

DOI:10.3969/j.issn.1674-8131.2012.04.06

基于微观非实验数据的政策 效应评估方法评价与比较*

卫梦星

(中国社会科学院研究生院,北京 102488)

摘要: 社会政策是政府发挥职能的重要手段,对政策实施的效果进行量化评估,有助于政府决策的科学化、合理化。本文基于非实验数据,归纳了目前微观政策领域各种致力于构建“反事实”的政策效应评估方法,包括工具变量法(IV)、断点回归(RD)、双重差分(DID)和匹配方法(Matching),阐述了其各自的优势、劣势及适用条件,并对实际政策评估中各模型的选用进行了简要述评。

关键词: 政策效应评估;反事实;准实验方法;匹配方法;工具变量法;断点回归;双重差分

中图分类号: F224 **文献标志码:** A **文章编号:** 1674-8131(2012)04-0042-08

Review and Comparison of Researches on Policy Evaluation Based on Micro Non-experimental Data

WEI Meng-xing

(Graduate School, Chinese Academy of Social Science, Beijing 102488, China)

Abstract: Social policy is an important means for governments to realize their function. Quantitative evaluation on the effect of policy implementation can help the governments make decisions scientifically and reasonably. Based on non-experimental data, this paper summarizes the evaluation methods for all kinds of the “counterfactual” policy effects deliberately constructed in current micro-policy field, such as instrumental variable method, regression discontinuity method, difference-in-difference method and matching method, elaborates the advantages, disadvantages and applicable conditions of each method, and makes brief review of the selection of each model in the practice of policy evaluation.

Key words: policy effect evaluation; counterfactual; quasi-experiment method; matching method; instrumental variable method; regression discontinuity method; difference-in-difference method

进行政策评价是计量经济学研究的核心目的之一。中国经济正处于高速发展阶段,社会改革也在逐步深化,政策、规则的改变频频发生,这为政策评价方法与理论的发展提供了一个绝佳的背景。

对政策实施的效果进行量化评估,有助于政府决策的科学化、合理化。通过构建“反事实”对政策效应进行评估是当前政策评价研究中的热点,从搜集到的文献来看,目前的研究成果主要集中在以劳动经

* 收稿日期:2012-06-06;修回日期:2012-07-01

基金项目: 中国社会科学院重大课题(YZDA2010)“我国应对金融危机的政策效应分析及退出机制设计”;中国社会科学院经济政策与模拟重点实验室项目

作者简介: 卫梦星(1984—),女,山西运城人;博士研究生,在中国社会科学院数量经济与技术经济研究所学习,主要从事政策评价研究;Tel:15201294715, E-mail: weimengxing1223@126.com。

济学为主的微观领域,在宏观领域并未取得太多进展。有鉴于此,本文将对国内外微观政策评估领域的研究方法做一综述,阐述各种方法在应用上的优势及劣势,总结其适用条件,并对实证研究中各模型的选用进行简要述评。

一、政策效应评估的基本问题

1. “反事实”

假定总体中有两组或多组群体,能够在两期或多期观测到其中的个体数据,并且在某些时期某些个体受到了一项新政策的“干预”(treatment)。在政策评估中,通常把接受政策干预的样本称为实验组,把未受政策干预的样本称为控制组。要估计一项政策的实施效应,一个简单的逻辑就是对所有个体在政策实施前后的结果变量做比较,将差值作为个体的处置效应(treatment effect),进而可推知政策实施效应。

在二元选择的情况下,定义指示变量 D ,当 $D_i = 1$ 时,认为个体 i 选择接受政策的影响;反之,当 $D_i = 0$ 时则视为个体选择不接受政策影响。 D_i 由一系列个体特征因素决定,不同个体之间的选择相互独立。若定义时期 0 为政策未实施的时期,1 为政策实施后的时期,则个体 i 的处置效应可以写作:

$$\tau_i = Y_i(1) - Y_i(0)$$

但由于道德因素、社会成本等原因,现实中的政策只有极少数采用随机实验的方式实施,研究者可观测到的数据多是非实验数据。这就意味着社会实验政策实施以后,研究者不可能同时观测到同一个体在接受和不接受政策干预时的表现。对于实验组中的个体,研究者无法观测其在未接受政策干预时的表现;而对于控制组中的个体,研究者也无法预测其在接受政策干预时的表现。也就是说,存在着两组“反事实”(counterfactual)。因此,如何构建“反事实”就成为政策效应评估中的关键。本文所讨论的各种评估方法的不同,本质上是其构建“反事实”的方法不同。

