

公共政策评估的社会实验方法:理论综述

余向荣

摘要:对公共政策可信的评估为政府决策和公众参与提供必要的依据。社会实验通过随机分配的方法将社会实验的参与者分成实验组和控制组,进而考察社会项目对参与者的影响,是进行公共项目评估的基本方法。但是,和一般的非实验计量方法相比,该方法本身也存在难以克服的缺陷,例如随机分配以及此过程中实验参与者行为的改变会引起随机偏误等问题。因此,我们应该谨慎地解释实验数据,尤其当控制组的参与者存在可替代的社会项目,以及实验组的部分参与者为追求更好的项目而退出实验的时候。社会实验方法在中国的研究和实践才刚刚起步,但有着广泛的应用前景。

关键词:随机 社会实验 社会项目 经济计量

一、引言

公共政策的实施效果与社会福利紧密联系。对公共政策进行可信的评估是政府决策的重要依据。如何将有限的预算资金分配到最有效的社会项目上,政府需要在统计和数量上对项目进行细致的评估。而对于公众来说,决定是否参与社会项目也依赖于该项目的价值,依赖于该项目是否能够改变参与者的收入和福利水平。

社会实验方法和经济计量方法(非实验方法)是进行社会项目评估的主要手段。前者通过随机分配的方法将社会实验的参与者分成实验组和控制组,通过对照两组参与者的经济表现(如收入)进而考察社会项目的价值;而后者则使用大量的微观数据、统计方法和计量模型来比较项目的参与者和未参与者之间的收入差异。我们要讨论和评介的主要是实验方法。为了使本文的讨论更加具体,我们主要以美国《职业训练协作法》(Job Training Partnership Act, JTPA)下的一系列社会项目为例。美国的“国家JTPA研究”(National JTPA Study)为了对JTPA项目进行评估,在16个职业培训中心进行了实验,获得了很好的实验数据。在社会实验领域内的很多研究也都是建立在该数据的基础上(如 Heckman, Hohmann and Smith, 2000; Heckman, Ichimura, Smith and Todd, 1998; Heckman and Smith, 1997)。

在考虑培训项目的价值时,我们首先要问,培训对于受训者有没有影响,影响有多少?无论哪种评估方法都必须回答这一基本的问题,但是就参与者

个体而言,这是一个假设性问题。因为要考察培训对受训者的影响,那就应该比较培训之后参与者的工作收入和他在未受培训状况下能获得的收入,但未受培训的状态其实是假设的:在这个同一时点上,参与者个体不可能同时经历受训和不受训两种状态。简单地说,我们也许可以观察到一个人在培训之后挣得的收入,但我们并不知道此时,如果没有受培训,他能挣多少收入。因此,直接的比较是不可行的。社会实验的思路在于以实验组和控制组的期望收入分别代表两种状态下参与者的收入水平,然后用期望收入的差值来估计项目的实际影响。这相当于是用统计的方法,将估计培训项目对个体的影响转化为考虑项目对受训群体期望收入的影响。因此,这个方法的可信程度,显然严重地依赖于如何将社会实验的参与者随机地分为实验组和控制组。这种古典实验的方法被 Heckman 和 Smith (1995a) 称为实验方法的“黑箱”版本(Black-box Version)。

无论用哪种方法对社会项目进行评估,它的准确性同时依赖于数据的质量和模型的假设,这需要适当地权衡。社会实验方法,一方面解决了计量方法中无法避免的选择性偏倚(Selection Bias)问题,但另一方面它在随机分配过程中产生的随机偏误(Randomization Bias)却也是其本身难以超越的。例如,它假设随机分配不会改变参与者的行为,评估的准确性很大程度上依赖于这一行为假设的可接受程度。事实上,行为一旦发生改变,评估就会产生随机偏误。社会实验方法还假定控制组的参与者不存在相关或类似于实验组培训项目的可替代性选择,实

实验组的参与者也不会中途退出、去参与更好的社会项目。显然,这些假定不成立的情况也是常常发生的,因此就可能造成替代性偏误(Substitution Bias)和退出问题(Dropping out Problem)。所以,我们强调,对实验数据应该进行谨慎的解释。此外,社会实验还面临其他一些问题,如数据上可能与总体分布不一致,制度和伦理层次上也有诸多限制等。而且,社会实验的方法并不能就公共政策的各个方面给出满意的答案,它提供的评估只能是有限的一个或几个方面,更多我们感兴趣的话题,至少借助一种实验方法是无法单独讨论的。例如,JTPA项目中,参与者的家庭背景,受教育程度等都可能影响参与者的经济表现,但是在随机分配过程中,我们实际上很难保证每个特征都是随机的。这些问题在原则上可以通过随机实验进行评估,但是往往缺乏操作的可能性。如,我们也许可以按照性别等因素将参与者随机分为两组,但是家庭背景等因素就很难做到随机了。因此,无论作为一种单独的评估方法,还是作为实验数据的分析工具,非实验方法(计量方法)也同等重要。

