

随机实地实验:方法、趋势和展望

陆方文*

摘要: 随机实地实验作为识别社会现象中因果关系的重要手段,在国外经济学界正得到日益广泛的应用,而国内学界对这一重要方法的应用还刚刚起步。结合前沿文献和自己的研究实践,本文介绍了随机实地实验在可信性和创新性上的两大优势;系统地探讨随机实地实验的关键技术环节,以期降低开展实验的技术性壁垒;并对随机实地实验在国际上的研究趋势进行归纳和整理,希望能够引导国内研究实现后发赶超,推动国内随机实地实验研究的发展。

关键词: 随机实地实验;方法;趋势

一、引言

现代经济学发展的一个重要趋势是随机实地实验(random field experiment)方法的迅速崛起。经济学的科学化趋势不仅要求构建规范严谨的理论模型,而且还要求运用经验数据对经济理论进行实证检验。传统的检验方法主要依赖计量经济方法和数值模拟技术。其中计量经济方法的可信性依赖于数据的可获得性和可靠性,但变量的有限性和广泛存在的相关性导致在因果关系的评估中容易出现难以克服的遗漏变量偏误问题。数值模拟技术在很大程度上受制于研究者的假设设定,随意性较大。近几十年迅速发展起来的实验经济学方法能够在很大程度上弥补这些传统实证方法的不足,它有助于进一步打通经济学理论研究和现实世界的隔阂,并且有效避免数值模拟的主观性和数据生成的客观限制,从而促进经济学研究的快速发展。正如 Guala (2005) 所说,实验经济学方法已经成为经济学家检验经济学理论、获得新知识、创造新思想的重要来源。

迄今为止,实验经济学方法主要形成了实验室实验和实地实验两大流派,二者互相补充,相互促进,各有优劣。Levitt 和 List (2009) 以及 Al-Ubaydli 和 List (2015) 等详细对比分析了两大流派的优缺点:实验室实验在操作上相对简便容易,但在满足外部有效性方面有所欠缺;而实地实验则在满足外部有效性方面具有明显的优势,但实施环节更为复杂。他们把实地实验具体细分为人工实地实验、框架实地实验、自然实地实验等。人工实地实验类似于把实验室的游戏搬到实地的工人、农民或者管理者中去,是在现实社会中完成实验室实验设

* 陆方文,中国人民大学经济学院,邮政编码:100872,电子信箱:lufangwen@ruc.edu.cn。

非常感谢周业安、李辉文和付明卫以及五位匿名评审专家在本文写作过程中给予的建议和修改。本文受国家社会科学基金青年项目(项目编号:13CJY031)经费资助。文责自负。

计,其中的实验情景是完全人为设计出来的,而并非自然生成。框架实地实验则指在现实的环境中针对实地的参与者探讨现实的行为、信息、影响效果等的实验,但参与者对实验开展是知晓的。自然实地实验则在框架实地实验的基础上更加贴近现实,参与者不知道实验的开展。后两者的核心区别是实验被试是否知情。由于这三类实验都具有随机安排干预措施的特点,因而也被称为随机实地实验。本文重点关注自然实地实验,一定程度上兼顾框架实地实验。

随机实地实验方法在欧美经济学界已经得到迅速发展,不少学者系统运用这一方法开展研究,帮助欠发达经济体摆脱贫困等各种社会困境,并取得了令人信服的成效。Esther Duflo 因为将随机实地实验方法引入发展经济学而被授予 2010 年的克拉克奖。但国内学界对这种方法还比较陌生,迄今开展的研究也非常少。我国社会和经济转型面临大量的社会困境难题,其中很多通过传统的经济学方法难以获取答案,因此普及和推广随机实地实验方法就显得尤为重要。

Harrison 和 List(2004)、Levitt 和 List(2009)、陈玉梅和陈雪梅(2012)、姜树广和谯倩(2012)、张生玲和周晔馨(2012)、罗俊(2014)、罗俊等(2015)等就实地实验研究进行了一些综述性讨论,Duflo 等(2008)为如何开展实地实验提供了实用指南。在前沿文献和自己研究实践的基础上,本文的贡献主要表现在两个方面:一是对随机实地实验的技术性问题进行系统探讨,有助于降低开展实验的技术性壁垒;二是对随机实地实验在国际上的研究趋势进行归纳和整理,有助于引导国内研究实现后发赶超,推动国内随机实地实验研究的发展。

本文余下部分的安排是:第二部分简要说明随机实地实验的基本过程并概括其两大优势;第三部分探讨实验设计的重要技术环节;第四部分讨论随机实地实验方法的局限性和对策;第五部分归纳和整理随机实地实验在国外的发展趋势;最后一部分探讨随机实地实验在国内的研究现状和发展契机。

二、随机实地实验的基本过程及两大优势

随机实地实验的基本过程一般包括五个步骤。第一步,确定实验人群。以探讨节能宣传对居民消费影响的实验为例,因为涉及居民的能源消费,我们首先要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。第二步,进行随机分组。实验者需要将实验对象分成一个控制组以及一个或者多个干预组。随机分组的目的是通过消除实验前不同组别之间的系统性差异以实现可比性。在这个例子中就是要求,如果没有实验干预的话,不同组居民之间的能源消费行为是可比的,这样我们就可以把干预后不同组之间的行为差异解释成干预的效果。随机分组是随机实地实验的关键环节。第三步,进行实验干预。控制组的居民像往常一样的生活,而干预组的居民会受到明确界定的干预措施影响。举例来说,干预措施可能包括发放节能传单或者进行上门游说,宣传为公益而节能或者强调节约自我成本等等。基于媒介或内容的不同可以形成不同的干预措施。干预措施可以是现实当中已有的政策或方法,也可以由研究人员自己创新设计。第四步,数据收集。这可以通过问卷、实地检测 and 经济学游戏等多种手段实现。第五步,也是最后一步,是对数据进行处理和分析,并得出结论。

