

自然实验在政治科学领域的应用与反思

法 卉*

2021年10月,戴维·卡德(David Card)、乔舒亚·安格里斯特(Joshua Angrist)和吉多·因本斯(Guido Imbens)三位经济学家因在自然实验(natural experiment)的运用方面做出了“挑战传统智慧”的开创性工作,被授予了诺贝尔经济学奖。与西方经济学重视理论、通过抽象假设来推导结论的传统思维不同,这三位获奖者相信来自现实世界的数据才能够直观且真实地揭示外在因素对经济行为产生的影响。他们通过创新的研究设计和对统计学的巧妙运用,在经济学领域引领了一场“实证主义革命”,旨在现实世界中寻找可靠的因果关系证据以回答教育、健康、就业等重要的政策问题。

卡德和他的同事艾伦·克鲁格(Alan Krueger)以新泽西州出台提高最低工资的法令为契机进行了自然实验,试图验证提高最低工资是否会如同传统经济学中所预测的那样,降低雇主对工人的需求,从而导致失业率的提升。然而,对新泽西州以及与其临近的宾夕法尼亚州东部的400多家快餐店进行调查后,他们发现提高最低工资并没有对就业率产生影响。^①这一研究结果为之后美国各州提高最低工资铺平了道路,大量工人因此受益。而安格里斯特和因本斯则利用工具变量(instrumental variables)的构造和二阶段最小二乘估计法解决了受教育年限与未来工资水平模型构造的内生性问题。他们的研究选取了与家庭背景或是先

* 法卉,社会心理学博士,复旦大学社会科学高等研究院博士后。

① David Card and Alan B. Krueger, “Minimum Wages and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania”, *The American Economic Review*, Vol.84, no.4 (1994).

天能力无关的出生年份作为工具变量,有效证明了受教育年限的延长对于未来工资水平的正向作用,为美国教育财政投入及政策提供了重要依据和建议。^①

尽管这三位获奖者的学术成果都曾因为创新性和对既有研究的批判遭受过主流经济学界的质疑和谴责,但他们一次次通过严谨的因果分析和缜密的研究结果成功进行了反击,将经济学从固化思想和既有认知的束缚中逐步解放出来,引领经济学迈向更加多元化的发展方向。而这场“实证主义革命”并未就此止步于经济学界。由于能够通过简洁的研究设计对两个变量之间的因果机制进行有效推断,并能够针对广泛且复杂的社会问题提供更有效的政策建议,“自然实验”这一概念近年来也逐渐渗透到政治科学领域,越发受到主流学界和社会的重视。

鉴于尚没有文献专门对自然实验在政治科学领域的现状与应用进行过回顾与总结,本文试图较系统地从事背景、方法技术、研究实例和方法反思这四个方面对自然实验法的现状与趋势进行梳理,希望能够为相关研究者提供一些借鉴和参考。

一、自然实验法的发展背景

自然实验法属于实验政治学领域的一个重要方法论分支。虽然如今在政治学领域,定量实验方法的使用日益受到推崇。但回首实验政治学科的发展历史,可谓艰辛曲折。本节将在简要梳理实验政治学科的发展进程后,对包含自然实验在内的三类研究设计进行介绍。

(一) 实验政治学的兴起与发展

20 世纪初期,当自然科学领域的学者们正热火朝天地进行着各类实验时,政治学等社会科学领域的学者依旧对于进行实验研究的伦理性和必要性存有

^① 参见 Joshua D. Angrist and Alan B. Krueger, “Does Compulsory School Attendance Affect Schooling and Earnings?” *The Quarterly Journal of Economics*, Vol.106, no.4(1991), pp.979—1014; Joshua D. Angrist, Guido W. Imbens and Donald B. Rubin, “Identification of Causal Effects Using Instrumental Variables”, *Journal of the American Statistical Association*, Vol.91(1996), pp.434, 444—455。

质疑。与以客观物质和材料为对象的自然科学实验不同,社会科学实验无法保证实验操作不会对其对象——具有主观能动性和情感的人——带来任何的负面后果。同时,主流政治学学者一致认为,由于研究者无法在实验环境中还原出真实、复杂的社会系统,其研究结果对于现实世界也缺乏指导性和实践性。^①

转机出现在 20 世纪 50 年代——行为主义革命 (behavioral revolution) 的兴起成了实验政治学科发展的分水岭。在这场变革中,政治学者对传统政治研究中抽象思辨和演绎的方法进行了反思与创新,认为应该将政治科学研究的对象锁定到实际存在且可观察到的政治行为上。而其研究的最终目的应当是对政治行为的解释、预测和控制,而并非对社会政治现象的静态描述和简单写实。而为了实现这一目的,就必须将自然科学和社会科学的理论方法引入政治研究,采用新的思维方式和分析方法重新认识和解释社会政治现象。由此,区别于规范研究和阐释研究、通过科学方法进行研究的政治科学开始成为一门独立学科。问卷调查以及将人群按其国别、宗教或是生活年代等背景进行划分后进行比较以验证理论假设的观察研究法 (observational methods) 成为当时最受推崇的研究方法。^②在这样的背景下,实验政治学也开始逐渐得到发展。

进入 20 世纪 70 年代后,受到心理学及其方法论的影响,以实验室实验为主要研究方法的政治心理学作为政治科学领域的一个新兴学科,开始逐步成型并得到快速发展。密歇根大学、加利福尼亚大学和纽约州立大学都纷纷设立了政治心理学专用的实验室及相关课程,培养了一批掌握实验研究方法的政治学者。此后,政治学领域的实验研究开始如雨后春笋般不断涌现。到了 20 世纪 90 年代,三大政治学期刊——《美国政治学评论》(American Political Science Review, APSR)、《美国政治科学期刊》(American Journal of Political Science, AJPS) 及《政治学杂志》(Journal of Politics, JOP) 上刊登的实验研究将近 50 篇,几乎是 1950 年至 1980 年刊登的所有实验研究数量的总和。^③

① Abbott L. Lowell, "The Physiology of Politics", *American Political Science Review*, Vol.4, no.1(1910), pp.1—15.

② Damien Bol, "Putting Politics in the Lab: A Review of Lab Experiments in Political Science", *Government and Opposition*, Vol.54, no.1(2018), pp.167—190.

③ See Rebecca B. Morton and Kenneth C. Williams, *Experimental Political Science and the Study of Causality: From Nature to the Lab*, Cambridge: Cambridge University Press, 2010.

但即便如此,政治学界对于实验研究的质疑依旧没有消除。其中最大的质疑声还是在于实验环境与真实社会环境的脱节。与研究物体客观规律的物理学以及研究人们内在意识的心理学不同,政治学的研究对象是人们在现实世界中的自然行为表现,因此传统政治学界对于存在人为介入的实验方法始终有所排斥。而在同一时期,因为相同的理由而与实验政治学同样面临困境的还有实验经济学。但随着2002年弗农·史密斯(Vernon Smith)、2007年罗杰·迈尔森(Roger Myerson)两位经济学者接连因为其实验性研究在经济领域所作出的开创性贡献而获得诺贝尔奖(其中迈尔森也针对投票体制进行过实验研究),实验经济学终于一跃登上主流舞台,受到学术界关注。而在这样的背景之下,实验研究也终于在步入21世纪后逐渐被学术界接纳认可,成为政治科学研究的主流方法之一。

当然,实验政治学的迅速崛起并不仅仅是因为沾了实验经济学的光。伴随着政治现象和政治过程的日益复杂化和交互化,经典规范政治理论在解释力和对政治实践的指导能力上都表现出诸多局限。^①仅仅依靠观察研究或是非实验性研究(如问卷调查)不足以帮助研究者们锁定变量或因素间的因果关系,也无法为诸如全球治理、国家能力与治理、数字民主与网络暴政、恐怖主义与文化冲突等现实政治问题的解决提供经验建议。因此,能够判定因果效应、检验因果机制的实验研究方法必然会受到政治科学领域的青睐。

(二) 实验政治学的研究设计:实验室实验、田野实验以及自然实验

从实验形式来看,实验政治学主要包括实验室实验(conventional lab experiments)、田野实验(field experiments)和自然实验等。这些研究设计的核心思想都在于将对人的行为或态度产生影响的各类因素进行控制后,识别和评估干预条件(自变量)与人的行为或态度(因变量)之间的因果关系。但在不同类型的研究设计中,被研究对象受到干预的方式和过程有所不同,干预的程度和受控制的程度也差异甚大。表1列举了实验室实验、田野实验、自然实验以及非实验研究之间的差别。各类实验设计都存在着优势和劣势,研究者需要根据研究目

^① 孟天广:《从因果效应到因果机制:实验政治学的中国路径》,载《探索》2017年第5期。

的与可行性来选择合适的实验设计。

表 1 实验分类及相关比较

	被研究对象	实验设计 及材料	激励	实验环境	被研究对象 知晓实验	研究者 进行干预
实验室实验	大学生	虚构	诱导价值	实验室	是	是
田野实验	各类人群	现实	真实激励	社会环境	否	是
自然实验	各类人群	现实	真实激励	社会环境	否	否
非实验研究	各类人群	现实	真实激励	社会环境	否	否

资料来源:罗俊、汪丁丁、叶航、陈叶烽:《走向真实世界的实验经济学——田野实验研究综述》,载《经济学》2015年第3期,表1。

实验室实验源于心理学实验的启发,往往以学分或是少量的金额作为报酬,招募在校大学生成为被研究对象。通过在实验室中设置虚构的情景和信息,研究者记录和观察被研究对象的某一特定行为或态度如何随着干预变量的变化而变化,以验证基于理论所推导出的假设。虽然实验室实验看似对影响变量有着较高的控制度,能够相对准确地推导出研究者所关心的两个变量之间的因果关系,但由于实验场景过于做作且被研究对象没有代表性,实验室实验研究经常会遭受社会科学研究者的质疑和批判。

