■岭南实证与计量经济学研究丛书

Semiparametric and Nonparametric Estimation of Tobit Models

Tobit模型的半参数和非参数估计

周先波◎著

科学出版社

岭南实证与计量经济学研究丛书

Tobit 模型的半参数和非参数估计

Semiparametric and Nonparametric Estimation of Tobit Models

本书得到国家自然科学基金项目“Tobit 模型非参数和半参数估计方法研究及应用”(70971143)的支持

科学出版社
北京

内 容 简 介

减少单变量 Tobit 模型设定偏误，得到一致的和渐近正态的、且具有较好小样本表现的估计量是微观计量经济学方法论研究的重要领域；双变量 Tobit 模型一直缺乏计算上可行的、至少满足一致性的估计量。目前文献中已有的估计量存在一些缺陷。为此，本书首先基于受限因变量条件生存函数，构造单变量 Tobit 模型非参数或半参数估计量，与前人构造的估计量相比，其对误差项具有较弱约束限制，但有较好的小样本表现；其次构造双变量 Tobit 模型一致的和渐近正态的半参数估量，得到该模型的一个可行性估计，克服最大似然估计量和 Amemiya 估计量的缺陷；再次构造 Tobit 模型一致的渐近正态的非参数估计量；最后设计模拟实验，对上述各估计量进行模拟，并与文献中相关估计量进行比较。

本书方法性强，对模型估计的论证比较详细，操作性强，提供估计量模拟的详细程序，适于对计量经济 Tobit 模型估计方法和应用研究感兴趣的科研工作者、学生、实证经济分析者学习与参考。

图书在版编目(CIP)数据

Tobit 模型的半参数和非参数估计 / 周先波著．--北京：科学出版社，2015
（岭南实证与计量经济学研究丛书）
ISBN 978-7-03-043844-7

Ⅰ.①T… Ⅱ.①周… Ⅲ.①计量经济模型—半参数模型—研究②计量经济模型—非参数法—估计—研究 Ⅳ.①F224.0

中国版本图书馆 CIP 数据核字(2015)第 054907 号

责任编辑：李　莉 / 责任校对：贾如想
责任印制：李　利 / 封面设计：无极书装

科 学 出 版 社 出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com

文林印务有限公司　印刷

科学出版社发行　各地新华书店经销

*

2015 年 3 月第　一　版　开本：720×1000　1/16
2015 年 3 月第一次印刷　印张：14 1/4
字数：287 000

定价：68.00 元

（如有印装质量问题，我社负责调换）

总　序

中国改革开放30多年来沧桑巨变，对经济学科最大的影响之一是计量经济学的引入以及与此相对应的经济学研究范式的极大变化。1980年夏天，克莱因等国际知名计量经济学家举办了颐和园讲习班，这是我国学者系统性学习和应用计量经济学的开端。从无到有，从弱到强，经过30余年的发展，颐和园讲习班播下的计量经济学种子长成了学科之林；《计量经济学》已被教育部确定为高等学校经济学门类各专业的共同核心课程，其相关模型方法更是成为经济学经验研究的主流工具，这是我国经济学教学和科研走向现代化和科学化的重要标志。

中山大学岭南(大学)学院自1988年成立至今，一直努力尝试在教学上与国外著名大学实质性接轨，加强现代经济学的训练，作育英才、服务社会。从本世纪初开始，计量经济学进行了本科、硕士研究生以及博士研究生的分层次课程体系建设，从课程设置、教材选择以及教学实践活动等方面全面提升学生对计量经济学学科前沿知识的认识和了解，学生从更高的层面切入计量经济学的应用与研究。同时，成立岭南实证与计量经济研究中心，聘请岭南学院1993届毕业生、斯坦福大学洪瀚教授担任中心主任。洪瀚教授不遗余力，参与岭南学院计量经济学科发展和研究生培养计划的制订，还多次亲临学院上课、开讲座。在洪瀚教授的指导下，岭南实证与计量经济研究中心建设了能够远程操作、24小时不间断运行的以AFS分布式文件系统为核心的计算机集群平台，为计量经济学的教学和研究提供了良好的条件。此外，得益于毗邻香港的方便，岭南学院能够从香港高校获得计量经济学的师资培养，并展开密切的学术交流与合作研究。所有这些，都直接推动和促进岭南学院在计量经济学的理论研究和应用研究领域长足发展，相关成果发表于*Journal of Econometrics*、《经济研究》、《经济学》(季刊)、《数量经济技术经济研究》、《统计研究》和《管理世界》等国内外权威期刊。

这套《岭南实证与计量经济学研究丛书》是过去10多年间，岭南学院在计量经济学的理论研究和应用研究领域所取得的阶段性成果。丛书选题涵盖理论与应用计量，学术特色主要有两个：一是理论创新明显。理论研究反映了计量经济学某一方面的最新发展状况，形成多个具有学术价值和应用前景的创新点，比如Tobit模型的非参数与半参数估计、动态博弈模型估计、面板单位根与协整检验统计量构造，空间相关模型的有效估计等理论计量前沿性问题；另一个是实证研

究视角独特。实证研究密切围绕中国社会经济发展转型中的重要问题，努力提炼出具有学术性的一般问题，综合运用前沿实证分析方法与工具进行规范的实证研究，致力于科研成果国际化。

当前，中国计量经济学已迈入“国际化”的新阶段，需要越来越多的年轻学者致力于理论计量的基础性创新研究，以使得中国在国际计量经济学界占有一席之地；需要越来越多的新模型、新方法应用到中国的经济学实证研究，从而将中国经济故事以科学的经济学语言呈现出来；需要越来越多地关注实证应用研究的规范化问题，全面吸收和借鉴国际计量经济学界对于可信性问题的成果，改进现行的研究模式和教学模式。本丛书的出版，表明一批掌握了现代经济学的研究方法和计量经济学分析工具的青年学者在成长；他们知识结构新，学术素养好，已成为高等院校和研究机构教学、科研的生力军。

