arXiv:2304.04904v1 [stat.ME] 11 Apr 2023

纵向中介分析的靶向最大似然估计

作者：ZEYI WANG[1,*]、LARS VAN DER LAAN[2]、MAYA PETERSEN[1,†]、THOMAS GERDS[3]、KAJSA KVIST[4]、ANDMARK VAN DER LAAN[1,‡]，

[1]加州大学伯克利分校公共卫生学院生物统计学部，美国伯克利，
*wangzeyi@berkeley.edu;†mayaliv@berkeley.edu; ‡laan@stat.berkeley.edu

[2]加州大学伯克利分校公共卫生学院环境卫生科学部，美国伯克利，vanderlaanlars@yahoo.com

[3]哥本哈根大学公共卫生系生物统计学科，丹麦哥本哈根，tag@biostat.ku.dk[4]诺和诺德，丹麦索堡，tekk@novonordisk.com

采用随机干预的因果中介分析已成为理解纵向和生存结局中时变效应的重要研究领域。为应对复杂纵向数据结构所带来的因果与统计挑战——包括时变混杂因素、竞争风险和信息性删失——人们普遍希望将机器学习技术与半参数理论相结合。本文重点研究基于随机干预定义的纵向自然直接效应和间接效应的靶向最大似然估计（TMLE）。所提出的估计量具有多重稳健性和局部有效性，并直接估计和更新用于因子分解数据似然的条件密度。我们利用高度自适应套索（HAL）和投影表示，推导出纵向中介问题有效影响曲线的新估计量（HAL-EIC），并提出一种基于 HAL-EIC 的快速单步 TMLE 算法，同时保留其渐近性质。所提出的方法可推广至其他作为数据似然平滑函数的纵向因果参数，从而提供一种新颖而灵活的统计工具箱。

1. 引言。 越来越需要能够分析随时间变化的效应机制的方法，例如，在临床试验和电子健康研究中

关键词和短语：纵向中介分析，随机干预，随机干预，靶向最大似然估计（目标最大似然估计），有效影响曲线，有效估计量，高度自适应套索。

记录数据（Vansteelandt 等，2019；Buse 等，2020；Lai 等，2020）。例如，体重管理计划可能包括按计划安排的重复访视，在这些访视中收集体重指数（BMI）、血压、胆固醇水平以及其他健康状况的测量值。人们可能希望分析减重对胆固醇（连续型结局）或心血管事件风险（生存结局）的影响中，有多大比例是由血压控制得多好所介-导的。进一步而言，研究这一比例如何随时间变化也可能具有意义。中介分析提供了能够回答这些问题的估计目标。然而，在没有额外模型假设的情况下，当前最先进的基于静态干预的中介分析在适当调整时变混杂方面的能力有限；参见Avin、Shpitser 和Pearl（2005）关于存在“反悔见证者”时自然效应无法进行非参数识别的讨论。总体而言，纵向中介是一个具有挑战性的问题（Avin、Shpitser 和 Pearl，2005；VanderWeele 和 Vansteelandt，2009；Tchetgen 和 Shpitser，2012；Zheng 和 van der Laan，2012，2017；Van-derWeele 和 Tchetgen，2017），而当纵向因果中介目标参数（如自然直接效应和间接效应）被定义为静态干预时，调整时变混杂的任务会进一步复杂化（VanderWeele 和 Tchetgen，2017）。

近年来，基于随机干预的中介分析已被提出（Didelez、Dawid 和 Geneletti，2006；Díaz 和 Hejazi，2020；VanderWeele 和 Tchetgen，2017；Zheng 和 van der Laan，2012，2017）。结构因果模型中不再确定性地设定特定的中介变量取值（Pearl，2009），而是通过设定中介变量分布来定义随机干预的因果目标。这使得时变混杂因素能够得到灵活调整；相比之下，当存在任何干预后中介变量-结局混杂因素时，使用静态干预的自然效应无法实现非参数识别（Avin、Shpitser 和 Pearl，2005）。在不存在受干预影响的中介变量-结局混杂因素（即所谓的反悔见证者）的非纵向情境中，随机干预目标可以通过与相应经典静态干预相同的 g-计算公式来识别。然而，随机干预目标使我们能够放宽自然直接效应和间接效应（NDE 和 NIE）的跨世界识别假设，即不同静态干预下定义的反事实中介变量与结局之间不可检验的条件独立性（Robins 和 Richardson，2010）。

在本手稿中，我们针对纵向中介参数发展了基于靶向似然的方法（vander Laan，2010a,b），并构造了目标最大似然估计量（TMLE）。我们推导了 TMLE 保持一致且渐近线性的条件。我们还为纵向中介问题的有效影响曲线提供了投影表示（HAL-EIC），并利用它推导出一种快速的一步式靶向最大似然估计算法。在整个手稿中，我们将重点关注NIE 和 NDE 的纵向分析，但该方法可立即应用于受控直接效应（CDE）（见附录A.1）。这种灵活性源于：这些目标参数的 g-计算公式可以表示为同一个观测数据似然的函数。为了解释我们针对多元目标参数的基于靶向似然的方法的主要思想，考虑一个具有 K时间点的离散时间尺度。令 $O = (X_0, X_1, \ldots, X_K)$表示观测数据，并令 $O_1, \ldots, O_n \sim P_0$为一个独立同分布样本，且令 $p = dP/d\mu$ 为相对于分布 P的支配测度的密度函数。一个g-计算公式可以通过 O 在干预 g下的反事实分布 $P(g)$ 进行标记，其中， $P(g)$ 被识别为分解后的观测数据似然$p_0(O) = p_{0,X_0}(X_0) \prod_{k=1}^{K} p_{0,X_k}(X_k | X_0, \ldots, X_{k-1})$的函数。假设分析的目标不只是单个目标参数，而是一组目标参数 $\{\Psi_s(P_0)\ s = 1, \ldots, S\}$，其中$S$ 是目标参数的总数。当存在多个干预和/或多个结局受到关注时，就会出现这种情况，此时每一种干预与结局的组合都定义一个目标参数。我们要求每个目标参数都是路径可微的，并将 $D_s(P)$ 记为 Ψ_s的有效影响曲线。目标参数列表例如可以包含针对多个干预以及一系列评估时间点的 NIE和 NDE。因此，我们的两阶段基于靶向似然的估计方法首先从 p_0的完整似然 p_n^0 的初始估计开始，然后寻找似然 p_n^* 的更新估计，使其同时求解所有目标参数的有效影响曲线方程 $P_n D_s(p_n^*) = 0, s = 1, \ldots, S$。我们在本手稿中表明，由似然 p_n^* 的同一个更新估计推导出的代入估计量 $\Psi_s(p_n^*)\ s = 1, \ldots, S$（TMLE）是一致且渐近有效的。我们还表明，TMLE同时遵守所有目标的参数空间；也就是说，根据Robins 等人（2007）的意义，估计值保持在统计模型所蕴含的所有参数映射的范围内。

我们认为，与近期研究相比，我们针对纵向中介参数提出的统一似然方法具有优势。近期研究包括侧重于受限边际模型的方法（VanderWeele 和 Tchetgen，2017），在这类模型中，自然效应不再分解总效应，且不易适用于生存结局；也包括序贯回归分量方法（Zheng 和 van der Laan，2017），其条件期望通过迭代定义，类似于 Bang 和 Robins（2005），但不同目标参数之间并不共享这些条件期望。如上所述，我们基于完整似然的方法自然且灵活，适用于跨多个时间点和多个目标进行同步分析。在存在时变混杂因素、生存结局、信息性右删失和竞争风险的情形下，由同一更新后似然推导出的代入估计量将符合所有目标的参数空间。序贯回归或许能显著降低估计的计算成本，也可能符合参数空间（Robins 等人，2007），但仅在各个维度上分别成立。这使得采用一系列候选倾向评分模型及相应的 TMLE 更新进行基于损失的协同调整变得更加困难（van der Laan 和 Gruber，2010）。采用基于似然的估计时，即使对于多变量目标，也可以通过最小化同一组条件密度因子的负对数似然损失，自然地定义损失和调整过程。若将推断目标识别并表示为观测数据分布的函数，而无需额外的迭代定义和不直观的限制，识别假设和多重稳健性条件便更易于阐述。我们注意到，第7节中的多重稳健性条件仅涉及观测数据的条件密度，因此我们的方法无需指定序贯回归模型。简洁直接的稳健性条件更便于进行概念验证，从而能够更有效地利用专家的数据知识，提高研究的可靠性。

本文通过提出 HAL-EIC 表示，并开发一种使用 HAL-EIC 的快速单步 TMLE 算法，为纵向中介分析领域作出了创新性贡献。高度自适应套索（HAL）（Benkeser 和 van der Laan，2016；Bibaut 和 van der Laan，2019；van der Laan 等人，2018）是一种针对 d元实值右连续且有左极限函数的通用最大似然估计量（MLE），这些函数的截面变差范数有界；其收敛至真实函数的速率为 $n^{-2/3}(\log n)^d$，衡量标准为基于损失的差异度。近期进展（van der Laan、Rose 和 van der Laan，2018）

已证明，对于许多路径可微目标参数，HAL 估计量为有效影响曲线（EIC）提供了一种颇具前景的替代表示。在本文中，我们提出将 HAL-EIC 表示用于纵向中介问题，并开发一种快速的一步式靶向最大似然估计算法，利用 HAL-EIC 同时保留渐近性质。

本手稿的提纲如下。第2节介绍离散时间中介分析的符号和数据结构，其结局可能适用于所提出的方法，例如生存结局、右删失和竞争风险。第3节定义因果中介目标参数，并使用g-计算公式将其确定为统计目标，同时给出高效影响曲线。第4-5节讨论两阶段目标最大似然估计程序。第4节简要讨论构建初始密度估计量的步骤和注意事项。第5节给出迭代式和一步式靶向最大似然估计算法（vanderLaan and Gruber, 2016），并讨论局部效率的正则条件。我们在当前设定下回顾Dudoit和van derLaan（2008）提出的关于多元目标参数的同时推断方法。第6节中，我们使用高度自适应套索（HAL）（Benkeser and van der Laan, 2016；Hejazi,Coyle andvan der Laan, 2020）为高效影响曲线提出一种投影表示（HAL-EIC），并提出基于HAL-EIC的一步式靶向最大似然估计算法。在以下各节中，我们使用模拟数据分析多重稳健性性质（Tchetgen, 2009；Molina etal., 2017；Luedtke et al., 2017；Zheng and van der Laan,2017），并展示所提出的算法在有限样本挑战下的结果，例如正性假设近乎不成立的情况。

2. 数据与模型。 在离散时间尺度上， $\{0,1,\ldots,K\}$，我们用 X_t 表示所有 $0 \leq t \leq K$ 截至时间 t 测得的随机变量向量，即 $X_t = (X_k : \ k \in \{0,\ldots,t\})$。类似地，对于 $k \leq t$，我们定义 $\bar{X}_k^t = (X_k, \ldots\ldots, X_t)$。下文始终考虑以下数据结构：

$$O = (L_0, A_1, R_1, Z_1, L_1, \ldots, A_t, R_t, Z_t, L_t \ldots, A_K, R_K, Z_K, L_K),$$

其中每个节点都是一个随机变量或随机向量。 L_0 是基线协变量， A_K 是处理变量， Z_K 是时变中介变量， R_K 和 L_K 是在中介节点之前和之后的时变变量。向量 (A_t, R_t, Z_t, L_t) 是观测值

在时间点 t。鉴于这一时间顺序，我们将 X_t 的父变量记为$\text{Pa}(X_t)$，即位于X_t之前的变量，并将X_t 的子变量记为$\text{Ch}(X_t)$，即位于 X_t之后的变量。对于 O 取值范围内的一个实现 $o=(l_0,a_1\ldots,l_K)$ 的父节点和子节点，也采用类似的记法。例如，$\text{Pa}(L_1)=(L_0,A_1,R_1,Z_1)$，而 $\text{Ch}(a_K)=(r_K,z_K,l_K)$。在设定处理值的静态干预 $\bar{a}_K$的背景下，我们将 X_t 在干预 $A_K=\bar{a}_K$ 下的父节点记为 $\text{Pa}(X_t|\bar{a}_K)$，例如，$\text{Pa}(L_k|\bar{a}_k)=(L_0,a_1,R_1,Z_1,L_1,\ldots,a_k,R_k,Z_k)$，这是一个由观测到的随机变量和插补的处理值组成的有序向量。所关注的结局被定义为 L_K 的函数向量，即，对于多维泛函 $\psi=(\psi_s:s=1,\ldots,S)$，记为 $Y=\psi(L_K)$。示例包括不同时间点的累积测量值或生存状态。一个简单的特例是，在最后一个时间点测量的单变量 Y，使得 $Y\in L_K$。

设 $\mathcal{M}$ 是一个关于 O 分布的统计模型，该分布由测度μ 支配，因此对于 $P\in\mathcal{M}$，其密度为 $p=dP/d\mu$。为便于表示，我们假设 μ 是计数测度，因此像（2）和（3）这样的积分可简化为求和；不过，结果也适用于由勒贝格测度支配的连续概率分布及相应的积分。我们将 $P\in\mathcal{M}$ 与其对应的密度 p等同起来，并将密度相同的分布视为等价。我们考虑 n 个独立同分布的副本 O_1、.....、O_n，它们来自真实的数据生成分布 $P_0\in\mathcal{M}$。联合数据似然可按条件密度分解为，

$$(1)\quad p(O)=p_{L_0}(L_0)\prod_{k=1}^{K}\Big[p_{A_k}(A_k|\text{Pa}(A_k))p_{R_k}(R_k|\text{Pa}(R_k))\\ p_{Z_k}(Z_k|\text{Pa}(Z_k))p_{L_k}(L_k|\text{Pa}(L_k))\Big].$$

我们用记号 $Pf(O)=\mathbb{E}_P[f(O)]$表示关于 P的期望；同理，$P_nf(O)=\frac{1}{n}\sum_{i=1}^{n}f(O_i)$ 表示经验分布 P_n。

2.1. 结构因果模型和随机干预。 考虑以下结构因果模型（Pearl，2009；van der Laan 和 Rose，2011），其中观测数据的随机性 O 由所谓的外生节点 U_X的确定性映射来捕捉，并且 f_X 对于每个内生变量而言都是确定性的：

$$X_t=f_{X_t}(\text{Pa}(X_t),U_{X_t})$$

¯对于某些静态处理 $A_K = \bar{a}_K$，我们通过对结构因果模型进行静态干预来定义反事实。

$$
\begin{aligned}
L_0 =& f_{L_0}(U_{L_0}), \quad A_t = a_t \\
R_t(\bar{a}) =& f_{R_t}(\bar{a}_t, \bar{R}_{t-1}(\bar{a}), \bar{Z}_{t-1}(\bar{a}), \bar{L}_{t-1}(\bar{a}), U_{R_t}) \\
Z_t(\bar{a}) =& f_{Z_t}(\bar{a}_t, \bar{R}_t(\bar{a}), \bar{Z}_{t-1}(\bar{a}), \bar{L}_{t-1}(\bar{a}), U_{Z_t}) \\
L_t(\bar{a}) =& f_{L_t}(\bar{a}_t, \bar{R}_t(\bar{a}), \bar{Z}_t(\bar{a}), \bar{L}_{t-1}(\bar{a}), U_{L_t}).
\end{aligned}
$$

对于一对静态治疗干预和对照干预 $\bar{a}$ 和 $\bar{a}'$，用 $\Gamma_t^{\bar{a}'}$ 表示在 $\bar{a}'$下给定父节点时，对照干预反事实 $Z_t(\bar{a}')$ 的条件密度，即 $\Gamma_t^{\bar{a}'}(z_t|\bar{r}_t, \bar{z}_{t-1}, \bar{l}_{t-1}) = p(Z_t(\bar{a}') = z_t|\bar{R}_t(\bar{a}') = \bar{r}_t, \bar{Z}_{t-1}(\bar{a}') = \bar{z}_{t-1}, \bar{L}_{t-1}(\bar{a}') = \bar{l}_{t-1})$。因此，向量 $\bar{\Gamma}^{\bar{a}'} = \left(\Gamma_t^{\bar{a}'} \colon t = 1, \ldots, K\right)$ 表示对照干预下中介过程的一系列反事实条件密度。我们将 $(\bar{a}, \bar{\Gamma}^{\bar{a}'})$ 下的随机干预反事实定义为：强制中介变量遵循对照干预反事实的分布， $Z_t(\bar{a}, \bar{\Gamma}^{\bar{a}'}) \sim \Gamma_t^{\bar{a}'}$；对于节点 $X_t \in \bar{R} \cup \bar{L}$，插入 $\bar{A}_t = \bar{a}_t$，使得 $X_t(\bar{a}, \bar{\Gamma}^{\bar{a}'}) = f_{X_t}(\text{Pa}(X_t(\bar{a}, \bar{\Gamma}^{\bar{a}'})|\bar{a})\ U_X)$。也就是说，

$$
\begin{aligned}
L_0(\bar{a}, \bar{\Gamma}^{\bar{a}'}) =& f_{L_0}(U_{L_0}), \quad A_t = a_t \\
R_t(\bar{a}, \bar{\Gamma}^{\bar{a}'}) =& f_{R_t}(\bar{a}_t, \bar{R}_{t-1}(\bar{a}, \bar{\Gamma}^{\bar{a}'}), \bar{Z}_{t-1}(\bar{a}, \bar{\Gamma}^{\bar{a}'}), \bar{L}_{t-1}(\bar{a}, \bar{\Gamma}^{\bar{a}'}), U_{R_t}) \\
Z_t(\bar{a}, \bar{\Gamma}^{\bar{a}'}) \sim& \Gamma_t^{\bar{a}'}(z|\bar{R}_t(\bar{a}, \bar{\Gamma}^{\bar{a}'}), \bar{Z}_{t-1}(\bar{a}, \bar{\Gamma}^{\bar{a}'}), \bar{L}_{t-1}(\bar{a}, \bar{\Gamma}^{\bar{a}'})) \\
L_t(\bar{a}, \bar{\Gamma}^{\bar{a}'}) =& f_{L_t}(\bar{a}_t, \bar{R}_t(\bar{a}, \bar{\Gamma}^{\bar{a}'}), \bar{Z}_t(\bar{a}, \bar{\Gamma}^{\bar{a}'}), \bar{L}_{t-1}(\bar{a}, \bar{\Gamma}^{\bar{a}'}), U_{L_t}).
\end{aligned}
$$

