

吉多·因本斯对因果关系分析的方法论贡献^{*}

——2021 年度诺贝尔经济学奖得主学术贡献评介

洪霓 于冷

编者按:瑞典皇家科学院于 2021 年 10 月 11 日宣布,将 2021 年度诺贝尔经济学奖授予三位美国经济学家戴维·卡德(David Card)、乔舒亚·安格里斯特(Joshua D. Angrist)和吉多·因本斯(Guido W. Imbens)**。其中,表彰卡德“对劳动经济学的实证研究贡献”;表彰安格里斯特和因本斯“对因果关系分析的方法论贡献”。本刊曾先后刊登《戴维·卡德与艾伦·克鲁格对经验微观经济学的贡献》(《经济学动态》2015 年第 4 期)、《约书亚·安格里斯特对经验微观经济学的贡献》(《经济学动态》2015 年第 3 期)两篇论文,分别对卡德和安格里斯特的学术贡献进行了评介。本刊在此继续刊登《吉多·因本斯对因果关系分析的方法论贡献》一文,对因本斯的学术贡献进行评介。

摘要:吉多·因本斯由于在因果关系分析的方法论贡献而分享 2021 年度诺贝尔经济学奖。本文主要从三个方面来介绍因本斯的学术成就:第一,因本斯针对自然实验中的异质性和非依从性问题,将工具变量与潜在结果框架相结合,建立基于研究设计的局部平均处理效应框架,并对得到的估计值进行了清晰的阐释。第二,因本斯将 LATE 框架进行拓展,使之应用范围更广,同时对 LATE 的外部有效性做出了解释。第三,因本斯在倾向得分和匹配估计、断点回归设计、双重差分法和机器学习等因果推断方法上做出了突出贡献。社会科学中许多重大问题都涉及因果关系,因本斯的研究提高了人们回答因果问题的能力,对学术界大有裨益。

关键词:吉多·因本斯 诺贝尔经济学奖 因果推断 局部平均处理效应 工具变量

2021 年度诺贝尔经济学奖被授予三位美国经济学家戴维·卡德(David Card)、乔舒亚·安格里斯特(Joshua D. Angrist)和吉多·因本斯(Guido W. Imbens),以表彰卡德“对劳动经济学的实证研究贡献”,表彰安格里斯特和因本斯“对因果关系分析的方法论贡献”。本文侧重于评介因本斯对经济学的学术贡献。

因本斯是荷兰裔美国经济学家。他于 1963 年出生,先后就读于荷兰伊拉斯姆斯大学(Erasmus University)、赫尔大学(The University of Hull),最终于 1991 年在美国布朗大学获得经济学博士学位。毕业后,曾在哈佛大学、加州大学洛杉矶分校、加州大学伯克利分校任教,2012 年至今执教于斯坦福大学商学院。因本斯主攻计量经济学,尤其是因果推断方法。本文重点讨论其有关局部平均处

^{*} 洪霓,上海交通大学安泰经济与管理学院,邮政编码:200030,电子邮箱:hn_neon@163.com;于冷,上海交通大学安泰经济与管理学院,邮政编码:200030,电子邮箱:yuleng@sjtu.edu.cn。感谢上海交通大学安泰经济与管理学院王昆仑博士在论文修改过程中的帮助,感谢匿名审稿人提出的建设性意见。当然,文责自负。

^{**} 2021 年度诺贝尔经济学奖得主 David Card、Joshua D. Angrist、Guido W. Imbens 又被分别译为“大卫·卡德”“约书亚·安格里斯特”“圭多·因本斯”,本刊也曾采用其中的“约书亚·安格里斯特”译名,现依据新华社报道中采用的译名,分别译为“戴维·卡德”“乔舒亚·安格里斯特”“吉多·因本斯”。

理效应(LATE)分析框架的建构及其拓展与应用外,也对其在倾向得分匹配、断点回归设计、双重差分法、机器学习等计量经济学领域的主要贡献进行评介。

一、因果关系分析框架的建立

因本斯和安格里斯特的获奖成就在于建立了改变实证研究范式的因果分析框架,使研究人员能通过合理的研究设计,利用自然实验或随机试验收集到的数据进行合理的因果推断。

(一)从随机控制实验到自然实验

一直以来,经济学家都热衷于对某一项目或政策方案的效果做出估计和解释,如何准确识别因果效应是经济学实证研究的热点及难点。因果推断的关键是比较同一个体在接受和不接受政策处理时的两种结果,但问题在于,人们在现实情况下只能选择接受或拒绝处理,研究者也只能观察到一种结果,Holland(1986)将此称为“因果推断的基本问题”。随机控制实验(randomized controlled trials, RCT)能够通过将实验对象随机划分为处理组和对照组来解决这一问题,比较处理组和控制组的结果差异并求取均值,就可以得到平均处理效应(average treatment effect, ATE)。2019年诺贝尔经济学奖也授予了阿比吉特·班纳吉、埃丝特·迪弗洛和迈克尔·克雷默,表彰他们将随机控制实验应用于贫困领域的工作。然而,由于受到成本、道德、伦理等多方制约,许多政策相关的RCT难以实施或大范围地开展。现实中能观测到的大多数经济数据都不是在严格的实验条件下得到的,经济学家们无法单纯通过控制变量的方法来推测某个处理对结果的影响。

从20世纪90年代开始,“自然实验”(natural experiments)逐渐兴起,为揭示变量间的因果关系提供了新思路。在某些情况下,人们受到自然、政策或制度改变的影响,就如同被随机分配一般,这使得研究人员可以把那些受处理影响的人和不受影响的人分开。到了20世纪90年代中期,因本斯和安格里斯特进一步发展了基于自然实验的因果推断方法,他们将经济学中的工具变量(instrumental variable)框架和统计学中因果推断的潜在结果框架进行整合,阐明了基于工具变量的因果推断中所需的基本识别假设。同时,他们也对估计中存在的异质性问题和非依从问题进行了分析,给出了局部平均处理效应(local average treatment effect, LATE)框架。

(二)早期自然实验的不足

早期将自然实验用于政策研究的论文主要集中在劳动经济学领域(Angrist, 1990; Card, 1990; Angrist & Krueger, 1991)。它们对最低工资、移民和教育等政策效果进行了探讨,使得学术界对劳动力市场的运作方式有了进一步了解。最初一批自然实验研究也发现了一些问题。(1)异质性问题。传统的工具变量法假设处理对所有个体的影响是相同的,但实际情况下处理对不同特征的个体影响往往是不同的。例如,接受同样的教育,高能力的孩子会比低能力的孩子学到更多知识,从而获得更多的教育回报。安格里斯特等(Angrist & Krueger, 1991)认为,如果人与人之间存在能力的偏差,传统的OLS估计法会低估教育回报。(2)依从性问题。研究者无法确保被分配到处理组的人一定会严格按照政策处理改变自己的行为。例如,政策计划将义务教育年限延长一年,但对于政策范围内的一部分人来说,他们在该政策出台前,接受教育的时间就远超过义务教育规定的年限,因此该政策不会影响他们的教育行为。这导致总体平均教育程度实际增加不足一年,教育回报的估计值会产生偏误。这种情况在医学研究上同样存在,一直以来,研究者们估计的是意向治疗(intention to treat, ITT)结果,即无论个体最终是否按分配规则接受了治疗,都当作被分配到治疗组来分析个体的治疗效果。虽然估计ITT的方法简单,不需要额外的假设,但它并未估计出真正接受治疗的个体的治疗效果。那么,当人群的异质性和非依从性存在时,研究者应该如何估计政策的处理效应呢?

