文章编号:1671-4229(2025)02-0074-13

幸存者平均因果效应的尖锐边界研究

肖恒毅，陈爱贤
(广州大学 经济与统计学院，广东 广州　510006)

摘　要：幸存者平均因果效应(Survivor Average Causal Effect，SACE)可以用来度量任何处理下都能存活的受试者接受不同处理的影响差异,是因果推断中的一个重要研究方向。由于处理组和对照组中总是存活的受试者样本不能直接观测,SACE 通常是不可识别的,只能得到 SACE 的边界。已有文献中 SACE 尖锐边界的主流求解方法依赖于多参数线性规划,通过枚举对偶问题的约束多边形的所有顶点来产生封闭形式的解。如果单调性和随机占优等条件不成立,则无法采用枚举法求解该多参数线性规划问题。文章基于主分层框架考虑了“死亡截断”、稳定个体处理效应和可忽略性假设下 SACE 的尖锐边界问题,其中,优化问题的求解是基于一阶 KKT(Kraush-Kuhn-Tucker)条件所对应的多项式方程组。实证选取美国国家支持工作示范项目(National Supported Work Demonstration,NSW)中的 Lalonde 数据集,计算了“永远幸存者”(always-survivor)在完整协变量情形下的 SACE 尖锐边界。

关键词：幸存者平均因果效应；多参数线性规划；死亡截断；一阶 KKT 条件；尖锐边界
中图分类号：O212　　**文献标志码**：A

Sharp bounds study on survivor average causal effects

XIAO Heng-yi，*CHEN Ai-xian*
(School of Economics and Statistics，Guangzhou University，Guangzhou 510006，China)

Abstract：The survivor average causal effect (SACE) can be used to measure the differential impact of receiving different treatments on subjects who would survive under any treatment condition. It represents an important area of research in causal inference. Since the samples of subjects who would always survive in both the treatment and control groups cannot be directly observed，SACE is typically unidentifiable，allowing only for the estimation of its bounds. In the literature，mainstream methods for deriving sharp bounds on SACE rely on multi-parameter linear programming，generating closed-form solutions by enumerating all vertices of the constraint polyhedron in the dual problem. If monotonicity and stochastic dominance conditions are not satisfied，the enumeration method cannot be used to solve this multi-parameter linear programming problem. This paper addresses the sharp bounds of SACE within the principal stratification framework，considering “truncated by death”，stable unit treatment effect，and ignorability assumptions. The optimization problem is solved based on polynomial equations derived from first-order Karush-Kuhn-Tucker (KKT) conditions. The LaLonde dataset from the National Supported Work Demonstration (NSW) project in the United States was empirically selected to calculate the sharp bounds of the SACE under complete covariates for the “always-survivor”.

Key words：survivor average causal effect；multi-parameter linear programming；truncated by death；first-order KKT conditions；sharp bounds

收稿日期：2024－12－11；**修回日期**：2025－01－17

作者简介：肖恒毅(1999—)，男，硕士研究生. E-mail：xiaohengyi@e.gzhu.edu.cn

引文格式：肖恒毅，陈爱贤. 幸存者平均因果效应的尖锐边界研究[J]. 广州大学学报(自然科学版)，2025，24(2)：74-86.

在数据分析中,因果关系的探索是一个复杂且重要的问题。确定因果关系的最常用方法是进行对照实验研究。当实验研究面临一些限制时,也可以借助观察研究和统计建模等方法来推断因果关系。随着现代统计学的发展,因果关系可用数学语言准确描述,并被统计学家归纳为“因果推断”。因果推断经历了一段时间的低谷后,在近 40 年的发展中有了较为明显的进展,尤其是最近 20 年,它在得到广泛认可的同时,吸引了众多学者展开深入研究,成为统计学、经济学、医学等多个领域的重要课题。其中,平均因果效应(Average Causal Effect, ACE)是指所有个体接受和不接受处理的平均结果差异,幸存者平均因果效应(SACE)是在任何处理下都能存活的亚群平均因果效应。随着线性规划、机器学习等技术的蓬勃发展,从可观测数据来识别和估计 ACE 及 SACE 已成为不少学者的研究方向。对于因果效应的研究不仅揭示了事物间的关系,还在政策评估、药品研发等实际应用方面起着至关重要的作用,这也印证了因果推断是现代统计学中的一大重要思想。

因果推断中最常用的模型之一是 Rubin[1] 提出的潜在结果模型(Rubin Causal Model, RCM)。在 RCM 模型的理论框架下,个体在不同处理下会产生对应的潜在结果。而在实际中,每个个体只能接受一种处理,其余处理的潜在结果无法被观测到。由于潜在结果的缺失,个体的因果作用是不可识别的。为了解决这一问题,研究者们提出了一系列尝试,包括随机对照试验、倾向得分匹配、工具变量法等。随机对照试验是一种最为广泛应用的方法,它通过随机化来消除混杂因素的影响。考虑到理想的随机化难以实现,研究者们又针对可观测数据提出因果推断方法,如双重稳健估计、合成控制法等。

在实验中,经常会发生“死亡截断”问题,即部分受试者在观测到结果前死亡。“死亡截断”问题由 Frangakis 等[2] 首先提出解决方案,即主分层法,他们选择在处理和对照下都可观测到的“永远幸存者”计算因果效应,并定义了 SACE。SACE 在弱假设下通常是不可识别的,只能推断其边界。弱假设包括稳定个体处理效应假设(Stable Unit Treatment Value Assumption, SUTVA)、可忽略性(ignorability assumption)等,还可以通过附加假设获得更紧密的尖锐边界。

在“死亡截断”的背景下,Zhang 等[3-4] 基于主分层法,开创性地为“永远幸存者”推导了 SACE 的大样本边界,其边界涉及主分层的极值优化问题,并将结果应用于培训计划对“永远幸存者”收入的影响评估;Imai[5] 将“死亡截断”问题视为受污染数据,根据 Horowitz 等[6] 在破损数据以及受污染数据下的识别结果,计算出无假设和附加假设下的 SACE 尖锐边界;Freiman 等[7] 以 Zhang 等[3] 的边界为基础,在弱可忽略假设下引入二元协变量,尽管不能有效缩短边界,但提高了结果的可推广性。因果效应的识别性和估计方法也包括参数模型,Ding 等[8] 在半参数或非参数模型的假设下,通过协变量识别感兴趣的因果参数,其中,协变量不需要满足 Zhang 等提出的混合正态分布;以 Ding 等的研究工作为前提,Wang 等[9] 提出替代变量作为解决途径,在不引入协变量的条件下识别 SACE,同时,还进一步考虑了附加假设失效的情况。另外,因果推断中经常会假定结果变量是离散有限或连续有界的;Balke 等[10] 在边界求解中提出 Balke-Pearl 多参数线性规划法,计算离散有限结果的 ACE 尖锐边界;Beresteanu 等[11] 运用随机集理论,简化并扩展了 Balke 等以及 Manski[12] 的部分识别结果;Shan 等[13] 应用 Balke-Pearl 线性规划方法,给定单调性假设,计算了二元结果变量下的 SACE 边界。

尽管 Balke-Pearl 多参数线性规划法能计算因果效应的边界,但有时会因为条件概率间的关系导致极为复杂的约束,从而出现难以直接应用求解的情形。Gabriel 等[14] 提出两种不同的代数方法,在二元处理变量和结果变量下详细推导了 ACE 非参数边界,但局限于所研究的 8 类因果图。Chen 等[15] 将感兴趣的结果放宽至离散或连续情形,不假定因果图,提出了不同假设下 3 个主分层的 ACE 边界推导框架,并拓展至依从者平均因果效应(Complier Average Causal Effects,CACE)的问题。但在约束进一步复杂时,所推导的框架可能会因为可行解空间的非凸性而难以使用。

SACE 的边界求解也可以转化为多参数线性规划问题。多参数线性规划(Multi-parametric Linear Programming,mp-LP)的概念首次由 Gal 等[16] 提出,奠定了多参数线性规划问题的基本框架。

在二元结果变量下,多参数混合整数线性规划(Multi-parametric Mixed Integer Linear Programming,mp-MILP)问题由 Acevedo 等[17]、Dua 等[18]及 Li 等[19]进行了探讨,Dua 等将 mp-MILP 分解为 mp-LP 和 MILP 问题,其方法在求解速度上最具优势。在多参数线性规划进一步复杂化之后,为了系统地处理全局不确定性的情况,Charitopoulos 等[20-21]针对一般形式的 mp-LP 问题,基于一阶 KKT 条件和 Groebner Bases 理论提出新算法,保证了 mp-LP 问题的最优解以及关键区域是精确且完整的。同时,考虑到 mp-LP 问题中的等式约束,将该算法延伸至对数导致的非线性情形。在已有框架上,Charitopoulos 等[22]还研究了 mp-MILP 问题,通过添加整数性约束提高算法的适用度。