在此框架中，用 $\mathbb{E}\left[Y(\bar{a}, \bar{\Gamma}^{\bar{a}'})\right]$ 表示描述治疗对结局 Y 影响之中介作用的因果目标参数。随后，总效应可分解为自然间接效应（NIE）和自然直接效应（NDE），即】【。

$$
\begin{aligned}
\mathbb{E}\left[Y(\bar{a}, \bar{\Gamma}^{\bar{a}})\right] - \mathbb{E}\left[Y(\bar{a}', \bar{\Gamma}^{\bar{a}'})\right] =& \left(\mathbb{E}\left[Y(\bar{a}, \bar{\Gamma}^{\bar{a}})\right] - \mathbb{E}\left[Y(\bar{a}, \bar{\Gamma}^{\bar{a}'})\right]\right) \\
&+ \left(\mathbb{E}\left[Y(\bar{a}, \bar{\Gamma}^{\bar{a}'})\right] - \mathbb{E}\left[Y(\bar{a}', \bar{\Gamma}^{\bar{a}'})\right]\right).
\end{aligned}
$$

为了说明跨世界假设，我们还定义了在静态干预 $\bar{A} = \bar{a}$ 和 $\bar{Z} = \bar{z}$ 下的反事实 $X(\bar{a}, \bar{z})$。请注意，如果在根据 $(\bar{a}, \bar{\Gamma}^{\bar{a}'})$ 随机抽取后，我们有 $\bar{Z}(\bar{a}, \bar{\Gamma}^{\bar{a}'}) = \bar{z}$，那么结构因果模型意味着 $X(\bar{a}, \bar{\Gamma}^{\bar{a}'}) = X(\bar{a}, \bar{z})$。然而，在不使用随机干预的情况下定义自然间接效应和自然直接效应通常需要额外的假设—

等条件独立性，例如 $L(\bar{a},\ \bar{z})$ 以及 $Z(\bar{a}')$ 在两个反事实世界之间。这类“跨世界”假设并不理想，因为它们既无法通过经验验证，也无法从概念上验证，参见 Andrews 和 Didelez(2020)。跨世界假-设也与随时间变化的协变量 R不相容。另一方面，随机干预框架不需要“跨世界”假设。在本文中，我们关注静态处理规则 $A=\bar{a}$ 或 $\bar{a}'$，但该方法可以推广到动态处理方案，其中对 A_t 的处理和对照干预由可用历史 Pa（A_t）的确定性函数决定。

2.2. 右删失生存结局。 假设 T 是感兴趣事件发生的时间点。令 $Y_t=\mathbb{I}_{\{T>t\}}\in L_t$ 表示研究中在第 t个时间点仍未发生事件的单调过程。计数过程 Y_t 在 $t=0$ 时从1 开始，且仅当事件在时间点 t发生时才跳变为0 。如果事件在时间点 t发生，那么对于所有 $X\in \mathsf{Ch}(Y_t)\setminus Y$，给定 $Y_t=0$，我们令 X 为一个退化离散变量，使得 $X=\varnothing$ 以1的条件概率成立。感兴趣的结局可以是超过研究时长 K仍存活，此时目标参数的形式为

$$\mathbb{E}\left[Y_K(\bar{a},\bar{\Gamma}^{\bar{a}'})\right]=P(T(\bar{a},\bar{\Gamma}^{\bar{a}'})>K).$$

在实际应用中，通常只能观测到 $T=\min\{T,C\}$，其中 C是删失时间。为纳入删失数据，我们构建双变量处理节点 $A_t=(A_t^C,A_t^E)$，由保持未删失状态的单调过程 $A_t^C=\mathbb{I}_{\{C>t\}}$和处理分配 A_t^E组成。过程 A_t^C的初始值为1，并在发生删失的时间点跳变为0 。给定 $C_t=0$，对于所有 $X\in\mathsf{Ch}(C_t)$，我们令 X 为一个离散变量，它以1的条件概率取退化值 $\varnothing$；也就是说，删失发生后，观测数据似然不再包含任何信息。

2.3. 竞争风险。 考虑一个竞争事件框架（Andersen 等人，1993；Benkeser、Carone 和 Gilbert，2018；Rytgaard 和 van der Laan，2022），在时间 T，当个体进入多个吸收状态之一时，观测事件类型 $\Delta\in\{1,2,\ldots,J\}$。例如，$\Delta=1$ 可能表示癌症发病，而 $\Delta=2$ 表示死于其他原因。可以定义一个离散时间的多维计数过程，例如，当 $J=2$由 $\left\{Y_t=(N_{(1)t},N_{(2)t}):t=1,\ldots,K\right\}$时，使得 $N_{(1)t}=\mathbb{I}_{\{T\le t,\Delta=1)\}}$, $N_{(2)t}=\mathbb{I}_{\{T\le t,\Delta=2)\}}$。注意，$1-N_{(1)t}-N_{(2)t}=\mathbb{I}_{\{T>t\}}$现在表示个体仍然存活

¯且未发生事件。假设 $Y_t \in L_t$。对于 $Y_t = (0, 0)$，我们令 $X = \varnothing$ 对任意 $X \in \text{Ch}(Y_t) \setminus Y$ 以概率 1 成立。如果 $N_t^{(j)} = 1$ 对应第 j 种事件，则还令 $N(j)_{t'} = 1$ 和 $N^{(j')}{}_{t'} = \varnothing$ 对所有 $j' \neq j, t' \geq t$ 成立。竞争风险框架中的目标参数通常是多维的，例如，可以是在所有时间点上 $(\bar{a}, \bar{\Gamma}^{\bar{a}'})$ 下所有事件的风险，$\mathbb{E}\left[\bar{Y}(\bar{a}, \bar{\Gamma}^{\bar{a}'})\right] = \left(\mathbb{E}\left[N(j)_t(\bar{a}, \bar{\Gamma}^{\bar{a}'})\right] : t = 1, \ldots, K;\ j = 1, \ldots, J\right)$。

3. 目标参数。

3.1. 自然直接效应和间接效应。我们为感兴趣的多维结局定义自然直接效应（NDE）和自¯ ¯然间接效应（NIE），并采用第 $Y = \psi(L_K)$ 节中定义的随机干预2：

$$\text{NIE: } \mathbb{E}\left[\bar{Y}(\bar{a}, \bar{\Gamma}^{\bar{a}}) - \bar{Y}(\bar{a}, \bar{\Gamma}^{\bar{a}'})\right],$$

$$\text{NDE: } \mathbb{E}\left[\bar{Y}(\bar{a}, \bar{\Gamma}^{\bar{a}'}) - \bar{Y}(\bar{a}', \bar{\Gamma}^{\bar{a}'})\right].$$

¯选择 Y 的示例包括：$Y \in L_K$ 对于单变量结局变量，参见第2.1，$Y_K = \mathbb{I}_{\{T > K\}} \in L_K$ 对于删失生存终点，参见第2.2，以及 $\bar{Y} = \left(N^{(1)t},\ N_t^{(2)} :\ t = 1, \ldots,\ K\right)$ 对于竞争风险结局，参见第 2.3。

自然直接效应（NDE）是指在强制中介变量具有与其对照组反事实相同的分布时，治疗的直接效应。自然间接效应（NIE）是指不改变治疗值，而是改变中介变量分布所实现的间接治疗效应。第2.1 节的结构因果模型意味着，自然直接效应（NDE）和自然间接效应（NIE）分解了总效应，总效应可以在有或没有随机干预的情况下定义：

$$\text{NIE} + \text{NDE} = \mathbb{E}\left[\bar{Y}(\bar{a}, \bar{\Gamma}^{\bar{a}})\right] - \mathbb{E}\left[\bar{Y}(\bar{a}', \bar{\Gamma}^{\bar{a}'})\right] = \mathbb{E}\left[\bar{Y}(\bar{a})\right] - \mathbb{E}\left[\bar{Y}(\bar{a}')\right].$$

3.2. 识别。为了将因果中介目标识别为观测数据分布的统计参数，我们采用 Zheng 和 van der Laan （2017）提出的以下假设。对于任何感兴趣的随机干预 $(\bar{a}, \bar{\Gamma}^{\bar{a}'})$：

(A1) 序贯可交换性：$\bar{R}_t^K(\bar{a}')\ \bar{Z}_t^K(\bar{a}')\ \bar{L}_t^K(\bar{a}')\ \bar{R}_t^K(\bar{a}, \bar{z})\ \bar{L}_t^K(\bar{a}, \bar{z}) \perp A_t | \text{Pa}(A_t)$。(A2) 中介变量随机化：$\bar{R}_{t+1}^K(\bar{a}, \bar{z})\ \bar{L}_t^K(\bar{a}, \bar{z}) \perp Z_t | \text{Pa}(Z_t)$。(A3)（强）正值性：$p_0(a_t^* | \text{Pa}(a_t^* | \bar{a}^*)) > 0$ 当 $p_0(\text{Pa}(a_t^* | \bar{a}^*)) > 0$ 对于 $\bar{a}^* = \bar{a}$ 或 $\bar{a}'$成立；此外，$p_0(r_t | \text{Pa}(r_t | \bar{a}_t')) > 0$ 当 $p_0(r_t | \text{Pa}(r_t | \bar{a}_t)) > 0$成立，$p_0(l_t | \text{Pa}(l_t | \bar{a}_t')) > 0$ 当 $p_0(l_t | \text{Pa}(l_t | \bar{a}_t)) > 0$成立，且 $p_0(z_t | \text{Pa}(z_t | \bar{a}_t)) > 0$ 当 $p_0(z_t | \text{Pa}(z_t | \bar{a}_t')) > 0$成立。

注意，一致性假设（Cole和 Frangakis，2009）是通过结构因果模型隐含作出的。以下定理是将 Zheng 和 van der Laan （2017）中的引理1直接应用于多变量结局 $Y=(Y_s:\ s)$ 的每个维度。

定理 3.1（G计算公式）。设感兴趣的多维结局 $Y=(Y_s: s=1,\ldots,S)$ 其中 $Y_s=\psi_s(L_K)$ 由可测函数ψ_s定义。对于任意两个干预 $\bar{a}$， $\bar{a}'$ 在第*3.2*节列出的识别假设 （*A1*）–（*A3*）下，我们有

$$\begin{aligned}\Psi_s^{\bar{a},\bar{a}'}(P) \equiv& \mathbb{E}\left[Y_s(\bar{a},\bar{\Gamma}^{\bar{a}'})\right]\\ =&\sum_{\bar{l}}\psi_s(\bar{l}_K)p(\bar{L}_K(\bar{a},\bar{\Gamma}^{\bar{a}'})=\bar{l}_K)\\ =&\sum_{\bar{l},\bar{z},\bar{r}}\psi_s(\bar{l}_K)p(L_0=l_0)\prod_{t=1}^{K}p_{L_t}(l_t|Pa(l_t|\bar{a}_t))p_{Z_t}(z_t|Pa(z_t|\bar{a}'_t))p_{R_t}(r_t|Pa(r_t|\bar{a}_t)).\end{aligned}\tag{2}$$

此外，对于结局的每个维度对应的自然间接效应和自然直接效应，

$$\begin{aligned}NIE_s(P)=&\Psi_s^{\bar{a},\bar{a}}(P)-\Psi_s^{\bar{a},\bar{a}'}(P)\\ NDE_s(P)=&\Psi_s^{\bar{a},\bar{a}'}(P)-\Psi_s^{\bar{a}',\bar{a}'}(P).\end{aligned}$$

3.3. 与序贯回归的比较。我们已将目标参数定义并识别为观测数据似然 p 的函数，接下来的章节将聚焦于这种全似然表示。本小节将其与序贯回归（或迭代条件期望）方法进行比较（Bang 和 Robins，2005；Zheng 和 van der Laan，2017），该方法将相同的 G计算公式改写为迭代定义的回归分量的函数。

对于任意一对干预 $\bar{a}$ 和 $\bar{a}'$，在第 $P^{\bar{a},\ \bar{a}'}$ 节中的假设下，可以识别随机干预下反事实节点的干预后分布 $(\bar{a},\ \bar{\Gamma}^{\bar{a}'})$ 即，对于任意数据实现 3.2，只要该实现强制干预节点取值 o，便有 $\bar{a}_K$：

$$\begin{aligned}p^{\bar{a},\bar{a}'}(O=o)=&p(O(\bar{a},\Gamma^{\bar{a}'})=o)\\ =&p(\bar{L}_K(\bar{a},\Gamma^{\bar{a}'})=\bar{l}_K,\bar{Z}_K(\bar{a},\Gamma^{\bar{a}'})=\bar{z}_K,\bar{R}_K(\bar{a},\Gamma^{\bar{a}'})=\bar{r}_K)\\ =&\prod_{t=1}^{K}p_{L_t}(l_t|\mathrm{Pa}(l_t|\bar{a}_t))p_{Z_t}(z_t|\mathrm{Pa}(z_t|\bar{a}'_t))p_{R_t}(r_t|\mathrm{Pa}(r_t|\bar{a}_t)).\end{aligned}$$

对于统计目标参数的每个维度 s，都可以通过以下一系列回归表达式重写 g-计算公式（除最后一个方程外，省略了 Q 对 P 的依赖关系）：

$$
\begin{aligned}
Q^{\bar{a},\bar{a}'}_{s,R_{K+1}} =& Y_s = \psi_s(\bar{L}_K) \\
Q^{\bar{a},\bar{a}'}_{s,L_t}\left(\bar{R}_t, \bar{Z}_t, \bar{L}_{t-1}\right) =& \mathbb{E}_{P^{\bar{a},\bar{a}'}}\left[Q^{\bar{a},\bar{a}'}_{s,R_{t+1}} | \bar{R}_t, \bar{Z}_t, \bar{L}_{t-1}\right] = \mathbb{E}_P\left[Q^{\bar{a},\bar{a}'}_{s,R_{t+1}} | \bar{R}_t, \bar{Z}_t, \bar{L}_{t-1}, \bar{A}_t = \bar{a}_t\right] \\
Q^{\bar{a},\bar{a}'}_{s,Z_t}\left(\bar{R}_t, \bar{Z}_{t-1}, \bar{L}_{t-1}\right) =& \mathbb{E}_{P^{\bar{a},\bar{a}'}}\left[Q^{\bar{a},\bar{a}'}_{s,L_t} | \bar{R}_t, \bar{Z}_{t-1}, \bar{L}_{t-1}\right] = \mathbb{E}_P\left[Q^{\bar{a},\bar{a}'}_{s,L_t} | \bar{R}_t, \bar{Z}_{t-1}, \bar{L}_{t-1}, \bar{A}_t = \bar{a}'_t\right] \\
Q^{\bar{a},\bar{a}'}_{s,R_t}\left(\bar{R}_{t-1}, \bar{Z}_{t-1}, \bar{L}_{t-1}\right) =& \mathbb{E}_{P^{\bar{a},\bar{a}'}}\left[Q^{\bar{a},\bar{a}'}_{s,Z_t} | \bar{R}_{t-1}, \bar{Z}_{t-1}, \bar{L}_{t-1}\right] \\
=& \mathbb{E}_P\left[Q^{\bar{a},\bar{a}'}_{s,Z_t} | \bar{R}_{t-1}, \bar{Z}_{t-1}, \bar{L}_{t-1}, \bar{A}_t = \bar{a}_t\right] \\
\Psi^{\bar{a},\bar{a}'}_s(P) =& \mathbb{E}_P Q^{\bar{a},\bar{a}'}_{s,R_1}(P)(L_0).
\end{aligned}
$$

请注意，通过解析地计算嵌套条件期望，每个回归表达式也可以写成条件密度的函数。例如，为简单起见，假设 L_k 是一个离散变量，我们有

$$
\begin{aligned}
(3)\quad Q^{\bar{a},\bar{a}'}_{s,L_K} &= \mathbb{E}_P\left[Q^{\bar{a},\bar{a}'}_{s,R_{K+1}} | \bar{R}_K, \bar{Z}_K, \bar{L}_{K-1}, \bar{A}_t = \bar{a}_t\right] \\
&= \sum_{l_K} \psi_s(\bar{L}_{K-1}, l_K) p_{L_K}(l_k | \bar{R}_K, \bar{Z}_K, \bar{L}_{K-1}, \bar{A}_K = \bar{a}_K).
\end{aligned}
$$

因此，序贯回归表述为 $\Psi^{\bar{a},\bar{a}'}_s(P)$ 的 G计算公式提供了另一种将其表示为...

...的泛函的方式

$$
\bar{Q}^{\bar{a},\bar{a}'}_s(P) = \left(Q^{\bar{a},\bar{a}'}_{s,X}(P) : X \in \{L_t, Z_t, R_t : t = 1, \ldots, K\}\right).
$$

它只需要估计回归模型，$Q^{\bar{a},\bar{a}'}_{s,X}(P)$，而非 P因子化数据似然（1），这可能提升可扩展性。然而，这也意味着无法从数据分布中获取那么多信息 P，且一次只能处理一个目标参数。必须针对感兴趣结局的每个维度以及每个感兴趣的随机干预 $\bar{Q}^{\bar{a},\bar{a}'}_s(P)$ 定义并估计一组不同的中间回归分量，s 这一做法并不直观，而且可能导致结果不符合参数空间。尽管由于替代估计，所得估计值仍满足目标参数各维度对应的边际参数空间，但序贯回归无法保证联合参数空间的“有界性”性质（Robins 等人，2007）。在纵向中介分析中，目标参数通常是多维的——

¯tion分析。例如，当所关注的结局是完整的生存过程 $Y=(Y_t=\mathbb{I}_{\{T>t\}}:\ t=1,\ldots,K)$时，$\mathbb{E}\left[Y_t(\bar{a},\bar{\Gamma}^{\bar{a}'})\right]$ 的估计值应当随 $t=1$、2、. . .、 K 单调变化。另一个例子是第 2.3 节中的竞争风险，其中估计值也需要满足： $\mathbb{E}\left[Y(1)_t(\bar{a},\bar{\Gamma}^{\bar{a}'})\right]+\mathbb{E}\left[Y(2)_t(\bar{a},\bar{\Gamma}^{\bar{a}'})\right]=\mathbb{E}\left[Y_t^*(\bar{a},\bar{\Gamma}^{\bar{a}'})\right]\in[0,1]$，并且随 $t=1$、. ..、 K增大而增加。另一方面，我们的目标最大似然估计算法从同一组估计的似然因子中推导出所有目标参数的估计值，因此实现了所需的联合有界性。