早期对于这一问题的解决思路,大多是通过施加特殊的条件来估计参与处理个体的因果效应(Chamberlain 1986; Heckman 1990)。因本斯和安格里斯特(Imbens & Angrist, 1994)则另辟蹊径,他们提出的方法不需要对参与者的行为施加条件限制,只需要使用一组基本假设,就能够通过工具变量在自然实验中估计出依从者(complier)群体的平均因果效应,即局部平均处理效应(LATE)。

这一研究成果将工具变量回归结果纳入因果推断和政策效果评估的框架,同时赋予工具变量回归结果清晰的解释,成为该领域开创性的论文之一。

(三)LATE 框架的建立

因本斯和安格里斯特(1994)提出的方法论框架建立在潜在结果框架之上,该框架的核心是分配机制(assignment mechanism)。由于大部分的处理都不是随机的,因此讨论处理的分配机制十分重要。分配机制是指个体接受处理的条件概率,按分析的复杂程度可以分为三类。第一类是随机实验,其中个体分配到处理的概率是协变量的已知函数,且不随潜在结果的变化而变化。第二类分配机制则保留了处理概率不随潜在结果变化的条件,但不再是协变量的已知函数。第三类分配机制包括了与潜在结果有关的情况,对此最主要的分析方法就是工具变量、断点回归和双重差分(Imbens & Wooldridge, 2009)。因本斯和安格里斯特的工作就是将分配机制和工具联系起来。工具既可以是人为随机化的结果,比如随机对照实验,也可以来自自然实验。本文将通过如下例子,对他们提出的框架进行简述。

假设研究者想探究一个二值处理的效果,例如“高中学业完成与否对收入的影响”。将 Y_i 记为个体 i 的收入, $D_i=1$ 表示个体 i 完成了高中学业, $D_i=0$ 表示没有完成高中教育,分别对应两种潜在的收入结果: $Y_i(1)$ 和 $Y_i(0)$ 。在理想状况下,可以通过识别总样本的平均处理效应 $ATE = E[Y_i(1) - Y_i(0)]$ 来估计出完成学业对收入的影响。其中, $Y_i(1) - Y_i(0)$ 为两个潜在结果的差值,表示单一个体的因果效应。但是,如前所述,同一个体不可能出现既上完高中又没有上完高中的情况,因此个体的处理效应是无法被识别的,需要一个合理的“工具”——出生日期,来帮助我们估计完成高中学业对收入的影响。假设孩子仅在满 6 岁当年的 9 月 1 日获得一次入学的机会,在年满 18 岁后才可以选择辍学。那么,只有 7 月 1 日之前出生的孩子能够在高中毕业前年满 18 岁,有机会选择退学,因此出生在 7—12 月的孩子比出生在 1—6 月的孩子更有可能完成高中学业,出生日期的随机性导致了受教育程度的差异。研究者将出生日期定义为工具变量 Z ,对于出生在 7—12 月的个体,工具变量 Z 取值为 1,出生在 1—6 月的个体, Z 取值为 0。那些无论出生日期如何都会完成高中学业的人被称为“总是接受者”(always-taker);那些无论出生日期如何都不会完成高中学业的人被称为“从不接受者”(never-taker);出生在下半年并完成了高中学业的人和出生在上半年没有完成高中学业的人被称为“依从者”;处理值与工具变量值相反者被称为“违背者”(defier)。有了二元工具变量 Z_i ,就能得到两个潜在处理 $D_i(0)$ 和 $D_i(1)$ 以及四种潜在结果 $Y_i(z, d)$,其中 $z=0, 1, d=0, 1$ 。在因本斯和安格里斯特的分析框架下,出生日期作为有效工具变量需要满足以下最基本的假设。

(1)随机分配假设。该假设要求潜在结果与工具变量之间无关。在上述例子中表现为,出生日期是完全随机的,个体潜在收入和其出生日期之间是完全独立的。该假设成立时,研究者能够通过 Z 对 Y 的回归得到工具变量与潜在结果间的因果效应,也就是 RCT 中的 ITT 效应。

(2)排他性假设。该假设要求工具变量对于 Y 的影响只能通过 X 而不能有其他途径。在例子中表现为,出生日期只能通过影响个人是否完成学业而影响收入,而不会通过其他渠道影响收入。值得一提的是,假设(1)和假设(2)在一起才构成传统的外生性假设。因本斯和安格里斯特等(Angrist, Imbens & Rubin, 1996)也对这两个假设进行了严格的区分,违反这两个假设的原因和后果各不相同。

(3)相关性假设。该假设要求处理和工具变量之间的相关性不为 0,在例子中表现为出生日期会影响个体是否完成高中学业,下半年出生的人更有可能完成。

(4)单调性假设。即工具变量对潜在处理的影响是单调的, $D_i(1) \geq D_i(0)$ 或 $D_i(1) \leq D_i(0)$ 。在例子中表现为,如果个体出生在下半年,那么其完成高中学业的概率会比其出生在上半年时大。

在实际情况满足假设(1)~(4)的时候,最终得到的工具变量估计量是一个比值。其中,分子是所有被随机分配接受处理的群体的平均因果效应,也就是上文中提到的 ITT;分母是分配的人中实际依从了处理的群体的比例:

$$\frac{E[Y_i | Z_i = 1] - E[Y_i | Z_i = 0]}{E[D_i | Z_i = 1] - E[D_i | Z_i = 0]} = E[(Y_i(1) - Y_i(0)) | (D_i(1) - D_i(0)) = 1]$$

因本斯和安格里斯特将这种因果效应称为局部平均处理效应(LATE)。这些严格根据工具值的分配来改变自己处理状态的个人即为依从者,因此也将 LATE 称为依从者的平均因果效应(complier average causal effect,CACE)。

二、LATE 框架的拓展

(一)拓展型 LATE 框架

因本斯和安格里斯特(Imbens & Angrist,1994)除讨论了二值工具变量外,他们还证明,当工具变量取多个值或者是离散值时,工具变量的估计值是一个加权的局部平均处理效应(加权 LATE)。其中,两阶段最小二乘(two-stage least squares,2SLS)估计值作为最有效的工具变量估计值,是加权 LATE 的一致估计值,同时还具有渐进正态的良好性质,其方差可以用稳健估计值来估计。此前关于工具变量和 2SLS 的应用都是基于系数不变的回归模型,因本斯和安格里斯特(Angrist & Imbens, 1995)展示了如何利用 2SLS 来估计可变处理强度、多工具和协变量情况下的因果效应。他们引入了因果反应加权函数,该函数的长度可用于确定每单位处理效果的加权平均数。之后,他们以安格里斯特等人(Angrist & Krueger, 1991)的工作为基础,利用 2SLS 来估计受教育年限对收入的影响,其中因果反应加权函数的权重反映了个人在学校教育分布的不同点上受工具影响的程度。假设个体的收入为 Y_j , j 是其受教育的年限, $j=0,1,2,\dots,J$,虽然每个人都有完整的 Y_j 集,但我们只能观察到其中的一个变量。个人的 Y_j 集与其他人的处理状态和结果都无关。这种潜在结果或反事实结果在个体间的独立性被称为稳定单位处理值假设(stable unit treatment value assumption, SUTVA)。只要研究者接受反事实的结果概念,因果推断的目标是得到关于 $Y_j - Y_{j-1}$ 分布的信息,即为第 j 年学校教育的因果效应。相对于 Rubin(1974)的分析框架,该研究的创新之处有二:一是研究可针对多个处理值;二是加入了反事实的处理状态。

