文的发文单位与科学家简历中的工作机构匹配,且论文发表年份与其工作年份相差不超过 $n_2$ 年的论文构成 $P_2$ ,最终的 $P_2$ 便是科学家的论文集合。

参数 $n$ 是上述过程中唯一的参数。笔者手工收集了50位科学家的2,798篇论文作为准确集,并按照7:3的比例分成了训练集和测试集。使用F1-score作为评价指标,在训练集上的最优参数为 $n_1=2$ ;  $n_2=5$ ,在测试集上模型的F1值为0.91。将该消歧算法应用于全部的观测科学家样本,获得了最终的两组共506位科学家的17,727条论文数据,平均每位科学家发表论文35篇。论文集中包含科学家回国前发表的文章共6,719篇,回国后发表的文章共11,116篇。样本集中506位科学家涵盖数理科学、化学、工程与材料科学、生命科学、地球与环境科学和信息科学六个领域,其学科领域分布见图1。

基于上述数据,本研究在回国前发文数量、回国前加权JIF(Journal impact factor 期刊影响因子)的发文数量、回国前第一作者发文数量和回国前最后作者发文数量四个维度上进行匹配,以保证两组科学家回国前在以上四维度上均不存在显著差异。最终,得到161位实验组科学家和345位对照组科学家作为观测样本。

#### 4.2 研究方法

本研究的核心研究问题是探析青年科学家跨国流动对其知识扩散的影响,利用CEM匹配的样本结

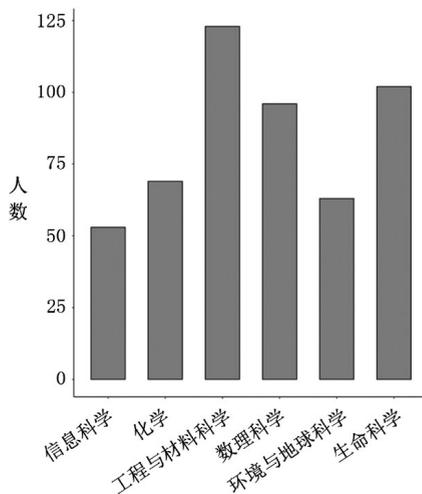


图1 学科领域分布图

合DID模型推断主效应。本文的观测单位是科学家一人一年,核心CEM-DID分析模型如下:

$$E[y_{i,t} | X_{i,t}] = \beta_0 + \beta_1 \times \text{treat} \times \text{post} + \beta_2 X_{i,t} + \eta_i + \delta_t + \varepsilon_{i,t}$$

本研究在粗糙完全匹配的基础上,运用了DID双重差分模型探析因果效应。其中, $y_{i,t}$ 是本研究的核心因变量,即科学家每人每年被国内知识机构引用的次数。 $\text{treat} \times \text{post}$ 是本研究关心的核心自变量, $\text{treat}$ 为是否获得青千, $\text{post}$ 是回国前后的时间差异,其交互项系数反映了政策实施的净效应。 $X$ 为控制变量,包含表征科学家特征的协变量,即性别、博士院校排名、博士毕业国家、本科院校排名、本科毕业国家、本科是否在海外就读以及本科毕业年份。模型中也包括了学科和年份的固定效应。表1为回归

表1 主回归模型的变量定义

变量	定义与取值
$\text{Treat}_i$	入选青千回国的实验组为1,对照组为0
$\text{Post}_t$	年份在回国年份之前为0,之后为1
$\text{Gender}_i$	科学家 $i$ 为男性取1,女性取0
$\text{Discipline}_i$	科学家 $i$ 所在学科领域
$\text{Year}_t$	$t$ 的年份
$\text{Return\_year}_i$	科学家 $i$ 回国时的年份
$\text{PhD\_rank}_i$	科学家 $i$ 博士院校排名,分类变量:1—排名1-10,2—排名11-50,3—排名51-100,4—排名101-200,5—排名201+
$\text{PhD\_country}_i$	科学家 $i$ 博士学校所在国家/地区,分类变量:1—美国,2—欧洲,3—亚洲国家/地区,4—其他国家
$\text{Bachelor\_country}_i$	科学家 $i$ 本科学校所在国家/地区,分类变量:0—中国,1—美国,2—欧洲,3—亚洲其他国家/地区,4—其他国家
$\text{Bachelor\_overseas}_i$	科学家 $i$ 本科学位在海外获得取1,在国内获得取0
$\text{Bachelor\_graduation\_year}_i$	科学家 $i$ 的本科毕业年份
$\text{No. of Chinese cite } 1_i$	如果一篇论文的被引文章有一个来自中国的机构,那么则视为被国内机构引用
$\text{No. of Chinese cite } 2_i$	如果一篇论文的被引文章全部来自中国的机构,那么则视为被国内机构引用