田野实验和自然实验则通过改进和加强研究方法,帮助研究者在更真实的世界中检验理论、分析人们的政治态度和行为。田野实验是在真实社会环境中开展的实验研究,任何人都可能成为被研究对象。与实验室实验不同的是,被研究对象对于自己正在参与实验一事并不知情,因此,他们在追求社会福利、支持政党当选等真实激励的驱动下,发乎自然地对实验任务进行反应,这在很大程度上减少了实验室实验中人为干预性可能对结果产生的影响。但是,在现实世界中进行实验研究的难点在于,能够用于向被研究对象实施实验干预的手段变得更少(例如在选举期间,向选民随机发放印有不同宣传标语的传单),而且研究者也无法逐一确认被研究对象是否真的受到干预条件影响,或是否对干预做出反应。这可能会对因果推论的准确性产生一定的干扰。

相比先确定被研究对象,再进行人为干预的实验室实验与田野实验,自然实验的特殊性在于,当政策变化等非人为干预或是随机化的实验条件在社会生

活中自然形成之后,研究者们再通过研究设计去寻找出符合要求的被研究对象,进而探讨自变量与因变量之间的因果关系。在下一节中,我们将对自然实验的概念与基本研究设计进行更为具体的介绍。

二、自然实验的概念、基本研究设计与优势

自然实验,如同其字面意思一般,是一场自然发生的实验。因此,它具有随机分配实验(randomized controlled experiment)的部分特征^①,例如都存在某一项或是一组干预措施(implementation/treatment),且该干预措施的有无是导致被研究对象的态度或行为发生变化的直接原因。但两者之间的根本区别就在于:自然实验中的“干预”是随自然环境(比如地震、海啸等灾害的发生)或是社会环境(比如新政策或是法案的颁布)的改变而自然发生的,其发生的时间、地点或是方式都无法受到任何人工性的干预或安排。

莱瑟代尔指出,自然实验具备以下两个特点:第一,干预措施的实施并不取决于研究者是否事先制定了评估该干预措施的计划;第二,出于伦理或政治因素的考量,无法保证干预措施能够达到完全的随机分配。^②因此,当实验条件在社会生活中自然形成时,研究者必须先运用“识别策略”(identification strategy),即通过进行一些精妙的研究设计来确保被研究对象随机受到干预措施影响,并对其所关心的两个变量之间存在因果关系的理由进行充分的说明和假设后,才能开展数据的收集和统计,确保因果推断的有效性。^③

(一) 自然实验的基本研究设计

作为当代西方经济学界最被广泛接受的研究方法之一,自然实验目前已经

^① 参见 David Freedman, Robert Pisani and Roger Purves, *Statistics*, 3rd ed. New York: Norton, 1997, pp.4—8。

^② Scott T. Leatherdale, “Natural Experiment Methodology for Research: A Review of How Different Methods can Support Real-World Research”, *International Journal of Social Research Methodology*, Vol.22, no.1(2019), pp.19—35.

^③ Cyrus Samii, “Causal Empiricism in Quantitative Research”, *The Journal of Politics*, Vol.78, no.3(2016), pp.941—955.

演化出了许多成熟的研究设计,包括工具变量、双重差分(difference in difference)、倾向值匹配(propensity score matching)和断点回归设计(regression discontinuity)等。

1. 工具变量

工具变量是社会科学定量分析中解决内生性问题的重要手段之一。其原理在于寻找一个会与自变量(原因)密切相关,但与因变量(结果)毫无关系的外生变量作为“工具”,以探讨自变量与因变量之间的因果关系。具体来说,如果自变量与因变量之间存在因果关系,那么作为外作用力的工具变量对自变量产生的影响也就势必传递到因变量。因此,只要工具变量对因变量的间接影响是显著的,那么我们就可以推断出自变量与因变量之间必然存在因果关系。例如,亨德森和布鲁克斯将雨天作为工具变量,探讨了选民投票成本与投票行动之间的关系。^①显然,下雨天出行会增加选民出门投票时的成本(例如需要更多的准备时间等),但是否下雨与投票行为本身并没有直接关联。因此,如果与晴天相比,雨天时的选票显著减少,那我们就可以推测出投票成本与投票行为之间存在因果关系:投票成本升高会抑制投票行为。

虽然工具变量能够较好地解决内生性问题,但要在现实世界中寻找到符合研究需要且合格的工具变量并不容易。目前,寻找工具变量的思路有以下几个方面:(1)来自自然界的物候天象——降雨、龙卷风、地震等自然现象在一定地域范围内具有高度的随机、外生特性,但同时又可能对社会进程产生影响;(2)来自生理上的自然现象——性别、死亡率、出生日期等有机体的自然历程既具有随机性,又往往和特定的经济社会进程相关;(3)来自外生性政策、事件干预——从社会规章制度(例如通过抽签决定是否服军役)到恐怖袭击的发生,这些事件的发生是完全随机的,但又能对个体或是群体今后的收入、健康或是政治态度等产生影响。^②

^① John Henderson and John Brooks, “Mediating the Electoral Connection: the Information Effects of Voter Signals on Legislative Behavior”, *The Journal of Politics*, Vol. 78, no. 3 (2016), pp. 653—669.

^② 陈云松:《逻辑、想象和诠释:工具变量在社会科学因果推断中的应用》,载《社会学研究》2012年第6期。

2. 双重差分

双重差分主要用于研究某项政策实施前后的效果评估。其基本思想是将受到政策影响的地区或城市作为干预组,并选取一个相对同质但未受到政策影响的地区或城市作为控制组,接着通过比较两组样本数据来估计政策的实际效应。例如,如果想要评价A市新颁布的人才引进政策效果如何,首先要选取一个符合平行趋势假定的B市作为控制组,即在没有实行新政策的情况下,A、B两市在1—2年内人才落户(因变量)的增长率应该一致。如果A市的人才落户数量在政策实施前后有所变化,我们无法断定该变化是由政策,还是由时间引起,也就是说即便没有受到政策影响,随时间推移A市的人才落户数量也可能产生同样的变化。此时,为了剔除时间效应的影响,就需要引入控制组以观察在A市开始执行政策后的同一时间段内,未受到政策影响的B市的人才落户数量会发生怎样的变化。接着,在A市政策实施前后人才落户数量的变化量中,再扣除通过B市估计出的,因时间而产生的变化量,就能得到政策干预所带来的净效应,从而进行因果关系的估计(见图1)。

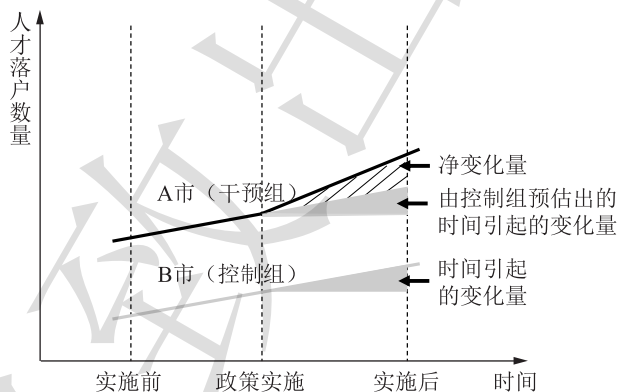


图1 双重差分法示例

3. 倾向值匹配

匹配的核心思想是运用统计学方法、人为地选取一个控制组,通过可观测特征(observable characteristics)为每一个身处干预组的被研究对象“匹配”一个未受干预的对象进行比较。而倾向值匹配就是保证匹配起来的个体之间有着相等或近似的倾向。^①这一方法的优势在于其能够很好地解决存在多混淆变量

^① 胡安宁:《倾向值匹配与因果推论:方法论述评》,载《社会学研究》2012年第1期。

(confounding variables)时的控制问题。例如当我们要探讨接受大学教育对个人收入的影响时,年龄、性别以及地区差异等混淆变量都可能对研究结果产生影响。为了对这些混淆变量进行控制,从而真正确定接受大学教育与个人收入两者之间的因果关系,我们可以将混淆变量纳入 Logistic 回归模型以产生一个预测个体受到自变量影响的概率,并从有相同概率(即上大学的概率)的群组中选出受过大学教育和没有受过大学教育的个体进行匹配,对其个人收入进行比较研究。因为已有的混淆变量已经在基于倾向值的配对过程中得到控制,因此两组个体收入上产生的差异就只能归因于大学教育的有无,从而明确两者之间的因果关系。

4. 断点回归设计

所谓“断点”,是指研究者预测存在某一个会对因变量产生直接影响的临界点。超越该临界点与否决定了被研究对象是否受到干预。同时,为了确保被研究对象的同质性、确保干预条件分配的随机性,研究者往往在临界点附近选取研究样本,即将略高于临界点的样本作为控制组、略低于临界点的样本作为干预组进行比较分析,以确定干预效应与因变量之间的因果关系。例如,当我们希望探讨进入重点大学对个人收入的影响时,录取分数线便成为一个精确的“断点”——高于此分数线学生确定能够进入重点大学,而低于此分数线学生则确定无法进入重点大学。这样能够将样本进行“一刀切”的断点回归设计被称为精确断点回归(sharp regression discontinuity)。通过对比略高、略低于录取分数线的两组学生毕业后的收入状况,就能够最大程度地排除学生本身的能力、家境等可能对个人收入产生影响的其他因素,更为准确地估计出进入重点大学与收入的因果效应。

但在其他国家,根据教育制度和模式的不同,录取分数线将不再是判断学生能否入学的明确标准,达到重点大学的录取分数线只是显著地提高了学生被录取的概率,而不能确保其一定被录取。这种情况下就需要采取模糊断点回归设计(fuzzy regression discontinuity),将是否高于录取分数线定义为一个分组虚拟变量,将其作为影响干预效应的工具变量来使用,进而估计因果效应。