二、基本模型

评估的问题产生于个人不可能同时经历培训和未受培训两种状态。如果一个人接受了培训,我们就只能观察到他在培训状态下的收入,而无从猜测他在未受训状态下的收入,但是只有比较两种状态下的收入水平,我们才能评价该培训项目对参与者是不是有效。本节中我们将讨论社会实验方法如何解决这一假设性的政策评估问题,基本的模型基于Heckman(1997)和 Heckman与 Smith(1995a,1997)。

首先,我们定义一个变量 $D: D=1$ 表示个体 $I(I=0, 1, 2, \dots, n)$ 参与了某个具体的社会培训项目,获得 Y_1 的收入; $D=0$ 表示个体 I 未参与该项目,获得收入 Y_0 ,即:

$$Y = \begin{cases} Y_1 & \text{如果 } D=1 \\ Y_0 & \text{如果 } D=0 \end{cases} \quad (1)$$

假定收入水平 Y 由 X 决定,其中 X 穷尽了 Y 的所有决定因素,既包括与培训相关的各方面,也包括家庭背景、个人条件以及市场条件等等。如果考虑线性情形,则有:

$$E(Y_1|X) = g_1(X) = X \beta_1 \quad (2a)$$

$$E(Y_0|X) = g_0(X) = X \beta_0 \quad (2b)$$

这里,我们分别用条件均值和函数形式来表示 Y 与 X 的关系,由条件均值的性质可知,条件均值分别等于 Y_1 和 Y_0 ,因此有: $Y_1 = E(Y_1^{1/2} | X)$, $Y_0 = E(Y_0^{1/2} | X)$ 。

用变量 Z 代表个体 I 决定是否参与该培训项目时考虑的因素,包括个人、家庭、市场和制度等因素,

则有:

$$Z \in C, D=1; Z \notin C, D=0 \quad (3)$$

其中 C 为 Z 的可行集,当 I 的条件落入该集合时,他就参加培训项目,否则不参加。简便起见,我们用 $C = (X, Z)$ 表示所有解释变量,其中 X 和 Z 也可能是共同的变量或因素。

假定没有不可观测的因素,而且解释变量的样本元素有足够的变差,那么我们可以用实验组的数据决定 g_1 ,用控制组的数据决定 g_0 ,用联合样本的数据来决定 F 。由此,我们才可能倒推出每个人的 Y_1 和 Y_0 ,然后评估出样本中每个人参与培训能获得的净收益。只有了解 g_1, g_0 和 F ,我们在可行的样本中才能回答 Heckman 的三个基本问题:

(1) 培训项目对受训者的影响有多大?

(2) 培训项目对随机的参与者影响有多大?

(3) 补贴、广告、劳动力市场、家庭收入、种族和性别如何影响人们的申请决定?培训中心的业绩标准和劳动力市场条件如何影响培训中心的录取决定?

但是事实上,我们不可能观测到 C 中的所有元素,因为我们无法预知 Y_1 和 Y_0 ,也无法预知 D 的值。而且就算可以观测到每个人 C 集合中的所有元素,可行的样本中这些矢量元素也可能没有足够的变差。无法观测的因素和是否有足够的变差是引起评估问题的主要原因。正是这些缺失的元素,迫使我们以可行的信息集为条件,将 Y_1, Y_0 和 D 看作随机的变量。而这一随机化的过程排除了确定每个人的 Y_1 和 Y_0 的方法的可能性。替代的是,用统计的方法强调以可行的信息集为条件来估计的 Y_1, Y_0 和 D 联合分布情况(条件分布)。在部分信息可获得的条件下,我们可以得到 Y_1, Y_0 和 D 的联合分布函数为:

$$F(y_0, y_1, d|c) = \Pr(Y_0 \leq y_0, Y_1 \leq y_1, D=d|C=c) \quad (4)$$

其中,大写字母表示随机变量,小写字母表示它们的真实值。如果(4)式可以确定,而 C 的分布情况可知,那么我们可能在统计意义上回答上面三个基本问题:我们可以决定 Y_1, Y_0 的人口分布状况,以及培训的总收益 $E(Y_1 - Y_0)$ 的人口分布状况,并在 C 给定时决定 $D=1$ 的概率。具体地,在参与者预期收入的联合分布函数 $F(y_0, y_1, D=1, c)$ 已知的情况下,问题(1)可以在群体(Population)的意义上进行回答,因为此时我们可以决定影响 $E(Y_1 - Y_0)$ 的分布状况, $F(D=1, c)$ 。从期望值的角度,也可以给出答案:

$$E(Y_1 - Y_0 | D=1, c) = E(Y_1 - Y_0 | D=1, c) \quad (5)$$

同样道理,我们可以从(4)式以及各解释变量的分布状况通过初等的概率运算推导出所有人预期收入的联合分布函数 $F(y_0, y_1 | c)$,进而在所有人的意义上回答问题(2)。即,我们可以确定随机地把一个人从一种分布状态 $F(y_0 | c)$ 移到另一种分布状态 $F(y_1 | c)$ 的净所得,(6)式定义了净所得的期望。此外,而我们通过(4)式还可计算出一个人参与的概率 $\Pr(D=1 | c)$,以此回答问题(3)。

$$E(c | c) = E(Y_1 - Y_0 | c) \dots\dots\dots (6)$$

除了一般的文献中广泛讨论的培训项目对期望收入的影响,Heckman 和 Smith(1997)还讨论了分布函数 $F(y_0, y_1 | c)$ 和 $F(y_0 | c)$ 的其他一些重要指标。

(a) 从项目获得正收益的人占参与者的比例:
 $\Pr(Y_1 > Y_0 | D=1)$;

(b) 参与者培训净收益分布函数的分位数:
 $\inf\{q : F(Y_1 - Y_0 | D=1) = q\}$, 其中 q 为净收益分布函数的一个分位数;

(c) 在非参与者收入分布函数上选定一点,相对应应该点,参与者培训净收益的分布情况:
 $F(Y_1 - Y_0 | D=1, Y_0 = y_0)$

(d) 最大化两种收入的期望价值:
 $E[\max(Y_1, Y_0)]$

Heckman和 Smith指出,问题(a)考察了项目参与者中获得正收益的广泛程度,问题(b)使我们了解培训对从项目中获益最多(或者最少)的人的影响。问题(c)回答了培训效果的分布如何依赖于未受训状况下的收入。问题(d)则关心如果个人可以在 Y_1 和 Y_0 之间进行选择,例如提供劳动补贴的社会项目(此时他也可以选择不要补贴),在这种情况下, $E[\max(Y_1, Y_0)] - E(Y_0)$ 是培训项目产生的选择价值(Option Value)。进一步,即使(5)式和(6)式的期望影响为零,知道项目参与者中或总人口中哪部分人从项目中获益了,也是有意义的。我们可以通过 $F(Y_1 > Y_0 | D=1, c)$ 来计算参与者中获得正收益的概率,即 $\Pr(Y_1 > Y_0 | D=1, c)$ 。为了证实培训中心是否总是将最好的人选到培训项目中,我们必须知道 Y_1 和 Y_0 的联合或随机的依赖情况(Correlation or Stochastic Dependence),当然,这需要了解联合分布函数 $F(y_0, y_1 | D=1, c)$ 和 $F(y_0, y_1 | c)$ 的相关性质(Heckman and Smith, 1997)。

显然,所有的这些参数都需要了解 (Y_1, Y_0) 的联合分布函数。但是,即使是最理想的实验数据也无法决定(4)式的联合分布函数,还是因为一个实验不可能同时观测到任何人的 Y_1 和 Y_0 。下面我们将讨论实验方法如何解决这一评估的问题。但是,随机社会实验的方法几乎都是为了回答问题(1)和问题(2),只有极少的例外,例如 Bjorklund 和 Regner

(1996)使用实验评估了在项目的规模上边际变化带来的影响。

社会实验的方法中,我们引入新的变量 D^* ,类似于前面定义的 D ,表示在随机分配存在的情况下,一个人申请项目并被项目接受的情况。 $D^*=1$ 表示在随机分配的情况下,一个人申请参加到项目中;反之, $D^*=0$ 。我们用 $R=1$,表示这个人被分配到实验组,即接受了培训,反之, $R=0$ 。这里我们假定所有被分配到实验组的每个人都会参加培训,不会中途退出。在 $D^*=1$ 时, $R=1$ 的概率我们用 p 表示,显然,当 $p=1$ 时表示不存在随机分配。从直觉上我们需要强调,在存在随机分配的情况下,一个人参与项目,但不一定会受到培训,因为他可能被分配到控制组,此时,我们通常可以把参加项目的人称为子总体(Subpopulation)。而且,如果随机分配会影响个人的参与决策,那么 D 和 D^* 并不总是相等的。社会实验的标准方法中,随机分配是在 D^* 已知的情况下进行的,即,假如 $D^*=1$,那么 R 是独立于 (Y_1, Y_0, C) 的。