模型的全部变量与定义。特别地,关于核心因变量,本文采用两种定义方式测度科学家知识向国内学术界扩散的程度:①以No. of Chinese cite1 计算论文被国内机构引用的次数,即如果一篇论文的被引文章有一个来自中国的机构,那么则视为被国内机构引用;②采用No. of Chinese cite2 进行稳健性检验,即如果一篇论文的被引文章全部来自中国的机构,那么则视为被国内机构引用。

## 5 实证结果

### 5.1 描述性统计

对观测科学家匹配后样本的核心自变量进行了

描述性统计,从结果表2可以看出,匹配后的实验组和对照组在学科、博士院校排名、博士毕业国家、本科院校排名和本科毕业时间上均值十分接近,且两组在回国前6年的发文数量、回国前6年加权平均期刊影响因子的发文数量、回国前6年的第一作者发文量和回国前6年的最后作者的发文数量四个维度上没有显著差异,即最大程度上在个人层面控制了科学家的差异,为回国的青千实验组找到了非常相似但没有回国的“反事实”对照组。

### 5.2 科学家流动对知识扩散的影响

表3列出了本文的主效应,即青年科学家跨国流

表2 匹配后核心变量T检验

变量	Welch Two Samplet-test				
	Group	mean	t-statistic	df	Pvalue
学科	Treat	3.680	0.946	839.630	0.344
	Control	3.533			
博士院校排名	Treat	2.848	0.235	806.880	0.814
	Control	2.818			
博士毕业时间	Treat	2,007.520	0.096	505.000	0.924
	Control	2,002.063			
本科院校排名	Treat	3.870	0.078	657.120	0.938
	Control	3.859			
本科毕业时间	Treat	1,999.962	0.095	505.010	0.924
	Control	1,994.556			
回国前6年的发文数量	Treat	8.056	-0.849	495.260	0.396
	Control	8.678			
回国前6年加权平均JIF的发文数量	Treat	56.096	1.512	477.240	0.131
	Control	47.245			
回国前6年的第一作者发文数量	Treat	3.242	0.211	410.630	0.833
	Control	3.182			
回国前6年的最后作者发文数量	Treat	0.702	-1.633	343.350	0.102
	Control	0.955			

表3 回国前后科学家知识扩散的主效应

	Dependent variable: No. of Chinese cite 1					
	Full pub		Pub before return		Pub after return	
	OLS (1)	Poisson (2)	OLS (3)	Poisson (4)	OLS (5)	Poisson (6)
Treat×Post	59.393*** (7.874)	0.701*** (0.079)	33.178*** (5.302)	0.596*** (0.079)	26.946*** (3.807)	0.848*** (0.098)
Gender	14.792*** (2.728)	0.413*** (0.080)	8.394*** (1.847)	0.383*** (0.079)	5.056*** (1.096)	0.567*** (0.121)
Constant	-13.545 (22.679)	-15.792*** (0.671)	-18.559** (8.201)	-21.497*** (2.267)	-13.139** (6.487)	-23.546*** (3.024)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Discipline FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,062	6,062	6,062	6,062	6,062	6,062
Adjusted R <sup>2</sup>	0.184		0.137		0.183	
Log Likelihood		-216,961.500		-152,614.100		-69,913.970

注:\*\*\*p < 0.01, \*\*p < 0.05; 括号中汇报的是稳健标准误差;其他控制变量包括本科毕业学校排名与国家,本科毕业年份和博士毕业学校排名。