分组的随机性和干预措施的可设计性为随机实地实验带来两大优势:一是可信性优势,二是创新性优势。经典的Rubin因果模型指出,潜在结果的概念对于探讨干预的因果效应非常重要;只有知道被干预个体如果不被干预时的表现,才能够准确地衡量干预的效果(Rubin,1974)。但同一个体无法同时存在被干预和不被干预的两种状态,因而潜在结果必须通过其他具有可比性个体去获得。计量经济学的各种方法,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,都是为了更好地解决可比性的问题;但各有适用的条件,并且各自存在着局限性。随机实地实验的一个核心环节是进行随机分组,随机分组使得控制组和干预组在期望上都能代表整体,实现可比性。如果没有干预措施,控制组和干预组无论在过去还是在潜在的未来都是可比的,因而,在干预实施后控制组和干预组之间的差异都可以归因于干预措施。控制组和干预组安排的随机性,为实验结果的可信性提供了有力的保障。

除了可信性优势之外,人为的实验干预措施还产生了随机实地实验的第二个优势——创新性优势。有一些公众关心但现实当中还没有实施的政策,或者甚至公众没有意识到、但从理论中可以推导出的有效措施,是无法通过观察性研究进行定量政策评估的,但这些情况都可以通过实地实验、即设计相关的干预措施进行探讨。此外,有一些在现实当中无法分解的机制,也可以在实验中设定相关的情境进行分解。干预措施的可设计性为实地实验提供了创新性的源泉。

三、随机实地实验方法的若干关键环节

随机实地实验在技术层面有很强的实验设计要求。概览近年来的众多文献,一项随机实地实验主要涉及七个关键环节:确定研究主题、考察实地背景、选择随机方法、确定样本量、安排干预措施、收集实验数据和妥善处理实验中的道德问题等。本部分将对这七个关键环节进行详细的讨论。

(一) 确定研究主题

研究主题的类别直接影响到随机实地实验研究中干预措施的设计策略。随机实地实验的研究主题从功能上大致可以分成两类:

一类侧重探讨新方法。扶贫、教育、健康干预等方面的大多数研究都属于这一类。这类实验的目的,是从实验当中发现一些有效的潜在干预措施,这些措施可以进行大规模的社会推广并提升社会福利。以此为出发点的随机实地实验对于干预措施的有效性和可行性有着双重要求:一方面要求设计出的干预措施具有现实有效性,这好比爱迪生的实验,虽然发现一些不能用做灯丝的材质的确是有益的,但最终发现能够发光的灯丝对于社会的影响更为重要;另一方面又要求这些干预措施在现实当中是潜在可推广的,否则实验也会丧失现实意义。

另一类实地实验研究则侧重于识别现实状况。很多关于歧视和机制探讨的文章都属于这一类别。比如,Bertrand和Mullainathan(2004)试图探讨在美国就业市场上是否存在基于白人和黑人的种族歧视,其目标并非要通过简历的设置获得歧视的结果,而是要让实验中白人和黑人的简历尽可能地模拟现实劳动力市场中的简历,从而探讨在现实劳动力市场上是

否存在歧视问题。Atkin 等(2017)目的不是为了寻找更好的外贸手段,而是探讨通常的来料加工对企业生产的影响,因而其关于如何续订单等机制的设置需要尽可能地符合现实中的一般情况。这类实验的第一要义是干预措施尽可能地模拟现实。

(二) 考察实地背景

如果将随机实地实验比作创业,那么研究主题就是创业的点子,实地背景则构成创业的环境。List (2011)特别强调,实验人员一定要熟悉所研究的市场。考察实地背景,需要回答以下几个主要问题:第一,实地是否存在所预想的待解决的问题,或者存在提升的空间?第二,当前存在这个问题的原因大致是什么?第三,干预措施理论上能否解决这个问题?如果能够解决,可能的机制大概是什么?第四,这个实地背景是否允许干预措施有效地进行?尤其,控制组和干预组之间能否有效隔离?第五,能否进行有效的数据收集?尤其如何进行长期的追踪?

(三) 选择随机方法

随机的好坏直接影响到研究的可信性,是实地实验的核心环节。随机的目的是为了实现在控制组和实验组个体之间的可比性:不仅在过去是可比的,而且在未来如果没有实验干预措施,也是可比的。

在方法上,随机大致可以分为三种方法。第一种方法是简单随机,指的是不管样本特性如何,直接做随机分组。第二种方法是分组随机,指的是首先根据个体的特征进行分组,然后在每一个小组的内部进行随机分组,把各小组的干预组合并起来就是总体的干预组。第三种方法是匹配随机,指在每两个个体中,随机获得一个干预组个体和一个控制组个体。匹配随机是分组随机的极致,对小样本的随机分组有重要的应用价值。

