

贫困、RCTs 及其争议

——2019 年诺贝尔经济学奖简评

罗必良

(华南农业大学 国家农业制度与发展研究院,广东 广州 510642)



摘要 2019 年诺贝尔经济学奖授给了巴纳吉(Banerjee)、迪弗洛(Duflo)和克雷默(Kremer)三位经济学家。文章介绍了获奖者对贫困与发展问题的细致研究及其贡献,简要说明了微观实证中的 RCTs 方法及其应用价值,并对由 RCTs 方法所引发的争议进行了梳理与评价。在此基础上强调:尽管众多经济学家已经对经济学的实验主义方法论反复做过批判,但应该承认,经济学诺奖的技术性或技巧性倾向越来越明显,与鼓励基础性的理论突破或原创性的知识贡献的初衷背道而驰。经济学的实证研究是重要的,RCTs 方法对于提升对现实问题的洞察力也是重要的,但并非是解决发展或贫困问题的“万用良药”。

关键词 诺贝尔经济学奖; 贫困; RCTs; 争议

中图分类号:F 061.3; C 33 **文献标识码:**A **文章编号:**1008-3456(2020)01-0001-08

DOI 编码:10.13300/j.cnki.hnwkxb.2020.01.001

瑞典皇家科学院将 2019 年度诺贝尔经济学奖授给了三位关注贫困问题研究的经济学家,即 MIT 的巴纳吉(Abhijit Banerjee)和迪弗洛(Esther Duflo),以及最早采用实验方法研究贫困问题的克雷默(Michael Kremer),以表彰他们为减轻全球贫困所采用的实验性方法而做出的贡献。因为他们所用的方法及其研究,大大提高了应对全球贫困的能力,他们基于实验的新方法改变了发展经济学,并使之重新成为一个生机勃勃的研究领域。尽管每年的经济学诺奖都会引发一些争论,但 2019 年的评奖结果一经公布,却引起了广泛的争议,从而为本届诺奖抹上了一丝淡雾而不似以往那样光彩夺目。

一、对贫困问题的细分研究

在发展经济学研究领域,最难解的课题是贫困问题。尽管学者们为此做出了不懈的努力,但理论与方法的“贫困”丝毫没有减轻。所以,将 2019 年诺奖授予对有关贫困问题的研究者,体现了对发展经济学的重视,毕竟世界 85% 的人口生活在发展中国家,也表达了评奖委员会对全球重大问题的关切。

发展经济学或发展中国家经济学,有时也被戏称为“穷国的经济学”,是专门研究发展中国家问题的一个经济学分支。严格来说,它并非标准的学科分支,而是经济学在发展中国家中的应用。发展经济学曾在 20 世纪 50—70 年代因为提出了各种雄心勃勃的发展战略而广受青睐,从中产生的数位诺贝尔奖得主,例如 1974 年获奖的缪尔达尔(Myrdal)、1979 年获奖的刘易斯(Lewis)和舒尔茨(Schultz),都是发展经济学家中的佼佼者。

可惜好景不长。由于发展经济学家们深受凯恩斯主义的影响,并根据哈罗德-多马模型、索洛模型以及两部门增长模型,认为只要通过政府干预、加大资本投入,并与本国丰富而廉价的劳动力资源

收稿日期:2019-10-20

基金项目:国家自然科学基金政策研究重点支持项目“农地确权的现实背景、政策目标及效果评价”(71742003);国家自然科学基金青年项目“确权、信任与农地流转契约选择”(71703041)。

作者简介:罗必良(1962-),男,教授,长江学者,博士;研究方向:农业经济与制度经济。

相结合,经济起飞就指日可待。由于缺乏微观经济学的学理基础,这些发展战略差不多均以失败而告终。一方面因为发展经济学家开出的“药方”,在众多发展中国家并没有收到他们预期的效果,甚至产生了负面效果;另一方面,由于发展经济学理论大多着眼于宏观层面,且其研究方法又主要集中于理论模型的推演,过于空泛又缺乏实际应用价值,所以有人一度宣告发展经济学已经“死亡”。