动对其知识扩散的影响。以知识扩散的第一种测度方式来检测主效应,即当科学家的论文被引文章中有一个中国机构,则视为被国内机构引用。其中模型(1)、(3)、(5)采用OLS模型对主效应进行检验,模型(2)、(4)、(6)采用Poisson模型对主效应进行稳健性检验。此外,笔者不仅对科学家的全部文章进行回归,还进一步区分了科学家回国前的文章和回国后的文章被国内学术界引用的结果。本文的回归模型中全部采用稳健性标准误,并且包含了学科、博士院校排名、本科毕业国家、本科院校排名和本科毕业年份等控制变量。

从模型(1)、(2)的样本回归结果来看,交互项Treat×Post的估计系数显著为正,且在1%的水平上显著通过检验,即总体来讲,科学家回国后其知识会显著的在国内学术界得到传播与扩散,且正效应在OLS模型和Poisson模型中均得到检验,因此表明科学家跨国流动会显著促进知识扩散。模型(3)、(4)回归结果显示,对于科学家回国前的文章,国内学术界对其引用也显著增加,且这种正效应稳健地在1%的显著水平上通过检验,表明对于科学家回国前的文

章,也存在流动对知识扩散的正向效应。模型(5)、(6)的结果显示,对于科学家回国后的文章,在两个模型中均在1%的显著性水平上呈现正向效应,表明回国后的文章也受到国内学术界的引用和扩散。

### 5.3 稳健性检验

为了检验模型的准确性,进一步进行了稳健性检验,采用知识扩散的第二种测度方式来检测主效应,即当科学家的论文被引文章中全部来自中国机构,则视为被国内机构引用,并且同样地分别运用了OLS回归和Poisson回归对科学家的全部文章、科学家回国前的文章和回国后的文章进行了检验。从表4的回归结果可以看出,总体上与表3的主效应基本一致,交互效应的系数在全部发文、回国前发文和回国后发文均在1%的水平上呈现显著的正向效应。稳健性的结果表明了科学回国后,会增加国内学术界对其知识的引用程度,揭示了流动对知识扩散的正向效应。

### 5.4 异质性分析

科学家的不同学科和所在机构的差异和教育水平的不同会对本文的研究结果产生不同的影响,因

表4 回国前后科学家知识扩散的稳健性检验

	Dependent variable: No. of Chinese cite2					
	Full pub		Pub before return		Pub after return	
	OLS (1)	Poisson (2)	OLS (3)	Poisson (4)	OLS (5)	Poisson (6)
Treat×Post	34.006*** (4.097)	0.735*** (0.073)	18.508*** (2.698)	0.624*** (0.078)	15.498*** (2.161)	0.939*** (0.101)
Gender	7.146*** (1.192)	0.436*** (0.071)	4.653*** (0.775)	0.409*** (0.067)	2.494*** (0.638)	0.497*** (0.126)
Constant	-20.355*** (6.534)	-21.477*** (2.055)	-13.464*** (4.100)	-32.673* (19.515)	-6.891* (3.603)	-25.621*** (3.479)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Discipline FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,062	6,062	6,062	6,062	6,062	6,062
Adjusted R <sup>2</sup>	0181		0136		0177	
Log Likelihood		-106,161.300		-83,389.960		-37,872.390