进一步,我们将存在随机分配情况下的收入定义为: (Y_0^*, Y_1^*) ,并且假定:

$$\Pr(D=1 | c) = \Pr(D^*=1 | c, p) \dots\dots\dots (AS-1)$$

$$F(y_1^*, y_0^* | c | D^*=1) = F(y_1, y_0 | c | D=1) \dots\dots\dots (AS-2)$$

(AS-1)式表示随机分配不会改变人们选择申请、参与项目的概率。(AS-2)式表明无论随机分配存在还是不存在,收入的分布情况都是相同的,从而 $(y_0^*, y_1^*) = (y_0, y_1)$ 。

在(AS-1)式和(AS-2)式成立时,我们有:

$$F(y_1, c | D^*=1, R=1) = F(y_1, c | D^*=1) = F(y_1, c | D=1) \dots\dots\dots (7a)$$

$$F(y_0, c | D^*=1, R=0) = F(y_0, c | D^*=1) = F(y_0, c | D=1) \dots\dots\dots (7b)$$

因此,

$$E(Y_1 | R=1) - E(Y_0 | R=0) = E(Y_1 - Y_0 | D=1) \dots\dots\dots (8)$$

至此,我们在期望的意义上给出了问题(1)的答案:只要简单地比较子总体中受训者和未受训者(他们被随机地分配到控制组)的收入水平即可得到培训者的影响。这是因为解释变量 C 以 R 为条件的条件分布对于样本中的受训者和非受训者是相同的。社会实验通过将项目的参与者进行随机分配直接得到了缺失的假设状态下的收入分布状况: $F(y_0, c | D=1)$,从而解决评估的问题。而且,社会实验方法得到的结果是理想的,因为即使未受过统计训练的政治家和管理当局都可以理解这一简明的表

达,也无须对数据进行精密的统计上的调整或者函数形式的假定。对于强制性参与的项目来说,(8)式也是进行成本-收益分析的基础,当然关于社会成本的计算是另一个重要话题。

三、社会实验中的随机偏误

社会实验运用随机分配的方法得到控制组的收入,并用它替代实验组在未受培训的假定状态下的收入水平,然后通过直接的比较来评估社会项目对参与者的影响。通过这一方法得到的结果,其可信程度严重依赖于方法本身的假定。但是这些假定往往是值得怀疑的,也可能是不符合现实的。如果(AS-1)式不成立,即存在随机分配时项目参与者的特性不同于没有随机分配的情形,那么我们从实验数据中得到只是 $E(D^*=1,c)$,而不是 $E(D=1,c)$,此时只有被评估项目本身就通过随机分配执行, $E(D^*=1,c)$ 才反映该项目的实际效果。我们可以认为该假定对于一些 p 是成立的,或者在实践中是成立的,但是我们同样也有足够的理由去怀疑它的正确性。不可避免地,如果随机分配影响了参与决策,那么两个参数事实上完全不是一回事。在这种情况下得出的评估是存在偏误的。我们把这种由随机分配造成的偏误称为随机偏误。由于随机过程是社会实验方法的核心,因此随机偏误的问题是实验方法本身无法超越的。

同样道理,(AS-2)式也可能不成立,引起随机偏误。一方面随机分配可能改变参与者的行为,例如被分配到实验组的人可能会更加努力,进而改变收入水平,此时得到的结果不仅包含了项目的影响,而且包含了随机分配的激励作用,因此评估是扭曲的。另一方面随机分配也可能改变培训中心的行为,例如,随机分配将使得一部分人成为控制组,培训中心可能不得不通过各种途径增加项目的申请者,因此申请者的个人条件会发生变化,收入水平也会产生偏误,又如随机分配的存在,录取决定可能会更加随意,此时受训人员的条件也可能不同于没有随机分配的情形,也会产生随机偏误。

事实上,正如 Heckman和 Smith(1995a)所强调的,随机偏误不仅仅是一个理论问题,在实际的社会实验中确实广泛存在,但遗憾的是它们往往被忽视了,相关的研究也还很少。

四、替代偏误与退出问题

社会实验的方法假定实验组和控制组除了是否得到培训之外所有条件的分布状况都是相同的。但是毕竟不是自然实验,即使是培训本身,社会实验也只能做到部分控制。控制组和实验组的区别只在于

他们是否可以选择不参加我们所要评估的这个社会项目,但是并不能排除他们去选择其他类似的项目。具体地以JTPA项目为例,控制组的人员可能选择其他类似的培训项目,或者是私人企业提供的,或者是公益组织提供的,而不是如我们设想的那样不接受任何其他相关或类似的培训;同样道理,实验组的人员也可能因为其他更好的培训项目而退出JTPA,或者在JTPA之外又接受了额外的培训。无论是替代还是退出都会影响实验组和控制组的收入差距,如果我们仍然把这一差距作为对JTPA价值的评估,那么便会产生替代偏误和退出问题。

Heckman, Smith和 Taber(1994)较早讨论了社会实验中退出问题,并且提供了对实验数据进行统计、经济调整的方法。Heckman, Hohmann和 Smith(2000)在真实实验的基础上第一次系统地检测了实验组退出和控制组替代问题的严重程度,他们首先认为应该在两个基本的问题上做出区分:

问题(I):在存在其他类似的培训选择时,JTPA培训对期望收入的影响是多少?

