Chinese Association of Quantitative Economics

中国数量经济学会

中 国 数 量 经 济 学 会

21世纪
数量经济学

Quantitative Economics in the 21st Century

第19卷

◎主　编 李　平 石　磊
◎副主编 李富强　向其凤

经济管理出版社
ECONOMY & MANAGEMENT PUBLISHING HOUSE

中国数量经济学会

21世纪

数量经济学

Quantitative Economics in the 21st Century

第19卷

◎主　编　李　平　石　磊

◎副主编　李富强　向其凤

经济管理出版社
ECONOMY & MANAGEMENT PUBLISHING HOUSE

图书在版编目（CIP）数据

21 世纪数量经济学．第 19 卷/李平，石磊主编．—北京：经济管理出版社，2019.6
ISBN 978 - 7 - 5096 - 6468 - 1

Ⅰ．①2…　Ⅱ．①李…②石…　Ⅲ．①数量经济学—文集　Ⅳ．①F224.0 - 53

中国版本图书馆 CIP 数据核字（2019）第 054369 号

组稿编辑：陈　力
责任编辑：杨国强　张瑞军
责任印制：黄章平
责任校对：王淑卿

出版发行：经济管理出版社
（北京市海淀区北蜂窝 8 号中雅大厦 A 座 11 层　100038）
网　　址：www. E - mp. com. cn
电　　话：（010）51915602
印　　刷：三河市延风印装有限公司
经　　销：新华书店
开　　本：710mm × 1000mm/16
印　　张：30. 75
字　　数：504 千字
版　　次：2019 年 6 月第 1 版　　2019 年 6 月第 1 次印刷
书　　号：ISBN 978 - 7 - 5096 - 6468 - 1
定　　价：128. 00 元

前言

本书是《21世纪数量经济学》丛书第19卷。

中国数量经济学会2018年年会于2018年10月26~28日在云南昆明召开，年会由中国数量经济学会与云南财经大学主办，云南财经大学统计与数学学院承办。来自政府部门、研究机构、大专院校和企业的536位数量经济学专家、学者参加了本次年会，会议共收到学术论文305篇。

在本次年会上，美国普林斯顿大学教授范剑青教授、中国人民大学艾春荣教授、云南财经大学石磊教授分别做了高水平的学术报告。天津财经大学白仲林教授、厦门大学郑挺国教授、云南财经大学陈昆亭教授在名家论坛上做了学术报告。会议分为：数量经济理论与方法，经济增长、宏观经济，金融、保险，资本市场，财政、税收，投资、贸易，区域经济、产业经济，环境、资源，大数据理论及方法，实验经济学及其他学科10个小组进行了专题讨论，100多名学者在小组学术交流会上介绍了自己的最新研究成果，会议收到了良好的效果。

本书是在本次年会提交的论文中遴选出来的优秀论文集结成册的，共22篇，分为5个部分：数量经济理论与方法，宏观经济、财政税收，金融、资本市场，企业、产业经济，绿色经济、实验经济学。入选的这些论文均有较高的学术水平，具有一定的理论意义和实践意义。

囿于编者的水平和能力所限，本书一定存在不少错误和疏漏之处，欢迎广大读者批评指正。

目　录

1. 数量经济理论与方法

2. 宏观经济、财政税收

3. 金融、资本市场

4. 企业、产业经济

5. 绿色经济、实验经济学

1. 数量经济理论与方法

Heckman模型截距项的半参数估计

潘哲文
（中山大学岭南学院，广州 510275）

摘 要 Heckman模型是处理样本选择问题的标准工具，其截距项的估计可以应用于平均处理效应估计、工资差异分解等问题。无穷处识别方法是Heckman模型截距项的一种重要的半参数估计方法，但存在窗宽参数难以选取的问题。本文通过把无穷处识别转化为边界处识别，提出Heckman模型截距项的一种核估计方法，并利用经验法则解决窗宽选取问题。模拟结果显示，利用经验法则选取的窗宽在不同的模型设定下均能达到或接近最优水平。