3.4. 有效影响曲线（*EIC*）。

定理 3.2（有效影响曲线（EIC））。对于每一对干预 $\bar{a}$， $\bar{a}'$ 以及结局向量的第 s 个元素 $\bar{Y}$， $D^{\bar{a}}$的有效影响曲线 $_{s}{}^{\bar{a}'}$ （ P ）在 $\Psi^{\bar{a}}{}_{s}{}^{\bar{a}'}$ 处的值为 P （省略了对 P 的部分依赖）：

$$
\begin{aligned}
D_{s,L_t}^{\bar{a},\bar{a}'}=&\frac{\mathbb{I}_{\{\bar{A}_t=\bar{a}_t\}}}{\prod_{j=1}^{t}p_A(a_j|Pa(A_j|\bar{a}_{j-1}))}\\
&\prod_{j=1}^{t}\frac{p_Z(Z_j|Pa(Z_j|\bar{a}'_j))}{p_Z(Z_j|Pa(Z_j|\bar{a}_j))}\left\{Q_{s,R_{t+1}}^{\bar{a},\bar{a}'}(\bar{R}_t,\bar{Z}_t,\bar{L}_t)-Q_{s,L_t}^{\bar{a},\bar{a}'}(\bar{R}_t,\bar{Z}_t,\bar{L}_{t-1})\right\}\\
D_{s,Z_t}^{\bar{a},\bar{a}'}=&\frac{\mathbb{I}_{\{\bar{A}_t=\bar{a}'_t\}}}{\prod_{j=1}^{t}p_A(a'_j|Pa(A_J|\bar{a}'_{j-1}))}\prod_{j=1}^{t-1}\frac{p_L(L_j|Pa(L_j|\bar{a}_j))}{p_L(L_j|Pa(L_j|\bar{a}'_j))}\prod_{j=1}^{t}\frac{p_R(R_j|Pa(R_j|\bar{a}_j))}{p_R(R_j|Pa(R_j|\bar{a}'_j))}\\
&\left\{Q_{L_t}^{\bar{a},\bar{a}'}(\bar{R}_t,\bar{Z}_t,\bar{L}_{t-1})-Q_{Z_t}^{\bar{a},\bar{a}'}(\bar{R}_t,\bar{Z}_{t-1},\bar{L}_{t-1})\right\}\\
D_{s,R_t}^{\bar{a},\bar{a}'}=&\frac{\mathbb{I}_{\{\bar{A}_t=\bar{a}_t\}}}{\prod_{j=1}^{t}p_A(a_j|Pa(A_j|\bar{a}_{j-1}))}\\
&\prod_{j=1}^{t-1}\frac{p_Z(Z_j|Pa(Z_j|\bar{a}'_j))}{p_Z(Z_j|Pa(Z_j|\bar{a}_j))}\left\{Q_{s,Z_t}^{\bar{a},\bar{a}'}(\bar{R}_t,\bar{Z}_{t-1},\bar{L}_{t-1})-Q_{s,R_t}^{\bar{a},\bar{a}'}(\bar{R}_{t-1},\bar{Z}_{t-1},\bar{L}_{t-1})\right\}\\
D_s^{\bar{a},\bar{a}'}=&D_{s,L_0}^{\bar{a},\bar{a}'}+\sum_{t=1}^{K}D_{s,L_t}^{\bar{a},\bar{a}'}+D_{s,Z_t}^{\bar{a},\bar{a}'}+D_{s,R_t}^{\bar{a},\bar{a}'}.
\end{aligned}
$$

证明见附录A。

注意，这里的Q泛函被定义为条件密度的函数，方法是按照方程（3）迭代地执行所有积分。例如，

$$
Q_{s,L_t}^{\bar{a},\bar{a}'}(\bar{R}_t,\bar{Z}_t,\bar{L}_{t-1})=\sum_{l_t,\mathrm{Ch}(l_t)\setminus\bar{a}}\psi_s(\bar{L}_{t-1},\bar{l}_t^K)\prod_{j=t}^{K}p_L(l_j|\bar{R}_t,\bar{Z}_t,\bar{L}_{t-1},\bar{a}_j,\bar{r}_{t+1}^j,\bar{z}_{t+1}^j,\bar{l}_t^{j-1})
$$

$$\prod_{j=t+1}^{K} p_Z(z_j|\bar{R}_t, \bar{Z}_t, \bar{L}_{t-1}, \bar{a}'_j, \bar{r}^j_{t+1}, \bar{z}^{j-1}_{t+1}, \bar{l}^{j-1}_t)$$

$$p_R(r_j|\bar{R}_t, \bar{Z}_t, \bar{L}_{t-1}, \bar{a}_j, \bar{r}^{j-1}_{t+1}, \bar{z}^{j-1}_{t+1}, \bar{l}^{j-1}_t).$$

4.数据似然的初始估计。在本节中，我们简要描述构建初始估计量的步骤和注意事项，

$$p^0_n = p^0_{n,L_0} \prod_{t=1}^{K} p^0_{n,A_t} \prod_{t=1}^{K} p^0_{n,R_t} p^0_{n,Z_t} p^0_{n,L_t},$$

对于观测数据似然中涉及的条件密度（1），在 P_0。

对于每个二元变量 $X \in A \cup R \cup Z \cup L$，我们有 $p_X(1|\text{Pa}(X)) = \mathbb{E}[X|\text{Pa}(X)]$,以及 $p_X(0|\text{Pa}(X)) = 1 - \mathbb{E}[X|\text{Pa}(X)]$。对于分类变量，我们定义一系列二元虚拟变量，形式为 $X^j = \mathbb{I}_{\{X=j\}}$ 或 $X^j = \mathbb{I}_{\{X=j, X\notin\{1,\ldots,j-1\}\}}$，其中令 $\text{Pa}(X^j) = \text{Pa}(X) \cup \{X^s : s = 1, \ldots, j-1\}$，然后可以通过对条件期望集合建模来估计 p_X，即 $\left\{\mathbb{E}\left[X^j|\text{Pa}(X^j)\right] : j\right\}$。对于连续变量，条件密度的建模更具挑战性。建模选项包括参数假设，以及离散化的条件密度；后者可采用 Díaz Muñoz 和 van der Laan (2011) 所述的合并风险回归进行建模，其中需针对覆盖连续变量取值范围的网格 $(\alpha_v : v = 0, 1, \ldots)$，指定 $\mathbb{E}[\mathbb{I}_{\{X\in[\alpha_{v-1},\alpha_v)\}}|A \geq \alpha_{v-1}, Pa(X)]$ 的模型。基于所有变量的模型库，可以应用超级学习器（van der Laan、Polley 和 Hubbard，2007）来估计 $\mathbb{E}[X|\text{Pa}(X)]$，方法是将候选学习器库组合成凸集成（凸学习器），或选取候选学习器中的最佳估计器（离散学习器）；具体选择依据交叉验证损失表现。根据有限样本预言不等式（Van Der Laan 和 Dudoit，2003；van der Laan、Polley 和 Hubbard，2007），超级学习器通常在渐近意义上等价于预言学习器；预言学习器定义为候选学习器中在真实数据生成分布下损失最小的学习器。

高度自适应套索（HAL）（Benkeser 和 van der Laan，2016；Bibaut 和 van der Laan，2019）是一种具有快速收敛率的非参数估计量，可用于对条件期望对象 $\mathbb{E}[X|\text{Pa}(X)]$建模。已有研究表明，在真实条件期望为càdlàg 且在截面变差范数下有界的假设下，相应的 HAL 估计量 $\mathbb{E}[X|\text{Pa}(X)]$ 的收敛速率为$\left\|\hat{}\mathbb{E}[X|\text{Pa}(X)] - \mathbb{E}[X|\text{Pa}(X)]\right\|_{P_0} = O_P(n^{-1/3})$</style}，精确到 $\log(n)$ 因子的范围内。因此，任何二元

或者分类密度估计器$p^0_{n,X}$，该估计器源自HAL（或将 HAL 纳入候选学习器库的超级学习器）估计器 $\mathbb{E}[X|\text{Pa}(X)]$，将保留 $\left\|p_{0,X} - p^0_{n,X}\right\|_{P_0} = O_P(n^{-1/3})$ 的相同收敛速率，最多相差 $\log(n)$ 因子，其中$p_0 = p_{0,L_0}\prod_{t=1}^K p_{0,A_t}\prod_{t=1}^K p_{0,R_t}p_{0,Z_t}p_{0,L_t}$ 是在真实数据分布 P_0 下的因子化联合密度。

最后，我们指出，最近由`haldensify` （Hejazi、Benkeser 和van der Laan，2022）开展的工作允许基于 HAL 对离散变量和连续变量的条件密度进行灵活估计。建议考虑使用一个包含基于 HAL的估计器以及其他密度估计器（如基于核的估计器或基于神经网络的估计器）的库来构建超级学习器。

5.靶向最大似然估计。给定初始密度估计量 $p^0_n = p^0_{n,L_0}\prod_{t=1}^K p^0_{n,A_t}\prod_{t=1}^K p^0_{n,R_t}p^0_{n,Z_t}p^0_{n,L_t}$，我们现在定义目标最大似然估计更新$p^*_n = p^*_{n,L_0}\prod_{t=1}^K p^*_{n,A_t}\prod_{t=1}^K p^*_{n,R_t}p^*_{n,Z_t}p^*_{n,L_t}$，使得代入估计量$\Big(\Psi^{\bar{a},\bar{a}}_s(P^*_n)\ \Psi^{\bar{a}}_{s\ \bar{a}'}(P^*_n)\ \Psi^{\bar{a}'}_{s\ \bar{a}'}(P^*_n): s = 1,\ldots,S\Big)$在正则性假设下是一致且渐近线性的。

5.1. 迭代 *TMLE*。 为似然的各个组成部分定义以下（局部）最不利路径：

$$\tilde{p}_{R_t}(p_{R_t},\bar{\epsilon}_{R_t}) = (1 + \sum_{s=1}^S \epsilon^{\bar{a},\bar{a}}_{s,R_t} D^{\bar{a},\bar{a}}_{s,R_t}(P) + \epsilon^{\bar{a},\bar{a}'}_{s,R_t} D^{\bar{a},\bar{a}'}_{s,R_t}(P) + \epsilon^{\bar{a}',\bar{a}'}_{s,R_t} D^{\bar{a}',\bar{a}'}_{s,R_t}(P))p_{R_t}$$

$$\tilde{p}_{Z_t}(p_{Z_t},\bar{\epsilon}_{Z_t}) = (1 + \sum_{s=1}^S \epsilon^{\bar{a},\bar{a}}_{s,Z_t} D^{\bar{a},\bar{a}}_{s,Z_t}(P) + \epsilon^{\bar{a},\bar{a}'}_{s,Z_t} D^{\bar{a},\bar{a}'}_{s,Z_t}(P) + \epsilon^{\bar{a}',\bar{a}'}_{s,Z_t} D^{\bar{a}',\bar{a}'}_{s,Z_t}(P))p_{Z_t}$$

$$\tilde{p}_{L_t}(p_{L_t},\bar{\epsilon}_{L_t}) = (1 + \sum_{s=1}^S \epsilon^{\bar{a},\bar{a}}_{s,L_t} D^{\bar{a},\bar{a}'}_{s,L_t}(P) + \epsilon^{\bar{a},\bar{a}'}_{s,L_t} D^{\bar{a},\bar{a}'}_{s,L_t}(P) + \epsilon^{\bar{a}',\bar{a}'}_{s,L_t} D^{\bar{a}',\bar{a}'}_{s,L_t}(P))p_{L_t}.$$

假设在每个节点 $X \in R \cup Z \cup L$，我们都构造了一个关于 $\mathcal{M}$ 的子模型，该子模型具有多维参数 $\bar{\epsilon}_X = (\epsilon^{\bar{a}}_{s\ \bar{a},X}, \epsilon^{\bar{a}}_{s\ \bar{a}',X}, \epsilon^{\bar{a}'}_{s\ \bar{a}',X} : s = 1,\ldots,S)$，其中 $\tilde{p}_X(p_X, \bar{\epsilon}_X = \bar{0}) = p_X$，并且在 $\bar{\epsilon}_X = 0$ 处，得分 $\frac{d}{d\bar{\epsilon}_X}\log \tilde{p}_X(p_X,\bar{\epsilon}_X)|_{\bar{\epsilon}_X=0}$ 等于有效影响曲线 $\Big(D^{\bar{a},\bar{a}}_{s,X}, D^{\bar{a},\bar{a}'}_{s,X}, D^{\bar{a}',\bar{a}'}_{s,X}$ 的相应分量组成的向量：$s = 1,\ldots,S\Big)$。在 $p = p^0_n$ 处，参数子模型的最大似然解为 $t = 1,\ldots,K$，

$$\bar{\epsilon}_{R_t,n} = \text{argmax}_{\bar{\epsilon}_{R_t}} P_n \log \tilde{p}_{R_t}(p^0_{n,R_t},\bar{\epsilon}_{R_t})$$

$$\bar{\epsilon}_{Z_t,n} = \text{argmax}_{\bar{\epsilon}_{Z_t}} P_n \log \tilde{p}_{Z_t}(p^0_{n,Z_t},\bar{\epsilon}_{Z_t})$$

$$\bar{\epsilon}_{L_t,n} = \text{argmax}_{\bar{\epsilon}_{L_t}} P_n \log \tilde{p}_{L_t}(p^0_{n,R_t}, \bar{\epsilon}_{L_t}),$$

从而进行第一轮 TMLE 更新

$$\tilde{p}_{n,R_t} = \tilde{p}_{R_t}(p^0_{n,R_t}, \bar{\epsilon}_{R_t,n})$$

$$\tilde{p}_{n,Z_t} = \tilde{p}_{Z_t}(p^0_{n,Z_t}, \bar{\epsilon}_{Z_t,n})$$

$$\tilde{p}_{n,L_t} = \tilde{p}_{L_t}(p^0_{n,L_t}, \bar{\epsilon}_{L_t,n}),$$

以及 $\tilde{p}_n = p^0_{n,L_0} \prod_{t=1}^K p^0_{n,A_t} \prod_{t=1}^K \tilde{p}_{n,R_t} \tilde{p}_{n,Z_t} \tilde{p}_{n,L_t}$。

在将初始估计量 p^0_n 替换为上一次迭代的TMLE 更新 $\tilde{p}_n$ 后，重复该过程，直到 $P_n D^{\bar{a}}_s, \bar{a}(\tilde{P}_n) \to 0$, $P_n D^{\bar{a}}_s, \bar{a}'(\tilde{P}_n) \to 0$, $P_n D^{\bar{a}'}_s, \bar{a}'(\tilde{P}_n) \to 0$, 对于 $s = 1, \ldots, S$。现在，我们将最终 TMLE 更新 p^*_n 定义为第 I次迭代的更新，其中 $I = I(n)$ 足够大，使得 $P_n D^{\bar{a}}_s, \bar{a}(P^*_n)$、$P_n D^{\bar{a}}_s, \bar{a}'(P^*_n)$、$P_n D^{\bar{a}'}_s, \bar{a}'(P^*_n)$ 都是 $o_P(n^{-1/2})$。最后一个陈述是在正则条件下对vander Laan和Rubin（2006）中的结果1 和定理 1 的应用，这些正则条件要求 $\tilde{p}_n$ 位于 $\mathcal{M}$ 的内部，从而使得 $P_n D^{\bar{a}}_{s,X}, (\tilde{P}_n) = P_n D^{\bar{a}}_{s,X} (\tilde{P}_n) \tilde{p}_n, P_n D^{\bar{a}'}_{s,X}, (\tilde{P}_n) = P_n D^{\bar{a}'}_{s,X} (\tilde{P}_n) \tilde{p}_n, P_n D^{\bar{a}'}_{s,X}, (\tilde{P}_n) \tilde{p}_{n,X} = 0$, $s = 1, \ldots, S$, 并且对于所有 $X \in R \cup Z \cup L$, $\bar{\epsilon}_{X,n} \to 0$, 因为对数似然分量 $P_n \log \tilde{p}_{n,X}$ 随着迭代增加并收敛。

算法 1迭代 TMLE

1: 重复 2: 对于 $X \in R \cup Z \cup L$ 执行 3: 计算聪明协变量 $H^I_{s,X} = D^I_{s,X}(P^0_n)$，它们等于第 s 个结局 $s = 1, \ldots, S$ 以及干预 $\sum_s = (\bar{a}^I_s, \bar{a})$, $(\bar{a}, \bar{a}')$, $(\bar{a}', \bar{a}')$ 的 EIC 分量。 4: 沿多变量子模型 $(1 + \bar{\epsilon}\ \bar{H}) p^{\bar{\epsilon}} X$, $n = \arg\max_{\bar{\epsilon}_X} P_n \log(\ _s, I\ , X H^I_{s,X}) p^0_{n,X}$，求 $\bar{\epsilon}_X = (\epsilon^I_{s,X} : s, I)$ 的最大似然估计。 5: 将 $\bar{\epsilon}_X$, n 代回子模型，定义 $p^0_{n,X}$ 的更新值 $\tilde{p}_{n,X}$。 6: 结束循环 7: 将 p^0_n 替换为更新后的 $\tilde{p}_n$。 8: 直到对于所有 s, I, $P_n D^I_s(\tilde{P}_n) \le \sqrt{\text{var}_n D^I_s(\tilde{P}_n)_n} / \log n$。令 $P^*_n = \tilde{P}_n$。

5.2. 单步 *TMLE*。 我们希望限制搜索最大似然估计的算法在每次迭代中 $\bar{\epsilon}$ 的步长。这应该能提高算法的性能。事实上，对于某个足够小的 $dx > 0$，根据欧几里得范数 $\|\bar{\epsilon}\| < dx$，由于

线性近似，我们可以得到最大似然估计的闭式解

$$\bar{\epsilon}_{X,n}=\left(P_nD^{\bar{a},\bar{a}}_{s,X}(P^0_n),P_nD^{\bar{a},\bar{a}'}_{s,X}(P^0_n),P_nD^{\bar{a}',\bar{a}'}_{s,X}(P^0_n):s=1,\ldots,S\right)dx/\left\|P_n\bar{D}(P^0_n)\right\|$$

，其中 $\bar{\epsilon}=(\epsilon_X:\ X\in\bar{R}\cup\bar{Z}\cup\bar{L})$ 是所有 ϵ_X 的向量，且 $\bar{D}=\left(D^{\bar{a},\bar{a}}_{s,X}(P^0_n)\ D^{\bar{a},}_{s\ X}(P^0_n)\ D^{\bar{a},\bar{a}'}_{s,X}(P^0_n)\ D^{\bar{a}',}_{s\ X}(P^0_n)\ D^{\bar{a}',\bar{a}'}_{s,X}(P^0_n):\ X\in\bar{R}\cup\bar{Z}\cup\bar{L},\ s=1,\ldots\ldots,\ S\right)$ 是所有涉及的 EIC 分量的向量（van der Laan 和 Gruber (2016) 第 8.2 节）。

将 P_{dx} 记为在上述 $\|\bar{\epsilon}\|<dx$ 的限制下对 P 的 TMLE 更新，并将 $P_{2dx},P_{3dx},\ldots$ 记为后续迭代的 TMLE 更新。然后，我们可以设计一个一元通用最有利路径 $P_{ulfm}(P,x)$，其解析表达式如下：

$$p_{ulfm}(p,x)=p\exp\left\{\int_0^x\frac{\left\{P_n\bar{D}(P_{ulfm}(p,u))\right\}^\top\bar{D}(P_{ulfm}(p,u))}{\left\|P_n\bar{D}(P_{ulfm}(p,u))\right\|}du\right\}$$