在经历一段时间的“萧条”之后,发展经济学开始着眼于实证研究,但却依然是宏观层面的。一方面,即便是实证研究,大多也专注于宏观框架、基本制度等方面的逻辑检验,对于穷人的行为理解太少;另一方面,尽管众多研究基于庞大的数据处理,对比不同国家与地区的经济发展,试图发现影响经济发展及其国别差异的关键因素,但由于这类研究过于一般化,依然难以为缓解贫困问题提供“对症下药”的有效方案与手段。正是在这种背景下,寻找发展经济学研究的新路径,尤其是通过排除外部干扰,明确识别政策与行为结果关联性的研究范式,就具有重要的理论与实践价值。以巴纳吉、迪弗洛、克雷默为代表的新一代发展经济学家们由此脱颖而出。由于他们的杰出工作,使得实地田野试验,或者更准确地说是随机控制试验(Randomized controlled trials,即 RCTs),成为重要的研究方法。令人惊讶的是,由于他们的引领,发展经济学的研究迅速从宏观转向微观,从“一般政策”转向“扶贫措施”,从“上层”转向“基层”,从多样化的实证工具转向随机统计学尤其是 RCTs 方法,呈现出一边倒的趋势。今天,RCTs 几乎主导了发展经济学并被视为最为成功的研究方法。

与以往发展经济学的学术传统不同,三位获奖者将贫困与发展问题的研究从宏观转向微观,进一步将若干相互关联的问题划分为较小、易于处理的问题,并逐一进行细致的解析与实证检验。其研究的主要特点在于:一是关注微观问题,像管道工一样解决问题。比如营养对健康的影响、健康对收入的影响、土地自耕与雇工的勤奋差异等;二是通过实验方法解决所谓的内生性问题,从而探寻问题表象背后的真正根源并以此获得消除贫困的最佳答案。

基于 RCTs 分析工具,三位经济学家的研究成果因其可操作性,有效提高了人们与贫困作斗争的实践能力。他们通过精细的研究,对一些流行的观点进行了证实或证伪。例如,他们关于疟疾除虫的一系列实验,揭示了健康扶贫的重要性;免费午餐并不能显著提高学生的学习成绩,而课本也仅仅只对最优秀的学生有效;无论在短期还是中期,针对性的辅导均可显著提高学生的学习成绩,等等。诸如此类,他们在教育、健康、小微金融、学校管理、政府治理等众多领域开展了广泛的随机实地实验研究。其突出的贡献在于三个方面:

首先,从 20 世纪 90 年代中期开始,克雷默和他的同事们在肯尼亚开展了一系列的田野实验,以理清教育生产函数的各个组成部分。该方法的关键是把如何促进人力资本积累的问题分解成更细小、更易于管理的主题,从而使得每个主题都可以通过专门设计的随机对照试验进行严格研究。此后,巴纳吉和迪弗洛等人进一步扩展到教育主题的范围,并就其他几个主题展开了田野实验。类似研究基本涵盖了发展经济学的所有分支并发挥了支撑性作用。

其次,在一系列的贡献中,巴纳吉和迪弗洛的一项工作是将微观经济发展问题同发展中国家较低的总人均收入联系起来,建立关键性概念并进行经验观察:低收入和中等收入国家同一生产要素的回报率存在很大的不均匀性,利用有利可图的投资机会的程度也存在很大的差异。这种分配不当的程度可能严重到足以帮助解释低收入和高收入国家之间巨大的全要素生产率差距。巴纳吉和迪弗洛进一步认为,发展研究文献中记录的市场和政府缺陷——无论是政府失败、信贷约束、保险失败、外部性、家庭动态或行为问题——都有助于解释分配不当。RCTs 工具有助于寻找解决这些问题的精准对策。

再次,通过设计新的实验研究方法,为新一代研究人员制定了明确的前进方向。创立于 MIT 的阿卜杜勒·拉蒂夫·贾米尔贫困行动实验室(J-PAL),在许多国家开展 RCTs 研究(2003 年以来进行了多达 978 项的试验),并促进了学界与政界对此类试验结果的推广与应用。

为了进一步说明他们的具体研究工作,下面援引姚宇所介绍的迪弗洛在健康、教育及小额信贷方面的研究成果^[1]。

健康研究。免疫接种是一种经济有效改善儿童生存发展的手段,但全球每年至少有 2 700 万儿童和 40 万孕妇没有得到这一基本服务,导致有二三百万人死于可以进行预防的疾病。在一些经济欠

发达的国家,即使政府有相关的政策为人们提供预防接种,但是居民因为认识不足,而不愿意主动参与接种的现象也很普遍。为了解决现实中的这个问题,经过实验,迪弗洛提出了增加激励措施的操作性方案。她提出特别是在欠发达地区,只有提高免疫接种的基础设施供应,并配套以增加激励措施的政策,才能够在公共服务领域产生积极效果。她在印度拉贾斯坦邦农村开展的干预实验中,把人群分成两组,一组开展完整的 5 个课时的预防免疫的课程教育;另一组也开展授课,但同时给每个参与者每人发一千克小扁豆,结果发现:在一些村庄强制推行的免疫接种率只有 6.2%,对于第一干预组,免疫接种率上升到了 16.6%,但第二组却显著地上升到了 38.3%。这一研究说明,在资源比较匮乏的地区,尝试用一些小额的奖励或很低成本激励措施,就能够让一些看似难以推行的公共服务政策效果得到显著的提升,这在现实社会中具有非常重要的政策意义。