关键词 样本选择；截距项；半参数估计；无穷处识别；窗宽选取；经验法则

Semiparametric Estimation for the Intercept of the Heckman Model

Pan Zhewen
(Lingnan College, Sun Yat - sen University 510275)

Abstract: The Heckman model is a standard tool in sample selection correction, and its intercept is indispensable in applications such as the estimation of average treatment effect and the decomposition of wage difference. The identification at infinity is an important semiparametric estimation method for the intercept of the Heckman model. However, it is difficult to choose appropriate bandwidths for the identification - at - infinity estimation. This study rephrases the identification - at - infinity problem into an identification - at - boundary one, and accordingly proposes a kernel regression method to semiparametrically estimate the intercept of the Heckman model. A virtue of the kernel method is that the bandwidths can be chosen by the rule of thumb. A Monte Carlo simulation shows that the rule - of - thumb bandwidths are optimal or nearly optimal in various designs.

Key Words: Sample Selection; The Intercept; Semiparametric Estimation; Identification at Infinity; Bandwidth Selection; The Rule of Thumb

一、引言

样本选择问题源自样本的非随机性。除了调查样本选取不当的原因外，样本的非随机性主要来源于样本个体的自选择行为。例如，在经典的工资方程研究中，只有在职劳动者的工资数据可以被观测，但是否参加工作是劳动者的自选择行为，因此基于在职劳动者的随机抽样必然是一个非随机样本。又如，在医疗保险对老年人医疗支出影响的研究中（黄枫和甘犁，2010），发现存在大量的“零医疗”支出数据。如果这部分老年人是出于医疗价格或便利程度的考虑而主动放弃就医，那么医疗支出是否为零就是个体自选择行为的结果，而“非零医疗”支出的样本则是一个非随机样本。实际上，样本选择问题普遍存在于微观调查数据以及经济学实证研究，正如 Lee（2009）所述：“In the real world, the sample selection problem seems to be the rule rather than the exception.”（在现实世界中，样本选择问题似乎是普遍规律而非个别现象。）

样本选择问题的存在有可能导致对经济变量间真实关系的错误推断，从而误导社会经济实践。目前，用于解决或减轻样本选择问题的主要工具是 Heckman 样本选择模型（以下简称 Heckman 模型），该模型通过对自选择行为进行单独建模，可以有效修正非随机样本的不利影响。Heckman 模型的传统估计方法是最大似然估计和 Heckman 两步法，但这两种方法要求模

型扰动项服从（联合）正态分布，否则相应的最大似然估计量和 Heckman 两步估计量均不具有相合性①。鉴于此，大量研究通过放松扰动项的参数性分布假定，提出 Heckman 模型的半参数估计量，从而避免了分布误设所导致的非相合性。代表性的研究包括 Gallant 和 Nychka（1987）、Ahn 和 Powell（1993）、Chen 和 Lee（1998）、Newey（2009）、Chen 和 Zhou（2010）、王亚峰（2012）等。但是，这些研究主要关注 Heckman 模型的斜率系数而在一定程度上忽视了截距项的半参数估计问题。

Heckman 模型截距项在平均处理效应估计、工资差异分解等应用中具有关键作用。与 Heckman 模型斜率系数的大量研究相比，其截距项的半参数估计方法相对缺乏，目前最主要的估计方法是 Heckman（1990），Andrews 和 Schafgans（1998）所提出的无穷处识别（Identification - at - infinity）方法。但是，无穷处识别方法的理论窗宽依赖于不可观测的扰动项，因此存在窗宽选取困难的问题，成为限制其实际应用的主要障碍。Schafgans（2004）通过数值模拟考察了不同窗宽选择下无穷处识别估计量的有限样本表现，发现随着窗宽偏离最优水平，估计量的均方误差②会增大；而当偏离程度大于某一阈值时，其均方误差甚至会超过非相合的 Heckman 两步估计量。Schafgans 的研究反映了合适的窗宽选取在实际应用中的重要性，但截至目前，对无穷处识别估计量窗宽选取方法的研究仍未有明显突破。