使得对于所有 $x\in(0,a)$，存在某个 $a>0$，$\frac{d}{dx}P_n\log p_{ulfm}(p^0_n,x)=\left\|P_n\bar{D}(P_{ulfm}(P^0_n,x))\right\|$。此外，$\{P_{jdx}:\ j\}$ 以如下意义逼近通用路径：对于 $j=1$、2、……，$P_{ulfm}(P^0_n,jdx)=P_{jdx}$；通过在 $x=0,dx,2dx,\ldots\ldots$ 处对 $p_{ulfm}(p^0_n,x)$ 进行泰勒展开，可以在温和的正则条件下验证这一点，而对数似然 $P_n\log p_{jdx}$ 则随着 $j\to\infty$ 增大。

我们假设 $\left\{P_{ulfm}(P^0_n,x):\ x\in(0,a)\right\}\subset\mathcal{M}$ 是包含在 $\mathcal{M}$ 中的子模型，并且对数似然 $P_n\log p_{ulfm}(P^0_n,x)\ x\in(0,a)$ 沿着 $x=dx,2dx,\ldots$ 单调不减，并在某个 $x_m<a$ 处达到最接近的局部最大值；对于足够小的 dx 和某个 j_0，有 $j_0dx=x_m$ 和 $\left\|P_n\bar{D}(P_{j_0dx})\right\|=0$。可以选择最终 TMLE 更新为 $P^*_n=P_{Idx}$，并取足够大的 $I=I(n)$，使得 $\left\|P_n\bar{D}(P_{Idx})\right\|=o_P(n^{-1/2})$。这是一个单步过程，用于搜索 $x=0$ 附近 $P_n\log p_{ulfm}(p^0_n,x)$ 的最接近局部最大值 x_m；其中 $\left\|P_n\bar{D}(P)\right\|=0$ 在 $P=P_{ulfm}(P^0_n,x_m)$ 处精确求解，因此称为“单步 TMLE”。

与 $\bar{\epsilon}$ 不受限制的迭代 TMLE 相比，单步 TMLE 在较弱的假设下求解效率得分方程，因为只要 P^0_n 是 $\mathcal{M}$ 的内部点，并且 dx 选取得足够小，P_{jdx} 就会保持在 $\mathcal{M}$ 的内部。单步 TMLE 通过从概念上寻找单变量子模型的局部似然最大值，并在实践中沿已知方向进行更新，而不是迭代求解多维最大似然估计，不仅实现了更高的稳定性，还降低了计算成本。

5.3. 渐近效率和同时推断。对于纵向中介分析，假设目标参数是一个向量 $\bar{\Psi}^F(P)$，其各不

算法 2单步 TMLE

1: **重复** 2: **对于** $X \in R \cup Z \cup L$ **执行**3: 计算巧妙协变量 $H_s^I,\ X = P_n D_{s,\ X}^I(P_n^0)$，对应第 s 个结局 $s = 1,\ \ldots\ldots,\ S$ 以及干预 $I = (\bar{a},\ \bar{a}),\ (\bar{a},\ \bar{a}'),\ (\bar{a}',\ \bar{a}')$。4: **结束循环** 5: **对于** $X \in R \cup Z \cup L$ **执行**6: 定义 $\bar{\epsilon}_X = (\epsilon_s^I,\ X\colon\ s,\ I)$ 的已知多元最大似然估计量，其中 dx 足够小:

$$\epsilon_{s,X,n}^I = H_{s,X}^I dx / \left\| \left(\sum_X H_{s,X}^I : s, I \right) \right\|$$

7: 通过将 $\bar{\epsilon}_{X,n}$ 代回子模型，定义 $\tilde{p}_{n,X}$ 的更新。8: **结束循环**9: 将 p_n^0 替换为更新后的 $\tilde{p}_n$。10: **直到** $P_n D_s^I(\tilde{P}_n) \le \sqrt{\text{方差}_n D_s^I(\tilde{P}_n)_n}\ /\sqrt{n}$ 对所有 s, I 均为 $\log\ n$。令 ~~$P_n^* = \tilde{P}_n$~~。

从 $\left(\Psi_s^{\bar{a},\bar{a}}(P)\ \Psi_{s}^{\bar{a}},\ {}_{\bar{a}'}(P)\ \Psi_{s}^{\bar{a}'},\ {}_{\bar{a}'}(P)\colon\ s = 1, \ldots, S\right)$ 的可微变换，其中 EIC 向量 $\bar{D}^F(P)$ 根据泛函 delta 方法（参见 van der Laan 和 Rose (2011) 的 A.3 以及 van der Laan、Dudoit 和 Keles (2004) 的第 3 节）计算为 $\left(D_s^{\bar{a},\bar{a}}(P)\ D^{\bar{a}},\ {}_{s\ \bar{a}'}(P)\ D^{\bar{a}'},\ {}_{s\ \bar{a}'}(P)\colon s = 1, \ldots, S\right)$ 的函数。本节将 P_n^* 视为迭代式或单步 TMLE，如第 5.1 节和第 5.2 节所述；在满足正则条件时，它可实现 $P_n D^{\bar{a}},\ {}_{s\ \bar{a}}(P_n^*) = o_P(n^{-1/2})$，$P_n D^{\bar{a}},\ {}_{s\ \bar{a}'}(P_n^*) = o_P(n^{-1/2})$，$P_n D^{\bar{a}'},\ {}_{s\ \bar{a}'}(P_n^*) = o_P(n^{-1/2})$，适用于 $s = 1, \ldots, S$。

我们将目标参数的余项定义为

$$\bar{R}^F(P, P_0) = \bar{\Psi}^F(P) - \bar{\Psi}^F(P_0) + P_0 \bar{D}^F(P).$$

目标参数的余项是余项$\left(R_s^{\bar{a},\bar{a}}(P, P_0)\ R^{\bar{a}},\ {}_{s\ \bar{a}'}(P, P_0)\ R^{\bar{a}'},\ {}_{s\ \bar{a}'}(P, P_0)\colon\ s = 1, \ldots, S\right)$ 的函数，分别对应于 $\left(\Psi_s^{\bar{a},\bar{a}}(P)\ \Psi_{s}^{\bar{a}},\ {}_{\bar{a}'}(P)\ \Psi_{s}^{\bar{a}'},\ {}_{\bar{a}'}(P)\colon\ s = 1, \ldots, S\right)$中的各个目标。TMLE P_n^*满足以下展开:

$$\begin{aligned}
\bar{\Psi}^F(P_n^*) - \bar{\Psi}^F(P_0) =& \bar{R}^F(P_n^*, P_0) - P_0 \bar{D}^F(P_n^*) \\
=& (P_n - P_0)\bar{D}^F(P_1) + (P_n - P_0)(\bar{D}^F(P_n^*) - \bar{D}^F(P_1)) \\
& - P_n \bar{D}^F(P_n^*) + \bar{R}^F(P_n^*, P_0). \qquad (4)
\end{aligned}$$

对于 $\bar{\Psi}^F(P)$，在以下条件下可实现 P_n^*的渐近线性:

(B1) EIC 向量 $\left(D_s^{\bar{a},\bar{a}}(P)\ D^{\bar{a}},\ {}_{s\ \bar{a}'}(P)\ D^{\bar{a}'},\ {}_{s\ \bar{a}'}(P)\colon\ s = 1,\ \ldots\ldots S\right)$ 在 $P = P_n^*$处收敛至其在 $P = P_1$ 处的极限，在每个维度上均服从 $L^2(P_0)$ 范数，且属于一个 P_0-Donsker 类;

(B2) $\left(R_s^{\bar{a},\bar{a}}(P_n^0,\ P_0)\ R^{\bar{a},}{}_s{}^{\bar{a}'}(P_n^0,\ P_0)\ R^{\bar{a}',}{}_s{}^{\bar{a}'}(P_n^0,\ P_0)\colon\ s=1,\ \ldots\ldots,\ S\right)$ 均为$o_P(n^{-1/2})$。

在满足以下附加条件时，可达到渐近效率：

(B3) 假设 (B1) 中的极限在 $P_1=P_0$处达到。

随后，我们可以基于渐近线性和正态极限分布构造同时置信区间（Dudoit 和 van der Laan，2008；Rose 和 van der Laan，2018）。请注意，在上述条件下，以下经验协方差矩阵，

$$\bar{\Sigma}(P_n,P_n^*)=P_n\bar{D}^F(P_n^*)\bar{D}^F(P_n^*)^\top,$$

¯提供极限协方差矩阵的一致估计 $\Sigma(P_0,\ P_0)$。因此，可以构造95%同时置信区间，形式为

$$\bar{\Psi}^F(P_n^*)\pm q_{0.95,n}\bar{\sigma}_n/\sqrt{n},$$

其中，$\bar{\sigma}_n$是$\left\{\text{diag}\left(\bar{\Sigma}(P_n,P_n^*)\right)\right\}^{1/2}$的对角元素向量，而$q_{0.95}$、$_n$可以是随机向量 Z 的逐元素绝对值¯ ¯最大值的0.95分位数的蒙特卡洛估计，该随机向量服从均值为0、协方差矩阵为$\left\{\text{diag}\left(\bar{\Sigma}(P_n,P_n^*)\right)\right\}^{-1/2}\bar{\Sigma}(P_n,P_n^*)\left\{\text{diag}\left(\bar{\Sigma}(P_n,P_n^*)\right)\right\}^{-1/2}$的多元正态分布。

6. 数值改进。同时对似然的所有因子进行目标最大似然估计更新所面临的一项计算挑战，是需要反复计算关于条件密度的嵌套积分，如方程（3）所示。本文推导了 EIC 的另一种投影表示，称为 HAL-EIC，并展示了在温和条件下，对 HAL-EIC 进行数值近似可以降低计算成本，同时保留渐近性质。本节中，为了简化符号，我们聚焦于单个实值目标参数的 EIC $D^{\bar{a},\bar{a}'}$。不过，这些方法同样适用于我们所设定的多维目标参数。

¯6.1.基于HAL-EIC的 EIC 数值近似。对于任意变量 $X\in R\cup\bar{Z}\cup\bar{L}$令 $T_X(P)=\left\{f\in L^2(P)\colon\ \mathbb{E}\left[f|\text{Pa}(X)\right]=0\right\}$表示切空间，并对于任意$f\in L^2(P)$定义关于 $L^2(P)$范数的到 $T_X(P)$ 上的投影为 $\prod(f|T_X(P))=\arg\min_{h\in T_X(P)}P\{f-h\}^2$。

引理 6.1（EIC 的投影表示）。将初始映射定义为

$$G_{LR}^{\bar{a},\bar{a}'} = G_{L}^{\bar{a},\bar{a}'} = G_{R}^{\bar{a},\bar{a}'} = \frac{Y\mathbb{I}_{\{\bar{A}=\bar{a}\}}}{\prod_{j=1}^{K} p_A(a_j|Pa(A_j|\bar{a}_{j-1}))} \frac{\prod_{j=1}^{K} p_Z(Z_j|Pa(Z_j|\bar{a}'_j))}{\prod_{j=1}^{K} p_Z(Z_j|Pa(Z_j|\bar{a}_j))}$$

$$G_{Z}^{\bar{a},\bar{a}'} = \frac{Y\mathbb{I}_{\{\bar{A}=\bar{a}'\}}}{\prod_{j=1}^{K} p_A(a'_j|Pa(A_j|\bar{a}'_{j-1}))} \frac{\prod_{j=1}^{K} p_L(L_j|Pa(L_j|\bar{a}_j))}{\prod_{j=1}^{K} p_L(L_j|Pa(L_j|\bar{a}'_j))} \frac{\prod_{j=1}^{K} p_R(R_j|Pa(R_j|\bar{a}_j))}{\prod_{j=1}^{K} p_R(R_j|Pa(R_j|\bar{a}'_j))}.$$

以下投影表示适用于 $t=1$、……、 K:

$$D_{L_t}^{\bar{a},\bar{a}'}(P) = \prod(G_{L}^{\bar{a},\bar{a}'}(P)|T_{L_t}(P))$$

$$D_{R_t}^{\bar{a},\bar{a}'}(P) = \prod(G_{R}^{\bar{a},\bar{a}'}(P)|T_{R_t}(P))$$

$$D_{Z_t}^{\bar{a},\bar{a}'}(P) = \prod(G_{Z}^{\bar{a},\bar{a}'}(P)|T_{Z_t}(P)).$$

证明见附录B。

请注意，引理6.1中的投影项可视为切空间子空间中的真实风险最小化解，前提是将 P视为“真实”分布，并且风险 $\left\|G^{\bar{a},\bar{a}'}_{X} - f\right\|_P^2 = P\left\{G^{\bar{a},\bar{a}'}_{X} - f\right\}^2$对所有 $f \in T_X(P) \subset L^2(P)$定义。给定一个服从 P的独立同分布样本，可通过在包含真实 EIC 的函数类上求经验风险最小化解，来近似这些投影项。此处我们采用一个额外的正则条件，以引入 EIC 的 HAL 近似并实现快速收敛率（Bibaut 和 van der Laan,2019）：

(B4)对于 $X \in R, Z, L$，在真实分布 P_0、初始估计量 P_n^0以及任意TMLE更新 $\tilde{P}_n$处，相应的 EIC 分量均为具有有界截面变差范数的右连续且有左极限函数。

在条件 (B4) 下，我们构造中心化 HAL 基如下

$$\phi_{j,X}(P) = \mathbb{I}_{\{\text{Pa}(X)\geq \text{Pa}(x)(\mu_j)\}}\left(\mathbb{I}_{\{X\geq x(\mu_j)\}} - P(X \geq x(\mu_j)|\text{Pa}(X))\right),$$

其中，$\{u_j:\ j\}$是（X, Pa(X)）上的结点集合，而 u_j =(x(u_j), Pa(x)(u_j)) 是相应的子向量。请注意，这些中心化 HAL 基满足（正如我们在附录 C 中所示）：1） $T_X(P)$由集合 $\{\phi_{j,X}(P):\ j\}$ 张成；2）给定一个来自 P、大小为 N 的独立同分布样本，采用 $G^{\bar{a},\bar{a}'}_{X}(P)$ 的 Lasso 回归，并使用 $\{\phi_{j,X}(P):\ j\}$ 的有限子集，且其截面变差范数的界通过交叉验证确定，可达到有保证的收敛速率，

其中，Lasso 估计量

$$\tilde{D}^{\bar{a},\bar{a}'}_X(P) = \sum_{j=1}^{J} \hat{\beta}^{\bar{a},\bar{a}'}_{j,X}(P)\phi_{j,X}(P)$$

使用 $\left\|\tilde{D}^{\bar{a},\ \bar{a}'}_{X}(P) - D^{\bar{a},\ \bar{a}'}_{X}(P)\right\|_P = O_P(N^{-1/3}(\log N)^{d/2})$近似 EIC。我们将这一数值近似称为

$$\tilde{D}^{\bar{a},\bar{a}'}(P) = \sum_{X\in\bar{R}\cup\bar{Z}\cup\bar{L}} \tilde{D}^{\bar{a},\bar{a}'}_X(P)$$

$D^{\bar{a},\bar{a}'}(P)$的 HAL-EIC。下文中，我们分别将以 HAL-EIC 替代 EIC 的迭代 TMLE 和单步 TMLE 称为迭代 HAL-EIC TMLE 和单步 HAL-EIC TMLE。

请注意，近似 $\tilde{D}^{\bar{a},\ \bar{a}'}_{X}(\cdot) \approx D^{\bar{a},\ \bar{a}'}_{X}(\cdot)$ 既可针对初始估计 P_n^0 获得，也可针对 TMLE 更新 $\tilde{P}_n$获得。但 LASSO 估计量的系数 $\hat{\beta}^{\bar{a},\ \bar{a}',}_{j\ X}(P)$ 是 P的函数，因此，对 β 系数的估计需要一个由具有联合分布 P的 N随机向量构成的独立同分布样本。重抽样规模 N 可以选择为大于观测样本量 n。

为实现计算速度更快的 HAL-EIC 目标最大似然估计更新，我们注意到，随着 P 的变化，并不需要实时地对所有 P 进行大小为 N 的数据重采样。相反，我们通过以下方式定义采用延迟系数估计的 HAL-EIC：

$$\tilde{D}^{\bar{a},\bar{a}'}_{X,P_n^0}(P) = \sum_{j=1}^{J} \hat{\beta}^{\bar{a},\bar{a}'}_{j,X}(P_n^0)\phi_{j,X}(P),$$

and

$$\tilde{D}^{\bar{a},\bar{a}'}_{P_n^0}(P) = \sum_{X\in\bar{R}\cup\bar{Z}\cup\bar{L}} \tilde{D}^{\bar{a},\bar{a}'}_{X,P_n^0}(P),$$

其中，我们相对于给定的初始估计值 P_n^0，保持 β 系数不变。只有当 P_n^0 的值发生变化时，才会对这些 β 系数进行重新抽样和重新估计。对于快速版 HAL-EIC TMLE，将 $\tilde{P}_n$ 定义为 P_n^0 的（迭代或单步）TMLE 更新，它满足 $P_n\tilde{D}^{\bar{a},\ \bar{a}',P_n^0}_{X}(\tilde{P}_n) = o_P(n^{-1/2})$。重复此过程 I 次迭代，并在每次迭代结束时将 P_n^0 替换为 $\tilde{P}_n$。然后，在满足迭代或单步 TMLE 的相同正则条件时，存在一个足够大的整数 $I = I(n)$，使得在第 I 次迭代时，我们有 $P_n\tilde{D}^{\bar{a},\ \bar{a}'}_{X}(\tilde{P}_n) = P_n\tilde{D}^{\bar{a},\ \bar{a}',\tilde{P}_n}_{X}(\tilde{P}_n) = o_P(n^{-1/2})$，并且，我们将最终 TMLE 更新 P_n^* 定义为第 I 次迭代的 TMLE 更新 P_n。

算法 3HAL-EIC（一步法）目标最大似然估计。

为 $X \in \quad \bar{R} \cup \bar{Z} \cup \bar{L}$ **do**

初始化数值 HAL-EIC $\tilde{D}_s^I$，X，$P_{\text{fixed}}(P_n^0)$，用于 $X \in \bar{R} \cup \bar{Z} \cup \bar{L}$，并拟合固定的 β 系数在从 P固定$= P_n^0$生成的独立同分布样本中，对于第 s个结局 $s = 1, \ldots, S$和干预 $I = (\bar{a}, \bar{a}), (\bar{a}, \bar{a}'), (\bar{a}', \bar{a}')$。计算巧妙协变量 $H_s^I{}_{,X}(P_n^0) = P_n \tilde{D}_s^I{}_{,X,P_{\text{fixed}}}(P_n^0)$。

结束 for

重复

重复

for $X \in \bar{R} \cup \bar{Z} \cup \bar{L}$ **do**

定义已知的多元最大似然估计量 $\bar{\epsilon}_X =$（$\epsilon_{s,}^I X$： s， I）其中 dx足够小：

$$\epsilon_{s,X,n}^I = H_{s,X}^I(P_n^0)dx / \left\| \left(\sum_X H_{s,X}^I(P_n^0) : s, I \right) \right\|$$