注:\*\*\*p < 0.01, \*p < 0.1; 括号中汇报的是稳健标准误差;其他控制变量包括本科毕业学校排名与国家,本科毕业年份和博士毕业学校排名。

此本文从四个方面对总体样本观测值进行了异质性检验(表5)。将全部科学家的学科按照中组部的引进

学科大类分成了数理科学、化学、地球与环境科学、信息科学、工程与材料科学和生命科学六个学科。根据

表5 回国前后科学家知识扩散的异质性检验

	Cited_by_Chinese_nst1			Cited_by_Chinese_nst2			Observations
	Full pub	Pub before return	Pub after return	Full pub	Pub before return	Pub after return	
	Poisson (1)	Poisson (2)	Poisson (3)	Poisson (4)	Poisson (5)	Poisson (6)	
Main effect model	0.677*** (0.075)	0.596*** (0.079)	0.848*** (0.098)	0.735*** (0.073)	0.624*** (0.078)	0.939*** (0.101)	6,062
Mathematics and Physics	0.965*** (0.144)	0.881*** (0.156)	1.138*** (0.183)	0.941*** (0.166)	0.932*** (0.174)	0.976*** (0.221)	1,151
Chemistry	0.637*** (0.130)	0.532*** (0.117)	0.756*** (0.171)	0.656*** (0.135)	0.518*** (0.119)	0.795*** (0.167)	815
Earth and environmental science	0.790*** (0.182)	0.770*** (0.170)	0.974*** (0.273)	1.092*** (0.180)	0.978*** (0.166)	1.515*** (0.316)	754
Information science	0.386** (0.195)	0.270 (0.193)	0.509** (0.251)	0.648*** (0.212)	0.549*** (0.200)	0.720*** (0.267)	625
Engineering and Material science	0.116 (0.106)	-0.032 (0.115)	0.418*** (0.152)	0.382*** (0.109)	0.148 (0.132)	0.898*** (0.141)	1,492
Life science	0.392*** (0.139)	0.089 (0.139)	1.016*** (0.179)	0.314*** (0.119)	0.108 (0.124)	0.828*** (0.171)	1,225
Return to Tsinghua & Peking University	0.267* (0.159)	-0.231 (0.201)	1.220*** (0.198)	0.210 (0.171)	-0.241 (0.212)	1.235*** (0.193)	1,142
Return to other C9 Universities	0.829*** (0.112)	0.694*** (0.126)	1.111*** (0.157)	0.837*** (0.126)	0.688*** (0.141)	1.095*** (0.187)	1,784
Return to Chinese Academy Science systems	0.326 (0.246)	0.514** (0.246)	-0.104 (0.291)	0.295* (0.162)	0.467*** (0.171)	-0.147 (0.195)	1,183
Return to other universities	0.776*** (0.105)	0.728*** (0.107)	0.867*** (0.147)	0.923*** (0.116)	0.882*** (0.108)	1.013*** (0.177)	1,953
PhD University Rank 1-10	0.740*** (0.201)	0.715*** (0.217)	0.778*** (0.230)	0.769*** (0.169)	0.707*** (0.191)	0.883*** (0.220)	1,114
PhD University Rank 11-50	0.486*** (0.127)	0.394*** (0.137)	0.640*** (0.162)	0.592*** (0.136)	0.470*** (0.135)	0.801*** (0.188)	2,067
PhD University Rank 51-100	0.216 (0.246)	0.248 (0.262)	0.100 (0.288)	0.266 (0.199)	0.248 (0.226)	0.275 (0.246)	823
PhD University Rank 101-200	1.251*** (0.189)	1.176*** (0.204)	1.401*** (0.246)	1.475*** (0.186)	1.383*** (0.198)	1.657*** (0.266)	878
PhD University Rank 200+	0.636*** (0.147)	0.386*** (0.144)	1.108*** (0.213)	0.599*** (0.146)	0.329** (0.143)	1.115*** (0.217)	1,117
PhD location in US	0.670*** (0.087)	0.608*** (0.096)	0.775*** (0.106)	0.741*** (0.085)	0.653*** (0.093)	0.896*** (0.116)	5,397
PhD location in Europe	1.526*** (0.482)	0.220 (0.297)	2.825*** (0.721)	1.734*** (0.547)	-0.157 (0.313)	3.139*** (0.782)	102
PhD location in Asian countries	0.326 (0.248)	0.124 (0.248)	0.656* (0.337)	-0.047 (0.263)	-0.136 (0.260)	0.108 (0.350)	303
PhD location in other countries	0.590*** (0.194)	0.644*** (0.205)	0.506* (0.292)	0.804*** (0.205)	0.916*** (0.223)	0.549* (0.286)	260