我们希望本丛书的出版，能够为对相关领域感兴趣的博士研究生以及研究人员提供一些有用的思考或参照，为有关经济政策的制定提供依据，为中国计量经济学的学科建设添砖加瓦。作为探索性研究的成果，书中肯定有很多需要改进的地方，希望读者批评指正。

王美今　徐现祥

2013 年 5 月

目 录

第一章

Tobit 模型估计概述及准备知识

自 Tobin 于 1958 年具开创性地对居民耐用品消费支出计量模型进行研究之后，Tobit 模型的估计问题一直是微观计量经济学的重要研究领域。20 世纪 80 年代之前，此类模型基本上被设定为扰动项服从某种已知分布(如正态分布)的参数模型，然后由最大似然方法或二阶段最小二乘法对模型中的参数进行估计。但是，Aramazar 和 Schmidt (1982)的研究结果表明，扰动项正态性的假定对于 Tobit 模型的一致性估计是至关重要的；当模型设定产生偏误且扰动项不服从正态分布时，基于最大似然或二阶段最小二乘法的参数估计是不一致的，其估计结果和统计推断可能会对经济分析产生错误导向。

本章以具有固定归并点的归并(或删失)数据回归模型(censored regression model)以及相应的截断数据回归模型(truncated regression model)为例，回顾较早期文献对它们的参数和半参数估计方法；然后引入本书所用方法，并给出本书估计方法(主要是半参数估计)研究的相关准备知识，主要是关于经验过程和 U 统计量分解的相关概念和定理。

假设真实的数据生成过程为

$$y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta}_0 + u_i, \quad i=1,\ 2,\ \cdots,\ n \tag{1.1}$$

其中，$\boldsymbol{x}$ 是 k 维列向量；$\boldsymbol{\beta}_0$ 是 k 维参数向量。但实际观察到的数据为 $\{y_i,\ \boldsymbol{x}_i:\ i=1,\ 2,\ \cdots,\ n\}$，其中，

$$y_i = \max(y_i^*,\ 0) \tag{1.2}$$

即 y_i^* 只在大于 0 的时候才能被观察，所有小于 0 的观测值都被归并到 0 这一点，称 y_i 为归并因变量。当归并点不为 0 时，可以通过被解释变量的一个平移变换得到模型(1.1)和模型(1.2)。

如果 y_i^* 和 $\boldsymbol{x}_i$ 都只在 $y_i^*>0$ 的时候才能被观察，则得到截断数据回归模型，此时的实际观测数据为 $\{y_i,\ \boldsymbol{x}_i:\ i=1,\ 2,\ \cdots,\ n_0\}$，其中，$y_i>0$ 表示截断因变量，而 n_0 表示 y_i^* 可观测的样本数量。归并数据回归模型比截断数据回归模型包含更多的数据信息，表现在两个方面：第一，截断数据回归模型并不知道被

截断的数据量有多少，即它只能观察到 n_0 个样本而不能观察 n 个全样本；第二，对于 $y_i^* \leqslant 0$ 的样本，归并数据回归模型仍可观察到 $\boldsymbol{x}_i$ 的值，而截断数据回归模型则丢失了这部分信息。因此，截断数据回归模型的估计方法均可用于归并数据回归模型的估计；反之不然。但是，如果能对归并数据回归模型新增的数据信息加以利用，应能得到参数更优的估计。归并数据回归模型和截断数据回归模型是经典 Tobit 模型的两个主要类型。

假设模型 $y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta}_0 + u_i$ 满足 Gauss-Markov 基本假定的要求。我们感兴趣的是参数向量 $\boldsymbol{\beta}_0$的估计，即 $\boldsymbol{x}$ 对潜变量 y_i^* 的边际影响的估计。现在的问题是，y_i^* 不被观察，我们仅可观察到 y_i。故从数据可得性上看，我们不可以由 y_i^* 关于 $\boldsymbol{x}_i$ 的回归得到$\boldsymbol{\beta}_0$的估计，而应由可观察的 y_i 出发，探讨 $\boldsymbol{\beta}_0$的估计。

对于归并因变量 y 来说，其条件数学期望是

$$\begin{aligned} E[y \mid \boldsymbol{x}] &= P[y > 0 \mid \boldsymbol{x}] \cdot E[y \mid y > 0, \boldsymbol{x}] \\ &= P[u > -\boldsymbol{x}'\boldsymbol{\beta}_0 \mid \boldsymbol{x}] \cdot E[\boldsymbol{x}'\boldsymbol{\beta}_0 + u \mid u > -\boldsymbol{x}'\boldsymbol{\beta}_0, \boldsymbol{x}] \\ &= [1 - F_{u|\boldsymbol{x}}(-\boldsymbol{x}'\boldsymbol{\beta}_0)]\boldsymbol{x}'\boldsymbol{\beta}_0 + \int_{-\boldsymbol{x}'\boldsymbol{\beta}_0}^{\infty} t f_{u|x}(t)\mathrm{d}t \end{aligned}$$