8：通过将 $\bar{\epsilon}_X$、n 代入子模型，定义 p_n^0、X 的更新 $\tilde{p}_n$、X。9：结束循环 10：将 p_n^0 替换为更新后的 $\tilde{p}_n$。11：直到对所有 s、I，$\tilde{P}_n$ 的 $P_n \tilde{D}_s^I$、$P_{\text{fixed}} \leq \sqrt{\text{var}^n \tilde{D}_s^I{}_{、}{}_{P_{\text{fixed}}}(\tilde{P}_n)}_n \quad / \log n$。

12：在数值 EIC 对象 $\tilde{D}_s^I$、X、$P_{\text{fixed}}(P)$ 中重新拟合 β 系数，并通过将 P_n^0 刚刚被最新的 TMLE 更新 $\tilde{P}_n$ 替换的位置上的 P 固定$= P_n^0$，同时更新 $H_s^I{}_{、}{}_X(P_n^0) = \tilde{D}_s^I{}_{、}{}_{X、}{}_{P_{\text{fixed}}}(P_n^0)$。13：直到使用更新后的 β 系数，对所有 s、I，$\tilde{P}_n$ 的 $P_n \tilde{D}_s^I$、$P_{\text{fixed}} \leq \sqrt{\text{var}^n \tilde{D}_s^I{}_{、}{}_{P_{\text{fixed}}}(\tilde{P}_n)}_n \quad / \log n$。

HAL-EIC TMLE 的渐近线性和效率在重抽样规模 N 对于套索估计量的 β 系数（见附录D）的以下附加条件下得以保持：

(B5) $N = n$ 和 $P_0 \{p_0 - \tilde{p}_n\}^2 = o_P$（$n^{-1/3}$），或 N（n）增长速度快于 $n^{3/2}$，在强正性条件下，对于某个 $\delta > 0$，p_0（o）$> \delta > 0$ 和 p_n^*（o）$> \delta > 0$ 在其支持集上成立。

6.2. 使用汇总协变量建模。由于维度灾难以及对条件密度进行建模所面临的挑战，在实践中，考虑通过引入汇总协变量来进行降维是很有意义的。例如，对于每个节点 $X \in \{\bar{A}, \bar{R}, \bar{Z}, \bar{L}\}$，假设存在一个向量值确定性函数 C_X：Pa（X）$\mapsto C_X$（Pa（X），该函数能够概括 Pa（X）中的信息，并且希望不会丢失信息，使得对于所有 $P \in \mathcal{M}$:

$$p_X(X|\text{Pa}(X)) = p_X(X|C_X(\text{Pa}(X))). \tag{5}$$

方程（5）可视为对统计模型施加了额外的限制 $\mathcal{M}$，其中每个条件密度中的自变量个数现在可以根据汇总向量的维数来确定

而不必随着时间点的增加而增加。

汇总协变量 C_X 可以通过数据自适应的方式选取，从而在假设（5）下进行分析。例如，对于一个离散化的分类变量 X，其可能取值为1、.....、d_X，值得注意的是，C_X 的一个自然的预言机选择为

$$C_X(\text{Pa}(X)) = (P(X = x|\text{Pa}(X)) : x \in \{1, \ldots, d_X\}),$$

根据迭代期望可知，这满足（5）。直观而言，这一观察与倾向评分匹配或协变量调整有关（Rosenbaum and Rubin，1983；D'Agostino Jr，1998）。尽管在实践中无法观测到这样的 C_X，且使用估计的条件概率也可能引入偏差，但可以在 C_X 中加入额外项，同时利用 HAL 将条件密度建模为 C_X的函数，从而保留所需的渐近性质。有关理论细节，请参阅 meta-HAL super-learner 的技术报告。在温和条件下，可以从训练样本中获得汇总协变量 C_X，所得的 CV-TMLE（Zheng and van der Laan，2011；Hubbard, Kennedy and van derLaan，2018）对于感兴趣的目标参数仍是局部有效的。

实际上，条件密度模型也可以通过运用对数据生成过程的实际知识来构建。例如，如果已知开具某种药物的倾向仅取决于特定症状近期的发作情况以及开具处方时已有的病症，那么在这种情况下，C_{A_t} 取向量 $\text{Pa}(A_t)$的一个子集。建议将 $\text{Pa}(A_t)$ 替换为其与 t 之前最近时间点的交互集。然后，g- 计算公式、 EIC 和切空间子空间的相同表达式仍然成立，前几节中的算法也适用。

7. 多重稳健性。若以下三组条件密度估计量中有两组正确，或至少以 $o_P(n^{-1/4})$ 范数达到一致性误差率，则所提出的估计量具有多重稳健性（ Díaz 和 van der Laan，2017；Luedtke 等，2017）： 1) p^0_{n,A_t}， 2) p^0_{n,Z_t}， 3) p^0_{n,R_t} 和 p^0_{n,L_t}。此时，$\left(R_s^{\bar{a},\bar{a}}(P, P_0)\ R^{\bar{a}}、\ {}_s{}^{\bar{a}'}(P, P_0)\ R^{\bar{a}'}、\ {}_s{}^{\bar{a}'}(P, P_0)\text{：}\ s = 1，\ .....，\ S\right)$ 中定义的余项均为 $o_P(n^{-1/2})$（见附录 E）。然而，在实践中，即使在随机对照试验中，中介分析的多重稳健性条件也并非轻易就能满足，因为只有 p^0_{n,A_t} 才能保证

正确或得到一致估计。因此，建议将 HAL 纳入针对 P_n^0的超级学习器估计量之一，以满足所有误差率条件（Bibaut 和 van der Laan，2019；van der Laan、Polley 和 Hubbard，2007；van der Laan、Dudoit 和 Keles，2004）。

作为当前设定的一种变体，如果定义一个数据自适应框架，使中介随机干预 $\Gamma_t^{\bar{a}'}$ $(z_t|\bar{r}_t,\bar{z}_{t-1},\bar{l}_{t-1}) = p(Z_t(\bar{a}') = z_t|\bar{R}_t(\bar{a}') = \bar{r}_t, Z_{t-1}(\bar{a}') = \bar{z}_{t-1}, L_{t-1}(\bar{a}') = l_{t-1}) = p_{Z_t}(z_t|\text{Pa}(z_t|\bar{a}'_t))$ 被替换为例如估计得到的对照组中介分布 $\Gamma_n^{\bar{a}'},_t(z_t|\bar{r}_t,\bar{z}_{t-1},\bar{l}_{t-1}) = p^0_{n,Z_t}(z_t|\text{Pa}(z_t|\bar{a}'_t))$，那么反事实 $X(\bar{a},\bar{\Gamma}_n^{\bar{a}'})$ 和由此得到的目标参数也会变为数据自适应（Hubbard、Kennedy 和 van der Laan，2018）。若进一步假设 $\Gamma_t^{\bar{a}'} = \Gamma_n^{\bar{a}'},_t$，那么这就是 van der Laan 和 Petersen（2008）方法向纵向数据的推广。在这些推广中，对于已知治疗随机化和动态规则的随机试验，多重稳健性条件可能会得到可靠满足；但对于新的推断目标，其解释有所不同，因为这些目标现在取决于中介干预的选择。此类广义随机直接效应和间接效应的影响曲线及其实施方法见附录 A.1。

8. 模拟。在本节中，我们研究所提出算法在模拟数据中的性质。在本节中，我们聚焦于以下数据结构，

$$O = (L_{01}, L_{02}, A_1^C, A_1^E, R_1, Z_1, Y_1, A_2^C, A_2^E, R_2, Z_2, Y_2) \sim P_0.$$

我们聚焦于生存结局 $Y = (Y_1, Y_2)$，其中 Y_1 和 Y_2 为二元变量，且事件 $\{Y_1 = 1\}$ 蕴含 $\{Y_2 = 1\}$。我们关注多元参数$\left(\mathbb{E}\left[\bar{Y}(\bar{1},\Gamma^{\bar{1}})\right], \mathbb{E}\left[\bar{Y}(\bar{1},\Gamma^{\bar{0}})\right], \mathbb{E}\left[\bar{Y}(\bar{0},\Gamma^{\bar{0}})\right]\right)$。我们使用精确EIC 和 HAL-EIC 计算目标最大似然估计（采用受限步长的单步TMLE），并使用不同的正确设定或错误设定的初始估计量计算 g-公式插件估计。此小节旨在验证基于精确EIC或HAL-EIC 的目标最大似然估计的一致性、渐近线性和多重稳健性。

在每次迭代中，我们根据以下数据生成过程生成一个大小为 $n = 1000$ 的独立同分布样本：

L_{01} ~伯努利(0.4)

L_{02} ~伯努利(0.6)

A_t^C ~伯努利(expit($1.5 - 0.4L_{01} - 0.8L_{02} + \mathbb{I}_{\{t>1\}}0.5A_{t-1}^E$))$A_t^E$ ~伯努利(expit(λ($-0.55 + 0.35L_{01} + 0.6L_{02} - \mathbb{I}_{\{t>1\}}0.05A_{t-1}^E$)))$R_t$ ~伯努利(expit($-0.8 + 0.1L_{01} + \mathbb{I}_{\{t=1\}}0.3L_{02} + \mathbb{I}_{\{t>1\}}0.3R_{t-1} + A_t^E$))$Z_t$ ~伯努利(expit($-0.25 + 0.4L_{02} + 0.4A_t^E + 0.5R_t$))$Y_t$ ~伯努利(expit($0.05 + 0.375L_{02} + 0.25R_t - 0.075A_t^E - 0.075Z_t - \mathbb{I}_{\{t>1\}}0.025R_{t-1}$))，其中我们还将倾向评分缩放因子 λ 的值从1变为5，以模拟不同程度的有限样本下正值性假设近似违背情况。每种情景迭代1000次，详细结果见附录F。

8.1. 多重稳健性。我们给出了模拟研究的结果，该研究是对精确 EIC TMLE 的基本多重稳健性以及HAL-EICTMLE可比性能的概念验证研究。模型设定错误被强制设置为：1）无条件密度估计器；2）初始 p_A 估计器；3）初始 p_Z 估计器；4）初始 p_Y估计器。正确的条件密度模型通过正确的主项逻辑回归进行拟合。模型设定错误的模型将给定过去信息的每个变量的条件期望设为观测样本均值，并额外加入 0.05 的偏差，同时将其限制在(0.01,0.99)之间。基于精确 EIC和 HAL-EIC 的目标最大似然估计在四种情景中均表现出相似的性能（见表 1-4）。

8.2. 有限样本正值性挑战。 随着我们将正值性参数 λ从 1 调整到 5，真实的处理倾向得分越来越接近 0，因此在有限样本中，我们发现初始估计值 p_n^0和后续更新值 $\tilde{p}_n$ 违反了有界性假设（B2）、（B4）和（B5）。

有趣的是，在所有情景（$\lambda = 1$、2、3、4、5）以及目标参数的所有维度下，与精确 EIC TMLE 相比，HAL-EIC TMLE 的均方误差增幅更小，置信区间覆盖率降幅也更小（图1-2）。这体现了基于 HAL 的投影表示的一个潜在优势：HAL 算法会自动寻找有界的 EIC 近似，同时保持最终 TMLE 估计的渐近线性。这是一项理想特性，也比任意设定 IPW 值或影响曲线估计的界限更具吸引力。后者往往会造成渐近偏差，而 HAL-EIC 能够保持所需的多重稳健性。

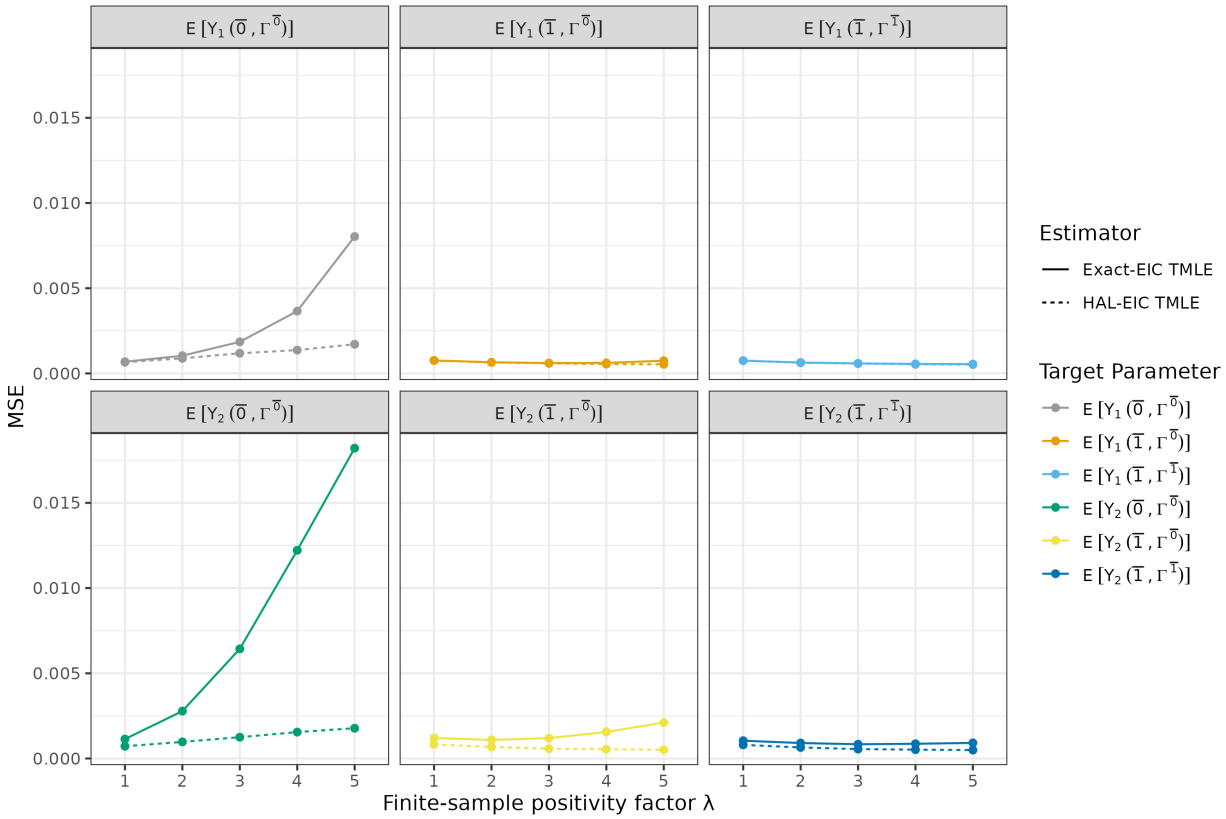


图 1. 均方误差比较：随着有限样本正值性违背程度增加。

仍需进一步研究 HAL-EIC 在更复杂情景下的表现，例如罕见事件与有限样本正值性违背相结合的情形。这项工作以及进一步的数值优化将在未来的研究中加以解决。

9.讨论。 在本文中，我们构建了一个可扩展的、基于似然和随机干预的纵向中介估计框架。我们框架的一个优势在于，作为一种基于似然的方法，它具有灵活性。这带来了一个能够同时针对多维参数的估计程序，其方法是更新纵向似然过程的同一估计量的估计器。此外，我们的 HAL-EIC近似进一步减轻了实践者的负担，因为算法所需的输入减少为对条件分布的建模约束，这对于数据科学家而言是一项众所周知的任务，并且无需进一步理解 EIC 的解析计算或序贯回归方程即可实现。

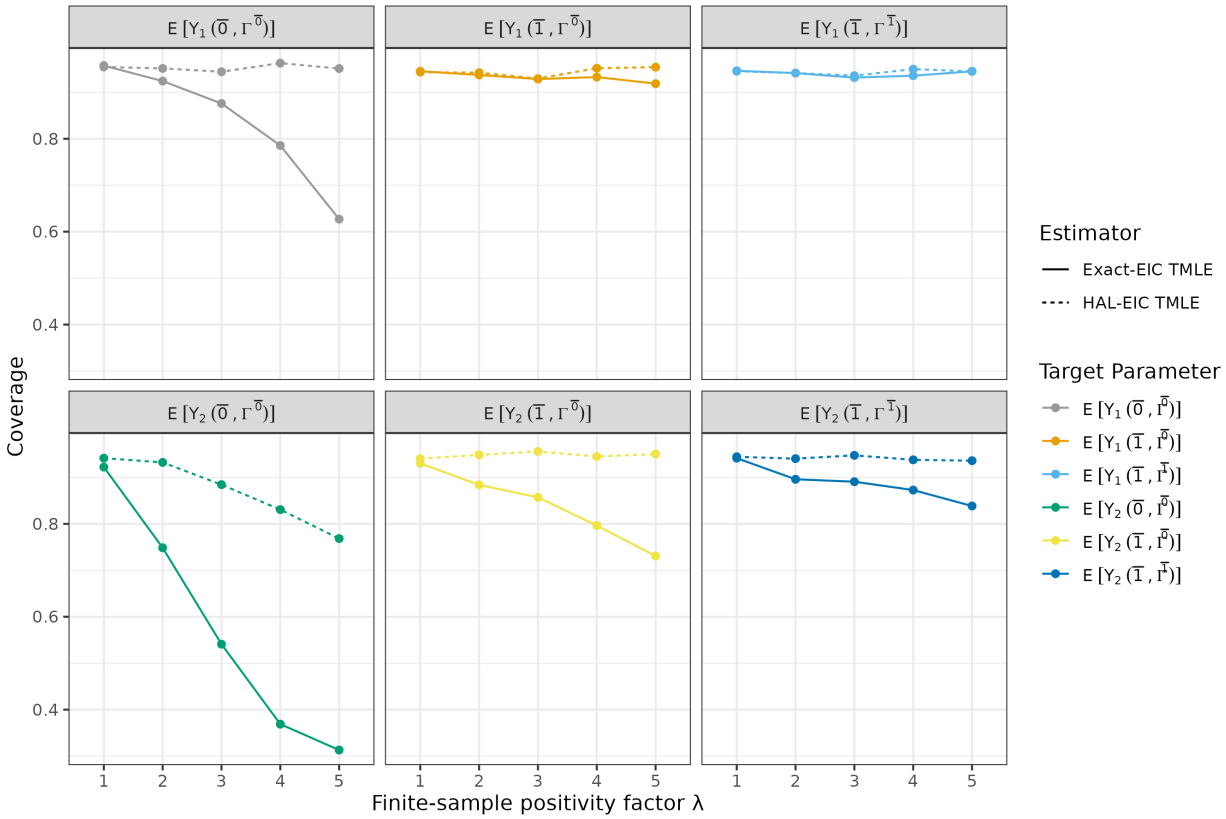


图 2. 随着有限样本正值性违背加剧的覆盖率比较。

我们这项工作的一项重要拓展，是应用协作式目标最大似然估计（van der Laan 和 Gruber，2010），这在有限样本正值性违背的纵向问题中可能至关重要（Petersen 等人，2011）。虽然第 6.1 节提出的数值 HAL-EIC 表示法隐式地寻求 EIC 各分量截面变差范数的最优界限，并在模拟中展现了稳健性，但将协作式目标最大似然估计与所提出的估计量结合起来，或许能够实现更加可靠的推断。先前的研究结果表明，这是优化有限样本性能的一个有前景的方向（Ju、Schwab 和 van der Laan，2019）。