注:\*\*\*p < 0.01, \*\*p < 0.05, \*p < 0.1; 括号中汇报的是稳健标准误差。

上海交通大学世界大学学术排名(ARWU, Academic Ranking of World Universities), 本文将博士毕业院校分为世界排名1-10、排名11-50、排名51-100、排名101-200以及排名200+共五组。按照科学家流入机构的性质, 将其分为清华大学与北京大学、除清北外其他C9类高校、中科院系统和其他高校四类。对于科学家的博士院校毕业国家, 将其分为美国、欧洲、亚洲国家或地区和其他国家四类。

### (1) 学科异质性分析

在所有观测的科学家样本中, 数理科学专业科学家96人, 共1,151条观测值, 占样本总量的18.99%; 化学专业科学家69人, 共815条观测值, 占样本总量的13.44%; 地球与环境科学专业科学家63人, 共754条观测值, 占样本总量的12.44%; 信息科学专业科学家53人, 共625条观测值, 占样本总量的10.31%; 工程与材料科学专业科学家123人, 共1,492条观测值, 占样本总量的24.61%; 生命科学专业科学家102人, 共1,225条观测值, 占样本总量的20.21%。从表5的结果可以看出, 数理科学、化学和地理与环境科学的知识扩散模型与全样本主效应模型结果保持一致, 来自这三个学科的科学家的全部文章、回国前文章和回国后文章均在1%的水平上知识扩散效应显著。信息科学和生命科学领域的科学家在其全部发文和回国后发文的模型中稳健地存在显著正向知识扩散效应。工程与材料科学领域的科学家其回国后的知识在1%的显著性水平上呈现正向扩散效应。

### (2) 流入机构异质性分析

科学家的流入机构会对科学家回国后的知识扩散的程度带来差异。通过对样本观测值的整理分类, 其中回国后回到清华大学和北京大学的科学家94位, 共计1,142条观测值, 占总观测值的18.84%; 在除清北外其他C9高校工作的科学家148位, 共计1784条观测值, 占总观测值的29.43%; 回到中科院系统的科学家100位, 共计1,183条观测值, 占总观测值的19.52%; 回到其他高校工作的科学家164位, 共计1,953条观测值, 占总观测值的32.22%。表5的模型结果表明, 回国且回到其他C9高校和其他高校的科学家结果与全样本模型的回归基本一致, 基本保持在1%的显著性水平上呈现正向知识扩散效

应。回到清华北大的科学家其回国后的文章更多地在国内学术界得以传播扩散, 在1%的水平上呈现显著的正向效应。回国且回到中科院系统的科学家, 其回国前的文章存在显著的正向知识扩散效应, 且结果稳健。

### (3) 博士院校异质性分析

科学家的教育背景差异, 尤其是博士阶段的教育, 对其知识扩散的效果会产生不同的影响。根据世界大学学术排名, 全部样本中博士毕业于世界顶尖院校排名1-10的科学家93位, 共计1,114个观测值, 占总观测值的18.38%; 博士院校排名11-50的科学家170位, 共计2,067条观测值, 占总体的34.1%; 博士院校排名51-100的科学家69位, 共计823条观测值, 占总体13.58%; 博士院校排名101-200的科学家74位, 共计878条观测值, 占总体的14.48%; 排名200+院校的科学家95位, 共有1,117个观测值, 占全部观测值的18.43%。表5的回归结果表明, 博士毕业于顶尖院校排名1-10、排名11-50、排名101-200和排名200+的科学家回国后论文受到国内学术界的引用显著增长, 其正向效应在1%的显著性水平上通过检验, 与主效应结果保持一致。博士毕业于排名51-100院校的科学家在全部文章、回国前和回国后的模型上扩散结果均不显著。

### (4) 博士国家异质性分析

不同国家的博士培养方式存在差异, 因此本文分析具有不同国家博士学位的科学家其知识扩散效果的差异。总体样本中, 博士毕业于美国的科学家占了大多数, 共计449人, 有5,397条观测值, 占总观测值的89.03%; 博士毕业于欧洲的科学家8人, 有102个观测值, 占总观测值的1.68%; 博士毕业于亚洲国家或地区的科学家27人, 有303个观测值, 占总体观测值的5%; 博士毕业于其他国家的科学家22人, 有260个观测值, 占总体观测值的4.29%。表5的分组回归结果中, 博士毕业于美国院校的科学家在回国后其知识被国内学术界大量地引用, 在1%的显著水平上呈正效应。博士毕业于欧洲院校的科学家在全部发文和回国后发文的模型呈现显著正效应, 然而其回国前发文的正向扩散效应不稳健, 这可能是样本过少的原因。博士毕业于亚洲国家或地区的