除了精确断点回归与模糊断点回归以外,还存在一种比较特殊的断点回归设计——地理断点回归(geographic regression discontinuity),即以实在的区域分界线作为断点进行实验设计。例如,戴尔等人以越南两组曾经从属不同政体的

相邻村庄为研究对象,以1968年形成的疆界为地理断点进行自然实验设计,探讨了历史政体对地区长远发展可能产生的影响。^①

虽然断点回归设计发展较晚,在近几十年才为社会科学界所重视和接受,但这一实验设计被认为是最接近于随机实验方法的拟随机实验方法,在微观政策评价方面进行因果推断就有较大的优势。

(二) 自然实验的优势

综上所述,各类研究设计使得自然实验在最大程度上克服了传统的实验室实验和田野实验中的人为操作性,在自然产生的政治情景之中对变量之间的因果关系进行检验,兼顾了现实性与科学性。因此,自然实验对于决策者来说尤为重要,因为它既能够为政策建议提供有效依据;又能够对政策效果进行有效且及时的评估。

1. 为政策建议提供有效依据的自然实验

当灾害等社会突发事件发生时,自然实验能够成为研究者检验其经验推论有效性的一种重要工具,也为政策建议的提出提供现实依据。1854年,麻醉医师约翰·斯诺(John Snow)就通过一个简单的自然实验首次证明了霍乱的传播途径。^②19世纪的英国霍乱持续肆虐,当时的普遍观点是,霍乱是由瘴气引起的。而斯诺基于自身行医经验,认为霍乱是通过被污染的水源而非空气进行传播。但这一假设与当时盛行的瘴气学说背道而驰,因此他必须拿出确凿证据来证明自己的观点。

当时的伦敦有两家主要供水公司——兰贝思水务公司(the Lambeth Company)与索思沃克和沃克斯霍尔水务公司(the Southwark and Vauxhall Company)。在1852年之前,这两家公司都从被污染的泰晤士河下游取水送至居民家中。但是在1852年,兰贝思公司将其供水管道的源头移到了更为干净的泰晤士河上

^① Melissa Dell, Nathaniel Lane and Pablo Querubin, “The Historical State, Local Collective Action, and Economic Development in Vietnam”, *Econometrica*, Vol. 86, no. 6 (2018), pp. 2083—2121.

^② 参见 David Freedman, “From Association to Causation: Some Remarks on the History of Statistics”, *Statistical Science*, Vol. 14 (1999), pp. 243—258; David Freedman, *Statistical Models: Theory and Practice*, Cambridge: Cambridge University Press, 2005。

游。这一改变为斯诺提供了实施自然实验的机会。他认为伦敦超 30 万户不同年龄和社会阶层的居民在“非自主选择,且不知情的情况下被分为了两组,其中一组居民的供水中含有污染物,甚至还可能含有来自霍乱患者的东西,而另一组居民的供水则完全没有这种杂质”^①。

在确认了居民被分配为干预组和对照组的随机性后,斯诺绘制了一张简单的交叉表以展示 1853—1854 年各户的霍乱死亡率与负责其供水的公司。数据表明,兰贝思公司供水范围内的居民霍乱死亡率更低,仅为另一家公司的八分之一左右。这充分验证了斯诺的判断:霍乱是通过污水而非瘴气传播。这一早期的自然研究案例研究充分说明了在一些特殊情景下,通过自然实验以进行有效的因果推断,能够为社会科学研究者所关心的实质性问题提供解决方案或启发。

2. 对政策进行有效评估的自然实验

对于政策制定者来说,基于自然实验的实证数据能够为其政策发展战略提供可靠依据。任何一个政府或是部门都希望有充分的证据能够表明自己即将出台的政策是切合实际且行之有效的。但传统的实证研究法(例如实验室实验)中所得到的结论由于受限较多,往往很难外推至真实世界。而从伦理角度考量,研究者也无法通过进行田野实验来检验一项政策的效果。

基于此,自然实验便成为了能够对政策效果进行评价的最佳工具。^②首先,政策评价不仅能够向民众公开政策执行效果,由研究者(而非政策制定者)所进行的第三方评估还能够完善行政问责制度。其次,基于自然实验的评价能够对政策效果进行及时且可靠的反馈。尽管许多政策制定者认为系统评估(systematic reviews)才能为政策评价提供最为准确、科学的依据^③,但相较传统的干预研究(intervention study),此类研究往往需要收集到足够数量的个体案例才能进行整体性评估,因此要花费较长时间。而据奥多诺休-詹金斯研究团队

^① 参见 John Snow, *On the Mode of Communication of Cholera*, 2nd edition, London: John Churchill, 1855, p.75。

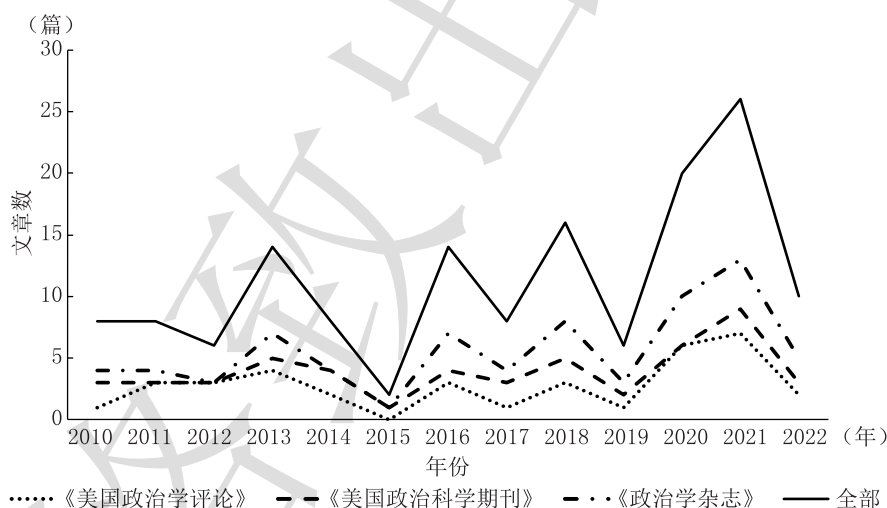
^② Scott T. Leatherdale, “Natural Experiment Methodology for Research: A Review of How Different Methods can Support Real-World Research”, *International Journal of Social Research Methodology*, Vol.22, no.1(2019), pp.19—35。

^③ Richard W. Puddy and Natalie Wilkins, *Understanding Evidence Part 1: Best Available Research Evidence. A Guide to The Continuum of Evidence of Effectiveness*, Atlanta, GA: Centers for Disease Control and Prevention, 2011。

报告,95%的政策制定者声称,他们在为政策制定寻找依据时重视的是实时性——例如其是否能够证明当前在其他领域所执行的类似政策的有效性——而不是该依据的质量或科学价值。^①因此,通过自然实验对政策效果做出的及时评价,能够为发展战略的制定提供可靠依据。

三、自然实验在政治科学领域的具体应用

在本节中,我们对自2010年以来发表在政治学三大期刊(《美国政治学评论》《美国政治科学期刊》及《政治学杂志》)上的73篇自然实验研究实例进行了梳理和归纳(图2),并对具有代表性的研究实例进行介绍。这些研究根据其主题主要可以分为:(1)选民及竞选相关;(2)民众的政治态度与政治参与相关;(3)政策效能评价相关;(4)执政、执法者行为相关;(5)民族/种族歧视问题相关;(6)冲突与暴力相关;(7)其他(图3)。除此之外,还有一些文献针对自然实验研究进行了方法论上的探讨与反思,我们将在下一节中对这些文献进行介绍。



注:该图展示了从2010年1月起至2022年3月收录于政治学三大期刊中、标题或是摘要中包含“natural experiment”是“quasi-experiment”的研究文献数量。

图2 2010—2022年在政治学三大期刊上发表的自然实验研究文献数量

^① Lily O'Donoghue-Jenkins, Paul M. Kelly, Nicolas Cherbuin and Kaarin J. Anstey, “Evaluating and Using Observational Evidence: The Contrasting Views of Policy Makers and Epidemiologists”, *Frontiers in Public Health*, Dec.6 2016, doi: 10.3389/fpubh.2016.00267.

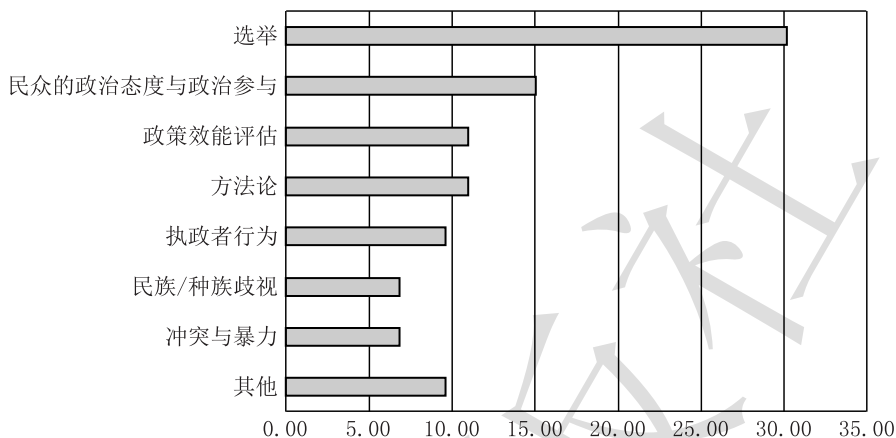


图3 2010—2022年在政治学三大期刊上发表的自然实验研究中各研究主题分布

(一) 探讨选民及竞选的自然实验

如何提高投票率和选民支持度是西方政治学领域中的永恒主题之一。事实上,政治学领域的实验研究也正是起源于20世纪40—50年代的竞选动员研究。^①而这一传统也体现在自然实验研究之中——根据我们的统计,以选民及竞选为研究主题的自然实验研究数量占比最多,超过了文献总数的三分之一。此类研究主要分为两个方向:(1)探讨影响投票率的因素;(2)探讨对选举结果产生影响的因素。

1. 影响投票率的因素

与心理学不同,自然实验无法通过控制外在变量来探究人们的内在动机。因此,自然实验研究主要探讨了自然环境、家庭社会环境等客观因素对投票率可能产生的影响。布雷迪和麦克纳尔蒂的研究以同时出现在2002年、2003年的投票名单上的300万名洛杉矶选民为对象,考察了投票点的设置对投票率的影响。^②他们将投票地点连续两年没有发生变动的选民作为对照组,由于选区合并而导致投票地点发生了变动的选民作为干预组(由于洛杉矶市全市都存在选区合并的情况,因此投票点发生变动这一干预是随机的),研究结果表明,相比

① 孟天广:《从因果效应到因果机制:实验政治学的中国路径》,载《探索》2017年第5期。

② Henry E. Brady and Jonh E. McNuty, "Turning out to Vote: The Costs of Finding and Getting to the Polling Place", *American Political Science Review*, Vol.105, no.1(2011), pp.115—134.