有时,政策制定者除了关心某项政策在群体中的平均因果效应外,更关注政策对于群体中分布在不同分位点上的人的影响。例如,技能培训政策对那些相对贫穷的人收入的提升是否更大。因本斯等(Imbens & Rubin, 1997)证明,在工具变量满足 SUTVA、排他性、严格单调性和随机性这四个假设时,研究者除了可以估计出依从者的平均处理效应外,还可以从数据中估计出依从者在不同处理下结果的边际分布。该结果使得政策制定者在评估和比较政策方案时,不仅可以考虑政策平均效应的差异,还可以考虑结果对人群的分散性差异。此外,标准的工具变量估计值在估计结果分布时是没有施加非负条件的(nonnegativity condition),在加入非负条件后,LATE 的估计效果会有很大改变。随后,因本斯和安格里斯特等(Abadie, Angrist & Imbens, 2002)再次讨论了如何利用工具变量识别分位数处理效果(quantile treatment effect, QTE)问题。当处理是外生确定的时候,QTE 的估计方法可以简化为分位数回归,运用凸线性规划问题来计算结果。该文在实证中估计了 JTPA (job training partnership act) 培训项目对不同收入分布人群的影响,结果显示,培训项目对低分位的女性收入提升效果较大,对高分位的男性收入提升较大。

(二)“弱工具”与“无效的工具”

安格里斯特等(Angrist & Krueger, 1991)在利用出生季度作为工具变量对美国教育回报率进行估计时得出结论:2SLS 和 OLS 估计的结果差异非常小,使用 OLS 估计教育回报率存在偏差的可能性并不大。有学者利用相同的数据针对这一结论提出了质疑——出生季度与受教育年限的区别之间似乎只有微弱的相关性,存在弱工具变量问题。该问题使得 2SLS 估计不仅难以矫正 OLS 估计的偏误,还会增大估计的标准差,降低估计效率。

因本斯等(Imbens & Chamberlain,2004)随后提出了一种新的估计方法。他们针对一个内生变量和多个工具变量的模型,在内生变量和工具变量间设置了一个方差为未知参数的随机系数结构,

得到的估计值被称为随机效应准极大似然估计值(random-effects quasi-maximum likelihood estimator, REQML)。这种方法也解释了有限信息极大似然估计(limited-information maximum likelihood, LIML)和 2SLS 在许多弱工具变量下的不佳表现。REQML 估计值的风险通常小于 LIML 估计值的风险,同时在限制条件下,REQML 估计值是渐进于 LIML 和 2SLS 的。

与讨论工具变量与处理的相关性强弱不同,因本斯等(Kolesár, Chetty, Friedman, Glaeser & Imbens, 2015)给出了另一种新方法——“无效的工具”。传统的工具变量要求其与内生变量相关且对因变量没有直接的影响,而无效的工具变量会对结果有直接影响。新方法的假设需要“无效工具对结果的直接影响”与“无效工具对内生变量的影响”之间是无关的。无效工具对结果的直接影响存在时,LIML 估计值不再具有一致性,但调整偏差校正的两阶段最小二乘(modified bias correct two-stage least square, MBTSLS)估计值是一致的。文章认为可以使用由 Cragg & Donald(1993)提出的统计量来检验无效工具的直接效应,并强烈建议研究者在实证研究中检验直接效应是否存在,同时对 LIML 估计值和 MBTSLS 估计值进行比较。

(三)LATE 的外部有效性

有学者质疑(Deaton, 2010; Heckman & Urzua, 2010),LATE 的内部有效性是以牺牲外部有效性为代价的。在四个基本假设下,LATE 可以准确识别依从者的因果效应;但是 LATE 估计值是建立在明确依从者的个人特征上的,当依从者群体的特征发生变化时,估计值是否还有效?在工具变量局部因果的分析框架下,能否估计整个群体的平均因果效应?这就需要验证 LATE 的外部有效效应。

因本斯(Imbens, 2010)给出了一个例子,加利福尼亚州的政策制定者打算将 4 年级班级的规模削减 10%,希望能通过提高分数之类的途径来改善学生的劳动力市场前景。因此,他们想对减少班级规模所需的成本和带来收益进行比较。有不少学者研究了班级规模对学生教育结果的影响(Krueger, 1999)。虽然这些文章选取的研究对象都是加州之外的学生,但因本斯认为它们的结果是可供政策制定者参考的,参考程度取决于研究的外部性和内部有效性。外部有效性依赖于很多因素,例如:(1)之前研究的时间相对于目标政策而言,越早的研究权重越低;(2)研究人群与加州目标人群之间的差异;(3)研究结果与加州项目之间的差异。而内部有效性主要取决于研究设计是否能得到可信的因果解释。因本斯认为,总体来说,虽然单独设计一项实验来研究加州的问题会得到更好的结果,但实验需要很长的时间,无法对当下的政策制定提供帮助。因此,单一研究虽然不能为政策提供一个明确和全面的预测值,但基于不同人群和不同环境的若干此类研究的组合可以为有效的政策措施的制定和实施提供指导。

学术界关于 LATE 的适用性还有如下讨论。很多时候依从者的身份是无法直接观察到的,但可以用观察到的协变量来描述依从者群体(Abadie, 2003; Angrist, 2004),这种描述足以阐明依从者的特殊性。例如,如果政策目标是低收入者,且其在依从者中占比足够高,那么 LATE 就会与政策目的高度相关。因本斯等(Imbens & Rubin, 1997)研究表明,可以从依从者的潜在结果分布中得出依从者间的效果异质性。当工具变量不止一个时,可以测试对工具异质性反应的程度。如果异质性不严重,估计值就可以推广到更广泛的群体中。赫克曼等(Heckman et al, 2001, 2003)和安格里斯特(Angrist, 2004)运用潜指数模型(latent-index models)来处理多值工具变量,识别并比较不同群体的因果效应,例如,比较总体、受处理群体和依从者的因果效应。同时,Vytlacil(2002)也肯定地认为,在不对潜指数模型施加参数化的函数形式或分布假设的情况下,LATE 假设并不弱于潜指数模型的假设。这表明,因本斯和安格里斯特的 LATE 框架可以被看作是潜指数模型的应用,他们的研究可以与关于选择模型的计量经济学文献直接联系起来。DiNardo & Lee(2011)证明可以将单调性假设放宽为概率单调性(probabilistic monotonicity),放宽后仍然可以得到明确的因果处理效应。概率单调性要求所有随机接受处理个体的处理概率都增加,此时的工具变量估计值可以解释为个体因果效应的加权平均值,其中权重反映了工具对每个(类型)个体的处理概率的影响。这一扩展意味着工具变量估计值可以不局限于依从者,其应用更有普遍性。

三、LATE 框架的应用

因本斯和安格里斯特将经济学中的工具变量分析框架和统计学中的潜在结果框架相结合,反过来产生了更通用的推断框架,使识别假设的性质更透明,研究者能够以此评估实证设计偏离假设时的统计敏感性。多重优点使 LATE 框架成为经济学中实验性工作的主导框架,它不仅能告诉研究者怎样的实验设计可以得到合理的工具变量估计值,还能帮助研究者检验在使用因果推断方法时,识别假设是否合理。本节将通过一些具体的例子来说明 LATE 框架的应用。