2. 个体的异质性决策

在政策实践中,个体并非通过随机委派的方式接受政策干预。多数情况下政策的实施是区域性的,区域中的个体会预测自己接受政策干预时可能获得的净收益,进而通过迁移等反应来决定是否接受政策影响。也就是说,个体存在私人信息,并根据私人信息进行参与决策。

个体的私人信息与个体特征紧密相关,具有异质性,不仅会影响个体的参与决策,间接影响结果变量,还可能直接对结果变量施加影响,是混杂因素(confounding factor)。如果研究者不能控制混杂因素对个体结果变量的直接影响,所测得的政策效应就会出现偏差,称为选择偏差(selection bias)。不可观测的混杂因素的存在是选择偏差产生的主要原因。但在政策评估的实践中,抽样方法、样本选择等的不规范会使得控制组与实验组的可观测变量分布并不相同,这时就会出现可观测变量的不匹配引起的偏差。Heckman 等(1997)指出,可观测变量的不匹配引起的偏差比不可观测变量引起的偏差要大得多。

个体异质性决策以及混杂因素的存在,使得微观政策效应的评估既要解决内生性问题,又要解决混杂因素的干扰问题,传统上仅将结果变量与协变量和指示变量 D 做简单回归的方法就不再可行了,必须寻找新的方法来解决这一问题。

3. 相关统计量

如上分析可知,由于存在着“反事实”,估计单个个体的处置效应是不可能的,只能关注平均处置效应(average treatment effect, ATE)或实验组的平均处置效应(average treatment effect on the treated, ATT)。其中,ATE 表示从总体中随机选择一个具有特征 X 的个体接受干预时所能获得的平均回报,ATT 表示通过自我选择接受干预的个体相比于其未接受干预时能获得的平均回报。ATT 与 ATE 回答了两个完全不同的问题,一般情况下二者是不同的。但当政策对所有个体的影响相同,或者当受干预的个体是随机选取,且研究者关注政策对总体的平均影响时,ATT 与 ATE 等价。ATE 与 ATT 的计算方法分别如下所示:

$$\tau_{ATE} = E(\tau) = E[Y(1) - Y(0)]$$

$$\begin{aligned} \tau_{ATT} &= E(\tau | D = 1) \\ &= E[Y(1) | D = 1] - E[Y(0) | D = 1] \end{aligned}$$

此外,Bjorklund 等(1987)还提出了边际政策效应(marginal treatment effect, MTE)的概念。MTE 指处于接受或不接受干预临界状态的个体最终选择接受政策干预时的平均回报。可见,研究者要根据实际需要选择合适的估计量进行测算。

因为 ATE 所包含的研究对象来自于总体,包含了很多实际未被纳入政策实施范围的个体,MTE 将处于临界状态的个体作为研究对象,二者都不是决

策者关注的重点。ATT 直接将实际接受政策干预的个体的平均处置效应作为关注点,符合决策者的政策意图,有利于评估目标群体从政策中获得的平均回报,从而帮助判断政策是否达到预期目标。因此,实际中 ATT 的应用较为普遍。

二、使用非实验数据构建“反事实”的方法

根据数据产生方式的不同,微观政策效应评估的方法可分为随机受控实验(randomized controlled trial)、准实验(quasi-experiment)以及匹配(matching)等方法。

1. 随机受控实验

从某种意义上讲,随机受控实验是最可信的一种政策评估方法。它通过将符合项目参与条件的部分个体随机地排除至项目之外,为实验组对象提供一个自动生成的控制组。但是,随机受控实验有其自身的局限性:(1)在经济中很少遇到,且实施成本高,时间长;(2)不能用于外推,不能轻易用作政策实施前的决策依据;(3)实验要求控制组完全不受政策的影响,但实际上很难排除政策的溢出效应、替代效应等。基于以上原因,随机受控实验在实践中应用很少。尽管如此,这一方法仍可帮助我们判断其它评估方法的效力。

2. 准实验

准实验方法将政策改革视为一项实验,试图在精心设计的实验条件下,为实验组找到一个自然产生的控制组。准实验方法包括工具变量(instrumental variables, IV)、双重差分(difference-in-differences, DID)以及断点回归(regression discontinuity, RD)等方法。其中,IV 是该领域最早使用的方法;RD 在 2000 年后才开始使用,目前使用还较少;DID 是目前应用最多的方法,可信度也较高。