对照组,干预组的投票率下降了1.85个百分点,研究者认为这一降低源于投票成本的上升——干预组的选民需要花费更多的时间和精力去寻找新的投票点。同样,亨德森和布鲁克斯的研究也再次验证了这一观点:他们将1956年至2008年美国众议院选举期间的降雨作为工具变量实施了自然实验,发现每增加1英寸的降雨量,就会导致民主党的得票降低1.4—2.0个百分点。^①这些研究结果提醒了研究者不应忽视这些看似无关的外在因素可能会为选举格局带来的改变。

除了环境因素的影响以外,家庭社会环境因素也会对投票率产生影响。先天基因传承与后天家庭教育,究竟哪一个对人们的政治态度以及政治参与行为影响更大?切萨里尼等学者的自然实验研究试图对这一问题进行解答。^②他们将被收养者的亲生兄弟姐妹以及其收养家庭中的兄弟姐妹作为对照组,被收养者作为干预组,对他们成人后的投票行为进行了比较。结果表明,先天基因传承与后天家庭教育对于人们的影响同样显著:当被收养人的亲生父母履行过投票义务时,被收养人参与投票的概率更高(被收养人与其亲生父母不曾有过联系);另一方面,当收养家庭中的父母至少有一人受过大学教育时,被收养人的投票率会增加4.3个百分点,且收养家庭中母亲的政治态度与行为对被收养人的影响更大。

2. 影响选举结果的因素

既有的自然实验研究还分别从选民、候选人(或政党)以及选举制度这三个角度出发,对可能左右选举结果的因素进行了分析与探究。

首先,研究表明选民的生活经历会影响其投票倾向。马歇尔的研究探讨了选民受教育年限与政党支持的关联。^③1947年,英国在离校年龄上做出了重大制度改革,这一改革使得近半数人口的受教育年限增加了至少一年。而通过模

① John Henderson and John Brooks, "Mediating the Electoral Connection: the Information Effects of Voter Signals on Legislative Behavior", *The Journal of Politics*, Vol. 78, no. 3 (2016), pp. 653—669.

② David Cesarini, Magnus Johannesson and Sven Oskarsson, "Pre-Birth Factors, Post-Birth Factors, and Voting: Evidence from Swedish Adoption Data", *American Political Science Review*, Vol. 108, no. 1 (2014), pp. 71—87.

③ John Marshall, "Education and Voting Conservative: Evidence from a Major Schooling Reform in Great Britain", *The Journal of Politics*, Vol. 78, no. 2 (2016), pp. 382—395.

糊断点回归设计,研究者发现人们在高中就读的时间越长,其日后支持保守党的可能性就越大。其背后的主要原因在于受教育程度与工作收入成正比,而工作收入越高的人,越会倾向于保守主义经济政策,最终转变为对保守党的支持。

一些重大的社会事件也会对选民政党支持的偏好产生影响。2016年,历时15小时的英国脱欧公投出现了令人意想不到的结果——英国以51.89%对48.11%的微弱多数票决定其将脱离欧盟。基于这一社会民意,执政的保守党也不得不将其政策风向在一夜之间从“亲欧”转变为“疑欧”。借由这一事件,舍恩费尔德和温特-利维探讨了选民究竟是忠于政策还是忠于其党派身份。^①利用中断时间序列设计(interrupted time series designs),他们比较了选民在脱欧公投前后几个时间点对于保守党的支持度。结果显示选民更受政策,而非其党派身份驱动:“疑欧”的选民们纷纷涌向转变为“疑欧”的保守党,而“亲欧”的保守党支持者则选择离开政党。据统计,保守党在脱欧问题上突然的态度转变导致其失去了8%的支持者,毫无疑问,这将对今后的选举结果产生重大影响。此外,格拉西亚-蒙托亚等学者的研究则发现,在美国总统选举前经历过校园枪击案的县,支持严格控枪的民主党得票率平均增加近5个百分点。^②以上的自然实验研究充分说明,难以预料的偶发事件都可能对选民的政治行为产生一定的影响。

其次,具备何种特征的候选人或政党更容易赢得选举也是研究者关心的话题。例如,福元健太郎和三轮洋文的自然实验研究别出心裁地探讨了日本参议院选举时,议员候选人的姓氏与其得票数的关系。^③参议院选举时,选民一人两票,即除了投给地方选区议员一票以外,还需要投一票给全国选区议员。在日本,地方选区的议员候选人常常通过在其选区用车载喇叭循环播放自己姓氏的方式来进行“拉票”,以强化自己在选民心中的印象。而它的影响一直延续到了全国选区议员的选举上——同一位全国选区的议员候选人在其姓名识别度高

① Bryan Schonfeld and Sam Winter-Levy, “Policy or Partisanship in the United Kingdom? Quasi-Experiment Evidence from Brexit”, *The Journal of Politics*, Vol.83, no.4(2021), pp.1450—1461.

② Laura Garcia-Nontoya, Ana Arjona and Matthew Lacombe, “Violence and Voting in the United States: How School Shootings Affect Elections”, *American Political Science Review*, Vol.116, no.3(2021), pp.807—826, <https://doi.org/10.1017/S0003055421001179>.

③ Fukumoto Kentaro and Miwa Hirofumi, “Share the Name, Share the Vote: A Natural Experiment of Name Recognition”, *The Journal of Politics*, Vol.80, no.2(2018), pp.726—730.

的地区(即该地区有与该议员姓氏相同的地区候选人)的得票数与在其姓名识别度低的地区(即该地区没有与其姓氏相同的地区候选人)相比,整整高出了69个百分点。基于这一结果,研究者指出强化选民对候选人姓氏的认知不失为一个有效的竞选策略。而洛温等学者的研究则发现政党想要在选举中获胜还需要一些好运气。^①加拿大的平民院(the Canadian House of Commons)通过抽签的方法从非内阁大臣中随机挑选人员对其议案进行审议。研究结果表明,与没有机会将自己的议案提交审议的人相比,有幸赢得机会的非内阁大臣更容易因为提出议案这一行为而受到选民青睐,从而增加其所属政党在选举中的获胜率。

候选人或政党在临近选举时往往会出台各类政策以拉拢选民。墨托利尼亚于2012—2018年在墨西哥进行的、以连任可能性不同的各组议员为对象的准实验研究则发现,为了迎合选民以增加连任机会,议员在选举临近前会增加与“特殊主义的立法”(particularistic legislation)相关的发言,即强调自身将立法以照顾少数民众的特殊需求(如基础建设、医疗需求等),且后续追踪调查也证实了他们在一定程度上履行了自己的承诺。^②但是,从另一方面来看,在一些发展高度不均衡的国家,执政政党为了连任而实施增加定向支出以赢得大批低收入选民的支持这一竞选政策,从长远来说,是低效且不利于国家全体发展的。阿马特和贝拉门蒂利用巴西联邦政府随机进行的反腐败审计进行了准实验研究,发现在大选前夕进行的审计能够有效遏制其地方政府的资金滥用,但相比未进行审计的地区,这些地区的选民投票率以及现任者的连任可能性都显著降低。^③

最后,还有一些研究聚焦于选举方式对选举结果的影响:在选举规则完全相同的情况下,投票方式也可能会影响选票结构和选举结果。^④一项在澳大利亚

① Peter J. Loewen, Royce Koop, Jaime Settle and James H. Fowler, "A Natural Experiment in Proposal Power and Electoral Success", *American Journal of Political Science*, Vol.58, no.1(2014), pp.189—196.

② Lucia Motolinia, "Electoral Accountability and Particularly Legislation: Evidence from an Electoral Reform in Mexico", *American Political Science Review*, Vol.115, no.1(2021), pp.97—113.

③ Francesc Amat and Pablo Beramendi, "Democracy under High Inequality: Capacity, Spending, and Participation", *The Journal of Politics*, Vol.82, no.3(2019), pp.1—40.

④ Tiffany D. Barnes, Carolina Tchintian and Santiago Alles, "Assessing Ballot Structure and Split Ticket Voting: Evidence from a Quasi-Experiment", *The Journal of Politics*, Vol.79, no.2(2017), pp.439—456.