本文在稳定个体处理效应假设和条件可忽略性的假设下,基于主分层框架,将 SACE 边界问题转化为多参数线性规划问题,并求解一阶 KKT 条件对应的多项式方程组,从而得到 SACE 的尖锐边界。计算结果可以推广到其他分层,并允许模型中有离散或者连续的协变量。本文还通过丰富的模拟验证了所提方法的可行性。

1 幸存者平均因果效应

在分析感兴趣的结果受到某项指标影响时,经常会发生受试者在研究人员测量结果前已经死亡的情形,即"死亡截断"。当实验研究或实际应用中遭遇受试者死亡的状况时,相应的因果效应即为 SACE,估计 SACE 是非常具有挑战性的。同时,因为仅在可观测到的幸存者中对处理组和对照组进行比较,其结果不能准确反映整个受试群体的实际情况,从而导致选择性偏差。为了解决"死亡截断"引发的偏差问题,本文以 Frangakis 等提出的主分层框架为基础,研究 SACE 的尖锐边界。

1.1 符号定义与基本假设

对于受试者个体 k,假定处理变量 Z_k 是二元的,$Z_k=1$ 表示个体 k 分配到处理组,$Z_k=0$ 时表示个体 k 分配到对照组。令 Y_k 为感兴趣的二元结果变量,$Y_k(1)$、$Y_k(0)$分别表示个体 k 在处理 $Z_k=1$ 及 $Z_k=0$ 下的潜在结果。令 S_k 为生存状态,$S_k(z)$为个体 k 在 $Z_k=z$ 下的潜在生存状态,$S_k(z)=1$ 表示个体 k 存活,$S_k(z)=0$ 表示其死亡。由于仅可观测到存活受试者的结果,因此,仅当 $S_k(z)=1$ 时,$Y_k(z)$才有定义,否则会遭遇"死亡截断"问题,感兴趣的结果此时是无法定义的。另外,用 C_k 表示协变量,该协变量可以包括离散型和连续型随机变量。特别地,当连续型随机变量中 0 占比较大时,转化为 0-1 变量,用 1 表示不为 0 的数据。

在 Frangakis 等提出的主分层框架中,整个受试者群体划分为 4 个主分层。用 G_k 表示生存类型,其中,L 表示存活,D 表示死亡,每种类型简记为 $g_j(j=0,1,2,3)$。表 1 反映了 G_k 和潜在二元生存状态$\{S_k(1),S_k(0)\}$的一一对应关系,可以理解为 G_k 是$\{S_k(1),S_k(0)\}$的简要表达。

表 1 受试者生存类型

Table 1 Survival types of subjects

$S_k(1)$	$S_k(0)$	G_k	生存类型	生存类型描述
1	1	g_0	永远幸存者(LL)	无论接受何种处理,都会存活
1	0	g_1	处理受益者(LD)	接受治疗时存活,对照时死亡
0	1	g_2	处理受害者(DL)	接受对照时存活,治疗时死亡
0	0	g_3	注定死亡者(DD)	无论接受何种处理,都会死亡

根据 Pearl[23]提出的一致性定理,对于个体 k,可观测的生存状态 S_k 满足等式 $S_k=Z_k\,S_k(1)+(1-Z_k)S_k(0)$。当且仅当 $S_k=1$ 时,感兴趣结果 Y_k 满足等式 $Y_k=Z_k\,Y_k(1)+(1-Z_k)Y_k(0)$。若希望从可观测的数据中计算某个受试群体的因果效应,可给出简单直观的条件期望:$E\{Y_k(1)-Y_k(0)\mid S_k=1\}=E\{Y_k(1)\mid S_k=1\}-E\{Y_k(0)\mid S_k=1\}$。但从表 1 可以注意到,$S_k(1)=1$ 混合了"永远幸存者"和"处理受益者",$S_k(0)=1$ 则混合了"永远幸存者"和"处理受害者",因此,$E\{Y_k(1)-Y_k(0)\mid S_k=1\}$的计算结果在因果上是没有意义的,因为可观测到的受试者很可能不是来自同一

个分层。为应对该问题,在每个主分层中定义 ACE:

$$\Delta_{g_j}=E\{Y_k(1)-Y_k(0)\mid G_k=g_j\}。\quad(1)$$

当 $G_k=g_0$ 时,式(1)中的 Δ_{g_0}被称为 SACE,即该分层内处理组与对照组感兴趣结果的均值差,这也是本文首要研究的因果效应。当 $G_k=g_3$ 时,由于“注定死亡者”的性质,受试者无论接受任何处理都会死亡,缺乏可观测的结果,因此,$\Delta_{g_3}=E\{Y_k(1)-Y_k(0)\mid G_k=g_3\}$不作进一步研究。并且在缺乏强有力的假设条件下,主分层 $G_k=g_j(j=0,1,2)$不能被直接观测到,难以给出各分层因果效应的点估计。本文在弱假设条件下推导因果效应的尖锐边界,假设条件如下:

(1)稳定个体处理效应假设:$S_k(z)\perp Z_l,Y_k(z)\perp Z_l,\forall k\neq l,z=0,1$。该假设表示对于任意的受试者个体 k、生存状态 S_k 和感兴趣结果 Y_k 仅取决于自身分配的处理 Z_k,其他个体接受任何处理都不会带来影响。而且每个个体只有一种处理方式,不存在多个版本[23]。若该假设仅在给定协变量 C_k 下才成立,则生存状态 S_k 和结果 Y_k 的联合分布与分配处理 Z_k 在协变量 C_k 给定时相互独立。

(2)条件可忽略性假设:$Z_k\perp\{S_k(1),S_k(0),Y_k(1),Y_k(0)\}\mid G_k,C_k$,其中,$G_k=\{g_0,g_1,g_2,g_3\}$。在该假设下,若给定生存类型 G_k 和协变量 C_k,对于任意的个体 k,分配的处理 Z_k 与潜在生存状态 $S_k(z)$和潜在结果 $Y_k(z)$是相互独立的。

1.2　SACE 的边界求解问题

在基本假设都成立时,给定协变量 C_k,定义主分层占比 $\pi_{j\cdot c}=P(G_k=g_j\mid C_k=c)$,不同处理下的存活及死亡占比 $P_{s\cdot zc}=P(S_k=s\mid Z_k=z,C_k=c)$。根据表 1 和基本假设 2,可以推出 $\pi_{j\cdot c}$和 $P_{s\cdot zc}$隐含的等式关系,由方程组(2)表示:

$$\begin{cases}\pi_{0\cdot c}+\pi_{1\cdot c}=P_{1\cdot 1c},\\ \pi_{0\cdot c}+\pi_{2\cdot c}=P_{1\cdot 0c},\\ \pi_{2\cdot c}+\pi_{3\cdot c}=P_{0\cdot 1c},\\ \pi_{1\cdot c}+\pi_{3\cdot c}=P_{0\cdot 0c}。\end{cases}\quad(2)$$

此时,若想知道主分层占比 $\pi_{j\cdot c}$的取值范围,以 $\pi_{0\cdot c}$为例,可以看作是求解线性规划 min or max $\pi_{0\cdot c}$,其中,方程组(2)和 $\pi_{j\cdot c}\geqslant 0\ (j=0,1,2,3)$是线性规划的约束。对于其他的主分层,使用相同的线性规划法求解,计算结果整理为

$$\max\{0,P_{1\cdot 1c}-P_{0\cdot 0c}\}\leqslant\pi_{0\cdot c}\leqslant\min\{P_{1\cdot 1c},P_{1\cdot 0c}\},\quad(3)$$

$$\max\{0,P_{1\cdot 1c}-P_{1\cdot 0c}\}\leqslant\pi_{1\cdot c}\leqslant\min\{P_{1\cdot 1c},P_{0\cdot 0c}\},\quad(4)$$

$$\max\{0,P_{1\cdot 0c}-P_{1\cdot 1c}\}\leqslant\pi_{2\cdot c}\leqslant\min\{P_{0\cdot 1c},P_{1\cdot 0c}\},\quad(5)$$

$$\max\{0,P_{0\cdot 1c}-P_{1\cdot 0c}\}\leqslant\pi_{3\cdot c}\leqslant\min\{P_{0\cdot 1c},P_{0\cdot 0c}\}。\quad(6)$$

