



简历分析法：一种教育实证研究新方法

刘进^{1,2}, 于宜田¹, 杨莉³, 林松月⁴, 李志峰⁵, 高媛², 陈恺哲¹

(1. 北京理工大学 人文与社会科学学院, 北京 100081; 2. 南方科技大学 高等教育研究中心, 深圳 518055;
3. 清华大学 教育研究院, 北京 100084; 4. 香港中文大学 教育学院, 香港 999077;
5. 武汉理工大学 法学与人文社会学院, 武汉 430064)

摘要:虽然教育学属于独立的一级学科,但十分缺乏对本学科发展起支撑作用的、区别于其他学科的专门研究方法,而且由于大量引入其他学科的研究方法,又进一步限制了其学科方法体系、知识体系和理论体系的发展。近年来,简历分析法在全球兴起,试图通过对简历数据的分析,融合大数据等方法,综合研判人才的教育与成长规律、项目/政策实施成效等,这既为教育科学研究提供了实证新方法,也为教育学科建设提供了新契机。进入以 ChatGPT 为代表的人工智能时代,生成式人工智能模型+算力(超级计算)+应用场景的新研究范式逐步成熟,简历分析法也在教育研究应用上迎来新突破,有望成为教育学科的专门研究方法。在中国教育场域中,蕴含着各级各类教育主体的简历和其他大数据资源,借助简历分析法的技术突破,有望全面解构本土教育规律,构建具有中国特色、符合中国国情的教育学科研究新范式,甚至有望通过碎片化专门知识累积,形成新的本土学科体系。最近十年,学术界涌现出不少基于简历分析法的研究成果,但已有研究对简历分析法这一方法本身的研究并不充分,尤其是对这一方法的学科归属、是不是教育学科的专门研究方法、如何与传统研究方法以及大数据、人工智能等新研究方法的融合等探讨不够。同时,我国已有研究在简历分析法概念使用、研究流程方面也存在很大差异,这表明简历分析法在中国教育学科的应用仍不成熟。为此,研究从方法论层面展开对简历分析法的

致谢:本文写作得到北京大学沈文钦副教授的帮助,特此感谢,文责自负。

修回日期:20231115

基金项目:国家自然科学基金面上项目“‘帽子’政策促进还是抑制了学术人才回流?——基于对 580 万份简历大数据库的人工智能(准)因果推断”(72374023);国家自然科学基金面上项目“‘一带一路’学术人才向中国流动的开放式‘推-拉’模型研究——人工智能方法的运用”(71774015);国家自然科学基金面上项目“政府奖学金能否提升来华留学生质量?——基于机器学习方法的‘一带一路’国家因果推断”(71974012);科技部国家外国专家项目高端外国专家引进计划“‘卡脖子’学术人才对华流动研究”(G20221780040)

作者简介:刘进,男,江苏东海人,北京理工大学人文与社会科学学院副研究员,博士生导师,南方科技大学高等教育研究中心访问教授,主要从事高等教育国际化和教育大数据与人工智能研究;

于宜田,女,河北邢台人,北京理工大学人文与社会科学学院博士生,主要从事高等教育国际化研究;

林松月,女,河北邢台人,香港中文大学教育学院博士生,主要从事高等教育国际化研究;

李志峰,男,湖北武汉人,武汉理工大学法学与人文社会学院教授,博士,主要从事学术职业与高校教师发展研究;

高媛,女,天津人,南方科技大学高等教育研究中心助理教授,博士,主要从事高等教育国际化、跨文化教育和高等教育理论研究;

陈恺哲,男,浙江杭州人,北京理工大学人文与社会科学学院博士生,主要从事高等教育研究。

通信作者:杨莉,女,湖南衡阳人,清华大学教育研究院博士生,主要从事研究生教育研究。

引用格式:刘进,于宜田,杨莉,等. 简历分析法:一种教育实证研究新方法[J]. 重庆高教研究,2024,12(2):87127.

Citation format: LIU Jin, YU Yitian, YANG Li, et al. Curriculum vitae analysis: a new approach to empirical research in education [J]. Chongqing higher education research, 2024, 12(2): 87127.

剖析,通过大量研究案例尝试全面解构这一研究方法,并尝试对其内涵与外延、具体操作流程等进行规范,对简历分析法在我国使用中存在的问题进行梳理,最后对该方法在我国的应用前景进行探讨,以期促进简历分析法在我国教育学科的规范使用和全面推广。

关键词:简历分析法;教育研究方法;实证研究;大数据;人工智能

[中图分类号]G640;G40032 [文献标志码]A [文章编号]16738012(2024)02008741

一、简历分析法有望成为教育研究的专门方法

从本质上讲,简历是“对各类资本积累过程的记录”^[1]。简历资料中既有相对个性化的人力资本积累,也有相对统一化的社会资本积累。简历存在狭义上的“简历”概念和广义上的“简历”概念。狭义上的“简历”,通常是指一份总结和记录高等教育专业人员研究活动信息的存储文件,是一种映射高等教育从业者个体轨迹的资源^[1]。随着简历分析法应用场景的不断扩大,广义上的“简历”应运而生,并与狭义上的“简历”出现两个主要区别:一是广义上的“简历”不仅包含高等教育专业人员的研究活动信息,也包含高等教育专业人员的各类非研究活动信息(如大学教师婚姻、子女状况等);二是广义上的“简历”开始从高等教育专业人员简历逐步推广到其他非高等教育专业人员,以及其他各级各类教育活动参与人员(如中小学教师、公务员等)的简历信息。无论是狭义上的“简历”还是广义上的“简历”,都是人们通常所认为的“个人简历”。此外,近年来“泛简历”数据也逐步被纳入简历分析法的研究范畴,比如储存于专门数据库的个人学术发表、专利申请等成果性记录,个人传记、采访报道、网络发言等观点性记录,学校管理部门保存的研究对象的学业表现,组织人事部门掌握的研究对象的职业表现信息等。如无特殊说明,本文所指的“简历”同时包括狭义上的“简历”、广义上的“简历”以及“泛简历”的范畴。

本课题组对简历分析法的定义是,利用研究对象的履历记录,获得纵向的事实数据,通过量化研究方法,形成面向特定研究问题的研究结论。简历分析法的主要目标是实现对研究对象“可输出内容”(如出版物、专利等)的量化分析,以评估研究对象的各类长期能力(如知识生产能力)。需要注意的是,也可以使用质性等研究方法对简历数据进行分析,但截至目前,学术界对简历分析法的使用仍主要停留在量化阶段。

本课题组认为简历分析法有望成为教育科学研究的专门方法。教育学通常被认为是我国社会科学领域基础较为薄弱的学科之一,加之教育理论与教育实践割裂^[2],教育学甚至被指责为“无用之学”“书斋学问”^[3]。教育研究质量不高的一个重要原因是科学研究方法的缺位。因此,教育学科迫切需要通过方法论创新提升学科地位^[4]——关键在于教育学科是否拥有区别于其他学科的专门研究方法^[5]。从此种意义上讲,探讨简历分析法是否有望成为教育研究的专门方法具有重要意义。

(一)简历分析法较早出现在教育学科,并在教育研究领域得到广泛应用

刘进在《简历分析法及其在高等教育研究中的应用》一文中对简历分析法产生与发展过程中几项具有标志性的研究成果进行了详细介绍^[6]。从起源来看,20世纪50年代,Knapp等分析了美国科学家本科毕业高校的来源,他们的数据来源渠道是1944年出版的《科学美国人》(*American Men of Science*)辞典^[7],虽然该辞典并不是通常人们所认为的科学家简历,但它实质上是一类非常重要的“泛简历”数据。因此,一些研究认为,20世纪50年代的这项研究已经具备简历分析法的基本雏形。该文研究对象主要为高等教育专业人员,可以认为简历分析法初创于高等教育、科技政策、科学学等交叉

学科领域。自这项研究发表至今已有 70 余年历史,简历分析法在教育学科的应用场景、数据来源、与其配合使用的研究方法、使用中存在的问题及其解决方式等方面均有较大变化与发展,但研究对象仍然主要是高等教育领域的专业人员,高等教育研究的一些理论、方法和思维经常被应用于其中。由此,简历分析法与教育学科的联系也愈加紧密。

在应用场景方面,简历分析法最初主要用于研究科技人才的群体特征、职业发展轨迹、流动特征等。随着除科技人才之外的其他各行各业人才简历数据的可获得性不断增强,简历分析法的研究对象也由科技人才逐渐拓展到在校生以及由学校教育培养出来的进入劳动力市场的各种人才,如学生^[8]、企业高管^[9]、政界官员^[10]等。此外,简历分析法也逐渐被应用到项目/政策效果评估领域。早在 1988 年,Gerrard 等对加拿大某大学一个医学学士学位项目的实施效果进行评估就使用了简历分析法,该研究系统比较了医学学士学位项目毕业生与其配对组的毕业生(配对依据包括性别、入学年龄、高中所在地等)在职业发展结果上的差异^[11]。

在数据来源与研究方法方面,早期简历分析法通常只采用单一渠道的简历数据源。本课题组对近十年来采用简历分析法所形成的相关文献的梳理发现,从多渠道采集简历数据或综合使用简历数据与非简历数据正变得日益普遍。例如,关于人才政策的评估研究就同时使用了简历数据与地区生产总值数据等。此外,近十年来,单篇文献中与简历分析法配合使用的实证研究方法的数量一般多于 1 种。同时,从总体特征来看,实证研究方法较为多元,既包括描述统计分析方法,也包括常见的因果推断方法(如双重差分、断点回归、工具变量法与倾向得分匹配法)、社会网络分析、序列分析(sequence analysis)和生存分析法等研究方法。

当前关于简历分析法是否成熟还存在一定的争论。一些研究认为,简历分析法存在简历可获得性差、简历信息缺失、简历标准化程度低、编码难度大等^[6,12]缺陷,但随着大数据、人工智能等技术的迅速发展,简历分析法的一些缺陷正在得到改善。例如,针对简历标准化程度低这一问题,一些国家已经开始在国家层面建立标准化简历数据库,如葡萄牙、西班牙、挪威和巴西等^[1];对于编码困难,一些研究将人工智能自然语义分析方法应用于简历数据编码,也大大提高了编码效率。

本课题组通过对简历分析法自 20 世纪 50 年代以来在应用场景、与其他数据源及方法的结合使用、存在的问题及其完善过程等方面的梳理发现,教育学科是最早应用简历分析法的学科,且过去 70 年来简历分析法在教育研究领域的使用范围不断扩大、应用场景也不断得到拓展。此外,简历分析法也具有很好的方法延展性,与其他数据源、研究方法有较好的融合能力,即简历分析法可以对传统教育研究较难解决的问题做出科学分析,有望提升教育研究科学化水平,进而发展成为教育学科的专门研究方法。

(二)简历分析法是一种科学的教育研究方法

涂尔干认为,一门学科如欲发展成为合乎科学的学术,首先必须有确定的研究主题与对象,其次必须有科学的研究方法^[13]。袁振国也认为,科学化程度是衡量学科成熟的唯一标准,实证研究是教育学走向科学的必要途径。此处的实证研究指那些基于事实和证据的研究,其对立面是毫无依据的主观判断与情绪性的观点宣泄^[14]。因此,简历分析法的科学化水平是判断其能否成为教育研究专属方法的关键依据之一。

教育研究方法通常指的是在从事教育研究过程中所遵循的基本规则和采取的具体方式以及采取这种方式所依据的理论基础^[15]。它包括 3 个层次:(1)方法论(methodology)层次,表示研究者对教育的基本看法,涉及该方法的哲学基础,而且这种看法对研究者选择教育研究路径与具体研究方法具有指导作用。(2)研究路径或研究方式(approaches)层次,表示研究者认识事物的基本策略或途径,如定量研究与定性研究。定量研究的方法论基础是实证主义,主张认识过程可重复;它的哲学基础是

经验主义,认为世界是可感知的。定性研究恰好相反。(3)具体方法(methods)层次,表示研究方式的具体化,如调查法^[15]。从这3个层次分类来看,某种具体教育研究方法的科学性首先体现在它必须具有明确的方法论基础与哲学基础^[15]。就简历分析法而言,它主要强调对简历数据的量化分析,其在研究方式层次上属于定量研究,方法论基础与哲学基础分别是实证主义和经验主义^[15]。从这个角度看,简历分析法是一种科学的教育研究方法。

其次,作为一种具体的研究方法,简历分析法同样具有科学性。这种具体方法层次的科学性指的是,教育研究方法在研究问题提出、研究方案设计、数据收集和研究问题验证等各个环节均应该符合科学的规范,以认识事物真相^[16]。根据这一标准,简历分析法的科学性具体体现在以下4个方面。在研究问题提出方面,简历分析法非常适用于分析两类研究问题:一是简历分析法擅长的研究领域,包括人的职业生涯轨迹研究和人的流动研究;二是采用传统教育研究方法容易陷入困境(如面临研究数据缺乏或数据质量不高)的研究议题,传统研究一般采用问卷调查数据估计教育收益率^[17],但这种数据可能存在研究对象自我报告的最高学历、收入数据与真实情况不相符等问题^[18]。与之相比,简历分析法通过对“泛简历”数据(学校管理部门保存的研究对象接受教育与学位获得等教育背景信息,以及税务部门掌握的纳税记录)等的量化分析,可以较好地规避调查数据带来的回答误差,实现对教育收益率更为精确的估计。在研究方案设计方面,由于简历是个人经历过的所有教育阶段与职业发展阶段的长时间跨度数据,因此,简历信息具有纵向数据的特征与优势,这使得简历分析法能够对研究对象各项表现进行长跨度评估与(准)因果推断估计,从而有效弥补各类横断面数据及其相关研究设计的缺陷^[19]。在数据收集方面,简历分析法本质上是一种非介入式研究方法^[6]。对于人才流动等较为敏感的话题,传统问卷、访谈法等获得的数据可能因为被试的隐私心理、逆反心理而存在回答偏差^[20],而简历分析法通过非接触性方式^[21]收集学术人才流动的事实性数据,可以避免因获取的数据与实际情况存在偏差而导致研究结论科学性不足等问题。此外,不同于单一的问卷调查法面临调查规模小、问卷回收率低^[22]、样本代表性不足等问题,简历分析法通过与问卷调查、大数据、人工智能等研究方法的融合使用,能大幅增加研究的样本量,甚至能覆盖研究对象的全样本信息,这将显著降低基于抽样而造成的各类研究误差,提高研究的科学性水平。例如,可以根据简历信息中的姓名、机构等信息,精准面向全体研究对象开展问卷调查,实现超大规模问卷调查目标。与此同时,通过引入区块链技术,将基于网络简历构建的简历数据库与问卷调查数据进行匹配(无需调查研究对象简历中的已有信息),这将大幅减少问卷篇幅,有效提高问卷回收率。在研究问题验证方面,简历分析法的结果呈现方式更为简单,可以基于研究对象的全样本纵向信息,简单明了地呈现研究对象的群体性规律和个体性特征,而且更容易形成(准)因果推断式研究结论,这比传统研究方法构建各类研究模型、设置较多前置条件、主要呈现相关性结论更有优势。这种(准)因果推断式研究结论的呈现大大提高了教育研究的科学性水平,可以为教育决策提供更为直接的学理依据。

综上,本课题组认为简历分析法无论是在方法论层次还是具体方法层次均具有科学性的特征,简历分析法是一种科学的教育研究方法。

(三)简历分析法与教育学具有一致的研究对象与研究目的

如前所述,研究方法的科学性对于学科发展至关重要。同时,一门学科要确立自己的学科地位还应拥有区别于其他学科的专门研究方法。由于简历分析法与教育学具有一致的研究对象与研究目的,因此它还是教育学科的专门研究方法。

科学方法是“人们揭示客观世界奥秘、获得新知识、探索真理的工具”^[23]。根据适用范围或概括程度的差异,科学的研究方法可以分为3个层次:(1)哲学方法,适用于所有科学部门且应用最普遍

的方法;(2)一般研究方法,适用于部分学科的研究方法;(3)特殊研究方法,个别学科或领域专门使用的研究方法,如物理学中的光谱分析方法、化学中的电解法、心理学中的精神分析法、人口学中的人口普查法等^[23-24]。从具体的研究对象看,属于个别学科的专门研究方法是指各学科为研究其特定的研究对象而采用的特殊方法^[25]。根据科学方法与个别学科的专门研究方法的定义,在判断某种研究方法是否属于某个学科的专门研究方法或特殊研究方法时,需要同时考察两个条件:一是这种研究方法与某个学科的研究对象是否一致;二是二者的研究目的是否一致。基于这两个判断标准,邱均平等论证了网络数据挖掘法是网络信息计量学的专门研究方法^[26]。

借鉴上述分析思路,本课题组认为简历分析法可以成为教育学的专门研究方法。一方面,简历分析法和教育学的研究对象具有一致性。教育学以教育现象为研究对象^[27-28];简历分析法自创立之初,就主要用于教育和科技问题研究。目前,简历分析法已经被广泛应用于对各级各类教育现象的研究。例如,在高等教育领域,除可用于分析人才流动、职业生涯轨迹等研究问题外,简历分析法还可以广泛应用于大量相关领域,如项目评估、学生学科专业选择、高等教育国际化与全球化等。此外,简历分析法还可以广泛用于基础教育、终身教育、干部教育,以及学校教育对毕业生长期发展的影响等各级各类教育现象的研究。学生基础信息、学业表现信息、学校信息等构成天然的学生简历大数据库。虽然有时该数据库并不直接以传统“简历”的方式呈现,但仍然能通过简历分析法对学生各类简历信息进行提取、加工和分析,从而得到各类教育科学研究发现。简而言之,简历数据在教育研究领域几乎无处不在,简历分析法可以对各类教育现象和教育研究问题进行深入细致的分析。另一方面,简历分析法和教育学的研究目的也是一致的。相较于传统教育研究方法,简历分析法通过对教育领域广泛存在的简历信息及相关数据的分析、挖掘,能够从新的视角分析教育现象,揭示教育规律,指导教育实践。这也正是教育学研究的最终目的^[27,29]。

二、简历分析法的教育应用场景

前文主要从理论层面论证了简历分析法不仅是一种科学的教育实证研究新方法,而且有望成为教育学科的专门研究方法。这一部分将通过详细介绍简历分析法在教育领域应用的研究成果,从实践层面进一步佐证这一观点。近十年,随着大数据、人工智能等技术的快速发展及其在教育领域的推广应用,以及随机控制实验、双重差分、断点回归、工具变量法、倾向得分匹配法等因果推断方法在社会科学研究应用的不断增加,简历分析法与其他数据源及研究方法的混合使用日渐普遍。因此,本文将基于简历分析法擅长的研究领域与教育学科的重点难点问题,重点介绍简历分析法最近十年在人才流动、职业生涯轨迹、项目/政策评估和学校人才培养成效4个领域的应用情况(如图1)。此外,采用简历分析法的研究成果非常丰富,限于篇幅,本文仅选取这4个研究领域中部分具有代表性且能够反映简历分析法的广泛应用前景与最新发展趋势的研究成果进行介绍,以期在教育研究思路和研究设计等方面对读者有所启发^①。

(一)人才流动研究

简历通常包含学术人员在各阶段的履历资料。因此,简历分析法自创立以来,用途最广、成果最多、被认为最具潜力的研究领域就是人才流动研究。近年来,在一些涉及人才流动的学位论文中,通常将简历分析法作为论文的主要研究方法。简历分析法在人才流动领域的研究方向主要包括人才流

^① 本文在介绍相关文献的研究方法时,将重点介绍除简历分析法之外的其他定量或质性研究方法。这主要考虑两个因素:首先,该部分所选取的研究案例均采用了简历分析法;其次,在使用了简历分析法的研究成果中,它们可能并没有明确提及简历分析法这一研究方法,却较为详细地介绍了对简历及相关数据进行量化或质性分析的方法以及其他未使用简历数据的实证研究方法。