教育研究。在印度,政府在普及初级教育的计划中鼓励地方领导人和学生的父母共同组织委员会来对公共教育资源进行分配,并且监管学校的教育管理。然而,在基线调查中迪弗洛发现,人们不知道存在这样的政策。这种情况在很多发展中国家是普遍存在的现象。她通过实验证明了在社区层面上培训骨干人员来推动这件事是无效的,在学校里宣传这些政策内容也是无效的。但是她在实验中发现,通过对志愿者进行培训,然后让志愿者进入学校对孩子及家长进行的干预活动,却产生了效果。研究结果表明,通过新的干预手段,公民参与对教育活动的监管积极性增加了,家长们变得更加重视教育并积极改善教育管理。

小额信贷。小额信贷是否真的有利于穷人,因缺乏严格的证据而成为一个有争论的话题。迪弗洛在摩洛哥选择了在村庄水平开展小额信贷的实验,并且在一些村庄鼓励家庭女性参与家庭申请贷款的决策,最后她发现,虽然获得贷款农户的生产活动明显增加,但对于贫困改善、地区贫困发生率以及居民消费水平的变化都没有确定的影响。她的这一发现为人们反思和调整小额信贷政策再次提供了经验支持。

总体来说,三位获奖者的成果具有重要的启迪意义:第一,比援助资金更需要实施援助的是致富的能力和办法;第二,贫困的形成是来自他们自己对贫困的认知、形成贫困的生活习惯没有警觉,也没有改变贫困的行动力量;第三,解决大多数人面临的问题并通过专业有效的方法加以缓解,是经济学家应该肩负起的十分重要且刻不容缓的责任^[2]。

二、RCTs 方法及其应用价值

一门学科的发展一般都需要经历三个阶段:第一阶段是对基本事实和关键问题的发现,第二阶段是提出解释事实和新理论,第三阶段是对新理论的检验、运用和预测^[3]。欧美国家的经济学大体已经处于第三阶段。例如,基于对美国经济学会克拉克奖获得者成果的分析表明,从理论转向实证已成为基本趋势。1995 年以前,该奖 80% 的获得者的主要研究领域为理论,20% 为实证;1996—2005 年,主要研究领域为理论的获奖者比例已降至 50%,而主要研究领域为实证的获奖者比例增至 50%;2006—2016 年,只有 33% 获奖者的主要研究领域为理论,其余 67% 均为实证^[3]。

经济学是一门社会科学。通常认为,社会科学区别于自然科学的一个重要特征是它不能进行可控制的随机试验。但是,只有理论却没有实验的科学不能被视为真正的科学,所以经济学理论的现实价值与经验意义总是被人怀疑。长期以来,在对经济增长与发展(包括贫困问题)的研究中,经济学家总在努力寻找其内在的规律,而通过技术手段如统计学方法从数据中发现规律性现象,则是备受推崇的处理方式。规律性最核心的表达是因果关系,但在因果关系中如何寻找和识别其内生的决定因素,一直是经济学家面临的重大挑战。事实上,至少从马歇尔(Marshall)开始,经济学家们就试图依照物理学为模板来塑造自己的学科,从而将其尽可能科学化。经济学家们从没有放弃过对实验的追求。例如,“垄断竞争理论”创始人张伯伦(Chamberlin)就曾在自己的课堂上进行实验以模拟市场竞争,用来验证我们熟悉的供给需求理论。2017 年的诺奖得主塞勒(Thaler)关于禀赋效应的经典实验也是在课堂上完成的。随着心理学对经济学影响的加深,一部分经济学家也开始借助心理学家的工具,在实验室研究人的行为。