另一个有待深入研究的方向，是针对高维时变协变量的应用，自适应地选择降维算法。这样可以在保持目标参数估计量渐近性质的同时，进一步提高可扩展性。

未来的研究还应考虑将连续时间 TMLE 推广(Rytgaard、Gerds 和 van der Laan，2021），以应用于事件连续发生并以连续时间记录的数据结构，从而开展纵向中介分析。应采用高阶 TMLE（van der Laan、Wang 和 van der Laan，2021）来提高效率。

致谢。 本研究部分由NIH资助项目 R01AI074345-10A1 提供资金。诺和诺德向加州大学伯克利分校靶向机器学习与因果推断中心提供了一笔慈善捐赠，用于额外资助。

附录 A：高效影响曲线（EIC）

为证明第 3.4 节中的 EIC 表示，我们利用$\Psi^{\bar{a}、\ \bar{a}'}_{s}(P)$的路径可微性，并得到（省略其对 $\bar{a}$、$\bar{a}'$、s 的依赖，并令 $Y=\psi_s(\bar{L}_K)$）

$$\frac{\mathrm{d}\Psi(P_\epsilon^h)}{\mathrm{d}\epsilon}_{|\epsilon=0}=\langle D(P),h\rangle_P$$

对于所有子模型 P_ϵ^h至 P，满足 $h=\mathrm{d}\,\mathrm{d}\epsilon|_{\epsilon=0}\log p_\epsilon^h$。注意， $h\in T(P)$可以分解为 $h=\sum_X h_X$，其中 $X\in L\cup R\cup Z$，这对应于子模型的因子分解

$$p_\epsilon^h=\prod_{X\in\bar{L}\cup\bar{R}\cup\bar{Z}}p_{\epsilon,X}^h=\prod_{X\in\bar{L}\cup\bar{R}\cup\bar{Z}}p_X(1+\epsilon h_X).$$

对于所有满足 $h_{Z_t}=0$ 的 $h\in T(P)$，对所有 $t=1,\ldots,K$，我们有

$$\begin{aligned}\frac{\mathrm{d}\Psi(P_\epsilon^h)}{\mathrm{d}\epsilon}_{|\epsilon=0}=&\frac{\mathrm{d}\Psi(\prod_{t=1}^K p_{Z_t}\prod_{X\in\bar{L}\cup\bar{R}}p_{\epsilon,X}^h)}{\mathrm{d}\epsilon}_{|\epsilon=0}\\=&\frac{\mathrm{d}\sum_{\bar{l},\bar{z},\bar{r}}y\prod_{X\in\bar{L}\cup\bar{R}}p_{\epsilon,X}^{h_X}(x|\mathrm{Pa}(x|\bar{a}))\prod_{t=1}^K\mathbb{I}_{\{A_t=a_t\}}\prod_{t=1}^K p_{Z_t}(z_t|\mathrm{Pa}(z_t|\bar{a}'_t))}{\mathrm{d}\epsilon}_{|\epsilon=0}\\=&\langle\sum_{X\in\bar{L}\cup\bar{R}}D_X(P),\sum_{X\in\bar{L}\cup\bar{R}}h_X\rangle_P\end{aligned}$$

这对应于干预 $A_t=a_t$和 $Z_t\sim p_{Z_t}(Z_t|\mathrm{Pa}(Z_t|\bar{a}'))$ 下治疗特异性均值的路径导数。应用 $\Gamma_t(Z_t|\mathrm{Pa}(Z_t))=p_{Z_t}(Z_t|\mathrm{Pa}(Z_t|\bar{a}'))$ 于引理 A.1。由此得到高效影响函数 $\sum_{X\in\bar{L}\cup\bar{R}}D_X(P)$，如第 3.4 节所述。

另一方面，要计算 $D_{Z_t}(P)$，考虑 $h\in T(P)$，使得对于所有 $X\in L\cup R$均有 $h_X=0$。然后

$$\begin{aligned}\frac{\mathrm{d}\Psi(P_\epsilon^h)}{\mathrm{d}\epsilon}_{|\epsilon=0}=&\frac{\mathrm{d}\sum_{\bar{l},\bar{z},\bar{r}}y\prod_{t=1}^K\mathbb{I}_{\{A_t=a'_t\}}\prod_{X\in\bar{L}\cup\bar{R}}p_X(x|\mathrm{Pa}(x|\bar{a}))\prod_{t=1}^K p_{\epsilon,Z_t}^{h_{Z_t}}(z_t|\mathrm{Pa}(z_t|\bar{a}'_t))}{\mathrm{d}\epsilon}_{|\epsilon=0}\\=&\langle\sum_{t=1}^K D_{Z_t}(P),\sum_{t=1}^K h_{Z_t}\rangle_P,\end{aligned}$$

这对应于在干预 下治疗特异性均值的 $A_t=a'_t$路径导数，其中干预 $X\sim p_X(X|\mathrm{Pa}(X|\bar{a}))$对所有 $X\in L\cup R$均成立。由此证明了第 $\sum_{t=1}^K D_{Z_t}(P)$ 节 3.4 中其余 EIC 分量。

最后，请注意，正交分解给出$\langle D_{Z_t}(P)\sum_{X\in\bar{L}\cup\bar{R}}h_X\rangle_P=0$以及$\langle D_{R_t}(P)+D_{L_t}(P)\sum_{X\in\bar{Z}}h_X\rangle_P=0$。因此，$\frac{\mathrm{d}\Psi(P_\epsilon^h}{\epsilon}_{|\epsilon=0}=\langle\sum_{t=1}^K D_{Z_t})\mathrm{d}(P)+D_{R_t}(P)+D_{L_t}(P)h\rangle_P$对所有 $h\in T(P)$。这证明了 $D(P)=\sum_{X\in\bar{L}\cup\bar{R}\cup\bar{Z}}D_X(P)$

是梯度，因此也是非参数切空间中的唯一典范梯度$T(P)$。

A.1. 其他纵向干预的 EIC。

受控直接效应（CDE）。假设我们将 2.1 中的随机干预$Z_t(\bar{a},\bar{\Gamma}^{\bar{a}'}) \sim \Gamma_t^{\bar{a}'}$替换为强制进行一次随机抽取 $Z_t(\bar{a},\bar{\Gamma}) \sim \Gamma_t$，且该抽取不依赖于对照组的反事实。一个典型的选择是令 Γ_t 为退化的离散密度函数，其以概率 1 取某个特定值 z_t；在这种情况下，$Y(\bar{a},\Gamma)$ 的 g-计算¯公式与静态干预下 $Y(\bar{a},\bar{z})$ 的公式相同。另一个例子是，我们认为某个估计量 Γ_t 可以很好地代表未观测到的$\bar{\Gamma}^{\bar{a}'}$，于是关注 $Y(\bar{a},\bar{\Gamma}=\bar{\Gamma}\hat{})$，尽管需要注意，当¯ˆΓ 与 $\bar{\Gamma}^{\bar{a}'}$偏离过大时，其解释可能受到限制。一般而言，这些对应于对 $\bar{A}$ 和Z施加某种联合干预，其中（随机/随机化）CDE,

$$\mathbb{E}\left[Y(\bar{a},\bar{\Gamma}) - Y(\bar{a}',\bar{\Gamma})\right],$$

在对 $\bar{A}$ 和 $\bar{Z}$ 进行随机干预时，变为标准的平均处理效应（ATE）参数。对于任一 $\Psi(P) = \mathbb{E}\left[Y(\bar{a},\bar{\Gamma})\right]$ 或 $\mathbb{E}\left[Y(\bar{a}',\bar{\Gamma})\right]$，$p_{Z_t}$均可视为干扰项，因为任何带有得分 $h \in T_Z(P)$ 的子模型 P_ϵ^h 都会导致 $\mathrm{d}\Psi(P^h_\epsilon)\,\mathrm{d}\epsilon = 0$。这将导致 $D(P) \perp \overline{T_Z(P)}$ 以及 $D_Z(P) = 0$。对于 $X \in L \cup R$，只需将 $Z_t \sim p_{Z_t}(Z_t|\mathrm{Pa}(Z_t|\bar{a}'))$ 替换为 $Z_t \sim \Gamma_t(Z_t|\mathrm{Pa}(Z_t))$，即可对 $X \in L \cup R$ 应用类似的 EIC 分解证明（见引理 A.1）。NIE 和自然直接效应的证明方法相同，只是此时 EIC $D(P)$ 具有 $D_{Z_t} = 0$，并且在 $D_{R_t}(P)$ 和 $D_{L_t}(P)$ 中，由已知且固定的函数 $\Gamma_t(Z_t|\mathrm{Pa}(Z_t))$ 替代 $p_{Z_t}(Z_t|\mathrm{Pa}(Z_t|\bar{a}'))$。

一般纵向随机干预。对于一般纵向随机干预（可以是固定干预和随机干预的混合）的平均处理效应，假设 A 表示静态干预节点，Z 表示第2.1 节 SCM 上的随机干预节点。在对比$(\bar{a},\bar{\Gamma}^{(1)})$和$(\bar{a}',\bar{\Gamma}^{(2)})$时，若 $\bar{\Gamma}^{(1)}$ 和 $\bar{\Gamma}^{(2)}$互不相同，则可将此混合干预下的广义 ATE 定义为

$$\mathbb{E}\left[Y(\bar{a},\bar{\Gamma}^{(1)}) - Y(\bar{a}',\bar{\Gamma}^{(2)})\right].$$

对于 $\bar{\Gamma} = \bar{\Gamma}^{(1)}$或 $\bar{\Gamma}^{(2)}$，我们有以下关于 $\Psi(P) = \mathbb{E}\left[Y(\bar{a},\bar{\Gamma})\right]$的引理。

引理 *A.1.* 对于 $\bar{X} = \bar{R} \text{ or } \bar{L}$ *and* $\Psi(P) = \mathbb{E}\left[Y(\bar{a}, \bar{\Gamma})\right]$，其中 $\bar{\Gamma}$ 是一个已知的密度函数集合，用于 Z （其含义是 $Z(O)$ 是 O 的固定函数），假设 $\mathbb{E}\left[Y(\bar{a}, \bar{\Gamma})\right]$ 可由与方程 *2* 类似的 *G*计算公式识别，但将 $p_Z(Z_t|Pa(Z_t|\bar{a}'))$ 替换为 $\Gamma_t(Z_t|Pa(Z_t))$。定义。

$$
\begin{aligned}
Q_{R_{K+1}}^{\bar{a},\bar{\Gamma}} =& Y \\
Q_{L_t}^{\bar{a},\bar{\Gamma}}\left(\bar{R}_t, \bar{Z}_t, \bar{L}_{t-1}\right) =& \mathbb{E}_{P^{\bar{a},\bar{\Gamma}}}\left[Q_{R_{t+1}}^{\bar{a},\bar{\Gamma}} | \bar{R}_t, \bar{Z}_t, \bar{L}_{t-1}\right] = \mathbb{E}_P\left[Q_{R_{t+1}}^{\bar{a},\bar{\Gamma}} | \bar{R}_t, \bar{Z}_t, \bar{L}_{t-1}, \bar{A}_t = \bar{a}_t\right] \\
Q_{Z_t}^{\bar{a},\bar{\Gamma}}\left(\bar{R}_t, \bar{Z}_{t-1}, \bar{L}_{t-1}\right) =& \mathbb{E}_{P^{\bar{a},\bar{\Gamma}}}\left[Q_{L_t}^{\bar{a},\bar{\Gamma}} | \bar{R}_t, \bar{Z}_{t-1}, \bar{L}_{t-1}\right] = \sum_{z_t} Q_{L_t}^{\bar{a},\bar{\Gamma}} \Gamma_t(z_t|Pa(Z_t)) \\
Q_{R_t}^{\bar{a},\bar{\Gamma}}\left(\bar{R}_{t-1}, \bar{Z}_{t-1}, \bar{L}_{t-1}\right) =& \mathbb{E}_{P^{\bar{a},\bar{\Gamma}}}\left[Q_{Z_t}^{\bar{a},\bar{\Gamma}} | \bar{R}_{t-1}, \bar{Z}_{t-1}, \bar{L}_{t-1}\right] = \mathbb{E}_P\left[Q_{Z_t}^{\bar{a},\bar{\Gamma}} | \bar{R}_{t-1}, \bar{Z}_{t-1}, \bar{L}_{t-1}, \bar{A}_t = \bar{a}_t\right] \\
\Psi^{\bar{a},\bar{\Gamma}}(P) =& \mathbb{E}_P Q_{R_1}^{\bar{a},\bar{\Gamma}}(P)(L_0),
\end{aligned}
$$

and

$$
D_{L_t}^{\bar{a},\bar{\Gamma}} = \frac{\mathbb{I}_{\{\bar{A}_t = \bar{a}_t\}}}{\prod_{j=1}^t p_A(a_j|Pa(A_j|\bar{a}_{j-1}))} \prod_{j=1}^t \frac{\Gamma_t(Z_j|Pa(Z_j))}{p_Z(Z_j|Pa(Z_j|\bar{a}_j))} \left\{ Q_{R_{t+1}}^{\bar{a},\bar{\Gamma}}(\bar{R}_t, \bar{Z}_t, \bar{L}_t) - Q_{L_t}^{\bar{a},\bar{\Gamma}}(\bar{R}_t, \bar{Z}_t, \bar{L}_{t-1}) \right\}
$$

$$
D_{R_t}^{\bar{a},\bar{\Gamma}} = \frac{\mathbb{I}_{\{\bar{A}_t = \bar{a}_t\}}}{\prod_{j=1}^t p_A(a_j|Pa(A_j|\bar{a}_{j-1}))} \prod_{j=1}^{t-1} \frac{\Gamma_t(Z_j|Pa(Z_j))}{p_Z(Z_j|Pa(Z_j|\bar{a}_j))} \left\{ Q_{Z_t}^{\bar{a},\bar{\Gamma}}(\bar{R}_t, \bar{Z}_{t-1}, \bar{L}_{t-1}) - Q_{R_t}^{\bar{a},\bar{\Gamma}}(\bar{R}_{t-1}, \bar{Z}_{t-1}, \bar{L}_{t-1}) \right\}.
$$

那么，$\Psi(P)$ 的 *EIC* 为 $D^{\bar{a},\bar{\Gamma}}(P) = \sum_{X_t \in \bar{R} \cup \bar{L}} D^{\bar{a},}_{X_t,\bar{\Gamma}}(P)$。

附录 B：EIC 的投影表示法

对于引理 6.1，回顾一下，在附录A中我们证明了 $\sum_{X \in \bar{R} \cup \bar{L}} D_X(P)$ 是在干预 $A_t = a_t$ 和$Z_t \sim p_{Z_t}(Z_t|\text{Pa}(Z_t|\bar{a}'))$ 下治疗特异性均值（TSM）的影响曲线。对于此类 TSM 参数，在假设 p_A 和 p_Z 已知的半参数模型中，可以从 IPW 估计量 $\frac{1}{n}\sum_{t=1}^K \frac{1_{\{\bar{A}=\bar{a}\}}}{\prod_{j=1}^K p_A(a_j|\text{Pa}(A_j|\bar{a}_{j-1}))} \cdot \frac{\prod_{j=1}^K p_Z(Z_j|\text{Pa}(Z_j|\bar{a}'_j))}{\prod_{j=1}^K p_Z(Z_j|\text{Pa}(Z_j|\bar{a}_j))}$ 推导出影响曲线（vander Laan，2018）。然后，非参数模型中的影响函数是该半参数影响曲线到切空间上的投影。因此，对于 $X \in R \cup L$,

$$
D_X^{\bar{a},\bar{a}'}(P) = \prod\left(\prod\left(G_{LR}^{\bar{a},\bar{a}'}|L_0^2(P)\right)|T_X(P)\right) = \prod\left(G_{LR}^{\bar{a},\bar{a}'}|T_X(P)\right),
$$

其中 $L_0^2(P)$ 是满足 $f(O)$ 的有限方差函数 $\mathbb{E}_P[f(O)] = 0$的集合。同样适用于 $\sum_{X \in \bar{Z}} D_X(P)$，它是在干预 $A_t = a'_t, R_t \sim p_{R_t}(R_t|\text{Pa}(R_t|\bar{a}))$下作为 TSM 的影响曲线推导得到的，以及 $L_t \sim p_{L_t}(L_t|\text{Pa}(L_t|\bar{a}))$。

引理 6.1 也可以通过使用 $\prod(f|T_X(P)) = \mathbb{E}_P[f|X,\text{Pa}(X)] - \mathbb{E}_P[f|\text{Pa}(X)]$ 并展开条件期望的积分项，通过代数计算加以验证。

附录 C：中心化 HAL 基

假设 $D^*_X(P)$ 是第3.4 节中列出的与切空间 $T_X(P) = \{h(X|\text{Pa}(X)) : \mathbb{E}_P\{h(X|\text{Pa}(X))|\text{Pa}(X)\}) = 0\}$相关联的 EIC 组成部分之一。在本节中，我们将 $D^*_X(P)$ 表示为中心化 HAL 基函数的线性组合，并介绍基于该近似的 TMLE 更新。

如果我们有一个初始梯度 $G_X(P)$，使得其在切空间$\prod(G_X(P)|T_X(P))$上的投影等于 EIC 分量 $D^*_X(P)$（例如引理 6.1），并且如果 $T_X(P)$可以由一组基函数 $\{\phi_{j,X}(P) : j\}$的线性张成空间很好地近似，那么我们可以得到以下 EIC 表示

$$D^*_X(P) = \sum_j \beta_{j,X}(P)\phi_{j,X}(P),$$

其中，系数 $\beta_X(P) = (\beta_{j,X}(P) : j)$ 由最小平方投影定义

$$\beta_X(P) = \arg\min_\beta P\left\{G_X(P) - \sum_j \beta_j \phi_{j,X}(P)\right\}^2.$$

在实践中，可以从 P生成一个样本量为 N的大样本，再针对 $\{\phi_{j,X}(P) : j\}$ 的一个子集进行交叉验证 Lasso 回归。该子集的大小为 J，这将决定以下近似表示中至多有 $N-1$ 个非零系数；该表示具有有限的截面变差范数

$$\hat{D}^*_X(P) = \sum_{j=1}^{J} \hat{\beta}_{j,X}(P)\phi_{j,X}(P).$$

在假设 (B4)，$T_X(P)$ 是$(X,\text{Pa}(X))$ 上右连续左极限函数空间的子空间，且这些函数具有有界的截面变差范数；后者是由以下形式的（未中心化）HAL 基张成的空间（参见 van derLaan 等人(2018)）：

$$\phi_{j,X} = \mathbb{I}_{\{U(X)\geq u_j\}},$$

其中 $U(X) = (X,\text{Pa}(X))$。u_j是 $U(X)$某个子向量取值范围中的一个节点，其中 u_j 的每个元素都取值于 $\mathbb{R}\cup\{-\infty\}$。分两步将其投影到 $T_X(P)$ 上，得到