由式(3)~式(6)的结果可知,即便给定 2 个基本假设和协变量 C_k,仍无法得出 $\pi_{j\cdot c}$的点估计,但可以得到分层占比的上下边界。记存在协变量的幸存者平均因果效应为 SACE_c,对 SACE_c 的点估计进行计算,定义任意分层感兴趣结果的条件均值 $\mu_{zjc}=E(Y_k\mid Z_k=z,G_k=g_j,C_k=c)$,根据式(1)和条件可忽略性假设,得到 SACE_c 为

$$\mathrm{SACE}_c=\mu_{10c}-\mu_{00c}=E(Y_k\mid Z_k=1,G_k=g_0,C_k=c)-E(Y_k\mid Z_k=0,G_k=g_0,C_k=c),\quad(7)$$

其中,由于主分层的不可观测性,μ_{10c}和 μ_{00c}的点估计是无法识别的,转为考虑 SACE_c 的边界问题。为了得到 SACE_c 的尖锐边界,用 T_k 表示结果 Y_k 在处理 Z_k 下的潜在响应类型,具体为

$$T_k=\begin{cases}t_0:\{Y_k(1)=1,Y_k(0)=1\},\\ t_1:\{Y_k(1)=1,Y_k(0)=0\},\\ t_2:\{Y_k(1)=0,Y_k(0)=1\},\\ t_3:\{Y_k(1)=0,Y_k(0)=0\}。\end{cases}$$

G_k 表示生存状态 S_k 在处理 Z_k 下的生存类型,详见表 1。定义 T_k 和 G_k 的条件联合分布 $F_{ij\cdot c}=P(T_k=t_i,G_k=g_j\mid C_k=c),i,j=0,1,2,3$。此时,$\mathrm{SACE}_c$ 可以改写为

$$\mathrm{SACE}_c=\frac{F_{20\cdot c}-F_{10\cdot c}}{F_{00\cdot c}+F_{10\cdot c}+F_{20\cdot c}+F_{30\cdot c}}。\quad(8)$$

基于第 2 条基本假设,此处定义包含可观测部分的条件联合分布:$F_{is\cdot zc}=P(Y_k=i,S_k=s\mid Z_k=z,C_k=c),i,s,z=0,1$。可以发现,$F_{is\cdot zc}$和 $F_{ij\cdot c}$存在线性关系,如 $F_{01\cdot 0c}=F_{10\cdot c}+F_{30\cdot c}+F_{12\cdot c}+F_{32\cdot c}$。为方便描述,记 $F_{0\cdot 1c}=F_{01\cdot c}+F_{11\cdot c}$,$F_{3\cdot 1c}=F_{21\cdot c}+F_{31\cdot c}$,$F_{1\cdot 2c}=F_{02\cdot c}+F_{22\cdot c}$,$F_{2\cdot 2c}=F_{12\cdot c}+F_{32\cdot c}$,$F_{\cdot 3c}=F_{03\cdot c}+F_{13\cdot c}+F_{23\cdot c}+F_{33\cdot c}$。求解 SACE_c 的边界可以转化为求解优化问题,即 min or max

$\dfrac{F_{20\cdot c}-F_{10\cdot c}}{F_{00\cdot c}+F_{10\cdot c}+F_{20\cdot c}+F_{30\cdot c}}$,约束如下:

$$\begin{cases} F_{11\cdot 0c} = F_{00\cdot c} + F_{20\cdot c} + F_{1\cdot 2c}, \\ F_{01\cdot 0c} = F_{10\cdot c} + F_{30\cdot c} + F_{2\cdot 2c}, \\ F_{11\cdot 1c} = F_{00\cdot c} + F_{10\cdot c} + F_{0\cdot 1c}, \\ F_{01\cdot 1c} = F_{20\cdot c} + F_{30\cdot c} + F_{3\cdot 1c}, \\ 1 = \sum_{i=0}^{3} F_{i0\cdot c} + F_{0\cdot 1c} + F_{3\cdot 1c} + F_{1\cdot 2c} + F_{2\cdot 2c} + F_{\cdot 3c}, \\ F_{i0\cdot c} \geqslant 0 (i = 0,1,2,3), \\ F_{0\cdot 1c} \geqslant 0, F_{3\cdot 1c} \geqslant 0, F_{1\cdot 2c} \geqslant 0, F_{2\cdot 2c} \geqslant 0, F_{\cdot 3c} \geqslant 0。 \end{cases} \tag{9}$$

根据 Schaible 等[24]提出的定理以及 Chen 等[15]的推导框架,进一步转化为符号线性规划问题,即 min or max $(f_{20\cdot c}^{(0)} - f_{10\cdot c}^{(0)})$,约束如下:

$$\begin{cases} f_{00\cdot c}^{(0)} + f_{20\cdot c}^{(0)} + f_{1\cdot 2c}^{(0)} = F_{11\cdot 0c}/\pi_{0\cdot c}, \\ f_{10\cdot c}^{(0)} + f_{30\cdot c}^{(0)} + f_{2\cdot 2c}^{(0)} = F_{01\cdot 0c}/\pi_{0\cdot c}, \\ f_{00\cdot c}^{(0)} + f_{10\cdot c}^{(0)} + f_{0\cdot 1c}^{(0)} = F_{11\cdot 1c}/\pi_{0\cdot c}, \\ f_{20\cdot c}^{(0)} + f_{30\cdot c}^{(0)} + f_{3\cdot 1c}^{(0)} = F_{01\cdot 1c}/\pi_{0\cdot c}, \\ f_{00\cdot c}^{(0)} + f_{10\cdot c}^{(0)} + f_{20\cdot c}^{(0)} + f_{30\cdot c}^{(0)} = 1, \\ \sum_{i=0}^{3} f_{i0\cdot c}^{(0)} + f_{0\cdot 1c}^{(0)} + f_{3\cdot 1c}^{(0)} + f_{1\cdot 2c}^{(0)} + f_{2\cdot 2c}^{(0)} + f_{\cdot 3c}^{(0)} = 1/\pi_{0\cdot c}, \\ f_{i0\cdot c}^{(0)} \geqslant 0 \ (i = 0,1,2,3), \\ f_{0\cdot 1c}^{(0)} \geqslant 0, f_{3\cdot 1c}^{(0)} \geqslant 0, f_{1\cdot 2c}^{(0)} \geqslant 0, f_{2\cdot 2c}^{(0)} \geqslant 0, f_{\cdot 3c}^{(0)} \geqslant 0, \end{cases} \tag{10}$$

其中,$t^{(0)} = 1/(F_{00\cdot c} + F_{10\cdot c} + F_{20\cdot c} + F_{30\cdot c}) = 1/\pi_{0\cdot c}$,则有 $f_{ij\cdot c}^{(0)} = t^{(0)} F_{ij\cdot c}$,$f_{0\cdot 1c}^{(0)} = t^{(0)} F_{0\cdot 1c}$,$f_{3\cdot 1c}^{(0)} = t^{(0)} F_{3\cdot 1c}$,$f_{1\cdot 2c}^{(0)} = t^{(0)} F_{1\cdot 2c}$,$f_{2\cdot 2c}^{(0)} = t^{(0)} F_{2\cdot 2c}$,$f_{\cdot 3c}^{(0)} = t^{(0)} F_{\cdot 3c}$。在约束(10)当中,符号变量为 $F_{11\cdot 0c}$、$F_{01\cdot 0c}$、$F_{11\cdot 1c}$、$F_{01\cdot 1c}$,可以清晰地看出,约束等式的右侧有符号变量,计算复杂度会明显提升,难以直接应用传统的 Balke-Pearl 线性规划法求解。本文将结合多参数线性规划的求解方法,通过符号运算工具得到 SACE_c 尖锐边界的精确表达式。

2 基于多参数线性规划的 SACE_c 边界研究

2.1 多参数线性规划与 KKT 条件方程组

本文首先介绍多参数线性规划的基本形式。为保证问题的一般性,选取全局不确定性下的线性规划问题,即目标函数系数及约束的左侧和右侧均存在不确定参数,具体形式为

$$\text{mp}-\text{LP} = \begin{cases} z(\boldsymbol{\theta}) = \min_{\boldsymbol{x}} \boldsymbol{c}^{\mathrm{T}}(\boldsymbol{\theta})\boldsymbol{x}, \\ \text{subject to: } \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{x} \leqslant \boldsymbol{b} + \boldsymbol{F}(\boldsymbol{\theta}), \\ \boldsymbol{h}(\boldsymbol{x},\boldsymbol{\theta}) = 0, \\ \boldsymbol{x} \in X \triangleq \{x \in R^{n_x} \mid x_i^{\min} \leqslant x_i \leqslant x_i^{\max}, i \leqslant n_x\}, \\ \theta \in \Theta_i \{\theta \in R^{n_\theta} \mid \theta_j^{\min} \leqslant \theta_j \leqslant \theta_j^{\max}, j \leqslant n_\theta\}, \end{cases} \tag{11}$$