动特征与规律研究、人才流动的影响因素研究和人才流动的影响效应研究(见表1)。

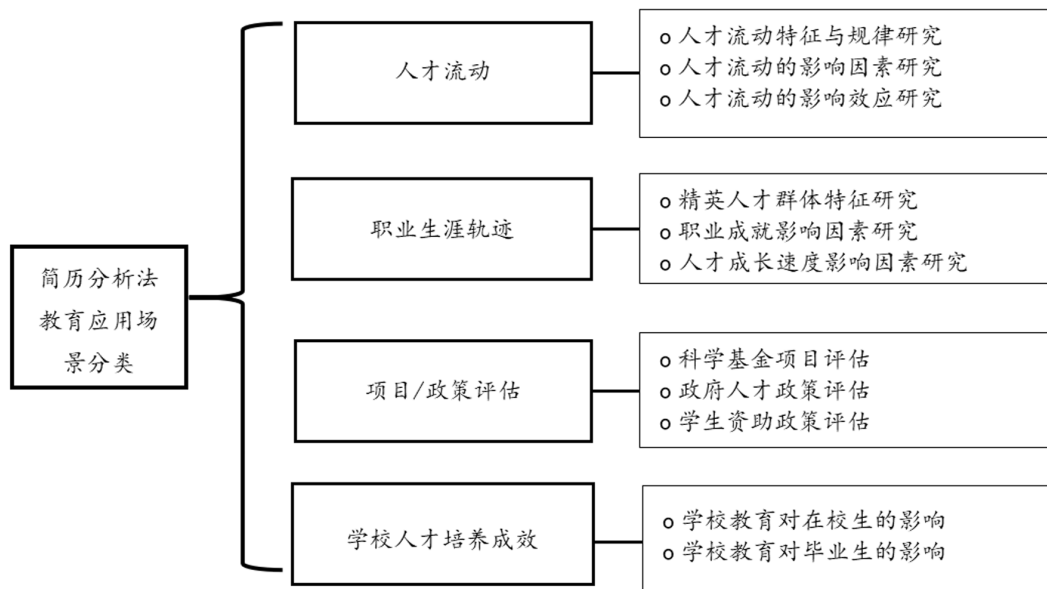


图1 简历分析法教育应用场景分类

1. 人才流动特征与规律研究

科学家是国家技术创新和经济增长的核心要素,简历分析法在发展初期的研究对象主要是科技人才。因此,这部分选取的前3项研究主要分析了科学家等学术人才的流动特征与规律。此外,简历分析法目前的研究对象已经拓展到非学术人才。例如,Gomez等使用ORCID(开放研究者和贡献者ID)数据描绘了全球科学家的流动规律。ORCID数据提供了科学家的发表历史,以及科学家的教育和就业信息。与常用于追踪科学家流动的调查数据相比,该数据在消除歧视文献数据方面具有更为明显的优势^[30]。Gomez等先对ORCID中2 814 360个ORCID传记资料数据进行筛选,得到满足以下条件的样本116 400个:(1)持有博士学位的研究者;(2)在1980—2010年获得博士学位;(3)在获得博士学位后有一些流动数据。然后他们基于筛选后的数据对全球科学家流动现状进行描述统计分析。研究发现,那些在职业生涯中选择国际流动的科学家,更加倾向在同一跨国地区内流动。此外,与国内科学家相比,尽管从20世纪80年代开始全球科学家的国际流动并没有增加,但科学家选择国际流动的频率更高、流动距离更短,而且在这些流动的科学家中,欧洲和亚洲的博士学位持有者占大部分^[30]。该研究的样本量为116 400个,这是传统的问卷调查研究难以实现的,表明简历分析法在研究样本量方面具有明显的优势。

Shi等对中国科学家的流动网络进行了分析^[31]。与Gomez等的研究采用大样本不同,Shi等的研究以2 406名中国科学院和中国工程院院士为中国科学家的代表,首先基于网络信息检索(中国科学院和中国工程院的网站以及工作单位的网站)、文献查询(如传记相关资料)、访谈和电子邮件查询等途径获取数据,使用简历分析法确定了这些院士的6条发展阶段信息(出生地、本科毕业地、最高学位获得地、第一份工作地点、获得院士头衔的工作单位所在地以及现职工作地),并构建了中国科学家空间信息数据库。然后,作者采用空间统计、复杂网络分析,以及修正的位置商数模型揭示中国科学家在每个发展阶段的空间分布特征,描绘了中国科学家的流动网络,分析了中国科学家流动背后的驱动机制。他们的主要研究结论为:(1)在不同发展阶段,中国科学家的空间分布趋于收敛。(2)从网络特性看,中国科学家成长的空间异质性明显,只有少数城市(如北京、上海和南京)具有较高的节点中心性、强度中心性和介数中心性,因此他们的网络层次结构呈金字塔形,“核心—外围”结构显

著。(3)在驱动机制方面,中国科学家的空间迁移受到国家政策、区域经济水平、教育水平和个人特质的综合影响^[31]。

表1 人才流动研究成果一览

参考文献	数据及其来源	实证研究方法	主要研究问题
[30]	发表、教育和就业信息;ORCID 数据	描述统计分析	全球科学家的流动规律
[31]	发展阶段信息:网络信息检索(工作单位的网站等)、文献查询(如传记相关资料)、访谈和电子邮件查询	空间统计、复杂网络分析,以及修正的位置商数模型	中国科学家(中国科学院和中国工程院院士)的空间分布特征、流动网络,以及流动背后的驱动机制
[32]	简历:高校官方网站	描述统计和倾向值得分匹配法	我国博士研究生就业地域特征
[33]	履历信息:中国经济金融研究数据库(CSMAR)	社会网络分析与修正型复合指标划分方法	长三角地区企业经营管理人才流动规律与特征
[34]	构建的独特数据集:美国国立卫生研究院(NIH)的个人简介、Who's Who(名人录)个人资料、医学期刊中的荣誉/讣告、国家科学院生平回忆录,以及Google 搜索的数据	离散时间危害率模型(discrete-time hazard rate models)	美国生命科学家的职业流动性特征与影响因素
[35]	简历信息:安达卢西亚科学信息系统(SICA)	t 检验、p 检验和 Mann-Whitney 检验	西班牙博士学位持有者国际流动的性别差异
[36]	四类信息(个人信息、工作和毕业大学的机构信息、大学的位置、中国教育部对环境科学/工程项目质量的官方评估):公开网站手动收集	线性回归模型、Heckman 偏差修正方法、中介效应模型	中国环境科学和工程领域研究人员毕业后的流动特征,以及院士对上述群体流动的影响
[37]	全国城市空气质量日报的数据:中国生态环境部 简历和企业层面的控制变量:中国经济金融研究数据库(CSMAR) 区域层面控制变量:中国国家统计局的统计年鉴	断点回归方法	空气质量与上市企业高管流动的关系
[38]	简历数据(文章未提及具体来源)	独立样本 t 检验、泊松回归	国际学术流动与获得长江学者头衔速度之间的关系
[39]	传记信息:研究人员个人网页公开的简历及教育、专业经历等信息 出版数据:Scopus 数据库	配对两样本学生 t 检验	年轻俄罗斯研究人员的国际流动对其未来出版活动的影响
[40]	简历;ORCID 网站 出版物记录:Web of Science 核心合集数据库	线性回归模型	美国相关科学家的流动在何种程度上破坏了他们研究合作的稳定性
[41]	简历:研究对象所在实验室官方网站上发布的信息 出版数据:Scopus 数据库	多期双重差分(DID)模型和负二项回归分析	招聘回国的“青年千人计划”科学家对同事(同一实验室的其他科学家)学术产出的因果效应
[42]	工作流动经历:中国经济金融研究数据库(CSMAR) 其他数据:万得(WIND)数据库	二阶段最小二乘法(2SLS)、双重差分法(DID)、倾向得分匹配(PSM)	高级管理团队的工作流动经历与企业创新行为之间的关系

注:由于每篇文章均采用了简历分析法,所以“实证研究方法”这一栏只介绍除简历分析法之外的其他实证研究方法(量化方法或质性方法),下同。

李和章等从博士毕业生就业去向的视角分析了我国学术人才的流动特征。他们基于国内各高校

官方网站公开的 97 134 条大学教师简历,采用描述统计和倾向值得分匹配法(PSM)分析了我国不同区域博士研究生就业地域分布的总体特征与差异。研究发现,从博士来源看,在各地区博士研究生中,本地区籍贯的学生占比最高。从博士就业去向看,各地区博士毕业生属地就业特征较为明显;相较于东部,中部和西部地区博士毕业生属地就业比例更低,表明这两个地区存在一定的人才流失问题。从博士就业区域差异看,与东、中部地区相比,西部地区博士人才流失更多,而且总体上呈现逐渐增强的趋势^[32]。该研究表明,采用简历分析法探究我国博士人才的地域分布及人才流失问题具有明显优势。原因在于,相较于问卷法、访谈法,该研究的样本量更大,而且是对博士人才实际流动行为(而非流动意向)的分析。

除了用于分析学术人才流动特征外,简历分析法还被用于分析各个领域人才的流动与聚集规律。例如,朱鹏程等采用履历分析、社会网络分析和修正型复合指标划分方法,探究了长三角地区企业经营管理人才的流动规律与特征。文章的研究对象是公司的“董监高”团队成员,包括公司董事、监事和高级管理人员(如总经理、副总经理等)。作者从中国经济金融研究数据库(CSMAR)采集了这些群体的履历信息,该数据库是我国目前规模最大、信息最精准的金融经济数据库。最终用于分析的有效样本为 9 493 个。研究发现:(1)企业经营管理人才的空间分布存在明显的地域差异。(2)企业经营管理人才流动的网络节点层级显著,网络链接呈“Z”字形空间分布。(3)长三角地区企业人才流动比率整体偏低,整体呈现“一主四副”多核心的网络结构。(4)结合流动方向与活跃程度,长三角地区存在 4 种人才流动地域类型,且总体呈现“破碎化”的流动特征^[33]。

2. 人才流动的影响因素研究

这部分选取的前 3 项研究分别以美国生物学家、西班牙博士学位持有者和中国环境科学与工程领域研究人员为研究对象,最后 1 项研究则以企业界的高管为分析对象。这些研究较好地展现了简历分析法在不同国家、不同职业领域的应用情况。

Azoulay 等先综合美国国立卫生研究院(NIH)的个人简介、Who's Who(名人录)个人资料、医学期刊中的荣誉/讣告、国家科学院生平回忆录,以及 Google 搜索的数据,编制了美国 10 051 名精英生命科学家的独特数据集,然后采用离散时间危害率模型(discrete-time hazard rate models)描绘了生命科学家的职业流动性特征及其影响因素。研究发现,高产的科学家更有可能流动,但这种流动受到美国 NIH 资助的约束,如果他们最近获得了美国 NIH 资助,则不太可能流动。此外,家庭结构对科学家流动也有重要影响,科学家似乎不愿意在他们的孩子处于 14-17 岁时流动。相较于父亲,这种情况对于母亲来说更为明显^[34]。从这篇文章的数据采集方式来看,“泛简历”数据的来源渠道广泛。

与 Azoulay 等的研究综合了各个来源的简历数据不同,Cañibano 等的研究直接调用了安达卢西亚科学信息系统(SICA)储存的标准化格式的简历信息。SICA 源于新信息技术在西班牙公共研究机构和资助机构中的广泛使用。该系统以电子化和标准化的格式存储研究人员的简历信息,当申请公共资助、认证和其他评估时,研究者有义务将他们简历中的数据输入系统。这意味着存储在 SICA 中的简历信息基本覆盖了在该地区公共组织工作的所有研究者,以及与他们合作的私人机构的研究者。因此,该系统被广泛应用于学术研究。Cañibano 等使用来自该系统的 9 个研究领域的 10 349 个有效数据,采用 t 检验、p 检验和 Mann-Whitney 检验等方法,分析了西班牙博士学位持有者在国际流动率、临时访问频率、持续时间和目的地等方面在不同性别之间的差异。研究发现,在所有领域的流动性方面,女性的国际流动率高于男性。与男性相比,女性的国际访问频率较低,访问多发生于更年轻的时候以及职业生涯早期阶段,访问的持续时间较短,访问的目的地离家也更近^[35]。这项研究表明,标准化简

历数据库能够为科研活动提供极大便利,大幅缩减研究人员用于简历数据收集与编码的时间。

上述两项研究分别关注了美国与西班牙学术人才流动的影响因素,Peng 等的研究则关注了院士群体对我国研究人员流动的影响。Peng 等以公开网站手动收集的中国环境科学与工程领域研究人员的数据(包括个人信息、工作和毕业大学的机构信息、大学所在位置,以及中国教育部对环境科学/工程项目质量的官方评估4类信息)为研究对象,采用线性回归模型、Heckman 偏差修正方法(Heckman bias correction method)、中介效应模型分析了中国环境科学和工程领域研究人员毕业后的流动特征。研究发现,像院士这样的精英科学家可能是造成大学聘用自己的毕业生(“近亲繁殖”)的因素之一,研究人员工作单位离他们的毕业机构越远,其出版生产力越低^[36]。

除学术人才之外,简历分析法还可以用于分析非学术人才流动的影响因素。例如,杨金玉采用断点回归方法探讨了中国地级市的空气质量与上市企业高管流动之间的关系。该研究使用的数据有3个来源:(1)中国生态环境部公布的全国城市空气质量日报;(2)从中国经济金融研究数据库(CSMAR)获得高管简历数据和企业层面的控制变量的信息;(3)从中国国家统计局统计年鉴获得区域层面的控制变量的信息。其中,作者从CSMAR数据库获取了2007—2017年439 246份上市企业高管的简历,这些简历包括高管的职位类别、高管工作经历等详细信息。文章最终用于分析的样本为47 637位高管,共计179 393个“个体—年度”观测样本。研究发现:(1)空气污染显著提高了上市企业高管流动的可能性,他们更有可能向空气质量好的城市流动。(2)能力强的上市企业高管更有可能由空气污染城市向空气质量好的城市迁移;企业薪酬补偿、股权补偿和外界经济政策的不确定性则削弱了空气污染对上市企业高管流动可能性的影响。(3)不同类型的污染物均提高了上市企业高管流动的可能性,而且细微颗粒物(PM10和PM2.5)对上市企业高管流动可能性的正向影响更强^[37]。这篇文章的数据来源渠道多元,既包括大数据公司的数据库,又包括我国不同行政部门的数据;既包括个人层面的简历数据,又包括企业层面与区域层面的非简历数据。因此,该研究较好地展现了简历分析法具有与其他各类数据源进行融合应用的潜力。

3. 人才流动的影响效应研究

运用简历分析法不仅能够分析人才流动对流动者本人职业发展、工作成果的影响,还可以分析其对同事的绩效、工作单位的创新行为的影响。

Li等利用1 447名长江学者(不包括艺术、人文和社会科学领域的长江学者)的简历数据(文章未提及具体来源),采用独立样本t检验、泊松回归分析了国际学术流动与获得长江学者头衔速度之间的关系。研究表明,本地联系(被聘用为长江学者的单位是否与获得学士、硕士或博士学位的单位相同)加速了中国学者的职业发展,而国际学术流动对中国学者职业发展的影响可以忽略不计,甚至减慢了学者后期职业发展的速度;与本地同行相比,回国学者获得长江学者头衔的时间更长^[38]。

流动对流动者的影响不仅包括获得某项学术头衔的速度,还包括学术发表。例如,Kotsemir等的研究分析了研究人员的国际流动对其未来出版活动的影响。他们以在海外工作、实习或学习超过3个月后返回俄罗斯的年轻俄罗斯研究人员为主要研究对象,利用一个综合了传记信息(来自研究人员个人网页公开的简历及教育、专业经历等信息)和出版数据(来自Scopus)的数据库开展该项研究。文章采用了配对样本的研究思路:首先为每一位流动的研究人员寻找一位在科研领域、年龄和教育背景方面均非常接近的非流动样本(非流动双胞胎),共获得119组配对样本,然后采用配对两样本学生t检验比较这两类群体在各个科研产出指标上的均值差异。研究发现,流动的研究人员不仅发表的文章更多,而且还在更高级别的期刊上发表更多的论文,并获得更多的引用。此外,这种影响效应

在不同学科、不同流动持续时间和不同目的地之间均存在显著差异^[39]。该研究的一个非常显著的特征是采用了配对样本的研究思路,这表明简历分析法也可以设置实验组与对照组,通过比较两组特征相似的样本在结果变量上的差异,从而得到(准)因果推断式的研究发现。

Zhao 等分析了科学家的流动在何种程度上破坏了他们研究合作的稳定性(用“流动之后原作者的比例”衡量)。他们先从 ORCID(开放研究者和贡献者 ID)网站上获取 4 343 名与美国相关的科学家(至少在两个机构工作过,其中至少有一个机构在美国)的简历,并从 Web of Science 核心合集数据库中获取这些科学家的出版物记录,然后采用线性回归模型对数据进行量化分析。研究发现,科学家流动的次數越多,越倾向于与以前的合作者共同撰写文章;与国内流动相比,跨国流动对科学家研究合作稳定性的破坏更大^[40]。

上述 3 个研究主要采用简历分析法分析了学术人才流动对其自身的影响,比如学术人才获得学术头衔的速度、科研生产力,以及研究合作的稳定性,接下来要介绍的两项研究则采用简历分析法研究了人才流动对其同事及所任职企业的影响。例如, Yang 等以 45 个中国国家重点实验室(State Key Laboratories, SKLs)的 1 006 名科学家为研究对象,采用多期双重差分(DID)模型和负二项回归分析探讨了招聘回国的“青年千人计划”(TYTS)科学家对同事(同一实验室的其他科学家)学术产出的因果效应。文章所用数据有两个来源:一是作者根据 45 个中国国家重点实验室官方网站发布的信息,收集这些实验室所有中青年科学家的简历,然后采用简历分析法建立科学家背景信息(包括性别、年龄、工作经验、学术职位等)数据库。其次,作者从 Scopus 数据库收集每一位科学家 2009—2018 年的出版数据。研究发现,虽然回国的“青年千人计划”科学家对同事的论文发表总数不存在显著影响,但可以增加同事的高被引论文数量。异质性分析发现,对于那些更能适应本地学术环境的科学家而言,“青年千人计划”科学家的同伴效应更强^[41]。该研究表明,采用简历分析法评估我国海外人才政策的实施效果,不仅可以分析人才政策对回国人才本身的影响,还可以分析这类群体回国后对其同事的影响,这有利于全面评估政策实施成效。

Jia 等采用二阶段最小二乘法(2SLS)、基于高级管理团队(TMT)成员突然死亡的双重差分法(DID)以及倾向得分匹配(PSM)探讨了 TMT 成员的工作流动经历与企业创新行为之间的关系。该研究不仅表明简历分析法可以用于分析人才流动对企业的影响,还说明简历分析法的研究对象已经拓展到非学术人才领域。他们的原始数据来自万得(WIND)数据库和中国经济金融研究数据库(CSMAR),其中,工作流动经历的数据是从 CSMAR 提供的 TMT 成员的简历中提取。作者对 TMT 成员的平均企业变更次数、平均岗位变更次数和平均职能变更次数进行主成分分析得到一个工作流动经历指数,并用这一指数综合衡量 TMT 成员的工作流动经历。研究发现,那些 TMT 成员的工作流动经历更丰富的公司具有更好的创新能力。此外,工作流动经历对探索性创新策略的投入有积极影响,这些策略能够产生新的知识^[42]。同时,这篇文章还使用了简历分析法和 3 种因果推断方法,充分体现了简历分析法可以与其他方法融合应用。

(二) 职业生涯轨迹研究

简历分析法早期主要用于科技人才的职业生涯轨迹研究,被认为是评估科学家和工程师职业生涯发展的“独特的、潜在的、有用的方法”^[43]。随着简历分析法的不断发展,其应用范围也不断拓展到对各行各业人才生涯轨迹的研究。下文将从精英人才群体特征、职业成就影响因素和人才成长速度影响因素 3 个方面对职业生涯轨迹相关研究成果逐一介绍(见表 2)。