实施的自然实验研究结果表明,相比使用传统纸质邮寄投票的选区,在导入了更为简洁便利的电子投票系统的选区出现分裂投票(ballot splitting)的概率更高,因此其通过选举产生的立法机构的构成也更具有包容性。此外,辛格的研究表明,在实行强制投票制(compulsory voting,是一项旨在命令有选举权的公民必须参加投票的法律或者规范制度)的国家,参选政党将减少贿选等行为,将拉票重心转向基于选民意向的政策制定。^①

(二) 探讨民众的政治态度及政治参与的自然实验

政治态度指民众对政治目标和政治情景的认知、情感和评价。^②个体的政治态度往往决定了其政治参与的意愿。而政治参与是维持和促进政府合法性的有效手段,也是人们表达意见和维护自身权益的重要途径。因此,把握民众政治态度、探究影响其政治参与行为的因素也是政治科学领域中的一个重要议题。

1. 影响政治态度的因素

人们的政治态度早在童年时期就开始逐步形成。既有研究已经证明,个人的政治认同与政治态度继承于父母,且其影响可以持续40年之久。^③在此基础上,希利和马尔霍特拉则通过自然实验,进一步探究了兄弟姐妹对美国人的性别角色以及政治态度的影响。^④他们认为,拥有姐妹的年轻男性会更偏向于传统的性别角色观念(因为他们往往不会被分配做洗碗、收拾房间等家务),从而逐步树立保守的政治态度。为了验证这一假设,他们将有妹妹的男性(干预组)和有弟弟的男性(对照组)的性别角色观念与政治态度进行比较——虽然父母是否再要一个孩子的决定并不是随机的,但下一个孩子的性别是完全随机的,因

① Shane P. Singh, "Compulsory Voting and Parties' Vote-Seeking Strategies", *American Journal of Political Science*, Vol.63, no.1(2019), pp.37—52.

② Gordon W. Allport, *A Handbook of Social Psychology*, Worcester, Mass.: Clark University, 1935, pp.798—844.

③ M. Kent Jennings, Laura Stoker and Jake Bowers, "Politics across Generations: Family Transmission Reexamined", *The Journal of Politics*, Vol.71, no.3(2009), pp.782—799.

④ Andrew Healy and Neil Malhotra, "Childhood Socialization and Political Attitudes: Evidence from a Natural Experiment", *The Journal of Politics*, Vol.75, no.4(2013), pp.1023—1037.

此该研究将“下一个孩子的性别”作为干预变量进行了实验分组。结果验证了研究者的假设:有妹妹的年轻男性有着更为传统的性别角色观念,也更倾向于支持共和党。而是否拥有妹妹或弟弟则不会对年轻女性的政治态度产生影响。除了家庭环境之外,植根于大学中的阶级文化规范(class culture norm)也会对处于社会化过程中的大学生的政治态度产生影响。^①相比身处其他环境中的同龄人,出生富裕家庭的大学生在大学校园中享受过几年富裕无忧的生活后,会更倾向于支持保守主义经济政策。

如若没有获得民众的信赖和认可,政府提出的有益政策可能也会遭到反对。例如在缺少有效沟通的情况下,当地社区居民会以妨碍发展为由,反对风力涡轮机等节能基建,从而导致政府支持率的下降。^②因此,如何增加民众对政治体制、政府机构、政府绩效等的认同和支持也是一个极具应用价值的研究课题。库克等学者将已收到过美国政府下发的《社会保障声明》(Statements of Social Security)的居民作为干预组,未曾收到该文件的居民作为对照组,比较了两组居民的对于社会保障制度的相关知识以及对该制度的信心(例如:你在多大程度上认可该制度能够为退休工人或因公死亡人员家庭的生活带来保障?)。^③结果表明,相比对照组,干预组的居民获得了更多关于社会保障制度的知识,且对该制度的实施表现出更高的信心。这说明政府向民众传达信息的质量和数量会正向影响民众对政府机构的评价。而弗赖伊和鲍里索娃则发现,尽管选举本身对俄罗斯民众的政府信任度没有显著影响,但选举后发生的大规模抗议活动反而增加了反对派选民对于政府的信任——政府认可反对派进行游行这一决定本身是使得他们态度发生转变的关键。^④因此,研究者认为“专制”

① Tali Mendelberg, Katherine T. McCabe and Adam Thal, “College Socialization and the Economic Views of Affluent Americans”, *American Journal of Political Science*, Vol.61, no.3 (2016), pp.606—623.

② Leah C. Stokes, “Electoral Backlash against Climate Policy: A Natural Experiment on Retrospective Voting and Local Resistance to Public Policy”, *American Journal of Political Science*, Vol. 60, no.4 (2016), pp.958—974.

③ Fay L. Cook, Lawrence R. Jacobs and Dukhong Kim, “Trusting What You Know: Information, Knowledge, and Confidence in Social Security”, *The Journal of Politics*, Vol.72, no.2 (2010), pp.397—412.

④ Timothy Frye and Ekaterina Borisova, “Elections, Protest, and Trust in Government: A Natural Experiment from Russia”, *The Journal of Politics*, Vol.81, no.3 (2019), doi.org/10.1086/702944.

政府也许可以通过允许抗议活动以增加民众对政府的信任。此外,行政辖区的面积大小会也直接影响公民的内部政治效能感(internal political efficacy):行政辖区面积越大,公民会认为自己能够影响政治决策的机会越小,从而导致内部政治效能感的降低。^①

另一方面,国内所发生的政治事件还可能会对该国的国际形象产生影响。以四个拉丁美洲国家的民众为对象,在美国大选前后实施的准实验研究结果显示,在特朗普出人意料地当选美国总统后,国外民众对于美国政府的信任度骤降。^②

2. 影响政治参与的因素

通过政治参与,公民能够直接或间接影响政治决策制定与所执行的政治行为。因此,政治参与也是公民的政治权利得以实现的重要方式。但既有研究证实,在政治参与中存在着显著的性别鸿沟。即便在美国这样公民权利与政治权利高度强调的国家,女性对政治生活也显示出较小的兴趣、参与度以及知识。^③而在父权制文化的国家,例如印度,女性更是被以男性为中心的价值标准所主导和制约,基本政治权利无法得到保障。普里拉曼认为,女性社交网络的扩大能够有效降低女性政治行为的成本,从而促进政治参与。^④为了验证这一假设,他们将非政府组织 PRADAN 的妇女自助公益会(Self-Help Groups, SHGs)看作干预措施,在印度中央邦的农村地区开展了自然实验研究。妇女自助公益会组织同村 10—20 名女性为一个小组,通过每周一次的会议向她们提供非正式的储蓄、借贷等金融服务,有时也会教授她们相关的农业知识。该研究将这些开展

① David D. Lassen and Søren Serritzlew, "Jurisdiction Size and Local Democracy: Evidence on Internal Political Efficacy from Large-Scale Municipal Reform", *American Political Science Review*, Vol. 105, no.2 (2011), pp.238—258.

② Regina Bateson and Michael Weintraub, "The 2016 Election and America's Standing Abroad: Quasi-Experimental Evidence of a Trump Effect", *The Journal of Politics*, (2021), <http://doi.org/10.1086/718209>.

③ Sidney Verba, Nancy Burns and Key L. Schlozman, "Knowing and Caring about Politics: Gender and Political Engagement", *The Journal of Politics*, Vol.59, no.4 (1997), pp.1051—1072.

④ Soledad A. Prillaman, "Strength in Numbers: How Women's Groups Close India's Political Gender Gap", *American Journal of Political Science*, https://www.povertyactionlab.org/sites/default/files/research-paper/StrengtheningNumbers_Oct2017.pdf.

妇女自助公益活动的村子作为干预组,并根据这些村子的人口、女性比例等特性寻找与其地理位置最接近的村子匹配为对照组,将两组女性的参选行为、意愿等政治参与行为进行了比较。结果表明,参与妇女自助公益活动能够通过扩展社交网络、提高集体行动能力以及培养公民技能这三方面来增加女性的社会参与。基于这一研究结果,研究者指出,尽管在西方及高收入国家,多数研究聚焦如何通过给予女性如工作、金钱等社会资本,来增加其政治话语权和政治资本,但在中低收入国家——尤其是那些女性在父权制文化影响下被社会孤立的国家,相比增加社会资本,重建她们与社会的联系可能会更有效地促进女性政治参与的行为和意愿。

(三) 探讨执政、执法者行为的自然实验

除了上述以民众为对象的自然实验研究以外,还有一些研究将焦点置于执政者、执法者,旨在找出可能会对执政、执法行为产生影响的因素。例如,美国各个县的治安官由选举产生,因此人们担心他们执法时会迎合所属政党的政治倾向,而不是将维护该地区的安全放在首位。为此,汤普森以治安官为研究对象,进行了一项自然实验研究,证明了治安官的党派并不会对其执法行为产生影响:无论其属于推动移民政策改革的民主党还是属于反对移民的共和党,治安官在面对移民海关执法局提出扣押非法移民的要求时,其服从率并没有显著差别。^①虽然执法者的政治倾向并没有对其执法行为产生影响,但格林和森以美国上诉法院的法官为对象的准实验研究则发现,执法者的人际关系与生活经验——例如是否有女儿——会对案件判决产生影响:有女儿的法官只有在处理与女性议题有关的案件时(例如针对女性的就业歧视、堕胎行为等),会倾向于作出更为开明的判决(liberal decision)。^②其原因可能在于家有女儿会促使法官在日常生活中加深对于女性议题的关心和思考。

① Daniel M. Thompson, "How Partisan is Local Law Enforcement? Evidence from Sheriff Cooperation with Immigration Authorities", *American Political Science Review*, Vol. 114, no. 1 (2019), pp. 222—236.

② Adam N. Glynn and Maya Sen, "Identifying Judicial Empathy: Does Having Daughters Cause Judge to Rule for Women's Issues?" *American Journal of Political Science*, Vol. 59, no. 1 (2015), pp. 37—54.