其中,$\boldsymbol{b} = [b_1, \cdots, b_m]^{\mathrm{T}}$ 为 m 维常数向量,$\boldsymbol{c}(\boldsymbol{\theta}) = [c_1, \cdots, c_{n_x}]^{\mathrm{T}}$ 为 n_x 维目标函数系数向量,$\boldsymbol{A}(\boldsymbol{\theta}) = [a_{k,i}(\theta)]_{m \times n_x}$ 为约束左侧系数矩阵,$\boldsymbol{F}(\boldsymbol{\theta}) = [f_{k,j}(\boldsymbol{\theta})]_{m \times n_\theta}$ 为约束右侧参数矩阵,$\boldsymbol{h}(\boldsymbol{x},\boldsymbol{\theta}) = [h_1, \cdots, h_n]^{\mathrm{T}}$ 为 n 维等式约束。即在问题(11)中,线性规划问题的约束用矩阵形式表达。记不等式约束的拉格朗日乘子为 λ_k,等式约束的拉格朗日乘子为 μ_l,对应向量分别为 $\boldsymbol{\lambda}$ 和 $\boldsymbol{\mu}$,则拉格朗日函数为

$$L(\boldsymbol{x},\boldsymbol{\theta},\boldsymbol{\mu},\lambda) = \boldsymbol{c}^{\mathrm{T}}(\boldsymbol{\theta})\boldsymbol{x} + \sum_{k=1}^{m} \lambda_k \left\{ \sum_{i=1}^{n_x} a_{k,i}(\boldsymbol{\theta}) x_i - b_k - \sum_{j=1}^{n_\theta} f_{k,j}(\boldsymbol{\theta}) \right\} + \sum_{l=1}^{n} \mu_l \left\{ \sum_{i=1}^{n_x} c_{l,i}(\boldsymbol{\theta}) - \sum_{j=1}^{n_\theta} g_{l,j}(\boldsymbol{\theta}) \right\}。$$

问题(11)的一阶 KKT 条件如下:

$$\text{一阶 KKT 条件} = \begin{cases} h_l(\boldsymbol{x},\boldsymbol{\theta}) = \sum_{i=1}^{n_x} c_{l,i}(\boldsymbol{\theta}) - \sum_{j=1}^{n_\theta} g_{l,j}(\boldsymbol{\theta}) = 0, l = 1, \cdots n, \\ \lambda_k \{ \sum_{i=1}^{n_x} a_{k,i}(\boldsymbol{\theta}) x_i - b_k - \sum_{j=1}^{n_\theta} f_{k,j}(\boldsymbol{\theta}) \} = 0, k = 1, \cdots m, \\ \nabla_x L(\boldsymbol{x},\boldsymbol{\theta},\boldsymbol{\mu},\boldsymbol{\lambda}) = 0。 \end{cases}$$

一阶 KKT 条件由原约束乘以拉格朗日乘子和拉格朗日函数梯度组成,$\nabla_x L(\boldsymbol{x},\boldsymbol{\theta},\boldsymbol{\mu},\boldsymbol{\lambda}) = 0$ 表示拉格朗日函数梯度为0。问题(11)满足一阶 KKT 条件的解为候选解[20],但不一定是参数的全局或局部最优解,因此,可以给定两个筛选条件,其中,可行性条件为 $r_k = \sum_{i=1}^{n_x} a_{k,i}(\boldsymbol{\theta}) - b_k - \sum_{j=1}^{n_\theta} f_{k,j}(\boldsymbol{\theta}) \leqslant 0, k = 1, \cdots m$,对偶性条件为 $\lambda_k \geqslant 0, k = 1, \cdots, m$。

2.2 SACE_c 边界研究及推广

令 $LB_0(c)$、$UB_0(c)$ 分别为 SACE_c 的上界和下界,其中,$LB_0(c)$ 可以通过 $\min(f_{20\cdot c}^{(0)} - f_{10\cdot c}^{(0)})$ 满足约束(3)和约束(9)得到,$UB_0(c)$ 则可以通过 $\max(f_{20\cdot c}^{(0)} - f_{10\cdot c}^{(0)})$ 满足约束(3)和约束(9)得到。为了

求解 $LB_0(c)$ 和 $UB_0(c)$,通过线性变换把约束(9)写成约束(10)的形式。以 $\min(f^{(0)}_{20\cdot c}-f^{(0)}_{10\cdot c})$ 为目标函数,可以将约束(10)转化成多参数线性规划问题(11),其中,

$$\boldsymbol{x}=(f^{(0)}_{00\cdot c},f^{(0)}_{10\cdot c},f^{(0)}_{20\cdot c},f^{(0)}_{30\cdot c},f^{(0)}_{0\cdot 1c},f^{(0)}_{3\cdot 1c},f^{(0)}_{1\cdot 2c},f^{(0)}_{2\cdot 2c},f^{(0)}_{\cdot 3c},\pi_{0\cdot c})^{\mathrm{T}},$$

$$\boldsymbol{c}(\boldsymbol{\theta})=(0,-1,1,0,0,0,0,0,0,0)^{\mathrm{T}},$$

$$\boldsymbol{\theta}=(F_{11\cdot 0c},F_{01\cdot 0c},F_{11\cdot 1c},F_{01\cdot 1c},P_{1\cdot 1c},P_{1\cdot 0c},P_{0\cdot 0c})^{\mathrm{T}}。$$

$\boldsymbol{b}$ 和 $\boldsymbol{F}(\boldsymbol{\theta})$ 都是维数为 13 的 0 向量,$\boldsymbol{h}(\boldsymbol{x},\boldsymbol{\theta})$ 为约束(10)中的前 6 行等式约束,系数矩阵 $\boldsymbol{A}(\boldsymbol{\theta})$ 详见附录一。该线性规划问题的一阶 KKT 条件对应方程组包括了 10 个未知数和 29 个多项式方程。基于符号运算软件 Mathematica 求解该多项式方程组可得到 221 个可行解,根据可行性条件和对偶性条件可以筛选出 14 个最优解。该问题的多项式方程组详见附录二,最优解示例详见附录三。对于优化问题 min or max$(f^{(0)}_{20\cdot c}-f^{(0)}_{10\cdot c})$ 求解,可以得到:

$$LB_0(c)=\max\left\{-1,-\frac{F_{01\cdot 0c}+F_{11\cdot 1c}}{P_{1\cdot 1c}-P_{0\cdot 0c}}+1,-\frac{F_{01\cdot 0c}}{P_{1\cdot 1c}-P_{0\cdot 0c}},-\frac{F_{11\cdot 1c}}{P_{1\cdot 1c}-P_{0\cdot 0c}}\right\},$$

$$UB_0(c)=\min\left\{1,\frac{F_{01\cdot 1c}+F_{11\cdot 0c}}{P_{1\cdot 1c}-P_{0\cdot 0c}}-1,\frac{F_{01\cdot 1c}}{P_{1\cdot 1c}-P_{0\cdot 0c}},\frac{F_{11\cdot 0c}}{P_{1\cdot 1c}-P_{0\cdot 0c}}\right\}。$$

从而

$$LB_0(c)\leqslant \mathrm{SACE}_c\leqslant UB_0(c)。\tag{12}$$

为了推导 SACE 的边界,根据全期望公式可得

$$\mathrm{SACE}=E(\pi_{0\cdot c}\mathrm{SACE}_c)/E(\pi_{0\cdot c})=E\{(P_{1\cdot 1c}-P_{0\cdot 0c})\mathrm{SACE}_c\}/E(P_{1\cdot 1c}-P_{0\cdot 0c})。$$

令 LB_0、UB_0 分别为 SACE 的上界和下界,则 SACE 有以下范围:

$$LB_0\leqslant \mathrm{SACE}\leqslant UB_0,\tag{13}$$

其中,$LB_0=\dfrac{E\{(P_{1\cdot 1c}-P_{0\cdot 0c})LB_0(c)\}}{E(P_{1\cdot 1c}-P_{0\cdot 0c})}$,$UB_0=\dfrac{E\{(P_{1\cdot 1c}-P_{0\cdot 0c})UB_0(c)\}}{E(P4_{1\cdot 1c}-P_{0\cdot 0c})}$。