其中，$F_{u|\boldsymbol{x}}(t)$是扰动项 u 的条件分布函数；$f_{u|\boldsymbol{x}}(t)$是相应的条件密度函数。可见，利用这种归并(censored)的数据，将 y 关于 $\boldsymbol{x}$ 回归得到的普通最小二乘(ordinary least squares，OLS)估计量 $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$不是参数 $\boldsymbol{\beta}_0$的一致性估计。同样，对于截断因变量来说，其条件数学期望是

$$\begin{aligned} E[y \mid \boldsymbol{x}] &= E[y \mid y>0, \boldsymbol{x}] \\ &= E[\boldsymbol{x}'\boldsymbol{\beta}_0 + u \mid u > -\boldsymbol{x}'\boldsymbol{\beta}_0, \boldsymbol{x}] = \boldsymbol{x}'\boldsymbol{\beta}_0 + \int_{-\boldsymbol{x}'\boldsymbol{\beta}_0}^{\infty} t f_{u|\boldsymbol{x}}(t)\mathrm{d}t \end{aligned}$$

可见，利用这种截断(truncated)数据，将 y 关于 $\boldsymbol{x}$ 回归存在遗漏变量问题，从而 OLS 估计量 $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$不是参数 $\boldsymbol{\beta}_0$ 的一致性估计。

Tobit 模型(1.1)和模型(1.2)的传统估计方法为参数的最大似然估计方法(Amemiya，1973)以及基于似然函数(likelihood-based)的其他一些参数方法(Heckman，1979)。这些方法假设扰动项服从某个具体的参数分布，然后由最大似然估计方法估计。例如，在扰动项 u_i 服从正态分布 $\mathrm{N}(0, \sigma^2)$之下，归并数据模型(1.1)和模型(1.2)的似然目标函数是

$$\prod_{d_i=0} [1 - \Phi(\boldsymbol{x}_i'\boldsymbol{\beta}/\sigma)] \prod_{d_i=1} [\sigma^{-1}\phi((y_i - \boldsymbol{x}_i'\boldsymbol{\beta})/\sigma)]$$

其中，$d_i = 1\{y_i > 0\}$是反映被解释变量观察值没有被归并的结果的示性函数，而第一项反映了被解释变量被归并时其似然值对似然目标函数的贡献；$\phi(\cdot)$和 $\Phi(\cdot)$分别表示标准正态分布的密度函数和累积分布函数。参数($\boldsymbol{\beta}_0$，σ^2)的最大似然估计量是由上述目标函数最大化问题的一阶条件求解得到的(需迭代算法求解)。

二步法估计使用可观察变量 y_i 的条件期望：

$$E[y_i \mid \boldsymbol{x}_i]=\Phi(\boldsymbol{x}_i'\boldsymbol{\beta}_0/\sigma)\boldsymbol{x}_i'\boldsymbol{\beta}_0+\sigma\phi(\boldsymbol{x}_i'\boldsymbol{\beta}_0/\sigma)$$

首先，由 $d_i\equiv 1\{y_i>0\}$ 关于 $\boldsymbol{x}_i$ 作 Probit 估计得 $\boldsymbol{\beta}_0/\sigma$ 的估计 $\hat{\alpha}$，其对应的似然函数是 $\prod\limits_{d_i=0}[1-\Phi(\boldsymbol{x}_i'\boldsymbol{\beta}/\sigma)]\prod\limits_{d_i=1}\Phi(\boldsymbol{x}_i'\boldsymbol{\beta}/\sigma)$；记 $\Phi_i\equiv\Phi(\boldsymbol{x}_i'\hat{\alpha})$ 和 $\phi_i\equiv\phi(\boldsymbol{x}_i'\hat{\alpha})$。其次，将 y_i(包括 0 观察点)关于 $\Phi_i\boldsymbol{x}_i$ 和 ϕ_i 作回归得 $\boldsymbol{\beta}_0$ 和 σ 的估计 $\hat{\boldsymbol{\beta}}$ 和 $\hat{\sigma}$。

可以证明，当扰动项分布设定正确，且具有条件同方差时，最大似然估计量(maximum likelihood estimator，MLE)和二步法估计量具有一致性和渐近正态性。但是，这些良好的性质依赖于对扰动项分布的正确设定。如果设定不正确，MLE 一般是不一致的(Aramazar and Schmidt，1982)。另外，如果扰动项的条件异方差被忽略，那么即使其分布被正确设定，参数最大似然方法也会得到不一致的估计(Maddala and Nelson，1975)。这些结论激发了人们对 Tobit 模型的半参数估计方法的研究，以寻求更稳健的估计量。本章所指的半参数方法是指在保留回归方程的线性形式(参数部分)的同时，放松对扰动项分布的参数设定，允许扰动项服从某个未知分布(非参数部分)，只要这个分布满足一定的正则条件即可。这些正则条件一般只对分布的光滑性以及矩的存在性做出要求，可包含一大类分布，当然包括正态分布。由于不需要对扰动项的具体分布进行人为设定，因此半参数估计量的性质对于扰动项的不同分布是稳健的。

第一节　早期的半参数估计量

早期的 Tobit 模型半参数估计方法主要包括 Miller(1976)、Buckley 和 James(1979)及 Koul 等(1981)，它们的核心都在于对 OLS 估计进行修正，以适用于归并数据回归模型的估计。这三种方法均针对随机归并模型而提出，其中 Koul 等(1981)的估计方法涉及归并点变量的随机分布，不能直接应用于具有固定归并点的模型(1.1)和模型(1.2)，因此本节只简述前两种方法。