本研究的主要目标在于解决无穷处识别方法的窗宽选取难题，为实际应用提供理论指导。通过把无穷处识别转化为边界处识别，本研究提出 Heckman 模型截距项的一种具有核回归形式的半参数估计量，这个新的估计量包含无穷处识别估计量作为特例。更重要的是，核回归文献中有着较为成熟的窗宽选取方法可供借鉴。例如，经验法则（The Rule of Thumb）以及交叉验证（Cross Validation）方法。通过数值模拟，本研究发现由经验法则选取的窗宽在不同的模型设定下均接近最优水平，表明简单的经验法则已经能基本解决无穷处识别方法的窗宽选取难题，满足实证研究者的需要。

---

① 相合性（consistency）指的是估计量依概率收敛于参数真实值。一般来说，非相合估计量的特点是渐进偏误收敛于一个非零常数，而渐进方差（以$\sqrt{n}$的速度）收敛于零，因此当样本量很大时，参数真实值几乎总是会落在95%的置信区间外，从而产生误导性的统计推断结论。可见，相合性是对估计量的最低要求。

② 估计量的均方误差（mean square error）等于估计量偏误的平方加上估计量的方差，是衡量估计量优劣的主要标准。均方误差越小，估计量与参数真实值越接近。

本研究的关键在于发现无穷处识别问题与非参数回归问题之间的联系，从而成功地将无穷处识别方法置于经典的核回归分析框架下进行研究。本研究至少在两个方面补充和发展了无穷处识别文献：一是扩展了无穷处识别估计量，提出了一种更一般化的估计量，而且这种估计量具有核回归估计量的形式；二是利用核回归文献中的经验法则方法解决了无穷处识别估计量的窗宽选取难题。

本文余下部分的内容安排如下：第二节简要介绍 Heckman 模型，并通过两个例子说明 Heckman 模型截距项的重要性。第三节以核回归方法为基础，给出 Heckman 模型截距项的一种新的半参数估计量。第四节设计数值模拟实验，考察估计量的有限样本性质，并与 Heckman 两步估计量和无穷处识别估计量进行比较。第五节是结论。

二、Heckman 模型

Heckman 模型由选择方程和结果方程构成，形式如下：

$$
\begin{aligned}
&D = 1\{X'\beta_0 > \varepsilon\} \\
&Y^* = \mu_0 + Z'\theta_0 + U \qquad (1) \\
&Y = DY^*
\end{aligned}
$$

式中，可观测的变量为（D，Y，X，Z），1｛A｝是示性函数，当 A 事件成立时取值为 1，否则为 0。D 是选择变量，取值为 1 和 0，用于标记是否发生样本选择。Y^* 是我们真正关心的变量，但我们只能在 $D=1$ 时观测到它，即 Y^* 只能被部分地观测。为方便表述，定义一个可观测的变量 Y，当 Y^* 可以被观测时（$D=1$ 时）Y 就等于 Y^*，而当 Y^* 不能被观测时 Y 就取值为 0。在工资方程的例子中，$D=1$ 表示参加工作，Y^* 代表工作所能获得的市场工资，但我们只能观察到在职劳动者的工资，即 Y。X 和 Z 分别是选择方程和结果方程的协变量组成的列向量，在 $D=0$ 的时候也可以被观测；ε 和 U 是随机扰动项，其中 U 的均值假设为零。本文不对 ε 和 U 施加任何参数性分布假定，即本文考虑 Heckman 模型的半参数估计。注意到选择方程参数 β_0 可以用二元选择模型的半参数估计方法进行估计，如 Powell 等（1989）、Klein 和 Spady（1993）、Chen 和 Zhang（2015）等，因此 Heckman 模型文献主要关注结果方程参数 θ_0 和 μ_0 的估计问题。其中，斜率系数 θ_0