科学家在回国后发文的模型中呈现了正向效应。

## 6 结论与讨论

本文选取了青年千人获得者为样本,通过手工收集2011-2013年721位科学家的简历,准确核实了全职回国且仍在学术界工作的599位中国学者。运用WoS数据库检索科学家的论文,通过粗糙完全匹配和双重差分模型,分析了青年科学家跨国流动对其知识扩散的影响。研究发现,总体来讲,科学家回国后,其知识扩散呈现显著的正向效应。这一正向效应既存在于青千回国后发表的论文,也存在于回国前发表的论文。进一步的稳健性检验证实了模型的准确性和稳定性,表明科学家的文章随着其回国被国内学术界的引用显著增长。研究结论对于目前政策实践具有十分重要的意义,揭示了青年科学家的流动对知识扩散的影响存在延续性,验证了人才流动对知识转移的贡献。

为进一步探究知识扩散效应在不同学科、回国机构、博士毕业院校和博士就读国家间的差异,本文进行了异质性分析。结果发现,数理科学、化学和地球与环境科学学科的科学家回国后的知识扩散效应与主效应一致,存在显著正向效应。回国且回到其他C9高校和其他高校工作的科学家知识扩散的效果十分显著。博士毕业院校排名在1-10、11-50、101-200和200+的科学家回国后其文章在国内学术界得以显著扩散。博士毕业于美国高校的科学家,回国对其知识扩散影响显著增强。异质性的差异体现在两个方面:①研究的组织开展形式不同,主要依托于个人独立展开研究的学科,科学家回国前的文章能够显著地被国内学术界引用,如数学学科;而主要以实验室、团队的形式展开研究的学科,科学家回国对知识扩散的影响并不稳健,如工程与材料科学、生命科学等学科。②科研环境对不同学科的重要性也存在差异,对于一些高精尖学科,实验设备和材料的重要性对科研成果的产生至关重要。学科的差异对于我国未来人才引进政策有重要的启示:国内高校和科研机构在引进海外青年学者时应综合考虑到不同学科的人才规模与发展趋势,鼓励从事实验研究或应用研究学科的优秀科学家回国,促进知识的全球扩散。

本研究也存在一定的不足:①本研究基于科学家个体层面进行匹配以控制实验组和对照组在回国前的差异,进而研究知识扩散的净效应,然而知识扩散的效果体现在科学家论文的引用轨迹上,因此后续研究需要在文章层面进行匹配以控制文章的质量差异;②科学家回国前后知识扩散的影响存在着不同的机制路径,例如回国后合作网络、师承关系等均影响知识扩散的广度和深度,对机制的深入探讨有助于深入理解知识生产过程和科学研究的基本规律。未来的研究需要为海外人才引进的制度设计做出贡献,既要加强对海外优秀科学家的甄选、识别,还需要完善现有的评价体系,综合考虑科学家在其领域内的科研能力、学术潜力、新知识与新技术的转移扩散能力等各方面的能力。

### 注释:

①粗糙完全匹配由lacus等人<sup>[27]</sup>提出,是基于协变量匹配的一种特殊的匹配方式。其基本思路是,首先根据连续变量的分布将其分成若干区间,然后与离散变量一起将空间分成若干格子。每一个观测样本唯一落在一个格子里,最终保留同时有实验组和对照组的格子。

②本文结合WoS数据库中作者的论文署名形式,将作者姓名形式分为三类。例如作者zhangyuanbo,那么其三类姓名形式为zhangyuanbo、zhangyb和zhang y。

### 参考文献:

[1]OECD. The OECD Innovation Strategy: Getting a Head Start on Tomorrow[M]. Paris: Oecd Press, 2010:145.

[2]Fallick B, Fleischman C A, Rebitzer J B. Job-Hopping in Silicon Valley: Some Evidence Concerning the Microfoundations of a High-Technology Cluster[J]. The Review of Economics and Statistics, 2006, 88(3): 472-481.