表 2 职业生涯轨迹研究成果一览

参考文献	数据及其来源	实证研究方法	主要研究问题
[44]	文章所有数据:学校官网、教育部网站、中国知网和 Web of Science 等数据库	描述统计、关键词可视化分析	我国教育学科高层次人才(长江学者)的成长规律
[45]	经济学学术精英数据:诺贝尔奖官网、经济学家个人网页的简历、经济学名人录、各种传记词典 社会学学术精英数据:简历、美国国家传记、《大英百科全书》和《社会行为科学国际百科全书》、Marquis Who's Who in America 等	序列分析(sequence analysis)	比较经济学和社会学学术精英的职业轨迹与流动模式
[46]	履历信息:“科学探索奖”官网、获奖者工作单位官网、新闻报道和论文数据库	描述统计	我国获得“科学探索奖”的青年科技人才的成长特征与规律
[47]	履历数据:校长的个人主页、维基百科、百度百科、面访询问,中国知网、Web of Science、ORCID 等数据库	统计手段与社会网络分析法	我国高层次科技人才(“双一流”建设高校的校长)的成长规律
[48]	简历:BoardEx 和 Bloomberg 数据库, BusinessWeek、公司网站、美国证券交易委员会的文件和其他公开可用的信息	序列分析、Ward's linkage 的层次聚类分析	福布斯 100 强公司首席执行官的整个职业生涯的就业轨迹
[49]	履历:官方公布 领军人才的学术文章、公开讲话和网络博客(文章未提及来源)	词频统计、描述统计	对我国传统媒体和新媒体领军人才的特质进行对比分析
[50]	作者姓名、作者所在机构、发表年份、期刊名称以及引用频次等:Web of Science 数据库	随机效应负二项模型(random-effects negative binomial model)	认知开放性和结构开放性对商业学者(business scholars)研究影响力的影响如何随着职业发展阶段的变化而变化
[51]	人口统计数据:手工收集公司年度报告(最主要的渠道)、公司网站、传记数据库、直接联系投资者关系部门或高级管理者/高级管理者办公室 CEO 薪酬数据、公司和行业数据(除外国销售数据之外):Standard & Poor's Capital IQ 外国销售的信息:汤森路透 公司董事会的数据:BoardEx	回归分析	美国化(Americanization)与欧洲企业 CEO 薪酬水平之间的关系,以及 CEO 权力如何调节这种关系
[52]	简历和出版物信息:社会学家的个人网页	嵌套的 Cox 回归(nested Cox regressions)	绩效因素(学术生产力)以及非绩效因素(如性别等先赋特征)如何影响德国社会学终身教职的获得
[53]	朱卷履历:《清代朱卷集成》和《未刊清代朱卷集成》	逻辑斯特回归、线性回归、单因素方差分析	教育条件、家庭文化资本对寒门子弟考取科举功名的影响
[54]	候选人与提名人的基本信息:诺贝尔奖官网 候选人论文相关的信息:Web of Science 网站	Kaplan—Meier 分析法和 Cox 回归模型	诺贝尔奖候选人获奖概率的影响因素
[55]	出版数据:Web of Science 职业数据:商学院网站公布的简历、ProQuest 数据库的论文和论文索引、LinkedIn 的个人资料	负二项回归和中介效应模型	流动、研究职业资本(research-career capital)、性别与欧洲商学院教职员工学术晋升速度之间的关系

续表

参考文献	数据及其来源	实证研究方法	主要研究问题
[56]	简历数据:ReaD、科学家个人网站、问卷 调查数据 出版物数据:Web of Science	Cox 比例风险模型 (Coxproportional hazard model)	国际研究访问(international research visits)对日本科学家晋升的影响
[57]	简历信息:中国工程院官网、高等院校 和科研机构的官网	生存分析法	院士成长速度的影响因素
[58]	文章所有数据:公司年报中的顶级管理 者的简历、公司网站、传记数据库、Boar- dEx, Who's Who、Lexis Nexis	Ward's linkage 的层次 聚类分析和普通最小 二乘法	职业多样性与德国 DAX-30 公司的 顶级管理者达到董事会层级职位 所需时间之间的关系

1. 精英人才群体特征研究

精英人才主要指具有如下特征之一的人才:(1)获得行业内的人才称号或国际国内重要奖项。例如,在学术界,我国的人才称号包括院士、长江学者等;国际重要奖项包括诺贝尔奖等。(2)在工作单位担任高级领导,如学校校长、企业 CEO 等。(3)其他各行各业公认的领军人才。上述归类的主要目的并不是为了对精英人才进行概念界定,而是为了突出这一部分所选取的几项研究具有一定的内在联系及其各自的代表性,同时也为了说明简历分析法的应用场景广泛——能对不同精英群体的职业发展轨迹(工作经历)、人口统计学、教育背景等特征进行较为深入细致的分析。

在获得人才称号的精英人才方面,宋晓欣等以 22 位教育学科的长江学者为研究对象,采用描述统计、关键词可视化分析的方法探讨了我国教育学科高层次人才的成长规律。作者从研究对象所在学校官网、教育部网站,以及中国知网和 Web of Science 等数据库获取研究所需的数据。研究发现,教育学科高层次人才成长的内部条件包括关注我国教育发展过程中的重点问题、具有专注的科研精神(第一篇高质量论文与目前研究方向一致),以及在学习与实践中的学习与反思 3 个方面;外部环境主要包括高等教育阶段具有交叉学科背景、研究生阶段就读知名高校的优势学科和博士阶段师从名师 3 个方面^[44]。该研究的一个明显特征是样本量小,仅有 22 个样本,但作者对这些样本进行深入分析并获得了有价值的研究发现。这意味着,简历分析法具有成本可控的优势^[6],即在研究预算有限的情况下,同样能够采用简历分析法开展高质量的研究。

在获得国际重要奖项的精英人才方面,Korom 以 79 位经济学诺贝尔奖获得者和 51 位高被引社会学家为研究对象,采用序列分析(sequence analysis)方法比较了这两个学科学术精英的职业轨迹与流动模式。该研究从多个渠道采集了学术精英的传记数据并创建了全面的传记数据库。其中,经济学学术精英的数据来自诺贝尔奖官网的官方传记、经济学家个人网页的简历、经济学名人录(Who's who in economics)的各种版本,以及各种传记词典。社会学学术精英的数据来源包括:简历、美国国家传记、《大英百科全书》和《社会行为科学国际百科全书》的传记条目,以及 Marquis Who's Who in America 的多个版本、芝加哥大学图书馆对某个著名社会学家档案论文的指南中包含的“传记笔记”、奥地利社会学历史档案馆的在线传记、国家科学院学者的“传记回忆录”等。研究发现:(1)与社会学家相比,经济学家在获得博士学位后会更早地获得第一个教授职位,他们的职业生涯以向上流动为主要特征;在社会学中,博士学位获得者的职业生涯水平的流动更加明显。(2)经济学家在后续的教授职位获得上花费的时间更少。(3)在职业目标和流动倾向方面,经济学家更加倾向在该学科全球排名前五的系所担任教授职位并进行研究访问,而社会学的学术职业则不那么依赖于有声望的研究机构^[45]。可以发现,该研究的数据来源渠道广泛,能够在简历数据收集方面为研究人员提供参考。

在获得国内重要奖项的精英人才方面,高瑞等以我国 150 名在 2019—2021 年获得“科学探索奖”的青年科技人才为研究对象,采用履历分析法、描述统计等方法,从基本特征、高等教育经历、工作经

历和成长周期4个方面揭示了我国青年科技人才的成长特征与规律。作者先从“科学探索奖”官网获取上述获奖者名单,然后从“科学探索奖”官网、获奖者工作单位官网的“个人简介”,以及新闻报道和论文数据库等途径搜集获奖者的履历信息,并建立获奖者履历数据库,最后对数据进行量化分析。研究发现:(1)在基本特征方面,“科学探索奖”获得者的女性比例远低于男性,获奖者的年龄集中在40~45岁。(2)在高等教育经历方面,约77%的获奖者本科就读于我国一流大学建设高校(A类),在国内获得博士学位的比例约为63%,在同一所学校完成本—硕—博3个阶段教育的比例较低,为23%。(3)在工作经历方面,约七成获奖者具有博士后工作经历,其中仅有20%左右的获奖者具有职业流动经历。(4)在成长周期方面,获奖者的成长周期主要集中在毕业后11~16年^[46]。

在学校担任领导的精英人才方面,屠羽等探讨了我国高层次科技人才(高校校长)的成长规律。他们以137位中国“双一流”建设高校的校长为研究对象,首先通过校长的个人主页、维基百科、百度百科、面访问问等方式获取这些校长的履历数据,然后基于中国知网、Web of Science、ORCID等数据库中的科研成果的作者、单位等信息对缺失内容进行补充,并建立校长履历数据库,最后综合采用统计手段与社会网络分析法对“双一流”建设高校校长的成长规律进行量化分析。研究发现:(1)“双一流”建设高校校长在境内大学获得博士学位的比例非常高,在境外从事博士后研究工作的比例(71%)远高于境内(11%)。(2)大多数“双一流”建设高校校长拥有在政府(67%)、其他高校或科研院所(64%)的工作经历,但所有校长均没有企业工作经历。(3)美国是“双一流”建设高校校长境外流动的主要目的国家和中介国家^[47]。

除了关注学术精英人才的群体特征外,简历分析法还可以用来分析非学术精英群体,下面以企业的首席执行官(CEO)和新闻行业的领军人才为例加以说明。Koch等研究了福布斯排行100强企业CEO的整个职业生涯的就业轨迹。作者从BoardEx和Bloomberg数据库收集关于企业CEO的简历数据,并与来自BusinessWeek、公司网站、美国证券交易委员会的文件和其他公开可用的信息进行交叉检查和补充。文章的分析思路分为两个步骤:首先采用序列分析(sequence analysis)确定这些CEO的职业发展路径特征及模式,然后采用层次聚类分析(hierarchical cluster analysis with Ward's linkage)识别导致个体达到最高管理层级的职业路径的不同组别。结果显示,福布斯排行100强企业CEO的职业生涯大体上遵循传统的职业发展路径,这些发展路径的主要特征是:职责稳步提升,较少在企业 and 行业之间流动,重视一般管理职能^[48]。

柯春晓对新闻领军人才的特质进行了对比分析。他们以官方公布的传统媒体和新媒体(各30位)领军人才的完整履历,以及领军人才的学术文章、公开讲话和网络博客为研究对象,首先采用词频统计(一种内容分析方法)对上述领军人才的简历进行运算,以发现两类领军人才的总体特征,然后从不同方面对领军人才的特征进行更为细致的分析与比较。研究发现,在总体特征方面,传统媒体的领军人才履历文本具有“行政事业单位”话语体系的特征,而新媒体领军人物的履历信息则更加互联网化;传统媒体和新媒体领军人才都具有中国媒体人的共同特征。具体来看,在两类领军人才中,男性占比均远高于女性,而且新媒体领军人才更加年轻化;“科班出身”、高学历、新闻报道能力是两类人才的共同特征,二者在改革创新能力和工作变迁方面存在差异;传统媒体和新媒体领军人才在身份地位特征、影响公众的方式上有较大差异;二者在品行特征、人际特征方面具有较多共性^[49]。该研究表明,简历分析法既可以对简历数据进行量化分析,又可以对简历数据进行质性分析。

2. 职业成就影响因素研究

职业成就包括两层含义:第一,职业发展过程中产生的实际工作成果,如学术人员的学术产出(数量与质量)。第二,外界对工作成果的评价。其评价指标包括薪酬、岗位升迁、人才称号或奖项的

获得等^①。这一部分所选取的案例涉及的职业成就主要包括学者的研究影响力、企业员工的薪酬、学术人员终身教职获得、中国古代举人考取进士,以及学术人才获得诺贝尔奖等。这些文献运用简历分析法探讨了影响上述职业成就的因素,以下是相关研究成果的具体介绍。

Belkhouja 等采用随机效应负二项模型(random-effects negative binomial model)研究了认知开放性和结构开放性对商业学者(business scholars)研究影响力(research impact)的影响如何随着职业发展阶段变化而变化。文章数据来自 Web of Science 数据库中经过同行评审发表在期刊上的论文,这些论文包括作者姓名、作者机构、发表年份、期刊名称以及引用次数等,最终得到在 1997—2013 年至少发表过两篇文章的 35 296 位商业学者的“个体—年度”层面的数据。认知开放性采用每个作者截至当前年份所发表的期刊论文涵盖的不同学科数量来衡量。结构开放性是指学者与自己所在机构以外的学者进行合作的程度。研究影响力指学者在某一年所发表论文的引用次数。研究发现:(1)商业学者的认知开放性和结构开放性与其研究影响力之间存在非线性关系。(2)适度的认知开放性和结构开放性有助于增加年轻学者的研究影响力,而高水平的认知开放性和低水平的结构开放性则对年长学者的研究影响力具有正向影响^[50]。该研究表明,综合使用简历分析法与其他实证研究方法,不仅可以对全样本展开分析,还可以比较不同群体之间存在的差异。

Schmid 等采用回归分析法研究了美国化(Americanization)与欧洲企业 CEO 薪酬水平之间的关系,以及 CEO 权力如何调节这种关系。他们通过不同途径采集研究数据:(1)人口统计数据包括 CEO 的国籍、教育背景、工作经验和董事会任命次数,这些数据是由手工收集公司年度报告(最主要的渠道)、公司网站、传记数据库(如欧洲的 Who is Who 和 BoardEx),以及直接联系投资者关系部门或高级管理者/高级管理者办公室得到。(2)CEO 薪酬数据、公司和行业数据(除外国销售数据之外)来自 Standard & Poor's Capital IQ。(3)外国销售的信息来自汤森路透(Thomson Reuters)。(4)公司董事会的数据来自 BoardEx。最终得到具有完整信息的样本为 303 个。作者根据 CEO 是否为美国国籍、在美国接受高等教育年限、在美国的工作经验年限和被任命为美国公司董事会的次数 4 个指标,构建了一个复合指数用于衡量 CEO 的美国化程度。研究发现,CEO 的美国化、公司的美国化和行业的美国化均能显著提高欧洲企业 CEO 的薪酬水平。当 CEO 权力较大时,美国化对欧洲企业 CEO 薪酬的正向影响可能更大^[51]。该研究表明,多样化的数据来源渠道和数据种类(个人、公司与行业层面)有利于拓展简历分析法的应用场景。

Lutter 等采用嵌套的 Cox 回归(nested Cox regressions)分析了绩效因素(学术生产力)以及非绩效因素(如性别等先赋特征)如何影响社会学终身教职的获得。文章的研究对象是当前在德国大学任职(2013 年)的社会学家。作者从这些社会学家的个人网页获取他们的简历和出版物信息,最终得到一个包含 1 260 位社会学家的 28 545 篇出版物的非平衡面板数据集。研究发现:学术产出与能否获得社会学终身教职之间存在密切关系,每一篇经过同行评审的期刊论文(refereed journal article)和每一部专著会使社会学家获得终身教职的机会增加 10%~15%;对于非绩效因素而言,网络规模、个人声誉、性别均与终身教职的获得有关;与男性相比,女性在获得首个终身教职时平均出版物数量比男性少 23%~44%;在其他条件相同的情况下,女性获得终身教职的可能性大约是男性的 1.4 倍^[52]。

与上述研究主要分析当代人物职业成就的影响因素不同,林上洪等分析了教育条件、家庭文化资本对寒门子弟考取科举功名的影响。这篇文章的科举功名指通过科举考试获得的有做官资格的科名

^① 需要说明的是,工作成果只是影响薪酬、岗位升迁、人才称号或奖项获得的重要因素之一,这也是研究工作成就的各种可能影响因素的前提。

身份。朱卷履历类似于记录科举功名者学习成长经历的简历,是人文教育研究领域可用于量化研究的重要素材。林上洪等从《清代朱卷集成》和《未刊清代朱卷集成》两套科举文献中获取 1 194 个浙江籍进士功名者和举人功名者的朱卷履历,采用受业师功名值(综合考虑受业师的数量与受业师的功名身份)衡量科举功名者的教育条件,采用家庭文化资本值(父亲和祖父的功名身份量化值之和)作为家庭文化资本的衡量指标。在此基础上,采用逻辑斯特回归分析影响寒门子弟考取科举功名的因素。研究发现,教育条件、家庭文化资本均能够显著提高举人考取进士的概率。此外,线性回归分析发现,家庭文化资本对教育条件具有显著的正向影响。单因素方差分析结果表明,与贡生家庭、举人家庭和进士家庭的子弟相比,考取科举功名的寒门子弟的教育条件更差^[53]。该文章的独特之处在于,它将简历分析法的研究对象从当代人物拓展到历史人物,这意味着简历分析法在教育领域的应用前景广阔,未来可能会有更多学者应用简历分析法对历史人物展开研究。此外,该研究未来可以深化的一个可能方向是,将简历分析法的研究样本从目前的浙江籍科举功名者扩展到全国所有科举功名者,并比较不同省份或地区的差异,这或许能够得到更加有意思、有价值的研究发现。

职业成就除了包括前文提到的研究影响力、薪酬、岗位晋升、考取进士外,还包括获奖。例如,孙玉涛等分析了影响诺贝尔奖候选人获奖概率的因素。他们从诺贝尔奖官网获取诺贝尔物理学奖候选人与提名人的基本信息,包括诺贝尔奖候选人的姓名、出生日期、被提名时所在的国家、是否获奖,以及提名人的姓名、身份(学术头衔或职务)等。在此基础上,作者还通过 Web of Science 网站收集了诺贝尔物理学奖候选人论文相关的信息。文章采用生存分析法中的 Kaplan-Meier 分析法和 Cox 回归模型对数据进行了分析。研究发现,诺贝尔物理学奖候选人的单篇最高被引频次,以及诺贝尔物理学奖候选人提名人身份是瑞典皇家科学院院士或诺贝尔委员会成员的数量,均能够显著提高诺贝尔物理学奖候选人获奖的概率^[54]。

3. 人才成长速度影响因素研究

人才成长速度是指人才晋升到某个更高级别的岗位或获得某种人才称号或奖项的速度。这部分的前两项研究主要运用简历分析法分析学术人员晋升速度的影响因素,第三项研究探讨影响我国院士成长速度的因素,最后一项研究聚焦企业的顶级管理者获得董事会层级职位的速度。这些研究的分析对象分别来自欧洲、日本、中国和德国。

Ryazanova 等关注了流动、研究职业资本(research-career capital)、性别与学术晋升速度之间的关系。研究职业资本是指一个学者所有的同行评审出版物。他们采用负二项回归和中介效应模型对来自 10 个国家的 20 所高研究影响力(highly research-visible)的欧洲商学院 376 名终身教职员工样本进行了分析。文章的数据来源包括:(1)每位学者在 2018 年 9 月之前整个职业生涯的出版数据来自 Web of Science。(2)职业数据主要来自商学院网站公布的简历;对于缺失的职业信息,则使用 ProQuest 数据库的论文和论文索引、LinkedIn 的个人资料进行补充。研究发现,流动对研究职业资本有正面影响,但多次流动会延迟学术晋升;在性别差异方面,女性教职员工不太可能实现国际流动;女性学者晋升到终身教职和正教授的时间也更长^[55]。

与 Ryazanova 等的研究以 10 个国家的学者为研究对象不同,Lawson 等以 370 名日本生物科学教授为研究对象,探讨国际研究访问(international research visits)对科学家晋升的影响。他们从不同的渠道采集研究所需的数据。其中,简历数据主要从 ReaD 数据库获得——这是一个由政府机构创建的职业生涯数据库,科学家可以自愿存放他们的简历信息。此外,由于在 ReaD 数据库中存放简历并不是强制性的,作者还使用科学家个人网站的信息对简历数据进行补充,同时将所有简历都与通过问卷调查收集的信息进行核实。出版物数据由 Web of Science 中获得。文章采用 Cox 比例风险模型(Cox-

proportional hazard model)对上述数据集进行分析后发现,国际研究访问对科学家的晋升具有积极影响,能将等待晋升的时间缩短1年。对于同时改变工作的学者来说,这种积极的研究访问效应较弱,但对于近亲繁殖的学者来说,这种积极的研究访问效应较强^[56]。

前两项研究主要分析学术人员晋升速度的影响因素,与之不同,瞿群臻等利用250名中国工程院院士的简历数据,采用生存分析法研究了院士成长速度的影响因素。他们从中国工程院官网、高等院校和科研机构的官网获取院士的简历信息,包括个人信息(如姓名、性别、年龄)、教育信息(如高等教育经历、海外留学经历)、工作信息(如工作单位类型、职称),以及学术信息(如发表的论文数量)等。研究发现:(1)具有海外留学经历、最高学历授予单位为科研院所、发表SCI和EI论文的数量之和等均能显著提高院士从博士毕业或硕士毕业晋升到教授的速度。(2)学位类型为医学、获得国家最高科学技术奖和国家自然科学奖等5大奖数量之和也均能显著提高院士从教授增选为院士的速度^[57]。

此外,简历分析法还可以用于对非学术人才成长速度的研究。例如,Schmid等以德国DAX-30公司的256名顶级管理者为研究对象,采用Ward's linkage的层次聚类分析(hierarchical cluster analysis with Ward's linkage)和普通最小二乘法(OLS),研究了那些达到顶尖的管理者在被任命为董事会成员之前的职业多样性的差异,以及这种职业多样性与他们获得董事会层级职位所需时间之间的关系。文章的数据主要来自公司年报中的顶级管理者的简历、公司网站和传记数据库;如果上述方式得到的数据不完整,则进一步参考传记数据库,如BoardEx,Who's Who, Lexis Nexis。研究表明,存在4种顶级管理者群体的职业发展路径,这说明他们在进入董事会层级职位之前,其职业晋升路径具有多样性。此外,更高级别的职业多样性并不能缩短一个人晋升到公司高层的时间,相反,它会延长一个人晋升到高层的时间^[58]。