交替分组,也是现实当中经常使用的方法,它是指按照列表顺序,若分两组按 0/1/0/1 顺序交替分组,若分多组则按 1/2/3/4... /1/2/3/4... 分组。虽然经常被使用,但交替分组不是随机分组。比如,Miguel 和 Kremer (2004)在肯尼亚进行的疟疾除虫实验中,对学校基于特征分组之后,在每一小组中将学校名称按照字母排列,然后按照 1/2/3/1/2/3... 进行交替分组。在动态获得实验样本的时候,也常常采取交替分组。比如,在商场门口想获得穿高跟鞋的女士样本,那么可以按照样本出现的顺序进行交替分组(Guéguen, 2015)。交替分组在有些时候可以实现各组之间可比性的目的,但也存在一定的风险。比如,如果样本列表本身是男/女/男/女交替顺序的,我们在 0/1/0/1 的分组,将会得到所有的 0 是男生 1 是女生。对此,更值得推荐的方法是,在有序名单的基础上嫁接匹配分组。比如,Miguel 和 Kremer (2004)在获得学校名单字母排序后,按照顺序把每三个学校作为一个小组,更优的做法是从每一个小组随机确定 123,最后分组的结果可能是 123,321,231,132...,这可以极大地降低风险。

(四) 确定样本量

一个好的实验设计要具有鉴别度。如果一个干预措施有六分的效果,而且这六分的效果具有重要的现实意义,在实验假设检验的时候,就必须做到让这六分的效果显著不为零。这涉及两方面的问题:一是对实验措施的效果进行预估,二是对分析的误差进行预估。后者

与样本量紧密相关。在一个干预组和一个控制组的简单对比分析中,干预措施的效果可以由 $Y_i = \alpha + \beta T_i + \varepsilon_i$ 方程中的 β 表示,在独立同方差性的假设下, $\hat{\beta}$ 的方差可以表示为: $Var(\hat{\beta}) = \frac{\sigma^2}{SST_T} = \frac{\sigma^2}{N \times Var(T_i)} = \frac{\sigma^2}{NP(1-P)}$ 。其大小受到误差的方差 σ^2 、实验组样本占总样本的比例 P 和

样本量 N 的影响。减小 $\hat{\beta}$ 的方差,主要有三个方式:一是通过增加控制变量来减小 σ^2 减小 $\hat{\beta}$ 的方差。二是确定合适的干预组占比(P)。干预组的比例(P)和控制组的比例($1-P$)对方差的影响是一样的,如果二者的实验成本是相似的,那么 $P=0.5$ 的时候 $P(1-P)$ 取最大值;如

果干预组的成本远大于控制组的成本,那么可以适当地调整控制组的大小, $\frac{P}{1-P} = \sqrt{\frac{C_c}{C_t}}$,其

中, C_c 和 C_t 分别是控制组和干预组每一个样本的成本(Duflo et al., 2008)。三是确定足够的样本量 N 。给定干预效果 E ,使用通常的0.05的显著性水平和80%的功效,拒绝 $E=0$ 的虚

拟假设要求 $N \geq \frac{7.84 \times \sigma^2}{E^2 \times P(1-P)}$,其中 $7.84 = (1.96 + 0.84)^2 = |t_{0.025} + t_{0.2}|^2$ 。上述公式中,干预效

果 E 取决于对干预效果的预估, P 根据实施成本设定, σ^2 的值则有一定的复杂性。如果有前期数据基础, σ^2 可以明确地计算出来。如果不知道 σ ,经验而言,干预效果的大小一般定义:小效果 0.2σ 、中效果 0.5σ 、大效果 0.8σ 。如果是在群体层面上进行随机分组,群体成员之间的相关性会使得 σ 迅速增加,对样本量会提出更高的要求。

(五) 安排干预措施

干预措施的随机安排可以在两个层次上进行:一是个体层次,另一个是群体层次。比如,我们想探讨农村医疗保险对村民医疗支出和健康的影响。假设当地还没有医疗保险,我们可以在家庭层次上进行随机干预,即把每一个村庄里的家庭分成有保险的家庭和没有保险的家庭;也可以在村庄的层面上干预,把村庄分成有保险的村庄和没有保险的村庄,在有保险的村庄里面的家庭就都有保险。究竟是以家庭为单位还是以村庄为单位做随机安排,需要进行至少四个方面的考察。第一,关于样本量的考察。如果以村庄为单位随机安排医保,需要很多的村庄实现随机分组的可比性,因为随机是大样本性质;同时由于同一村庄内部家庭之间的相关性,也需要更多的村庄获得估计的精确性。第二,关于溢出效应的考虑。如果有医保的村民会代没有医保的村民开药,那么就会导致医保的溢出效应;如果不同的家庭看同一医生,也会通过医生产生溢出效应。溢出效应会导致对比同一个村庄内部的村民无法准确地评估医保的效果。但如果村庄之间的溢出效应很小,应以村庄作为随机干预的单位。第三,关于干预措施实施的考虑。如果医保涉及到特殊药品的供应或者不同表格的填写,让基层医生在不同的药品或者表格单据上来回切换可能很麻烦。因而,在干预的实施中,可能以村庄为单位在操作上更容易。第四,关于控制组家庭合作的考虑。以家庭为单位,同一个村庄内的控制组家庭更可能知道实验,可能会对调研人员产生敌视,也不利于收集控制组的相关数据。

如果有若干个干预措施,他们可以并列实施或者是交互实施。如下表所示, $T1/T2$ 为干预措施, N 和 M 表示样本量。并列实施指的是各干预措施独立作用于样本,而交互实施指的