近年来,政坛女性的崛起也已经成为一种全球性趋势。为了回应女性的政治参与需求,争取女性选民以及妇女组织的支持,西方政党在党内生活中逐渐重视女性,在议员候选人选拔领域实行性别配额制(gender quota,指选举中当选的女性数量需要占到一定的比例),旨在推进男女平等、发展党内民主。但是,有一部分研究者认为,女性即便依靠配额制快速地成为描述性代表(descriptive representation),基于性别的歧视和质疑也可能会对她们的职业发展前景产生负面影响:女性可能难以连任或是成为党内领导。奥布莱恩和里克内在瑞典进行的自然实验研究使用源于现实生活的数据打消了这一疑虑。^①通过检验性别配额与女性政治领导力的因果关系,他们发现性别配额制度不但能够推动更多高素质女性担任公职,还能帮助女性快速升职为党内领导,对其连任率也没有产生消极影响。该研究确认了配额制的实施对于推动女性参政的积极影响。但是另一方面,通过对比英国首相特蕾莎·梅在2017年曼彻斯特恐怖袭击事件前后的民众支持率,霍尔曼等学者发现,遭遇恐怖袭击会降低民众——尤其是对女性抱有负面印象的人——对女性首相的支持和信心。^②因此,虽然女性参政率已经明显提高,但要在政治领域达到真正的男女平等还有很长的一段路要走。

(四) 探讨公共政策效能的自然实验

自然实验研究在评价公共政策效能方面发挥了积极作用。卡内基和纳里诺夫利用欧盟理事会轮值主席制度,检验了国际援助对于帮助受援助国健全民主制度、提高人权保障水平的效果。^③在对各轮值主席国在国际援助上的拨款进行统计后,研究者发现当曾拥有殖民地的国家当值时,它们倾向于为自己曾经

① Diana Z. O'Brien and Johanna Rickne, "Gender Quotas and Women's Political Leadership", *American Political Science Review*, Vol.110, no.1(2016), pp.112—126.

② Mirya R. Holman, Jennifer L. Merolla and Elizabeth J. Zechmeister, "The Curious Case of Theresa May and the Public that did not Rally: Gendered Reactions to Terrorist Attacks can Cause Slumps not Bumps", *American Political Science Review*, Vol.116, no.1(2022), pp.249—264.

③ Allison Carnegie and Nikolay Marinov, "Foreign Aid, Human Rights, and Democracy Promotion: Evidence from a Natural Experiment", *American Journal of Political Science*, Vol.61, no.3(2017), pp.671—683.

的殖民地国提供更多的经济援助:例如 2001 年比利时成为轮值主席时,欧盟为非洲,尤其是刚果民主共和国,提供了高额援助;2002 年作为轮值主席的西班牙则为其在拉丁美洲的前殖民地提供了特别援助。而收到援助的国家会在短时间内开展一系列针对民主和人权的改革,但由于援助是暂时性的,因此改革效果并不会持久。基于此,研究者认为如果欧盟能够为受援助国提供长期且稳固的援助计划,那么其在人权和民主方面就能获得持久改善。

由于长期存在的系统性种族主义,种族不平等问题早已渗入美国社会生活的方方面面。虽然美国政府为了促进种族融合和制定了一系列的干预政策,如取消“种族隔离学校”,提高非洲裔与拉丁裔学生的受教育机会,以实现教育资源的公平配置。但基钦斯在对比了 1995 年至 2011 年取消种族隔离令对美国近 12 000 个学区中公立中小学教育资源分配所产生的影响后发现,这一命令会促使公立学校 6%—12% 的白人孩子退学——父母选择将他们送去种族隔离程度较高的私立学校受教。^①这则会减少社区居民将税收用于提升公立中小学教育质量的意愿,从而导致师资力量短缺,学生的成绩下滑。由此可见,取消种族隔离对于美国社区的公共教育资源分配并没有达到预计的效果。基于这一结果,研究者建议采用适当的种族隔离教育政策,或是由联邦政府直接拨款以支持种族融合的公立中小学校以求改善当前局面。

(五) 探讨民族/种族歧视的自然实验

民族/种族歧视俨然成了一个全球性的问题。如何预防和干预民族冲突、实现种族和平融合已经成为政治发展研究中的一个核心问题。除了不同文化集团之间的冲突以外,同一文化集团内部也会因为其社会阶层的划分而导致各阶层之间的矛盾与不平等——以印度为代表的种姓制度就是一个典型的例子。既有的自然实验研究主要聚焦于移民、难民以及表列种姓(scheduled castes)这三类群体,深入探讨了抑制或加剧民族/种姓歧视的社会因素。

值得一提的是,通过实验室实验或是问卷调查的方法探究人们对待移民、

^① Karin Kitchens, “Exit or Invest: Segregation Increases Investment in Public Schools”, *The Journal of Politics*, Vol.83, no.1(2021), doi.org/10.1086/708916.

难民或是弱势群体态度的研究并不在少数。^①但由于这是一个比较敏感的话题,人们基于道德伦理和社会规范的约束或是感知自己正在被实验员所观察的情况下,可能会隐藏心中真实所想,对自己的回答或是行为进行美化。而完全基于真实世界的自然实验研究则能解决这一难题,帮助研究者们掌握人们真实的行为或想法。

1. 聚焦移民群体的研究

近年来,许多国家内部在移民政策这一议题上产生分歧,而其中最具争论性的问题在于移民的归化入籍问题。尽管在一些发达国家,移民人口已经占据总人口的10%,不少政客依旧反对移民入籍。他们认为移民会对本国人的工作和收入产生威胁,并破坏当地的传统价值观,因此这些“外来者”不应享受与本国同等的政治和社会权利。

为了更好地理解人们为什么排斥移民、会排斥什么类型的移民,以及归化入籍对移民融入当地社会的影响,海因米勒和汉加特纳利用瑞士1970年至2003年间独特的移民入籍制度进行了自然实验研究。^②2003年以前,瑞士的一些城市采取“投票箱”(ballot box)制度,通过市民投票来自主决定是否同意移民归化入籍。具体来说,市民会收到由市政局寄来的匿名投票信和一份记载了所有申请入籍的移民信息名单,上面详细记载了移民的性别、年龄、出生地、移民历史以及语言能力等信息。依据这些信息,市民对每一位申请人做出同意与否的判断后将选票寄回。通过票数统计,获得多数认可的移民将获得入籍资格。他们通过断点回归设计,以是否获得半数以上否定票为基准,将申请人分为了两组进行了比较。结果表明,与来自较为富裕的北欧及西欧国家的申请人相

① 参见 Colleen Ward and Anne-Marie Masgoret, "Attitudes towards Immigrants, Immigration, and Multiculturalism in New Zealand: A Social Psychological Analysis", *International Migration Review*, Vol.42, no.1 (2008), pp.227—248; Cícero Pereira, Jorge Vala and Rui Costa-Lopes, "From Prejudice to Discrimination: The Legitimizing Role of Perceived Threat in Discrimination Against Immigrants", *European Journal of Social Psychology*, Vol.40, no.7 (2010), pp.1231—1250.

② Jens Hainmuller and Dominik Hangartner, "Who Gets a Swiss Passport? A Natural Experiment in Immigrant Discrimination", *American Political Science Review*, Vol.107, no.1 (2013), pp.159—187; Jens Hainmuller, Dominik Hangartner and Giuseppe Pietrantuono, "Catalyst or Crown: Does Naturalization Promote the Long-Term Social Integration of Immigrants?" *American Political Science Review*, Vol.111, no.2 (2017), pp.256—276.

比,来自原南斯拉夫和土耳其的申请人获得否定票的概率要高出近 40% 左右。同时,研究还发现,该市的排外性越高,其市民对移民的原籍歧视越严重。基于这一结果,研究者认为,在决定移民入籍的问题上,需要依据的应该是相对公正的法律,而非市民的判断。而在之后的研究中,他们对归化入籍如何影响移民的社会融合这一问题做了进一步的探讨。结果表明,获得瑞士公民身份对于帮助移民更好融入当地社会有着长期影响。因此,归化不仅仅是表彰移民达到入籍标准的“奖励品”,还应当被视为促进社会融合的重要催化剂。

2. 聚焦难民群体的研究

相较于移民群体,难民群体的处境显得更为艰难。2015 年以来,来自叙利亚、阿富汗以及伊朗等穆斯林国家的超 300 万难民向欧洲申请了政治庇护。虽然自 2016 年欧盟与土耳其达成难民协议之后,申请人数略有下降,但试图通过申请政治庇护以逃离战争与迫害的人数始终处于高位。尽管大多数情况下难民的滞留时间不会超过 24 小时,不会对当地人的工作或是文化带来威胁,但许多人依然认为难民“搅乱”了他们的生活——例如为难民提供医药和公共卫生服务会使当地医疗系统产生混乱,甚至还可能爆发传染病。在这样的背景下,汉加特纳等学者在被认为是受难民危机影响最为严重的欧洲国家希腊进行了自然实验研究,探讨了难民危机的爆发如何影响希腊人对于难民群体以及政治庇护政策的态度。^①

为了排除与难民危机无关因素的影响(如当地人的排外性以及对移民政策的原有态度偏好等),该研究选取了希腊各岛屿至土耳其的距离这一客观地理因素作为工具变量:离土耳其越近的希腊岛屿,登陆的难民数量越多。与此同时,由于难民选择登陆地点的时候还受到地形、可行性等因素的限制,因此也会出现与土耳其距离相近的两个岛屿接收难民数量不同的情况。这也为锁定难民危机带来的影响提供了合适的研究条件。研究结果显示,相较于与难民直接接触较少的岛屿,难民登陆数量多的岛屿上的居民都表现出了更高的反难民情绪。而这种由难民引发的对外排斥性态度还会波及其他少数群体,如经济移民

^① Dominik Hangartner, Elias Dinas, Moritz Marbach, Konstantinos Matakos and Dimitrios Xeriferis, “Does Exposure to the Refugee Crisis Make Natives More Hostile?” *American Political Science Review*, Vol.113, no.2(2018), pp.442—455.