2002 年的诺贝尔经济学奖首次颁给了实验经济学。卡尼曼(Kahneman)是因为“把心理学研究和经济学研究结合在一起,特别是与不确定状况下的决策制定有关的研究”而得奖。史密斯(Smith)是因为“通过实验室试验进行经济方面的经验性分析,特别是对各种市场机制的研究”而得奖。该年的诺奖,可以视为一个里程碑,一是表达了当代经济学的“行为转向”,二是表达了经济学前沿的“实验转向”。从此,实验分析成为经济学中的热门工具,经济学甚至日益成为一种实证经济学。的确,从 2003 年开始,班纳吉与迪弗洛就在 MIT 建设了贫困行为实验室,哈佛大学的克雷默则是他们重要的合作者。以这个实验室为基地,他们在世界各地开展田野实验研究,掀起一阵学术风潮,在顶尖学报上发表了数量惊人的论文。因为对穷人的研究并在实践中取得了良好效果,迪弗洛在 2010 年获得了有“小诺贝尔奖”之称的克拉克奖。班纳吉与迪弗洛合作的《贫困经济学》(中译为《贫穷的本质》)也成为畅销书,获得了当年的高盛年度商业图书奖。

经济学家的试验大体包括两种类型。一是所谓的实验室试验。在行为经济学领域被广泛采用。其基本的逻辑是借助于一些外生的冲击来构建对照组和实验组,然后来对它们的表现进行比较。二是在田野环境下的随机性控制试验。严格地说,实验室试验只是一种准实验,因为它离现实生活非常遥远。RCTs 由于能够更好地模拟真实世界,往往被视为真正意义上的试验。其优点就在于无需为可能存在的干扰因子而担忧(否则就存在“内生性问题”)。从统计学角度来看,如果“零假设”(原假设)被拒绝,就能够确认干预手段的有效性。

2004 年,Harrison 等发表了介绍田野实验的标志性论文^[4]。截至 2019 年 10 月 18 日,该文在 Google 学术搜索上的被引次数已达到 2 742 次,足见该方法在经济学领域的影响力。国内学者张林秀^[5]、罗俊等^[6]也先后对相关方法进行了较为完整的介绍(如表 1、表 2)。

表 1 非实验性方法与实验方法的比较

方法	方法描述	研究特点	假设前提	所需数据
前后比较	测量参与个体在项目参与前后的变化	参与项目的个体本身参与前后的情况	项目是造成参与个体前后变化的唯一因素	项目参与个体参与项目前后的指标变量
倍差分析	测量干预组及对照组的前后变化并进行比较	干预组和对照组两类个体进行比较	如果没有干预,干预组和对照组的个体会有一样的前后变化	干预组和对照组的前后 2 期数据
统计匹配	在对照组中找一个或多个与干预组有相似特征的个体进行对比	对参与个体同非参与但具有相似特征的个体进行比较	没有用作匹配的变量不会对结果造成偏差	参与和非参与个体的“匹配变量”和最终结果变量
非实验方法	全体个体根据特定标准进行排序,并设定个体参与项目的准入条件,将参与者与非参与者进行比较	靠近准入标准但又不够资格参加项目的个体与参与项目的个体	靠近准入标准但不够资格的个体同那些刚好达到标准的个体之间没有显著的差异	最终结果变量、准入标准变量和其他控制变量
断点回归	参与情况可由偶发事件或工具变量来预测;但该变量同最终的结果变量并不关联	将工具变量预测的参与项目的个体同预测的未参与项目的个体进行比较	如果工具变量对项目参与情况的预测能力有限,该“工具”变量就缺乏解释效力	最终结果变量、工具变量和其他控制变量
实验方法	测量 2 个变量之间的因果联系	随机分配到干预组和对照组中的个体	随机性得到满足;干预组和对照组在统计上必须保持同质	干预组和对照组在干预前后的数据;样本量必须足够大

注:参见张林秀^[5]。

表 2 不同实验方法及其比较

	被试类型	被试信息	激励	实验环境	被试者知晓	研究者干预	外生改变
实验室实验	大学生	抽象	诱导	实验室	是	是	否
人为的田野实验	各类人群	抽象	诱导	实验室	是	是	否
框架的田野实验	各类人群	情境	真实激励	自然环境	是	是	否
自然的田野实验	各类人群	情境	真实激励	自然环境	否	是	否
自然试验	各类人群	情境	真实激励	自然环境	否	否	是
非试验数据	各类人群	情境	真实激励	自然环境	否	否	否