是两个维度的干预措施交互作用于样本。交互实施有两个优点:一是能够用来探讨不同干预措施之间的交互作用,二是可以样本共用,节约样本量。在探讨 $T1$ 的影响时,控制组和干预组的样本量分别是 $M1+M2+M3$ 和 $M4+M5+M6$;同理,在探讨 $T2$ 的影响时,控制组以及干预组 1 和 2 样本量分别是 $M1+M4$ 、 $M2+M5$ 和 $M3+M6$ 。

表 1 并列实施			表 2 交互实施			
$T=0$	$T1=1$	$T2=1$		$T2=0$	$T2=1$	$T2=2$
$N0$	$N1$	$N2$	$T1=0$	$M1$	$M2$	$M3$
			$T1=1$	$M4$	$M5$	$M6$

(六) 收集实验数据

实验数据收集包括干预前数据收集和干预后数据收集。因为干预措施是随机安排的,所以在结果分析当中,即使不控制任何其他变量,依然能够得到关于干预效果的无偏性估计。但即便如此,干预前数据对随机实验仍至少具有以下四个方面的作用:一是在分组随机和匹配随机的情况下,为随机分组提供信息;二是通过干预前变量的可比性,验证干预措施安排的随机性;三是在结果变量回归分析中,增加控制变量有利于提高估计的精确度;四是检测在不同特征群体上的干预效果,探讨干预措施的异质性效果。

干预前数据的收集可以通过基线问卷调查进行,但做一轮基线问卷可能会极大地增加成本。通过后期间卷来回顾,可以用于对干预前客观信息收集。此外,合作机构出于业务等需求收集的行政性数据也是干预前数据的重要来源。陆方文等探讨学生座位干预对学习成绩的影响,基线成绩来自学校数据,而关于学生家庭的信息则来自后期间卷的回顾(Lu and Anderson, 2015)。

干预后数据收集方法日益多元化。问卷是最经常的方式,此外还有实地检测以及各种经济学游戏。在收集干预后数据过程中,非常重要的是,控制组和干预组必须使用完全相同的数据收集方式。比如,Olken (2007)在关于印度尼西亚修路的研究中,其中有一个维度的干预措施是预先告知加强审计。因为干预后数据也是从审计中得来,控制组的数据必须是在同种方式下收集。

(七) 实验中的道德问题

科学实验得遵守一定的伦理道德准则,实地实验也不例外。几个重要的原则包括:第一,不能给实验对象带来直接伤害,尤其是关于儿童的,以及涉及到进入身体的,如食品、药品、注射等。第二,避免给社会带来可能的伤害。在 Bertrand 等(2007)关于印度驾驶证的研究中,有些人用实验的激励资金购买了驾驶证。不合格的驾驶员上路,可能对社会造成伤害,因而实验完成后这些人的驾驶证被购回了。第三,不可以欺骗,除了审计实验法有例外规定(陆方文, 2014)。第四,实验对象可以随时退出,不可以强制。

四、随机实地实验的局限性和对策

随机实地实验方法因其在实证研究中能够获得较高的可信度而日益受到经济学家的青睐,同时也因为干预措施的设计和实施易于与现实社会对接,故在经济制度和政策的设计和实施中日益得到倚重。不过与其他实证方法一样,随机实地实验方法也有自身的局限性。

首先,实验可能改变行为,从而直接影响实验结果的内部有效性。“霍桑效应”就是著名的例子。此外,Zwane 等(2011)指出,进行实验前的基线问卷调查有可能影响实验结果,因为基线问卷的调查人员通常会非常有礼貌地询问对方的问题并倾听回答,从而建立起信任,而信任会影响后续的行为。对于这一问题的解决之道,一方面是要尽量地模拟现实当中操作的情况,另一方面可以设定干预措施去捕获这一效果。譬如,在 Lu 等(2016)关于交警给司机发送手机短信影响司机驾驶行为的研究中,由于收到来自交警的短信本身就可能影响司机的行为,为了更好地区分收到短信这一事件和短信内容的影响,研究者们增加了一个干预组,其短信内容是警察提醒安全驾驶,但不包含任何实质的信息。实验结果表明,提醒安全驾驶并不能影响司机的行为,因而排除了可能的霍桑效应解释。但如果提醒安全驾驶能够改变行为,在探讨具体信息对行为的影响时,需要在干预效果中将提醒的作用差分掉。

其次,实地实验在探讨局部均衡效应上具有优势,但在捕捉一般均衡效应问题上有些捉襟见肘。随机实验通常是小规模实验,捕捉的往往是局部均衡效应。比如,Angrist 等(2002)探讨哥伦比亚学费优惠项目的影响,对比摇号中拿到和没有拿到学费优惠的学生在未来求学和收入上的区别,这就是典型的局部均衡效应;而从一般均衡的视角看,还需要考虑以下情形:因为优惠项目有一定的门槛,所有学生为了获得优惠都会更加努力学习;优惠项目资助学生去上私立学校,把好学生从公立学校选拔出去,等等。关于一般均衡的问题,并非实验所特有,很多实证分析都存在类似的问题。比如,探讨医保对健康的影响,很多实证研究在有医保的家庭和没有医保的家庭之间做对比。Finkelstein (2007)指出,当一个地区医保比例增多,医生会增加设备等投入,也会影响没有医保的患者;因而在地区层面上做对比,才能够体现出一般均衡效应。