在此之前，有必要对归并数据下分布函数的乘积限估计量(product limit estimator)作一介绍。假设$\{V_1, V_2, \cdots, V_n\}$是随机变量 V 的一个随机样本，V 的密度函数和分布函数分别为 $g(v)$ 和 $G(v)$，但实际可观测的是一个归并样本$\{W_1, W_2, \cdots, W_n\}$，其中，$W_i=\max(V_i, c_i)$，$c_i$ 是已知常数。现在希望用归并样本$\{W_1, W_2, \cdots, W_n\}$构造 $g(v)$ 和 $G(v)$ 的估计量。如果$\{V_1, V_2, \cdots, V_n\}$可以被观测，则 $g(v)$ 的一个直观估计是

$$g_n(v)=\begin{cases}\dfrac{1}{n}, & v=V_i;\ i=1,2,\cdots,n\\ 0, & \text{其他}\end{cases}$$

即用一个离散分布去逼近真实分布，并赋予每个样本点以相同的概率质量

(probability mass)。此时，如果记 $V_{(1)} \leqslant V_{(2)} \leqslant \cdots \leqslant V_{(n)}$ 为 $\{V_1, V_2, \cdots, V_n\}$ 所对应的顺序统计量，且令 $V_{(0)} = -\infty$ 和 $V_{(n+1)} = +\infty$，那么 $G(v)$ 的估计量为

$$G_n(v) = \sum_{j=0}^{n} \frac{j}{n} 1\{V_{(j)} \leqslant v < V_{(j+1)}\}$$

但是，当某些样本点被归并时，仍然赋予这些样本点以 $1/n$ 的概率质量显然是不妥当的。乘积限估计的思想在于把这些被归并的样本点的概率质量平均分配给比它小的所有样本点，这是因为，当样本点 V_i 被归并而观察值为 $W_i = c_i$ 时，我们只知道 $V_i \leqslant W_i = c_i$，但并不知道 V_i 的具体数值，因此 V_i 有可能等于满足 $W_j < W_i$ 的所有 W_j。为得到 $G(v)$ 的估计 $\hat{G}(v)$，在操作上，首先对可观测的样本 $\{W_1, W_2, \cdots, W_n\}$ 进行排序得到 $W_{(1)} \leqslant W_{(2)} \leqslant \cdots \leqslant W_{(n)}$。记 $\delta_i \equiv 1\{V_i > c_i\}$，$\delta_i = 1$ 表示第 i 个样本没有被归并。相应地，记 $\delta_{(i)} \equiv 1\{W_{(i)} = V_{(i)}\}$ 为 $W_{(i)}$ 所对应的 δ 值，则归纳可证，$G(v)$ 的乘积限估计量为

$$\hat{G}(v) = \prod_{i:\, W_{(i)} > v} \left(\frac{i-1}{i}\right)^{\delta_{(i)}} \tag{1.3}$$

Kaplan 和 Meier(1958)证明了乘积限估计量 $\hat{G}(v)$ 是分布函数 $G(v)$ 在归并数据下的非参数 MLE，且可以通过解方程

$$\widetilde{G}(v) = \frac{1}{n} \sum_{i=1}^{n} \left[1\{W_{(i)} \leqslant v\} + (1-\delta_{(i)}) 1\{W_{(i)} > v\} \frac{\widetilde{G}(v)}{\widetilde{G}(W_{(i)})}\right] \tag{1.4}$$

得到。方程(1.4)称为自一致性(self-consistency)方程。Efron(1967)证明方程(1.4)有且仅有唯一解[式(1.3)]。

现在回到归并数据回归模型(1.1)和模型(1.2)。先看 Miller(1976)的估计。Miller 方法的思想在于对线性回归模型 OLS 估计的目标函数进行修正以适用于归并数据回归模型。Miller(1976)指出，分布函数的乘积限估计量[式(1.3)]可用于归并数据回归模型的半参数估计。假定 $Eu_i = 0$，否则令其均值 $\alpha_0 \equiv Eu_i$ 作为常数与原来模型中的截距项合并至回归函数中。如果 $\{y_i^*\}$ 可以被观测，我们可以通过最小化

$$\int u^2 \mathrm{d}G_n(u; \boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^{n} u_i^2(\boldsymbol{\beta})$$

得到 $\boldsymbol{\beta}_0$ 的 OLS 估计量，其中，$u_i(\boldsymbol{\beta}) = y_i^* - \boldsymbol{x}_i'\boldsymbol{\beta}$，$G_n(u; \boldsymbol{\beta}) = \frac{j}{n}$，$u_{(j)}(\boldsymbol{\beta}) \leqslant u < u_{(j+1)}(\boldsymbol{\beta})$，$\{u_{(i)}(\boldsymbol{\beta})\}$ 是 $\{u_i(\boldsymbol{\beta})\}$ 对应的顺序统计量。但在归并数据回归模型中，我们只能观测到 $y_i = \max(y_i^*, 0)$ 以及 $e_i(\boldsymbol{\beta}) \equiv y_i - \boldsymbol{x}_i'\boldsymbol{\beta} = \max(u_i(\boldsymbol{\beta}), -\boldsymbol{x}_i'\boldsymbol{\beta})$，即 $\{u_i(\boldsymbol{\beta})\}$ 的一个归并样本。因此，很自然的一个思路是把上述 OLS 估计量目标函数中的 $G_n(u; \boldsymbol{\beta})$ 替换成乘积限估计量：

$$\hat{G}(e;\ \boldsymbol{\beta})=\prod_{i:\ e_{(i)}(\boldsymbol{\beta})>u}\left(\frac{i-1}{i}\right)^{d_{(i)}}$$

其中，$d_i=1\{y_i>0\}$；$d_{(i)}$是$e_{(i)}(\boldsymbol{\beta})$所对应的$d$值，再通过最小化

$$\int e^2\mathrm{d}\hat{G}(e;\ \boldsymbol{\beta})=\sum_{i=1}^{n}(y_i-\boldsymbol{x}'_i\boldsymbol{\beta})^2\hat{g}(e_i(\boldsymbol{\beta})) \tag{1.5}$$