注:参见罗俊等^[6]。

RCTs 试验的基本过程可以分为五个步骤。一是确定试验人群。以节水宣传为例。由于节水涉及居民的资源消费,需要先确定进行实验的居民区:是一个小区还是若干个小区;若干个小区是集中在一起还是分布在城市不同方位,或者是在不同的城市。二是随机分组。试验者将试验对象随机分成一个控制组以及一个或多个干预组。随机分组的目的是要消除不同组别之间的系统性差异,从而保证可比性。其中,一个重要的工作是基线调查并经过统计检验,以确保组别之间没有显著差异。三是试验干预。控制组的居民像他们往常一样的生活,而干预组的居民会受到明确界定的干预措施影响。例如,干预措施可能包括发放节水传单或进行上门游说等多种方式。关键是要确保干预组受到干预,且干预措施没有溢出效应,控制组没有受到“污染”。四是数据收集。可以通过问卷、实地检测和经济学游戏等多种手段实现。五是对数据进行处理和分析,并得出结论。通过与控制组的对比,观察干预组的行为变化,由此揭示干预(或政策措施)的效果及其行为发生学意义。

以往的实证研究存在两个突出的问题。一方面,尽管越来越多的微观计量方法逐渐被开发出来,从一元回归到多元回归,进而到样本选择模型、固定效应模型、双重差分法、工具变量法、断点回归法等更高深的计量模型,试图解决研究对象的可比性问题,但各自均存在局限性。尤其是遗漏变量、双向因果等问题的广泛存在,使受到某种干预的个体和没有受到的个体之间常常不具有可比性,因而存在不可信的问题。另一方面,由于政策的颁布与实施总是滞后于现实需求,尽管有些备选政策具有理论上的逻辑合理性,但是否具有实际可操作性往往是未知的,尤其是人们无法观察到新的政策选项与已实施的政策是否存在消长或冲突,更是不确定的。所以,RCTs 方法在理解政策绩效的可信性与确定性方面具有不可比拟的工具性优势。RCTs 方法的“实地”特性,因为更加贴近现实,所以成为检验和认识已有经济理论、逻辑假说、政策策略的重要工具。这类方法无疑能够鼓励和激发“直面现实”和“问题导向”的微观研究。由 Smith 等^[7]发展的实验室实验研究方法于 20 世纪中叶逐渐兴起,而田野实验尤其是 RCTs 方法基于对实验室实验的反思与改进,现已成为一个重要的经济学研究领域和数据来源。甚至有人将其称之为“21 世纪的经济方法”。

三、对 RCTs 方法的争议

2019 年诺奖的颁奖指向,表面看起来关注的是扶贫(研究对象),但实质上是关注于内生性问题的处理方法(研究方法)。事实上,RCTs 方法并非三位新科获奖者的新发明,它的应用由来已久。该方法最初是一种对医疗卫生服务中的某种疗法或药物的效果进行检测的手段,常常用于医学、药学、护理学,以及生物学与农学的研究中,在司法、教育、社会科学等其他领域也有所应用。只不过是最近十多年在研究贫困问题的发展经济学圈子中,该方法格外盛行而成为“标准”与时尚。应该说,对于 RCTs 方法的有效性,一直存在争论。只是 2019 年的经济学诺奖使得这些争议格外引人注目。

事实上,即使 RCTs 在医学领域应用广泛,但其方法的有效性至今依然没有达成共识。不久前 Frieden 在久负盛名的《New England Journal of Medicine》发表文章,从公共卫生的角度对 RCTs 的局限性进行了较为全面的评论^[8]。他强调:第一,尽管 RCTs 的内部效度很强,但无法将研究结果推广至研究人群以外。第二,RCTs 的研究期限或样本大小通常不足以评估治疗效应的持续时间(例如,疫苗免疫效果下降)或发现治疗中出现的罕见严重不良反应,这些通常在上市后监测和长期随访中变得明显,但在 RCTs 中实际无法进行评估。第三,RCTs 越来越高的成本和时间限制也可能造成研究者的选择性设计。例如,选择高危人群进行研究,但这些人群可能与更广泛的目标人群无关。第四,RCTs 评估个体化治疗效应的能力也有限,如外科手术的不同操作,同时 RCTs 对罕见疾病而言通常不切实际。因此,他认为,对于健康干预研究,并没有千篇一律的最佳方法;没有研究设计是天衣无缝的,所有类型的研究都可能出现相互矛盾的研究结果。他提供多个例子说明,不同的数据来源均有其优点与缺点,临床和公共卫生决策的数据依据几乎总是不完整的。特别要提醒的是,在 RCTs 数据之外,绝对不能排除观察性数据的重要性。

