

# 科学的经济学与经济学的科学贡献

## ——从微观计量经济学的发展看经济学的科学贡献

陈永清

**摘要:** 美国经济学家詹姆斯·赫克曼和丹尼尔·麦克法登因在微观计量经济学中的杰出贡献而荣获 2000 年度诺贝尔经济学奖。本文在一个相对统一的框架下阐述了赫克曼校正法,也即针对具有“选择性偏倚”和“自选择”特征的数据样本进行统计分析的方法,其次分析了麦克法登的基于多种选择的定性反应计量模型。从迄今为止的理论成果来看,他们的研究成果不仅在经济学和统计学之间架起了一座桥梁,而且极大地拓展了经济计量方法的应用领域,具有广泛的适用性。

**关键词:** 微观计量经济学 选择性偏倚 定性反应模型

“经济学是否能够称为一门严格的科学?”这是人们在评价经济学研究成果时经常会想到的一个问题。经济学因其研究对象的复杂性,常常使得研究者感叹经济学的理论方法、研究手段、分析工具与研究需要之间的巨大差距。经济学也被称为黑板上的科学,经济学家更是被称为“模型中人”——其结果是经济学家提出的理论似乎只能运用于抽象的经济模型之中,有如古希腊人只是存在于圣经故事中一样。

但是,经济学理论发展至今,显然得益于人类的理性思考,经济学也因其理性思考和严密的推断而逐渐成为一门科学。一年一度的诺贝尔经济学奖对获奖者的肯定重在其理论的科学贡献。瑞典皇家科学院在宣布 2000 年度诺贝尔经济学奖时称,詹姆斯·赫克曼和丹尼尔·麦克法登两位教授因在微观计量经济学领域所作出的杰出贡献而荣获本年度诺贝尔经济科学奖。两位获奖者在微观计量经济学领域的主要贡献是,他们发展了已被广泛用于对个人和家庭行为进行统计分析的理论和方法,其中赫克曼发展了针对选择性抽样数据进行分析的理论和方法,麦克法登发展了关于离散选择分析和定性反应分析的理论和方法。他们的工作和努力不仅使得经济学家能够更好地分析微观决策行为,如决定人们是否去工作以及工作多长时间的因素是什么,经济因素是如何影响人们对教育、职业和居住地所进行的选择,甚至也可以分析人们究竟是如何在开车上班和乘坐公共交通工具上班之间进行选择的。与此同时,他们的研究成果还可以广泛地运用于其他领域的科学研究,选择性抽样数据的分析方法和定性分析理论虽然主要源于,并发展了微观计量经济分析,但同时也直接发展了统计分析的理论和方法,具有广泛的适用性。

—

微观计量经济学是介于经济学和统计学之间的边缘科学,通过研究微观层面(个人、家庭、厂商等)的经济数据来揭示微观经济的信息、特征和反应。采集微观数据是微观计量经济学研究的基础和前提,由于微观数据的大量获取只是到 20 世纪 70 年代中期以后才逐渐成为可能,因此,赫克曼和麦克法登的研究成果基本都产生在 70 年代中期以后,而且大

部分集中在 80 年代,显示微观计量经济学至今仍是一门新兴学科。

依据统计数据来分析和推断对象总体的特征是统计分析的基本思路和一般方法。经济学出于研究的需要,总是要不断将抽象的理论分析与现实的统计数据相对照,从而判定理论与现实的拟合程度。通常人们总是期望一个好的理论(包括其前提、特征和结论)能够在现实的经济数据中找到其对应,对此几乎不会有人提出异议。然而赫克曼教授所提出的问题是,既然经济学研究如此倚重于对实证数据的分析,那么微观经济数据本身的特性就值得我们重视,是否存在这种可能:由于研究中对于微观数据特征的忽视而致使惯常的分析思路和处理方法带给我们的结论是存在系统偏误的?赫克曼的研究表明产生这种偏误的可能是存在的,并且导致系统性偏误的原因正是由于微观经济数据往往存在“选择性偏倚”(Selection Bias)和“自选择”(Self-Selection)的特点。出于纠正这种系统偏误的需要,赫克曼发展了针对“选择性偏倚”和“自选择”数据样本的统计分析方法:赫克曼校正法(也称两阶段法,赫克曼法则或赫克曼 $\lambda$ 法),迄今这一方法不仅早已成为微观计量经济学的一个标准工具,而且还被广泛地运用到了其他分析领域。