3. 匹配方法

匹配方法是利用非实验数据进行政策评估时常用的方法。匹配的目的非常简单,就是要找出足够的可观测变量,找到与实验组个体足够相近的个体作为控制组,二者结果的不同完全取决于是否接受政策干预。这样,如果每个实验组个体都可以找到一个或多个与其具有相同协变量且未参与项目的控制组个体相匹配,那么实验组的每个个体的处置效应就可以计算,ATT 也可相应地计算出来。

计量经济学方法依托于数据,不同来源的数据

具有不同的特征。在微观经济政策评估领域,常处理的数据类型主要有以下几种:截面数据(cross-section data)、纵向数据(longitudinal data)和重复截面数据(repeated cross-section data)。依托非实验数据进行政策评估时,必须根据数据特征选择恰当的评估方法。数据结构某种程度上决定了方法的选择,而数据质量则关乎实证结果的可靠性和正确性。由于随机受控实验在实际中较少遇到,本文将着重探讨准实验方法及匹配等其他方法的原理及使用。

三、利用非实验数据进行微观政策效应评估

1. 工具变量法

标准的计量经济学提供了一种处理内生性问题的方法——IV 法。Ehrlich(1975, 1977)运用时间序列数据和截面数据就美国执行死刑对降低谋杀率的影响进行的研究具有典型性。Ehrlich 认识到谋杀率与死刑执行率之间的双向因果关系,并试图应用 IV 来解决其内生解释变量和遗漏解释变量的问题。他选择了此项政策支出的滞后量、总的政府支出、人口、非白人比例等变量作为 IV,但并没有解释为什么这些变量是好的 IV,所选出的这些 IV 与内生的解释变量之间又具有怎样的关联。直至 Ehrlich(1987, 1996)的研究出版,其选择 IV 的考虑及相关的因果识别问题才得到详细的阐述。Angrist(1990)和 Angrist 等(1991)分别用 IV 研究了参加越战对老兵收入的影响和教育背景对收入的影响,从而充分显现了运用 IV 进行因果推断的价值。Card 等(1992a, 1992b)将学生的出生州与出生队列作为 IV,研究了教育投入对教育质量的影响,从而使得教育产出、教育质量领域的研究出现了重大转折。Bound 等(1995)指出了 Angrist 等(1991)研究中存在的弱工具变量的问题,从而将 IV 的效率问题以及 IV 的选取准则引入研究。此后,有关 IV 研究的理论问题都主要集中在如何寻找最优的工具变量上。

工具变量法是一个相对简单的估计方法,但是有两个重要的缺陷:(1)工具变量的选择问题。在政策评估问题中,要找出满足条件的工具变量并不容易。在实践中,尤其是当纵向数据和政策实施前的数据可以获得时,研究者多使用因变量的滞后变量作为工具变量。但是,这同样会引发相关性,并不能从根本上解决问题。(2)如果个体对于政策的

反应不同,只有当个体对政策反应的异质性并不影响参与决策时,工具变量才能识别 ATT、ATE。但这是一个很强的假定,有时研究者不得不假定非理性,或者忽略研究对象的行为异质性(Heckman, 1997)。

2. 断点回归

RD 是一种类似于随机受控实验的准实验方法。它的主要思想是:当个体的某一关键变量的值大于临界值时,个体接受政策干预;反之,则不接受政策干预。一般而言,个体在接受干预的情况下,无法观测到其没有接受干预的情况。而在 RD 中,小于临界值的个体可以作为一个很好的控制组来反映个体没有接受干预时的情况,尤其是在变量连续的情况下,临界值附近样本的差别可以很好地反映干预和结果变量之间的因果联系,进而计算出 ATE、ATT 等政策效应变量。

RD 最早是由美国西北大学的心理学家 Campbell 于 1958 年首先发展设计出来的,但一直没有得到广泛的应用。Hahn 等(2001)为断点回归的模型识别和模型估计进行了严格意义上的理论证明,并提出了相应的估计方法。自此之后, RD 在经济学上的应用才开始盛行。到目前为止,对这一方法的研究成果还主要集中在劳动经济学领域。国内学者运用 RD 进行分析的研究还比较少见,余静文等(2010a)研究了城市圈所产生的集聚效应、辐射效应,考察了其对于地区收入差距的影响;余静文等(2010b)还使用该方法考察了城市圈对区域经济效益的影响,发现城市圈产生的辐射效应和政府治理结构转变所带来的效应会促进城市圈区域的经济的发展。