得到$\boldsymbol{\beta}_0$的估计，这就是 Miller(1976)提出的估计量。如果所有的观测值都没有发生归并，即所得的数据为随机样本，则对所有$i=1,\ 2,\ \cdots,\ n$都有$d_i=1$，从而$\hat{G}(e;\ \boldsymbol{\beta})=G_n(u;\ \boldsymbol{\beta})$，Miller 估计量退化为 OLS 估计量。注意到目标函数(1.5)对$\boldsymbol{\beta}$不连续，因此微分方法不可用。Miller(1976)建议用迭代方法求解$\boldsymbol{\beta}_0$的估计$\hat{\boldsymbol{\beta}}$，即

$$\hat{\boldsymbol{\beta}}_{k+1}=\left[\sum_{i=1}^{n}\hat{G}(e_i(\hat{\boldsymbol{\beta}}_k))\boldsymbol{x}_i\boldsymbol{x}'_i\right]^{-1}\left[\sum_{i=1}^{n}\hat{G}(e_i(\hat{\boldsymbol{\beta}}_k))\boldsymbol{x}_iy_i\right]$$

直到收敛为止。初始的$\hat{\boldsymbol{\beta}}_0$可以选用非归并数据$y_i$对$\boldsymbol{x}_i$作回归的 OLS 估计。

与 Miller 方法不同，Buckley 和 James(1979)通过修正正规方程组(目标函数的一阶条件)来构造归并数据模型的半参数估计量。在归并数据回归模型中，由于$E[y_i\,|\,\boldsymbol{x}_i]\neq\boldsymbol{x}'_i\boldsymbol{\beta}_0$，因此不能直接解最小二乘(least square，LS)正规方程组来估计模型中的未知参数$\boldsymbol{\beta}_0$。但是，如果用潜在因变量y_i^*的期望值替代被归并的观测值$y_i=0$，即定义

$$\tilde{y}_i=\begin{cases}y_i, & d_i=1\\ E[y_i^*\,|\,y_i^*\leqslant 0,\ \boldsymbol{x}_i], & d_i=0\end{cases}$$

则有

$$\begin{aligned}E[\tilde{y}_i\,|\,\boldsymbol{x}_i]&=E[y_i1\{y_i^*>0\}\,|\,\boldsymbol{x}_i]+E[E[y_i^*\,|\,y_i^*\leqslant 0,\ \boldsymbol{x}_i]1\{y_i^*\leqslant 0\}\,|\,\boldsymbol{x}_i]\\&=E[y_i1\{y_i^*>0\}\,|\,\boldsymbol{x}_i]+E\left[\frac{E[y_i^*1\{y_i^*\leqslant 0\}\,|\,\boldsymbol{x}_i]}{P(y_i^*\leqslant 0\,|\,\boldsymbol{x}_i)}1\{y_i^*\leqslant 0\}\,|\,\boldsymbol{x}_i\right]\\&=E[y_i^*1\{y_i^*>0\}\,|\,\boldsymbol{x}_i]+E[y_i^*1\{y_i^*\leqslant 0\}\,|\,\boldsymbol{x}_i]=E[y_i^*\,|\,\boldsymbol{x}_i]=\boldsymbol{x}'_i\boldsymbol{\beta}_0\end{aligned}$$

从而正规方程组

$$\sum_{i=1}^{n}\boldsymbol{x}_i(\tilde{y}_i-\boldsymbol{x}'_i\boldsymbol{\beta})=0$$

的解可以作为未知参数$\boldsymbol{\beta}_0$的合理估计，即

$$\tilde{\boldsymbol{\beta}}=\left[\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}'_i\right]^{-1}\sum_{i=1}^{n}\boldsymbol{x}_i\tilde{y}_i \tag{1.6}$$

因为$E[y_i^*\,|\,y_i^*\leqslant 0,\ \boldsymbol{x}_i]$未知，所以需要用迭代方法得到参数$\boldsymbol{\beta}_0$的可行性估计。假设$\boldsymbol{\beta}_0$的第$k$步估计为$\hat{\boldsymbol{\beta}}_k$，由于

$$E[y_i^*\,|\,y_i^*\leqslant 0,\ \boldsymbol{x}_i]=\boldsymbol{x}'_i\boldsymbol{\beta}_0+E[u_i\,|\,u_i\leqslant -\boldsymbol{x}'_i\boldsymbol{\beta}_0,\ \boldsymbol{x}_i]$$

$$= \boldsymbol{x}_i'\boldsymbol{\beta}_0 + \frac{E[u_i 1\{u_i \leqslant -\boldsymbol{x}_i'\boldsymbol{\beta}_0\} \mid \boldsymbol{x}_i]}{P(u_i \leqslant -\boldsymbol{x}_i'\boldsymbol{\beta}_0 \mid \boldsymbol{x}_i)}$$

我们先用

$$\hat{y}_i(\hat{\boldsymbol{\beta}}_k) \equiv \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_k + \frac{\sum_{j=1}^{n} e_j(\hat{\boldsymbol{\beta}}_k) 1\{e_j(\hat{\boldsymbol{\beta}}_k) < e_i(\hat{\boldsymbol{\beta}}_k)\} \hat{g}_j}{\hat{G}(-\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_k;\ \hat{\boldsymbol{\beta}}_k)}$$