所谓“选择性偏倚”是指,样本数据的形成不是来源于真正的随机抽样,而是具有选择性的。这种选择性或者是来源于数据采集者自觉或非自觉的选择,或者是源于数据生成主体的自我决策(也即自选择问题),同时也可能源于制度性约束。统计分析的目的主要是希望通过对样本数据的分析来推断对象总体的特征,在社会和经济分析中尤其如此,因为毕竟我们不可能搜罗到相关问题的所有分析对象,因此我们总是希望(通常更多的是假设)所收集的样本数据具有足够的代表性,能够充分代表对象总体本身。从统计方法上看,一般要求样本的形成具有真正的随机性。但是,经济数据的生成过程极少具有所要求的随机性,撇开统计工作者在收集数据时掺杂的主观选择以外,制度性约束和经济行为主体(个人、家庭、企业甚至包括政府部门在内)的决策行为更是造成经济数据存在选择性偏倚的重要原因,例如:在对工作时间和

工资水平进行数据采集时, 样本数据反映的只是被观察者选择参与工作时的数据, 那么, 当他不参与工作时的数据以及那些没有工作的个人的相应数据显然无法获得; 又如: 为了分析社会成员受教育程度与其收入水平的关系, 以是否完成大学学业为例, 我们比较的通常只是大学毕业生与未上过大学的个人的收入水平, 但是, 真正有意义的比较是大学毕业生的现有收入水平与他(她)如果未上过大学的状态下的收入水平, 显然后者在行为主体选择完成大学学业后根本无法获得。这些都是典型的带有选择性偏倚和自选择特征的数据样本, 而这些又是经济分析中最为司空见惯的形式, 因此选择性偏倚和自选择问题是广泛存在的, 忽视其存在就会导致经济分析的系统偏误。

赫克曼的主要贡献就是针对导致这种系统性偏误的原因, 提出了被称为“赫克曼校正法”(Heckman's Correction)的处理方法, 从经济计量学的理论发展来看, 赫克曼的工作直接发展了托宾(James Tobin)关于删截数据样本的计量分析理论, 因而也被冠以“赫克特”(Heckit)的名称, 与“托比特”(Tobit)相对应, 足见理论界对其研究工作的肯定。为说明赫克曼方法的基本思路, 我们以托宾关于家庭耐用消费品开支的经典分析为出发点。

为了考察社会各家庭在耐用消费品上的开支状况, 分析耐用消费品开支与家庭收入的相关性是一种当然的研究选择。为此, 我们可以收集到许许多多家庭关于耐用消费品的支出状况, 以及各家庭的收入状况。用  $y_i$  表示第  $i$  个家庭在某些耐用消费品上的支出,  $x_i$  为影响各家庭耐用消费品开支的因素(其中包括家庭的收入水平), 按照惯常的计量分析方法, 我们立即可以针对下式进行分析。

$$y_i = \beta_i + \mu_i \quad (1.1)$$

但是, 问题在于: 无论我们如何抽取样本, 总会有相当数量的家庭在前述耐用消费品上的支出为 0, 并且原因并不是收入过低, 也即总会有一定数量的家庭对该项耐用消费品不感兴趣, 这显然是家庭个人选择(Self-Selection)的结果。由此也决定了我们的数据样本是带有选择性偏倚的。

为了揭示这种选择性偏倚对我们分析结论的影响, 我们需要将导致这种选择性偏倚的原因纳入分析框架。各家庭在进行支出决策时, 是否购买一定数量的耐用消费品, 其决策的依据一般会源于家庭效用的最大化, 例如: 最大化其效用函数  $u_i(y_i, z_i)$ ,  $z_i$  代表除耐用消费品外的其他开支。假定社会上最优惠的耐用消费品价格为  $y_0$ , 那么, 一旦某个家庭决定购买前述耐用消费品, 他们在耐用消费品上的支出  $y_i$  就会大于或等于  $y_0$ , 否则,  $y_i = 0$ 。