RD 是一种类似于随机实验的方法,也是准实验方法中最具有可信性的方法。Lee(2008)认为在随机实验不可得的情况下,断点回归能够避免参数估计的内生性问题,从而真实反映出变量之间的因果关系。RD 方法应用的关键假设是要求在断点附近的个体的特征相同,这一假设可以通过统计分析得到检验。由此可见, RD 的吸引力不仅在于它的实验性,还在于它的因果推断可以方便地得到检验。

3. 双重差分

近年来, DID 在政策评估研究得到了广泛应用。DID 处理选择偏差的基本思想是:允许存在不可观测因素的影响,但假定它们是不随时间变化的。假

定不可观测因素 U_{it} 可分解为 $U_{it} = \varphi_i + \theta_t + \mu_{it}$, 其中 φ_i 是个体固定效应,不随时间变化; θ_t 是个体所处的共同的环境带来的效应,对于所有个体而言都相同; μ_{it} 是个体时点效应。DID 假定实验组和控制组在研究的区间内具有相同的个体时点效应,也就是说 μ_{it} 相同,因此通过对截面单位在项目实施前后的结果取差值,就能排除 φ_i 、 θ_t 的影响。反之,若在政策实施条件下,个体时点效应 μ_{it} 不相同,则 DID 就不再是一致估计量。

应用 DID 评估政策效应的基本步骤是利用面板数据建立双固定效应模型并估计参数:

$$Y_{it} = \beta_0 + \beta_1 T_{it} + \beta_2 A_{it} + \beta_3 T_{it} A_{it} + \varepsilon_{it}$$

其中, $T_{it} = 1$ 表示实验组对象,反之则表示控制组对象; $A_{it} = 1$ 表示政策实施后的区段,反之亦然; $T_{it} A_{it}$ 是交叉项,其系数 β_3 表示实验组对象在接受政策后结果变量的变动程度,反映了政策变动的效应,是目标变量。参数 β_1 表示没有政策干预时,实验组与控制组的经济行为如何随时间变动,而参数 β_2 则反映实验组和控制组中任何不随时间变动的差异。应用 DID 评估政策效应的一个关键假设是:当不存在政策干预时, $\beta_3 = 0$, 这一假设只有在实验组与控制组性质非常接近时才是合理的。表 1 概括了 DID 方法的基本原理。

表 1 DID 方法的基本原理

	政策 变化前	政策变化后	Difference
Treatment	$\beta_0 + \beta_1$	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	$\Delta Y_t = \beta_2 + \beta_3$
Control	β_0	$\beta_0 + \beta_2$	$\Delta Y_c = \beta_2$
Difference	/	/	$\Delta \Delta Y = \Delta Y_t - \Delta Y_c = \beta_3$

Heckman 等(1985, 1986)最早提出使用 DID 方法对社会公共政策的实施效应进行评估,此后对 DID 方法的研究和应用成果层出不穷,典型的有: Card 等(1990, 1994)对移民政策、最低工资制度对工资和就业的影响进行了研究; Puhani(2000)对波兰 1991 年实施的失业救济政策改革对失业持续期的影响进行了评估; Stewart(2004)对英国 1999—2001 年引入的最低工资制度对就业的影响进行了评估; Donohue 等(2005)发现美国与加拿大的谋杀率之间具有相同的变化趋势,从而以取消死刑的加拿大作为控制组,评估了美国恢复死刑制度对降低

谋杀率的影响,结果表明美国执行的死刑政策并未对社会谋杀发生率起到遏制作用;Chen 等(2008)使用中国 2000 个家庭的数据对世界银行发展项目的效应进行了评估。

国内学者近年来也开始运用 DID 方法对政策效应进行评估,主要的研究有:周黎安等(2005)就农村税费改革对农民收入增长所产生的影响进行了评估;朱宁宁等(2008)对我国建筑节能政策的实施效应进行了评估;黄清(2009)对 2002—2005 年电力行业放松规制的政策效应进行了实证检验和研究;刘生龙等(2009)评估了西部大开发对于西部地区经济增长及中国区域经济收敛的作用;聂辉华等(2009)使用全国层面的企业数据,胥佚萱等(2011)使用上市公司数据分别对 2004 年开始在东北地区实行的增值税转型政策的影响进行了研究;俞红海等(2010)基于上市公司数据,对股权分置改革的有效性进行了实证分析;李楠等(2010)利用中国工业行业数据,对国有企业改革的绩效进行了评估。