(一) 参军经历对收入的影响

在越南战争期间,大量美国男子受到征兵的影响。在 1950—1952 年间出生的男性中,大约有 38% 的人会被抽到有参军的资格。在此背景下,参军经历是否会对劳动力市场存在一定作用? 如果参军经历对收入的作用有限,政府应该向退伍军人提供多少补偿? 在研究中,单纯将劳动力市场的结果与实际的退伍军人身份进行回归,得到的“参军对收入的影响”的因果效应很明显是具有误导性的,因为服兵役的人可能是根据观察到的和未观察到的特征而选择的。安格里斯特(Angrist, 1990)用随机分配的征兵资格作为在越南战争中服役的工具变量来估计退伍军人身份对收入的影响。研究规定参军的概率随着抽到的号码的增大而变小,单调性假设要求,在抽签号码为 K 的情况下服兵役的人,也会在抽签号为 $L(L < K)$ 的情况下参军,这确保了几乎不存在违背者。排他性假设要求,不管是否参军,潜在收入都与抽签号码无关。如果抽签号码通过其他变量对收入产生了影响,则会违反排他性假设。对总是接受者来说,无论抽签号码如何,他们一定会参军,所以抽签号码对收入没有直接影响是合理的。但是对于从不接受者来说,排他性假设不一定刚好成立。如果被征召者采用留在学校或移居国外等方法来免除兵役,这些行为可能会对收入产生直接影响,从而违反排他性假设。因本斯和安格里斯特等(Angrist, Imbens & Rubin, 1996)也用了类似的方法,以“越南战争时期参军对死亡率的影响”为例,再次向我们展示了一个合理的工具变量所需要满足的假设,并详细分析了工具变量估计在排他性假设和单调性假设违反时的敏感性。

(二) 法官判决对劳动成果的影响

将法官和判案者的判决用作工具变量,称之为“法官的宽恕”设计。鉴于需要决策者判决的案件是随机分配的,且他们各自的宽容程度不同,宽容程度可以用于工具变量设计中来研究“判决”对结果的影响。其中,有关监禁时间对其后劳动成果影响的研究难点在于,长期监禁的人和短期监禁的人的个体特征是不同的,这些不同可能会对其后的劳动收入产生影响。据此,Kling(2006)利用案件是根据提交的时间和地点随机分配给法官,且法官在判案时的宽严程度不同这一特点,将案件的随机分配作为工具变量。Dobbie et al(2018)的研究使用随机分配的法官的拘留决策来估计审讯前拘留对后续结果的因果效应。当决策者和被判决案件之间没有交互时,排他性假设很容易被满足。但是当决策者和个人之间存在互动时,就需要注意排除性假设是否成立。此外,单调性假设也是问题之一,Dobbie et al(2018)等认为,单调性要求“如果犯人在面对严厉的法官审判时,没有被拘禁,那面对宽容的法官也不会被拘禁”。但是,法官不仅在判案的宽严程度上有差异,在偏好上也会有不同,一些法官可能会对某些类型的被告做出更严厉的判罚,此时就要慎重考虑单调性假设是否成立。

(三) 在医学领域的应用

因本斯和安格里斯特开发的 LATE 框架不仅广泛应用于经济学和其他社会科学研究,而且在流行病学和医学等学科中的应用也越来越多。本小节讨论 LATE 估计作为 ITT 分析的补充在医学研究领域的运用。

医学界常通过解释性实验和实用性实验来研究某种药物或治疗方法的因果效应。解释性实验在理想条件或控制条件下进行,用于估计治疗措施的疗效(efficacy,一般是药物或治疗方法的疗效)。实用性实验则在现实条件或临床条件下进行,用于检验治疗措施的有效性(effectiveness,一般是根据实验数据,由主管机构做出判断)。在实用性实验中,对既定的治疗不完全依从的现象是很常见

的,传统的方法是主要使用 ITT 分析,此时 ITT 估计是用于测量分配的因果效应而不是治疗的因果效应。医学上使用 ITT 分析的理由有二:其一,ITT 效应可能是与实际政策效应最相关的,因为我们不能强迫人们接受一种治疗;其二,在排他性限制下,ITT 估计值是被低估的,因此它可以作为一种评估新治疗方法的保守策略。

然而,Hernán & Hernandez-Diaz(2012)认为这些使用 ITT 的理由并非完全合理。首先,当研究者需要对两种依从率不同的治疗方法进行比较时,ITT 分析就不一定适用。一种治疗方法在估计结果上表现得更有效很可能只是因为被试更坚持使用这种治疗。其次,在估计治疗的负面效应时,被低估的负面效果可能会使不安全的治疗在结果上显得安全。因此,国际临床试验管理规范(ICH-GCP, 1999)建议,对负面效果的分析应该根据被试实际接受治疗与否来进行。Dodd et al(2012)回顾 2008 年发表在医学顶刊上关于 RCT 中的不遵从治疗方案的研究发现,除 ITT 外,一半的研究还包括“接受治疗分析”(as-treated)或“完成治疗分析”(per-protocol),试图解决不遵从治疗造成的问题。其中,接受治疗分析是将接受治疗的人与不接受治疗的人进行比较;完成治疗分析则排除了没有遵守治疗方案的人。同样,但这两种方法对治疗效果的估计仍然存在偏差。然而,当研究者可以得到参与治疗者的信息时,就能使用 LATE 框架来估计实验中遵从治疗方案病人的治疗效果。因此,许多学者认为在医学领域应广泛运用 LATE 来估计治疗的平均效果。

四、因本斯的其他贡献

因本斯的贡献一方面在于提出了新的因果推断方法,将传统工具变量结果纳入因果推断分析框架重新解读,对既有方法加以改进和完善。另一方面,他在倾向得分和匹配估计法、断点回归设计等方面也颇有建树。近年来,他还与其夫人苏珊·阿西(Susan Athey)及同事一起将机器学习等方法引入到因果推断分析框架。

(一)倾向得分和匹配估计法

倾向得分法最初是由 Rosenbaum & Rubin(1983a, 1983b, 1985)提出,是使用非实验数据或观测数据进行处理效应分析的一种方法,其目的是消除数据偏误和混杂变量的影响,使处理组 and 对照组之间更加可比。因本斯在该领域的研究绝大多数是和麻省理工学院的计量经济学家阿巴迪(A. Abadie)共同完成的。

Rosenbaum & Rubin(1983a)认为只需要对倾向得分进行调整,就可以消除处理组和控制组之间的处理前变量差异带来的所有偏差,但这种方法只能针对处理仅取两个值的情况。因本斯将其拓展到估计多值处理的平均处理效应(Imbens, 2000)。之后又证明,通过对倾向得分非参数估计的倒数进行加权,可以得到平均处理效应的有效估计值(Hirano, Imbens & Ridder, 2003)。在研究非劳动所得的收入对劳动供给、工资、储蓄和消费的效应时,因本斯等(Imbens, Rubin & Sacerdote, 2001)就采用了改进的倾向得分方法,对被研究者中奖之前的差异进行调整。随后,在多值处理的基础上,因本斯等(Imbens & Hirano, 2004)又讨论了连续政策干预的情况,提出了广义倾向得分(generalized propensity score, GPS)的概念。与二值的情况一样,GPS 具有二值处理倾向得分的许多良好性质,通过对协变量的标量函数进行调整,可以消除与协变量差异相关的所有偏差。此外,GPS 还具有平衡特性,可用于评估该分数在特定情况下是否充分。