在经济学领域,更是存在类似的质疑。Mader 等怀疑 RCTs 能否产生适当的证据,因为该方法看

起来往往是由实验方案决定了问题^[9]。在某些社会政策辩论中,“寻求随机化的东西”导致了方法论的原教旨主义,随机主义者经常将试验(理想情况下是他们自己的实验)作为唯一真正可信的证据来衡量。科学哲学家卡特赖特(Cartwright)一直指责称,RCTs并不能告诉我们究竟什么变量在真正发挥作用^[10]。近期,2015年经济学诺奖得主迪顿(Deaton)等又专门撰文批评经济学领域的RCTs方法,并认为进行试验的研究人员和各种组织通常会夸大所得证据的价值,所以不应该将该方法当作研究领域的“黄金标准”^[11-12]。正如由15位领先的发展经济学家于2018年共同发布的一份公开信所说的那样,随机对照试验和行为主义方法设计实际上忽视了更宏观的图景^[9]。

综合各种批评意见,主要涉及7个方面:(1)社会干预发生的实际情景通常极其复杂,其中涉及大量发挥作用的影响因素。而且环境一旦发生变化,这些影响因素很容易消失不见。(2)RCTs作为因果识别的方法,仅仅是多种方法中的一种,当然有其缺陷。例如,新药品的RCTs实验通常只能说明是否有效果,既无法保证项目参与者的同质性,也不能彻底揭示其内在机理与作用机制。(3)如果说RCTs对于解决严格控制的小样本事件有价值,但对于大样本或不同情境下的大样本、特别是涉及公共政策的重大问题来说,局限性就很大。对微观小问题研究的有效性,并不意味着对解决宏观大问题的适用性。(4)随机对照试验方法论和科学复制存在严重的问题^①。(5)贫困问题不仅仅是个人行为问题,与公共政策密切相关,而公共政策的非排他性,几乎不可能采用RCTs方法进行识别。(6)贫困与发展并非是微观问题的全部,它涉及传统文化、经济结构、制度安排等多方面的问题,微观的RCTs显然对此无能为力。(7)由于随机对照试验非常昂贵,这不仅将会使研究变缓(时间的延滞必然导致情境变化),而且往往无法得出结论性的结果。所以学者们强调,对特定随机对照试验的细节进行详细持久的监督和审查、讨论什么时候应该开展随机对照试验,什么时候应该采用其他研究方法,是需要谨慎对待的问题。林毅夫教授曾经指出,RCTs这种试图“以科学经验为基础的政策来减贫”的方法尽管对于理解一些特定的微观发展项目颇有帮助,但受成本所限,实验方法只能去探讨一些小问题,而对于类似产业政策、贸易开放这种大的政策问题却无能为力,因此这类研究通常不能满足我们对知识空白的最紧迫需求^[13]。

在各种批评的声音中,迪顿的批评是最为专业且系统的^[11, 14]。归纳起来包括:第一,相比观察性数据,RCTs数据一般样本比较小,而且大多是针对一个局部地区或者行业的研究,所以缺乏代表性和外部有效性;第二,RCTs研究只是改善了系数估计的无偏性,但无法确保因果关系的建立。在很多时候,无偏性未必比降低方差重要;第三,不同的RCTs设计对于同一个问题可能产生不同结果,并不能解决政策制定中的争议;第四,因为样本比较小,所以RCTs较难产生关于宏大主题的思考 and 伟大思想;第五,RCTs大多是富人对穷人、富国对穷国做的,背后可能体现出高高在上的家长式思维;第六,RCTs提供的大多是小修小补、小打小闹式的政策改进,而不是长期解决方案。基于上述,迪顿教授强调,RCTs并不是解决发展或者扶贫问题的万灵之药或“魔术子弹”。相比基于大样本的观察性数据,它的优势可能没有倡导者说的那么大,甚至劣势更多。因此,它不具备替代基于观察数据的研究的可能,甚至在很多时候不如后者效果好。

迪顿教授的上述批评不一定都完全恰当,但他强调发展经济学的研究不能从一个极端走向另一个极端,则是恰当的。他在一次采访中表达了如下看法:在20世纪50-60年代,经济学家们开始借助回归分析进行计量研究。在他们眼中,回归分析是一个“神奇工具”(magic tool),可以揭示几乎所有事情^[15]。他们试图对各种事物进行多变量回归,并在一种因果关系框架内对其进行解释。在现在的发展经济学家们看来,他们也有一个“神奇工具”——RCTs,但对这个工具的弊端还不甚了解。当然,其他任何估计方法也都有其优点和缺点。但我认为RCTs的弊端很大,很少能达到人们日益对其高涨的预期。人们之所以选择RCTs,是因为他们厌倦了关于内生性、工具变量和样本选取等等的争

① 以小额信贷为例,2010年一个研究团队对J-PAL的研究工作进行了复制,结果发现数据收集和分析中存在众多缺陷,且足以使任何发现的有效性受到质疑。尽管证明了RCTs的缺陷是如此之明显,但复制研究的结果并未得到J-PAL的认可。Mader等^[9]为此感慨,RCTs研究是如何通过同行评审并广泛被学界视为可靠的呢?!

