

· 专题一: 双清论坛“行为科学与经济政策设计” ·

多维度溢出效应政策评估方法、思路与中国优势

——基于空间网络建模视角*

郭俊聪 郑维伟 瞿 茜**

上海交通大学 安泰经济与管理学院, 上海 200030

[摘 要] 在梳理固定效应模型、工具变量法、双重差分法、断点回归设计、合成控制法等传统政策效应评估方法基础上, 本文介绍了多维度溢出效应下的政策评估方法, 并强调了空间网络计量模型应对多维溢出效应挑战的独特优势。特别地, 本文重点探讨了空间网络计量模型在处理单一政策存在个体或区域间溢出、政策效应和溢出效应均存在动态变化、组合政策存在多维度效果溢出、政策效应存在于高维空间网络数据等四类情形时的政策评估方法、识别与估计问题。中国交通、互联网等硬件基础设施建设的不断完善, 以及全国统一大市场等软环境的日益改善, 成为提供中国学者深入探讨政策效应评估方法的原创沃土和独特优势。在政策效果溢出效应复杂化的当下, 本文的探讨可为政策评估方法未来的发展提供重要参考。

[关键词] 因果推断; 政策评估; 多维溢出效应; 空间网络计量模型; 中国优势

近年来, 党和国家对于“国家治理体系和治理能力现代化”的重视愈发凸显。党的十八届三中全会提出, 要把“完善和发展中国特色社会主义制度, 推进国家治理体系和治理能力现代化”确定为全面深化改革的总目标。党的十九届四中全会在党的历史上首次将“推进国家治理体系和治理能力现代化”作为大会的鲜明主题。鉴于国家治理体系和治理能力现代化建设依赖制度完善和政策执行的科学性、合理性, 机制研究需提供可靠的政策评估方法, 以满足政策制定者系统、直观地了解执行是否有力、政策是否有效, 进而有针对性地制定并实施相关改革措施的现实要求。

政策效果评估的科学性和重要性受到党和国家的日益重视。2020年10月29日, 党的十九届五中全会通过《中共中央关于制定国民经济和社会发展的第十四个五年规划和2035年远景目标的建议》, 建议明确指出我国要“健全重大政策事前评估和事后评价制度, 畅通参与政策制定的渠道, 提高决策科学化、民主化、法治化水平”。2022年10月26日, 党



瞿茜 上海交通大学安泰经济与管理学院教授、经济系主任, 上海市青年拔尖人才。研究方向包括计量经济理论、空间计量模型、区域与城市经济学、社交网络与同辈效应等。主持国家自然科学基金优秀青年科学基金项目, 在国内外顶级经济学期刊 *Journal of Political Economy*、*Journal of Econometrics*、*Journal of Development Economics* 等发表论文 20 余篇。



郭俊聪 上海交通大学安泰经济与管理学院助理研究员。研究领域包括应用空间计量经济学、区域经济学和劳动经济学; 研究方向包括使用空间计量经济学方法研究溢出效应和同伴效应。在 *Journal of Development Economics*、*Regional Science & Urban Economics*、*Economics Letters* 等期刊发表论文多篇。

的二十大报告中强调了“坚持精准扶贫、尽锐出战, 打赢了人类历史上规模最大的脱贫攻坚战”、“坚持精准治污、科学治污、依法治污, 持续深入打好蓝天、碧水、净土保卫战”。这反映出党和国家在提高政策实施精准性方面提出了更高的要求; 同时, 也从侧面

收稿日期: 2023-07-24; 修回日期: 2023-09-07

* 本文根据第 338 期“双清论坛”讨论的内容整理。

** 通信作者, Email: xiqu@sjtu.edu.cn

本文受到国家自然科学基金项目(72222007, 71973097)的资助。

体现出了精确、合理、客观评估公共政策评估方法的重要性。换句话说,精准评估政策效果是国家治理体系建设的重要内容,也是推进我国高质量发展的有力保障。

然而,伴随着我国交通基础设施的完善和互联网的快速发展,线下、线上沟通交流的便利程度得到大幅提高,这就决定了政策的实施效果并非仅仅受本地区、本行业因素影响,公共政策的溢出效应愈发凸显^[1]。互联网和交通基础设施同步发展的客观事实进一步扩大了政策在个体、行业、区域之间的影响范围。因此,一个预期定位精准的政策也可能产生广泛的溢出效应,进而对经济社会带来深远的影响,而对于此类政策的效果评估,因其具有复杂空间网络溢出效应,对传统政策评估方法提出了挑战。

为此,本文在对传统政策评估方法进行归纳总结的基础上,基于空间网络数据的特点,针对政策效应存在动态变化、多维度空间溢出等多种复杂情况,拓展现有的空间网络计量模型,提出并构建一种科学的政策评估方法,以期精准测度政策的实施效果。

1 政策评估方法的发展现状与趋势

1.1 政策评估方法的计量建模

从计量模型看,现有研究方法正逐步从“观察数据”到“拟实验数据”和“实验数据”的计量建模思路转变,总体形成从数据和结构中提炼因果关系,以达到政策评估的精细化和有效性的发展趋势。其中,固定效应模型(Fixed Effect, FE)、工具变量(Instrumental Variable, IV)、断点回归设计(Regression Discontinuity Design, RDD)、双重差分法(Difference in Differences, DID)、事件研究(Event Study)、合成控制法(Synthetic Control Method, SCM)、随机对照组实验(Randomized Controlled Trials, RCT)等为学术界主流的政策评估方法^[2]。在具体的政策评估中,研究者往往需要结合数据或场景自身的特点,选用合适的估计方法。