DID 方法允许不可观测因素的存在,而且允许不可观测因素对个体是否接受干预的决策产生影响,从而放松了政策评估的条件,使得政策评估的应用更接近于经济现实,因而应用更广。但是,研究者在应用中也应该充分认识到 DID 方法的局限性:

(1)数据要求更加苛刻。DID 方法以面板数据模型为基础,不仅需要横截面单位的数据,还需要研究个体的时间序列数据,特别是政策实施前的数据。因此,相比于 Matching, DID 方法要求更多的数据。

(2)个体时点效应 μ_{it} 未得到控制。DID 要求很强的识别假设,它要求在政策未实施时,实验组和控制组的结果变量随时间变化的路径平行,这一假设并没有考虑个体时点效应 μ_{it} 的影响。由于 μ_{it} 的影响,在项目实施前后,实验组和和控制组个体行为的结果变量并不平行,此时应用传统的 DID 方法就会出现系统性误差。

(3)未考虑个体所处的环境对个体的不同影

响。DID 方法假定环境因素的冲击对处于相同环境中的个体会产生相同的影响,即 θ_t 对所有个体都相同。但实际中,实验组和控制组个体可能因为某些不可观测因素的影响,使得其在面临相同的环境因素的冲击时做出不同的反应,此时 DID 的应用就会出现

问题。针对以上问题,国外学者在使用 DID 的过程中,逐步对其进行了扩展,扩展的方向主要有两个:一是考虑 DID 中未控制的因素,从而进一步放松其应用条件;二是将 DID 与 Matching 等其他政策评估方法结合起来,提出新的估计量。比如: Bell 等(1999)考虑了个体所处的环境对个体的不同影响,提出了经趋势调整的估计量;Heckman 等(1997)提出了“条件 DID”这一新的估计量(conditional DID estimator),将 Matching 与 DID 方法结合起来应用,不仅能大大降低选择偏差,且结果更为可信。但不容忽视的是,条件 DID 仍要满足“共同支撑域”假定。

4. 匹配方法

匹配是一种非实验方法,是对于一些没有采用或不方便采用实验方法区分实验组和控制组的数据采用的一种近似实验的方法。匹配方法假定,控制协变量之后,具有相同特征的个体对政策具有相同的反应。换句话说,不可观测因素不影响个体是否接受政策干预的决策,选择仅仅发生在可观测变量上。因此,对每一个实验组个体而言,可以根据可观测特征为其选择一个控制组个体构成反事实。

在实证分析中,根据选择控制组时匹配方法的不同,Matching 又可分为协变量匹配(covariant Matching, CVM)和倾向得分匹配(propensity score matching, PSM)等。其中,CVM 涉及多个协变量,会导致“维度灾难”、计算过于复杂等问题。Rosenbaum 等(1983)指出,如果协变量能使得条件独立假设(conditional independence assumption, CIA^①)成立,那么倾向得分(propensity score, PS)作为协变量的一个

① CIA 是由 Lechner 1999 年提出的概念,又被称为 unfoundedness (Rosenbaum et al, 1983)或 selection on observables (Heckman et al, 1985)。匹配方法的基本前提可表示为: $Y(0), Y(1) \perp D \mid X$ 或者 $Y(0), Y(1) \perp D \mid P(X)$, 二者分别是 CVM 和 PSM 下条件独立假设的不同表示形式。CIA 意味着所有能同时影响干预分配机制和结果变量的混杂因素都能被观察到,这样就可消除由不可观测因素引起的选择偏差,进而估计政策效应。