以及穆斯林群体。

如上述研究结果所示,即便在难民不造成任何威胁的情况下,仅仅是与他们“打照面”就足以提升当地民众的敌意。这使研究者有足够理由推测,当难民被认为会对国家安全产生威胁的时候,将会引发当地民众更大的恐慌与不满。布莱顿等人的准实验研究对这一假设进行了验证:他们比较了2015年11·13巴黎恐怖袭击事件前后一段时间内,加拿大民众对叙利亚难民的情感以及态度变化。^①研究发现,尽管在恐怖袭击发生的24小时内,人们对叙利亚难民可能带来的安全威胁表示担忧,且对移民安置政策表现出反对倾向,但经过15—20天以后,这些担忧与反对逐渐回落至恐怖袭击发生前的水平。这一源于现实生活的研究结果出乎了研究者的意料:这一起近十年来在西方发生的最大规模恐怖袭击事件并没有像政治学者和评论家们所预计的那样,对人们产生更为深刻且长远的影响。

3. 聚焦表列种姓群体的研究

印度是一个种姓制度森严的社会,这种源于印度教的制度将人由高至低分为四个不同等级:婆罗门、刹帝利、吠舍和首陀罗。除了四大种姓之外,还有一种被排除在种姓之外的人,即“不可接触者”或“贱民”,在官方文书中则被称为“表列种姓”。他们是印度社会中地位最低、最受歧视的群体,且这一情况在农村地区尤为严重。

为了保护表列种姓群体的利益,印度政府也采取了很多措施。例如1992年政府通过的第73次修正案就明确规定在村级委员会(panchayati)的选举中,必须按村中表列种姓人口占比,为其在委员会中分配同样比率的席位。同时,这一修正案还规定了在每隔5年进行一次的村委会选举中,25%的村委会主席必须由表列种姓群体担任,而具体是哪些村子则是按照政府名单顺序由上往下重复滚动。由于村委会主席有着较大的两项权力:一是有权为中央和邦的公共产品项目选定受益人;二是有权指定需要建设和维护的村级公共产品,如路灯、道路等。因此研究者认为由表列种姓担任村委会主席能使村中其他的低种姓群

^① Charles Breton and Gregory Eady, “Does International Terrorism Affect Public Attitudes toward Refugees? Evidence from A Large-Scale Natural Experiment”, *The Journal of Politics*, Vol.84, no.1(2022), <http://doi.org/10.1086/714781>.

体分配到更多资源,从而在一定程度上改善种姓间的资源不平等问题。^①与此同时,研究者还期待这一制度能够促进表列种姓与其他种姓群体之间的交流,从而改善表列种姓在印度社会的刻板印象,减少歧视。^②

通过自然实验,研究者对表列种姓群体担任村委会主席这一制度的效果进行了检验。他们将每两个人口、种姓构成等基本条件相同的村庄配对为一组——其唯一的区别在于是否有表列种姓群体担任过村委会主席。通过进一步的对比,邓宁等人发现,虽然村委会主席在住房、就业和福利等资源分配上有较大的自由裁量权,但让表列种姓担任村委会主席并没有明显增加村中低种姓群体成员获得的资源。在之后的问卷调查中,他们找到了其背后的原因所在:相比基于种姓的资源分配,低种姓群体更依赖基于所属党派和资源分配。与农村激烈的种姓矛盾相比,党派内往往能够做到较为和谐的多种姓共存。与此同时,为了拉拢民众选票,多数政党都会采取对低种姓更为有利的分配政策。因此,研究者得出结论:在印度,资源与福利的分配主要依靠党派,而非种姓。在这样的背景下,让表列种姓担任村委会委员这一制度对于解决农村的种姓不平等问题似乎并非那么有效。

另一方面,肖沙尔则发现,在表列种姓担任过村委会主席的村子里,虽然高种姓群体对于表列种姓持有的消极刻板印象并没有改善,但他们的歧视行为和意图有所减轻。赋予表列种姓政治地位和权力可能会使高种姓群体的社会规范感知产生变化——社会并不鼓励歧视和欺凌低种姓群体的行为,从而导致了高种姓群体行为和态度的转变。基于这一研究结果,研究者认为,如果社会边缘群体能够通过获得政治权力而引发群际关系的积极变化,那么无论该制度的再分配效应和政治影响如何,它对社会边缘群体的成员来说都是有益的。

① Thad Dunning and Janhavi Nilekani, "Ethnic Quotas and Political Mobilization: Caste, Parties, and Distribution in Indian Village Councils", *American Political Science Review*, Vol.107, no.1 (2013), pp.35—56.

② Simon Chauchard, "Can Descriptive Representation Change Beliefs about a Stigmatized Group? Evidence from Rural India", *American Political Science Review*, Vol.108, no.2 (2014), pp.403—422.

(六) 探讨冲突与暴力的自然实验

根据联合国统计,虽然自1946年以来因战争死亡的人数一直在下降,但全球范围的冲突和暴力在不断上升。尚未解决的区域紧张局势、法治崩溃、非法经济收益以及因气候变化而加剧的资源稀缺,已成为造成冲突的主要因素。而科技技术的进步也使机器人、无人机和网络袭击以及数据黑客得以武器化。新的和更复杂的安全威胁向政治科学研究提出了挑战。

近年来,冲突和暴力的性质已发生重大变化。冲突的杀伤力减少,但持续时间更长,而且更常发生在国内团体之间,而不是国家之间。布莱尔等人基于46项自然实验研究的元分析表明,农产品等劳动密集型产品的价格上涨,能够有效减少当地的武力冲突^①:例如自1993年至1998年,咖啡价格上涨190%,而在主要咖啡产地秘鲁与哥伦比亚,暴力袭击事件下降55%。其原因可能在于随着产品价格上涨,工人报酬也会增加,这会吸引大批工人投入生产,间接减少了冲突的发生。而另一方面,煤炭、汽油等资源密集型产品的价格上涨则会促进当地武力冲突的发生:1998年至2000年,原油价格增长115%,而同一期间,哥伦比亚产油城市发生的武装袭击增长16.5%——偷窃、抢占自然资源变得更为有利可图。

而政治不稳定也会导致有组织犯罪,包括对警察、妇女、移民和反对党支持者的有针对性的袭击。内利斯等人的自然实验研究聚焦于巴基斯坦,探讨了世俗政党执政对于宗教暴力的抑制作用。通过比较世俗政党在选举中以微弱优势获胜(干预组)或失败(对照组)的地区,研究者发现,世俗政党执政能够大幅减少当地宗教暴力事件的发生。^②进一步分析表明,来自核心选民的竞选压力促使世俗政党加大了平息宗教暴力的努力,且只有在其有足够的能力左右当地安全形势的情况下(例如警察数量较多),其努力才会有效。但这一研究结果也强调了改革的必要性:将平息宗教暴力作为竞选策略的一环,会使得安全部门行

^① Graeme Blair, Darin Christensen and Aaron Rudkin, “Do Commodity Price Shocks Cause Armed Conflict? A Meta-Analysis of Natural Experiments”, *American Political Science Review*, Vol. 115, no.2(2021), pp.709—716.

^② Gareth Nellis and Niloufer Siddiqui, “Secular Party Rule and Religious Violence in Pakistan”, *American Political Science Review*, Vol.112, no.1(2018), pp.49—67.

使职权时受到政治干扰,无法平等保障民众的人身安全。

(七) 其他

除去上述这些相对集中的研究主题以外,还有一些基于各国国情需要的自然实验研究实例。例如,达斯古普塔的研究探讨了印度的农业技术变革——“绿色革命”(the green revolution)的进行对于提升边缘化政党竞争力、推进民主进程所起到的作用。^①而艾哈迈德则通过自然实验研究验证了独裁国家会将外国援助以及海外务工人员汇款收入用于加固政府庇护结构,从而进一步稳固自身政权。^②另一方面,在政治极化趋势愈发严重的美国,列文达斯基提出假设,认为国民身份认同感的提升能够有效降低民众对于对立政党的厌恶与抵触情绪,从而缓和因政治极化而导致的社会分裂。^③为了验证这一假设,研究者在2008年美国总统选举年的4月至8月以近30 000名选民为对象开展了自然实验。结果表明,当美国独立日(7月4日)临近时,高涨的国民身份认同感会使得选民更倾向于将对立政党候选人看作同一集体中的成员,从而对其表现出更多积极情绪。但与此同时,研究者也承认,如何在日常生活中通过提升国民身份认同感以缓解政治极化趋势,依旧是一个有待解决的难题。

四、自然实验的局限性及其反思

相较于实验室实验与田野实验,自然实验往往能够更好地控制混淆因素的影响,实现可信和有效的因果推断,这能够为解释日益复杂化的政治现象和过程、指导政治实践起到重要作用。但是,自然实验的开展也必然有其局限性和不足之处。本节将对自然实验的基本研究设计及其作为政治科学研究法的合理性与局限性进行反思,希望能为今后自然实验在政治科学领域的实际运用提

^① Aditya Dasgupta, “Democratizing Effects of the Green Revolution in India”, *American Political Science Review*, Vol.112, no.4(2018), pp.918—938.

^② Faisal Z. Ahmed, “The Perils of Unearned Foreign Income: Aid, Remittances, and Government Survival”, *American Political Science Review*, Vol.106, no.1(2012), pp.146—165.

^③ Matthew S. Levendusky, “Americans, not Partisans: Can Priming American National Identity Reduce Affective Polarization?” *The Journal of Politics*, Vol.80, no.1(2018), pp.59—70.