(三)项目/政策评估研究

由于简历包含研究对象教育背景、工作经历、学业成就、获奖等各个方面的纵向信息,所以通过简历分析法可以系统比较研究对象在外部干预前后所发生的变化,以此评估项目或政策的实施效果。因此,简历分析法被认为在项目评估研究中非常有用,“甚至比其他研究方法更好”^[59]。简历分析法在项目/政策评估领域的主要研究方向包括:(1)科学基金项目评估;(2)政府人才政策评估;(3)学生资助政策评估。下文将从这3个方面对相关研究成果依次进行介绍(见表3)。

1. 科学基金项目评估

竞争性研究基金是大学教授的主要研究资助来源。下文将简要介绍采用简历分析法评估中国、美国、卢森堡和苏联科学基金项目实施效果的相关研究,以展现简历分析法在这一研究领域的适用性。

乔锦忠等采用双重差分法分析了我国杰出青年科学基金项目资助对学者学术产出的影响。文章的研究对象为257名国家杰出青年科学基金获得者(实验组)及257名对照组样本,对照组的选取标准是:(1)未获得国家杰出青年科学基金项目及其他重点项目资助;(2)与实验组的研究学科领域一致;(3)与实验组年龄相差在 ± 3 岁内;(4)与实验组博士毕业年份相差在 ± 3 年内。文章所使用的数据及其来源渠道包括:入选学者的基本信息来自学校官网中的学者个人主页和其他相关专业网站;所有研究对象的SCI论文信息来自Scopus数据库。研究发现:(1)项目资助对学者学术产出质量的提升作用在资助后6-10年更为显著,对学者学术影响力的提升作用则在资助期内(1-5年)更为明显。(2)学者在申请项目过程中存在显著的申请效应和考核效应。在学术影响力方面,项目资助前2年对学者的学术影响力最大,而后平稳下降,同时在资助期结束后2年再次上升。(3)项目资助对学者学术产出的影响在不同学科之间存在显著差异^[60]。

表3 项目/政策评估研究成果一览

参考文献	数据及其来源	实证研究方法	主要研究问题
[60]	入选学者的基本信息;学校官网中的学者个人主页和其他相关专业网站 所有研究对象的SCI论文信息;Scopus数据库	双重差分法	我国杰出青年科学基金项目资助对学者学术产出的影响
[61]	资助数据:NIH资助数据库和Dimensions 研究产出数据:Web of Science、PubMed数据库	两阶段最小二乘法回归、广义精确匹配(coarsened exact matching)、模糊断点回归	美国国立卫生研究院(NIH)资助对初级科学家未来职业成果的影响效应
[62]	教授名单、职位与研究单位信息:卢森堡大学人力资源部 研究资助信息:卢森堡国家研究基金的年度报告 其他数据:教授的简历和主页、Web of Science	双重差分方法	CORE资助对卢森堡大学教授科研产出的影响
[63]	出版数据:Web of Science 资助者信息:匈牙利布达佩斯国际科学基金会(International Science Foundation)档案	模糊断点回归	一项大规模的研究资助计划对苏联科学家学术发表和决策的影响
[64]	文章所有数据:千人计划官网、ProQuest全球博士硕士学位论文数据库、个人网页、实验室网页、LinkedIn和Google Scholar、Web of Science、Dimensions	广义精确匹配(coarsened exact matching)、差分法(DID)和泊松模型	中国“青年千人计划”在招募精英海外科学家以及培育归国科学家生产力方面的效果
[65]	文章所有数据:国家自然科学基金委网站、政府信息公开文件、各单位公示文件、Scopus数据库、研究报告	描述统计、清晰(sharp)断点回归	地方政府(陕西省)高层次人才政策的实施效果
[66]	个人简历数据:58同城 人才政策数据:长三角城市共260份报告中与人才相关的内容 地区生产总值(GDP)与在岗职工平均工资等数据:《中国城市统计年鉴》	负二项回归和双向固定效应模型	长三角城市群人才流动的特征、人才政策对人才跨区域流动的影响
[67]	文章所有数据:Scopus数据库、万方、中国知网、国家图书馆数据库	有序逻辑回归模型、负二项分布模型、普通最小二乘法	海外交流对联合培养博士生职业生涯的影响
[68]	被引情况和H指数:Scopus和Google Scholar 学术出版物:Lattes简历数据库 其他数据:FAPESP、CAPES、CNPq和Lattes简历数据库	精确匹配(exact matching)、广义线性模型和准泊松模型(Quasi-Poisson)	巴西博士生奖学金对博士学位持有者科研表现的影响
[69]	大学层面的变量:美国教育部的综合高等教育数据系统、大学评分卡数据和Barron的大学指南 大学入学、学士和副学士学位的获得数据:美国学位信息中心 其他数据:负责管理磁性奖学金的办公室	线性概率模型	Cal U磁性奖学金(Cal U Magnet Scholarship,由加利福尼亚的一所公立大学为大学生提供)对学生大学入学决策和学位完成的影响
[70]	文章所有数据:Knox Achieves的学生参与项目记录和全州的P-20追踪数据库	双重差分和倾向得分匹配方法	美国田纳西州的Tennessee Achieves资助对高中毕业生进入大学及其在大学坚持下来的可能影响
[71]	连续6年(2009/2010学年至2014/2015学年)学生层面的人口统计和测试成绩:印第安纳州教育部	精确匹配、普通最小二乘法	美国印第安纳州选择奖学金项目(Choice Scholarship Program)对初中和小学高年级学生学业成绩的影响

Wang 等以在 1990—2005 年申请美国国立卫生研究院(NIH)资助的初级科学家为研究对象,采用两阶段最小二乘法回归(2SLS)、广义精确匹配(Coarsened Exact Matching, CEM)、模糊断点回归(fuzzy Regression Discontinuity, fuzzy RDD)方法分析了 NIH 资助对初级科学家未来职业成果的因果效应。文章中与资助相关的数据来自 NIH 资助数据库和 Dimensions,与研究产出相关的数据来自 Web of Science、PubMed 数据库。最终的分析对象为 623 名评估得分接近资助线但未获资助的申请者,和 561 名评估得分略高于资助线的成功申请者。作者比较了这些评估得分接近资助线的失败申请者和微弱胜出的成功申请者的未来职业成果。研究发现,失败申请者在未来十年内离开学术界(用在 NIH 系统中是否继续活跃衡量)的比例显著高于成功申请者;获得资助对科学家未来论文发表数量不存在显著影响,但对他们未来 1-5 年、6-10 年和 1-10 年内发表热门论文的概率,以及每篇论文平均被引频次均具有显著的负面影响,即使考虑两组的不同流失率,这种负面影响仍然存在^[61]。

CORE 是卢森堡最重要的针对个人研究者的研究资助项目,Hussinger 等以在卢森堡大学工作的 213 名大学教授为研究对象,采用双重差分方法分析了 CORE 资助对大学教授科研产出的影响。文章的实验组样本为至少获得 1 项 CORE 资助的卢森堡大学教授。对照组样本包括两组:一组是除实验组样本之外所有的其他卢森堡大学教授;另一组只包括与实验组中某位教授最相似的卢森堡大学教授。文章所使用的数据有多个来源:(1)卢森堡大学人力资源部提供的大学教授名单,以及这些教授的职位(助理教授、副教授或正教授)、研究单位。(2)大学教授的其他信息(如获得博士学位日期)来自教授的简历和主页。(3)从 Web of Science 检索的出版物记录中对大学教授的数据集进行补充。(4)从卢森堡国家研究基金(FNR)2001—2019 年的年度报告中获得基于 CORE 项目的 FNR 的研究资助信息。研究发现,研究资助能够促使大学教授的出版物总数增加 31%,这相当于多发表 1 篇论文。此外,大学教授在获得资助后 5 年左右,其科研产出数量会再次下降^[62]。这项研究的一个数据来源渠道是卢森堡大学的人力资源部,表明高校人事部门是“泛简历”数据的重要来源渠道,学者在条件允许的情况下,可以充分利用高校人事部门的数据资源开展研究。

评估科学基金项目实施效果的指标除了前 3 项研究涉及的受资助者的科研产出、是否继续留在学术界外,还包括对移民决策、政策的实施和效果评估。例如,Ganguli 评估了一项大规模的研究资助计划对苏联科学家学术发表和决策的影响。作者使用汤森路透 Web of Science 的出版数据和来自匈牙利布达佩斯国际科学基金会(International Science Foundation, ISF)档案的资助者信息,创建了一个科学家年度层面的数据集,最终得到 929 个有效分析样本。这些科学家的条件刚好低于或刚好超过获得资助资格的分界点,换言之,获得与未获得资助的科学家具有较强的可比性。作者采用模糊断点回归(fuzzy Regression Discontinuity)方法对上述数据进行了估计。研究表明,这项资助对科学家的发表活动具有非常大的积极影响,在项目实施后的 3 年内,科学家发表的文章数量增加了 1 倍以上;这项资助还使得科学家长期留在科学领域,降低了科学家的移民率;这项资助对科学家移民的负面影响仅限于莫斯科人^[63]。

由上可见,该部分的 4 项研究均具有一个明显的共同特征:它们均采用了配对的设计思路,这有利于更加科学地评估科学基金项目的实施效果。

2. 政府人才政策评估

近年来,我国各级政府出台了大量人才优惠政策,“抢人大战”“人才争夺战”等新闻不时见诸报端。因此,无论是评估中央政府出台的高层次人才政策,还是地方政府实施的高层次人才政策或一般的人才引进政策,都具有非常重要的政策意义。采用简历分析法开展对政府人才政策评估研究具有独特的优势。

Shi 等分析了中国“青年千人计划”在招聘精英海外科学家以及培育归国科学家生产力方面的效

果。文章使用的数据集主要包括3个部分:(1)青年千人计划招聘者及其海外对照组的名单和职业信息;(2)每位科学家在回国前和回国后的出版物及引文记录;(3)每位科学家的研究团队和资助信息。作者从以下渠道收集这些信息:(1)千人计划官网。(2)ProQuest全球博士硕士学位论文数据库。(3)个人网页、实验室网页、LinkedIn和Google Scholar。(4)ISI的Web of Science SCI-Expanded和SSCI(WoS-SCI/SSCI)数据库。(5)Dimensions研究资助数据库,主要用来获取每位科学家的研究经费信息。该研究采用广义精确匹配(coarsened exact matching)、差分法(DID)和泊松模型进行数据分析。研究的思路是:将每个归国人员与可比的“留任者”(stayers)进行匹配,即那些在中国上大学,并与归国人员在相同时期(± 3 年)内从同一海外大学获得同一领域的博士学位,但仍然留在海外学术界的科学家。研究发现,青年千人计划科学家在研究方面基本处于较高水平,但作为一个群体,他们在归国前的生产力并未达到顶级水平。此外,青年千人计划科学家回国后的出版物数量在各个期刊质量层次上都有所增加,但这种增加主要发生在最后署名的出版物中,以及高素质(尽管不是顶级素质)的新人中^[64]。这项研究估计了“青年千人计划”对招聘回国的海外科学家本人科研产出的因果效应。未来,该研究可以借鉴Yang等的研究思路,采用简历分析法分析招聘回国的海外科学家对本单位同事的学术影响^[41]。

董新宇等以陕西省高层次人才政策为例,采用描述统计、清晰(sharp)断点回归等方法分析了地方政府高层次人才政策的实施效果。作者以陕西省4项人才工程(陕西省“百人计划”、“三秦学者创新团队计划”、陕西省“高层次人才特别支持计划”和“三秦工匠计划”)的入选者为分析对象,首先从国家自然科学基金委网站、政府公开文件、各单位公示文件、Scopus数据库、研究报告等途径采集了这些入选者的信息,然后将上述数据结果与陕西省科技资源管理服务一体化云平台科技人才数据库进行核对,最终得到用于分析的样本量为701个。研究发现,地方高层次人才即使获得了政府支持,也难以成长为国家级人才;青年学者是地方高层次人才群体中最具发展潜力的群体;地方政府对高层次人才的支持对他们的年度论文发表数、篇均论文被引数均不存在显著影响^[65]。

与前两项研究主要关注高层次人才不同,马双等的研究将人才定义为具有大专及以上学历的人群。他们利用“58同城”的2万份个人简历数据、长三角城市共260份报告中与人才相关的内容、《中国城市统计年鉴》中的地区生产总值(GDP)与在岗职工平均工资等数据,采用负二项回归和双向固定效应模型分析了长三角城市群人才流动的特征、人才政策对人才跨区域流动的影响,以及这种影响在不同经济发展水平和不同收入水平城市间的差异。研究发现,长三角城市群城际人才流动十分频繁;人才流动规律由大城市间流动转变为大城市间流动与邻近城市间流动并重的局面;人才政策能够吸引人才流入本地,但这种政策效应仅限于高经济发展水平和高收入水平的城市^[66]。

3. 学生资助政策评估

前文从理论层面提到,简历分析法能够用于对各级教育现象的研究,这一部分将从实践层面佐证这一观点。在学生资助政策评估研究领域,简历分析法不仅能用于对高等教育阶段奖助学金政策实施成效的评估,还能用于对基础教育阶段奖助学金政策实施成效的评估。接下来介绍的5项研究所评估的学生资助政策依次为:国家留学基金委员会(CSC)的联合培养博士生项目、巴西的一个博士生奖学金资助项目、美国一所公立大学为大学生提供的奖学金项目、美国田纳西州的一项面向高三学生的资助项目,以及美国印第安纳州的一项面向初中和小学高年级学生的奖学金项目。

梁镇涛等以国家留学基金委员会选派的联合培养博士生(以下简称“联培博士生”)为研究对象,分析了海外交流对联培博士生职业生涯的影响。作者运用配对的设计思路,采用有序逻辑回归模型、负二项分布模型、普通最小二乘法对配对后的数据进行分析。具体的配对思路、数据来源与处理方法包括:(1)识别受资助学者群体。由于国家留学基金委员会规定,与资助相关的科研成果应注明该成

果得到国家留学基金委员会资助等信息,因此,作者先根据 Scopus 数据库中的论文致谢信息初步确定受资助学者群体。(2)识别联培博士生群体。将博士学位论文(来自万方、中国知网、国家图书馆数据库)中的作者姓名、所在机构、毕业年份等信息与第一步得到的数据进行匹配,精确识别出获得国家留学基金委员会资助群体中的联培博士生群体。(3)构建对照组。为每个实验组样本(联培博士生)选取对照组样本的依据是:与联培博士生所在机构、导师、专业、毕业年份完全相同,但未出国。最终得到实验组样本 769 个,对照组样本 1 407 个,其中 131 个实验组样本只能获得 1 个对照组样本,其他均有 2 个对照组样本。(4)采集研究对象的论文信息与简历。研究发现,与对照组相比,联培博士生在毕业后有更大的可能性继续在科研机构从事研究工作,他们的职称和任职机构排名均更高。此外,联合培养经历对博士生论文的发表数量与质量均具有显著的正向影响^[67]。

圣保罗研究基金会(São Paulo Research Foundation, FAPESP)是巴西最大的州级研究资助机构。Bin 等以在 2003—2017 年申请了 FAPESP 的博士生奖学金(以下简称“FAPESP 奖学金”)的个体(包括获得资助与未获得资助的个体)为研究对象,分析了巴西博士生奖学金对博士学位持有者科研表现的影响。他们的研究采用了精确匹配(Exact Matching)这一准实验方法,以及广义线性模型(Generalized Linear Models, GLM)和准泊松模型(Quasi-Poisson)等估计方法。实验组包括完成博士学位的获奖个体。作者设置了两组对照组:(1)申请 FAPESP 奖学金被拒,但通过其他资助机构的奖学金完成了博士学位;(2)申请 FAPESP 奖学金被拒,并且未获得任何奖学金的博士学位持有者。此外,文章还比较了 FAPESP 奖学金获得者与申请被拒群体中表现最佳的个体之间的科研表现差异。文章的数据来源包括:(1)博士学位持有者学术成果的被引情况和 H 指数来自 Scopus 数据库和 Google Scholar;(2)申请 FAPESP 奖学金后被授予者和被拒绝者的名单来自 FAPESP;(3)学士、硕士和博士学习期间的信息同时参考了 FAPESP、CAPES 和 Lattes 简历数据库;(4)学术出版物的数据也来自 Lattes;(5)博士阶段的信息(如博士研究知识领域)来自 FAPESP、CAPES、CNPq。其中,Lattes 被认为是一个非常丰富且强大的数据库,它提供了巴西科学技术领域的学生、教授、研究人员和专业人士的简历。研究发现,FAPESP 奖学金获得者的出版物数量、平均被引频次和 H 指数均高于那些申请被拒的个体。此外,相较于那些没有获得任何资助机构支持完成博士学位的个体,FAPESP 奖学金的效应更为显著^[68]。除了其他常用数据来源渠道外,该研究还使用了 Lattes 简历数据库的简历信息,表明标准化科技人员简历库可以给简历分析法的发展应用带来新的机遇。

Firoozi 采用线性概率模型(LPM)分析了 Cal U 磁性奖学金对学生大学入学决策和学位完成的影响。Cal U 磁性奖学金(Cal U Magnet Scholarship)是 2015 年由加利福尼亚州的一所大型公立大学(被称为“Cal U”)为大学生提供的。从很多方面看,Cal U 磁性奖学金与 Cal U 在其他年份提供的奖学金非常类似,关键的区别是,由于预算限制,Cal U 在符合条件的申请者中随机分配了 Cal U 磁性奖学金,这说明该项目实质上是一个单盲随机对照试验(single-blind RCT)。获奖者在前两年的学习中每年可以获得 3 000 美元的奖学金,并在 4 年内保证有校园住宿。文章使用的数据及其来源包括:(1)大学层面的变量来自美国教育部的综合高等教育数据系统(IPEDS)、大学评分卡(College Scorecard)数据和 Barron 的大学指南(Barron's College Guide)。(2)大学入学、学士和副学士学位的获得数据来自美国学位信息中心(National Student Clearinghouse)。(3)其他数据主要来自磁性奖学金管理办公室的记录。上述数据覆盖 2015—2019 学年,最终用于分析的样本数量为 1 500 个。研究发现,Cal U 磁性奖学金能够显著增加 Cal U 招收的弱势学生人数,但会显著降低加利福尼亚州其他公立大学招收的总学生数与弱势学生人数,且对学生学位的完成不存在显著影响^[69]。这项研究的数据来源与之前提到的研究均不同,它同时使用了学校与不同教育行政机构个人层面的行政数据,而这些都是非常重要的“泛简历”数据来源。

使用简历分析法同样可以评估基础教育阶段学生资助政策的实施效果。例如, Carruthers 等采用双重差分和倾向得分匹配方法评估了 Tennessee Achieves 资助对高中毕业生进入大学及其在大学坚持下来的可能影响。Tennessee Achieves 是美国田纳西州的一项资助项目, 该项目承诺, 自 2015 年的高中毕业班开始, 田纳西州将为每一位高中毕业生提供大学辅导和“尾款援助”(last-dollar aid), 以便他们能够顺利进入社区学院。与其他资助方式不同, Tennessee Achieves 资格既不基于需求又不基于优秀表现(need-based or merit-based), 因此 Tennessee Achieves 资助消除了其他资助方式可能存在的申请障碍。文章的数据来自 Knox Achieves 的学生参与项目记录和全州的 P-20 追踪数据库。P-20 数据库整合了田纳西州教育部、田纳西州高等教育委员会以及田纳西州劳工和劳动力发展部的行政数据。研究发现, Tennessee Achieves 资助能显著提高学生高中按时毕业和直接进入大学的可能性, 尽管帮助学生进入四年制大学的可能性略低。此外, 在对学生直接进入大学及其在大学坚持下来的可能影响中, 低收入学生群体受到的影响最大^[70]。该研究使用的个人层面的行政数据不仅包括教育行政机构的, 还包括非教育行政机构的, 这进一步拓展了简历数据的来源渠道。