能否探讨一般均衡效应,主要取决于实验措施的干预范围和结果变量的观测范围。Crépon 等(2013)探讨就业帮扶对就业的影响,为我们提供了一个探讨一般均衡的实验范例。其干预措施在两个层次上进行随机安排:在不同城市,获得就业帮助的人数比例有 0%、25%、50%、75%和 100%之分;同一个城市,不同求职者随机地分配到获得帮扶组和没有帮扶组。该研究不仅检测获得就业帮助对求职者个人的影响,还检测一个地区获得就业帮助的比例对当地市场的影响,发现就业帮扶的社会净效益为零。

最后,实验结果的外部有效性问题。Harrison 和 List (2004)、Levitt 和 List (2007)以及黄湛冰和肖尔特(2012)对外部有效性进行了一定的探讨。外部有效性问题,首先表现在实验结果有从一个人群推广到另一个人群的问题。克服此问题的一个方法是,尽可能在更大范围上更具代表性的人群中做实验。此外,收集可能影响实验效果的背景变量,做差异性分析,也能增强实验结果解释的广度。在小规模实验向大规模政府政策推广时,外部有效性问题还会涉及到研究人员实施和公职人员推广的差异问题。Olken (2007)在印度尼西亚修路的实验中探讨审计的作用,实验中他自己聘用了审计人员,从而确保审计人员能够认真客观地审核;但政策推广时,审计人员甚至可能主动寻租。这个问题要求实验不能仅仅满足于做简单的项目评估,还要研究影响结果的核心机制。

总体上看,随机实地实验方法的这些局限很大程度上也是实验方法、甚至整个实证分析

方法的局限。即便存在这些局限,相较其他的实证分析,随机实地实验具有可信性和创新性的双重优势,这也是其能够被越来越多的经济学家重视的缘故。

五、随机实地实验在国际的发展趋势

近二十年来,随机实地实验在国外经济学领域获得迅猛的发展,涌现出大量的研究文献。早期的随机实地实验研究主要侧重经济政策的评估,且多集中于劳动经济学和发展经济学领域,其意图是测试实施单个政策的短期后果。但由于实验设计思路本身的局限,这类早期研究并没有对经济学产生广泛而深刻的影响。

随着随机实地实验方法论上的日益成熟,经济学家逐渐把这类实验方法拓展到劳动经济学和发展经济学以外的诸多领域,并不再局限于政策评估,而是把实验方法和理论研究相结合,通过随机干预措施来模拟现有的或者未来可能的制度安排和政策设计。随机实地实验研究近些年的变化可以归纳为以下几个方面特点:应用领域更广泛;机制探讨更深入;研究效果更长期;更重视干预措施的交互影响等。对于随机实地实验发展趋势的总结和归纳,有利于展现当前实地实验方法设计的前沿,促进国内实地实验相关研究的后发赶超。发展趋势详述如下:

(一) 应用领域更加广泛

随机实地实验已经从传统的劳动经济学和发展经济学等领域,扩展到机制设计、新政治经济学、公共经济学、环境经济学、国际贸易等众多经济学分支。劳动经济学和发展经济学是随机实地实验经典的运用领域,这主要是因为针对个体的、学生的和欠发达地区的实地实验成本相对较小、容易开展,而且这些领域的政策需求大。List 和 Rasul (2010) 按照人生历程分阶段地对劳动相关的实验文献进行了综述,议题覆盖了人力资本培育、就业市场、劳动激励等各个方面。在发展经济学领域,以现金转移支付为政策措施的扶贫实验在世界银行人员和资金的支持下从墨西哥开始,扩散到了大多数的拉美国家和很多非洲国家。此外,实验还涉及小额信贷、农业保险以及各种扶贫、教育、健康干预。

近年来,实地实验方法越来越多地运用到其他经济学领域。比如,在市场机制方面,Brown 和 Morgan (2009) 在 ebay 和 yahoo 两个不同的在线交易平台、并随机调整是否有最低价格、是否可延长拍卖的结束时间,探讨拍卖市场的均衡机制;在新政治经济学领域,Olken (2007) 通过在印度尼西亚的村庄修路探讨审计和公民参与对工程腐败的影响;关于税收征收,Pomeranz (2015) 通过提高审计强度探讨增值税的威慑机制;关于公共物品提供,Chen 等 (2010) 通过考察在线电影评价探讨社会比较对公共物品提供的影响;在环境监管问题上,Duflo 等 (2013) 通过改变支付主体探讨环境评估中的激励扭曲问题;在国际贸易领域,Atkin 等 (2017) 通过给小作坊安排订单探讨来料加工的外贸方式对企业经营的影响,等等。这些新的领域的拓展几乎涉及微观经济学的各大热门主题,展示出随机实地实验日益增长的影响力。

(二) 机制探讨更加深入

早期的随机实地实验侧重单纯的政策效果评估,近年来则越来越注重对政策机制的挖掘和评估,从而也更加符合经济学理论的逻辑。起初,随机实地实验主要用作政策评估的手

段,回答政策是否有效果以及效果大小的问题。但这种就事论事的探讨使得结论很难外推,因为一旦政策有一个小的调整,对政策效果的预估就可能存在很大的或然性。有鉴于此,后来越来越多的实验都更加关注对政策机制的探讨。比如,墨西哥实行有条件的现金转移扶贫项目,只有在子女上课出席率达标并且按期参加体检的条件下,政府才会给相应家庭支付扶贫款。Gertler (2004)探讨该项目对儿童教育和健康的影响,但这些影响是条件和扶贫款的共同作用。如果不设置条件要求(尤其核实条件增加了项目的执行成本),扶贫款的效果会怎样呢? Baird 等(2011)进一步对比了现金转移扶贫在有条件约束和没有条件约束这两种情况下的不同效果,指出了条件约束对辍学率和成绩的积极影响。再比如,很多文献都发现,由员工推荐的求职者要比独立的求职者更容易得到企业的雇佣,但其内在原因并不清楚。Pallais 和 Sands (2016)运用实验对为什么员工推荐可能发挥作用的三个可能机制进行探讨。他们发现,推荐在简历等文书之外提供额外信息,员工在和其推荐者一起工作的时候效率更高,但并没有发现员工会因为考虑到对推荐者的影响而更努力工作。关注机制的作用,将更有助于实验结果的外推和未来政策措施的设计。