问题(II):相对于没有任何培训时,JTPA培训对期望收入的贡献是多少?

问题(I)要比较的是接受JTPA培训下的收入和接受其他培训时的收入,评估的是JTPA项目的相对价值,即哪个项目更有效的问题。而问题(II)考察的是接受JTPA培训时的收入与完全没有培训时的收入,评估的JTPA的绝对价值,即JTPA是否有效(Work)的问题。因此,如果对实验组和控制组简单地进行收入比较,在存在可替代项目或退出问题时,会严重低估项目本身的价值,我们对实验数据应该进行谨慎的解释。随机偏误和退出问题的大小和社会项目本身的可替代程度有关,例如对于一般培训来说也许可替代的项目很多,但对于某些具体技能的培训可替代的项目要少得多。显然地,强制参与的社会项目不存在替代和退出的问题,而在自愿参与项目的评估中应该更多地考虑这些因素。

Heckman等(2000)在JTPA(CT-OS)项目的实验数据基础上估计出替代和退出使得实验组接受的培训对实验组和控制组之间收入差距的贡献为0.2,而不是通常认为的1.0。他们还将计量方法应用到替代和退出问题存在时实验数据的解释,他们对实验组的数据进行计量的结果表明,在培训的过程中培训对收入有大的负影响,而培训完成后,该项培训则表现出巨大的正的影响。他们的研究为JTPA项目的真实效果提供了乐观的看法。

作为本节的结尾,我们将简要考察社会实验方法其他方面的不足(也可参见 Bjorklund Regner,1996的概括)。首先,社会实验能够回答的问题非常有

限,也许只能给出公共政策一个或者几个方面的评价。在前文列举的问题中,到目前为止我们还只能在统计意义上给出问题(1)的答案,问题(2)(3)(a)(b)(c)(d)都还不能解决,或者至少不能通过一个实验单独解决。还有在本文中未列举、但在实践中非常重要的问题也可能无法用实验方法回答。其次,社会实验不仅经济上因为它的高成本而受到限制,而且在制度和伦理层面上也受到诸多的约束。一项有价值的社会实验需要足够广的范围、足够长的时间才能取得质量好的实验数据,成本巨大,往往需要有专门的机构来承担。在实际的操作中也不可避免遭受制度和伦理层次的诸多限制,例如一项关于个人收入、生活甚至是命运性的社会项目怎么可以由随机分配的方式决定谁去参与?这可能是令人难以接受的。Devine和 Heckman(1996)甚至对JTPA本身的合法性问题提出了怀疑,更不要说实验上的障碍了,他们强调他们的理论和经验研究表明,JTPA作为联邦最大的职业培训项目,在提供和执行的过程中导致了不公平、不一致、管理的分散等问题,甚至可以说是荒谬的(Perversity),违背了公共品的基本原则。他们认为这是为了追求经济效率的目标对公平价值的牺牲。再次,社会实验还可能受到官僚行为的影响,这与文化和社会环境等因素联系在一起。最后,实验数据本身是残缺的、甚至会出现不一致,因此除非进一步的假定,不能就项目效果的总体分布情况提供有用的信息。例如实验数据只给出了实验组和控制组收入的分布状况,没有给出两种状态下收入的联合分布情况,因此我们无从知道两种状态下二者的具体联系(HeckmanandSmith,1997)。

五、非实验方法:简要评论

对于 Y_1 可观测的人,他 Y_0 数据缺失的问题,使得我们考虑用未参与项目者的收入来推断参与者在假定未参加状态下的收入。社会实验是一种方法。而政府当局所公布的社会项目的记录数据,以及一般的非实验分组比较也使得我们可能确定参与者的收入分布 $F(y_1|D=1,c)$,和未参与者的收入分布 $F(y_0|D=0,c)$ 。从社会项目的记录数据,有时也可以获得个人参与项目的概率 $Pr(D=1|c)$ 。非实验方法或者计量方法试图通过这些数据解决评估的问题,当然,要确定(4)式和(5)式需借助进一步的假定。

一般地,对于相同的 $C=c$,不管是在参与的情况下,还是在假设没有参与的情况下,项目参与者的收入分布函数,都会不同于未受训者的收入分布。即:

$$F(y_0|D=1,c) \neq F(y_0|D=0,c)$$

$$F(y_1|D=1,c) \neq F(y_1|D=0,c) \dots\dots\dots (9)$$

这会引引起收入分布中的选择性偏倚问题,在期望的意义上,可以表示为:

$$E(Y_1|D=1,c) - E(Y_0|D=1,c) \neq E(Y_1|D=0,c) - E(Y_0|D=0,c) \dots\dots\dots (10)$$