作为 $E[y_i^* \mid y_i^* \leqslant 0,\ \boldsymbol{x}_i]$ 的估计，其中，$e_j(\hat{\boldsymbol{\beta}}_k) \equiv y_j - \boldsymbol{x}_j'\hat{\boldsymbol{\beta}}_k$；$\hat{G}(u;\ \boldsymbol{\beta})$ 是式(1.3)所定义的乘积限估计量；$\hat{g}_j$ 是按 $\hat{G}(u;\ \boldsymbol{\beta})$ 分配给第 j 个样本点的概率质量。然后，用 $\tilde{y}_i(\hat{\boldsymbol{\beta}}_k) \equiv d_i y_i + (1-d_i)\hat{y}_i(\hat{\boldsymbol{\beta}}_k)$ 代替正规方程组或式(1.6)中的 $\tilde{y}_i$，得到第 $k+1$ 步的估计：

$$\hat{\boldsymbol{\beta}}_{k+1} = \left(\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i \tilde{y}_i(\hat{\boldsymbol{\beta}}_k)$$

初始估计可使用非归并数据样本 y_i 对 x_i 作回归的 OLS 估计。

Miller 和 Halpern(1982)通过数值模拟比较了 Miller(1976)、Buckley 和 James(1979)及 Koul 等(1981)的估计量，发现在这三种估计量中，Buckley-James 估计量的小样本表现最好。James 和 Smith(1984)证明了 Buckley-James 估计量的弱一致性(即估计量依概率收敛于真正参数)。Lai 和 Ying(1991)对 Buckley-James 估计量进行了改良，克服了乘积限估计量在下尾处的不稳定性，并证明了改良后的 Buckley-James 估计量的一致性和渐近正态性。

Miller 估计量和 Buckley-James 估计量都可以推广到截断数据以及双边归并(doubly censored)数据情形，只需把估计量构造过程中出现的乘积限估计相应地置换成 Lynden-Bell 估计(Tsui et al.，1988)和 Turnbull(1974)的分布估计即可。

第二节 CLAD 估计量和 STLS 估计量

Miller 估计量和 Buckley-James 估计量的一致性和渐近正态性没有完全建立，且对扰动项的同分布假定具有敏感性。文献中对 Tobit 模型提出了其他一些半参数估计方法，其中估计量具有良好的渐近性质，且关于扰动项分布的一些约束有稳健性。Powell(1984)把标准线性回归模型的最小绝对偏误(least absolute deviation，LAD)估计方法推广至归并数据模型，构造了归并最小绝对偏误(censored LAD，CLAD)估计量。它的构造基于如下事实：对于任一随机变量 z，其分布的中位数是使目标函数 $E[\,|z-b| - |z|\,]$ 最小化的解。对于归并数据回归模型，如果扰动项 u_i 存在唯一的条件中位数零：$\text{Median}(u_i \mid \boldsymbol{x}_i) = 0$，则 y_i

的条件中位数函数是Median$(y_i \mid \boldsymbol{x}_i)=\max(0, \boldsymbol{x}_i'\boldsymbol{\beta}_0)$。由此，参数$\boldsymbol{\beta}_0$可由以下最小化问题

$$\min_{\boldsymbol{\beta}\in\boldsymbol{\Theta}}\sum_{i=1}^{n}\left| y_i - \max(0, \boldsymbol{x}_i'\boldsymbol{\beta})\right|$$

的解得到估计，即为 CLAD 估计量 $\hat{\boldsymbol{\beta}}_{\mathrm{CLAD}}$。Powell(1984)证明，在扰动项 u_i 独立同分布，中位数为零，并与 x_i 是相互独立的条件下，如果关于扰动项和解释变量的分布具有相应的正则性，该估计量具有强一致性(几乎处处收敛于真正的参数)和渐近正态性：$2f(0)M^{1/2}(\hat{\boldsymbol{\beta}}_{\mathrm{CLAD}}-\boldsymbol{\beta}_0)\to \mathrm{N}(0, \boldsymbol{I})$，其中，$f(0)>0$ 是 u_i 的密度函数在零处的取值；$M=E\left[n^{-1}\sum_i 1\{\boldsymbol{x}_i'\boldsymbol{\beta}_0>0\}\boldsymbol{x}_i\boldsymbol{x}_i'\right]$；$\boldsymbol{I}$ 是单位矩阵。这里对扰动项 u_i 的分布没有做具体限制，只要求其条件中位数为零。除常用的正态分布、Logistic 等具有对称性的分布外，此估计量对于更为广泛的扰动项分布都是一致的和渐近正态的，且关于异方差是稳健的。但是，目标函数(1.6)的构造利用了 $y_i=0$ 时解释变量的信息，因此 LAD 方法不能用于截断模型的估计。

Powell(1986a)基于对扰动项进行剪切(trimming)的思想提出了半参数截断模型的对称截断最小二乘(symmetrically truncated least squares, STLS)估计量。假设扰动项 u_i 的条件分布关于 0 对称。如果潜在因变量 y_i^* 可以被观测，那么扰动项的对称分布假设所隐含的零条件均值可以保证 OLS 的无偏性和一致性。但是，我们只能观察到正的 y_i^*，这意味着，即使$\boldsymbol{\beta}_0$已知，我们也只能观察到大于$-\boldsymbol{x}_i'\boldsymbol{\beta}_0$部分的 u_i，因此“可观测”的扰动不再具有对称性。但是，如果在$\boldsymbol{x}_i'\boldsymbol{\beta}_0$处对“可观测”的扰动再进行一次剪切，只保留小于$\boldsymbol{x}_i'\boldsymbol{\beta}_0$的扰动，那么就能恢复扰动项的对称性，使得最小二乘估计量一致。这样构造的估计量即为对称截断最小二乘估计量。具体地，解正规方程组

$$0=\sum_{i=1}^{n_0}1\{u_i<\boldsymbol{x}_i'\boldsymbol{\beta}_0\}\cdot(y_i-\boldsymbol{x}_i'\boldsymbol{\beta})\boldsymbol{x}_i=\sum_{i=1}^{n_0}1\{y_i<2\boldsymbol{x}_i'\boldsymbol{\beta}_0\}\cdot(y_i-\boldsymbol{x}_i'\boldsymbol{\beta})\boldsymbol{x}_i$$