(1) 固定效应模型。作为政策评估中识别因果关系最基础、发展最早的方法之一,其主要思想在于通过控制丰富的个体和时间固定效应,从而实现消

除或缓解共同因素对政策效应识别的干扰。从适用范围看,该方法主要应用于面板数据结构,研究者通常采取在模型中加入双重固定效应(Two-way Fixed Effects)^①的做法,以控制个体层面不随时间变化的因素,以及宏观层面只随时间变化但不随个体变化的因素。例如,探讨政府财政补贴对企业创新成果数量和质量的影响时,不同企业自身固有特点不同,不同时期企业共同面临的国际国内大环境也不同,而两者同样是影响企业创新的重要因素。因此,一方面有必要引入企业固定效应,将各企业本身固有的创新禀赋分离出来,另一方面又要控制不同时段的时间固定效应,从而得出更加合理、准确的政策效果。在此基础上,近年来学术界也提出了更一般化的方法。例如,将个体效应和时间效应相乘构建的交互固定效应模型(Interactive Fixed Effect)^[3]。控制固定效应目前已成为政策评估研究中的规范方法,并与其他方法相结合得到广泛应用。图1展示了1980年以来美国国家经济研究局(National Bureau of Economic Research, U.S., NBER)和2004年以来经济学五大期刊^②发表的论文采用各类方法的比例。根据图1A可知,固定效应模型始终是经济学研究中的主流方法,比例最高,且长期处于稳步增长趋势,2015年后占比甚至超过了50%。

(2) 工具变量法。作为政策评估中另一种识别因果关系常见的方法,尤其适用于当政策实施存在内生性的情况。例如,在制定和落实精准扶贫政策过程中,由于被扶贫地区的人均收入水平本来就较低,即收入水平往往决定了一个地区是否作为扶贫对象。此时,若只是简单对比扶贫对象和非扶贫对象的收入水平差距,极有可能会得出扶贫政策反而降低了居民收入水平的错误结论。因此,为克服政策实施的内生性问题,研究者需要为政策寻找合适的工具变量^[4]。同样以扶贫政策为例,考虑到贫困往往与当地的自然条件相关,满足相关性假定;而自然条件一般较为外生且与地区人均收入水平并不存在直接联系,满足外生性和排他性假定,因此,与自然条件相关的变量可以作为政策变量的工具变量。图1B展示了工具变量方法在经济学研究中的重要地位,尽管运用该方法的论文占比增速并不快,但近十几年来其比例一直处于高位。

① 即个体固定效应和时间固定效应。

② 包括 *American Economic Review*、*Econometrica*、*Journal of Political Economy*、*Quarterly Journal of Economics*、*Review of Economic Studies*。

(3) 断点回归设计。其基本思想在于：考虑存在一个连续变量，该变量可以决定个体在某一临界点两侧接受政策干预的概率。若政策的实施依赖于某个指标，且存在客观的断点，如个人所得税基本起征点、企业获得补贴的最低研发费用投入额等，那么断点回归设计无疑是评估此类政策的合理方法。该方法最早由 Thistlethwaite 和 Campbell^[5] 在 1960 年提出，但在提出的早期关注度并不高。1999 年 Angrist 和 Lavy^[6] 开始将其运用在经济学领域，研究班级规模对成绩的影响。直到 Hahn 等^[7] 于 2001 年对该方法进行了规范的计量经济学理论证明，学术界才开始大量涌现基于该方法的应用研究。如图 1C 所示，NBER 中运用断点回归设计方法的论文在 2000 年之后得到了快速增长，所占比例在 2018 年已达 10%；发表在经济学五大期刊上的相关论文占比也保持了类似的增长趋势。

(4) 双重差分法。又称为“倍差法”，由 2021 年诺贝尔经济学奖获得者 Card 及其合作者在研究最低工资政策时引入到主流经济学研究^[8]。该方法主要通过对比实验组和控制组在政策前后变化的差异来评估政策效果，因其涉及两次差分^①而得名。若政策具有一定随机性，且只在部分地区实施，那么双重差分法不失为一种有效的评估方法。近三十年来，相关的研究层出不穷，论文占比快速上涨。最具代表性的中文文献之一来自于 2005 年周黎安和陈烨^[9] 运用双重差分模型对中国农村税费改革的政策效果评估。如图 1D 所示，2018 年后该方法在经济学五大期刊上得到了超过 15% 的论文采用。不仅如此，近年来该方法在计量理论和应用层面继续有所创新和突破，例如交错双重差分法 (Staggered DID)^[10-12]、三重差分法 (Triple Difference)^[13, 14] 等。

(5) 合成控制法。在现实过程中，研究者还会不可避免的面临难以满足双重差分法所要求的处理组与控制组严格共同趋势假定，导致样本选择偏误问题。例如，实施了创新型城市试点政策的城市，极有可能也在研发费用达到一定额度后享受税收减免、地方政府人才引进政策力度也可能加大，这种“政策叠加效应”使得传统双重差分模型等方法难以处理潜在的内生性问题。进一步看，即便采用倾向得分匹配 (Propensity Score Matching, PSM) 与双重差分模型相结合的思路，为处理组寻找到了尽可能