论。但是,所有这些问题其实在 RCTs 中都以不同的形式重新出现。因此,RCTs 并未带来研究质量的提升,坏的研究仍然层出不穷。

实验研究的目的是“见微知著”,但将 RCTs 的研究方法引入经济学,最多只能做到“见微”,却无法做到“知著”^[16]。因为经济学的研究对象是人,人生活在社会之中,会受到各种各样因素的影响,要排除各种变化带来的影响进行控制性考察,这在理论上几乎是不可能的。变量的无法控制、看不见的数据无法统计,RCTs 这个在自然科学、工程领域相对有效的研究方法,在经济学这里却存在致命的硬伤。例如,迪弗洛等人发现,肯尼亚当地农民对化肥的使用较为不足,会等到农作物较为成熟的时候才开始购买化肥,而使用一些小小的价格折扣,就会激励他们在合适的时间施用化肥。这一研究对于以农为生的穷人来说,价格折扣激励或许能够改变他们的行为响应,但对于经济相对发达地区或兼业的农民来说,甚至在有着不同传统的村庄,该激励会同样有效吗?显然,RCTs 并不能对所有因素进行严格控制。此外,在信息发达与便利传播的今天,一个购买化肥的农民会不会通过微信、Twitter 或其他的推文方式向外发布信息呢?一个干预组的农民下地施肥难道不会被控制组的其他农民发现或观察到?如果控制组受到“污染”,研究结果就必然存在致命的内生性问题。一般来说,研究者通常为了验证自己预设的假说,很难真正洞悉当地人行为决策的全貌,可能会遗漏重要变量。因此,RCTs 不仅难以“知著”,有时甚至难以“见微”。

必须强调,通过 RCTs 方法理解影响贫困的关键变量,识别精准扶贫措施,当然有可取之处,但因其有着强烈的环境或特定条件的情境依赖性,所以其试验结果的普适性依然存在疑问。因此,强调 RCTs 仅仅是可选择的众多分析方法与手段中的一种,显然是有必要的。

四、一点余论

正如迪顿教授所批评的那样,在经济学研究中,总是一直存在明显的二律背反现象。从方法论来说,有两种似乎出自同一学派但实质不同的方法:一是弗里德曼著名的“只要预测准确,经济学假设的真实性无关紧要”的思想主张,被西蒙戏称为“轮椅里的经济学”方法论;二是科斯所倡导的“直面现实的经济学的”方法论,即悬置一切既有理论,以真实现象为出发点重新构建能够解释现象的理论,哪怕这重新构建的理论 with 既有理论相重合而显得“多此一举”。同样,在处理方式上,人们也很容易受到两种极端做法的影响:一种是集中提出相互依存的关系,把被认为是主要的那些关系汇集于一个分析框架。这种做法的缺点是,它只能提出一个过于简单的宏观图景,而不能使各种不同领域的专家认识到他们自己的问题,也难以为决策提供操作方案。另一种是凭借目前所能提供的大量专门知识分析特殊问题。在这种情况下,虽然具体方面的问题从未被忽视,但它却不能使不同领域相互关联中的那些困难得到解决。显然,早期发展经济学的研究属于前者,这次诺奖获得者的研究属于后者。

我猜想,可能是因为方法的时髦性与话题的重要性让评奖委员会在获奖对象的选择上有了偏向(因为贫困问题重要,复兴发展经济学也重要),而不完全是理论与方法的突破,从而让 2019 年诺奖的含金量打了折扣。事实上,三位获奖者的研究成果所表达的操作性工作及其研究结论,既没有证伪经济学的已有知识,也没有基本的创新性理论贡献,更不可能通过他们的实验策略而使缓解贫困问题发生根本性改观。尽管众多经济学家已经对经济学的实验主义方法论反复做过批判,但确实应该承认,近些年来经济学诺奖以及其导向下的经济学研究,技术性 or 技巧性倾向越来越明显,与基础性的理论突破或原创性的知识贡献的初衷已经背道而驰。迄今,尽管行为经济学的发展日新月异,但在不确定性情境下不同的偏好模型并不具有预测和解释能力的一致性;尽管实验经济学的发展如日中天,但不同情境下的实验结果矛盾重重也不具有逻辑的一致性。话题的时髦性似乎成为基本的研究导向。或许引用《黑天鹅》作者塔勒布(Taleb)的一句话是恰如其分的:社会科学理论的命运取决于其传染性,而不是其正确性。