即可得到$\boldsymbol{\beta}_0$的一致性估计。但是真实参数$\boldsymbol{\beta}_0$未知，因此一个可行的估计量是下面渐近“一阶条件”的解：

$$o_p(n_0^{1/2})=\sum_{i=1}^{n_0}1\{y_i<2\boldsymbol{x}_i'\boldsymbol{\beta}\}\cdot(y_i-\boldsymbol{x}_i'\boldsymbol{\beta})\boldsymbol{x}_i \tag{1.7}$$

一阶条件[式(1.7)]有可能存在多重解(multiple solutions)，为此由其右边积分，将问题转化为如下最小化问题的求解：

$$\min_{\boldsymbol{\beta}\in\boldsymbol{\Theta}}\left\{R(\boldsymbol{\beta})\equiv\sum_{i=1}^{n_0}\left(y_i-\max\left(\frac{1}{2}y_i, \boldsymbol{x}_i'\boldsymbol{\beta}\right)\right)^2\right\} \tag{1.8}$$

即 $\hat{\boldsymbol{\beta}}_{\mathrm{STLS}}=\arg\min\limits_{\boldsymbol{\beta}\in\boldsymbol{\Theta}}R(\boldsymbol{\beta})$。这种方法也可用于归并数据回归模型的估计。对于归并

模型，类似地定义一阶条件

$$0=\sum_{i=1}^{n}1\{\boldsymbol{x}_i'\boldsymbol{\beta}>0\}\cdot(\min(y_i,\ 2\boldsymbol{x}_i'\boldsymbol{\beta})-\boldsymbol{x}_i'\boldsymbol{\beta})\cdot\boldsymbol{x}_i$$

及其对应的最小化问题

$$\begin{aligned}\min_{\boldsymbol{\beta}\in\boldsymbol{\Theta}}S(\boldsymbol{\beta})\equiv&\sum_{i=1}^{n}\left(y_i-\max\left(\frac{1}{2}y_i,\ \boldsymbol{x}_i'\boldsymbol{\beta}\right)\right)^2\\&+\sum_{i=1}^{n}1\{y_i>2\boldsymbol{x}_i'\boldsymbol{\beta}\}\cdot\left[\left(\frac{1}{2}y_i\right)^2-(\max(0,\ \boldsymbol{x}_i'\boldsymbol{\beta}))^2\right]\end{aligned}\tag{1.9}$$

并定义对称归并最小二乘(symmetrically censored least squares，SCLS)估计量为最小化问题[式(1.9)]的解：$\hat{\boldsymbol{\beta}}_{\mathrm{SCLS}}=\arg\min\limits_{\boldsymbol{\beta}\in\boldsymbol{\Theta}}S(\boldsymbol{\beta})$。在扰动项的条件分布关于0对称的条件下(截断模型还要求扰动项分布具有单众数)，Powell(1986a)证明了STLS估计量和SCLS估计量的一致性和渐近正态性，且这些性质对于扰动项的不同分布以及不同形式的异方差都是稳健的。

有趣的是，如果把STLS估计量的目标函数$S(\boldsymbol{\beta})$中的平方损失函数换成绝对值损失函数，即定义

$$\begin{aligned}S^*(\beta)\equiv&\sum_{i=1}^{n}\left|y_i-\max\left(\frac{1}{2}y_i,\ \boldsymbol{x}_i'\boldsymbol{\beta}\right)\right|\\&+\sum_{i=1}^{n}1\{y_i>2\boldsymbol{x}_i'\boldsymbol{\beta}\}\cdot\left[\left|\frac{1}{2}y_i\right|-|\max(0,\ \boldsymbol{x}_i'\boldsymbol{\beta})|\right]\end{aligned}$$

那么$S^*(\boldsymbol{\beta})$恰好为CLAD估计量的目标函数：$S^*(\boldsymbol{\beta})=\sum\limits_{i=1}^{n}|y_i-\max(0,\ \boldsymbol{x}_i'\boldsymbol{\beta})|$，此时的STLS估计量即为CLAD估计量。如果把损失函数换成其他形式，还可以构造其他类型的估计量。另外需要指出的是，对称剪切的思想还可以用于非固定(但已知)归并(或截断)点情形以及双边归并(或截断)情形的模型估计。

M. J. Lee(1992)认为，SCLS估计量所使用的矩条件：

$$\begin{aligned}&E_{x_i}[1\{\boldsymbol{x}_i'\boldsymbol{\beta}_0>0\}\boldsymbol{x}_iE_{u_i\mid x_i}[1\{|u_i|\leqslant\boldsymbol{x}_i'\boldsymbol{\beta}_0\}u_i\\&\quad+(1\{u_i>\boldsymbol{x}_i'\boldsymbol{\beta}_0\}-1\{u_i<-\boldsymbol{x}_i'\boldsymbol{\beta}_0\})\boldsymbol{x}_i'\boldsymbol{\beta}_0]]=0\end{aligned}$$

并不是对OLS矩条件$E[\boldsymbol{x}_iu_i]=0$进行修正的唯一方式。假设w是一个大于0的常数，那么用w进行缩尾(winsorize)后的扰动项满足下面的矩条件：

$$E_{x_i}[1\{\boldsymbol{x}_i'\boldsymbol{\beta}_0>w\}\boldsymbol{x}_iE_{u_i\mid x_i}[1\{|u_i|\leqslant w\}u_i+(1\{u_i>w\}-1\{u_i<w\})w]]=0\tag{1.10}$$