两步给出

$$D_X^*(P) = \sum_j \beta_{j,X}(P)\phi_{j,X}(P),$$

其中 $\phi_{j,X}(P) = \Pi(\phi_{j,X}|T_X(P)) = \phi_{j,X} - \mathbb{E}_P[\phi_{j,X}|\text{Pa}(X)] = \mathbb{I}_{\{\text{Pa}(X)\geq \text{Pa}(x)(\mu_j)\}}\left(\mathbb{I}_{\{X\geq x(\mu_j)\}} - P(X \geq x(\mu_j)|\text{Pa}(X))\right)$. 根据有限样本 HAL 估计量的构造（van der Laan 等，2018），使用由 P生成的大小为 N 的独立同分布样本，可以从 $\{\phi_{j,X}(P)$ 中选取一个大小为 J 的有限子集，使得相应的交叉验证 Lasso 估计量 $\hat{}D_X^*(P) = \sum_j^J \hat{}\beta_{j,X}(P)\phi_{j,X}(P)$ 满足收敛速率 $\left\|\hat{}D_X^*(P) - D_X^*(P)\right\|_P = O_P(N^{-1/3}(\log N)^{d/2})$。

附录 D：HAL-EIC 目标最大似然估计

注意，在该展开式（4）中仅增加了一项，

$$\begin{aligned}&P_n\left\{\boldsymbol{D}^F(P_n^*) - \hat{\boldsymbol{D}}^F(P_n^*)\right\}\\ =&(P_n - P_0)\left\{\boldsymbol{D}^F(P_n^*) - \hat{\boldsymbol{D}}^F(P_n^*)\right\} + (P_0 - P_n^*)\left\{\boldsymbol{D}^F(P_n^*) - \hat{\boldsymbol{D}}^F(P_n^*)\right\}.\end{aligned}$$

为保持第 5.3 节所列假设下实现的渐近效率，1）第一部分在类似的 Donsker 条件下收敛；2）根据已知的 HAL 误差率（Bibaut 和 van der Laan，2019），$\left\|D_X^*(P) - \hat{}D_X^*(P)\right\|_P = O_P(N^{-1/3}(\log N)^{d/2})$，其中 $\|g(O)\|_P = \sqrt{\langle g,g\rangle_P} = \sqrt{Pg^2} = \sqrt{\int g(o)^2 p(o)d\mu}$ 是 $L^2(P)$ 范数，第二部分（假设下文的 $L^2(\mu)$ 范数按元素应用于向量）由柯西-施瓦茨不等式得出。

$$\left\|(P_0 - P_n^*)\left\{\boldsymbol{D}^F(P_n^*) - \hat{\boldsymbol{D}}^F(P_n^*)\right\}\right\| \leq \left\|\|p_0 - p_n^*\|_\mu \cdot \left\|\boldsymbol{D}^F(P_n^*) - \hat{\boldsymbol{D}}^F(P_n^*)\right\|_\mu\right\|,$$

只需要满足 (B5)，也就是说，在我们选择 $\|p_0 - p_n^*\|_{P_0} = o_P(n^{-1/6})$ $N = n$ 时需要满足该条件；或者，在 $N(n)$ $n^{3/2}$ 的增长速度快于 p_0 时，无需额外条件；以上均假设 $p_0(o) > \delta > 0$ 和 $p_n^*(o) > \delta > 0$ 在某个 $\delta > 0$ 的支持集上成立。

在实践中，我们可以使用 $N > n$ 进行模拟，以进一步改善有限样本性能，并且可以将 HAL 作为 P_n^0 的估计器之一纳入超级学习器，从而保证 $\|p_0 - p_n^*\|_{P_0} = O_P(n^{-1/3}(\text{对数}\, n)^{d/2})$ 得到保证。

附录 E：多重稳健性和精确余项

首先，聚焦于其中一个维度 $\Psi(P) = \Psi^{\bar{a}, \quad \bar{a}'}_{s}(P)$ 及其对应的精确余数，记为 $R(P, \ P_0) = R^{\bar{a}, \quad \bar{a}'}_{s}(P, \ P_0)$。然后根据第5.3节的定义，

$$\begin{aligned} R(P, P_0) &= \Psi(P) - \Psi(P_0) + P_0 D(P) \\ &= PQ_{R_1}(P) - P_0 Q_{R_1}(P_0) + \sum_{X \in \bar{R} \cup \bar{Z} \cup \bar{L}} P_0 D_X(P) \end{aligned}$$

定义以下广义倾向项：

$$\begin{aligned} \pi_A^k &= \prod_{t=1}^{k} p_A(A_t|\text{Pa}(A_t)), \ \pi_{0,A}^k = \prod_{t=1}^{k} p_{0,A}(A_t|\text{Pa}(A_t)) \\ \pi_R^k &= \prod_{t=1}^{k} p_R(R_t|\text{Pa}(R_t)), \ \pi_{0,R}^k = \prod_{t=1}^{k} p_{0,R}(R_t|\text{Pa}(R_t)) \\ \pi_Z^k &= \prod_{t=1}^{k} p_Z(Z_t|\text{Pa}(Z_t)), \ \pi_{0,Z}^k = \prod_{t=1}^{k} p_{0,Z}(Z_t|\text{Pa}(Z_t)) \\ \pi_L^k &= \prod_{t=1}^{k} p_L(L_t|\text{Pa}(L_t)), \ \pi_{0,L}^k = \prod_{t=1}^{k} p_{0,L}(L_t|\text{Pa}(L_t)) \\ \pi_A^{*k,\bar{a}} &= \prod_{t=1}^{k} \mathbb{I}_{\{A_t = a_t\}} \\ \pi_A^{*k,\bar{a}'} &= \prod_{t=1}^{k} \mathbb{I}_{\{A_t = a'_t\}} \\ \pi_R^{*k} &= \prod_{t=1}^{k} p_R(R_t|\text{Pa}(R_t|\bar{a}_t)), \ \pi_{0,R}^{*k} = \prod_{t=1}^{k} p_{0,R}(R_t|\text{Pa}(R_t|\bar{a}_t)) \\ \pi_Z^{*k} &= \prod_{t=1}^{k} p_Z(Z_t|\text{Pa}(Z_t|\bar{a}'_t)), \ \pi_{0,Z}^{*k} = \prod_{t=1}^{k} p_{0,Z}(Z_t|\text{Pa}(Z_t|\bar{a}'_t)) \\ \pi_L^{*k} &= \prod_{t=1}^{k} p_L(L_t|\text{Pa}(L_t|\bar{a}_t)), \ \pi_{0,L}^{*k} = \prod_{t=1}^{k} p_{0,L}(L_t|\text{Pa}(L_t|\bar{a}_t)), \end{aligned}$$

then

$$\begin{aligned} D_{L_t}(P) &= \frac{\pi_A^{*t,\bar{a}} \pi_Z^{*t}}{\pi_A^t \pi_Z^t}(Q_{R_{t+1}} - Q_{L_t}) \\ D_{Z_t}(P) &= \frac{\pi_A^{*t,\bar{a}'} \pi_R^{*t} \pi_L^{*,t-1}}{\pi_A^t \pi_R^t \pi_L^{t-1}}(Q_{L_t} - Q_{Z_t}) \\ D_{R_t}(P) &= \frac{\pi_A^{*t,\bar{a}} \pi_Z^{*,t-1}}{\pi_A^t \pi_Z^{t-1}}(Q_{Z_t} - Q_{R_t}). \end{aligned}$$

代入上面的精确余项（当未指定依赖关系时，令 $Q_X = Q_X(P)$），并注意到

$$P_0 Q_{R_1}(P_0) = P_0 G_L(P_0) = P_0 \frac{\pi_A^{*K,\bar{a}} \pi_{0,Z}^{*K}}{\pi_{0,A}^K \pi_{0,Z}^K} Y$$

$$P_0 D_{R_1}(P) = P_0(Q_{Z_1}(P) - Q_{R_1}(P))$$

$$P_0 Q_{R_1}(P_0) - P_0 Q_{R_1}(P) = P_0 \frac{\pi_A^{*K,\bar{a}} \pi_{0,Z}^{*K}}{\pi_{0,A}^K \pi_{0,Z}^K}(Y - Q_{L_K}(P)) + P_0 \frac{\pi_A^{*K,\bar{a}'} \pi_{0,R}^{*K} \pi_{0,L}^{*,K-1}}{\pi_{0,A}^K \pi_{0,R}^K \pi_{0,L}^{K-1}}(Q_{L_K}(P) - Q_{Z_K}(P)) + \cdots$$

$$+ P_0(Q_{Z_1}(P) - Q_{R_1}(P)),$$

因此（为清晰起见，仍记为 $Q_X = Q_X(P)$）

$$\begin{aligned} R(P, P_0) = & - P_0 Q_{R_1}(P_0) + P_0 Q_{R_1}(P) + \\ & \sum_{t=1}^{K} \left(P_0 \frac{\pi_A^{*t,\bar{a}} \pi_Z^{*t}}{\pi_A^t \pi_Z^t}(Q_{R_{t+1}} - Q_{L_t}) + P_0 \frac{\pi_A^{*t,\bar{a}'} \pi_R^{*t} \pi_L^{*,t-1}}{\pi_A^t \pi_R^t \pi_L^{t-1}}(Q_{L_t} - Q_{Z_t}) + P_0 \frac{\pi_A^{*t,\bar{a}} \pi_Z^{*,t-1}}{\pi_A^t \pi_Z^{t-1}}(Q_{Z_t} - Q_{R_t}) \right) \\ = & \sum_{t=1}^{K} \left(P_0 \left(\frac{\pi_A^{*t,\bar{a}} \pi_Z^{*t}}{\pi_A^t \pi_Z^t} - \frac{\pi_A^{*t,\bar{a}} \pi_{0,Z}^{*t}}{\pi_{0,A}^t \pi_{0,Z}^t} \right)(Q_{R_{t+1}} - Q_{L_t}) + \right. \\ & P_0 \left(\frac{\pi_A^{*t,\bar{a}'} \pi_R^{*t} \pi_L^{*,t-1}}{\pi_A^t \pi_R^t \pi_L^{t-1}} - \frac{\pi_A^{*t,\bar{a}'} \pi_{0,R}^{*t} \pi_{0,L}^{*,t-1}}{\pi_{0,A}^t \pi_{0,R}^t \pi_{0,L}^{t-1}} \right)(Q_{L_t} - Q_{Z_t}) + \\ & \left. P_0 \left(\frac{\pi_A^{*t,\bar{a}} \pi_{0,Z}^{*,t-1}}{\pi_A^t \pi_Z^{t-1}} - \frac{\pi_A^{*t,\bar{a}} \pi_Z^{*,t-1}}{\pi_{0,A}^t \pi_{0,Z}^{t-1}} \right)(Q_{Z_t} - Q_{R_t}) \right). \end{aligned}$$

由于 $\{Q_X : X \in R, Z, L\}$ 按顺序定义为 (p_R, p_Z, p_L) 的函数，可以验证 $p_{R_t} = p_{0,R_t}$ 会导致 $P_0\{Q_{Z_t} - Q_{R_t}\} = 0, p_{Z_t} = p_{0,Z_t}$，进而导致 $P_0\{Q_{L_t} - Q_{Z_t}\} = 0$，而 $p_{L_t} = p_{0,L_t}$ 会导致 $P_0\{Q_{R_{t+1}} - Q_{L_t}\} = 0$。在正值性假设下，这证明了以下三种情形之一成立时，我们有 $R(P, P_0) = 0$：1) $p_A = p_{0,A}$ 且 $p_Z = p_{0,Z}$；2) $p_A = p_{0,A}$ 且 $p_R = p_{0,R}, p_L = p_{0,L}$；或 3) $p_Z = p_{0,Z}$ 且 $p_R = p_{0,R}, p_L = p_{0,L}$。

此外，在满足(B4)和(B5)中规定的强正性和有界变差范数假设时，可应用柯西–施瓦茨不等式，从而放宽上述条件，只需满足 $\|p_X - p_{0,X}\|_P = o_P(n^{-1/4})$ 即可满足以下任一条件：1） $X \in \bar{A} \cup \bar{Z}$；2） $X \in A \cup R \cup L$；或 3） $X \in Z \cup R \cup L$。这证明了第7节中的多重稳健性结论。

附录 F：数值结果

$\mathbb{E}[Y_1(\bar{1}, \Gamma^1)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0001	0.0274	0.0008	0.9465	0.1064
HAL-EIC 初始估计	0.0001	0.0274	0.0008	0.9455	0.1060
精确 EIC TMLE	0.0000	0.0274	0.0008	0.9465	0.1064
HAL-EIC TMLE	0.0001	0.0274	0.0008	0.9455	0.1060
$\mathbb{E}[Y_2(\bar{1}, \Gamma^1)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0003	0.0282	0.0008	0.9687	0.1259
HAL-EIC 初始估计	0.0004	0.0283	0.0008	0.9444	0.1085
精确 EIC TMLE	0.0005	0.0323	0.0010	0.9414	0.1251
HAL-EIC TMLE	0.0004	0.0283	0.0008	0.9444	0.1085
$\mathbb{E}[Y_1(\bar{1}, \Gamma^0)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0001	0.0275	0.0008	0.9455	0.1086
HAL-EIC 初始估计	0.0000	0.0275	0.0008	0.9434	0.1070
精确 EIC TMLE	-0.0001	0.0276	0.0008	0.9455	0.1085
HAL-EIC TMLE	0.0000	0.0275	0.0008	0.9434	0.1070
$\mathbb{E}[Y_2(\bar{1}, \Gamma^0)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0003	0.0287	0.0008	0.9737	0.1327
HAL-EIC 初始估计	0.0003	0.0287	0.0008	0.9404	0.1108
精确 EIC TMLE	0.0002	0.0347	0.0012	0.9303	0.1306
HAL-EIC TMLE	0.0003	0.0287	0.0008	0.9404	0.1108
$\mathbb{E}[Y_1(\bar{0}, \Gamma^0)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0021	0.0255	0.0007	0.9606	0.1056
HAL-EIC 初始估计	0.0021	0.0255	0.0007	0.9545	0.1031
精确 EIC TMLE	0.0020	0.0261	0.0007	0.9576	0.1055
HAL-EIC TMLE	0.0021	0.0255	0.0007	0.9545	0.1031
$\mathbb{E}[Y_2(\bar{0}, \Gamma^0)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	-0.0006	0.0267	0.0007	0.9808	0.1290
HAL-EIC 初始估计	-0.0006	0.0268	0.0007	0.9414	0.1051
精确 EIC TMLE	-0.0005	0.0338	0.0011	0.9222	0.1266
HAL-EIC TMLE	-0.0006	0.0268	0.0007	0.9414	0.1051

表 1模型设定错误：

无。

$\mathbb{E}[Y_1(\bar{1}, \Gamma^{\bar{1}})]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0002	0.0274	0.0008	0.8983	0.0894
HAL-EIC 初始估计	0.0002	0.0275	0.0008	0.9003	0.0889
精确 EIC TMLE	0.0002	0.0274	0.0008	0.9003	0.0894
HAL-EIC TMLE	0.0002	0.0275	0.0008	0.9003	0.0889
$\mathbb{E}[Y_2(\bar{1}, \Gamma^{\bar{1}})]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0004	0.0284	0.0008	0.8983	0.0940
HAL-EIC 初始估计	0.0003	0.0284	0.0008	0.8602	0.0865
精确 EIC TMLE	0.0007	0.0318	0.0010	0.8592	0.0937
HAL-EIC TMLE	0.0003	0.0284	0.0008	0.8602	0.0865
$\mathbb{E}[Y_1(\bar{1}, \Gamma^{\bar{0}})]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0001	0.0275	0.0008	0.9075	0.0914
HAL-EIC 初始估计	0.0002	0.0276	0.0008	0.9034	0.0903
精确 EIC TMLE	0.0000	0.0277	0.0008	0.9096	0.0913
HAL-EIC TMLE	0.0002	0.0276	0.0008	0.9034	0.0903
$\mathbb{E}[Y_2(\bar{1}, \Gamma^{\bar{0}})]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0003	0.0289	0.0008	0.9013	0.0990
HAL-EIC 初始估计	0.0003	0.0289	0.0008	0.8654	0.0890
精确 EIC TMLE	0.0004	0.0333	0.0011	0.8602	0.0983
HAL-EIC TMLE	0.0003	0.0289	0.0008	0.8654	0.0890
$\mathbb{E}[Y_1(\bar{0}, \Gamma^{\bar{0}})]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0021	0.0255	0.0007	0.9609	0.1048
HAL-EIC 初始估计	0.0022	0.0255	0.0007	0.9579	0.1043
精确 EIC TMLE	0.0021	0.0255	0.0007	0.9568	0.1048
HAL-EIC TMLE	0.0022	0.0255	0.0007	0.9579	0.1043
$\mathbb{E}[Y_2(\bar{0}, \Gamma^{\bar{0}})]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	-0.0005	0.0268	0.0007	0.9712	0.1187
HAL-EIC 初始估计	-0.0004	0.0269	0.0007	0.9353	0.1012
精确 EIC TMLE	-0.0001	0.0298	0.0009	0.9466	0.1185
HAL-EIC TMLE	-0.0004	0.0269	0.0007	0.9353	0.1012

表 2模型设定错

误：*A.*

$\mathbb{E}[Y_1(\bar{1}, \Gamma^1)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	-0.0000	0.0274	0.0007	0.9477	0.1064
HAL-EIC 初始估计	0.0001	0.0274	0.0008	0.9443	0.1061
精确 EIC TMLE	0.0001	0.0274	0.0007	0.9477	0.1064
HAL-EIC TMLE	0.0002	0.0274	0.0008	0.9455	0.1061
$\mathbb{E}[Y_2(\bar{1}, \Gamma^1)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0003	0.0287	0.0008	0.9682	0.1260
HAL-EIC 初始估计	0.0003	0.0287	0.0008	0.9318	0.1087
精确 EIC TMLE	0.0007	0.0329	0.0011	0.9364	0.1253
HAL-EIC TMLE	0.0005	0.0287	0.0008	0.9352	0.1086
$\mathbb{E}[Y_1(\bar{1}, \Gamma^0)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0016	0.0274	0.0008	0.9511	0.1065
HAL-EIC 初始估计	0.0017	0.0274	0.0008	0.9455	0.1062
精确 EIC TMLE	0.0002	0.0277	0.0008	0.9455	0.1068
HAL-EIC TMLE	0.0008	0.0274	0.0008	0.9420	0.1062
$\mathbb{E}[Y_2(\bar{1}, \Gamma^0)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0017	0.0287	0.0008	0.9693	0.1261
HAL-EIC 初始估计	0.0017	0.0287	0.0008	0.9341	0.1094
精确 EIC TMLE	-0.0000	0.0345	0.0012	0.9273	0.1263
HAL-EIC TMLE	0.0008	0.0290	0.0008	0.9284	0.1094
$\mathbb{E}[Y_1(\bar{0}, \Gamma^0)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0041	0.0257	0.0007	0.9568	0.1055
HAL-EIC 初始估计	0.0041	0.0258	0.0007	0.9534	0.1031
精确 EIC TMLE	0.0029	0.0261	0.0007	0.9580	0.1055
HAL-EIC TMLE	0.0031	0.0255	0.0007	0.9591	0.1031
$\mathbb{E}[Y_2(\bar{0}, \Gamma^0)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0016	0.0277	0.0008	0.9773	0.1291
HAL-EIC 初始估计	0.0016	0.0278	0.0008	0.9352	0.1048
精确 EIC TMLE	0.0006	0.0342	0.0012	0.9136	0.1268
HAL-EIC TMLE	0.0006	0.0273	0.0007	0.9455	0.1048