有人说,近 20 多年来,在诺贝尔经济学奖的获奖者中,似乎已经很难找到哈耶克、科斯、弗里德曼那样的经济学思想巨匠了。当然,这与经济学的发展阶段有关,但显然与经济学(当然包括发展经济学)的实证研究导向有关。坚持“问题导向”的科学研究,强调直面现实的经济学的探索,无疑是正确的

(RCTs 往往是实验构建的空间,并不一定代表真实的世界)。但应该强调,好的研究既要回应现实需求,也必须推动理论进步。因此,2019 年诺奖能够给我们的启发是,“好的”经济学研究必须立足于基本事实、情景依赖与关联性证据,才能发现问题、筛选问题、研究问题和解决问题,进而在此基础上能够发现理论并推进理论创新。应该持有的基本态度是:

(1)解释和揭示新的经济社会现象,对已有的理论假说进行证实或证伪;

(2)用更严密或更规范的语言重新叙述已有的理论或假说,或者在方法论层次上批驳已有的理论;

(3)直面现实并基于系统性证据提出新的学理解释,或者建立新的假说与理论。因此,我同意 Mader 等人的看法^[9],即一个 21 世纪的先进的经济学应当摒弃行为主义的思维方式和方法论,并将其应用于那些导致贫困和不平等的政治、社会和文化问题的分析。真正的发现等待着那些拒绝狭义框架的“贫困”经济学家。

中国的经济学界已经开始了 RCTs 的应用研究,但如果忽视基本理论与学理性研究,过于倾向于实证研究并将其转换为数据挖掘工或管道修理工的技巧性工作,显然是不恰当的。所以我担心 2019 年诺奖可能带来的后遗症是:将某类方法与某类被关切的特殊问题结合而赢得人们的青睐,可能会滋生并鼓励学界在选题与研究方法上的机会主义(尽管已经存在)。

参 考 文 献

- [1] 姚宇. 埃丝特·迪弗洛对发展经济学的贡献[J]. 经济学动态, 2011(6): 87-91.
- [2] 孙立坚. “研究大多数人的问题”成为近年诺贝尔经济学奖的亮点[N]. 21 世纪经济报道, 2019-10-16(4).
- [3] 甘犁, 冯帅章. 以微观数据库建设助推中国经济学发展——第二届微观经济数据与经济学理论创新论坛综述[J]. 经济研究, 2019(4): 204-208.
- [4] HARRISON G W, LIST J A. Field experiments[J]. Journal of economic literature, 2004, 42(4): 1009-1055.
- [5] 张林秀. 随机干预试验——影响评估的前沿方法[J]. 地理科学进展, 2013, 32(6): 843-851.
- [6] 罗俊, 汪丁丁, 叶航, 等. 走向真实世界的实验经济学——田野实验研究综述[J]. 经济学(季刊), 2015(3): 853-884.
- [7] SMITH V L. An experimental study of competitive market behavior[J]. Journal of political economy, 1962, 70(2): 111-137.
- [8] FRIEDEN T R. Evidence for health decision making—beyond randomized, controlled trials[J]. New England journal of medicine, 2017, 377(5): 465-475.
- [9] MADER P, RICHARD J, MAREN D CK, et al. 2019 Nobel prize reveals the poverty of economics[C]. Institute of development studies(IDS), Published on 17 October, 2019.
- [10] CARTWRIGHT N. Are RCTs the gold standard? [J]. Biosocieties, 2007, 2(1): 11-20.
- [11] DEATON A, CARTWRIGHT N. Understanding and misunderstanding randomized controlled trials[J]. Social science & medicine, 2018(210): 2-21.
- [12] GUO S Y, FRASER M W. Propensity score analysis: Statistical methods and applications[C]. Thousand Oaks, CA: SAGE Publications, 2010.
- [13] 陈永伟. 诺贝尔经济学奖揭晓: 他们提高了全球减贫的能力[N]. 四平日报, 2019-10-17(8).
- [14] 包特. 随机对照实验的“新老诺奖得主之争”, 怎么看[EB/OL]. (2019-10-17)[2019-10-19]. <http://opinion.caixin.com/2019-10-17/101472302.html>.
- [15] 张维迎, 王勇. 2015 年诺贝尔经济学奖得主安格斯·迪顿对随机对照实验方法的批评[Z/OL]. “经济学原理”公众号, 2019-10-16.
- [16] 张是之. 最佳的田野调查在中国[Z/OL]. “经济学原理”公众号, 2019-10-17.

(责任编辑:陈万红)