自 2006 年开始,因本斯和阿巴迪合作发表了一系列文章,进一步讨论了采用匹配估计方法进行因果推断时,估计量的统计性质、其方差的正确计算方式等问题,为该方法在实证研究中的使用提供了实践指引。

因本斯和阿巴迪(Imbens & Abadie, 2006)推导了匹配估计值的大样本特性。标准的渐进扩展不适用于固定数量匹配的匹配估计值,虽然匹配估计方法在因果推断中已经广为使用,但它的大样本特性还没有确立。他们首先证明,简单的近邻匹配估计值包括一个条件偏差项,其收敛到零的速度可能比 \sqrt{n} 还要慢,匹配估计值在一般情况下不是 \sqrt{n} 一致的;其次,即使在匹配估计值是 \sqrt{n} 一致的情

况下,固定数量匹配的匹配估计值也不能达到半参数效率约束。针对此性质,之后因本斯和阿巴迪(Imbens & Abadie, 2011)提出了一种偏差校正的方法,渐进地消除了条件性偏误,使匹配估计值是 \sqrt{n} 一致的。同时,在不需要对未知函数进行一致的非参估计的情况下,得到新的条件方差估计值。他们通过模拟研究说明,使用了偏差校正的匹配估计值与那些简单的偏误估计值和回归估计值相比,在偏误和均方根方面都表现得要好。

如前文所述,在大样本下,匹配估计值只能在特定的情况下被推导出来(Abadie & Imbens, 2006)或进行偏差校正(Imbens & Abadie, 2011)。为此,因本斯和阿巴迪(Abadie & Imbens, 2012)建立了一个匹配估计值的鞅表示。该表示方法使用鞅极限定理来推导匹配估计值的大样本分布。为了说明该方法的实用性,他们推导出匹配估计值在不进行替换的情况下的渐进分布。同时还运用这一新方法推导出,在使用“hot deck”(一种常用于当前人口调查(CPS)和其他社会科学大型调查中的匹配推算法)进行缺失数据推算时,对样本均值的标准误差的修正的方法。使用“hot deck”标准误差修正的效果很好,而没有使用“hot deck”修正的标准误会出现严重的向下偏移。

在实证研究中,如果处理组和控制组之间存在系统性差异时,平均处理效应的估计结果就会不精确。因本斯等(Imbens et al, 2009)给出了可以最精确地估计平均处理效果的最佳子样本的特征。在某些条件下,最佳选择规则完全取决于倾向性得分。对于范围较广的分布,获得良好近似值的方法就是放弃所有估计倾向分数在 $[0.1, 0.9]$ 范围之外的单位。此外,针对传统的 bootstrap 方法,因本斯和阿巴迪(Imbens & Abadie, 2008)还指出,在匹配估计中运用 bootstrap 是无效的。由于最近邻匹配的极端非平滑性,导致 bootstrap 的标准条件不能满足,因此 bootstrap 方差与实际方差差异很大。这一模拟结果也证实了理论计算所预测的引导法置信区间的实际覆盖率和名义覆盖率之间的确存在差异。

因本斯等(Graham, Imbens & Ridder, 2014)还介绍了一种衡量重新分配政策对平均产出的影响的方法,用非参数方法分析分类匹配效应的互补和总体影响,并推导出该估计值的大样本特性和小样本特性。重新分配政策的特征是不涉及资源的增加。虽然无法在投入资源不可分割且总量固定时同时提高所有生产单位或企业的投入水平,但是对不可分割的投入进行重新分配可能会改变平均产出。例如,教师的能力就是不可分割的一种投入(一个老师只能教一个班的学生),如果教师总数固定不变,研究者就不可能将高能力的教师安排到所有班级。在这种情况下,将教师重新分配到各个班级中的效果就是我们所关注的。这种重新分配政策的平均因果效应被称为总体再分配效应(aggregate redistributive effects, AREs),该估计方法是完全非参数化的,且需要生产函数具有非加性和非线性。具体采用两步估计法:第一,非参数化的估计生产函数;第二,在由新分配规则确定的投入分布下,估计平均生产函数。

传统的倾向得分分类和匹配法并不能够拓展到有多层级处理的情况下,在最近的研究中,因本斯等(Yang, Imbens, et al, 2016)还将倾向得分和匹配法推广到有多层级处理时的平均处理效应估计。在新方法中,不需要通过调整倾向得分向量的方法来构建使处理和协变量相互独立的平衡单元集,仅需要调整个体接受某特定处理的概率就可以估计每个处理的潜在结果均值。他们还运用弱无干扰性和广义倾向评分的概念证明,对处理前变量的标量函数进行调整可以消除与观察到的预处理变量相关的所有偏差。

(二)断点回归设计

断点回归设计(regression discontinuity designs, RDD)是由 Thistlewaite & Campbell(1960)首次提出的,利用接受离散处理的激励或能力的断点来进行因果推断。这一设计的关键特征是定义了“参考变量”(forcing variable)——在其临界值处,接受处理的概率会发生不连续的变化。此时,我们假设接近临界值两端的个体是相似的,他们之间平均结果的差异可以归因于是否接受了处理。根据处理的分配规则,RDD 可以分为精确断点回归和模糊断点回归。前者是指处理的分配完全取决于参考变量是否超过临界值,超过临界值则表明个体一定会接受处理;反之,后者是指处理的分配不完

全取决于参考变量,还受到其他不能被观测到的因素影响,当参考变量超过临界值时,个体接受处理的概率会增大,但概率不等于1。

从20世纪90年代末开始,经济学中应用和扩展RDD的研究越来越多。Hahn(2001)进一步提出了对模糊断点回归设计估计的解释、处理效果的普遍异质性等关键的理论和概念。因本斯等(Imbens & Lemieux, 2008)所撰写的《断点回归实践指引》一文主要讨论了RDD中的几个实际问题:一是强调图形分析是阐明设计的有力方法;二是建议只使用靠近不连续点的观测值来进行局部线性回归;三是提供了渐进方差的两个估计值;四是讨论了三种情况下的设定检验和敏感性分析:协变量均值的不连续性、参考变量条件密度的不连续性和参考变量其他值平均结果的不连续性。该文成为之后应用该方法的实践宝典。RDD的一个重要环节在于断点邻域大小的选择,也就是带宽选择(bandwidth selection)问题。带宽选择过大,意味着有更多样本被纳入分析中,参数估计更准确,但同样也意味着样本随机性要求难以满足,内生性问题更严重。此前,很少有论文专门讨论RDD中的带宽选择问题,也没有文章给出带宽的最优性质。因本斯等(Imbens & Kalyanaraman, 2012)研究了断点回归估计值的带宽选择问题,他们考虑了RDD背景的特殊性,推导出平方误差损失下的渐进最优带宽,根据Li(1987)的准则,可以证明由此推出的带宽一定是最优的。同时,他们提供了一种完全依赖于研究者手中数据的最优渐进带宽选择方法,最优带宽取决于数据分布的未知函数的一致估计值。他们最后使用了Lee(2008)的数据及该数据集的模拟研究来验证带宽选择效果,模拟结果表明新规则表现良好。