函数,当然也能使得 CIA 成立。通过将协变量中蕴含的信息转移至 PS 中,PSM 可以克服 CVM 的劣势,成功降维,从而在实践中应用更多。

Rusenbaum 等(1983)提出的利用 PSM 来消除混杂因素所引起的偏差,在提出之初并没有受到很大关注,但是近些年被广泛应用于医药、经济、政策评估等领域,成为政策效应评价中最常用的方法。Perkins 等(2000)讨论了此方法在流行病药理学上的应用,Gilligan 等(2007)对在埃塞俄比亚农村实施的应急食品救援政策的效应进行了评估,Sandra 等(2009)对法国的一项再就业培训项目的效应进行了评估。国内学者近年来也开始运用 PSM 方法对社会公共政策的效应进行评估。陈玉萍等(2010)运用 PSM 方法研究了滇西南山区改良陆稻技术的采用对农户收入的影响,解决了应用传统方法分析农业技术效应时,因农户技术采用存在的自我选择而带来的因果干涉问题;李佳路(2010)运用 PSM 方法,采用 S 省 30 个国家扶贫开发重点县 2009 年的农村贫困监测数据,对扶贫项目的减贫效应进行了评估。

PSM 作为非参数方法,不需要对可观测因素的条件均值函数和不可观测因素的概率分布进行假设,因而相比参数方法具有优势。但是,PSM 也有局限性,主要表现为以下几点:

(1)极强的前提假设。PSM 的应用必须满足 CIA 和共同支撑域假定,这两个假定合起来称为“强可忽略性”假设。一旦违背这一假定,ATE 和 ATT 就会出现偏误。Heckman 等(1997)通过假设检验表明当“强可忽略性”假定不满足时,不宜应用 PSM 的方法对政策效应进行评估;Heckman(2008)通过理论分析认为当存在未观测到的混杂因素时,PSM 方法不仅不能消除系统误差,反而会带来新的偏差;刘凤芹等(2009)运用蒙特卡罗模拟实验的结果也表明 PSM 对强可忽略性假设非常敏感,即使是轻度的违背,PSM 的估计结果偏差也超过 50%;Kannika 等(2010)运用实际数据,对比参数方法与 PSM 方法的结果,进一步验证了 PSM 的应用需要满足“强可忽略性”假定。

(2)不能为所有的实验组个体找到对照组个体。匹配方法仅能为处在共同支撑域上的个体找

到合适的对照个体。如果对于不同个体而言,处置效应是同质的,那么共同支撑域的假定不会对政策效应的大小造成影响;反之,如果处置效应是不同质的,共同支撑域的假定使得某些实验组个体很难找到“反事实”,处置效应无法识别。换句话说,如果匹配过程损失了大量的观察值,处置效应的估计量就仅在共同支撑域上具有一致性特征。在异质性响应中,如果实验组个体的处置效应差别很大,估计出的 ATT 就不能代表政策的平均回报。

(3)数据量要求极大。Matching 方法往往应用于截面数据,为了保证条件独立假设成立,需要尽可能多地搜集协变量信息,将混杂因素分离出来。同时,为了保证能找到与实验组个体特征最为接近的控制组,研究者也需要收集大量的个体数据,以保证结果的精度。

(4)结果的稳健性受到多种挑战。PSM 方法计算得到的 ATE 或 ATT 的稳健性受到多种因素的影响,如干预分配机制方程的设定、匹配算法的选择等。刘凤芹等(2009)运用蒙特卡罗模拟实验的结果表明,PSM 对误差项分布不敏感,对隐指标方程的误设极为敏感;在共同支撑域较小时,PSM 对具体匹配方法的选择极其敏感。

综上所述,Matching 的应用必须满足很强的假设前提,并且要具有相当的数据量。如果研究者认为无法验证强可忽略性假定,手头的数据样本又不够大,就必须选用其他的政策评估方法,如 DID 和样本选择模型等,它们都明确允许有未被观测到的混杂因素的存在。

四、研究评述

本文对政策评价的基本原理做了简单阐述,指出构建“反事实”是政策效应评估的核心。目前利用微观非实验数据构建“反事实”的方法主要有 RD、IV、DID 和 Matching 等方法,其中,RD 类似于随机实验,因果推断最为清晰,结果最为可信,假设的可检验性也最强,但适用的范围还较小;IV、DID、Matching 是目前微观政策评价中的常用方法。本文从适用数据类型、对混杂因素的处理以及参数识别三方面对三种方法做简要比较,如表 2 所示。

表2 基于微观非实验数据的政策效应评估方法比较

方法	适用数据类型	对不可观测因素的假定及处理	异质性处置效应下参数的识别
IV	截面数据	不存在未被观测到的混杂因素,个体的参与决策都基于可观测变量。	仅当个体对政策反应的异质性不影响参与决策时,IV才能识别ATT、ATE,否则无法识别。
	纵向数据		
	重复截面数据		
DID	纵向数据	允许有未被观测到的混杂因素存在,且对个体的参与决策产生影响。	可以识别ATT,但无法识别ATE。
	重复截面数据		
Matching	截面数据	不存在未被观测到的混杂因素,个体的参与决策都基于可观测变量。	仅在共同支撑域上可以识别ATT,无法识别ATE。
	纵向数据		