的半参数估计研究相当丰富，而截距项 μ_0 的半参数估计方法则相对缺乏。但是，Heckman 模型的多个经济学应用依赖于对 μ_0 的估计。下面以平均处理效应估计和工资差异分解为例说明截距项 μ_0 在 Heckman 模型应用中的重要性，然后对 μ_0 现有的半参数估计方法进行简要回顾。

（一）截距项 μ_0 与平均处理效应估计

考虑以下处理效应模型：

$$\begin{aligned} &D = 1\{X'\beta_0 > \varepsilon\} \\ &Y_t^* = \mu_t + Z'\theta_t + U_t \\ &Y_u^* = \mu_u + Z'\theta_u + U_u \\ &Y = DY_t^* + (1-D)Y_u^* \end{aligned} \tag{2}$$

选择变量 D 用于标记是否接受处理，D = 1 表示进入处理组，D = 0 表示进入控制组，Y_t^* 和 Y_u^* 分别表示进入处理组和进入控制组的潜在结果。因为接受处理和不接受处理互为反事实，所以我们只能观察到其中一个结果，记为 Y。Y_t^* 和 Y_u^* 的差值称为处理效应，其均值 $E[Y_t^* - Y_u^*]$ 则称为平均处理效应（ATE），是模型（2）主要的研究对象①。例如，如果 D 表示是否接受高等教育，结果变量 Y 表示工资收入，那么 ATE 就是高等教育的工资回报率。把中间两式代入 $E[Y_t^* - Y_u^*]$，可得 $ATE = (\mu_t - \mu_u) + (\theta_t - \theta_u)'E[Z]$。可见，要估计 ATE，需要先得到结果方程参数的估计。实际上，模型（2）可以拆分成两个 Heckman 模型分别进行估计。若忽略 Y_u^* 的信息，即令 Y_u^* 恒等于 0，则可得到处理组的 Heckman 模型：

$$\begin{aligned} &D = 1\{X'\beta_0 > \varepsilon\} \\ &Y_t^* = \mu_t + Z'\theta_t + U_t \\ &Y = DY_t^* \end{aligned}$$

类似地，若忽略 Y_t^* 的信息，并定义 $\tilde{D} = 1 - D$，则可得到控制组的 Heckman 模型：

$$\tilde{D} = 1\{X'\beta_0 \leqslant \varepsilon\}$$

---

① 模型（2）关心的其他研究对象还包括处理组的平均处理效应 $E[Y_t^* - Y_u^* \mid D = 1]$、控制组的平均处理效应 $E[Y_t^* - Y_u^* \mid D = 0]$、中值处理效应 $Median(Y_t^* - Y_u^*)$ 等，参见 Imbens 和 Wooldridge（2009）。

$Y_u^* = \mu_u + Z'\theta_u + U_u$

$Y = \tilde{D}Y_u^*$

ATE 中的参数 μ_t 和 μ_u 分别对应于这两个 Heckman 模型的截距项。因此，Heckman 模型截距项的半参数估计是获得 ATE 半参数估计的关键步骤。

（二）截距项 μ_0 与工资差异分解

Oaxaca 分解（Oaxaca，1973；Oaxaca and Ransom，1994）是分析工资差异的标准工具，通常用于研究性别歧视（Schafgans，2000）、种族歧视（Schafgans，1998）等。下面以性别歧视为例说明 Heckman 模型在 Oaxaca 分解中的应用。考虑按性别分组的工资方程：

$D_k = 1\{X'_k\beta_k > \varepsilon_k\}$

$Y_k^* = \mu_k + Z'_k\theta_k + U_k$

$Y_k = D_kY_k^*$，$k = F$，M

下标 $k = F$ 表示女性组，$k = M$ 表示男性组，D_k 标记是否参加工作，Y_k^* 是潜在的市场工资，Y_k 是观测工资。记观测样本的平均工资为 $\bar{Y}_k = E[Y_k \mid D_k = 1]$，则可观测的工资差异定义为 $\bar{Y}_M - \bar{Y}_F$。记 $\bar{Z}_k = E[Z_k \mid D_k = 1]$ 和 $\bar{U}_k = E[U_k \mid D_k = 1]$，那么可观测的工资差异可以分解为三项之和：

$$\bar{Y}_M - \bar{Y}_F = \underbrace{(\mu_M - \mu_F) + \bar{Z}'_M(\theta_M - \theta_F)}_{\text{歧视效应}} + \underbrace{(\bar{Z}_M - \bar{Z}_F)'\theta_F}_{\text{禀赋效应}} + \underbrace{\bar{U}_M - \bar{U}_F}_{\text{样本选择效应}} \tag{3}$$