2019年后,因本斯再次撰写数篇论文,讨论了断点回归使用中存在的一些问题。

(1)参考变量的高次多项式。在之前的断点回归分析中,控制参考变量的三次、四次或更高次多项式是很常见的。但因本斯等(Gelman & Imbens, 2019)认为,在断点回归中控制全局高阶多项式(global high-order polynomials)是一种有缺陷的方法,主要有三个问题:一是估计有噪声;二是得到的因果效应结果对多项式的阶数敏感;三是置信区间的覆盖率不好。因此,他们建议研究人员改为使用基于局部线性、二次多项式或其他平滑函数的估计值。

(2)数值凸优化法。现有大多数研究都首先在处理分配的边界两侧拟合非参数回归模型,再报告处理因果效应的估计值。但在实际中,往往很难针对具体的目标来调整非参数回归。因本斯等(Imbens & Wager, 2019)给出了一种在断点回归设计中进行估计和统计推断的新方法。该方法使用数值凸优化直接得到断点回归中有限样本极大极小(finite-sample-minimax)的线性估计量,该估计量受限于条件响应函数的二阶导数界限。当二阶导数的界限确定时,这一方法完全受数据驱动,可以为离散或连续变量的断点回归参数提供一致的置信区间。该方法还可以应用于多个参考变量的情况。

(3)模糊断点回归设计中的外部有效性。现有的效应识别结果仅在参考变量的临界值附近可信,但是在很多情况下,研究人员还想将研究结果推广到不在临界值附近的群体,或是依从者之外的群体。因本斯等(Bertanha & Imbens, 2020)说明了何时LATE估计值能够代表总样本的因果效应,这需要进行两项检验:一是检验未接受处理的依从者和从不接受者的平均效应是否相等;二是检验接受处理的依从者和总是接受者的平均效应是否相等。如果这两个条件成立,则意味着由参考变量和处理状态决定的预期结果是连续的。他们还讨论了将参考变量临界值附近个体的识别效应推广到总样本的其他部分的方法。Stata和Matlab中已经有了新方法对应的命令。

(三)双重差分法

当条件独立性假设不成立,又难以找到合适的工具变量时,研究者亦可采用双重差分法(difference-in-differences, DID)来开展政策评估。经典的DID方法假设处理组和控制组拥有相同的线性趋势,经过两次差分后就可以得到所关注的平均处理效应。DID方法有两个新的进展:合成控制法(synthetic control method)以及非线性双重变换模型(nonlinear changes-in-changes method)。合成控制法的关键在于确定权重,一般是使用最小距离法(Abadie et al, 2010)。因本斯等(Doudchenko

& Imbens, 2016)提出一种更普遍的合成控制估计法,可以放松 ADH 方法中的一些限制。当备选控制单元数目很大时,可以采用最优子集回归、LASSO 和弹性网方法来确定更为有效的权重。双重变换模型也被称为非线性双重差分法(Athey & Imbens, 2006),与经典的 DID 不同,这一方法引入了一个非线性模型,允许不可观察因素的影响随时间变化,例如,工资水平以及劳动力市场的技能回报可能存在时间趋势。该方法提供了对整个反事实结果分布的估计,其得出的估计值具有渐进正态的良好性质。

(四)机器学习与因果推断

近年来,因本斯和许多学者一起,将机器学习方法引入到因果推断分析框架中。之前学界采用的因果推断方法,大都是通过“构造”与观测样本特征相同的反事实样本,比较二者实际观测值的差异,从而估计出因果效应。但是,过多的假设一定程度上限制了估计结果的准确性。而近年来兴起的机器学习方法可以减少传统因果推断分析框架下假设的束缚,对不可观测的研究对象做出精确的预测,使因果效应的估计更为可靠。

1. 因果树和因果森林。在很多情况下,一项政策干预应用在不同背景下可能产生不同的成本和收益。研究异质性的方法之一就是通过学习来识别不同处理效应的子群(subgroups),从而发现异质性的具体形式,而回归树模型是机器学习中最常见的应用之一。其基本思路是,根据样本的协变量特征与阈值进行比较,不断划分形成子群,最后比较处于同一个分组内控制组样本和处理组样本,得到研究所关注的因果效应。因本斯夫妇(Athey & Imbens, 2016)使用诚实分支(honest splitting)取代适应性分支(adaptive splitting),将回归树算法改进为“因果树”。因果树是基于回归树的机器学习方法,但在构建树时,以降低处理效应的均方差为准则。这一方法需要划分样本:一部分用来决定对协变量空间的最优划分,另一部分则用来估计每个子样本的处理效应,最终能得到子样本的无偏异质性因果效应。因果树的优点在于,研究人员可以使用更复杂的结构来估计因果效应,且无论估计中用到多少个协变量,所得到的置信区间都是有效的。该方法随后被 Wager & Athey(2018)拓展,他们将随机森林和因果推断中的潜在结果模型结合起来,创造了“因果森林”(causal forest)。它使研究者能够从现有数据中得到更多关于异质性的信息,极大促进了经济学家在因果效果异质性上的探索。

2. 面板数据因果效应的推断。因本斯夫妇等(Athey, Bayati, Imbens & Gu, 2019)将机器学习中的集成方法(ensemble methods)应用于面板数据,发现能够得到比传统方法更精确的结果。他们选取了三种现有的预测方法:一是假设单位之间的相关性随时间稳定;二是假设所有单位共有稳定的时间序列模式;三是使用因子模型。他们用这三种方法分别预测了美国 270 个季度的州级 GDP 数据的缺失值。随后探究了基于这三种不同方法的组合预测结果。结果显示,集成方法的预测值明显更实用有效。他们也呼吁研究者在之后的经济学实证工作中多加关注集成方法的应用。

3. 协变量的处理。之前很多研究对平均处理效应的估计都是在协变量数量一定时进行的。因本斯夫妇等(Athey, Imbens, et al, 2017)将之前文献中的估计方法扩展到协变量的数目很大的情形下,给出了四种固定协方差数量下的估计,并讨论了当存在多个预处理变量时几种较好的估计量,分别是双重选择估计量(double selection estimator)、均衡近似残差估计量(approximate residual balancing estimator)以及双重稳健估计量和双重机器学习估计量(doubly robust estimator & the double machine learning estimator)。它们都可以处理两类关联偏差:一是预处理变量和潜在结果之间的关联偏差;二是预处理变量和处理分配之间的偏差。他们在文章最后建议研究者除了报告点估计和标准误差外,还应该报告一些补充分析的结果,以便更好地帮助评价估计的可信度,此处的可信度并非指无干扰假设是否成立,而是指估计值是否针对协变量的差异进行了有效调整。

诚然,分析中包含的预处理变量越多,无干扰假设就会越合理,但过多的预处理变量可能会降低对处理效应估计的表现。因此,因本斯夫妇等(Athey, Imbens & Wager, 2018)提出了一种消除惩罚回归调整偏差的方法。在给定线性假设时,研究者不需要假设存在一致的倾向得分估计量或假设