家庭  $i$  的效用最大化过程意味着: 谋求  $u_i(y_i, z_i)$  的最大化, 同时使其支出决策满足家庭预算约束  $y_i + z_i \leq x_i$  和边界条件  $y_i \geq y_0$  (或  $y_i = 0$ )。假定  $y_i^*$  是这样一个最优化问题的解, 满足预算约束  $y_i + z_i \leq x_i$ , 但不一定满足  $y_i \geq y_0$  的要求。类似地, 我们也可以分析  $y_i^*$  与家庭收入  $x_i$  之间的关系,  $y_i^* = \beta_i + u_i$ 。由于  $y_i^*$  与我们现实观察到的支出  $y_i$  并不完全一致, 因此,  $y_i^*$  只是一个隐含变量或潜在变量(Latent Variable)。在引入  $y_i^*$  后我们可以将整个问题概括为:

$$y_i^* = \beta_i + u_i, \quad i=1,2, \dots, n \quad (1.2)$$

$$y_i = y_0, \quad \text{如果 } y_i^* \geq y_0 \quad (1.3)$$

$$= 0, \quad \text{如果 } y_i^* < y_0$$

为简化问题的表述, 我们可以设定  $y_0 = 0$ , 当  $y_0 > 0$  时, 实际上相当于从每一个  $y_i$  中减去  $y_0$ 。

扰动项  $u_i$  满足通常假定:  $i \text{ i d } N(0, \sigma^2)$ , 则基于  $n$  个观察值的似然函数为

$$\lambda = \prod_{i=1}^n [1 - \Phi(\beta_i/\sigma)] \cdot \sigma^{-n} \prod_{i=1}^n \phi[(y_i - \beta_i)/\sigma] \quad (1.4)$$

赫克曼的两阶段方法可以概括如下:

首先, 从(1.2)、(1.3)式可以看出

$$E(y_i | y_i > 0) = \beta_i + E[u_i | u_i > -\beta_i] \quad (1.5)$$

将(1.5)式与(1.1)式相比较我们可以发现, 如果忽视了数据样本的选择性偏倚, 按照(1.1)式进行的回归分析是存在系统偏误的, 因为  $E(y_i | y_i > 0) \neq E[y_i]$ 。将(1.5)式等号右边的第二项直接积分可得

$$E(y_i | y_i > 0) = \beta_i + \alpha \lambda(\beta_i/\sigma) \quad (1.6)$$

其中  $\lambda(\cdot)$  是穆尔比率(mill's ratio)的倒数,  $\lambda(\cdot) = \frac{\phi(\cdot)}{\Phi(\cdot)}$  在经济分析中通常被称为损害比例。这也是为什么赫克曼校正法又称为“赫克曼  $\lambda$  法”的原因, 赫克曼两阶段方法的第一步就是根据等式(1.4), 运用概率单位模型(Probit)获得  $\beta/\sigma$  的估计值。

将(1.4)式改写为

$$L = \prod_{i=1}^n [1 - \Phi(\beta_i/\sigma)] \cdot \prod_{i=1}^n [\Phi(\beta_i/\sigma)] \cdot \Phi(\beta_i/\sigma)^{-1} \sigma^{-n} \prod_{i=1}^n \phi[(y_i - \beta_i)/\sigma] \quad (1.7)$$

(1.7)式等号右边的前两项构成一个标准的概率单位模型, 后两项构成一个标准的截断数据模型(truncated model), 定义  $\alpha = \frac{\beta}{\sigma}$ , 则  $\alpha$  的估计值可由(1.7)式的前两项通过极大似然估计法获得。

第二步, 在获得  $\alpha = \frac{\beta}{\sigma}$  的估计后, 针对每一个  $x_i$ , 可以计算出  $\lambda$ , 然后以(1.6)式为基础采用普通最小二乘法就可获得模型其他参数的估计值。

从前述分析中不难看出, 赫克曼提出的两阶段分析法显然具有普遍的适用性, 它广泛地适用于存在选择性偏倚的数据分析中。从计量经济学的角度看, 赫克曼的分析方法真正地在经济学和统计学之间架起了一座桥梁, 使经济学家和社会学家能够正确地针对微观经济数据进行分析, 得出真正切合实际的分析结论。

与赫克曼的科学贡献类似, 麦克法登的贡献同样也是经济理论与计量方法紧密结合的产物。麦克法登主要发展了基于离散选择的计量分析方法, 他的研究成果不仅极大地促进了针对个人选择行为的经济计量方法, 而且从根本上改变了那种认为经济计量方法只适用于定量分析而不适用于定性分析的传统观点, 因此, 这种基于离散选择的经济计量模型通常也称为定性反应计量模型, 麦克法登的研究成果大大拓展了经济计量方法的应用领域。