3　模拟研究

3.1　条件概率估计

对于 SACE_c 边界表达式中的符号变量,根据 Li 等[25]提出的方法,在该部分给出核估计量。假设可观测到独立同分布的多元样本,对于任意个体 k,有 (X_k,Y_k,Z_k,S_k),$k=1,\cdots,n$。令 $I_k(y)=I(Y_k=y)$,$I_k(z)=I(Z_k=z)$,其中,$I(\cdot)$ 为示性函数,y、z 取值为 0、1。用 $\boldsymbol{C}^{\boldsymbol{d}}_{\boldsymbol{k}}$ 表示一个 $r\times 1$ 的离散型协变量,并且用 $\boldsymbol{C}^{\boldsymbol{c}}_{\boldsymbol{k}}\in R^q$ 表示剩余的连续型协变量。假设 C^d_{ks} 表示 $\boldsymbol{C}^{\boldsymbol{d}}_{\boldsymbol{k}}$ 的第 s 个分量,取值 $c_s\geqslant 2$,即 $C^d_{ks}\in\{0,1,\cdots,c_s-1\}$,其中,有 $s=\{1,\cdots,r\}$,并且定义 $\boldsymbol{C}_{\boldsymbol{k}}=(\boldsymbol{C}^{\boldsymbol{d}}_{\boldsymbol{k}},\boldsymbol{C}^{\boldsymbol{c}}_{\boldsymbol{k}})$。令 $\boldsymbol{c}^{\boldsymbol{c}}=(c^c_1,\cdots,c^c_q)$,多元核函数定义如下:

$$W_h(\boldsymbol{c}^{\boldsymbol{c}},\boldsymbol{C}^{\boldsymbol{c}}_{\boldsymbol{k}})=\frac{1}{h_1\cdots h_q}W\left(\frac{\boldsymbol{c}^{\boldsymbol{c}}-\boldsymbol{C}^{\boldsymbol{c}}_{\boldsymbol{k}}}{\boldsymbol{h}}\right)\equiv\prod_{s=1}^{q}\frac{1}{h_s}w\left(\frac{c^c_s-C^c_{ks}}{h_s}\right),$$

其中,w 是一个对称、非负的单变量核函数,常见的核函数有 Uniform 核 $w(u)=I(|u|<1)$ 以及高斯核 $w(u)=\dfrac{1}{\sqrt{2\pi}}\exp(-\dfrac{u^2}{2})$ 等。窗宽 $\boldsymbol{h}=(h_1,\cdots,h_q)$,分量 h_s 是 c^c_s 的平滑参数且 $h_s>0$。对于 $\boldsymbol{c}^{\boldsymbol{d}}=(c^d_1,\cdots,c^d_r)$,定义如下多元函数:

$$L(\boldsymbol{c}^{\boldsymbol{d}},\boldsymbol{C}^{\boldsymbol{d}}_{\boldsymbol{k}},\boldsymbol{\lambda})=\prod_{s=1}^{r}\lambda_s^{N_{ks(c)}},$$

其中,$N_{ks(c)}=1(C^d_{ks}\neq c^d_s)$ 是一个示性函数,当 $C^d_{ks}\neq c^d_s$ 时等于 1,否则为 0。$\boldsymbol{\lambda}=(\lambda_1,\cdots,\lambda_r)$,分量 λ_s 是 c^d_s 的平滑参数且满足 $0\leqslant\lambda_s\leqslant 1$。注意,当 $\lambda_s=1$ 时,$L(\boldsymbol{c}^{\boldsymbol{d}},\boldsymbol{C}^{\boldsymbol{d}}_{\boldsymbol{k}},1)$ 与 (c^d_s,C^d_{ks}) 相互独立(因为 c^d_s 被平滑了)。因此,混合变量向量 $\boldsymbol{c}=(\boldsymbol{c}^{\boldsymbol{c}},\boldsymbol{c}^{\boldsymbol{d}})$ 的核函数可以定义为 $W_h(\cdot)$ 和 $L(\cdot)$ 的乘积,即 $K_\gamma(\boldsymbol{c},\boldsymbol{C}_{\boldsymbol{k}})=W_h(\boldsymbol{c}^{\boldsymbol{c}},\boldsymbol{C}^{\boldsymbol{c}}_{\boldsymbol{k}})L(\boldsymbol{c}^{\boldsymbol{d}},\boldsymbol{C}^{\boldsymbol{d}}_{\boldsymbol{k}},\boldsymbol{\lambda})$,$\boldsymbol{\gamma}=(\boldsymbol{h},\boldsymbol{\lambda})$。因此,$LB_0(c)$ 和 $UB_0(c)$ 中符号表达式的核估计量为

$$\hat{P}_{1\cdot zc}=\sum\nolimits_{k=1}^{n}S_kI_k(z)K_\gamma(\boldsymbol{c},\boldsymbol{C}_{\boldsymbol{k}})/\sum\nolimits_{k=1}^{n}I_k(z)K_\gamma(\boldsymbol{c},\boldsymbol{C}_{\boldsymbol{k}}),\tag{14}$$

$$\hat{F}_{11\cdot zc}=\sum\nolimits_{k=1}^{n}S_kI_k(y)I_k(z)K_\gamma(\boldsymbol{c},\boldsymbol{C}_{\boldsymbol{k}})/$$

$$\sum_{k=1}^{n} I_k(z) K_\gamma(\boldsymbol{c}, \boldsymbol{C}_k)。 \quad (15)$$

根据 $P_{s\cdot zc}$ 和 $F_{ks\cdot zc}$ 的定义,可推出:$\hat{P}_{0\cdot zc} = 1 - \hat{P}_{1\cdot zc}$, $\hat{F}_{01\cdot zc} = \hat{P}_{1\cdot zc} - \hat{F}_{11\cdot zc}$。Lee[26] 和 Huber 等[27] 在求解 ACE 尖锐边界时也使用了核估计法,Lee 还证明了估计量具有良好的一致性和渐近正态性。

本文主要使用 R 内部的 np 包进行数值模拟,通过模拟分布生成数据和使用 npcdistbw 函数计算窗宽 h_k,结合 $F_{ks\cdot zc}$ 定义和 npcdist 函数估计 $\hat{F}_{ks\cdot zc}$,其中,npcdist 函数适用于离散与连续混合的协变量。并且,使用蒙特卡罗算法近似计算 $\hat{P}_{s\cdot zc}$,从而得到 SACE_c 边界,再通过全期望公式计算出 SACE 边界。

3.2 模拟结果

本文对"永远幸存者"的 SACE 尖锐边界进行研究,验证方法的可行性。由于理论假定二元结果,因此,选择二项分布进行模拟。从该分布生成样本量为 1 000 的样本数据,以 500 次实验的平均值作为上下边界估计量。对于模拟中的参数:不考虑协变量时,定义 $\mu_{zj} = E(Y \mid Z = z, G = g_j)$, $\pi_j = P(G = g_j)$;考虑协变量时,定义 $\mu_{zjc} = E(Y \mid Z = z, G = g_j, C = c)$, $\pi_{jc} = P(G = g_j \mid C = c)$。

模型不含协变量情形:假定结果变量服从二项分布,有 $F(Y \mid Z = z, G = g_j) \sim B(n, \mu_{zj})$,其中,$n$ 为 1 000。假定第一组参数为 $\mu_{10} = 0.6$, $\mu_{11} = 0.8$, $\mu_{00} = 0.2$, $\mu_{02} = 0.5$;第二组参数为 $\mu_{10} = 0.6$, $\mu_{11} = 0.8$, $\mu_{00} = 0.5$、$\mu_{02} = 0.2$。

模型包含离散和连续协变量情形:假定结果变量服从二项分布,则有 $F(Y \mid Z = z, G = g_j, C = c) \sim B(n, \mu_{zjc})$,其中,$n$ 为 1 000。假定离散型协变量 C^d 服从概率为 0.4 的两点分布,即 $B(1, 0.4)$;连续型协变量 C^c 服从均值为 4、方差为 1 的正态分布,即 $N(4, 1)$。当 $C^d = 1$ 时,2 组参数分别为 $\mu_{101c} = 0.6$, $\mu_{111c} = 0.8$, $\mu_{001c} = 0.2$, $\mu_{021c} = 0.5$; $\mu_{101c} = 0.6$, $\mu_{111c} = 0.8$, $\mu_{001c} = 0.5$, $\mu_{021c} = 0.2$。当 $C^d = 0$ 时,2 组参数分别为 $\mu_{100c} = 0.6$, $\mu_{110c} = 0.8$, $\mu_{000c} = 0.2$, $\mu_{020c} = 0.5$; $\mu_{100c} = 0.6$, $\mu_{110c} = 0.8$, $\mu_{000c} = 0.5$, $\mu_{020c} = 0.2$。