模型的稀疏性,就可以使用 LASSO 等回归方法对高维线性模型中的平均处理效应进行 \sqrt{n} 的一致推断。

4. 生成对抗网络(generative adversarial networks, GANs)。研究人员在开发新的计量经济学方法时,通常将新方法的性能与蒙特卡洛研究中现有方法的性能进行比较。但由于研究人员往往是酌情选取蒙特卡洛设计的结果,因此这种方法并不可信。基于此,因本斯夫妇等(Athey, Imbens, et al, 2021)建议使用生成对抗网络来提高可信度。该方法起源于机器学习文献,可以系统地生成与现有数据集非常相似的人工数据。因此结合现有的真实数据集,GANs 可以用于限制蒙特卡洛研究设计中的自由度,使比较的结果更为可信。如果研究人员比较关注特定统计方法作用在特定数据集上的效果,还可以使用这种 GANs 生成与特定数据集非常相似的模拟数据来评估所感兴趣的方法的效果。他们还给出了将 Wasserstein GANs 用于估计平均处理效应的例子,当没有一个估计量完全优于其他估计量时,研究人员应该根据具体情况调整分析方法。此时,系统的模拟研究和生成的模拟数据可以帮助研究人员选择更好的方法。

五、总结

2021 年度诺贝尔经济学奖获奖者的工作共同为基于研究设计的因果关系估计方法奠定了基础,在过去 30 年里,这种方法极大地改变了实证研究的范式。因本斯和安格里斯特的研究将工具变量与统计学中常用的因果推断的潜在结果框架联系起来,证明可以通过工具变量来估计因果效应。即使存在异质性和不完全依从现象,只要满足一系列最宽松的假设,研究者都可以估计出一个明确的因果处理效果,并能够对估计的结果进行清晰的阐释。因本斯和安格里斯特的分析框架使识别假设的本质更加清晰,提高了研究人员建立因果效应、评估其实证设计的敏感性和解释实证结果的能力。

本文详细评介了因本斯最主要的贡献之一,即因果推断和政策效果评估的分析框架,给出了估计 LATE 最宽松的假设条件,及识别和解释 LATE 的方法。同时,本文总结了因本斯及其他研究者对 LATE 分析框架的后续拓展工作,以及 LATE 在经济学及其他领域的应用。最后,本文梳理了因本斯在计量经济学倾向得分和匹配估计、断点回归设计、双重差分法和机器学习等方面的贡献。这些方法大多是为了更严谨、更精确地分析经济运行中的具体因果关系,为解决社会广泛的复杂问题提供更加可靠、可信的因果实证分析结果与经验证据。虽然计量经济学方法的不断改进和创新能够为研究者提供更强大的“工具”,但实证研究中容易存在投机取巧的现象,实证结果具有统计学意义的文章发更容易发表。经济学研究的可复制性和稳健性检验也应受到重视。中国经济学研究很容易受到传统经济学分析框架的影响,中国学者如何摆脱固有思维的束缚,结合中国实际情况开展适合中国国情的经济学研究,为当前存在的社会问题提供实证帮助,将是未来学者的努力方向。

参考文献:

- Abadie, A. (2003), “Semiparametric instrumental variable estimation of treatment response models”, *Journal of Econometrics* 113(2):231–263.
- Abadie, A. & G. W. Imbens(2006), “Large sample properties of matching estimators for average treatment effects”, *Econometrica* 74(1):235–267.
- Abadie, A. & G. W. Imbens(2012), “A martingale representation for matching estimators”, *Journal of the American Statistical Association* 107(498):833–843.
- Abadie, A., J. D. Angrist & G. W. Imbens (2002), “Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings”, *Econometrica* 70(1):91–117.
- Abadie, A. et al(2010), “Synthetic control methods for comparative case studies: Estimating the effect of California’s tobacco control program”, *Journal of the American Statistical Association* 105(490):493–505.
- Angrist, J. D. (1990), “Lifetime earnings and the Vietnam era draft lottery: Evidence from social security administrative records”, *American Economic Review* 80(3):313–336.

- Angrist, J. D. (2004), "Treatment effect heterogeneity in theory and practice", *Economic Journal* 114(494):C52—C83.
- Angrist, J. D. & G. W. Imbens(1995), "Two-stage least squares estimation of average causal effect in models with variable treatment intensity", *Journal of the American Statistical Association* 90(430):431—442.
- Angrist, J. D. , G. W. Imbens & D. Rubin(1996), "Identification of causal effects using instrumental variables", *Journal of the American Statistical Association* 91(434):444—455.
- Angrist, J. D. & A. Krueger(1991), "Does compulsory schooling attendance affect schooling and earnings?", *Quarterly Journal of Economics* 106(4):979—1014.
- Athey, S. & G. W. Imbens(2006), "Identification and inference in nonlinear difference-in-differences models", *Econometrica* 74(2):431—497.
- Athey, S. & G. W. Imbens(2016), "Recursive partitioning for heterogeneous causal effects", *Proceedings of the National Academy of Science* 113(27):7353—7360.
- Athey, S. , G. W. Imbens & S. Wager(2018), "Approximate residual balancing: Debiased inference of average treatment effects in high dimensions", *Journal of Royal Statistical Society-Series B* 80(4):597—623.
- Athey S. , M. Bayati, G. Imbens & Z. Qu(2019), "Ensemble methods for causal effects in panel data settings", *AEA Papers and Proceedings* 109:65—70.
- Athey, S. , G. Imbens, J. Metzger & E. Munro(2021), "Using Wasserstein generative adversarial networks for the design of Monte Carlo simulations", *Journal of Econometrics*, forthcoming.
- Athey, S. , G. Imbens, T. Pham & S. Wager(2017), "Estimating average treatment effects: Supplementary analyses and remaining challenges", *American Economic Review* 107(5):278—281.
- Bertanha, M. & G. W. Imbens(2020), "External validity in fuzzy regression discontinuity designs", *Journal of Business & Economic Statistics* 38(3):593—612.
- Card, D. (1990), "The impact of the Mariel boatlift on the Miami labor market", *Industrial and Labor Relations Review* 43(2):245—257.
- Chamberlain, G. (1986), "Asymptotic efficiency in semi-parametric models with censoring", *Journal of Econometrics* 32(2):189—218.
- Cragg, J. & S. Donald(1993), "Testing identifiability and specification in instrumental variable models", *Econometric Theory* 9(2):222—240.
- Deaton, A. (2010), "Instruments, randomization, and learning about development", *Journal of Economic Literature* 48(2):424—455.
- DiNardo, J. & D. Lee(2011), "Program evaluation and research designs", in: O. Ashenfelter & D. Card (eds), *Handbook of Labor Economics*, Vol. 4A, Amsterdam: Elsevier.
- Dobbie, W. et al(2018), "The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges", *American Economic Review* 108(2):201—240.
- Dodd, S. et al(2012), "Nonadherence to treatment protocol in published randomised controlled trials: A review", *Trials*, <https://trialsjournal.biomedcentral.com/articles/10.1186/1745-6215-13-84>.
- Doudchenko, N. & G. W. Imbens(2016), "Balancing, regression, difference-in-differences and synthetic control methods: A synthesis", NBER Working Papers, No. 22791.
- Gelman, A. & G. W. Imbens(2019), "Why high-order polynomials should not be used in regression discontinuity designs", *Journal of Business & Economic Statistics* 37(3):447—456.
- Graham, B. , G. W. Imbens & G. Ridder(2014), "Complementarity and aggregate implications of assortative matching: A nonparametric analysis", *Quantitative Economics* 5(1):29—66.
- Hahn, J. et al(2001), "Identification and estimation of treatment effects with a regression discontinuity design", *Econometrica* 69(1):201—209.
- Heckman, J. (1990), "Varieties of selection bias", *American Economic Review* 80(2):313—318.
- Heckman, J. et al(2001), "Four parameters of interest in the evaluation of social programs", *Southern Economic Journal* 68(2):210—223.
- Heckman, J. et al(2003), "Simple estimators for treatment parameters in a latent-variable framework", *Review of Economics and Statistics* 85(3):748—755.