简单地讲,即使所有可行的解释变量都相同,C相同,参加者在假使未受训的状态下的期望收入也不同未参与者在该状态下的预期收入,这是因为项目的参加者和未参加者是不同的。例如,项目的参与者可能更加努力,假如努力可以增加收入、影响参与的积极性,那么参与者将比未参与者在相同的条件下赚得更多收入。在评估中没有控制这一个明显的差别,将会把那些参与者因更加努力而获得的收入错误地归结为项目本身的贡献。这就是选择性偏倚问题。社会实验用随机分配的方法克服了这一问题,因为随机化使“是否参与”的决策在统计意义上独立于 (Y_1, Y_0, C) ,当然它本身也带来了随机偏误。计量方法受制于数据选取的方式本身,而无法超越选择性偏倚问题。Heckman, Ichimura, Smith 和 Todd(1998)运用半参数方法讨论了计量分析实验数据时的选择性偏倚问题。

在讨论了非实验方法面临的选择性偏倚问题之后,我们要简要地介绍一下几种可选择的计量方法(详细的讨论参见 HeckmanandSmith,1996)。在社会项目的评估中,一方面我们需要对面板数据进行分析,例如我们可能要考虑在培训时,各个解释变量对收入的贡献,也可能要考虑培训完成后各解释变量,包括培训累积的人力资本对受训者收入的贡献,这些都是截面的问题。工具变量方法在面板数据分析中应用非常广泛,在评估问题中,我们可以写出标准的回归方程:

$$Y = \beta_0 + \beta_1 D + U \dots\dots\dots (11)$$

其中 U 为误差项,期望为 0, $D (=0,1)$ 如前文定义的表示是否参与。当然,我们也还可以控制更多的变量,原则上,我们可以估测社会项目对收入的贡献,即项目的价值,但是这建立在对数据和方程的严格假设之上。另一方面,我们可能需要对社会项目的长期影响给出评价,这需要分析纵向的记录数据,需要借助纵向方法(LongitudinalMethods),一般地,我们可以表达为:

$$Y_{jt} = g_j(X_{jt}) + U_{jt}, j=0,1 \dots\dots\dots (12)$$

其中 X 如前文所定义,决定收入函数,我们也可以考虑收入函数为线性的具体形式。时间序列分析的许多方法在纵向数据的分析也适用。此外,固定影响方法(Fixed-effectMethodes)在社会项目的评估中也有应用的价值,一般的形式为: $X_{j,t} = \alpha_j + X_{j,0}$,其中 α_j 对所有人来说都是相同的。这种方法

的基本的假定是：

$$E(U_i - U_i | D, X) = 0, \quad \forall t > k > t \quad \dots \dots (13)$$

这些不同的方法建立在不同假设的基础上考察了社会项目的不同方面,那么如何在可选择的非实验方法中选择合适的来解决评估问题呢?LaLonde (1986) 在分析国家劳动培训示范项目(National Supported Work Demonstration, NSW) 的实验数据中认为这种选择是相当困难的、甚至是不可能的,因此需要借助社会实验方法。Heckman和 Smith(1995a) 反驳了这种看法,指出了 Lalonde(1986) 研究中的许多不足,认为这种非实验方法之间的选择是可能的。事实上,计量方法无论是作为独立的评估方法,还是作为实验方法的分析工具,都为社会项目的评估提供了洞见。

六、中国的研究与应用

应用随机社会实验的方法评价公共政策,无论在理论上,还是在实践上,都是重要的尝试。美国的 JTPA 项目得到的实验数据引发了大量的理论和实证研究,但是这些研究也告诉我们,仍然有大量的问题需要进一步探索。例如,在已有的文献中,讨论社会项目对参与者的影响,实际上都是考察它对收入(或者产出)的影响。但是就个人而言,他关心的是自身的效用是否因为接受培训得到了提高,即自身的福利水平。一个能够增进参与者收入水平的社会项目并不必然改善他的福利水平。简单地,只要引入闲暇即可印证这种情况,对于一个看重闲暇的人来说,参与项目带来收入的提高如果并不足以弥补他因此而放弃的闲暇,那么他的效用水平不会因为参与得到改善,而福利水平的改善才是社会项目真正的效果。因此,为了更加真实地考察社会项目对参与者的影响,我们需要引入效用函数进行更加细致的考察,至少在理论上这样的考虑是可行的。但是实验本身几乎没有任何操作的可能性。