第一项为歧视效应，是工资方程系数差异导致的性别工资差异；第二项为禀赋效应，是男女间特征差异导致的性别工资差异，例如，平均教育水平的差异、工作经验的差异等；第三项为样本选择效应，是样本选择性偏误导致的性别工资差异。与传统回归分解相比，基于 Heckman 模型的 Oaxaca 分解可以把不可观测的样本选择效应分解出来，修正样本选择问题的不利影响（寇恩惠和刘柏惠，2011）。式（3）中，性别歧视的严重程度定义为歧视效应占可观测的工资差异的比重，而歧视效应的估计依赖于分组 Heckman 模型结果方程参数的估计，其中包括截距项 μ_M 和 μ_F。因此，Heckman 模型截距项的半参数估计是工资差异分解半参数分析的重要组成部分。

（三）截距项 μ_0 现有的半参数估计方法

Heckman 模型文献所提出的大多数半参数估计方法只能得到斜率系数的

估计，只有 Gallant 和 Nychka（1987）的半参数最大似然估计方法、Lewbel（2007）的特殊回归元方法，以及 Chen、Zhou（2010）、Chen 等（2017）的方法可以同时得到截距项的估计。但是，这几种方法均存在一定的缺陷：Gallant 和 Nychka 估计量的渐进分布至今没有得到证明，导致其难以对参数进行假设检验和统计推断；Lewbel 的方法需要假设扰动项的支撑集有界，这个假设限制了它的应用范围；Chen 和 Zhou（2010）、Chen 等（2017b）的方法要求扰动项具有联合对称分布，而大部分数据并不满足这个要求。另外，Heckman（1990）、Andrews 和 Schafgans（1998）专门针对 Heckman 模型的截距项提出了一种称为无穷处识别的半参数估计方法，这种方法不需要对扰动项施加任何形状限制。

Heckman（1990）发现，如果 Heckman 模型（1）中的潜变量 Y^* 可以被观测，那么我们只要对（$Y^*-Z'\theta_0$）取样本平均即可得到截距项 μ_0 的相合估计，这是因为 $E[Y^*-Z'\theta_0]=\mu_0+E[U]=\mu_0$。但是，样本选择问题的存在使我们只能观测到 Y，也就是只有在 $D=1$ 时才能观测到 Y^*，而

$$E[Y-Z'\theta_0 \mid D=1]=\mu_0+E[U \mid D=1]=\mu_0+E[U \mid \varepsilon<X'\beta_0]\neq\mu_0$$

样本选择问题的两个特点：Y^* 的不可观测性以及 U 和 ε 的相关性共同造成了 μ_0 估计的困难。对于这个困难，无穷处识别方法的解决思路是找到 $X'\beta_0$ 的值等于正无穷的个体，因为对于这些个体，$D=1$ 总是成立，即 $D=1$ 对它们而言是一个确定性的事件，所以这类个体不存在自选择行为：

$$E[Y-Z'\theta_0 \mid D=1,\ X'\beta_0=+\infty]=E[Y-Z'\theta_0 \mid X'\beta_0=+\infty]=\mu_0+E[U \mid X'\beta_0=+\infty]=\mu_0+E[U]=\mu_0 \tag{4}$$

式（4）第三个等号的成立需要假设扰动项与协变量独立。

基于这个思想，Heckman（1990）给出了截距项 μ_0 的一个半参数估计量：

$$\hat{\mu}_n^H=\frac{\sum_{i=1}^{n}(Y_i-Z'_i\hat{\theta})D_i\cdot 1\{X'_i\hat{\beta}>\gamma_n\}}{\sum_{i=1}^{n}D_i\cdot 1\{X'_i\hat{\beta}>\gamma_n\}}$$