供有益的思路。

（一）基本研究设计的局限性

尽管自然实验中所运用到的计量设计都已经发展得较为成熟,但是各类型的研究设计也都有其局限性以及不足之处。

1. 工具变量

工具变量不仅为定量分析提供了因果推断的重要武器,还是一种十分巧妙且具有趣味性的研究设计。但是,想要找到一个合格的工具变量非常之难,其主要原因就在于它要求研究者变换思维角度,跳出假设模型的框架,找到一个能够帮助解决的外生变量。因此,寻觅工具变量不但要求研究者拥有严谨的逻辑力和诠释力,还往往需要其具备一定的社会科学想象力和灵感。^①同时,前人曾经使用过的工具变量往往只适用于特定的研究情景,虽然具有一定的参考性,但是很难推广、照搬至其他研究中。

2. 双重差分

双重差分主要用于评估政策效能。而陈林等人在综合了实验理论与前人研究后,总结出一个合格的双重差分研究必须满足两类假设^②:(1)随机性假设,即需要保证每个样本有同等机会接受干预处理,且实验发生的时间应当是随机的。满足这一假设条件主要是为了应对政策内生性问题,也就是说,基于宏观社会和经济环境,政策的实施本身可能就是系统内部运行的结果;(2)同质性假设,即保证干预组与对照组样本在统计意义上是同质个体。当以上两条基本假设无法得到满足时,双重差分估计量将会产生偏差,从而难以反映出政策的真实效果。因此,双重差分法并不适用于所有的政策效能评估,研究者在使用这一研究设计前必须首先对其合理性进行甄别。

3. 倾向值匹配

倾向值匹配能够很好地解决多混淆变量时的控制问题,帮助研究者得出因

① 陈云松:《逻辑、想象和诠释:工具变量在社会科学因果推断中的应用》,载《社会学研究》2012年第6期。

② 陈林、伍海军:《国内双重差分法的研究现状与潜在问题》,载《数量经济技术经济研究》2015年第7期。

果性结论。但是它的缺陷在于不考虑变量之间的交互作用,只关注某个自变量的效果。因此,当我们运用倾向值匹配法对某一因果效应进行检验时,就已经默认:(1)该自变量是导致因变量变化的唯一单因;(2)各变量之间不存在互相影响的效果。^①虽然它能够将复杂的政治现象的发生归结至某一个重要的关键性因素,具有逻辑上的简洁之美,但这种单因解释却也成为了倾向值匹配遭受质疑的理由:例如拉金认为,政治过程之所以复杂并难以解释,不仅因为有太多影响结果的变量,还因为在各类条件的共同作用下才会导致某一个特定结果的产生。^②因此,不支持多因分析的倾向值匹配法无法实现政治现象的多变量分析和联合分析。

4. 断点回归设计

虽然断点回归被认为是自然实验中揭示因果效应最有效的一种研究设计,但在运用过程中需要注意以下几点:第一,如果被研究对象自身能够精确地操作决定干预的关键变量,那么断点回归方法便是无效的;第二,除了干预条件以外,其他协变量在临界点附近应当都是连续的,这样才能保证因变量在临界点处的间断是由于干预条件的介入所引起的^③;第三,被研究对象应当是同质或是近似同质的,即若将对照组的个体放置于干预组中,其应该与原本就在干预组的个体产生相同的效应。^④

综上所述,尽管自然实验的研究设计能够帮助研究者在自然产生的政治情景之中检验变量之间的因果机制,更好地把握影响人们的政治行为和态度的因素、评估政策冲击带来的效果,但我们也必须对各类研究设计的局限性有充分认识。因此,研究者在实际使用各类研究设计时必须进行深入的事前分析,建立充分且合理的识别策略以考察被研究对象的分布,或是政策实施是否满足自然实验的各项基本特征与假设。

① 蒋建忠:《国际关系因果推论中的匹配方法——原理、类型与操作》,载《国际关系研究》2017年第3期。

② 参见 Charles C. Ragin, *The Comparative Method: Moving beyond Qualitative and Quantitative Strategies*, Berkeley and Los Angeles: University of California Press, 1987, p.26。

③ David S. Lee and Thomas Lemieux, "Regression Discontinuity Designs in Economics", *Journal of Economic Literature*, Vol.48, no.2(2010), pp.281—335。

④ 罗胜:《断点回归设计:基本逻辑、方法、应用述评》,载《统计与决策》2016年第10期。

(二) 自然实验法的局限性及其反思

虽然自然实验的研究设计为研究者提供了有效的研究工具,但不可否认的是,并非我们所关心的所有因果关系都在现实世界中有对应的自然实验,也就是说,自然实验是“可遇而不可求”的。同时,想要从现实世界中寻找到自然实验,对研究者也有着较高的要求。斯诺验证污水为霍乱感染源的自然实验研究看似构思简洁精巧,但他也是在进行了大量前期分析、排除掉其他各类可能性之后,才最终将霍乱传播的罪魁祸首锁定到污水。^①由此可见,研究者不但需要拥有高度的敏锐性,能够在错综复杂的形势中捕捉到合适的研究机会,还必须具备完善的相关知识体系与统计技术,才能尽力以逼近随机的方式对被研究对象进行挑选与分组。因此,要想设计出一场合理且有效的自然实验实属不易。

然而,正是这种“不易”会导致人们,尤其是研究者对自然实验研究结果的“过信”。当研究者在自己的逻辑体系中建立起一套合理的识别策略后,难免会产生一种自负感,认为其所设计的自然实验能够在避免内生性问题的情况下推断因果关系,从而忽视很多细节。^②这可能导致研究者遗漏关键的解释变量或控制变量,进而对研究结果的准确性产生影响。除了遗漏变量的问题以外,萨米还指出,自然实验的研究者对由回归分析等统计方法所得出的结果有进行过分解读的倾向,这会夸大他们所关心的两个变量之间的因果效应。^③

要想理性看待自然实验的研究结果,我们就必须要承认自然实验在方法论上的局限性。自然实验的研究设计面临着一个普遍的问题——局部平均干预效应问题(local average treatment effect, LATE)^④:在双重差分和倾向值匹配中,被研究对象是由研究者根据一定的规则进行人为挑选和分组,具有一定的特殊性;断点回归衡量的也只是在临界值附近的局部效应,而非整体效应;工具变量

① Thad Dunning, “Improving Causal Inference: Strengths and Limitations of Natural Experiments”, *Political Research Quarterly*, Vol.61, no.2(2008), pp.282—293.

② 陈云松:《逻辑、想象和诠释:工具变量在社会科学因果推断中的应用》,载《社会学研究》2012年第6期。

③ Cyrus Samii, “Causal Empiricism in Quantitative Research”, *The Journal of Politics*, Vol.78, no.3(2016), pp.941—955.

④ Guido W. Imbens and Joshua D. Angrist, “Identification and Estimation of Local Average Treatment Effects”, *Econometrica*, Vol.62, no.2(1994).

对于被研究对象的影响分布是否均匀也无法通过统计方法来进行确认。基于上述,自然实验在很多情况下反映的仅仅是某一局部样本的分析结果,同时,由于自然实验中的干预变量必须满足非人为性的条件,研究者也无法通过大规模复制研究以检验其结果的普遍性。因此,自然实验的研究结果并不具有普遍性,也很难推广到进一步的研究中去。在理解了这一特性以后,我们才能更理性地看待自然实验所得出的结果,并对其作出合理的政治、社会学解释。

如果自然实验的研究结果不具备普遍性,那它还能否为现实世界提供有效的实践指导呢?首先,自然实验毫无疑问能够加深我们在特定区域内对因果机制以及其影响参数的理解,其对于该区域政治现象的实践指导性是不可否认的。其次,我们可以通过实验室实验、田野实验等,来对从自然实验的结果中得到的可能性或启发进行进一步的验证。虽然一次自然实验只能检验某一特定干预条件会导致的结果(effect of causes),但通过搭配其他实验研究方法进行系列性研究,不断累积研究成果,便能逐步完善导致某一现象产生的各类原因(causes of effects),从而归纳总结出更具普遍性的理论规律。^①

五、总 结

随着2021年诺贝尔经济学奖的揭晓,自然实验法再次引起社会科学界的关注。作为当今西方经济学界最被广泛接受的研究方法之一,在真实世界中,因果机制的自然实验对于经济学作出的贡献有目共睹。但除此之外,基于实验路径的研究设计能够帮助研究者在现实政治情景中更好地控制干扰因素的影响,实现可信且有效的因果推断,因此,自然实验对于政治科学界的贡献也不容小觑。基于此,本文从发展背景、方法、在政治科学领域中的应用、局限性等几个方面对自然实验研究法进行了详细的介绍。

通过回顾与总结,我们发现,在西方主流政治学界,自然实验研究已经被广泛运用于诸多政治现象及过程,诸如国家能力与治理、数字民主与互联网安全、

^① Rebecca B. Morton and Kenneth C. Williams, *Experimental Political Science and the Study of Causality: From Nature to the Lab*, Cambridge: Cambridge University Press, 2010.

恐怖主义与种族冲突等,通过对现实政治情景的定量科学化研究,探求中观和微观层面政治现象或过程背后变量或因素间的因果联系,能够为解决现实政治问题提供有益的参考。但尽管如此,在进行自然实验研究时,研究者也必须始终保持审慎的头脑,在把握自然实验局限性的基础上,清楚地说明识别策略以及研究假设,并理性地看待研究结果。这样才能使自然实验发挥出最大效用,为社会科学定量分析中的因果推断提供更强的说服力和更高的可信度。

近年来,中国政治学领域对于因果机制、经验性政治理论的需求日益迫切。聚焦于中国政治和治理研究的国内外学者也开始广泛采取实验方法对中国治理模式、国家社会关系、协商民主和网络政治进行原创性研究。而适用于理解复杂社会中的政治现象的自然实验有助于从实践上推进国家治理的演进,故而逐渐受到国内学者的重视。同时,改革开放以来的中国改革实践赋予了自然实验研究独特应用场景和丰富机遇。本文对西方政治学领域自然实验研究实例的梳理及分类,有助于国内学者了解自然实验的应用条件和研究范围,从而为推进自然实验在中国政治科学领域的研究发展提供参考。今后,应加强国内学术机构与组织、机构之间的合作机制,充分利用大数据等丰富网络资源和技术以及促进政治学与社会学、心理学等其他学科的融合交流等,期待中国学者能够在掌握自然实验设计方法的同时,为这一前沿研究方法贡献更多有独创性且贴近现实世界的研究实例。