另一方面,我们也关心社会福利状况是否因为社会项目得到改善。最基本的我们希望给出一个成本-收益的简单分析,相当于要回答社会项目是不是值得的问题,这不同于前面讨论的项目对个人是否有影响的问题。在这里我们不仅要计算出社会项目对所有人影响的总和,而且需要关于项目计划、执行的社会成本。这都需要极为复杂的分析,单就社会成本的讨论就需要很多因素:基本的是项目需要资助的资金以及这些资金的机会成本,也应该考虑在执行过程中因讨价还价等行为带来的效率损失(简单地讲,即“交易费用”),还要考虑项目(例如补贴)对市场机制的扭曲带来的效率损失,何况还有信息、搜寻等问题。总之,无论在个人层面上,还是在

社会层面上,对社会项目的影响进行福利分析是我们感兴趣的话题,但是在实验的操作上还非常困难。要解决这些困难,应该要设计出更加理想的实验模型,并且借助更加精巧的分析工具,至少需要突破实验方法的“黑箱”版本。

事实上,社会项目评价的实验方法和非实验方法并不是相互排斥,非此即彼的。在实际的分析中,把两种分析方法结合起来使用,会得到更加精确的评估。例如,Heckman 等(1998)用实验的数据考虑了计量分析中的选择性偏倚问题,Heckman 等(2000)将计量方法引入到有替代偏误和退出问题的实验数据的分析中。实验方法和计量方法更加紧密地结合起来使用得到的评估结果往往比单纯使用一种方法更加理想、接近真实。

在实践中,用社会实验方法来评估公共政策并不多,美国的 JTPA 项目是一个典型。此外,在欧洲许多强调福利政策的国家里,社会实验方法的评估也少之又少,Bjorklund 和 Regner(1996) 是用实验方法研究欧洲劳动力市场政策的一个例子。

在中国,类似的研究和实践则几乎完全是空白。处于经济转型的过程中,政府在经济运行和发展中扮演的角色非常重要,发挥着巨大的作用。但是,公共政策的评价体系完全没有建立起来。无论是在理论研究,还是在经验研究中,这一领域的进展远远落后于现实经济的发展。例如,随着国有企业改制和劳动力市场的发展,下岗工人(失业)的数目非常庞大,政府为此出台了一系列的措施帮助下岗工人再就业,既包括技术培训、就业帮助等项目,也包括税收减免、提供补贴等优惠措施。那么这些政策是不是真正有效,多大程度上有效?我们还不能提供统计上或数量上答案。但是政策评估对中国政府的决策却是迫切需要的,一方面,公共政策的背后是资金支持,政府收入来自于纳税人的贡献,政府有责任把这些资金配置到最有效的项目上去,这需要提供数量上的评估,另一方面作为发展中国家,许多地方需要政府资金支持,政府需要把有限的资金配置到最能促进经济发展的项目上去,这也必须建立在可信的评价体系之上。因此,公共政策评估在中国是一个非常迫切而发展又相对滞后的领域。

某种意义上,中国经济改革本身就是一个最大的社会实验。政策“试点”的方法也体现了公共政策出台的谨慎性,这也是社会实验的一种思想——虽然它并没有建立在随机分配的基础上。例如农村税费改革在安徽试点,农村合作医疗在浙江试点,这样的试点可以为政府决策提供参考。但是,这只能是非常粗糙的估计,存在很大的偏差,毕竟其他省份的情况可能完全不同于安徽或者浙江,在试点省份非

常有效的办法未必能够解决全国的问题。因此,要为公共政策提供更加精确的评估,除了借助计量方法之外,随机实验方法为我们提供了很好的工具,由此得到的高质量数据也为精确的计量分析提供了素材。此外,随机实验方法的应用也面临很多困难。随机实验需要足够广的范围,足够长的时间才能获得可靠的数据,因此成本很高。个人的力量是无法完成这样大规模的实验的。在美国也是由劳动部资助两个专门的实验机构完成 JTPA 项目的评估实验的。在现有的条件下,中国要进行这样大型的实验需要借助各方面的力量。

七、结论性评述

公共政策的评估是政府决策和公众评价的重要工具,但是评估问题面临数据缺失的困难,这一困难在个体的意义上不可能超越,因为任何人不可能同时经历参与项目和不参与项目两种状态,我们自然也无法观测到两种状态下同一个人的收入状况。于是我们只能寻找这一假设性问题在群体意义上统计性的答案。随机分配的社会实验是一种方法,但是它本身也会带来随机偏误的问题,这是随机分配的方法本身所带来的。而在现实的实验中,我们只能部分控制实验过程,因此控制组的人员可能接受其他类似的培训,实验组人员也可能退出实验追求更好的项目,这便产生了替代性偏误和退出问题。通过计量方法的分析可以一定程度上解决这一问题,因此我们对于实验数据应该进行谨慎的解释。而非实验方法,无论是作为独立的评估手段,还是作为实验方法的分析工具,都是极为重要的,当然它本身也会面临选择性偏倚问题。在综述性地评介了 Heckman 等人在个人数据分析上贡献之后,我们还讨论了社会实验方法在评估公共政策时进一步研究的方向,包括福利分析等。而对于中国来说,政府在经济中扮演着重要的角色,用社会实验对公共政策进行评估非常迫切,尽管已有的研究非常有限,也存在诸多困难,但作为一个新的领域并不会因为这些阻碍而停滞。