Waddington 等研究了美国印第安纳州选择奖学金项目(Choice Scholarship Program)对初中和小学高年级阶段低收入学生的学业成绩的影响, 这些学生在该计划的前 4 年使用教育券(voucher)从公立学校转到私立学校。印第安纳州选择奖学金项目于 2011 年开始实施, 旨在为低收入和中等收入的印第安纳州家庭提供州政府的资助, 学生使用教育券可以抵消在参与该计划的私立学校就读的学费。文章使用的数据是连续 6 年(2009/2010 学年至 2014/2015 学年)学生层面的人口统计和测试成绩, 这些数据由印第安纳州教育部(Indiana Department of Education, IDOE)提供。作者首先采用精确匹配的方法为每一位获得奖学金并转入私立学校就读的学生(实验组样本)匹配了对照组样本, 每个实验组样本与其对照组样本在种族/族裔、性别、年级等方面完全相同, 且对照组样本来自与之配对的实验组样本之前就读的公立学校, 最终用于分析的样本包括 34 587 名对照组样本和 3 363 名实验组样本。然后, 作者采用普通最小二乘法(OLS)对上述配对数据进行了分析。研究发现, 与仍在公立学校就读的对照组学生相比, 使用教育券的学生在私立学校就读第一年时, 他们的数学成绩平均下降了 0.15 个标准差。无论在私立学校就读多长时间, 这种下降都持续存在。对于英语/语言艺术科目, 实验组与对照组的成绩之间均不存在显著差异^[71]。

(四) 学校人才培养成效研究

前文重点介绍了简历分析法最近十年在人才流动、职业生涯轨迹和项目/政策评估这 3 个领域的应用情况, 这一部分将详细介绍与学校人才培养成效相关的研究成果。采用简历分析法研究学校人才培养成效具有两大优势: (1) 教育系统能为简历分析法提供广泛的数据源。随着教育数字化的快速发展, 教育系统储存了海量的教育教学数据, 如学生行为、学业成绩、升学, 教师教学、职业发展, 学校的各类教学资源、教学条件与变迁, 以及教育行政部门的人事变动情况、未公开的政策文件等各类教育主体、各个层面的数据。采用简历分析法可以从这些数据中提取适用于各种不同教育研究议题的变量信息并开展相关研究。例如, 通过将不同主体的信息进行匹配, 可以分析教师、同伴或学校因素对学生学业表现及其行为的影响(短期影响), 以及对高等教育入学机会的影响(中期影响)。(2) 简历分析法还具有与其他各类数据整合应用的巨大潜力。如果将教育系统内部的行政数据与其他部门的行政数据(如公安部门的犯罪记录、税务部门的纳税记录), 以及实验数据、调查数据进行匹配, 可以进一步拓展简历分析法的应用场景, 如可以采用简历分析法探究学校教育对学生成年后犯罪、收入等的影响(长期影响)。

鉴于学校教育对学生发展的影响既有短期的, 又有中长期的, 因此, 这一部分将采用简历分析法开展的学校人才培养成效研究分为两类: 学校教育对在在校生的影响和学校教育对毕业生的影响(见表 4)。

表 4 学校人才培养成效研究成果一览

参考文献	数据及其来源	实证研究方法	主要研究问题
[72]	学生成绩:智利的教育质量评估系统(SIMCE) 学生家庭和学术背景;SIMCE 的调查数据 国家教师注册信息;智利教育部	增值模型	教师对学生成绩的影响是否因学生经济社会地位、种族和学生之前的成绩,以及学校而异
[73]	学生高中毕业信息、学生是否参加了大学入学考试和考试成绩等信息;美国田纳西州的 STAR 项目的数据 学生高中教育阶段后的入学与学位获得数据;美国学位信息中心 北卡罗来纳州公立学校的学生层面的纵向行政数据;北卡罗来纳州教育研究中心 其他数据:美国国家教育统计中心、美国劳工统计局	两阶段最小二乘法和 OLS 回归	拥有相同种族的小学老师对学生教育成就的影响
[74]	1985—2010 年丹麦所有进入高等教育阶段的学生的信息;多个行政登记册(administrative registers)	线性均值模型(linear-in-means model)	同伴能力对丹麦大学生退学决策的影响
[75]	大一学生的相关信息(如性别、户口类型、高考成绩)、校园卡餐饮消费记录;我国某大学行政管理部门	工具变量法	室友对我国大学生餐饮消费的影响
[77]	我国 J 市所有高中 2010 级全体学生的行政数据	普通最小二乘法和两水平线性模型	学校投入对学生学业成绩的影响及其群体差异
[78]	家庭背景信息;芬兰统计局的人口普查纵向数据 其他数据:年轻芬兰人心血管风险研究的数据;芬兰统计局的芬兰纵向雇主-雇员数据	OLS 模型和工具变量方法	受教育程度与成年人体育活动(physical activity)之间的关系
[79]	就读北卡罗来纳州公共学校的学生的个人层面数据;北卡罗来纳州公共教学部 个人层面的犯罪记录信息;北卡罗来纳州公共安全部门 选民登记和投票记录;北卡罗来纳州选举委员会 学校所在县的地方经济发展特征;美国社区调查的公开数据	线性概率模型和双重稳健的逆概率加权(inverse probability weights)方法	美国北卡罗来纳州中学特许学校对九年级学生的行为,以及年轻人犯罪和参加选举的可能性的影响
[80]	企业绩效数据;财经数据库 BvD-Osiris 简历;企业年报和官方网站	计量分析法	企业高管的一流大学背景对个人收入的影响
[81]	学生 18 岁以前的成绩和教育记录;国家学生数据(National Pupil Data) 学生在 18 岁之后所接受的职业/技术教育信息;个人学习记录(Individualised Learner Record) 大学教育数据;高等教育统计局 个人层面的收入信息;英国皇家税务和海关总署	OLS 回归和逆概率加权回归调整(inverse probability weighting regression adjustment)估计方法	接受高等技术教育(higher technical education)与接受大学学位(university degrees)教育的英国毕业生之间的个人收入差异
[82]	学位项目申请相关的数据;智利国家图书馆 公司领导层数据;智利证券和保险监管局 税务记录;智利税务局	回归断点设计(regression discontinuity design)和双重差分	智利精英商业学位项目对学生获得顶级职位和收入的影响

1. 学校教育对在校生的影响

从主体看,学校教育对在校生的主要影响因素包括教师、同伴和学校本身。为此,下文将依次介绍采用简历分析法探讨这3类因素对学生的影响的相关研究。需要说明的是,学生资助政策也是学校教育影响在校生的重要因素,但由于前文在项目/政策评估研究部分已经进行了介绍,此处不再赘述。

Canales 等采用智利的“学生-教师”匹配数据和增值模型(value-added models, VA models)评估了教师对学生成绩的影响是否因学生经济社会地位(socioeconomic status, SES)、种族和学生之前的成绩,以及学校而异。文章使用的数据来源包括:(1)学生的成绩数据是从智利的教育质量评估系统(Chile's Education Quality Measurement System, SIMCE)收集的。这是一个国家级的学生测试系统,用来评估各个年级学生的不同科目成绩。(2)SIMCE的调查数据。该数据能够提供关于学生家庭和学术背景的丰富信息。(3)国家教师注册信息来自智利教育部。文章最终的样本数量在以阅读成绩和数学成绩为因变量的模型中有所不同,在阅读领域,样本包括13 823名学生、905个班级、752名教师和358所学校;在数学领域,样本包括11 296名学生、716个班级、577名教师和274所学校。研究发现,在纳入所有控制变量的模型中,教师只会显著加剧不同学业基础学生之间,以及不同学校之间的数学成绩差距;教师对不同种族学生、不同SES学生的成绩差距,以及不同学业基础学生和不同学校之间的阅读成绩差距均不存在显著影响^[72]。与前文介绍的研究相比,该项研究在数据来源方面有一个新特点,即它同时采集了个人层面的行政数据与调查数据。

Gershenson 等采用两阶段最小二乘法(2SLS)和OLS回归分析了拥有相同种族的小学教师对学生教育成就的影响。文章使用的数据包括:(1)来自美国田纳西州的STAR(Student Teacher Achievement Ratio, STAR)项目(教育领域一个非常著名的关于班级规模的随机控制实验)数据,包括学生高中毕业信息、学生是否参加了大学入学考试和考试成绩等信息。(2)美国学位信息中心(National Student Clearinghouse, NSC)(NSC是一个非营利组织,是美国唯一一个提供学生层面的高中教育阶段后的入学和学位完成情况的行政数据来源)提供的学生高中教育阶段后的入学与学位获得数据。(3)北卡罗来纳州教育研究数据中心(North Carolina Education Research Data Center, NCERDC)提供的北卡罗来纳州公立学校学生层面的纵向行政数据。(4)美国国家教育统计中心的一些数据和美国劳工统计局的失业数据。研究发现,与没有黑人教师的黑人学生相比,在小学三年级拥有至少一位黑人教师的黑人学生更有可能从高中毕业,也更有可能进入大学;黑人教师对白人学生的长期影响并不显著^[73]。

前两项研究主要关注了教师对学生成绩、高中毕业与大学入学机会的影响,接下来两项研究分别分析了同伴对大学生退学决策和餐饮消费的影响。例如,Humlum 等采用线性均值模型(linear-in-means model)分析了同伴能力对大学生退学决策的影响。同伴是指在同一年被同一学校同一教育项目录取的学生,并且用高中GPA来衡量学生能力。作者采用1985—2010年丹麦所有进入高等教育阶段的学生信息。他们从行政登记册(administrative registers)采集并合并了这些数据,从而获得一个丰富的数据集,其中包括入学学生、学生的同伴和学生的父母,以及学生最终是否退学等信息。研究发现,同伴能力是大学生退学的重要影响因素,同伴的高中GPA每增加1个标准差,大学生退学的概率减少4.6个百分点。非线性同伴效应分析结果表明,低质量同伴比例越高,大学生退学的可能性越大,而高质量同伴比例有利于大学生坚持学业^[74]。

黄依梵等探讨了室友对大学生餐饮消费的影响。他们的数据来自我国某大学行政管理部门提供的两个年级共4 790名大一学生的相关信息(包括性别、户口类型、高考成绩等),以及这些学生的校

园卡餐饮消费记录。作者采用随机分配宿舍的自然实验与工具变量法对上述数据进行分析。研究发现,大学生校园卡餐饮消费存在显著的同伴效应,即室友的餐饮消费金额对大学生餐饮消费金额具有显著的正向影响,但这种同伴效应存在一定的群体差异,相较于城镇学生,农村学生受到室友同伴的影响较小^[75]。不同于前3项研究主要分析学校因素对学生学业表现相关变量的影响,该研究探讨了室友对大学生餐饮消费的影响,并且大学生餐饮消费数据来自他们的校园卡餐饮消费记录,这意味着大学生校园卡数据可以作为简历分析法的重要“泛简历”数据来源。目前,除这项研究之外,其他一些学者也采用校园卡数据开展了一些有价值的研究。例如,刘进等基于B大学N学院学生的校园卡刷卡记录和其他行政数据,分析了学生早起行为的区域差异,以及学生早起行为与其学业成绩之间的关系^[76]。

简历分析法除了可以分析教师与同伴对在校生的影响之外,还可以分析学校作为一个整体对在校生的影响。例如,王骏等研究了学校投入对学生学业成绩的影响及其群体差异。作者将重点高中作为学校投入的代理变量,研究使用的数据为全样本行政数据,包括J市所有高中2010级的全体学生,最终获得的有效样本数量为17373个。文章采用普通最小二乘法和两水平线性模型对数据进行分析。研究发现:(1)重点高中对学生的学业成绩具有显著的正向预测作用;(2)学生的学业成绩虽然受到教师质量和班级规模的显著影响,但重点高中对学业成绩的影响主要体现在学校投入上,重点高中能显著影响学生的理科学业成绩,而对学生的文科学业成绩的影响并不显著;(3)重点高中对学生学业成绩的影响存在群体差异,家庭社会经济地位较差、学习能力较强和学习基础较好的学生受到重点高中的影响更大一些^[77]。

总体来说,该部分所选取的文章在数据来源方面有一个共同特征,即行政数据尤其是教育部门个人层面的行政数据是文章主要的数据来源,而通常人们所认为的简历不被作为数据来源之一。

2. 学校教育对毕业生的影响

这一部分要介绍的研究主要采用简历分析法探讨学校教育对毕业生行为(包括体育活动、犯罪等)、收入与职业获得的影响。这些研究的样本来自芬兰、美国、英国与智利等国家。

Kari等采用OLS模型和工具变量方法(Mendelian randomization, MR)分析了受教育程度(以受教育年数衡量)与成年人体育活动(physical activity)之间的关系。文章以1651名年轻芬兰人心血管风险研究(Cardiovascular Risk in Young Finns Study, YFS)的参与者为分析对象。作者从以下渠道采集研究所需要的数据:(1)芬兰年轻人心血管风险研究的数据。该研究自1980年以来,已经进行了7次追踪研究。(2)芬兰统计局的芬兰纵向雇主-雇员数据(the Finnish Longitudinal Employer-Employee Data, FLEED)。FLEED是一个年度面板数据,覆盖芬兰的全部劳动年龄人口,记录了这些人口的劳动力市场结果(如收入和就业)和最高教育程度等信息。(3)芬兰统计局的人口普查纵向数据(the Longitudinal Population Census, LPC)。作者从这一数据中提取了研究对象的家庭背景(父母的教育)信息。研究发现,受教育年限与体育活动之间存在显著的正相关关系。平均来说,每多接受一年的教育,总体体育活动时间会增加0.62个单位,每周高强度活动时间会增加0.26小时,每天的步数会增加560步,每天的有氧步数会增加390步^[78]。

学校教育除了影响受教育者成年后的体育活动参与外,还会影响其他行为。例如,McEachin等采用线性概率模型(LPM)和双重稳健的逆概率加权(inverse probability weights)方法评估了美国北卡罗来纳州中学特许学校对九年级学生的行为,以及年轻人犯罪和参加选举的可能性的影响。研究使用的数据有4个主要来源:(1)由北卡罗来纳州公共教学部(North Carolina Department of Public Instruction, NCDPI)提供的纵向行政数据,这些数据包括在2004/2005学年至2015/2016学年就读北卡

罗来纳州公共学校的所有学生的基本信息,如性别、族裔、成绩、毕业、GPA和选课信息等,而且这些信息都是学生个人层面的。(2)北卡罗来纳州公共安全部门(North Carolina Department of Public Safety, NCDPS)公开的个人层面的犯罪记录信息。(3)北卡罗来纳州选举委员会(North Carolina Board of Elections, NCBOE)公开的选民登记和投票记录。(4)美国社区调查的公开数据,通过这一数据可以衡量特许公立学校和传统公立学校所在县的地方经济发展特征。研究发现,相较于留在传统公立学校(Traditional Public Schools, TPS)的同龄人,这些在九年级从传统公立学校转到特许公立学校的学生在所有结果变量上的表现都更好,而且长期缺勤和被停学的可能性更低,成年后犯罪的可能性更小,登记并参加选举的可能性更大^[79]。

此外,学校教育对毕业生的另一个非常重要的影响是个人收入。例如,陈沛等采用履历分析法和计量分析法探讨了企业高管的一流大学背景对个人收入的影响。文章的数据来源与处理思路是:(1)通过财经数据库BvD-Osiris确定了全球2015年营业收入排名前10000的企业,并将其高管作为研究对象;(2)从企业年报和官方网站采集高管的个人简历,包括首席执行官、总裁、副总裁、总经理、副总经理、首席财务官等,手工提取简历中的年龄、毕业院校、毕业时间等信息,对这些信息进行编码并建立数据库;(3)从BvD-Osiris数据库中获取企业绩效数据,并将这部分数据与上一步得到的数据进行匹配;(4)剔除教育背景和个人收入存在缺失的样本,最终得到可用于分析的38670名企业高管的数据。研究发现,拥有一流大学学位或知名大学学位的企业高管在收入方面显著高于拥有普通大学学位的企业高管,即企业高管教育收益中存在“一流大学效应”。这种效应对不同特征高管的影响不同,对于拥有学士和硕士学位的企业高管、发展中国家企业高管、收入相对位置较低的企业高管而言,“一流大学效应”更加明显^[80]。

学校教育对毕业生收入的影响不仅具有“一流大学效应”,还具有其他影响,如可以通过接受不同类型的高等教育方式来实现。例如,Espinoza等采用OLS回归和逆概率加权回归调整(Inverse Probability Weighting Regression Adjustment, IPWRA)的估计方法分析了接受高等技术教育(higher technical education)与接受大学学位(university degrees)教育的英国毕业生之间的个人收入差异。文章整合了多个来源的行政数据:(1)国家学生数据(National Pupil Data, NPD),包含学生18岁以前的成绩和教育记录。(2)个人学习记录(Individualised Learner Record, ILR),包含学生在18岁之后所接受的职业/技术教育的信息。(3)高等教育统计局(Higher Education Statistics Agency, HESA)的数据,涵盖学生的大学教育信息。(4)个人层面的收入信息,来自英国皇家税务和海关总署(Her Majesty's Revenue and Customs, HMRC)的数据。通过对上述数据的整合,作者最终获得了2002/2003学年毕业的622000名英国学生的有效样本。研究发现,高等技术教育带来的早期优势会随时间的推移而减弱。到30岁时,大多数拥有学位的人收入更高。然而,对于接受STEM领域高等技术教育的男性而言,他们的收入仍然明显高于其他领域拥有学位的人。对于女性而言,并不存在此类显著差异^[81]。

毕业生在劳动力市场的表现除了收入之外,还包括职业获得,即学校教育同样可以影响毕业生的职位获得。例如,Zimmerman采用回归断点设计(regression discontinuity design)和双重差分的方法分析了智利的精英商业学位项目(elite business-focused degree programs)对学生获得顶级职位和收入的影响。文章使用的数据包括:(1)申请数据。该数据包括3种类别:1980—2001年所有入学考试参加者的数据,1982—2001年25所传统大学学位项目的录取结果数据,以及1974—2001年精英法学、工程学和商学项目的录取结果数据。这些数据是通过对存储在智利国家图书馆已发布的申请和候补名单公告的纸质副本进行电子化得到的。(2)公司领导层数据。在智利,上市公司必须向证券和保险监管局(Superintendencia de Valores y Seguros, SVS)披露高层管理人员和董事会成员的身份,作者从SVS

网站获取了公司领导层数据。(3) 税务记录,由智利税务局(Chilean tax authority)提供。研究发现,精英商业学位项目促使这些学生担任领导职位的数量增加了44%,促使他们获得前0.1%收入的概率提高了51%。然而,这些收益仅限于曾就读于高学费私立高中的男性学生,而对于入学考试成绩相近的女性学生或其他类型学校的学生则不存在显著差异^[82]。

三、简历分析法在中国教育领域的应用前景

(一) 简历分析法在中国教育领域应用的现存问题

简历分析法的教育研究应用兼具优势与不足。其优势包括:(1)吸收了纵贯研究中(准)自然实验的优点;(2)采用非介入式方法使得研究结论具有科学性;(3)具有整合各类数据源及研究方法的潜力;(4)研究流程清晰,成本总体可控。简历分析法的不足主要体现在简历数据缺乏完整性、简历标准化程度低以及简历数据分析不准确3个方面^[6]。

近年来,尽管简历分析法在我国教育研究活动中的应用越来越普遍,但学术界对这一方法在中国教育场域下的适用性以及具体使用过程中存在的问题缺乏系统研究。一方面,与国际上其他国家相比,简历分析法更加适用于对中国教育问题的研究,其原因在于:中国的教育体系本质上是以前科层制为主导的符号论教育体系,可以通过人的各类符号判断其在教育活动中的表现。符号论存在于中国整个教育体系,从幼儿园到博士后,基于符号论的简历或“泛简历”信息极其丰富。因此,只要收集到研究对象的简历或“泛简历”信息,就可以采用简历分析法开展教育研究活动,从新的方法论视角分析和研究中国教育问题。另一方面,除了前文提到的简历分析法本身存在的局限外,简历分析法在我国教育领域的推广使用目前还面临以下问题。

第一,缺乏统一的命名。由于国内学术界尚未明确提出“简历分析法”的规范概念,所以研究者在使用过程中对这一方法的命名并不统一,如“简历法”“履历分析法”“简历研究法”“简历抓取法”“简历文本分析法”“简历数据分析法”“简历分析”等。命名方法的混乱,反映出一些研究者对于简历分析法的使用仍缺乏规范。对于上述简历分析法命名混乱的问题,我们应回到其英文原文进行理解。简历分析法的英文表达为“curriculum vitae analysis”,因此建议在中文语境下统一译为“简历分析法”。