接近的控制组，也无法确保得到创新型城市试点政策的“净”政策效应。反观基于“反事实估计框架”提出的合成控制法则是一个可行的估计思路^[15]。合

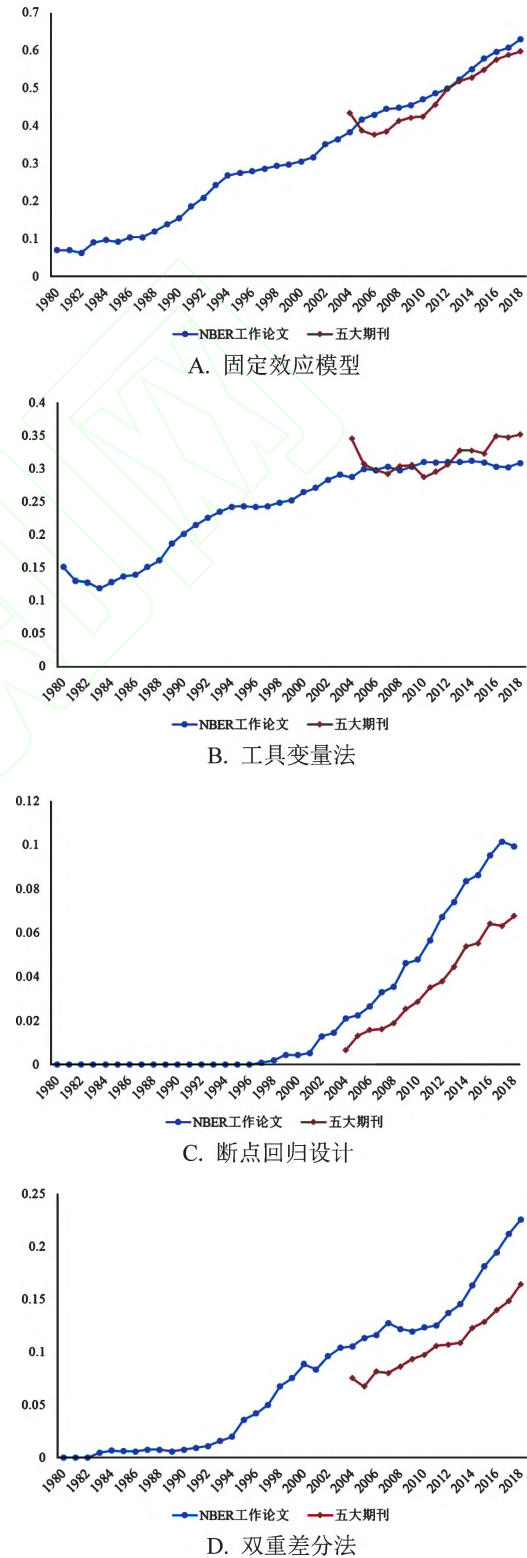


图 1 四种政策评估方法相关论文发表所占比例^[2]

① 即政策前后的差分、实验组和控制组的差分。

成控制法的基本思想在于:虽然通常不易找到与受政策干预地区完全相似的理想控制地区,但通过对潜在控制组样本库(Donor Pool)进行适当的线性加权组合,以构造出与受政策干预地区在政策实施前的相关预测变量完全类似的“合成控制组”,并将其作为受政策干预地区的反事实替身(Counterfactuals),进而可比较“真实处理组”与“合成控制组”在政策实施前后相关经济变量的变化情况,最终反映出政策净效应大小。合成控制法通过数据驱动确定“合成控制组”的最优权重,相较于DID等传统回归方法,避免了因控制组的主观选择而可能导致的政策内生性问题。

1.2 政策评估方法的应用场景

图2展示了1980—2018年美国NBER发表的有关政策评估方法应用的论文分类数量统计结果。可见,政策评估方法的应用场景非常广泛,涵盖经济、教育、医疗、环境等各种领域,最流行的应用场景是劳工研究和公共经济这类与公共政策密切相关的领域。不仅如此,政策评估方法的运用往往紧跟政策,并逐步从单一政策的评估演变为对政策组合和政策网络的综合评估,进而帮助政府和社会更好的了解政策的综合影响效果。综上,在众多应用场景中,有必要结合前文中各研究方法的适用范围有针对性地处理不同的政策识别问题。

表1展示了1992—2016年中文核心期刊和中文社会科学引文索引(Chinese Social Sciences Citation Index, CSSCI)期刊发表的公共政策研究领域文献的高频关键词统计结果^[16]。对该统计信息中相关主题分类汇总后,本文得到了该阶段中国学者所关注的政策评估方法主要应用场景(图3)。其中,以公共管理、公民参与和交通基础设施建设为代

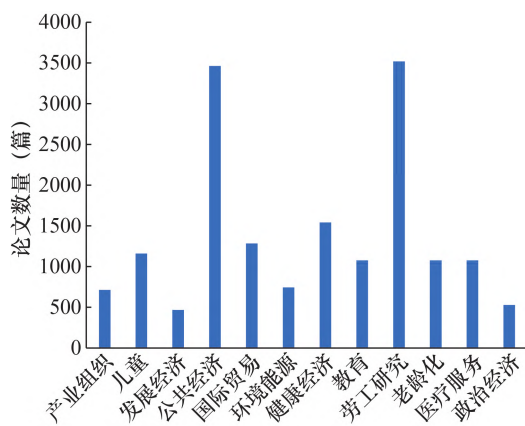


图2 1980—2018年NBER政策评估相关论文分类统计^[2]

表的公共政策执行效果评估等备受学者关注。此外,货币政策、财政政策、税收政策、金融机构、优惠政策以及产业政策等领域,同样是我国学者进行政策评估应用最流行的场景。并且,不少学者还积极从中国经济社会的发展要求出发,对我国的经济增长、经济发展以及创新水平等方面展开了深入的理论探讨和实证检验。不仅如此,以义务教育、教育公平、高等教育为主要内容的教育政策,以及农民工话题等同样占据了较大的比例,从侧面体现出中国学者所具有的浓厚人文关怀。同样地,为精准识别政策实施效果,有必要具体分析各公共政策潜在的内生性问题,寻找有效的工具变量,并立足于政策样本和数据的特点,使用合适的估计方法进行分析。