[3]Kahn S, Macgarvie M J. How Important Is US Location for Research in Science?[J]. Review of Economics and Statistics, 2016, 98(2):397-414.

[4]Jaffe A B, Trajtenberg M, Henderson R. Geographic

Localization of Knowledge Spillovers as Evidenced by Patent Citations[J]. *The Quarterly Journal of Economics*, 1993, 108(3): 577-598.

[5]Stephan P E. The Economics of Science. In *Handbook of the Economics of Innovation*[M]. Amsterdam: Elsevier Press, 2010, 1: 217-273.

[6]Partha D, David P A. Toward a New Economics of Science [J]. *Research Policy*, 1994, 23(5): 487-521.

[7]Merton R K. *The Sociology of Science: Theoretical and Empirical Investigations*[M]. Chicago: University of Chicago Press, 1973: 142-145.

[8]Romer P M. Endogenous Technological Change[J]. *Journal of Political Economy*, 1990, 98(5, Part 2): 71-102.

[9]Aghion P, Howitt P. A Model of Growth through Creative Destruction[J]. *Econometrica*, 1992, 60(2): 323-351.

[10]Krugman P. Increasing Returns and Economic Geography[J]. *Journal of Political Economy*, 1991, 99(3): 483-499.

[11][19]Azoulay P, Zivin J S, Sampat B N. The Diffusion of Scientific Knowledge across Time and Space: Evidence from Professional Transitions for the Superstars of Medicine[J]. *National Bureau of Economic Research*, 2011: 1-4.

[12]Jaffe A B, Trajtenberg M. International Knowledge Flows: Evidence from Patent Citations[J]. *Economics of Innovation and New Technology*, 1999, 8(1-2): 105-136.

[13]Branstetter L. Exploring the Link between Academic Science and Industrial Innovation[J]. *Annals of Economics and Statistics*, 2005, 79(80): 119-142.

[14]Belenzon S, Schankerman M. Spreading the Word: Geography, Policy, and Knowledge Spillovers[J]. *Review of Economics and Statistics*, 2013, 95(3): 884-903.

[15]盛智明 超越定量与定性研究法之争——KKV对定性研究设计的启发[J]. *公共行政评论*, 2015(04): 11-23.

[16]Wooldridge J M. *Econometric Analysis of Cross Section and Panel Data*[M]. Cambridge, MA: MIT Press, 2002: 108.

[17]Wooldridge J M. *Introductory Econometrics: A Modern Approach*. 3rd Edition[M]. Cincinnati: South-Western College Pub, 2005: 323-324.

[18]Cameron A C, Trivedi P K. *Microeconometrics: Methods and Applications*[M]. New York: Cambridge University Press, 2005: 22-23.

[20]Kahn S, MacGarvie M. Do Return Requirements Increase International Knowledge Diffusion? Evidence from the Fulbright Program[J]. *Research policy*, 2016, 45(6): 1304-1322.

[21]Ganguli I. Immigration and Ideas: What Did Russian Scientists “Bring to the United States?”[J]. *Journal of Labor Economics*, 2015, 33(1): 257-288.

[22]Lissoni F. Academic Inventors as Brokers[J]. *Research Policy*, 2010, 39(7): 843-857.

[23]刘云,程旖婕.基于文献引证的国际知识流动影响因素探究[J]. *科学学研究*, 2018(9): 12.

[24]李平,宫旭红,张庆昌.基于国际引文的技术知识扩散研究:来自中国的证据[J]. *管理世界*, 2011(12): 21-31

[25]林波,杨秀财,侯剑华.知识流动视角下中美专利技术扩散的比较分析——以碳纳米管技术为例[J]. *情报杂志*, 2019, 38(11): 43-49

[26]孟晋宇,刘超,陈向东.海归学者跨国知识转移绩效的测度研究[J]. *管理学报*, 2017, 14(12): 1855-1861

[27]Iacus S M, King G, Porro G. Causal Inference without Balance Checking: Coarsened Exact Matching[J]. *Political Analysis*, 2012, 20(1): 1-24.