注释:

该法案规定政府资助职业训练,其目的是帮助具有特殊就业障碍的人和提高劳动力水平、促进经济发展。前者完全由联邦政府资助,后者主要由州政府资助,大多由地方、教区的培训中心管理和执行。JTPA 的内容包括:针对能力低下青年和成人的训练(条款 A);夏季青年就业训练项目(条款 B);对不能返回原工作岗位的下岗工人的就业训练(条款 C)等。其中第一项又包括职业技术培训(Classroom Training in Occupational Skills, CT-OS),基本教育(通常为普通教育发展考试 General Educational Development Test 准备),为私人企业的在岗培训提供工资补贴,以及在寻找工作方面提

供帮助等项目。

问题(1)和问题(2)是不同的,在自愿参与的培训项目中,问题(1)回答的是项目对于自愿参与培训者的影响,而问题(2)回答的是项目对于随机选择的一个人能产生的影响,而不管他本来是否会自愿参与该培训。当然,在强制参与的社会项目中,例如必需的劳动保险等,不存在问题(1)和(2)的差别。

在本文中我们有的时候将项目对受训者的影响、受训者培训得到的总收益、项目的价值等若干概念混合起来使用,但是事实上对于部分个体来说,培训项目对他收入的影响是负的,而不是正的。

参见 Heckman and Smith (1997),他们集中定义、评估了 JTPA 的选择价值。

即 Y_1 和 Y_0 都很高的人,在这种情况下,培训中心可能有良好的业绩表现,但是培训项目本身可能违背了它的初衷。这事实上也是实验方法存在的主要理由。

关于福利分析可参见 Heckman 和 Smith (1995b)。

参考文献:

1. Bjorklund, Anders and Regner, Hakan, 1996. "Experimental Evaluation of European Labour Market Policy," in G. Schmid, J. O'Reilly and K. Schömann, eds., *International Handbook of Labour Market Policy and Evaluation*. London: Edward Elgar, pp. 89-114.
2. Devine, Theresa J. and Heckman, James J., 1996. "The Economics of Eligibility Rules for a Social Program: A Study of the Job Training Partnership Act (JTPA) — A Summary Report." *Canadian Journal of Economics*, 29 (Sp. Iss.), Part 1, pp. 99-104.
3. Heckman, James J., 1978. "Dummy Endogenous Variables in a Simultaneous Equation System." *Econometrica*, 46(4), 931-59.
4. Heckman, James J., 1997. "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations." *Journal of Human Resources*, 32 (3), 441-62.
5. Heckman, James J.; Hohmann, Neil and Smith, Jeffrey, 2000. "Substitution and Dropout Bias in Social Experiments: A Study of an Influential Social Experiment." *Quarterly Journal of Economics*, 115(2), pp. 651-94.
6. Heckman, James J.; Ichimura, Hidehiko; Smith, Jeffrey and Todd, Petra, 1998. "Characterizing Selection Bias Using Experimental Data." *Econometrica*, 66(5), pp. 1017-198.
7. Heckman, James J. and Smith, Jeffrey, 1995a. "Assessing the Case for Social Experiments." *Journal of Economic Perspectives*, 9 (2), pp. 85-110.
8. Heckman, James J. and Jeffrey Smith, 1995b. "Evaluating the Welfare State." *University of Chicago, Memo*.
9. Heckman, James J. and Smith, Jeffrey, 1996. "Experimental and Nonexperimental Evaluation," in G. Schmid, J. O'Reilly, and K. Schömann, eds., *International Handbook of Labour Market Policy and Evaluation*. London: Edward Elgar, pp. 37-88.
10. Heckman, James J. and Smith, Jeffrey, 1997. "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts." *Review of Economic Studies*, 64 (4), pp. 487-535.
11. Heckman, James J.; Smith, Jeffrey and Taber, Christopher, 1994. "Accounting for Dropouts in Evaluations of Social Experiments." *NBER Technical Working Paper*, No. 166.
12. LaLonde, Robert J., 1986. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *American Economic Review*, 76 (4), pp. 604-620.

(作者单位:北京大学经济学院 北京 100871)
(责任编辑:N、Q)