1.3 政策评估方法的最新进展

随着我国治理能力现代化、政策维度和实施精准性的不断提高,对政策评估方法的科学化、严谨化、精准化以及系统化等均提出了更高的要求,故而传统经典方法面临着适用性的重大挑战。值得注意的是,近年来大数据、文本分析、机器学习等新模型和技术发展为有效应对传统政策评估方法面临的挑战提供了可行的路径。正如Currie等^[2]总结和展望实证微观经济学的发展前景所得出的结论,新数据结构(如高维参数、网络结构、多元异构等)和新数据类型(如实验数据、文本数据、图像数据、社交网络数据等)必将引领新的计量建模理论变革。其中,空间网络计量模型的发展为评估具有潜在溢出效应的政策效果提供了理论基础;数据规模和类型的极大丰富为多维度和精准化政策评估提供了现实可能;机器学习方法的进步则为大数据背景下的政策评估提供了技术支撑。

具体而言,洪永淼和汪寿阳^[17]基于经济学视角研究发现,大数据背景下政策评估正经历着从模型驱动到数据驱动、从无偏估计到有偏估计、从低维建模到高维建模、从低频数据到高频甚至实时数据、从结构化数据到非结构化数据等新的理论计量模型研究范式。Leung^[18]以及Comola和Prina^[19]探讨了存在网络溢出和网络结构变化情况下,对政策处理效应的估计问题,为空间网络模型用于政策评估提供了初步的理论基础和思路。Chernozhukov等^[20]利用机器学习算法,就高维模型中普遍存在的结构性偏差问题进行了理论探索,并提出了自动化偏差矫正方法,为因果效应的识别和定量估计提供了稳健性等理论支撑。

表 1 1992—2016 年中文核心期刊和 CSSCI 期刊公共政策研究文献高频关键词统计*

序号	关键词	词频	序号	关键词	词频	序号	关键词	词频
1	公共政策	1666	21	治理	100	41	可持续发展	68
2	政策分析	789	22	政策工具	99	42	政策制定过程	68
3	政策过程	369	23	税收政策	99	43	中央银行	67
4	政策执行	351	24	协商民主	97	44	城市规划	66
5	政策	301	25	政策研究	92	45	高等教育	66
6	政策评估	295	26	社会政策	92	46	议程设置	66
7	货币政策	283	27	公共政策研究	90	47	公共服务	65
8	教育政策	228	28	产业政策	90	48	问题	64
9	公共决策	191	29	经济发展	88	49	教育公平	63
10	财政政策	167	30	优惠政策	87	50	农民工	62
11	政策网络	165	31	公共利益	86	51	政策执行过程	60
12	地方政府	153	32	影响因素	86	52	利益集团	60
13	公民参与	151	33	公共管理	84	53	义务教育	55
14	政策议程	131	34	政策执行情况	76	54	创新	55
15	公众参与	124	35	和谐社会	75	55	社会保障	54
16	美国	121	36	政策科学	73	56	欧盟	53
17	经济增长	120	37	政策制定	72	57	评估	53
18	政府	117	38	价值取向	72	58	政策执行力	53
19	对策	112	39	金融机构	71	59	政策创新	52
20	中国	110	40	公共政策执行	69	60	技术创新	52

从文本分析和机器学习两类前沿技术在经济学研究中的应用情况看(图 4),文本分析方法自 1990 年代开始便陆续被应用于经济学研究,并在 2008 年后得到了与日俱增的关注和运用;机器学习方法则在 21 世纪后才开始被引入到经济学研究中,直到 2013 年后才迎来了较多的实证应用。总体来看,尽管上述两类方法目前相关论文占比均不足 1%,但

可以预见的是,在未来的经济学研究中两者均具有较大的应用前景和潜力。

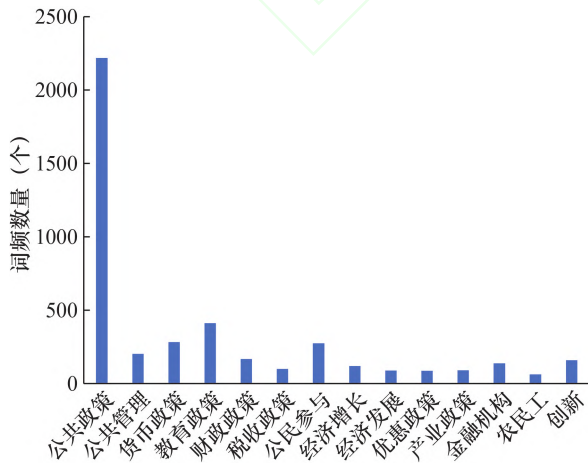


图 3 1992—2016 年中文核心和 CSSCI 期刊公共政策研究文献高频关键词统计^[16]

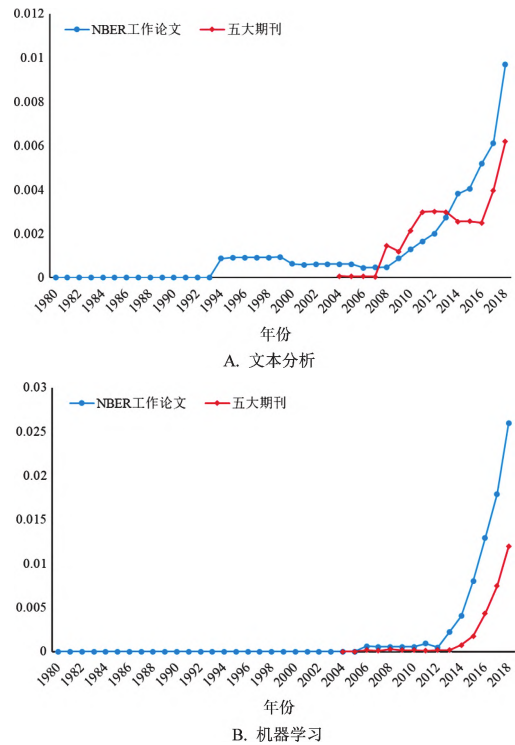


图 4 两种前沿技术相关论文发表所占比例^[2]