离散选择模型是相对于连续型计量模型而言的。传统上, 自新古典经济学发展以来, 为了细致刻画边际生产力、边际效用等概念, 微积分方法成为新古典经济学最重要的表现形式, 基于此, 无论是经济理论模型还是计量分析模型, 为了分析和演算的方便, 几乎都是连续型变量的形式出现的, 理论分析中也不断充斥着诸如“连续”、“无限细分”、“无限可加性”(σ—可加性)等等词汇。但是, 现实中经济主体的决策

行为却又是以离散选择为基础的,如个人对职业、居住地点、交通方式等的选择,备择集显然只是有限的,即使是个人的消费和支出也是离散变化的,并且通常是有限的。

离散选择模型早在经济学家广泛地使用之前,生物统计学者和心理分析专家就已开始运用这类模型分析定性反应问题,如测定一种杀虫剂的除虫效果,确定某种治疗方案对病人的治愈效果等。与经济学家所使用的模型和分析框架相比,前者要简单得多,对统计和计量分析方法并未构成真正的挑战。待到经济学家需要广泛地使用基于离散选择的定性反应模型时,由于经济体系的复杂性和经济主体、决策的独特性,往往使得问题的结构异常复杂,对统计理论和计量方法提出了更新、更高的要求,麦克法登的研究顺应了这种要求,因此他的理论研究不仅大大推动了定性反应模型在经济分析中的应用,同时也直接发展了与此有关的统计和计量分析的理论方法。在阐述麦克法登的理论贡献以前,让我们先来考察两个简单的定性反应模型,之所以简单是因为模型中只包含单一的解释变量和被解释变量,因此被称为单变量的定性反应模型,它构成我们分析麦克法登研究成果的基础和出发点。

第一个例子是以生物杀虫剂为背景的。假定我们给一批昆虫喷洒一定剂量的杀虫剂,  $x$ , 受试对象中的某一个如第  $i$  个昆虫是否如我们所期望地被杀死,取决于该昆虫对杀虫剂的耐受程度,  $y_i^*$ 。如果该昆虫耐受杀虫剂的剂量  $y_i^*$  小于喷洒剂量,  $y_i^* < x$ , 那么这只昆虫会被杀死,否则结果相反。通常,在昆虫的耐药度是由许多相互独立、但影响效果能够叠加的因素来决定时,  $y_i^*$  服从一定的正态分布(这是由中心极限定理保证的)。假定  $y_i^* \sim N(\mu, \sigma^2)$ , 我们定义  $y_i=1$  代表第  $i$  个昆虫最终被杀死,否则  $y_i=0$ , 那么昆虫被杀死的概率为

$$P(y_i=1) = P(y_i^* < x) = \Phi[(x - \mu)/\sigma] \quad (2.1)$$

$\Phi$  为标准正态分布的累积分布函数。如果  $y_i^*$  服从于逻辑分布函数,  $y_i^* \sim (\mu, \sigma)$ , 则

$$P(y_i=1) = P(y_i^* < x) = \Lambda\left[\frac{\pi}{4} \frac{x - \mu}{\sigma}\right] \quad (2.2)$$

(2.1) 式构成一个标准的普罗比特模型(Probit Model)(也称概率单位模型), (2.2) 式构成一个标准的逻辑特模型(Logit Model)(也称逻辑单位模型), 都可以用标准的统计方法得到待估参数。

第二个例子是一个经济分析问题,涉及个人对交通方式的选择。在考虑以何种方式上班时,他(她)的选择包括:开车上班(Car)、乘公共汽车(Bus)上班,决策的依据是要看哪种交通方式带来的效用更大。在有关交通方式选择中的个人效用取决于两部分内容,一项代表交通方式的影响,而另一项则代表除交通方式外的其他因素的影响。定义  $u_{i1}$  和  $u_{i0}$  分别为他(她)开车上班或乘车上班时的效用,并且效用函数具有如下线性形式:

$$u_{i0} = \alpha_0 + Z_{i0} \beta + W_i \gamma_0 + \epsilon_0 \quad (2.3)$$

$$u_{i1} = \alpha_1 + Z_{i1} \beta + W_i \gamma_1 + \epsilon_1 \quad (2.4)$$

$Z_{i0}$  与  $Z_{i1}$  分别代表乘车和开车的影响,  $W_i$  为除交通方式外的影响,个人效用最大化的结果表现为,如果  $u_{i0} > u_{i1}$ , 则乘车上班。由于  $\epsilon_0$  与  $\epsilon_1$  是不相关的随机误差项,所以  $u_{i0} = u_{i1}$  的概率为 0, 不存在无法抉择的情况。类似地,我们定义  $y_i=1$  代表开车上班,

$$\begin{aligned} P(y_i=1) &= P(u_{i1} > u_{i0}) \\ &= P[\epsilon_0 - \epsilon_1 < \alpha_1 - \alpha_0 + (Z_{i1} - Z_{i0}) \beta + W_i (\gamma_1 - \gamma_0)] \\ &= F[(\alpha_1 - \alpha_0) + (Z_{i1} - Z_{i0}) \beta + W_i (\gamma_1 - \gamma_0)] \end{aligned} \quad (2.5)$$

因此,只要我们指定扰动项  $\epsilon_0$  与  $\epsilon_1$  的分布函数  $F[\cdot]$ , 我们就可以获得待估参数的估计值。同样地,通常都是假定  $\epsilon_0 - \epsilon_1$  服从于正态分布或逻辑分布,这样,从统计分析和计量方法上看,第二个例子与第一个例子并无本质区别。

麦克法登的贡献是基于这种简明的分析框架细致分析了微观个体在面临离散选择时的决策行为。他首先将上述最基本的定性反应模型推广至基于多种选择(允许在  $n$  个备选方案中选择,而不是仅限于两种选择方案:昆虫被杀死或存活,开车上班或乘车上班)的定性反应模型;其次,他还细致分析了当备选方案之间存在关联时,微观个体、决策行为对计量分析的影响,从中推导出了既具有丰富的经济选择背景,同时又对既有统计和计量方法进行了推广的研究结果;最后,麦克法登突破特殊的分析背景,给出了定性反应模型的广义分析框架。

如前所述,经济学所分析的问题总是相对复杂,如果定性反应模型只能用于对两种备择方案的选择问题进行分析,那么它对经济分析的意义和价值就极为有限。麦克法登首先将简单的定性反应模型推广至可以基于多种选择的情形,为了便于表述,我们仍以前述有关交通方式的选取为分析背景,但此时的备选方案有三种:开车、乘公共汽车和乘地铁,定义与交通方式选择有关的效用函数为:

$$u_{ij} = \mu_{ij} + \epsilon_{ij}, \quad j=0,1,2 \quad (2.6)$$

$u_{ij}$  相当于前述模型中的  $\alpha + Z_{ij} \beta + W_i \gamma_j$ , 只是表现形式稍显简练。为了确定微观主体选择各种备选方案的概率,麦克法登仍需指定随机扰动项所遵从的分布律,他在分析中假定  $\{\epsilon_{ij}\}$  是一组独立同分布随机变量,具有所谓“I型极值分布函数”  $\exp[-\exp(-\epsilon_{ij})]$ 。根据效用最大化原则,微观个体选择乘地铁( $y_i=2$ )的概率为:

$$P(y_i=2) = P(u_{i2} > u_{i1}, u_{i2} > u_{i0}) \quad (2.7)$$

$$\begin{aligned} &= P(\epsilon_1 + \mu_2 - \mu_1 > \epsilon_1, \epsilon_1 + \mu_2 - \mu_0 > \epsilon_1) \\ &= \int_{-\infty}^{\infty} f(\epsilon_1) \int_{-\infty}^{\infty} f(\epsilon_2 + \mu_2 - \mu_1) f(\epsilon_2 + \mu_2 - \mu_0) f(\epsilon_1) d\epsilon_1 d\epsilon_2 \\ &= \int_{-\infty}^{\infty} \exp(-\epsilon_1) \exp[-\exp(-\epsilon_1)] \times \exp[-\exp(-\epsilon_1 - \mu_2 + \mu_1)] \times \exp[-\exp(-\epsilon_1 - \mu_2 + \mu_0)] d\epsilon_1 \\ &= \frac{\exp(\mu_2)}{\exp(\mu_0) + \exp(\mu_1) + \exp(\mu_2)} \end{aligned}$$