给定 $\pi_0 = \pi_{0c} = 0.5$, $\pi_1 = \pi_{1c} = 0.3$, $\pi_2 = \pi_{2c} = 0.1$,表 2 及表 3 为上述两种情形的边界结果。

表 2 SACE 边界(无协变量)

Table 2 Bounds on the SACE (without covariates)

二项分布参数	总体边界	真实值	边界宽度	覆盖率/%
$\mu_{10} = 0.6, \mu_{11} = 0.8, \mu_{00} = 0.2, \mu_{02} = 0.5$	[-0.402, 1.160]	0.4	1.562	100
$\mu_{10} = 0.6, \mu_{11} = 0.8, \mu_{00} = 0.5, \mu_{02} = 0.2$	[-0.654, 0.763]	0.1	1.417	100

表 3 SACE 边界(含离散和连续协变量)

Table 3 Bounds on the SACE (with discrete and continuous covariates)

二项分布参数	总体边界	真实值	边界宽度	覆盖率/%
$\mu_{101c} = 0.6, \mu_{111c} = 0.8$	[-0.299, 1.050]	0.4	1.349	100
$\mu_{001c} = 0.2, \mu_{021c} = 0.5$				
$\mu_{100c} = 0.6, \mu_{110c} = 0.8$	[-0.343, 1.119]	0.4	1.452	100
$\mu_{000c} = 0.2, \mu_{020c} = 0.5$				
$\mu_{101c} = 0.6, \mu_{111c} = 0.8$	[-0.597, 0.738]	0.1	1.335	100
$\mu_{001c} = 0.5, \mu_{021c} = 0.2$				
$\mu_{100c} = 0.6, \mu_{110c} = 0.8$	[-0.597, 0.760]	0.1	1.357	100
$\mu_{000c} = 0.5, \mu_{020c} = 0.2$				

根据表 2 和表 3 可知,即使二项分布的参数上有一定差异,在各个情形下的边界都包含了真实值,说明本文推导的 SACE 尖锐边界的真实值覆盖率能达到 100%。这也表示使用一阶 KKT 条件对应的多项式方程组求解 SACE 尖锐边界是可行的,方程组的最优解可以有效识别 SACE,因此,能

推广到实证中。同时,协变量的引入在一定程度上改善了边界宽度,相较于不含协变量的情形,混合离散和连续协变量的 SACE 边界宽度分别平均缩短了 10.3% 和 5.0%。需要留意的是,边界跨越了零点,表明难以确定一项政策或技术可以给感兴趣的研究结果带来正向或负向因果效应,在现实中可能提供较为有限的信息。

同时,假定 $P_{1\cdot 0}=P_{1\cdot 0c}=0.6$, $\pi_1=\pi_{1c}=0.3$, $\pi_3=\pi_{3c}=0.1$, π_0和π_{0c}的范围均为[0.3, 0.5],图 1 和图 2 分别为上述两种情形下 SACE 边界伴随 π_0或 π_{0c}的变化趋势。

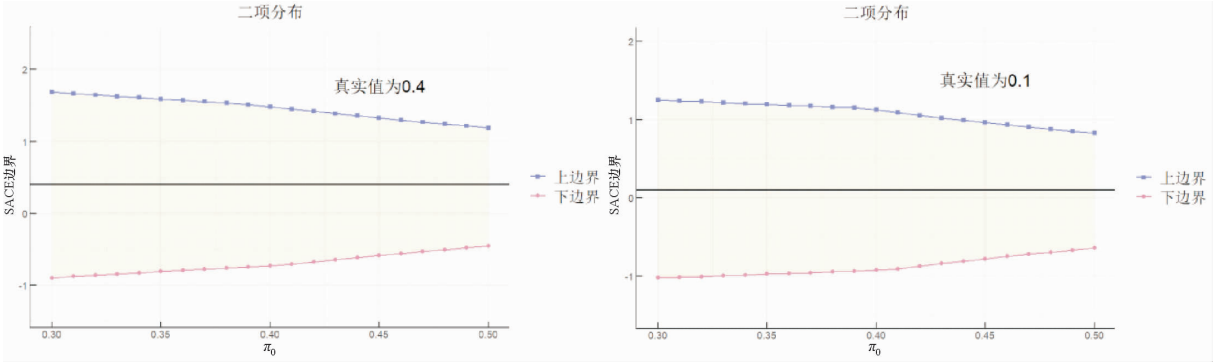

图 1　SACE 边界(无协变量)

Fig. 1　Bounds on the SACE (without covariates)

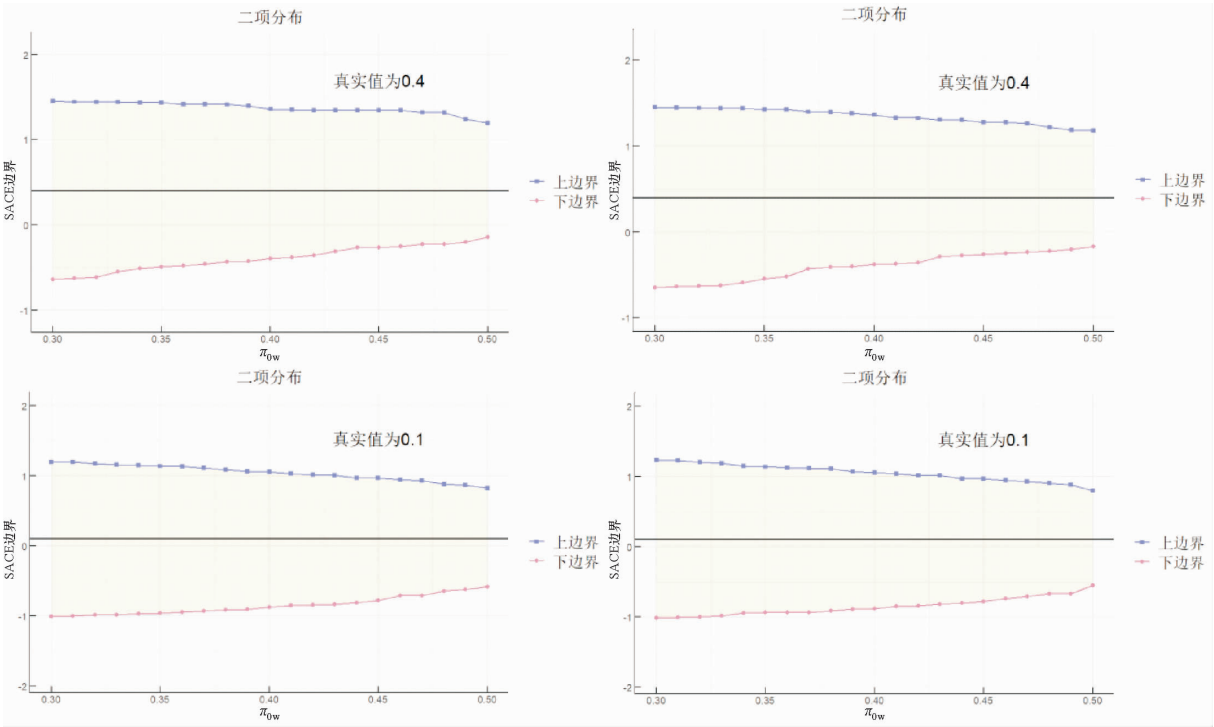

图 2　SACE 边界(含离散和连续协变量)

Fig. 2　Bounds on the SACE (with discrete and continuous covariates)

由图 1 可以清晰地看出,π_0在[0.3, 0.5]的范围中不断接近最大值时,SACE 边界逐渐向真实值靠近。不难发现,求解得到的 $LB_0(c)$及 $UB_0(c)$中的表达式在没有给定协变量的时候,是关于 π_0的单调函数,符合 SACE 边界的图像变化趋势。而根据图 2 可知,混合离散和连续协变量下的边界同样呈收敛趋势,符合 SACE 边界关于 π_{0c}具有单调性。

4 实证应用

4.1 数据集

实证部分使用的数据集来源于NSW项目,研究工作培训项目对于受试者工资的因果效应。该项目在20世纪70年代中期展开,人力示范研究公司在美国10处地点实施,接收了领取援助的女性、辍学的高中生及曾有案底的市民等群体作为受试者。该项目根据目标群体特点和居住地,保证分配到处理组的受试者在1976—1977年可以获得9~18个月的工作培训,对照组则需要自行解决就业问题,所有受试者在项目开始阶段及之后的每9个月记录一次真实收入。由于NSW项目可提供的就业资源有限,定期访谈调查会发现人数在逐渐下降,仅有65%的受试者参与了最后第27个月的访谈[28],未参加第27个月访谈的个体可看作遭遇了"死亡截断",他们的收入数据是不完整的。尽管该数据集跨越了半个世纪,但是在近几年的国际统计学和经济学期刊中,NSW项目的Lalonde和PSID数据集仍得到广泛使用[29-31],本文选取Lalonde数据集,其共有614名受试者,其中,185人随机分配到处理组中,剩余429人在对照组,并且该数据集的受试者均为男性。