2 多维度溢出效应下的政策评估方法

随着我国全面深化改革进入深水区,涉及问题之多、领域之广、矛盾之深前所未有。在此现实背景下,中央和地方政府陆续出台的改革政策“组合拳”与日俱增,政策叠加效应愈发凸显。为此,如何评估此类复杂政策的实施效果,成为今后较长时间内研究者所不得不面临和亟待解决的难题。同时,考虑到我国以高铁开通和提速为代表的交通基础设施建设,互联网信息基础设施的建设和改善,以及国内统一大市场建设方面的体制机制完善,加速了地区间人才、技术、资本等要素的流动和信息的快速传播,由此导致政策在地区间、不同政策间存在着难以避免且不容忽视的溢出效应,从而为政策评估带来挑战。因此,本文着重围绕以下四种情形中存在多维度溢出效应时,探讨传统政策评估方法的主要问题与难点,以及空间网络计量模型的建模、识别和估计思路,以满足复杂情形下综合性政策评估精准化的需要。

2.1 单一政策存在个体、区域间溢出时,政策效果如何评估

当一个政策在个体或区域间存在溢出效应时,直接对比实验组和控制组往往会低估政策的效果。例如,当某个促进经济增长的政策仅影响到实验组,使得实验组产生额外经济增长 $g_1 > 0$,那么此时的政策效果就是 g_1 。但若存在实验组的行为或表现可以溢出到控制组的情况(图 5A),使得控制组的经济也额外增长 $g_2 > 0$,那么直接对比实验组和控制组得到的政策效果将是 $g_1 - g_2 < g_1$,此时政策效果将被低估。同时,由于被溢出影响到的控制组又可能反过来影响到实验组(图 5A),从而政策可能具有乘数效应。在此情形下,如果忽视溢出效应,那么这种乘数效应也会被遗漏,导致政策效果评估存在明显偏差。

在客观证据表明确实存在溢出效应时,Butts^[21]以及 Huber 和 Steinmayr^[22]等均假设这种溢出是局部的,比如有距离限制或仅局限在某一个地区,即要求溢出范围有限且易观察(图 5B),在此基础上进行后续政策评估。然而,现实情况往往更加复杂,以至于不能较好地识别出溢出效应的有效范围。针对该问题,本文提出以下研究思路:

(1) 在政策实施前,政策制定者可提前试点,通过实验设计的手段控制溢出范围,结合双重差分模型以便更好的评估政策效果。以创新型城市试点政

策对地区创新水平的影响为例,可选择距离较远的试点城市与对照城市,使得对照城市不受溢出的影响,此时便相当于人为地控制了潜在的溢出效应范围。

(2) 在政策实施后,当无法确定溢出范围时,研究者可以构建空间计量模型与数据驱动相结合的政策评估方法,通过数据驱动判断溢出范围,通过空间计量模型控制溢出的影响。例如在分析低碳城市试点政策对企业碳减排及绿色创新的影响效应时,可以通过距离试点城市 100、200、500 公里等阈值,分别构建不同的广义邻接空间权重矩阵,进而采用空间杜宾模型等计量方法考察该政策在不同空间范围上的有效溢出范围。

2.2 政策效应和溢出均存在动态变化时,政策效果如何评估

当政策效应和溢出效应存在动态性时(图 6),政策效果的评估也将面临挑战。如果政策对实验组的影响程度随时间变化,那么静态对比实验和控制组无法捕捉政策效果的动态性;如果政策对控制组的溢出影响可能滞后且随时间变化,那么忽略溢出的滞后可能错误估计政策效果的动态变化。

目前文献中已有的研究往往通过事件研究方法分析动态变化^[23],但缺乏对溢出效应的讨论。为了同时将溢出效应的动态性纳入考量,本文提出以下研究思路:

(1) 研究者可以借鉴匹配(Matching)的思路,利用政策效应和溢出效应的时间差,构建基于多重差分或合成控制的空间网络计量模型。例如,在考察高铁开通是否有利于促进人才跨区域流动等问题时,考虑到北京和天津在 2008 年便开通了京津城际铁路,保定市于 2012 年底开通运营京广高速铁路京郑段,而直到 2019 年底张家口才迎来京张高铁的正式开通运营。结合上述信息,我们可以用 2008—2011 年进行一个标准的 DID 估计,以此得到北京和天津开通高铁后的政策效应 Y_1 。在此基础上,再以 2008—2018 年为样本观测期,此时仅张家口作为控

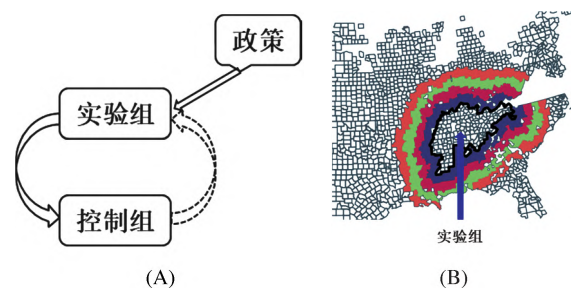


图 5 单个政策溢出效应示意图

制组,构建上述 4 个城市的空间网络计量模型,得到包含溢出效应在内的总政策效果 Y_2 。此时,只需要用 $Y_2 - Y_1$,即可将政策的溢出效应分离出来。

(2) 研究者可以利用溢出效应和政策效应不同的衰减特点,构建基于异质性动态趋势的空间网络模型。同样以创新型城市试点为代表,该思路的主要出发点在于:公共政策实施所带来的影响效果通常具有随着时间先增后减的特点;而受到时滞效应和相互作用的影响,该政策的溢出效应反而在政策实施后较长时间段内相对平稳。因此,可以尝试构建一个样本期较长的时变系数空间网络模型,以反映创新型城市试点政策本身及其溢出效应的动态变化,进而通过政策发生后各期影响系数大小及其变化趋势,分离出空间溢出效应大小。