(三) 研究效果更加长期化

过去的随机实地实验常常集中于讨论短期干预措施的效果,而近年来则更倾向于讨论长期效果的深度研究。这种转变有利于更加全面地刻画经济制度和政策的影响。一项经济制度和政策对人的行为的影响,不仅可能具有滞后效应,而且更重要的是可能改变总体的社会经济环境,从而参与人之间会重新开始一个博弈过程,这个过程趋于某个新的均衡可能需要较长时间。随机实地实验则为探讨中长期影响提供了机遇。墨西哥的有条件现金转移扶贫项目在1997年开始试行,Gertler (2004)探讨这一政策对当期儿童教育和健康的影响,Behrman 等(2011)则探讨了5年期的影响。Miguel 和 Kremer 等人在2000年左右开展了一项关于疟疾除虫的健康干预实验,并于2004发表了一篇关于除虫对当年健康和教育影响的论文。2016年他们再度探讨当年的除虫干预对于十多年之后人力资本发展的影响(Miguel and Kremer,2004; Baird et al.,2016)。这些研究为相关干预措施的中长期后果提供了科学的证据,从而使制度和政策设计的科学性大幅度提升。

(四) 更加重视干预措施的交互影响

现实中制度和政策往往具有互动特征,随机实地实验发展的新趋势是通过干预政策的组合设计,更有效地探讨干预措施的交互影响。比如,在教育激励方面,Kremer 等(2009)探讨给学生奖学金对学生成绩的影响,Duflo 等(2012)研究给老师的出勤奖对老师出勤以及最终对学生成绩的作用,此外还有很多单独研究给学生激励或者给老师激励的文章。Behrman 等(2015)则在探讨给学生奖学金和给老师绩效奖金的基础上,进一步考察二者的结合对学生成绩的影响,并且发现了显著的交互作用。这类研究无一例外都是在保证随机性的前提下,通过干预措施的合理组合,来刻画复杂制度和政策的作用机理。

六、随机实地实验在国内的发展现状和展望

随机实地实验在欧美已经得到广泛的运用,经济学家运用这种方法来帮助欠发达经济体如何改进诸如贫困、环保、节能减排等社会困境,并取得了显著的成效。对比之下,中国的

随机实地实验研究才刚刚起步,研究数量较少。代表性的有蔡洪滨、陈玉宇等人通过干预饭店的菜单考察消费者之间的社会学习,他们以及李洪彬、周黎安等人探讨牲畜保险对农民的影响(Cai et al.,2009; Cai et al.,2015)。张林秀团队在经济欠发达省份做了一系列随机实地实验,探讨提供营养品、眼镜和计算机辅助教学对学生的影响(Miller et al.,2012; Ho et al.,2013; Mo et al.,2013; Yi et al.,2015; Zhou et al.,2016)。李玲芳等探讨淘宝上的反馈机制(Li and Xiao,2014; Cabral and Li,2015)。陆方文等在学校进行了两项干预,分别探讨学生之间的同群效应以及担任班干部对学生的影响(Lu and Anderson,2015; Anderson and Lu,2016)。此外,陆方文等还对司机进行短信干预探讨如何降低交通违规(Lu et al.,2016)。

当前国内在开展随机实地实验上研究不足,一方面是由于国内科研经费的投入、尤其对年轻人的投入相对比较少,难以为实施实验研究提供充足的经费保障。另一方面,还因为在对随机实地实验方法的认识上存在着一些问题。首先,一些研究者对随机实地实验方法不太了解,不知道如何具体实施。其次,一些研究者直觉上认为做实验一定要花很多钱,因此望而生畏。与用观测数据所做的实证分析比较,实验的确需要更多的经费,而且有些实验也的确需要支出大额的经费,比如说给农民补贴、给学生免费午餐等等,实施一项实验也往往需要很长的周期,但并非所有实验都是如此。也有一些实验只需要较少的经费、较为简洁的流程和不太多的人力物力投入。最后,一些研究者仍然倾向于把随机实地实验简单地看成政策评估的工具,并没有把实验很好地融入到对理论的探讨和对机制的分析中去。解决上述的认识问题,是本文的重要目标所在。

我国正经历着深刻转型和快速发展,各种问题和矛盾以及内在的规律性都在短时期内不断产生和暴露出来。转型期的各种制度创新和政策试点层出不穷,这就形成了一个天然的大实验场,为随机实地实验方法的发展和应用提供了广阔的空间。但如果试点缺乏科学的设计和理论的支撑,没有控制和干预,试点主体的选择和政策措施存在模糊性,试点的数据采集和研究缺乏系统性,从而导致试点往往达不到预期的外部有效性。而随机实地实验的方法所强调的规范的实验方法能够解决这些问题,将有助于极大提高转型时期的制度和政策设计水平。