在(2.7)式的推导中第三个等号后的表达式源于  $\epsilon_j$  是一组独立同分布的随机变量,第四个等号是对上式直接积分的结果。类似的,可以推导出  $P(y_i=1)$ ,  $P(y_i=0)$ 。当备择方案大于 3 时,  $i$  取值于  $n$  个不同的数值,麦克法登的上述结果可以自然的得到推广

$$P(y_i = j) = \frac{\exp(\mu_{ij})}{\sum_{i=0}^{n-1} \exp(\mu_{ij})} \quad (2.8)$$

这就是由麦克法登首先提出的基于多种选择的定性反应模型,现在通常称为“多项式逻辑特模型”(Multinomial Logit Model)。显然,这种分析框架相比于简单的单变量定性反应模型具有更好的适用性。但是,麦克法登并未满足于这种分析框架,因为上述分析框架的一个明显不足就是要求各

备选方案相互之间不存在任何关联关系,理论界称之为“无关联选项间的独立性”(IIA),具体的表现是要求 $\{\epsilon_j\}$ 是一组独立同分布的随机变量,这种独立性要求不仅在统计分析上表现得过于严格,而且现实中往往也与实际情况不符。我们援引麦克法登关于“蓝色公共汽车”和“红色公共汽车”的经典示例,麦克法登指出:如果备选的三个方案改为乘红色公共汽车、乘蓝色公共汽车和乘地铁,那么前述独立性假设显然就不符合逻辑,因为对红色公共汽车的选择必然与对蓝色公共汽车的选择有关,更为突出的是,此时如果继续援用上述分析框架还会使得对乘地铁的选择概率产生低估,因为按照独立性假定计算出的 $P(u_2 > u_1, u_2 > u_0)$ 忽视了这种事实:一旦 $u_2 > u_1$ ,那么 $u_2$ 就极可能大于 $u_0$ 。

如果要使定性反应模型真正能够运用到经济分析中,那么分析框架必须能够涵盖这种关联性,麦克法登由此又发展出了所谓“套嵌式逻辑特模型”(Nested Logit Model)。

我们仍以前述交通方式的选择为背景,为了刻画备选方案间的关联性,麦克法登假定 $\{\epsilon_j\}$ 具有I型极值分布函数,同时 $\epsilon_0$ 与 $\epsilon_1$ 具有如下联合分布函数

$$F(\epsilon_0, \epsilon_1) = \exp\{-[\exp(-\rho^1 \epsilon_0) + \exp(-\rho^1 \epsilon_1)]^\rho\}, 0 < \rho \leq 1 \quad (2.9)$$

这种分布函数又称为冈贝尔B型极值分布函数,当 $\rho=1$ 时,结构退化为符合前述无关联选项间的独立原则,因此,前者成为一种特例。根据效用最大化原则,微观个体选择乘地铁的概率为

$$\begin{aligned} P(y_i=2) &= P(u_{12} > u_{11}, u_{12} > u_{10}) \quad (2.10) \\ &= P(\epsilon_0 + \mu_2 - \mu_1 > \epsilon_1, \epsilon_0 + \mu_2 - \mu_0 > \epsilon_1) \\ &= \int \int \left[ \exp(-\epsilon) \times \exp[-\exp(-\rho^1(\epsilon_0 + \mu_2 - \mu_0))] + \exp[-\rho^1(\epsilon_0 + \mu_2 - \mu_1)] \right]^\rho d\epsilon \\ &= \int \exp(-\epsilon) \exp[-\alpha \exp(-\epsilon)] d\epsilon \\ &= \alpha^{-1} \end{aligned}$$

其中 $\alpha = 1 + \exp(-\mu_2) [\exp(\rho^1 \mu_0) + \exp(\rho^1 \mu_1)]^\rho$

显然在考虑两个选项存在关联性时,选中第3种方案(乘地铁)的概率不同于此前的分析结果,同样也可以计算乘坐红色公共汽车和蓝色公共汽车的概率

$$P(y_i=1 | y_i=2) = P(u_1 > u_0 | u_1 > u_2 \text{ 或 } u_0 > u_2) = P(u_1 > u_0) \quad (2.11)$$