数据类型及容量对方法的选用至关重要。IV法多用于截面数据,但在纵向数据和重复截面数据中也有应用;(条件)DID方法对重复截面数据的处理效果最佳,但必须能够获得政策实施前的相关数据;Matching能很好地应用于截面数据和纵向数据,但需要大量的个体数据作为支撑,当数据量充足时,简单的PSM模型也能获得理想的政策效应参数。仅就数据要求而言,Matching所需的数据量最大。

不同方法对不可观测因素的处理也不相同。IV和Matching假定不存在未被观测到的混杂因素,所有能同时影响干预分配机制及结果变量的混杂因素都被观测到,个体接受或不接受政策干预都取决于可观测变量。从这个意义上讲,IV和Matching是基于可观测变量的评估方法。而DID则容忍不可观测因素的存在,且允许其对个体决策产生影响,通过假定及差分消除不可观测因素的影响。从对不可观测因素的处理而言,DID更为严谨。

当处置效应同质时,IV、DID和Matching均可识别ATT、ATE,但当处置效应异质时,各种方法对ATT、ATE的识别情况各不相同。此外,Matching是非参数方法,不用设定方程,更具一般性;DID方法不用施加严格的外生假设,只要处置效应能以可加方式进入结果方程,甚至都不用定义结果方程。相比而言,IV需要对方程施加严格的外生假定,实证分析中需要做更多的检验。

综上所述,各种方法的适用条件各有不同,优劣各异,在政策效应评估实践中,研究者必须以经济理论为基础,根据政策本身的特点,结合所能获取的数据特征选择适用的方法,并选择符合研究目的的政策效应参数进行识别和估计。

参考文献:

- 陈玉萍,吴海涛,陶大云,等. 2010. 基于倾向得分匹配法分析农业技术采用对农户收入的影响——以滇西南农户改良陆稻技术采用为例[J]. 中国农业科学(5):3667-3676.
- 黄清. 2009. 电力行业放松规制改革政策效果的实证研究——基于发电业数据的双重差分模型检验[J]. 山西财经大学学报(1):49-56.
- 刘凤芹,马慧. 2009. 倾向得分匹配方法的敏感性分析[J]. 统计与信息论坛(10):7-13.
- 刘生龙,王亚华,胡鞍钢. 2009. 西部大开发与区域经济收敛[J]. 经济研究(9):94-105.
- 李佳路. 2010. 扶贫项目的减贫效果评估:对30个国家扶贫开发重点县调查[J]. 改革(8):125-132.
- 李楠,乔榛. 2010. 国有企业改制政策效果的实证分析——基于双重差分模型的估计[J]. 数量经济技术经济研究(2):3-21.
- 聂辉华,方明月,李涛. 2009. 增值税转型对企业行为和绩效的影响[J]. 管理世界(5):17-35.
- 胥佚莹,林志伟. 2011. 增值税转型改革与企业固定资产投资决策——基于中国上市公司数据的面板双重差分模型分析[J]. 税务与经济(1):90-97.
- 俞红海,徐龙炳. 2010. 股权分置改革有效改善了公司绩效吗?——基于双重差分模型的估计[J]. 浙江工商大学学报(1):56-62.
- 余静文,赵大利. 2010a. 城市群崛起、经济绩效与区域收入差距[J]. 中南财经政法大学学报(4):30-38.
- 余静文,王春超. 2010b. 转型时期中国城市圈的经济绩效[J]. 当代经济科学(5):15-22.
- 周黎安,陈烨. 2005. 中国农村税费改革的政策效果:基于双重差分模型的估计[J]. 经济研究(8):44-53.
- 朱宁宁,朱建军,刘思峰,等. 2008. 我国政府建筑节能政策(措施)的实施效果评价[J]. 中国管理科学(10):576-580.
- ANGRIST, JOSHUA D. 1990. Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security