第二,方法的专有属性和概念边界不清晰。一方面,一些学术论文尤其是学位论文在研究方法介绍部分,普遍存在评阅人有关“简历分析法是不是专门的研究方法”的质疑。另一方面,对于实际使用了简历分析法的研究成果,虽然部分论文在方法介绍部分明确将简历分析法列为研究方法之一,但仍有较多研究成果并未直接使用“简历分析法”这一概念,这种情况在以“泛简历”数据作为论文的主要数据来源中尤为突出。例如,一些基于我国院士、杰青、长江学者简历数据的研究,还有一些有关官员、上市公司高管等教育经历的研究等。

第三,缺乏规范的操作流程。在简历分析法应用已有研究成果中,有的详细介绍了使用简历分析法的流程与相关细节,如简历数据筛选和编码的标准,但更多的研究并未详细介绍其研究过程,而是直接呈现研究结果。

第四,简历数据质量较低。以中国大学教师研究为例,教师简历的公开率低,部分青年教师不愿意在工作单位官网公布个人简历,本课题组在简历信息抓取过程中甚至发现,部分学校的整个系所甚至整个二级学院都没有一条教师简历信息。另外,教师简历的内容过于简略。比较而言,西方国家大学教师的简历信息非常丰富,不仅包括个人详细履历,还包括各种类型的学术成果,如书籍章节、会议论文。与之相反,中国大学教师简历的信息量较小。因此,在利用中国学者的个人公开简历数据开展

诸如学术产出研究时,通常需要匹配学术发表大数据库。然而,匹配后的数据可能仍然无法涵盖教师的全部学术工作。原因在于,学术发表数据库中一般不包括教师参编的书籍章节,以及艺术类教师的艺术作品等。此外,我国部分教师还存在不愿意在简历中提及第一学历等信息的问题。

第五,个人层面的行政数据的利用程度低。行政数据指为行政(非研究)目的收集的信息,这种数据是由政府部门和其他组织在提供服务过程中,为了登记、交易和记录保存而收集的^[83]。它可以来自各种行政系统,如教育、医疗保健、税务、住房系统等^[84]。从数据层次看,行政数据可以分为个人和宏观两个层面的行政数据。后者由前者通过合成或加总得到,如县级、地市级或省级层面的数据^[85]。根据这一分类标准,个人层面的行政数据是一种非常重要的“泛简历”数据。这些微观个体层面的数据对开展一些纵向追踪研究或(准)因果推断研究具有非常重要的价值。然而,相较于国外,我国教育学科的实证类文献对行政数据的利用程度较低,这种情况在使用非教育部门的行政数据方面尤为突出。例如,郭艳斌通过对国外19种高水平期刊1990—2021年发表的论文分析发现,国际学术界最近十年利用行政数据发表的教育类的量化研究论文明显增加^[86],其中一些论文还将非教育类的个人层面的行政数据作为数据来源之一——典型的研究成果包括:García等将美国政府部门提供的犯罪数据与其他数据匹配,分析了一项早期儿童教育项目给个人带来的终身收益^[87];Chetty等利用美国联邦所得税申报表(US federal income tax returns)数据研究了教师对学生发展的长期影响^[88]。与之形成鲜明对比的是,我国虽然有少数学者利用个别学校、县或省份的个人层面的行政数据开展了教育研究^[8,89-90],但目前公开发表的教育类的文献尚无一篇文章将中国政府部门掌握的个人层面的犯罪记录、纳税记录作为其中一个研究数据来源。

(二) 简历分析法在中国教育领域的应用展望

1. 运用简历分析法对我国教育领域的重点、难点问题开展研究

无论是在数据时间跨度、数据采集方式上,还是在与其他数据源和研究方法结合使用方面,简历分析法均具有独特的优势。因此,运用简历分析法对教育领域的重点、难点问题开展研究恰好能够充分发挥其优势。以我国2021年出台的“双减”(指有效减轻义务教育阶段学生过重作业负担和校外培训负担)政策为例,涉及的教育主体及其受到的影响非常复杂,如学生本人的学习投入、时间分配,学生家长的校内外教育投入选择,政府的教育经费投入与分配,学校教师的工作量、工作-家庭冲突,校外培训机构从业人员的职业流动等。此外,上述主体的规模庞大。2021年,我国义务教育在校生1.58亿人,专任教师1 057.19万人^[91],全国有7 743.1万名学生参加课后服务,534.5万名教师参与课后服务工作^[92]。根据2020年底的相关数据,我国教育培训行业从业人员已经超过1 000万人^[93]。对于涉及主体多元、相关主体规模庞大的“双减”政策,简历分析法有多个优势:(1)简历分析法能实现对所有研究对象的全样本估计。原因在于,教育行政部门与学校管理部门采集的师生层面的数据、税务部门的纳税记录等都是简历数据的重要组成部分,这些数据通常能够覆盖全部样本。(2)简历分析法可以分析“双减”政策对相关主体的长期影响。这主要得益于采用简历分析法可以实现多种数据源的准确匹配,如在分析“双减”政策对学生本人的长期影响时,可以将学生在义务教育阶段所接受的具体政策干预措施(如学生是否参加课后服务,提供课后服务的主体是不是本校老师,课后服务的内容、时长等)与其成年后的收入信息、婚姻登记信息、医疗保健信息、犯罪记录等进行匹配,分析不同政策干预措施的各种长期效果,从而为“双减”政策的进一步完善提供学理依据。简历分析法的上述优势是访谈研究、问卷调查等传统教育研究方法所不具备的。需要注意的是,能否切实发挥简历分析法在研究诸如“双减”政策评估等教育重点、难点问题上的优势,在很大程度上还取决于简历数据质量、“泛简历”数据的可获得性,以及简历分析法的规范使用等。这也是接下来要讨论的问题。

2. 推动标准化简历数据库建设

简历数据标准化程度低在很大程度上限制了简历分析法在教育研究领域的推广与应用。当前一些欧洲国家已经意识到简历数据库对于科研活动的重要价值,开始在国家层面进行标准化简历数据库建设。例如,葡萄牙科学技术基金会(The Foundation for Science and Technology)从2001年就开始建设电子简历数据库,即DeGóis Curriculum,该数据库通过结构化的简历表(structured CV form)在线收集葡萄牙研究人员的信息,以确保数据的高度标准化与准确性^[94]。此外,西班牙^[19]、巴西(Lattes平台)^[95]等国也在积极建设大规模标准化研究人员简历数据库。上述标准化简历数据库或类似系统/平台的建设为简历数据分析、研究评估和科学政策设计提供了新的可能^[1]。一方面,这些标准化简历在数据可比性和完整性方面具有明显的优势,因为简历数据编码没有任何成本^[94]。另一方面,电子简历信息系统的发展使得简历数据库和文献计量数据库之间的电子连接更加容易^[1],这为科研绩效相关研究提供了便利。然而,随着电子简历信息系统的发展,新的方法论问题也随之产生,如处理大量数据负载或定义下载优先级等^[1]。实践中,中国国家自然科学基金委近年来也开始收集申请人的标准化简历信息,并取得了一定成效。未来,为充分发挥简历数据库在中国教育科研事业发展中的重要作用,我国有关部门不仅要继续组织开展更高质量的标准化简历数据库建设,还要针对简历数据库建设过程中存在的问题加强研究攻关。

3. 提高个人层面的行政数据在简历分析法中的利用程度

个人层面的行政数据不仅是目前国际学术界教育类定量研究成果中比较常见的“泛简历”数据来源,而且还具有促进高质量和有影响力的社会科学研究发展的潜力^[84]。相较于国外,我国教育科学研究对微观个体层面行政数据的利用程度较低,其中一个可能的原因是国内外个人层面行政数据的公开程度不同。在北欧一些国家,如挪威、芬兰和瑞典,这些国家的统计局(National Statistical Institutes, NSIs)很多年前就已经开始使用CDROMs、DVDs或其他格式将匿名微观数据分发给大量的研究机构^[96]。此外,美国部分州在行政数据服务教育量化研究方面也做得较好^[86]。然而,在我国,研究人员难以获得个人层面的行政数据。以教育类的行政数据为例,我国教育行政主管部门仅公开宏观层面的行政数据,一般不公开微观个体层面的行政数据,这使得研究人员难以采用简历分析法对我国教育领域的一些重点、难点问题开展研究。对此,我国各级教育行政主管部门首先应该意识到行政数据所蕴藏的巨大科研价值,以及基于证据的科研成果对政策制定的积极意义;其次,应借鉴北欧国家以及美国的经验^[97],逐步建立完善的行政数据服务教育科学研究的体制机制,通过提高个人层面的行政数据在简历分析法中的利用程度,充分发挥个人层面的行政数据在推动我国教育科研事业高质量发展方面的巨大潜力和优势。

4. 推动简历分析法在我国基础教育研究领域的应用

简历分析法在我国基础教育领域应用的研究成果具有3个显著特征:(1)研究成果数量少。从粗略检索到的文献结果看,目前只有9篇期刊论文采用简历分析法对我国基础教育领域的相关议题开展了研究。(2)研究对象以高中生为主。在这9篇成果中,有6篇论文以高中生为研究对象^[77,90,98-101],以初中生和小学生为研究对象的成果分别有2篇^[8,102]和1篇^[103]。(3)研究主题较为单一。从因变量看,这些成果的因变量均与学业表现相关,分别是初中学业成绩、中考成绩、高中阶段教育选择(普通高中、职业高中或不上高中)、高中学业成绩、高中选科、是否参加高考、高考成绩和高考录取结果(“落榜”);从自变量看,这9篇成果的自变量只包括个体(如性别、民族、城乡户口)、家庭(如家庭贫困)、同伴、学校整体4个层面(如是不是重点高中)的变量,它们均没有涉及学校环境中的教师因素(有些文章将其作为控制变量)对中学生的影响。在国外,采用简历分析法研究基础教育的

成果涉及的研究话题更加多元,这些成果不仅包括国内这9篇成果所涉及的研究话题^[104],还包括教师在基础教育阶段的因素对学生非学业表现和长期发展的影响^[79,105]。例如,McEachin等分析了美国中学特许学校对学生成年后犯罪和参加选举的可能性的影响^[79]。造成上述差异的一个重要原因,可能还是个人层面的行政数据在国内外公开程度不同。基础教育领域的研究对象是在校生,他们一般没有人们通常认为的个人简历,而且他们的个人成长信息也较少公开。因此,采用简历分析法开展基础教育领域的研究应用,其主要的数据来源是个人层面的行政数据。然而,由于我国个人层面行政数据的公开程度远低于美国和欧洲的一些国家,这在一定程度上使得简历分析法在我国基础教育领域的研究成果无论是在数量上还是话题的多元性方面均逊色于西方国家。需要注意的是,我国已经建成世界上最大规模的教育体系^[106],基础教育领域实际上已经积累了海量个人层面的行政数据(“泛简历”数据)。因此,为充分发挥教育科研服务对于教育政策制定与教育实践的价值,让教育学不再被认为是“书斋学问”,需要畅通科研人员合理获取、规范使用个人层面行政数据的渠道,促进简历分析法在我国基础教育领域的推广与应用。

5. 从概念界定、操作流程与论文内容呈现3个方面规范简历分析法的使用

对于什么数据属于简历数据,目前学术界还没有统一的判断标准。结合本课题组使用简历分析法的经验以及对该领域文献的系统梳理,我们认为简历数据包括两类:(1)通常人们认为的简历,如研究对象在求职、申请项目经费时所使用的简历,这种简历一般包括研究对象教育背景、工作经历、技能、获奖等各方面的信息。(2)“泛简历”,指包含研究对象的姓名和其他信息、与简历类似的各类数据信息,如各级各类教育机构的学生信息(人口统计学信息、学业表现等)。基于上述对简历数据的概念界定,可以从以下两个方面判断一篇文章是否使用了简历分析法:(1)简历数据是不是文章的主要数据来源;(2)文章是否对简历数据进行实证分析(量化分析或质性分析)。如果一篇文章同时满足这两个条件,则作者应该在文章的方法介绍部分明确指出研究使用了简历分析法(建议在中文语境下统一使用“简历分析法”这一名称)。

对于使用了简历分析法的文章,还应该遵循规范的操作流程。一般来说,简历分析法的操作流程主要包括九大步骤:(1)研究问题的提出;(2)简历数据源寻找与论证;(3)简历数据库建设;(4)简历数据筛选;(5)简历数据提取(编码);(6)简历数据匹配;(7)简历数据运算;(8)研究质量检验;(9)形成研究结论。其中,步骤(6)涉及简历数据库与其他数据库的匹配问题,并非必要步骤。文章下一个部分将对简历分析法的九大操作步骤进行详细介绍,在此不再赘述。此外,对于大多数期刊论文来说,限于篇幅,作者可能无法在论文中呈现所有步骤,但从学术写作规范尤其是一篇量化类实证研究论文的结构框架看,作者应该在文章中呈现研究问题、数据来源渠道、数据使用是否符合法律与伦理要求(仅限于使用了非公开简历数据的文章)、数据采集时间、简历数据筛选与编码标准、变量的操作化定义、研究方法(包括简历分析法以及对处理后的数据进行量化分析的方法)、研究结论等基本要素及内容。

6. 加强简历分析法与传统研究方法及大数据、人工智能等新研究方法的融合应用

简历分析法与传统研究方法以及大数据、人工智能等新研究方法具有天然联系,有望成为教育学科面向大数据和人工智能时代进行方法变革的关键突破口。简历分析法与这些传统及新研究方法的融合应用体现在以下几个方面:(1)全样本信息或大样本信息。大数据包含4个特点:海量性、多样性、高塑性和异变性。传统观点认为,大数据通常掌握在少数大数据供应商或数据平台手中,但对于教育研究而言,大数据则更为丰富多样,需要根据教育研究实际进行大数据的概念界定和数据采集,以形成各类专题的简历大数据库。(2)ChatGPT等大语言模型的出现可以大幅提高简历数据编码的

准确率与有效信息提取率。近年来,人工智能自然语义分析方法被逐渐应用于简历数据编码,这种方法在编码效率等方面明显优于人工编码,但也存在编码准确率低、非结构化文本信息无法有效提取等问题。然而,随着开源的 ChatGPT 模型的发展及其研究应用,研究人员借此模型就可以训练一个专门用于阅读、分析、理解、翻译不同风格、不同语言的简历 ChatGPT 模型。这从根本上有利于解决传统自然语义分析方法在简历数据编码方面存在的技术瓶颈,提高简历数据编码的准确率与有效信息提取率。(3)便于进行多数据源挖掘和匹配。简历分析法除了可以对简历大数据的抓取和分析外,还可以实现对简历大数据与其他大数据资源、问卷调查数据、实验数据的挖掘、匹配与关联使用,这将极大拓展简历分析法在教育领域的应用范畴。(4)便于进行大数据计算或人工智能模型训练。传统统计学算法广泛应用于教育研究领域,其基本原理在于构建统计模型、降低抽样误差、形成相关性研究结论^[106-107]。简历分析法的引入及其在教育研究领域的应用,有望借助大数据算法替代传统统计学算法进行全样本计算而不是样本估算,进行精确计算而不是模糊推算,进行(准)因果关系计算而不是相关关系演算,甚至实现超级计算^[108]。

(三)简历分析法教育研究应用的操作流程

本课题组结合多年来使用简历分析法开展相关研究的经验,以及对相关文献的梳理,整理出一套规范的简历分析法操作流程。下文将对这些步骤逐一进行介绍,以期促进简历分析法在教育研究领域的规范、科学应用与全面推广(如图 2-4)。为表述方便,如无特殊说明,该部分中的“简历”均不包括“泛简历”的范畴。

1. 研究问题提出

运用简历分析法开展教育研究,最初始的步骤就是论证具体研究问题是否适用简历分析法,即研究问题需要首先对应用简历分析法的適切性进行分析。

从选题范围来看,简历分析法的应用主要包括两类研究问题:一类是简历分析法的专属研究领域,包括人的职业生涯轨迹研究和人的流动研究。具体到教育学领域,学业流动、职业流动、学科专业选择、淘汰与分流、劳动力市场、高等教育国际化与全球化,以及与人的学业发展、职业发展相关联的学业表现、学术产出、工作绩效、职业阶梯、职业倦怠等议题,均可以考虑应用简历分析法开展研究。另一类是使用传统教育研究方法容易陷入困境的研究议题。例如,在一些数据收集困难、无法进行纵贯研究的教育领域,可以尝试引入简历分析法。此外,简历分析法的适用范围既不限于人的职业生涯、流动等领域,也不限于高等教育研究领域。相反,简历分析法由于具有较好的纵贯研究属性,还可以对一些传统教育研究的重难点问题做出新的解释。例如,2020年山东省240多位高考冒名顶替事件实际上就形成了天然的实验组和对照组,研究者可以基于受害者、冒名顶替者两组人群的简历信息,使用简历分析法对“新读书无用论”以及高等教育投资回报率等传统研究议题展开纵贯式研究。

运用简历分析法进行教育研究大致有3种提出研究问题的方法:一是基于理论的问题提出;二是基于文献的问题提出;三是基于现实问题(实践)的问题提出。这与其他教育研究方法的问题提出过程基本一致。

2. 简历数据源寻找与论证

在确定使用简历分析法之后,最重要的工作就是数据源的寻找与论证。一般而言,简历数据源包含两大类:(1)直接的简历数据源,主要包括存储于互联网空间的公开简历信息;存储于各具体部门或单位的非公开简历信息;标准化简历数据库(如西班牙等国建立了标准化的学术人员简历库,中国国家自然科学基金项目申报系统中也要求提交完整的个人简历信息,类似这样的数据库可直接用于简历研究);各类市场化数据库(包括各类大数据公司、互联网公司、人力资源公司等通过各类渠道

收集整理各类简历数据库;其他简历来源(包括其他线上线下可以提供研究对象直接简历素材信息的有效渠道)。(2)间接的简历数据源,也可称之为“泛简历”,通常是指包含研究对象姓名和其他信息、近似简历的各类数据信息,如学术人才学术表现类数据库;各级各类教育机构学生信息;其他简历来源(包括其他线上线下可以提供研究对象间接简历素材信息的有效渠道)。

在具体研究活动中,在确定简历信息来源后,要做好以下工作:(1)对简历素材的真实性、有效性、时效性、可行性等进行论证,以判断相关简历素材是否满足研究需要。(2)简历素材使用的伦理判断,即论证相关简历素材在产权、法律、道德方面是否可以用于学术研究。(3)简历素材的成本论证,即相关简历素材的获得与使用成本是否可控。(4)简历素材的预分析,即选取小样本简历素材进行预研究,以判断简历素材的质量、编码可行性等,从而为最终简历库建设提供决策依据。

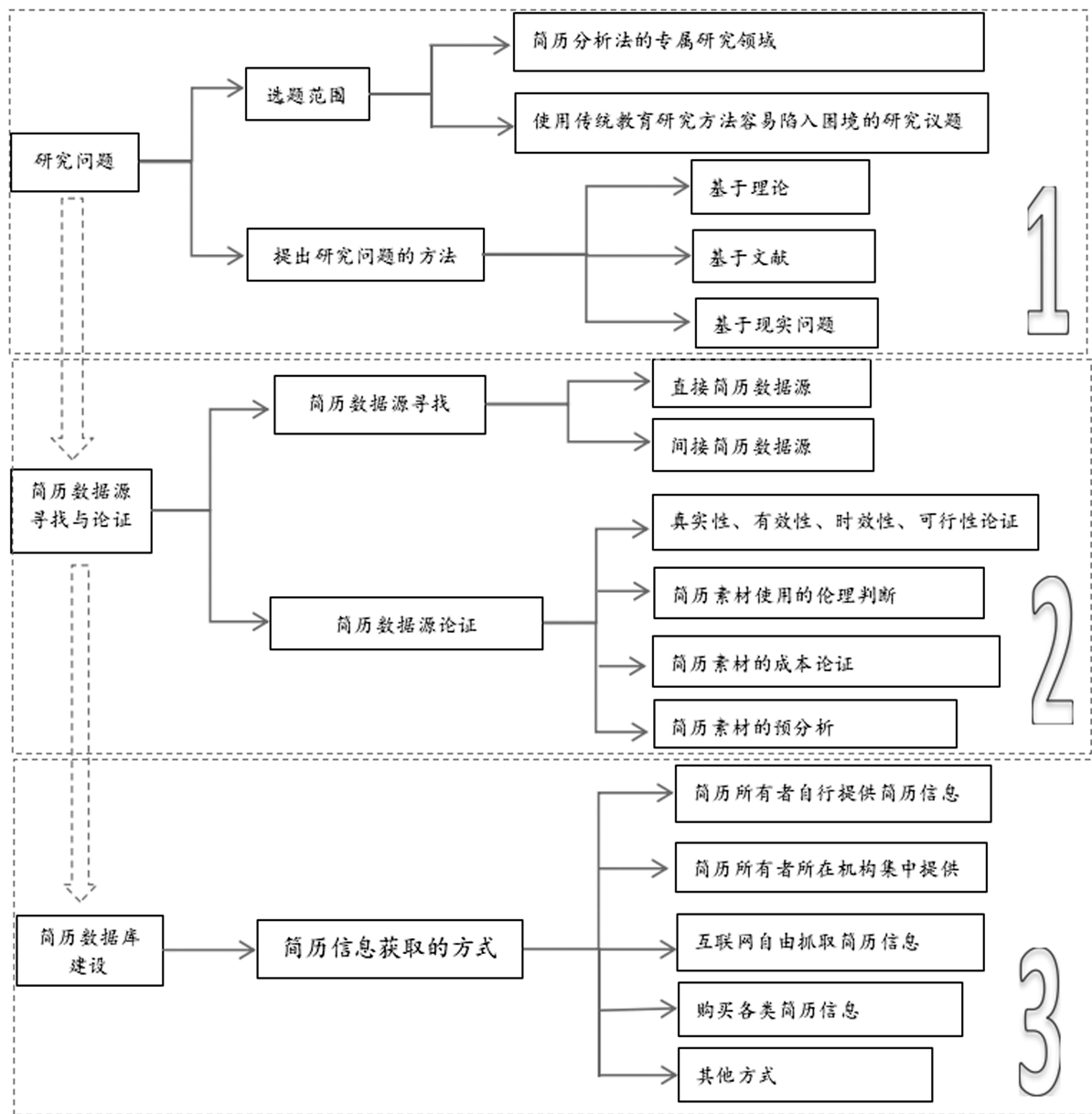


图 2 简历分析法的操作流程(步骤 1-3)