式中，$\hat{\beta}$ 和 $\hat{\theta}$ 是 β_0 和 θ_0 的相合估计，分别可以用二元选择模型和 Heckman 模型斜率系数的半参数估计方法得到；γ_n 是窗宽参数，用于调节剪截样本的比例，满足 $\gamma_n>0$ 且 $n\rightarrow+\infty$ 时 $\gamma_n\rightarrow+\infty$。不过，$\hat{\mu}_n^H$ 的定义中包含了非连续的示性函数，且示性函数中含有估计量 $\hat{\beta}$，我们无法通过泰勒

展开的方式对 $1\{X'_i\hat{\beta}>\gamma_n\}$ 进行处理，因此 $\hat{\mu}_n^H$ 的渐进性质难以推导。为克服这个困难，Andrews 和 Schafgans（1998）提出用光滑函数 s(·) 代替示性函数 1{·>0}，得到一个光滑化的修正估计量（以下简称 AS 估计量）：

$$\hat{\mu}_n^{AS}=\frac{\sum_{i=1}^{n}(Y_i-Z'_i\hat{\theta})D_i\cdot s(X'_i\hat{\beta}-\gamma_n)}{\sum_{i=1}^{n}D_i\cdot s(X'_i\hat{\beta}-\gamma_n)}$$

式中，s(·) 三阶连续可微，且满足 $t\leq0$ 时，$s(t)=0$，$0<t<b$ 时，$0<s(t)<1$、$t\geq b$ 时，$s(t)=1$，b 是一个正的常数。在若干正则条件下，Andrews 和 Schafgans（1998）证明了 $\hat{\mu}_n^{AS}$ 的相合性和渐进正态性，完善了无穷处识别方法的渐进理论。

AS 估计量 $\hat{\mu}_n^{AS}$ 的一个特点是不能达到 $1/\sqrt{n}$ 的收敛速度，但这并不妨碍 $\hat{\mu}_n^{AS}$ 的假设检验和统计推断。真正限制其应用于实证问题的是窗宽参数 γ_n 的选取难题。Andrews 和 Schafgans（1998）的理论结果表明 $\gamma_n\to+\infty$ 的速度应取决于 $X'\beta_0$ 和 ε 分布的尾部厚度，但目前为止仍缺乏半参数二元选择模型扰动项 ε 的估计方法，自然也就无法估计 ε 分布的尾部厚度。第三节通过把无穷处识别问题转化为非参数回归问题，提出截距项 μ_0 的一种核估计量，成功避开了对扰动项 ε 的估计，直接利用核估计文献中的经验法则解决窗宽选取问题。

三、截距项的核估计量

注意到式（4）中的 $E[Y-Z'\theta_0\mid D=1, X'\beta_0=+\infty]$ 是一个条件期望，但 $+\infty$ 并不是一个严格的点，所以我们无法构造它的核估计。本研究的思路是，如果 F(·) 是一个支撑集为实数集 R 的连续型随机变量的累积分布函数，即 F(·) 是在 R 上严格单调递增的绝对连续函数且满足 $F(-\infty)=0$ 和 $F(+\infty)=1$，那么条件 $X'\beta_0=+\infty$ 就等价于 $F(X'\beta_0)=1$，而条件期望 $E[Y-Z'\theta_0\mid D=1, F(X'\beta_0)=1]$ 可以通过非参数核回归的方法进行估计。根据这个想法，构造 μ_0 的核估计量为：

$$\hat{\mu}_n=\frac{\sum_{i=1}^{n}(Y_i-Z'_i\hat{\theta})D_i\cdot k\left(\frac{1-F(X'_i\hat{\beta})}{h_n}\right)}{\sum_{i=1}^{n}D_i\cdot k\left(\frac{1-F(X'_i\hat{\beta})}{h_n}\right)}\tag{5}$$