- Heckman, J. & S. Urzua(2010), “Comparing IV with structural models: What simple IV can and cannot identify”, *Journal of Econometrics* 156(1):27–37.
- Hernán, M. & S. Hernandez-Diaz(2012), “Beyond the intention-to-treat in comparative effectiveness research”, *Clinical Trials* 9(1):48–55.
- Hirano, K., G. W. Imbens & G. Ridder(2003), “Efficient estimation of average treatment effects using the estimated propensity score”, *Econometrica* 71(4):1161–1189.
- Holland, P.(1986), “Statistics and causal inference”, *Journal of the American Statistical Association* 81(396):945–970.
- ICHGCP(1999), “ICH harmonised tripartite guideline: Statistical principles for clinical trials”, *Statistics in Medicine* 18(15):1905–1942.
- Imbens, G. W. (2000), “The role of the propensity score in estimating dose-response functions”, *Biometrika* 87(3):706–710.
- Imbens, G. W. (2010), “Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009)”, *Journal of Economic Literature* 48(2):399–423.
- Imbens, G. W. & A. Abadie(2006), “Large sample properties of matching estimators for average treatment effects”, *Econometrica* 74(1):235–267.
- Imbens, G. W. & A. Abadie(2008), “On the failure of the bootstrap for matching estimators”, *Econometrica* 76(6):1537–1557.
- Imbens, G. W. & A. Abadie(2011), “Bias-corrected matching estimators for average treatment effects”, *Journal of Business and Economic Statistics* 29(1):1–11.
- Imbens, G. W. & J. D. Angrist(1994), “Identification and estimation of local average treatment effects”, *Econometrica* 62(2):467–475.
- Imbens, G. W. & S. Athey(2006), “Identification and inference in nonlinear difference-in-differences models”, *Econometrica* 74(2):431–497.
- Imbens, G. W. & G. Chamberlain(2004), “Random effects estimators with many instrumental variables”, *Econometrica* 72(1):295–306.
- Imbens, G. W. & K. Hirano(2004), “The propensity score with continuous treatments”, in: A. Gelman & X.-L. Meng(eds), *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, John Wiley & Sons, Ltd.
- Imbens, G. W. & K. Kalyanaraman(2012), “Optimal bandwidth choice for the regression discontinuity estimator”, *Review of Economic Studies* 79(3):933–959.
- Imbens, G. W. & T. Lemieux(2008), “Regression discontinuity designs: A guide to practice”, *Journal of Econometrics* 142(2):615–635.
- Imbens, G. W. & D. Rubin(1997), “Bayesian inference for causal effects in randomized experiments with noncompliance”, *Annals of Statistics* 25(1):305–327.
- Imbens, G. W. & D. Rubin(1997), “Estimating outcome distributions for compliers in instrumental variables models”, *Review of Economic Studies* 64(4):555–574.
- Imbens, G. W. & S. Wager(2019), “Optimized regression discontinuity designs”, *Review of Economics and Statistics* 101(2):264–278.
- Imbens, G. W. & J. Wooldridge(2009), “Recent developments in the economics of program evaluation”, *Journal of Economic Literature* 47(1):5–86.
- Imbens, G. W., K. Hirano & G. Ridder(2003), “Efficient estimation of average treatment effects using the estimated propensity score”, *Econometrica* 71(4):1161–1189.
- Imbens, G. W., D. Rubin & B. Sacerdote(2001), “Estimating the effect of unearned income on labor supply, earnings, savings and consumption: Evidence from a survey of lottery players”, *American Economic Review* 91(4):778–794.
- Imbens, G. W. et al(2009), “Dealing with limited overlap in estimation of average treatment effects”, *Biometrika* 96(1):187–199.
- Kling, J. (2006), “Incarceration length, employment, and earnings”, *American Economic Review* 96(3):863–876.
- Kolesár, M., R. Chetty, J. Friedman, E. Glaeser & G. W. Imbens(2015), “Identification and inference with many in-

- valid instruments”, *Journal of Business & Economic Statistics* 50(2):373–419.
- Krueger, A. (1999), “Experimental estimates of education production functions”, *Quarterly Journal of Economics* 114(2):497–532.
- Lee, D. (2008), “Randomized experiments from non-random selection in U. S. house elections”, *Journal of Econometrics* 142(2):675–697.
- Li, K. C. (1987), “Asymptotic optimality for C_p , C_L , cross-validation and generalized cross-validation: Discrete index set”, *Annals of Statistics* 15(3):958–975.
- Rosenbaum, P. & D. Rubin(1983a), “The central role of the propensity score in observational studies for causal effects”, *Biometrika* 70(1):41–55.
- Rosenbaum, P. & D. Rubin(1983b), “Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome”, *Journal of the Royal Statistical Society Series B* 45(2):212–218.
- Rosenbaum, P. & D. Rubin(1985), “Reducing bias in observational studies using subclassification on the propensity score”, *Journal of the American Statistical Association* 79(387):516–524.
- Rubin, D. B. (1974), “Estimating causal effects of treatments in randomized and non-randomized studies”, *Journal of Educational Psychology* 66(5):688–701.
- Thistlewaite, D. & D. Campbell(1960), “Regression-discontinuity analysis: An alternative to the ex-post facto experiment”, *Journal of Educational Psychology* 51(6):309–317.
- Vytlacil, E. (2002), “Independence, monotonicity, and latent index models: An equivalence result”, *Econometrica* 70(1):331–341.
- Wager, S. & S. Athey(2018), “Estimation and inference of heterogeneous treatment effects using random forests”, *Journal of the American Statistical Association* 113(523):1228–1242.
- Yang, S., G. W. Imbens, et al(2016), “Propensity score matching and subclassification in observational studies with multi-level treatments”, *Biometrics* 72(4):1055–1065.

Guido Imbens’ Methodological Contribution to the Analysis of Causal Relationships —A Review of the Academic Contribution of the 2021 Nobel Laureate in Economics

HONG Ni YU Leng

(Shanghai Jiaotong University, Shanghai, China)

Abstract: Guido W. Imbens was shared the 2021 Nobel Memorial Prize in Economics for their contributions to the methodology of causal analysis. This paper focuses on Imbens’ achievements in three aspects. Firstly, by combining instrumental variables with a potential outcome framework, Imbens proposed the design-based local average treatment effect (LATE) framework accounting for heterogeneity and incomplete compliance in natural experiments and provided a transparent interpretation of the estimates. Secondly, Imbens extended the LATE framework, enabling a wider range of applications, and also explained the external validity of LATE. Thirdly, Imbens made outstanding contributions to causal inference methods such as propensity score matching, regression discontinuity designs, difference-in-differences and machine learning. As many major issues in social sciences involve causality, Imbens’ research have greatly enhanced people’s ability to answer important causal questions, and is of great benefit to both academia and society.

Keywords: Guido Imbens; Nobel Memorial Prize in Economics; Causal Inference; Local Average Treatment Effect; Instrumental Variables

(责任编辑:李仁贵)

(校对:刘洪愧)