即在给定$\boldsymbol{x}_i'\boldsymbol{\beta}_0>w$的条件下，当$u_i$小于$-w$时把$u_i$换成$-w$，而当$u_i$大于$w$时把$u_i$换成$w$，这样缩尾后的扰动仍然满足零条件均值。对一阶条件[式(1.10)]进行积分可以得到目标函数：

$$Q(\boldsymbol{\beta})\equiv\sum_i\frac{1}{2}[y_i-\max(\boldsymbol{x}_i'\boldsymbol{\beta},\ w)]^21\{|y_i-\max(\boldsymbol{x}_i'\boldsymbol{\beta},\ w)|<w\}$$

$$+\sum_i [w \mid y_i - \max(\boldsymbol{x}_i'\boldsymbol{\beta},\ w) \mid - (w^2/2)] 1\{\mid y_i - \max(\boldsymbol{x}_i'\boldsymbol{\beta},\ w) \mid \geqslant w\} \tag{1.11}$$

通过最小化式(1.11)可以定义一个新的估计量，M. J. Lee(1992)将这个估计量称为缩尾均值(winsorized mean，WM)估计量。令 $w \to 0$，可以看到 CLAD 估计量是 WM 估计量的一个极限形式。与对称剪切估计量类似，将目标函数 $Q(\boldsymbol{\beta})$ 的第二项去掉，即可作为截断模型 WM 估计量的目标函数。M. J. Lee(1992)证明，只要扰动项 u_i 的分布在 $-w$ 到 w 之间对称，那么 WM 估计量就是一致的和渐近正态的。

第三节 成对差分估计量

Powell(1986a)的对称剪切最小二乘(symmetrically trimming least squares，STLS)估计量要求扰动项的分布具有对称性和单众数性。这两个约束条件较强，限制了对称剪切方法的应用范围。Honoré 和 Powell(1994)改进了这种方法，构造了成对差分(pairwise difference)估计量。成对差分方法所使用的事实是：如果两个随机变量独立同分布，那么它们相减后得到的随机变量的分布关于 0 对称。假设扰动序列$\{u_i\}$独立同分布。但是，由于我们只能观察到 y_i 而非 y_i^*，因此即使$\boldsymbol{\beta}_0$已知，我们也只能观察到 $e_i(\boldsymbol{\beta}_0) = y_i - \boldsymbol{x}_i'\boldsymbol{\beta}_0 = \max(u_i,\ -\boldsymbol{x}_i'\boldsymbol{\beta}_0)$，而可观测的扰动序列$\{e_i(\boldsymbol{\beta}_0)\}$并非独立同分布。Honoré 和 Powell(1994)注意到，可观测的扰动序列经剪切后仍然是独立同分布的。具体来说，就是利用第 j 个样本的 $-\boldsymbol{x}_j'\boldsymbol{\theta}_0$ 剪切第 i 个样本的可观测扰动 $e_i(\boldsymbol{\beta}_0)$，得到剪切扰动(trimmed error)序列 $e_{ij}(\boldsymbol{\beta}_0) \equiv \max(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}_0,\ -\boldsymbol{x}_j'\boldsymbol{\beta}_0) = \max(u_i,\ -\boldsymbol{x}_i'\boldsymbol{\beta}_0,\ -\boldsymbol{x}_j'\boldsymbol{\beta}_0)$，那么 $e_{ij}(\boldsymbol{\beta}_0)$ 与 $e_{ji}(\boldsymbol{\beta}_0)$在给定 $\boldsymbol{x}_i$ 和 $\boldsymbol{x}_j$ 的条件下独立同分布，因此 $e_{ij}(\boldsymbol{\beta}_0) - e_{ji}(\boldsymbol{\beta}_0)$的条件分布关于 0 对称，从而 $E[e_{ij}(\boldsymbol{\beta}_0) - e_{ji}(\boldsymbol{\beta}_0) \mid \boldsymbol{x}_i,\ \boldsymbol{x}_j] = 0$。实际上，对于任意奇函数 $\xi(\cdot)$，也有一阶条件矩成立：$E[\xi(e_{ij}(\boldsymbol{\beta}_0) - e_{ji}(\boldsymbol{\beta}_0)) \mid \boldsymbol{x}_i,\ \boldsymbol{x}_j] = 0$。以其作为最小化问题的一阶条件，且使用其样本形式，则真实参数 $\boldsymbol{\beta}_0$可由以下最小化问题的解：

$$\hat{\boldsymbol{\beta}}_{\mathrm{HP}} \equiv \arg\min_{\boldsymbol{\beta}} \binom{n}{2}^{-1} \sum_{i=1}^{n} \sum_{j=1,\ j\neq i}^{n} s(y_i,\ y_j,\ (\boldsymbol{x}_i - \boldsymbol{x}_j)'\boldsymbol{\beta}) \tag{1.12}$$

得到估计，其中，

$$\begin{aligned} s(y_i,\ y_j,\ \delta) = & (\Xi(y_i) - (y_j + \delta)\xi(y_i)) 1\{\delta \leqslant -y_j\} \\ & + \Xi(y_i - y_j - \delta) 1\{-y_j < \delta < y_i\} \\ & + (\Xi(-y_j) - (\delta - y_i)\xi(-y_j)) 1\{\delta \geqslant y_i\} \end{aligned}$$

这里，$\Xi(\cdot)$ 是 $\xi(\cdot)$ 的积分，它是一个偶函数。例如，若 $\xi(d) = 2d$，则