3. 简历数据库建设

在确定简历素材来源可靠、内容真实有效、能够满足研究需求、成本可控、符合法律和伦理要求等

基础上,才能进入简历数据库建设阶段。在此阶段,核心工作是简历信息的获取——具体获取方法需要根据数据来源不同而定,其主要的获取方式包括:(1)简历所有者自行提供简历信息,早期研究活动中大多直接发送邮件向研究对象索要个人简历,或借助招聘活动等方式收集个人简历信息。当然,在此过程中应注意标明简历类型,比如科学家通常拥有多个版本的简历,有的更适合申请科研基金项目资助,有的更适合就业求职,还有一些更适合专业咨询或服务^[43]。所以,应根据研究需要来收集特定类型的简历。从已有研究来看,直接索要简历的效率偏低。一方面存在邮件地址错误的问题,一些数据库中大约20%的电子邮件地址是过时或错误的^[43];另一方面邮件回复率偏低,在一项大规模邮件发放中,课题组向3799人发送了两次简历请求邮件,但总体回答率只有32%^[59]。(2)简历所有者所在机构集中提供简历信息。该简历数据由人事管理部门或简历管理部门批量化提供,如一项有关卢森堡大学教授的研究,大学教授的名单以及他们的职位(助理教授、副教授或正教授)、研究单位均由卢森堡大学人力资源部提供^[62]。(3)互联网自由抓取简历信息,即通过手工或自动化手段在互联网抓取研究对象的简历信息。在互联网搜索简历过程中,通常会显示大量无关的网页,如求职列表、学术通信、招聘启事或简历写作指南等,这些被统称为研究“噪音”——克服方法主要是引入更先进的搜索引擎。已有研究表明,互联网简历搜索需要在10个不同的搜索引擎上进行^[59]。(4)购买各类简历信息。大数据时代可以付费向互联网平台、大数据公司等购买满足研究目标需要的简历数据。(5)其他方式。

4. 简历数据筛选

在完成简历数据库建设之后,需要根据研究进行数据的筛选工作,主要筛选标准包括:(1)错误简历的剔除。要剔除误进入数据库不满足研究需要的简历数据。(2)简历的长短判断。要根据研究需要剔除篇幅不满足研究需要的简历,比如一些研究需要提供长简历,就需要根据长简历的具体标准进行简历筛选。(3)简历关键数据缺失判断。例如,研究问题为博士国际化流动规律,就需要剔除简历信息中未包含博士阶段信息的简历数据。(4)简历时效性判断。要根据研究需要剔除过于陈旧的简历数据。(5)其他判断标准。例如,应结合研究需要,根据被试年龄、性别、民族、职称、机构类型等进行数据筛选。除了剔除不合格简历外,还应对简历数据库中缺失关键要素的简历数据进行有效补充。例如,有研究根据需要进行了简历数据筛选:(1)从样本中剔除年龄在20岁以下和70岁以上的所有记录;(2)剔除那些没有明确的学习领域或者没有学位的记录;(3)剔除1950年之前或2009年之后获得博士学位的研究人员;(4)剔除没有明确目的地或停留时间的访问研究数据;(5)剔除在本国访问但简历里却标明“国际访问”的数据;(6)剔除所有被认为是不可靠的简历数据^[109]。

5. 简历数据提取(编码)

简历数据提取或者编码是整个简历分析法最为重要也是难度最大的环节。为清晰地呈现简历编码过程(见表5),本文将之分为3种编码方式:一是初级编码,也可称为简单编码,即研究任务相对简单明确,数据编码人员只需从简历信息中提取少数变量信息。例如,研究中国大学教师的学业流动与职业流动频率问题,就只需要编码人员进行21个变量的提取工作,如此这些变量编码内容就一目了然,不存在明显编码困难。因此,本文将编码数量低于50、编码内容简单、编码活动不因编码人员个人能力和熟练程度而产生重大差异的简历编码定义为“初级编码”。

二是高级编码,也可称为“复杂编码”,它与初级编码存在以下4个方面的不同。(1)编码数量更多。编码数量通常超过50个变量而又低于500个变量,如对于研究对象学术发表成果的编码。(2)编码内容更为复杂。编码内容可能涉及人口统计学指标、学业基础指标、职业基础指标、学术产出类指标等多种类型。(3)存在模糊领域编码。遇到此类情形,需要先进行协商。例如,有关学术人员的

学术成果认定,在编码实际执行过程中就可能存在模糊领域,如学术报告、书籍章节、学术奖项等是否纳入学术成果。对此,编码人员在编码前需要反复协商并确定统一标准。(4) 编码错误率更高。尤其是熟练编码人员与非熟练编码人员在编码效率和准确率方面可能存在较大差异,因此需要对编码人员进行规范系统的培训,以提升编码工作的效率和准确率。

三是大数据编码,也可称为“开放式编码”。近年来随着大数据和人工智能技术的突破,自然语义分析等技术开始应用到学术研究领域。一些简历分析研究已经开始尝试借助相关软件进行大数据编码,或者直接撰写代码以满足个性化的研究需要。大数据编码与初级编码、高级编码存在本质不同。前者主要依靠计算能力进行编码,后者主要依靠人力进行编码;前者在编码数量、编码效率等方面大大高于后者,但在准确率等方面可能弱于后者。

表 5 某简历分析法研究的部分编码标准举例^[110]

通用标准	<ul style="list-style-type: none"> ●使用日历年。例如,如果委员会领导在简历中填写的是 1992—1993 年,那么填写这两年 ●所有不清楚或似乎不完整的条目编码为 X
出版物的标准	<ul style="list-style-type: none"> ●同行评议的文章不包括书籍章节、信件、摘要和书评
演讲的标准	<ul style="list-style-type: none"> ●通过本地电话会议进行的演讲可以编码为同行评议的区域/全国性演讲 ●本地研究论坛不算作同行评议的区域/全国性演讲
领导职位的标准	<ul style="list-style-type: none"> ●领导职位“联合主任”编码为“主管职员”——如果其职责范围是区域性或全国性的,则为 E1;如果是在医学院,则为 F1 ●“协调员”的职位编码为“其他人员”(E2 或 F2)

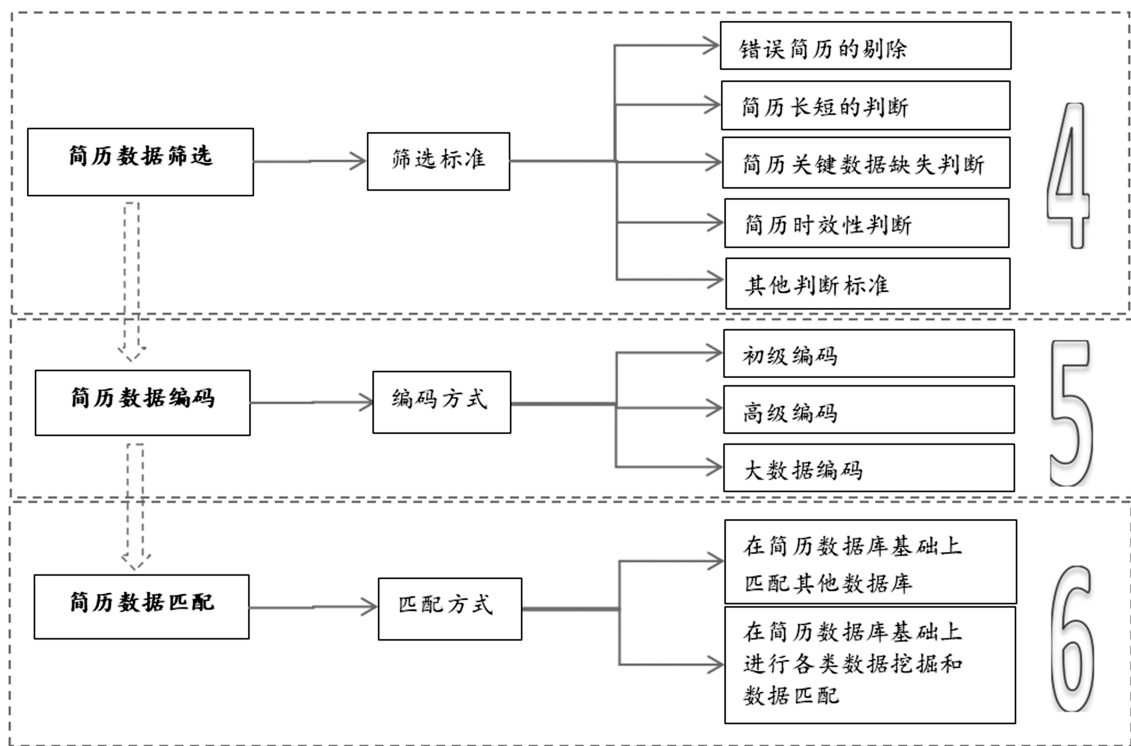


图 3 简历分析法的操作流程(步骤 4-6)

无论选取哪种编码方式,编码都只是简历数据提取的第一步,其最关键的步骤还在于进行编码准确率的分析。目前已有的简历分析法研究应用很少涉及编码准确率的量化分析,但 Dietz 等对此

有很大贡献。Dietz 等认为,一方面,简历分析应设定编码的可靠性“阈值水平”,任何信度低于 0.850 的编码都应被认为是有所问题的,任何低于 0.600 的编码都是完全不可接受的^[43]。该阈值水平测试的主要方法就是对编码人员进行背靠背抽样和打分,对不同编码人员对同一简历数据的编码结果进行一致性评估。另一方面,Dietz 等认为应引入 Crittenden 和 Hill 的编码器胡新都测量(Rs)作为测试编码器可靠性的工具,他们的研究从两组 10 个简历中对 37 个变量的子集进行编码抽样,发现编码器胡新都测量(Rs)结果是最有用和最相关的^[43]。

6. 简历数据匹配(挖掘)

简历分析法往往会缺失部分信息。解决这个问题的方法是利用其他数据源(如文献计量学和专利数据)来补充简历信息。例如, Lee 等通过将简历数据与调查数据进行互补^[111]; Woolley 等还对补充调查信息和附件简历信息的潜力进行了分析^[112]。

简历数据匹配通常有两种方式:一种是在简历数据库基础上匹配其他数据库,最典型的就是在学术人员简历数据库基础上,根据姓名、机构等进行学术发表、专利申请等数据库匹配,以此克服学者自我报告的简历数据存在的各种弊端及其带来的编码困难;另一种是在简历数据库基础上进行各类数据挖掘和数据匹配,如本课题组曾抓取 5 年内所有获得高校自主招生录取的学生名单,其中包括“中学名称”一栏,本课题组根据这一数据匹配了多项指标以满足研究需要。

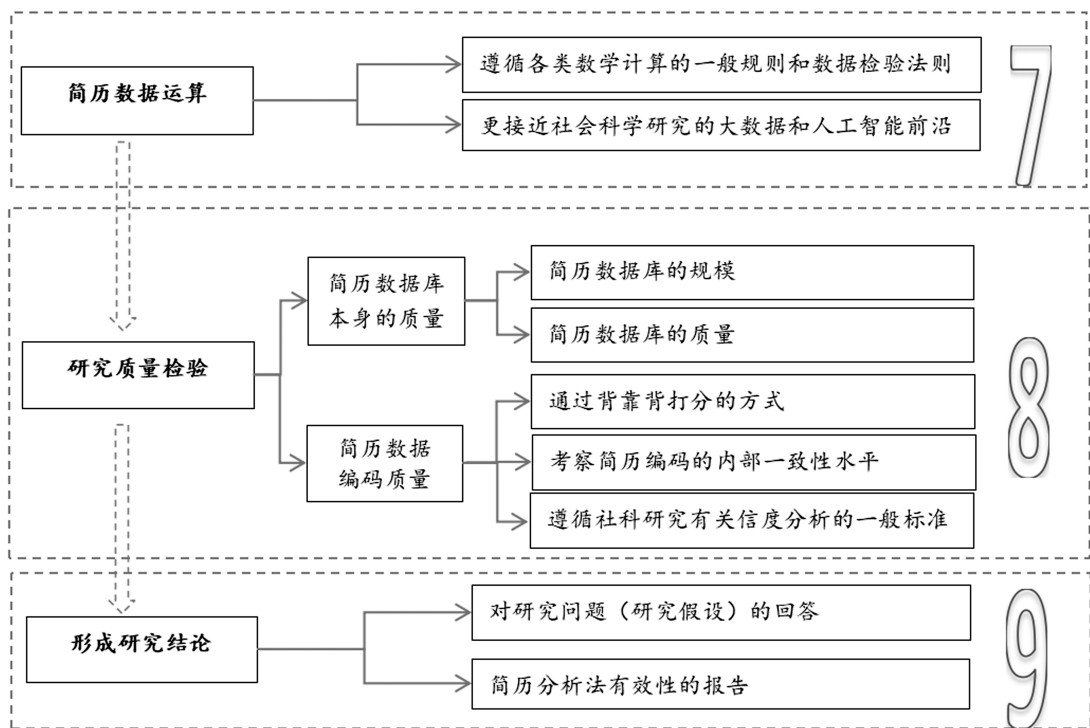


图 4 简历分析法的操作流程(步骤 7-9)

7. 简历数据运算

简历数据的计算需要遵循各类数学计算的一般规则和数据检验法则。唯一不同的是,简历分析法更接近社会科学研究的大数据和人工智能技术前沿,可能面临传统算力不足的问题,需要引入大数据和人工智能计算理念和计算能力。我们认为,以超级计算机为代表的计算能力的需求变化,是教育学科科学化水平提升的关键指标之一,中国教育研究一旦广泛使用超级计算,则代表着中国教育学科集体迈入科学化研究阶段,从而克服传统小样本研究在理念和技术方面的各种缺陷。

8. 研究质量检验

简历分析法的质量检验通常可以从以下两个方面展开:一是简历数据库本身的质量状况分析,包括简历数据库的规模与质量两个方面。其中,简历数据库质量主要包括简历信息真实性、完整性、时效性等。二是简历数据编码质量状况分析。编码质量主要通过背靠背打分的方式,考察简历编码的内部一致性水平,遵循社会科学研究有关信度分析的一般标准。

9. 形成研究结论

研究结论通常包含对研究问题(研究假设)的回答和简历分析法有效性报告两个部分。简历分析法在使用过程中通常涉及与其他研究方法的兼容问题。已有的简历分析法研究很少只采用简历分析这一单一方法。根据已有研究,与简历分析法配合使用的其他研究方法至少包括:(1)问卷调查法;(2)访谈法;(3)文献计量法;(4)地理信息系统方法;(5)文本分析法;(6)各类数据统计分析方法。需要注意的是,由于简历分析法的重要步骤之一是对整理好的数据进行量化或质性分析,因此,(4)-(6)这3类研究方法的使用实际上是简历分析法的一部分。简历分析法有效性报告目前主要使用的是传统数据有效性检验方法,未来需要开发专门的简历分析法有效性检验方法。

四、结 语

总体而言,在现阶段,虽然简历分析法在我国教育领域的推广使用仍然面临一些问题,但这些问题并不能掩盖简历分析法本身具备的研究优势。相较于问卷法、访谈法等传统教育研究方法,简历分析法通过非介入式数据采集方式客观收集研究对象的纵向与横向信息,实现对各类简历数据与非简历数据的追踪、挖掘、匹配和分析,得到各类相关性、(准)因果性、(准)实验性乃至开拓性的研究结论。从应用场景部分的文献所涉及的研究话题,以及大部分文献的数据来源渠道的多样性、研究方法的多元化也可以发现,简历分析法具有较强的延展性,它可以通过“多数据源+多研究方法”的研究范式对各级各类教育现象开展研究。

进入大数据和以 ChatGPT 为代表的通用人工智能时代,相较于其他学科,教育学科的信息化软硬件基础更好,大数据生成存储使用简历的可能性更大,教育大数据研究结果直接应用于教育教学改革前景更为广阔。因此,随着大数据与人工智能等新研究方法的不断发展与成熟,以及与简历分析法的混合使用越来越紧密,我们需要加强标准化简历数据库建设,提高个人层面行政数据的利用程度,规范简历分析法的使用等,促进简历分析法在教育研究领域的应用与推广。同时,只有更加充分地发挥简历分析法自身的独特优势,未来才可能有更多的学者在观念上认可简历分析法——既是一种教育实证研究新方法,又可作为教育学科的专门研究方法,进而在教育研究实践中不断拓展简历分析法的应用场景,这将大幅提高我国教育研究的科学性,提高教育学科在社会科学领域中的学科地位,提升教育研究成果服务于国家重大战略决策与教育教学实践的能力。

参考文献:

- [1] CAÑIBANO C, BOZEMAN B. Curriculum vitae method in science policy and research evaluation: the state-of-the-art [J]. Research evaluation. 2009,18(2):8694.
- [2] 叶澜. 思维在断裂处穿行:教育理论与教育实践关系的再寻找[J]. 中国教育学刊,2001(4):38.
- [3] 孙元涛. 论“参与式”教育研究及其价值承诺[J]. 高等教育研究,2007(5):6465.
- [4] 王洪才. 论教育研究的方法论特征[J]. 厦门大学学报(哲学社会科学版),2007(1):114122.
- [5] 隋玄华,纪晓平. 图书情报学特殊研究方法刍议[J]. 情报资料工作,2010(1):3941.
- [6] 刘进. 简历分析法及其在高等教育研究中的应用[J]. 高等教育研究,2022,43(6):4956.