参考文献:

1. 陈玉梅、陈雪梅,2012:《实验经济学的新突破:实地实验方法》,《经济学动态》第6期。
2. 黄湛冰、肖尔特,2012:《经济学实验结果外部效度排序探索研究》,《经济评论》第3期。
3. 姜树广、谯倩,2012:《实地实验及其在经济学中的应用》,《经济评论》第5期。
4. 陆方文,2014:《经济学中的审计实验法研究》,《教学与研究》第4期。
5. 罗俊,2014:《田野实验——现实世界中的经济学实验》,《南方经济》第6期。
6. 罗俊、汪丁丁、叶航、陈叶烽,2015:《走向真实世界的实验经济学——田野实验研究综述》,《经济学(季刊)》第2期。
7. 张生玲、周晔馨,2012:《资源环境问题的实验经济学研究评述》,《经济学动态》第9期。
8. Al-Ubaydli, O., and J. List. 2015. "On the Generalizability of Experimental Results in Economics." In *Handbook of Experimental Economic Methodology*. Edited by Guillaume R. Fréchette and Andrew Schotter, 240–262. Oxford: Oxford University Press.
9. Anderson, M., and F. Lu. 2016. "Learning to Manage and Managing to Learn: The Effects of Student Leadership Service." *Management Science*. <http://dx.doi.org/10.1287/mnsc.2016.2483>.
10. Angrist, J., E. Bettinger, E. Bloom, E. King, and M. Kremer. 2002. "Vouchers for Private Schooling in Colombia:

- Evidence from a Randomized Natural Experiment.” *American Economic Review* 92(5):1535–1558.
11. Atkin, D., A. Khandelwal, and A. Osman. Forthcoming. 2017. “Exporting and Firm Performance: Evidence from a Randomized Experiment.” *Quarterly Journal of Economics* 132(2):551–615.
12. Baird, S., J. Hicks, M. Kremer, and E. Miguel. 2016. “Worms at Work: Long-run Impacts of a Child Health Investment.” *Quarterly Journal of Economics* 131(4):1637–1680.
13. Baird, S., C. McIntosh, and B. Ozler. 2011. “Cash or Condition: Evidence from a Randomized Cash Transfer Program.” *Quarterly Journal of Economics* 126(4):1709–1753.
14. Behrman, J., S. Parker, and P. Todd. 2011. “Do Conditional Cash Transfers for Schooling Generate Lasting Benefits? A Five-year Followup of PROGRESA /Oportunidades.” *Journal of Human Resources* 46(1):203–236.
15. Behrman, J., S. Parker, P. Todd, and K. Wolpin. 2015. “Aligning Learning Incentives of Students and Teachers: Results from a Social Experiment in Mexican High Schools.” *Journal of Political Economy* 123(2):325–364.
16. Bertrand, M., S. Djankov, R. Hanna, and S. Mullainathan. 2007. “Obtaining a Driver’s License in India: An Experimental Approach to Studying Corruption.” *Quarterly Journal of Economics* 122(4):1639–1676.
17. Bertrand, M., and S. Mullainathan. 2004. “Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination.” *American Economic Review* 94(4):991–1013.
18. Brown, J., and J. Morgan. 2009. “How Much Is a Dollar Worth? Tipping versus Equilibrium Coexistence on Competing Online Auction Sites.” *Journal of Political Economy* 117(4):668–700.
19. Cabral, L., and L. Li. 2015. “A Dollar for Your Thoughts: Feedback-Conditional Rebates on eBay.” *Management Science* 61(9):2052–2063.
20. Cai, H., Y. Chen, and H. Fang. 2009. “Observational Learning: Evidence from a Randomized Natural Field Experiment.” *American Economic Review* 99(3):864–882.
21. Cai, H., Y. Chen, H. Fang, and L. Zhou. 2015. “The Effect of Microinsurance on Economic Activities: Evidence from a Randomized Field Experiment.” *Review of Economics and Statistics* 97(2):287–300.
22. Chen, Y., M. Harper, J. Konstan, and S. Li. 2010. “Social Comparisons and Contributions to Online Communities: A Field Experiment on MovieLens.” *American Economic Review* 100(4):1358–1398.
23. Crépon, B., E. Duflo, M. Gurgand, R. Rathelot, and P. Zamora. 2013. “Do Labor Market Policies Have Displacement Effects? Evidence from a Clustered Randomized Experiment.” *Quarterly Journal of Economics* 128(2):531–580.
24. Duflo, E., R. Glennerster, and M. Kremer. 2008. “Using Randomization in Development Economics Research: A Toolkit.” *Handbook of Development Economics* 4:3895–3962.
25. Duflo, E., M. Greenstone, R. Pande, and N. Ryan. 2013. “Truth-telling by Third-party Auditors and the Response of Polluting Firms: Experimental Evidence from India.” *Quarterly Journal of Economics* 128(4):1449–1498.
26. Duflo, E., R. Hanna, and S. Ryan. 2012. “Incentives Work: Getting Teachers to Come to School.” *American Economic Review* 102(4):1241–1278.
27. Finkelstein, A. 2007. “The Aggregate Effects of Health Insurance: Evidence from the Introduction of Medicare.” *Quarterly Journal of Economics* 122(3):1–37.
28. Gertler, P. 2004. “Do Conditional Cash Transfers Improve Child Health? Evidence from PROGRESA’s Control Randomized Experiment.” *American Economic Review* 94(2):336–341.
29. Guala, F. 2005. *The Methodology of Experimental Economics*. New York: Cambridge University Press.
30. Guéguen, N. 2015. “High Heels Increase Women’s Attractiveness.” *Archives of Sexual Behavior* 44(8):2227–2235.
31. Harrison, G., and J. List. 2004. “Field Experiments.” *Journal of Economic Literature* 42(4):1009–1055.
32. Ho, L., R. Luo, L. Zhang, and S. Rozelle. 2013. “The Impact of Vouchers on Preschool Attendance and Elementary School Readiness: A Randomized Controlled Trial in Rural China.” *Economics of Education Review* 35:53–65.
33. Kremer, M., E. Miguel, and R. Thornton. 2009. “Incentives to Learn.” *Review of Economics and Statistics* 91(3):437–456.
34. Levitt, S., and J. List. 2007. “Viewpoint: On the Generalizability of Lab Behavior to the Field.” *Canadian Journal of Economics* 40(2):347–370.
35. Levitt, S., and J. List. 2009. “Field Experiments in Economics: The Past, the Present, and the Future.” *European Economic Review* 53(1):1–18.
36. Li, L., and E. Xiao. 2014. “Money Talks? An Experimental Study of Rebate in Reputation System Design.” *Management Science* 60(8):2054–2072.
37. List, J. 2011. “Why Economists Should Conduct Field Experiments and 14 Tips for Pulling One off.” *Journal of*