由于乘坐地铁的选项与乘公共汽车的选项不相关,因此, $u_1 > u_2$ 或 $u_0 > u_2$ 涵盖了整个事件,所以条件概率等于无条件概率,第2个等号成立。从上述分析结果我们可以看出套嵌逻辑特模型的一些有趣的分析现象:对存在关联性的选项之一进行抉择时,该选项(乘红色公共汽车和蓝色公共汽车)被选中的概率与符合无关联选项独立性原则下的选中概率一致。而对第三种选项的选择概率,前两种选项间的关联性的影响表现为一种加权平均和的形式。基于此,麦克法登给出了定性反应模型的广义极值分布函数,从而使得基于离散选择的定性反应模型最终具有了真正广泛的实用性。麦克法登证明,只要备选方案的随机扰动项符合以下条件,各选项被选中的概率仍具有前述简单套嵌逻辑特模型的特征。

麦克法登所定义的广义极值分布函数具有如下的一般

形式

$$F(\epsilon_0, \epsilon_1, \dots, \epsilon_m) = \exp\{-G[\exp(-\epsilon_0), \exp(-\epsilon_1), \dots, \exp(-\epsilon_m)]\}$$

函数 $G()$ 满足如下条件

- (1)  $G(u_1, u_2, \dots, u_m) \geq 0, u_1, u_2, \dots, u_m \geq 0$
- (2)  $G(au_1, au_2, \dots, au_m) = aG(u_1, u_2, \dots, u_m)$
- (3)  $\frac{\partial G}{\partial a_{1k}} \geq 0$ , 当 $k$ 为基数时  
 $\leq 0$ , 当 $k$ 为偶数时

此时各选项的概率为

$$P_j = \frac{\exp(\mu_j) G_j[\exp(\mu_1), \exp(\mu_2), \dots, \exp(\mu_m)]}{G[\exp(\mu_1), \exp(\mu_2), \dots, \exp(\mu_m)]}, G_j \text{ 是函数 } G() \text{ 对第 } j \text{ 个变量的偏导数。}$$

注释:

赫克特(Heckit)与托比特(Tobit)的名称据称是来源于统计分析中概率单位分析(Probit)与逻辑单位分析(Logit)。

值得说明的是,按赫克曼第一步获得的估计值,会损失一些估计效率,因为估计中只利用了似然函数的一部分。

参考文献:

1. Heckman J. J. (1976), The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models, Annals of Economic and Social Measurement 5,475 ~ 492.
2. Heckman J. J. (1990), Varieties of Selection Bias, American Economic Review 80,313 ~ 318.
3. Heckman J. J. (1991), Identifying the Hand of the Past: Distinguishing State Dependence from Heterogeneity, American Economic Review 81,75 ~ 79.
4. Heckman J. J., R. LaLonde and J. Smith (1999), The Economics and Econometrics of Active Labor Market Programs, Handbook of Labor Economics, vol 3A, North - Holland.
5. Heckman J. J. and R. Robb (1985a), Alternative Methods for Evaluating the Impact of Interventions: An Overview, Journal of Econometrics 30,239 ~ 267.
6. Heckman J. J. and J. Smith (1995), Assessing the Case for Social Experiments, Journal of Economic Perspectives 9,85 ~ 110.
7. McFadden D. (1974a), Conditional Logit Analysis of Qualitative Choice Behavior, in P. Zarembka (ed.), Frontiers of Econometrics, Academic Press.
8. McFadden D. (1974b), The Measurement of Urban Travel Demand, Journal of Public Economics 3,303 ~ 328.
9. McFadden D. (1978), Modelling the Choice of Residential Location, in A. Kar - Iqvist, L. Lundqvist, F. Snickars and J. Weibull (eds.), Spatial Interaction Theory and Planning Models, North - Holland.
10. McFadden D. (1981), Econometric Models for Probabilistic Choice, in C. Manski and D. McFadden (eds.), Structural Analysis of Discrete Data with Econometric Applications, Harvard University Press.
11. McFadden D. (1987), Regression Based Specification Tests for the Multinomial Logit Model, Journal of Econometrics 34,63 ~ 82.
12. McFadden D. (1989), A Method for Simulated Moments for Estimation of Discrete Response Models without Numerical Integration, Econometrica 57,995 ~ 1026.

(作者单位: 武汉大学商学院 武汉 430072)

(责任编辑: 曾国安)