为了评估培训项目对工资的影响,选择受试者在1978年的收入作为感兴趣的结果变量Y,数据集中简记为"re78",此处通过中位数将其划分为高收入和低收入,$Y=1$表示高收入,$Y=0$表示低收入。数据集中的"treat"为处理变量Z,其中,$Z=1$表示接受工作培训,$Z=0$表示没有接受培训。S在此项目中为生存变量,实际为1978年的就业状态,$S=1$表示受试者在1978年因成功就业而有所收入,$S=0$则表示受试者由于失业而收入为0。记C为可用协变量,包括受试者的婚姻状况、个人所属种族、是否取得高中学历、1974年和1975年的收入状况、1977年的年龄以及受教育时长,在数据集中它们分别记为"married""race""nodegree""re74""re75""age""educ",其中,"re74""re75""age"和"educ"为连续型变量,其余为离散型变量。Lalonde数据集中所有连续型变量的描述性统计结果在表4中展示。

表4 描述性统计结果

Table 4 Descriptive statistical results

变量简称	样本量	均值	中位数	标准差	最大值	最小值
age	614	27.363	25	9.881	55	16
educ	614	10.269	11	2.628	18	0
re74	614	4 557.547	1 042.330	6 477.964	30 540.070	0
re75	614	2 184.938	601.548	3 295.679	25 142.240	0
re78	614	6 792.834	4 759.019	7 470.731	60 307.930	0

4.2 实证结果

2.2小节推导的SACE尖锐边界在模拟中证明了可行性,将应用于NSW项目的Lalonde数据集中。经过整理,发现1975年收入为0的数据较多,1974年有小部分收入为0的情况,因此,将数据集中的"re75"转化为0-1变量,1代表收入大于0的数据,"re74"则保留原始数据。同时,"race"中含有3种分类,选取其中的"black"和"hispan"作为0-1变量,取1属于该种族,不属于则取0。经过初步处理后,现有5个离散协变量和3个连续协变量。采用式(14)及式(15)来估计式(13)中的未知项,从而得到培训受试者的平均因果效应边界。在表5中,给出了"永远幸存者"在完整协变量下的边界。

表5 SACE边界

Table 5 Bounds on the SACE

假定情形	总体边界	边界宽度
完整协变量	[-0.255, 0.268]	0.523

观察表5中完整协变量下的SACE边界,数值结果为[-0.255, 0.268],其边界宽度相较于模拟中使用2个协变量时进一步缩短,更加接近于点估计,这也说明了所用的协变量对于受试者在1978年的真实收入有切实影响,引入它们可以提

供更准确的估计。但是,即使完整协变量的 SACE 边界已经较窄,它仍包含了零点,即处理组与对照组的平均收入差值同时包含正负数,难以确定 NSW 项目对接受培训的受试者收入有正向或负向效应。因此,本文结果仍需要进一步改善,才可以让推导的因果效应尖锐边界给出更有效的估计,进而在实际应用中提供更有价值的信息。

5　结　论

本文考虑到实验中会经常遭遇部分受试者死亡的影响,基于 Frangakis 等提出的主分层法,研究了“永远幸存者”的 SACE 尖锐边界,并推广到“处理受益者”和“处理受害者”分层[15]。将边界求解转化成多参数线性规划问题,求解其一阶 KKT 条件对应的多项式方程组,计算和筛选得到的最优解可以为 SACE 上下边界提供精确表达式。模拟结果表明,所得边界都包含了给定的真实值,可有效识别 SACE,而且边界结果在一定范围内随着“永远幸存者”占比的增大而逐渐缩短。最后,采用美国国家支持工作示范(NSW)项目中的 Lalonde 数据集进行实证研究,选择完整协变量作为假定情形,因为本文采取了较弱的假设条件,该情形下的 SACE 边界难以确定 NSW 项目对于受试者收入有正向或者负向效应。未来可考虑给定附加假设或使用更高阶的 KKT 条件改善结果,提高方法的实际应用价值。

参考文献:

[1] Rubin D. Estimating causal effects of treatments in randomized and nonrandomized studies[J]. Journal of Educational Psychology, 1974, 66(5): 688-701.

[2] Frangakis C E, Rubin D B. Principal stratification in causal inference[J]. Biometrics, 2002, 58(1): 21-29.

[3] Zhang J L, Rubin D B. Estimation of causal effects via principal stratification when some outcomes are truncated by “death”[J]. Journal of Educational and Behavioral Statistics, 2003, 28(4): 353-368.

[4] Zhang J L, Rubin D B, Mealli F. Evaluating the effects of job training programs on wages through principal stratification[J]. Advances in Econometrics, 2008, 21: 117-145.

[5] Imai K. Sharp bounds on the causal effects in randomized experiments with “truncation-by-death”[J]. Statistics & Probability Letters, 2008, 78(2): 144-149.

[6] Horowitz J L, Manski C F. Identification and robustness with contaminated and corrupted data[J]. Econometrica, 1995, 63(2): 281-302.

[7] Freiman M H, Small D S. Large sample bounds on the survivor average causal effect in the presence of a binary covariate with conditionally ignorable treatment assignment[J]. The International Journal of Biostatistics, 2014, 10(2): 143-163.

[8] Ding P, Geng Z, Yan W, et al. Identifiability and estimation of causal effects by principal stratification with outcomes truncated by death[J]. Journal of the American Statistical Association, 2011, 106(496): 1578-1591.

[9] Wang L B, Zhou X H, Richardson T S. Identification and estimation of causal effects with outcomes truncated by death[J]. Biometrika, 2017, 104(3): 597-612.

[10] Balke A, Pearl J. Bounds on treatment effects from studies with imperfect compliance[J]. Journal of the American Statistical Association, 1997, 92(439): 1171-1176.

[11] Beresteanu A, Molchanov I, Molinari F. Partial identification using random set theory[J]. Journal of Econometrics, 2012, 166(1): 17-32.

[12] Manski C F. Partial identification of probability distributions[M]. Berlin: Springer, 2003.

[13] Shan N, Dong X G, Xu P F, et al. Sharp bounds on survivor average causal effects when the outcome is binary and truncated by death[J]. ACM Transactions on Intelligent Systems and Technology, 2016, 7(2): 1-11.

[14] Gabriel E E, Sachs M C, Sjölander A. Causal bounds for outcome-dependent sampling in observational studies[J]. Journal of the American Statistical Association, 2022, 117(538): 939-950.

[15] Chen A X, Cui X, Yang G R. A unified framework for bounding causal effects on the always-survivor and other populations[EB/OL]. [2024-04-26](2024-08-25). http://arXiv.org/abs/2403.13398v2.

[16] Gal T, Nedoma, J. Multiparametric linear programming[J]. Management Science, 1972, 18(7): 406-422.
[17] Acevedo J, Pistikopoulos E N. An algorithm for multiparametric mixed-integer linear programming problems[J]. Operations Research Letters, 1999, 24(3): 139-148.
[18] Dua V, Pistikopoulos E N. An algorithm for the solution of multiparametric mixed integer linear programming problems[J]. Annals of Operations Research, 2000, 99(1): 123-139.
[19] Li Z K, Ierapetritou M G. A new methodology for the general multiparametric mixed-integer linear programming (MILP) problems[J]. Industrial & Engineering Chemistry Research, 2007, 46(15): 5141-5151.
[20] Charitopoulos V M, Papageorgiou L G, Dua V. Multi-parametric linear programming under global uncertainty[J]. AIChE Journal, 2017, 63(9): 3871-3895.
[21] Charitopoulos V M, Papageorgiou L G, Dua V. Nonlinear model-based process process operation under uncertainty using exact parametric programming[J]. Engineering, 2017, 3(2): 202-213.
[22] Charitopoulos V M, Papageorgiou L G, Dua V. Multi-parametric mixed integer linear programming under global uncertainty [J]. Computers & Chemical Engineering, 2018, 116: 279-295.
[23] Pearl J. Causality: models, reasoning and inference[M]. 2nd ed. Cambridge: Cambridge University Press, 2009.
[24] Schaible S, Ibaraki T. Fractional programming[J]. European Journal of Operational Research, 1983, 12(4): 325-338.
[25] Li Q, Racine J S. Nonparametric econometrics[M]. Princeton: Princeton University Press, 2007.
[26] Lee D S. Training, wages, and sample selection: estimating sharp bounds on treatment effects[J]. Review of Economic Studies, 2009, 76(3): 1071-1102.
[27] Huber M, Laffers L, Mellace G. Sharp Ⅳ bounds on average treatment effects on the treated and other populations under endogeneity and noncompliance[J]. Journal of Applied Econometrics, 2017, 32(1): 56-79.
[28] Lalonde R. Evaluating the economic evaluation of training programs with experimental data[J]. The American Economic Review, 1986, 76: 604-620.
[29] Armstrong T B, Kolesár M. Finite-sample optimal estimation and inference on average treatment effects under unconfoundedness[J]. Econometrica, 2021, 89(3): 1141-1177.
[30] Chernozhukov V, Newey W K, Singh R. Automatic debiased machine learning of causal and structural effects[J]. Econometrica, 2022, 90(3): 967-1027.
[31] Masten M A, Poirier A, Zhang L Q. Assessing sensitivity to unconfoundedness: estimation and inference[J]. Journal of Business & Economic Statistics, 2024, 42(1): 1-13.