2.3 组合政策存在多维度效果溢出时,政策效果如何评估

当存在组合政策时(图 7),政策效果通常体现在多个维度且具有溢出效应,直接评价单一政策的单一维度效果可能存在偏误。其主要原因在于:其一,单一政策可能影响到个体的多种行为,这些效果相互影响;其二,组合政策可能影响到个体的单一行为,这些政策相互叠加;其三,组合政策可能影响到个体的多种行为,效果和政策双重叠加。因而,多个政策多维度效应的相互影响和溢出给政策评估带来了困难。

然而,当前关于这一问题的探讨并不多,已有文献大多聚焦于单一政策下分别估计多维效果。为了有效处理这种复杂情形,本文从结构建模和反事实分析出发,提出以下潜在研究思路:

(1) 研究者可以将联立方程模型与空间计量模型相结合,拆解多重溢出背后的机制。例如,在考察互联网和交通基础设施各自通过线上和线下交流对

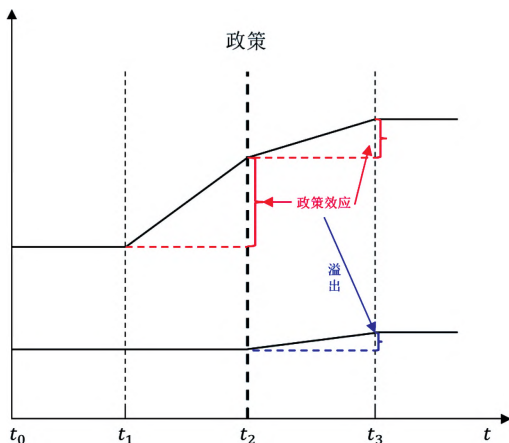


图 6 政策效应和溢出效应的动态变化

创新的影响及其溢出效应时,考虑到线上和线下交流两者本身极有可能存在相互影响,因此有必要引入联立方程模型,将这种潜在的互动关系进行控制。并在此基础上构建空间联立方程模型,实现对空间溢出效应的有效分离。

(2) 研究者还可以通过仿真模拟,测算单一政策的单一效果、多维度效果。假定作者重点关注高铁开通对区域创新成果的影响效应,则为剔除互联网政策可能带来的干扰,可以在联立方程估计的基础上,通过仿真模拟的方式,模拟在不同互联网发展水平下,各地区陆续开通高铁后的创新成果数量与质量提升情况。

2.4 存在复杂的高维空间网络数据时,政策效果如何评估

得益于数据收集和处理技术的发展,研究者对高维空间网络数据的获取成为可能,但由于这类高维空间网络数据的复杂性,与之对应地对政策评估方法的改进提出了更高的要求。其一,很多高维数据是以文本、图像甚至视频形式存在的、非结构化的数据类型,对数据的清洗和提取提出了新的更高的要求;第二,数据存在多个维度,例如数据可能呈现出类似“城市—产业—公司—产品—时间”等多个层面;第三,数据的多个维度间可能存在复杂的网络溢出效应,例如城市之间、产业之间、产品之间均可能存在不同的互补和替代效果。

现有文献对于此类情形的研究通常采用将多维度与溢出分开的思路,如使用多维度数据但不考虑网络和溢出^[24]、分别研究多个网络结构下的溢出^[25]等。为此,本文认为处理此类问题的潜在思路在于将经济学方法与计算机领域相关方法相结合。

(1) 利用文本分析、图像分析等方法提炼非结构化数据中的有效信息,进而运用双重差分模型和空间计量模型相结合的思路进行政策评估。例如运用文本分析方法对 2010 年以来各地方政府工作报告内容进行处理,提取出各地区各年份对人才引进的关注程度,及其具体的人才政策细节或吸引力,进而构建高维空间网络计量模型考察其对各城市各年

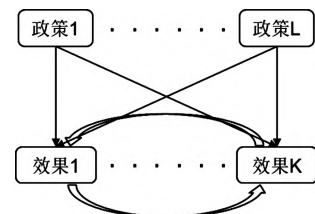


图 7 组合政策多维度效果溢出

份各产业中各公司的创新成果数量和质量的影响效应。与此同时,还可进一步重点关注各公司研发人员之间的知识交流共享情况对创新成果的潜在溢出效应。

(2) 基于机器学习等方法对复杂高维空间网络模型中的冗余参数进行降维,进而采用嵌套双重差分方法的空间计量模型实现政策效应和溢出效应的分离。例如,为得到科研人才集聚规模对创新成果的影响及其潜在溢出效应,该方法主要思路在于:利用机器学习方法对以各地区各年份各产业各公司的招聘信息为代表的大数据资源进行自动化分类等降维处理,从而得到不同地区、不同行业乃至不同企业对高科技人才的市场需求,并结合嵌套双重差分估计思路的高维空间网络计量模型进行更加深入具体的探讨。

3 中国的研究基础与优势

随着我国一系列交通基础设施、工业互联网基础设施、国家级新区、国家高新区、自由贸易区等公共政策的陆续实施,以及国内统一大市场建设方面的体制机制完善,为政策制定者和研究者系统把握各类型政策的实施效果提供了大量的数据支撑和政策样板。在我国线上、线下交流日益频繁带来政策效果溢出效应复杂化的现实背景下,既需新的政策评估方法助力政策的事前评估和事后评价,提高政策精准性和有效性,又为相关研究方法的提出和改进提供了研究基础和优势。具体而言,我国的研究基础和优势及其背后的传导机制主要体现在以下几个方面。