- Administrative Records [J]. *American Economic Review*, 80: 313-336.
- ANGRIST, JOSHUA D, ALAN B K. 1991. Does Compulsory School Attendance Affect Schooling and Earnings [J]. *Quarterly Journal of Economics*, 106: 976-1014.
- BELL B, BLUNDELL R, VAN REENEN J. 1999. Getting the unemployed back to work: an evaluation of the New Deal proposals [J]. *International Tax and Public Finance*, 6: 339-360.
- BJORKLUND A, MOFFITT R. 1987. The Estimation of Wage Gains and Welfare Gains in Self-Selection Models [J]. *Reviews of Economics and Statistics*, 69: 42-49.
- BOUND J, DAVID J, REGINA B. 1995. Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak [J]. *Journal of the American Statistical Association*, 90: 443-450.
- CARD, DAVID, ALAN B K. 1992a. Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States [J]. *Journal of Political Economy*, 100: 1-40.
- CARD, DAVID, ALAN B K. 1992b. School Quality and Black - White Relative Earnings: A Direct Assessment [J]. *Quarterly Journal of Economics*, 107: 151-200.
- DANIEL O G, JOHN H. 2008. The Impact of Ethiopia's Productive Safety Net Programme and its Linkages [C]. IFPRI Discussion Paper 00839.
- DONOHUE, JOHN J, JUSTIN W. 2005. Uses and Abuses of Empirical Evidence in the Death Penalty Debate [J]. *Stanford Law Review*, 58: 791-845.
- EHRlich, ISAAC. 1975. The Deterrent Effect of Capital Punishment: A Question of Life and Death [J]. *American Economic Review*, 65: 397-417.
- EHRlich, ISAAC. 1977. Capital Punishment and Deterrence: Some Further Thoughts and Additional Evidence [J]. *Journal of Political Economy*, 85: 741-788.
- EHRlich, ISAAC. 1987. On the Issue of Causality in the Economic Model of Crime and Law Enforcement: Some Theoretical Considerations and Experimental Evidence [J]. *American Economic Review*, 77: 99-106.
- EHRlich, ISAAC. 1996. Crime, Punishment, and the Market for Offenses [J]. *Journal of Economic Perspectives*, 10: 43-67.
- HAHN, TODD, VANDER K. 2001. Identification and estimation of treatment effects with a regression discontinuity design [J]. *Econometrica*, 69: 201-209.
- HECKMAN J J. 1997. Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations [J]. *The Journal of Human Resources*, 32: 441-462.
- HECKMAN J J. 2008. Econometric causality [J]. *International Statistical Review*, 52: 1-27.
- HECKMAN J J, H I. 1997. Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme [J]. *The Review of Economic Studies*, 70: 605-654.
- HECKMAN J J, ROBB R. 1985. Alternative Methods for Evaluating the Impact of Interventions // JAMES J HECKMAN, Singer B. *Longitudinal Analysis of Labor Market Data*. New York: Cambridge University Press; 352.
- HECKMAN J J, ROBB R. 1986. Alternative Methods For Solving The Problem of Selection Bias in Evaluating The Impact of Treatments on Outcomes [M] // Wainer H. *Drawing Inferences from Self-Selected Samples*. Berlin: Springer-Verlag.
- KANNIKA D, CHENG HSIAO, XUEYAN ZHAO. 2010. Decriminalization and Marijuana Smoking Prevalence: Evidence From Australia [J]. *Journal of Business and Econometrics*, 38: 344-356.
- LEE D S. 2008. Randomized Experiments from Non-random Selection in US House elections [J]. *Journal of Econometrics*, 144: 675-697.
- PERKINS S M, TW. 2000. For objective causal inference: design trumps analysis [J]. *The Annual of Applied Statistics*, 48: 93-101.
- PUHANI P A. 2000. Poland on the Dole: The Effect of Reducing the Unemployment Benefit Entitlement Period [J]. *Journal of Population Economics*, 59: 35-44.
- ROSENBAUM P RD. 1983. The central role of the propensity score in observational studies for causal effects [J]. *Biometrika*, 84: 41-55.
- SANDRA CAVACO DF. 2009. Estimating the Effect of a Retraining Program on the Re-Employment Rate of Displaced Workers [R]. *The Institute for the Study of Labor Working Paper*.
- SHAOHUA CHEN RM. 2008. Are There Lasting Impacts of Aid to Poor Areas [R]. *World Bank Policy Research Working Paper 4084*.
- STEWART MB. 2004. The Employment Effects of the National Minimum Wage [J]. *The Economic Journal*, 63: 110-116.

(责任编校:夏冬)