- Economic Perspectives* 25(3): 3–16.
38. List, J., and I. Rasul. 2010. “Field Experiments in Labor Economics.” *Handbook of Labor Economics* 4A: 103–228.
39. Lu, F., and M. Anderson. 2015. “Peer Effects in Microenvironments: The Benefits of Homogeneous Classroom Groups.” *Journal of Labor Economics* 33(1): 91–122.
40. Lu, F., J. Zhang, and J. Perloff. 2016. “General and Specific Information in Detering Traffic Violations: Evidence from a Randomized Experiment.” *Journal of Economic Behavior & Organization* 123(2): 97–107.
41. Miguel, E., and M. Kremer. 2004. “Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities.” *Econometrica* 72(1): 159–217.
42. Miller, G., R. Luo, L. Zhang, S. Sylvia, Y. Shi, P. Foo, Q. Zhao, R. Martorell, A. Medina, and S. Rozelle. 2012. “Effectiveness of Provider Incentives for Anaemia Reduction in Rural China: a Cluster Randomized Trial.” *British Medical Journal* 345: e4809. <https://doi.org/10.1136/bmj.e4809>.
43. Mo, D., J. Swinnen, L. Zhang, H. Yi, Q. Qu, M. Boswell, and S. Rozelle. 2013. “Can One-to-One Computing Narrow the Digital Divide and the Educational Gap in China? The Case of Beijing Migrant Schools.” *World Development* 46: 14–29.
44. Olken, B. 2007. “Monitoring Corruption: Evidence from a Field Experiment in Indonesia.” *Journal of Political Economy* 115(2): 200–249.
45. Pallais, A., and E. Sands. 2016. “Why the Referential Treatment? Evidence from Field Experiments on Referrals.” *Journal of Political Economy* 124(6): 1793–1828.
46. Pomeranz, D. 2015. “No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax.” *American Economic Review* 105(8): 2539–2569.
47. Rubin, D. 1974. “Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies.” *Journal of Educational Psychology* 66(5): 688–701.
48. Yi, H., H. Zhang, X. Ma, L. Zhang, X. Wang, L. Jin, K. Naidoo, H. Minto, H. Zou, L. Lu, S. Rozelle, and N. Congdon. 2015. “Impact of Free Glasses and a Teacher Incentive on Children’s Use of Eyeglasses: A Cluster-Randomized Controlled Trial.” *American Journal of Ophthalmology* 160(5): 889–896.
49. Zwane, A., J. Zinman, E. Dusen, W. Pariente, C. Null, E. Miguel, M. Kremer, D. Karlan, R. Hornbeck, X. Gine, E. Duflo, F. Devoto, B. Crepon, and A. Banerjee. 2011. “Being Surveyed Can Change Later Behavior and Related Parameter Estimates.” *Proceedings of the National Academy of Sciences* 108(5): 1821–1826.
50. Zhou, H., S. Sun, R. Luo, S. Sylvia, A. Yue, Y. Shi, L. Zhang, A. Medina, and S. Rozelle. 2016. “Impact of Text Message Reminders on Caregivers’ Adherence to a Home Fortification Program Against Child Anemia in Rural Western China: A Cluster-Randomized Controlled Trial.” *American Journal of Public Health* 106(7): 1256–1262.

Randomized Field Experiment: Method, Trend and Prospect

Lu Fangwen

(School of Economics, Renmin University of China)

Abstract: As an important method for identifying causality in social phenomena, randomized field experiments is being applied more and more widely in economics globally, however, the application of this method in the domestic academia has just started. Combining the frontier literature and my own research experience, this paper first introduces credibility and innovation as the two advantages of randomized field experiments. It then discusses systematically the key technical aspects of randomized field experiments, which is expected to reduce the technical barriers for conducting experiments. The paper also summarizes the global trends of randomized field experiments, which will help to guide the domestic research to realize the catching up and surpassing, and to promote the development of randomized field experiments in China.

Keywords: Randomized Field Experiment, Method, Trend

JEL Classification: C9, C8, B4

(责任编辑:彭爽)