(1) 我国铁路、公路、水路立体交通网发展迅速、增量可观,政策效应凸显。《国家综合立体交通网规划纲要》提出我国在2035年包括铁路、公路、水路等立体交通网里程要达到70万公里。中国交通基础设施增量可观的同时,也将通过同步实现数字化、智能化,共同驱动智慧交通行业繁荣发展。从交通基础设施促进人流、物流、信息流在区域间快速流动,促进知识的共享交流这一客观事实出发,包括公路通车、高铁网络等一系列相关基础设施建设带来的各方面生产力提升提供了丰富的数据和案例,为中国学者进行多维度溢出效应复杂化政策评估提供了现实条件。

(2) 互联网云计算、大数据、人工智能、物联网、工业互联网、区块链等数字经济重点产业兴起。《中华人民共和国国民经济和社会发展第十四个五年规

划和2035年远景目标纲要》围绕“打造数字经济新优势”,做出培育壮大“大数据”等新兴数字产业的明确部署。不仅如此,近年来我国众多政策均强调了要充分激发数据要素价值潜能,夯实产业发展基础,构建稳定高效产业链,统筹发展和安全,培育自主可控和开放合作的产业生态,打造数字经济发展新优势,为建设制造强国、网络强国、数字中国提供有力支撑。我国互联网基础设施建设不断加速,其通过线上交流传播显性知识,进而形成知识空间溢出效应,最终影响区域创新发展水平的作用机理愈发凸显。不仅如此,在大数据、云计算时代背景下,对于各种高维的非结构化数据获取已成为可能。从而,如何在此类复杂网络数据下有效识别出政策效果及其潜在的网络溢出效应,无疑成为未来构建和改进新型政策评估方法的重要出发点和落脚点。

(3) 全国统一大市场体制机制完善,市场资源整合成为知识空间溢出与创新发展新引擎。党的十九届五中全会通过的《中共中央关于制定国民经济和社会发展第十四个五年规划和二〇三五年远景目标的建议》提出,要加快建设现代化经济体系,加快构建以国内大循环为主体、国内国际双循环相互促进的新发展格局。在国际贸易保护主义抬头,经济全球化遭遇逆流的时代背景下,促进全国统一大市场体制机制的完善,既是做大做强“内循环”的现实要求,又是增强商品、资本与劳动力等市场资源整合能力,释放中国市场经济活力、潜力,发挥市场规模和体量优势,驱动区域创新发展水平提升的重要体现。总体而言,通过构建广阔和统一的国内大市场,不仅有利于形成产业链上下游关联,促进市场资源整合以及科技合作,而且还有利于扩大新产品的市场潜力,改善创新资源配置效率,最终促进科学技术的空间溢出和创新数量和质量双提升。

(4) 我国贯彻创新驱动发展战略,科技、人才、创新战略意义提升到全新高度。习近平总书记在党的二十大报告中指出:“必须坚持科技是第一生产力、人才是第一资源、创新是第一动力,深入实施科教兴国战略、人才强国战略、创新驱动发展战略,开辟发展新领域新赛道,不断塑造发展新动能新优势。”“加快实施创新驱动发展战略。加快实现高水平科技自立自强”。在全国各地区创新激励和人才引进政策层出不穷、科研人才集聚规模日益凸显的当下,亟需新的评估方法以有效精准地评估其在跨区域知识交流过程中对创新成果等的政策效果,以此加快中国建设成为创新型国家目标实现。

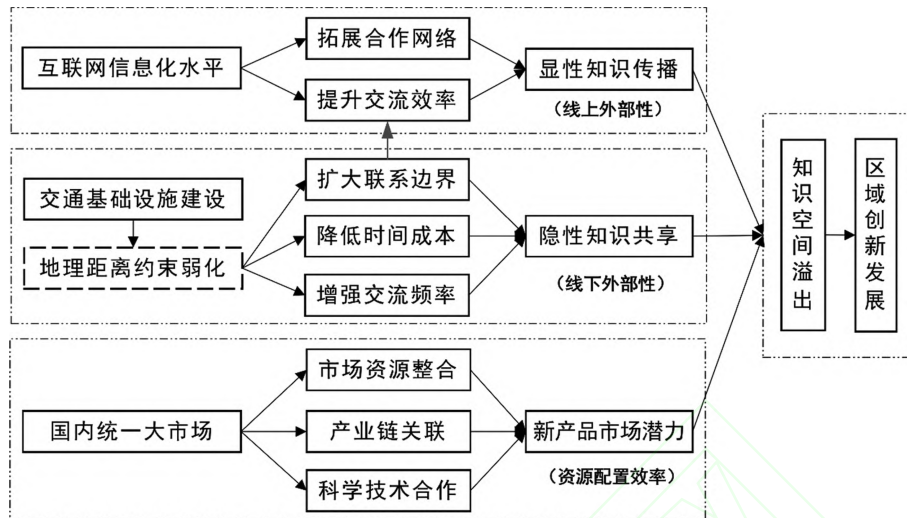


图 8 高铁、互联网发展、市场整合与人才促进知识溢出及创新发展的作用机理
(资料来源:作者根据郑维伟^[1]研究成果拓展绘制)

4 结 语

本文从服务国家治理体系现代化的角度指出了发展完善政策评估方法的重要性。文章首先回顾了传统政策评估方法中较为经典的固定效应模型、工具变量法、双重差分法、断点回归设计、合成控制等方法及其应用趋势,并强调了文本分析和机器学习在未来应对复杂空间网络数据时的重要应用场景,然后重点探讨了空间网络计量模型在应对多维溢出效应挑战下,进行科学化、严谨化、精准化和系统化政策评估的独特优势,以及其在建模、识别与估计层面潜在的研究思路。最后,本文结合我国在交通基础设施建设、互联网数字经济产业发展、全国统一大市场以及创新人才发展战略等方面探讨政策溢出效应的研究基础与优势,强调了我国众多的公共政策样本是未来考虑多维度溢出效应下进行政策评估方法的提出与改进的天然“沃土”。