【责任编辑：卓祯雨】

附录一

$$\boldsymbol{A}(\theta)=\begin{bmatrix}
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & -\pi_{0\cdot c} & -\pi_{0\cdot c} & 0 & 0\\
0 & 0 & 0 & 0 & -\pi_{0\cdot c} & -\pi_{0\cdot c} & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\pi_{0\cdot c} & 0
\end{bmatrix}。$$

附录二

梯度部分:

$$\begin{cases}(\mu_1+\mu_3+\mu_6)\pi_{0\cdot c}+\mu_5-\lambda_1=0,\\(\mu_2+\mu_3+\mu_6)\pi_{0\cdot c}+\mu_5-\lambda_2-1=0,\\(\mu_1+\mu_4+\mu_6)\pi_{0\cdot c}+\mu_5-\lambda_3+1=0,\\(\mu_2+\mu_4+\mu_6)\pi_{0\cdot c}+\mu_5-\lambda_4=0,\\(\mu_3+\mu_6)\pi_{0\cdot c}-\lambda_5=0,\\(\mu_4+\mu_6)\pi_{0\cdot c}-\lambda_6=0,\\(\mu_1+\mu_6)\pi_{0\cdot c}-\lambda_7=0,\\(\mu_2+\mu_6)\pi_{0\cdot c}-\lambda_8=0,\\\mu_6\pi_{0\cdot c}-\lambda_9=0,\\\mu_1F_{11\cdot 0c}+\mu_2F_{01\cdot 0c}+\mu_3F_{11\cdot 1c}+\mu_4F_{01\cdot 1c}+\mu_6+\lambda_{10}P_{1\cdot 0c}+\lambda_{11}P_{1\cdot 1c}-\lambda_{13}(P_{1\cdot 1c}-P_{0\cdot 0c})=0。\end{cases}$$

约束部分:

$$\begin{cases}\pi_{0\cdot c}(f^{(0)}_{00\cdot c}+f^{(0)}_{20\cdot c}+f^{(0)}_{1\cdot 2c})-F_{11\cdot 0c}=0,\\\pi_{0\cdot c}(f^{(0)}_{10\cdot c}+f^{(0)}30\cdot c+f^{(0)}_{2\cdot 2c})-F_{01\cdot 0c}=0,\\\pi_{0\cdot c}(f^{(0)}_{00\cdot c}+f^{(0)}_{10\cdot c}+f^{(0)}_{0\cdot 1c})-F_{11\cdot 1c}=0,\\\pi_{0\cdot c}(f^{(0)}_{20\cdot c}+f^{(0)}_{30\cdot c}+f^{(0)}_{3\cdot 1c})-F_{01\cdot 1c}=0,\\f^{(0)}_{00\cdot c}+f^{(0)}_{10\cdot c}+f^{(0)}_{20\cdot c}+f^{(0)}_{30\cdot c})=1,\\\pi_{0\cdot c}(\sum_{i=0}^{3}f^{(0)}_{i0\cdot c}+f^{(0)}_{0\cdot 1c}+f^{(0)}_{3\cdot 1c}+f^{(0)}_{1\cdot 2c}+f^{(0)}_{2\cdot 2c}+f^{(0)}_{\cdot 3c})-1=0,\\-\lambda_1f^{(0)}_{00\cdot c}=0,\\-\lambda_2f^{(0)}_{10\cdot c}=0,\\-\lambda_3f^{(0)}_{20\cdot c}=0,\\-\lambda_4f^{(0)}_{30\cdot c}=0,\\-\lambda_5f^{(0)}_{0\cdot 1c}=0,\\-\lambda_6f^{(0)}_{3\cdot 1c}=0,\\-\lambda_7f^{(0)}_{1\cdot 2c}=0,\\-\lambda_8f^{(0)}_{2\cdot 2c}=0,\\-\lambda_9f^{(0)}_{\cdot 3c}=0,\\\lambda_{10}(\pi_{0\cdot c}-P_{1\cdot 0c})=0,\\\lambda_{11}(\pi_{0\cdot c}-P_{1\cdot 1c})=0,\\-\lambda_{12}\pi_{0\cdot c}=0,\\\lambda_{13}(P_{1\cdot 1c}-P_{0\cdot 0c}-\pi_{0\cdot c})=0。\end{cases}$$

附录三

附表1 部分最优解

Appendix Table 1 A subset of optimal solutions

变量	最优解1	最优解2
$f_{00\cdot c}^{(0)}$	$F_{11\cdot 1c}/(P_{1\cdot 1c}-P_{0\cdot 0c})$	$1-F_{01\cdot 0c}/(P_{1\cdot 1c}-P_{0\cdot 0c})$
$f_{10\cdot c}^{(0)}$	0	$F_{01\cdot 0c}/(P_{1\cdot 1c}-P_{0\cdot 0c})$
$f_{20\cdot c}^{(0)}$	$1-(F_{01\cdot 0c}+F_{11\cdot 1c})/(P_{1\cdot 1c}-P_{0\cdot 0c})$	0
$f_{30\cdot c}^{(0)}$	$F_{01\cdot 0c}/(P_{1\cdot 1c}-P_{0\cdot 0c})$	0
$f_{0\cdot 1c}^{(0)}$	0	$-1+F_{11\cdot 1c}/(P_{1\cdot 1c}-P_{0\cdot 0c})$
$f_{3\cdot 1c}^{(0)}$	$-1+(F_{01\cdot 1c}+F_{11\cdot 1c})/(P_{1\cdot 1c}-P_{0\cdot 0c})$	$F_{01\cdot 1c}/(P_{1\cdot 1c}-P_{0\cdot 0c})$
$f_{1\cdot 2c}^{(0)}$	$-1+(F_{01\cdot 0c}+F_{11\cdot 0c})/(P_{1\cdot 1c}-P_{0\cdot 0c})$	$-1+(F_{01\cdot 0c}+(F_{11\cdot 0c})/(P_{1\cdot 1c}-P_{0\cdot 0c})$
$f_{2\cdot 2c}^{(0)}$	0	0
$f_{\cdot 3c}^{(0)}$	$1-[(F_{01\cdot 0c}+F_{01\cdot 1c}+F_{11\cdot 0c}+F_{11\cdot 1c}-1)/(P_{1\cdot 1c}-P_{0\cdot 0c})]$	$1-[(F_{01\cdot 0c}+F_{01\cdot 1c}+F_{11\cdot 0c}+F_{11\cdot 1c}-1)/(P_{1\cdot 1c}-P_{0\cdot 0c})]$
$\pi_{0\cdot c}$	$P_{1\cdot 1c}-P_{0\cdot 0c}$	$P_{1\cdot 1c}-P_{0\cdot 0c}$
μ_1	0	0
μ_2	$1/(P_{1\cdot 1c}-P_{0\cdot 0c})$	$1/(P_{1\cdot 1c}-P_{0\cdot 0c})$
μ_3	$1/(P_{1\cdot 1c}-P_{0\cdot 0c})$	0
μ_4	0	0
μ_5	-1	0
μ_6	0	0
λ_1	0	0
λ_2	0	0
λ_3	0	1
λ_4	0	1
λ_5	1	0
λ_6	0	0
λ_7	0	0
λ_8	1	1
λ_9	0	0
λ_{10}	0	0
λ_{11}	0	0
λ_{12}	0	0
λ_{13}	$(F_{01\cdot 0c}+F_{11\cdot 1c})/(P_{1\cdot 1c}-P_{0\cdot 0c})^2$	$F_{01\cdot 0c}/(P_{1\cdot 1c}-P_{0\cdot 0c})^2$