- [7] KNAPP R H, GOODRICH H B. The origins of American scientists[J]. *Science*, 1951, 113(2941): 543545.
- [8] 李芳华,肖寒. 贫困家庭高中阶段教育选择:基于行为经济学的视角[J]. *经济研究*, 2022, 57(12): 180200.
- [9] 何瑛,于文蕾,戴逸驰,等. 高管职业经历与企业创新[J]. *管理世界*, 2019, 35(11): 174192.
- [10] 祁凡骥,卢湘枚. 地级市市长任用规律:从“十五大”到“十九大”——基于 GDP 百强市市长相关数据的研究[J]. *国家行政学院学报*, 2018(5): 6975, 189.
- [11] GERRARD J M, FISH I, TATE R, et al. Evaluation of the careers of graduates of the University of Manitoba's BSc (Medicine) program[J]. *CMAJ: Canadian medical association journal*, 1988, 139(11): 10631068.
- [12] 周建中,肖小溪. 科技人才政策研究中应用 CV 方法的综述与启示[J]. *科学与科学技术管理*, 2011, 32(2): 151-156, 179.
- [13] 谭光鼎,王丽云. 教育社会学:人物与思想[M]. 上海:华东师范大学出版社, 2009: 37.
- [14] 袁振国. 实证研究是教育学走向科学的必要途径[J]. *华东师范大学学报(教育科学版)*, 2017, 35(3): 417, 168.
- [15] 王洪才. 教育研究的基本方法论[J]. *北京师范大学学报(社会科学版)*, 2006(6): 2427.
- [16] 姚计海. “文献法”是研究方法吗:兼谈研究整合法[J]. *国家教育行政学院学报*, 2017(7): 8994.
- [17] 刘泽云,刘佳璇. 中国教育收益率的元分析[J]. *北京师范大学学报(社会科学版)*, 2020(5): 1325.
- [18] 邓峰. 教育收益率估算中的计量偏误及调整方法的综述[J]. *教育与经济*, 2013(5): 4248.
- [19] CAÑIBANO C, OTAMENDI J, ANDÚJAR I. Measuring and assessing researcher mobility from CV analysis: the case of the Ramón y Cajal programme in Spain[J]. *Research evaluation*, 2008, 17(1): 1731.
- [20] 林松月,刘进,徐丽. 大学教师究竟因何流动:一种非接触式研究方法的引入[J]. *教师教育研究*, 2020, 32(1): 78-84, 102.
- [21] 卫垌圻,鲁晶晶,谭宗颖. 应用科研履历数据的情报研究进展[J]. *图书情报工作*, 2015, 59(7): 140146.
- [22] ROVELLI P, CURNIS C. The perks of narcissism: behaving like a star speeds up career advancement to the CEO position[J]. *The leadership quarterly*, 2021, 32(3): 101489.
- [23] 王海山. 科学方法辞典[M]. 杭州:浙江教育出版社, 1992: 50.
- [24] 中国百科大辞典编委会. 中国百科大辞典[M]. 北京:华夏出版社, 1990: 63.
- [25] 金胜勇,吴杏冉. 关于图书馆学专门研究方法之分析[J]. *图书馆理论与实践*, 2006(1): 1517.
- [26] 邱均平,张洋,赵蓉英. 网络信息计量学方法论[J]. *中国图书馆学报*, 2008(2): 2932, 41.
- [27] 刘蔚华. 方法大辞典[M]. 济南:山东人民出版社, 1991: 582.
- [28] 高鹏,杨兆山. “教育现象”何以是教育学的研究对象[J]. *教育研究*, 2014, 35(2): 5560.
- [29] 王江,杨全印. 关于教育研究目的、结果与思维方式的思考[J]. *江苏高教*, 2001(1): 6062.
- [30] GOMEZ C J, HERMAN A C, PARIGI P. Moving more, but closer: mapping the growing regionalization of global scientific mobility using ORCID[J]. *Journal of informetrics*, 2020, 14(3): 101044.
- [31] SHI W, DU D, YANG W. The flow network of chinese scientists and its driving mechanisms based on the spatial development path of CAS and CAE academicians[J]. *Sustainability*, 2019, 11(21): 5938.
- [32] 李和章,戚也,林松月,等. 我国博士研究生就业地域特征分析:基于 97134 条大学教师简历的实证分析[J]. *学位与研究生教育*, 2022(12): 3946.
- [33] 朱鹏程,张宇,曹卫东,等. 长三角企业经营管理人才空间分布及其地理流动网络:基于上市公司董监高团队数据分析[J]. *人文地理*, 2020, 35(4): 124129.
- [34] AZOULAY P, GANGULI I, ZIVIN J G. The mobility of elite life scientists: professional and personal determinants[J]. *Research policy*, 2017, 46(3): 573590.
- [35] CAÑIBANO C, FOX M F, OTAMENDI F J. Gender and patterns of temporary mobility among researchers[J]. *Science and public policy*, 2016, 43(3): 320331.
- [36] PENG C, LI Z, WU C. Researcher geographic mobility and publication productivity: an investigation into individual and institutional characteristics and the roles of academicians[J]. *Scientometrics*, 2023, 128(1): 379406.
- [37] 杨金玉. 空气污染与高管流动研究:基于秦岭—淮河线的断点回归设计[J]. *南开经济研究*, 2021(2): 143163.
- [38] LI F, TANG L. When international mobility meets local connections: evidence from China[J]. *Science and public policy*, 2019, 46(4): 518529.

- [39] KOTSEMIK M, DYACHENKO E, NEFEDOVA A. Mobile young researchers and their non-mobile 'twins': who is winning the academic race? [J]. *Scientometrics*, 2022, 127(12): 73077332.
- [40] ZHAO Z, BU Y, LI J. Does the mobility of scientists disrupt their collaboration stability? [J]. *Journal of information Science*, 2020, 48(2): 016555152094874.
- [41] YANG X, CAI X, LI T. Peer effects of the young returnee scientists: evidence from the state key laboratories in China [J]. *Science and public policy*, 2022, 49(5): 739750.
- [42] JIA J, LI Z, HU Y, et al. Does top management team's job mobility experience matter for corporate innovation? [J]. *Pacific accounting review*, 2022, 34(3): 426450.
- [43] DIETZ J, CHOMPALOV I, BOZEMAN B, et al. Using the curriculum vita to study the career paths of scientists and engineers: an exploratory assessment [J]. *Scientometrics*, 2000, 49(3): 419442.
- [44] 宋晓欣, 马陆亭, 赵世奎. 教育学科高层次人才成长规律探究: 以 22 位长江学者为例 [J]. *中国高教研究*, 2018(3): 5455, 87.
- [45] KOROM P. How do academic elites march through departments? a comparison of the most eminent economists and sociologists' career trajectories [J]. *Minerva*, 2020, 58(3): 343365.
- [46] 高端, 王彬. 中国杰出青年科技人才的成长过程及特征: 基于“科学探索奖”获得者的履历分析 [J]. *科学管理研究*, 2022, 40(2): 139146.
- [47] 屠羽, 吴维库, 彭本红. 如何成长为高层次科技人才: 基于“双一流”建设高校校长的履历分析 [J]. *科学管理研究*, 2020, 38(2): 117125.
- [48] KOCH M, FORGUES B, MONTIES V. The way to the top: career patterns of Fortune 100 CEOs [J]. *Human resource management*, 2017, 56(2): 267285.
- [49] 柯春晓. 众媒时代中国新闻领军人才特质分析 [J]. *中国广播电视学刊*, 2021(3): 5558, 123.
- [50] BELKHOUSA M, YOON H D. How does openness influence the impact of a scholar's research? an analysis of business scholars' citations over their careers [J]. *Research policy*, 2018, 47(10): 20372047.
- [51] SCHMID S, ALTFELD F, DAUTH T. Americanization as a driver of CEO pay in Europe: the moderating role of CEO power [J]. *Journal of world business*, 2018, 53(4): 433451.
- [52] LUTTER M, SCHRÖDER M. Who becomes a tenured professor, and why? panel data evidence from German sociology, 1980-2013 [J]. *Research policy*, 2016, 45(5): 9991013.
- [53] 林上洪, 刘海峰. 寒门子弟何以考取科举功名: 基于清代浙江科举人物朱卷履历的定量分析 [J]. *北京大学教育评论*, 2021, 19(3): 6280, 190.
- [54] 孙玉涛, 陈灵芝. 诺奖候选人学术影响、提名人身份与获奖概率 [J]. *科学学研究*, 2019, 37(9): 15501557.
- [55] RYAZANOVA O, MCNAMARA P. Choices and consequences: impact of mobility on research-career capital and promotion in business schools [J]. *Academy of management learning & education*, 2019, 18(2): 186212.
- [56] LAWSON C, SHIBAYAMA S. International research visits and careers: an analysis of bioscience academics in Japan [J]. *Science and public policy*, 2015, 42(5): 690710.
- [57] 瞿群臻, 高思玉, 汪鹏飞, 等. 基于生存分析视角的战略科学家成长路径研究 [J]. *中国科技论坛*, 2022(7): 157166.
- [58] SCHMID S, MITTERREITER S. Top managers' career variety and time to the top [J]. *European management review*, 2021, 18(4): 476499.
- [59] GAUGHAN M, BOZEMAN B. Using curriculum vitae to compare some impacts of NSF research grants with research center funding [J]. *Research evaluation*, 2002, 11(1): 1726.
- [60] 乔锦忠, 孙娜, 沈敬轩, 等. 科学基金对学术产出的长期影响研究 [J]. *中国科技论坛*, 2022(12): 8898.
- [61] WANG Y, JONES B F, WANG D. Early-career setback and future career impact [J]. *Nature communications*, 2019, 10(1): 4331.
- [62] HUSSINGER K, CARVALHO J N. The long-term effect of research grants on the scientific output of university professors [J]. *Industry and innovation*, 2022, 29(4): 463487.
- [63] GANGULI I. Saving Soviet science: the impact of grants when government R & D funding disappears [J]. *American economic journal: applied economics*, 2017, 9(2): 165201.

- [64] SHI D, LIU W, WANG Y. Has China's Young Thousand Talents program been successful in recruiting and nurturing top-caliber scientists? [J]. *Science*, 2023, 379(6627): 6265.
- [65] 董新宇, 鞠逸飞, 段雨欣. 地方政府高层次人才政策实施效果研究[J]. *中国科技论坛*, 2022(9): 128138.
- [66] 马双, 汪烽. 人才政策对人才跨区域流动的影响: 以长三角城市群为例[J]. *中国人口科学*, 2023(1): 10+113.
- [67] 梁镇涛, 毛进, 李纲. 海外交流对学者职业生涯的影响: 以国家建设高水平大学公派联合培养博士研究生为例[J]. *图书情报工作*, 2022, 66(24): 7683.
- [68] BIN A, SALLES-FILHO S, SPATTI A C, et al. How much does a Ph. D. scholarship program impact an emerging economy research performance? [J]. *Scientometrics*, 2022, 127(12): 69356960.
- [69] FIROOZI D. The impact of post-admission merit scholarships on enrollment decisions and degree attainment: evidence from randomization[J]. *Economics of education review*, 2022(86): 102221.
- [70] CARRUTHERS C K, FOX W F. Aid for all: college coaching, financial aid and post-secondary persistence in Tennessee [J]. *Economics of education review*, 2016(51): 97112.
- [71] WADDINGTON R J, BERENDS M. Impact of the indiana choice scholarship program: achievement effects for students in upper elementary and middle school[J]. *Journal of policy analysis and management*, 2018, 37(4): 783808.
- [72] CANALES A, MALDONADO L. Do teachers compensate for or compound disparities in academic achievement? [J]. *International journal of educational research*, 2023(117): 102138.
- [73] GERSHENSON S, HART C M D, HYMAN J, et al. The long-run impacts of same-race teachers[J]. *American economic journal: economic policy*, 2022, 14(4): 300342.
- [74] HUMLUM M K, THORSAGER M. The importance of peer quality for completion of higher education[J]. *economics of education review*, 2021(83): 102120.
- [75] 黄依梵, 马莉萍. 大学生餐饮消费的室友同伴效应: 基于校园卡餐饮消费的实证研究[J]. *青年研究*, 2022(3): 7484, 96.
- [76] 刘进, 林松月, 王艺蒙, 等. 西部生源大学生学业表现的多维对比分析: 基于 B 大学 N 学院学生的实证研究[J]. *重庆高教研究*, 2019, 7(4): 1229.
- [77] 王骏, 彭顺绪, 原莹. 重点高中、学校投入与学生学业成绩: 基于 J 市普通高中的—个经验研究[J]. *世界经济文汇*, 2017(3): 1745.
- [78] KARI J T, VIINIKAINEN J, BÖCKERMAN P, et al. Education leads to a more physically active lifestyle: evidence based on Mendelian randomization[J]. *Scandinavian journal of medicine & science in sports*, 2020, 30(7): 11941204.
- [79] MCEACHIN A, LAUEN D L, FULLER S C, et al. Social returns to private choice? effects of charter schools on behavioral outcomes, arrests, and civic participation[J]. *Economics of education review*, 2020(76): 101983.
- [80] 陈沛, 杨希. 杰出毕业生教育收益中的“一流大学效应”: 基于全球万家企业高管数据的量化研究[J]. *中国高教研究*, 2017(11): 1823.
- [81] ESPINOZA H, SPECKESSER S. A comparison of earnings related to higher technical and academic education[J]. *Education economics*, 2022, 30(6): 644659.
- [82] ZIMMERMAN S D. Elite colleges and upward mobility to top jobs and top incomes[J]. *American economic review*, 2019, 109(1): 147.
- [83] The Administrative Data Liaison Service. Administrative data introduction [EB/OL]. (2016-07-25) [2023-06-01]. <https://web.archive.org/web/20160725114045/http://www.adls.ac.uk/adls-resources/guidance/introduction/>.
- [84] CONNELLY R, PLAYFORD C J, GAYLE V, et al. The role of administrative data in the big data revolution in social science research[J]. *Social science research*, 2016(59): 12.
- [85] 马红梅, 肖雨桐, 张文军. 伴随式教育档案管理数据及其评价功能[J]. *华中师范大学学报(人文社会科学版)*, 2020, 59(2): 163168.
- [86] 郭艳斌. 行政数据在国外教育量化研究中的应用: 基于 1990—2021 年 19 种高水平期刊发表的论文分析[J]. *高等教育研究*, 2022, 43(4): 98109.
- [87] GARCÍA J L, HECKMAN J J, LEAF D E, et al. Quantifying the life-cycle benefits of an influential early-childhood program[J]. *Journal of political economy*, 2020, 128(7): 25022541.

- [88] CHETTY R, FRIEDMAN J N, ROCKOFF J E. Measuring the impacts of teachers II: teacher value-added and student outcomes in adulthood[J]. *American economic review*, 2014, 104(9): 2633-2679.
- [89] 张新亮, 石艳, 郑琦, 等. 公费师范生学业表现的实证研究: 基于某部属师范大学的行政数据[J]. *教育学报*, 2023, 19(1): 165-181.
- [90] 张燕, 丁延庆, 李虹辉, 等. 高考复读有用吗: 基于断点回归的因果推断[J]. *全球教育展望*, 2019, 48(7): 114-128.
- [91] 教育部. 2021年全国教育事业发展统计公报[EB/OL]. (2022-09-14)[2023-08-15]. http://www.moe.gov.cn/jyb_sjzl/sjzl_fztjgb/202209/t20220914_660850.html.
- [92] 周世祥. “双减”组合拳如何直击减负难点[N]. *光明日报*, 2021-10-08(08).
- [93] 龙煦霏. “双减”之下教培与资本的变奏曲[J]. *中国商界*, 2021(9): 34-35.
- [94] PIRRALHA A, FONTES M, ASSIS J. Assessing scientific mobility dynamics and impact: drawing on the potential of electronic CV databases[EB/OL]. (2009-02-01)[2023-08-23]. https://www.researchgate.net/publication/265159696_Assessing_Scientific_Mobility_Dynamics_and_Impact_drawing_on_the_potential_of_electronic_CV_databases.
- [95] MENA-CHALCO J P, JUNIOR R M C. scriptLattes: an open-source knowledge extraction system from the lattes platform[J]. *Journal of the brazilian computer society*, 2009, 15(4): 34-39.
- [96] United Nations Economic Commission for Europe. Register-based statistics in the nordic countries: review of best practices with focus on population and social statistics[M]. Geneva: United Nations Publication, 2007: 34.
- [97] 曹浩文. 美国教育行政数据集的建设现状、经验与启示[J]. *外国中小学教育*, 2017(5): 364-1.
- [98] DEE T, LAN X. The achievement and course-taking effects of magnet schools: regression-discontinuity evidence from urban China[J]. *Economics of education review*, 2015(47): 128-142.
- [99] 李代, 王一真. 高考录取中的不确定性与教育获得差异: 以X省为例[J]. *社会学研究*, 2020, 35(1): 104-125, 244.
- [100] 孙志军, 彭顺绪, 王骏, 等. 谁在学业竞赛中领先: 学业成绩的性别差异研究[J]. *北京师范大学学报(社会科学版)*, 2016(3): 385-1.
- [101] 陈晋玲, 丁延庆, 叶晓阳. 普通高中学生择校的因果效应[J]. *教育学术月刊*, 2014(5): 364-5.
- [102] DING W, LEHRER S F. Do peers affect student achievement in China's secondary schools? [J]. *The review of economics and statistics*, 2007, 89(2): 300-312.
- [103] LAI F. Are boys left behind? the evolution of the gender achievement gap in Beijing's middle schools[J]. *Economics of education review*, 2010, 29(3): 383-399.
- [104] MURPHY R, WEINHARDT F. Top of the class: the importance of ordinal rank[J]. *The review of economic studies*, 2020, 87(6): 2777-2826.
- [105] DOBBIE W, FRYER JR R G. The medium-term impacts of high-achieving charter schools[J]. *Journal of political economy*, 2015, 123(5): 985-1037.
- [106] 王思遥. 教育实证研究的理论依据、争议与去向[J]. *大学教育科学*, 2020(5): 12-17.
- [107] 王晓燕. 教育实证研究的立场问题与反思[J]. *大学教育科学*, 2020(5): 18-22.
- [108] 张海生. 人工智能与教育深度融合发展: 逻辑、困境与策略[J]. *当代教育论坛*, 2021(2): 576-5.
- [109] CAÑIBANO C, OTAMENDI F J, SOLÍS F. International temporary mobility of researchers: a cross-discipline study [J]. *Scientometrics*, 2011, 89(2): 653-675.
- [110] MORZINSKI J A, SCHUBOT D B. Evaluating faculty development outcomes by using curriculum vitae analysis[J]. *Family medicine*, 2000, 32(3): 185-189.
- [111] LEE S, BOZEMAN B. The impact of research collaboration on scientific productivity[J]. *Social studies of science*, 2005, 35(5): 673-702.
- [112] WOOLLEY R, TURPIN T. CV analysis as a complementary methodological approach: toward a deeper investigation of the mobility of scientists in the Asia-Pacific[J]. *Research evaluation*, 2009, 18(2): 143-151.

(责任编辑: 张海生 校对: 杨慷慨)

Curriculum Vitae Analysis: A New Approach to Empirical

Research in Education

LIU Jin^{1,2}, YU Yitian¹, YANG Li³, LIN Songyue⁴, LI Zhifeng⁵, GAO Yuan², CHEN Kaizhe¹

(1. *School of Humanities and Social Sciences, Beijing Institute of Technology, Beijing 100081, China;*

2. *Higher Education Research Center, Southern University of Science and Technology, Shenzhen 518055, China;*

3. *Institute of Education, Tsinghua University, Beijing 100084, China;*

4. *Faculty of Education, The Chinese University of Hong Kong, Hong Kong 999077, China;*

5. *School of Law and Humanities and Social Sciences, Wuhan University of Technology, Wuhan 430064, China)*

Abstract: Although it belongs to an independent first level discipline, education lacks specialized research methods distinct from other disciplines that could provide substantial support for its development. Moreover, the introduction of a large number of research methods from other disciplines further limits the development of its disciplinary methodology, knowledge system, and theoretical system. In recent years, the curriculum vitae analysis has emerged globally, attempting to comprehensively assess the education and growth patterns of talents, as well as the implementation effectiveness of projects/policies, through the analysis of resume data and the integration of big data and other methods, which not only provides empirical new methods for educational science research, but also provides new opportunities for the construction of educational disciplines. Entering the era of artificial intelligence represented by ChatGPT, the new research paradigm of generative artificial intelligence models+computing power(supercomputing)+application scenarios is gradually maturing. The curriculum vitae analysis has also ushered in new breakthroughs in educational research applications, and it is expected to become a specialized research method in the field of education. In the field of education in China, there are curriculum vitae and other big data resources of various levels and types of educational entities. With the technological breakthrough of curriculum vitae analysis, it is expected to comprehensively deconstruct the laws of local education, construct a new paradigm of educational discipline research with Chinese characteristics and in line with China's national conditions, and even accumulate fragmented specialized knowledge to form a new local discipline system. In the past decade, there have been many research achievements based on curriculum vitae analysis in the academic community, but there has been insufficient research on the method itself, especially in terms of its disciplinary affiliation, whether it is a specialized research method in the field of education, how to integrate with traditional research methods, as well as new research methods such as big data and artificial intelligence. At the same time, there are significant differences in the concept and research process of curriculum vitae analysis in existing studies in China, indicating that the application of curriculum vitae analysis in Chinese education disciplines is still immature. To this end, the study focuses on the methodology of curriculum vitae analysis method, attempting to comprehensively deconstruct this research method through a large number of research cases, and standardizing its connotation and extension, as well as specific operational processes. It also clarifies the problems that exist in the use of curriculum vitae analysis method in China, and finally explores the application prospects of the method in China, in order to promote the standardized use and comprehensive promotion of curriculum vitae analysis method in education disciplines in China.

Key words: curriculum vitae analysis; educational research methods; empirical research; big data; artificial intelligence