参 考 文 献

- [1] 郑维伟, 瞿茜, 刘耀彬, 等. 互联网时代“线下交流”不再重要了吗? ——基于“知识空间溢出”悖论与区域创新的检验. 财经研究, 2023, 49(3): 155—168.
- [2] Currie J, Kleven H, Zwiens E. Technology and big data are changing economics: mining text to track methods. AEA Papers and Proceedings, 2020, 110: 42—48.
- [3] Bai J. Panel data models with interactive fixed effects. Econometrica, 2009, 77(4): 1229—1279.
- [4] 林晨, 陈荣杰, 徐向宇. 外部产业投资与区域协调发展——来自“三线建设”地区的证据. 经济研究, 2022, 57(3): 173—190.

- [5] Thistlethwaite DL, Campbell DT. Regression-discontinuity analysis: an alternative to the ex post facto experiment. Journal of Educational Psychology, 1960, 51(6): 309—317.
- [6] Angrist JD, Lavy V. Using Maimonides' rule to estimate the effect of class size on scholastic achievement. The Quarterly Journal of Economics, 1999, 114(2): 533—575.
- [7] Hahn J, Todd P, Klaauw W. Identification and estimation of treatment effects with a regression-discontinuity design. Econometrica, 2001, 69(1): 201—209.
- [8] Card D, Krueger AB. Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. American Economic Review, 2000, 90(5): 1397—1420.
- [9] 周黎安, 陈焯. 中国农村税费改革的政策效果: 基于双重差分模型的估计. 经济研究, 2005, 40(8): 44—53.
- [10] Beck T, Levine R, Levkov A. Big bad banks? The winners and losers from bank deregulation in the United States. The Journal of Finance, 2010, 65(5): 1637—1667.
- [11] 黄炜, 张子尧, 刘安然. 从双重差分法到事件研究法. 产业经济评论, 2022(2): 17—36.
- [12] Roth J, Sant'Anna PHC, Bilinski A, et al. What's trending in difference-in-differences? A synthesis of the recent econometrics literature. Journal of Econometrics, 2023, 235(2): 2218—2244.
- [13] Olden A, Men J. The triple difference estimator. The Econometrics Journal, 2022, 25(3): 531—553.
- [14] 舒海兵, 叶五一, 李熠熠, 等. 非实验数据政策效应评估理论与实证研究方法. 中国管理科学, 2007, 15(6): 140—148.
- [15] Abadie A, Gardeazabal J. The economic costs of conflict: A case study of the Basque country. American Economic Review, 2003, 93(1): 113—132.
- [16] 邓汗清. 中国公共政策的知识图谱分析——基于文献计量学的研究. 深圳: 深圳大学, 2017.

- [17] 洪永森, 汪寿阳. 大数据如何改变经济学研究范式?. 管理世界, 2021, 37(10): 40—55, 72.
- [18] Leung MP. Treatment and spillover effects under network interference. *The Review of Economics and Statistics*, 2020, 102(2): 368—380.
- [19] Comola M, Prina S. Treatment effect accounting for network changes. *The Review of Economics and Statistics*, 2021, 103(3): 597—604.
- [20] Chernozhukov V, Newey WK, Singh R. Automatic debiased machine learning of causal and structural effects. *Econometrica*, 2022, 90(3): 967—1027.
- [21] Butts K. Difference-in-differences estimation with spatial spillovers. (2023-06-23)/[2023-08-23]. <https://arxiv.org/abs/2105.03737>.
- [22] Huber M, Steinmayr A. A framework for separating individual-level treatment effects from spillover effects. *Journal of Business & Economic Statistics*, 2021, 39(2): 422—436.
- [23] Sun LY, Abraham S. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 2021, 225(2): 175—199.
- [24] Liu Q, Lu Y. Firm investment and exporting: evidence from China's value-added tax reform. *Journal of International Economics*, 2015, 97(2): 392—403.
- [25] Bloom N, Schankerman M, Van Reenen J. Identifying technology spillovers and product market rivalry. *Econometrica*, 2013, 81(4): 1347—1393.

Policy Evaluation Methods, Strategies and China Advantages of Multi-dimensional Spillover Effects: Perspective of Spatial Network Econometric Models

Juncong Guo Weiwei Zheng Xi Qu*

Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai 200030

Abstract On the basis of combing the traditional policy effect evaluation methods such as fixed effect model, instrumental variable, difference in differences, regression discontinuity design and synthetic control method, this paper investigates policy evaluation methods under multidimensional spillover effects and emphasizes the unique advantages of spatial network models in addressing the challenges of multidimensional spillover effects. In particular, this paper focuses on the methods, identification and estimation problems of the spatial network econometric model when dealing with the inter-individual and inter-regional spillovers of a single policy, the dynamic changes of policy effects and spillover effects, the multi-dimensional spillover effects of combined policies, and the existence of policy effects in high-dimensional network spatial data. The continuous improvement of China's transportation, Internet and other hardware infrastructure constructions, as well as the increasing improvement of the soft environment such as the national unified market, provide an original fertile soil and China advantage for Chinese scholars to deeply explore the policy effect evaluation method. At a time when policy effect spillovers are becoming more complex, discussions in this paper can provide an important reference for the future development of policy evaluation methods.

Keywords causal inference; policy evaluation; multidimensional spillover effect; spatial network econometric model; China advantage

(责任编辑 姜钧译)

* Corresponding Author, Email: xiqu@sjtu.edu.cn