TABLE 3模型设

定错误：Z。

$\mathbb{E}[Y_1(\bar{1}, \Gamma^{\bar{1}})]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0532	0.0186	0.0032	0.5111	0.1072
HAL-EIC 初始估计	0.0533	0.0188	0.0032	0.5069	0.1070
精确 EIC TMLE	0.0003	0.0273	0.0007	0.9504	0.1068
HAL-EIC TMLE	0.0001	0.0276	0.0008	0.9409	0.1063
$\mathbb{E}[Y_2(\bar{1}, \Gamma^{\bar{1}})]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0494	0.0183	0.0028	0.8110	0.1281
HAL-EIC 初始估计	0.0494	0.0183	0.0028	0.6473	0.1105
精确 EIC TMLE	0.0005	0.0320	0.0010	0.9525	0.1263
HAL-EIC TMLE	-0.0005	0.0279	0.0008	0.9472	0.1118
$\mathbb{E}[Y_1(\bar{1}, \Gamma^{\bar{0}})]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0549	0.0186	0.0034	0.4984	0.1096
HAL-EIC 初始估计	0.0549	0.0187	0.0034	0.4794	0.1082
精确 EIC TMLE	0.0005	0.0276	0.0008	0.9493	0.1090
HAL-EIC TMLE	0.0009	0.0278	0.0008	0.9440	0.1074
$\mathbb{E}[Y_2(\bar{1}, \Gamma^{\bar{0}})]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0508	0.0183	0.0029	0.8332	0.1354
HAL-EIC 初始估计	0.0508	0.0183	0.0029	0.6547	0.1134
精确 EIC TMLE	0.0002	0.0342	0.0012	0.9356	0.1319
HAL-EIC TMLE	-0.0001	0.0283	0.0008	0.9535	0.1147
$\mathbb{E}[Y_1(\bar{0}, \Gamma^{\bar{0}})]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0590	0.0186	0.0038	0.3749	0.1067
HAL-EIC 初始估计	0.0590	0.0188	0.0038	0.3601	0.1045
精确 EIC TMLE	0.0021	0.0260	0.0007	0.9599	0.1057
HAL-EIC TMLE	0.0020	0.0255	0.0007	0.9588	0.1037
$\mathbb{E}[Y_2(\bar{0}, \Gamma^{\bar{0}})]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0536	0.0183	0.0032	0.7635	0.1330
HAL-EIC 初始估计	0.0537	0.0184	0.0032	0.5586	0.1105
精确 EIC TMLE	-0.0005	0.0342	0.0012	0.9166	0.1277
HAL-EIC TMLE	-0.0002	0.0271	0.0007	0.9599	0.1119

表 4模型设定错误：*Y*.

$\mathbb{E}[Y_1(\bar{1}, \Gamma^1)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	-0.0000	0.0251	0.0006	0.9394	0.0988
HAL-EIC 初始估计	-0.0001	0.0251	0.0006	0.9414	0.0978
精确 EIC TMLE	-0.0001	0.0252	0.0006	0.9455	0.0988
HAL-EIC TMLE	-0.0001	0.0251	0.0006	0.9414	0.0978
$\mathbb{E}[Y_2(\bar{1}, \Gamma^1)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	-0.0001	0.0254	0.0006	0.9697	0.1142
HAL-EIC 初始估计	-0.0002	0.0254	0.0006	0.9404	0.0988
精确 EIC TMLE	0.0004	0.0301	0.0009	0.9273	0.1129
HAL-EIC TMLE	-0.0002	0.0254	0.0006	0.9404	0.0988
$\mathbb{E}[Y_1(\bar{1}, \Gamma^0)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0001	0.0254	0.0006	0.9424	0.1010
HAL-EIC 初始估计	0.0001	0.0254	0.0006	0.9424	0.0990
精确 EIC TMLE	-0.0000	0.0256	0.0007	0.9414	0.1009
HAL-EIC TMLE	0.0001	0.0254	0.0006	0.9424	0.0990
$\mathbb{E}[Y_2(\bar{1}, \Gamma^0)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	-0.0000	0.0260	0.0007	0.9768	0.1209
HAL-EIC 初始估计	-0.0000	0.0260	0.0007	0.9485	0.1036
精确 EIC TMLE	0.0006	0.0330	0.0011	0.9253	0.1185
HAL-EIC TMLE	-0.0000	0.0260	0.0007	0.9485	0.1036
$\mathbb{E}[Y_1(\bar{0}, \Gamma^0)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0032	0.0296	0.0009	0.9677	0.1274
HAL-EIC 初始估计	0.0032	0.0296	0.0009	0.9515	0.1171
精确 EIC TMLE	0.0035	0.0321	0.0010	0.9475	0.1270
HAL-EIC TMLE	0.0032	0.0296	0.0009	0.9515	0.1171
$\mathbb{E}[Y_2(\bar{0}, \Gamma^0)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0007	0.0312	0.0010	0.9869	0.1896
HAL-EIC 初始估计	0.0008	0.0312	0.0010	0.9323	0.1227
精确 EIC TMLE	0.0004	0.0528	0.0028	0.7869	0.1613
HAL-EIC TMLE	0.0008	0.0312	0.0010	0.9323	0.1227

表 5模型设定错误：无，正值性缩放因子 $\lambda = 2$。

$\mathbb{E}[Y_1(\bar{1},\ \Gamma^1)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0010	0.0239	0.0006	0.9391	0.0948
HAL-EIC 初始估计	0.0012	0.0240	0.0006	0.9359	0.0930
精确 EIC TMLE	0.0008	0.0243	0.0006	0.9391	0.0948
HAL-EIC TMLE	0.0012	0.0240	0.0006	0.9359	0.0930
$\mathbb{E}[Y_2(\bar{1},\ \Gamma^1)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0012	0.0235	0.0006	0.9695	0.1100
HAL-EIC 初始估计	0.0013	0.0235	0.0006	0.9475	0.0931
精确 EIC TMLE	0.0015	0.0289	0.0008	0.9233	0.1077
HAL-EIC TMLE	0.0013	0.0235	0.0006	0.9475	0.0931
$\mathbb{E}[Y_1(\ \bar{1},\Gamma^0)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0011	0.0242	0.0006	0.9370	0.0970
HAL-EIC 初始估计	0.0013	0.0243	0.0006	0.9296	0.0947
精确 EIC TMLE	0.0007	0.0247	0.0006	0.9370	0.0971
HAL-EIC 目标最大似然估计	0.0013	0.0243	0.0006	0.9296	0.0947
$\mathbb{E}[Y_2(\ \bar{1},\Gamma^0)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0011	0.0239	0.0006	0.9737	0.1183
HAL-EIC 初始估计	0.0011	0.0240	0.0006	0.9559	0.1049
精确 EIC TMLE	-0.0008	0.0346	0.0012	0.8897	0.1172
HAL-EIC TMLE	0.0011	0.0240	0.0006	0.9559	0.1049
$\mathbb{E}[Y_1(\ \bar{0},\Gamma^0)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0023	0.0343	0.0012	0.9800	0.1650
HAL-EIC 初始估计	0.0023	0.0344	0.0012	0.9443	0.1334
精确 EIC TMLE	0.0010	0.0430	0.0019	0.9349	0.1624
HAL-EIC TMLE	0.0023	0.0344	0.0012	0.9443	0.1334
$\mathbb{E}[Y_2(\ \bar{0},\Gamma^0)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0004	0.0354	0.0013	0.9706	0.2595
HAL-EIC 初始估计	0.0004	0.0354	0.0013	0.8845	0.1411
精确 EIC TMLE	0.0028	0.0802	0.0064	0.6113	0.1732
HAL-EIC TMLE	0.0004	0.0354	0.0013	0.8845	0.1411

表 6模型设定错误：无，正值性缩放因子 $\lambda = 3$。

$\mathbb{E}[Y_1(\bar{1}, \Gamma^1)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0009	0.0230	0.0005	0.9619	0.0927
HAL-EIC 初始估计	0.0009	0.0231	0.0005	0.9499	0.0902
精确 EIC TMLE	0.0008	0.0236	0.0006	0.9459	0.0927
HAL-EIC TMLE	0.0009	0.0231	0.0005	0.9499	0.0902
$\mathbb{E}[Y_2(\bar{1}, \Gamma^1)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0008	0.0227	0.0005	0.9709	0.1094
HAL-EIC 初始估计	0.0008	0.0228	0.0005	0.9379	0.0894
精确 EIC TMLE	0.0008	0.0294	0.0009	0.8998	0.1054
HAL-EIC TMLE	0.0008	0.0228	0.0005	0.9379	0.0894
$\mathbb{E}[Y_1(\bar{1}, \Gamma^0)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0009	0.0234	0.0005	0.9619	0.0952
HAL-EIC 初始估计	0.0009	0.0234	0.0005	0.9519	0.0928
精确 EIC TMLE	0.0005	0.0249	0.0006	0.9439	0.0963
HAL-EIC TMLE	0.0009	0.0234	0.0005	0.9519	0.0928
$\mathbb{E}[Y_2(\bar{1}, \Gamma^0)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0007	0.0232	0.0005	0.9749	0.1221
HAL-EIC 初始估计	0.0006	0.0233	0.0005	0.9449	0.1082
精确 EIC TMLE	-0.0043	0.0392	0.0016	0.8347	0.1473
HAL-EIC TMLE	0.0006	0.0233	0.0005	0.9449	0.1082
$\mathbb{E}[Y_1(\bar{0}, \Gamma^0)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0025	0.0369	0.0014	0.9950	0.2265
HAL-EIC 初始估计	0.0025	0.0369	0.0014	0.9629	0.1506
精确 EIC TMLE	0.0028	0.0605	0.0037	0.8707	0.2123
HAL-EIC TMLE	0.0025	0.0369	0.0014	0.9629	0.1506
$\mathbb{E}[Y_2(\bar{0}, \Gamma^0)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0021	0.0393	0.0015	0.9509	0.3478
HAL-EIC 初始估计	0.0020	0.0393	0.0015	0.8307	0.1562
精确 EIC TMLE	0.0105	0.1101	0.0122	0.4559	0.2778
HAL-EIC TMLE	0.0020	0.0393	0.0015	0.8307	0.1562

表格 7模型设定错误：无，正值性缩放因子 $\lambda = 4$。

$\mathbb{E}[Y_1(\bar{1},\ \Gamma^1)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0013	0.0226	0.0005	0.9604	0.0919
HAL-EIC 初始估计	0.0013	0.0228	0.0005	0.9451	0.0886
精确 EIC TMLE	0.0010	0.0234	0.0005	0.9502	0.0918
HAL-EIC TMLE	0.0013	0.0228	0.0005	0.9451	0.0886
$\mathbb{E}[Y_2(\bar{1},\ \Gamma^1)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0010	0.0222	0.0005	0.9787	0.1111
HAL-EIC 初始估计	0.0010	0.0223	0.0005	0.9360	0.0871
精确 EIC TMLE	0.0010	0.0303	0.0009	0.8770	0.1054
HAL-EIC TMLE	0.0010	0.0223	0.0005	0.9360	0.0871
$\mathbb{E}[Y_1(\bar{1},\ \Gamma^0)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0014	0.0229	0.0005	0.9644	0.0945
HAL-EIC 初始估计	0.0013	0.0231	0.0005	0.9543	0.0925
精确 EIC TMLE	0.0004	0.0273	0.0007	0.9360	0.0970
HAL-EIC TMLE	0.0013	0.0231	0.0005	0.9543	0.0925
$\mathbb{E}[Y_2(\bar{1},\Gamma^0)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0009	0.0226	0.0005	0.9817	0.1211
HAL-EIC 初始估计	0.0009	0.0227	0.0005	0.9502	0.1141
精确 EIC TMLE	-0.0089	0.0451	0.0021	0.7764	0.1370
HAL-EIC TMLE	0.0009	0.0227	0.0005	0.9502	0.1141
$\mathbb{E}[Y_1(\bar{0},\Gamma^0)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0016	0.0413	0.0017	0.9919	0.2520
HAL-EIC 初始估计	0.0015	0.0414	0.0017	0.9512	0.1623
精确 EIC TMLE	0.0017	0.0897	0.0080	0.7083	0.2234
HAL-EIC 目标最大似然估计	0.0015	0.0414	0.0017	0.9512	0.1623
$\mathbb{E}[Y_2(\bar{0},\Gamma^0)]$	Bias	SD	MSE	覆盖率	宽度
精确 EIC 初始估计	0.0019	0.0421	0.0018	0.9350	0.3276
HAL-EIC 初始估计	0.0019	0.0421	0.0018	0.7683	0.1672
精确 EIC TMLE	0.0077	0.1348	0.0182	0.3780	0.2262
HAL-EIC TMLE	0.0019	0.0421	0.0018	0.7683	0.1672

表 8模型设定错误：无，正值性缩放因子 $\lambda = 5$。

附录 G：符号列表

数据与模型数据与模型数据与模型

$_X$ X的可用历史，例如（X_t： $t=0$、 $\ldots K$）$_X_t$ X截至 t的可用历史，例如（X_s： $s=0$、 $\ldots t$）$\bar{X}_s^t$ X从 s到 t的可用历史（假设 $s \leq t$）x 随机变量取值范围内的一个实现值 X Pa(X) 在 X之前的父节点X在给定变量排序下，之后的子节点 X 在给定变量排序下 $\bar{}$ $\bar{}$Pa($X|\bar{a}$) X 的父节点向量，但对Pa(X) 进行干预 $\cap A$ 继而 $A=\bar{a}$；例如，第 2节中的 Pa($R_1|\bar{a}$) = （L_0、 a_1）$\mathcal{M}$ 统计模型，即数据分布的集合P 数据分布P_n 独立同分布样本 O_1、.....、 O_n的经验分布Pf期望 $Pf=\mathbb{E}_P$（f） $=\int f$（o）dP（o）在 P下的 f（O）；例如， $P_n f=\frac{1}{n}\sum_{i=1}^{n} f$（$O_i$）$p$ 密度 $p=dP/d\mu$ 关于某个测度 μp_X（$X|$Pa(X)）给定 Pa(X)时变量 X 的条件密度P_0 真实数据分布$_X$（$\bar{a}$）在给定结构因果模型$\Gamma_t^{\bar{a}'}$ 的干预 $A=\bar{a}$ 下 X 的反事实 Z_t给定$\bar{R}_t$（$\bar{a}'$）时的条件密度$\bar{a}'$ $\bar{Z}_{t-1}$（$\bar{a}'$） $\bar{L}_{t-1}$（$\bar{a}'$）X（$\bar{a}$、$\bar{\Gamma}^{\bar{a}'}$）在干预 $\bar{A}=\bar{a}$ 和Z_t $\sim$下X 的反事实$\Gamma_t^{\bar{a}'}$（$z_t|\bar{R}_t$（$\bar{a}$、$\Gamma^{\bar{a}'}$） $\bar{Z}_{t-1}$（$\bar{a}$、$\Gamma^{\bar{a}'}$） $\bar{L}_{t-1}$（$\bar{a}$、$\Gamma^{\bar{a}'}$））对于 $t=1$、.....、 K；St

–

Statistical Estimation

Ψ 一个一般的 k 维目标参数 $\mathcal{M} \to \mathbb{R}^k{}_{\Psi^{\bar{a}}, \ \bar{a}'}(P)\,\mathbb{E}\left[Y_s(\bar{a}, \bar{\Gamma}^{\bar{a}'})\right]$ 其中 $Y_s = \psi(\bar{L}_K)$是第 s个结局；ψ是 $\bar{L}_K P^{\bar{a}}$的函数，$\bar{a}'$ 分布 P在干预$(\bar{a}, \bar{\Gamma}^{\bar{a}'})_{Q^{\bar{a}}}$下的反事实，$\bar{a}'$, $(\text{Pa}(X) \setminus \bar{A})\,\mathbb{E}_{P^{\bar{a}}, \bar{a}'}\left[Y_s | \text{Pa}(X) \setminus \bar{A}\right]$，给定 X 的父节点（排除 A）的 Y_s的条件均值（遵循 ${}^{s}P^{\bar{a}, \bar{a}'}_{X}$），$\bar{a}'\,(P)$是 $\Psi^{\bar{a}}$的典范梯度，${}_{s}{}^{\bar{a}'}(P)_{D^{\bar{a}}, \ {}_{s}{}^{\bar{a}'}_{X}}, (P) \prod \left(D_s^{\bar{a}, \bar{a}'}(P) | T_X(P)\right)$ X的梯度分量，定义为投影到^{s}切空间 $T_X(P) = \left\{f \in L^2(P) : \mathbb{E}\left[f | \text{Pa}(X)\right] = 0\right\}\tilde{p}_X(p_X, \bar{\epsilon}_X)$是穿过 p_X的（多变量）局部最不利路径P_n^0 P_0的初始分布估计量$\tilde{P}_n$ P_n^0的一次目标最大似然估计更新P_n^* 最终目标最大似然估计更新

参考文献

ANDERSEN, P. K., BORGAN, O., GILL, R. D. and KEIDING, N. (1993).基于计数过程的统计模型。*Springer* 统计丛书。Springer，纽约。 ANDREWS, R. M. and DIDELEZ, V. (2020). 对因果中介分析中跨世界独立性假设的深入探讨。《流行病学》 **32** 209–219. AVIN, C., SHPITSER, I. and PEARL, J. (2005). 路径特异性效应的可识别性。 BANG, H. and ROBINS, J. M. (2005). 缺失数据与因果推断模型中的双重稳健估计。《生物统计学》**61** 962–973. BENKESER, D., CARONE, M. and GILBERT, P. B. (2018). 罕见结局累积发生率的改进估计。 《医学统计学》**37** 280–293. BENKESER, D. and VAN DER LAAN, M. (2016).高度自适应套索估计量。载于*2016 IEEE* 数据科学与高级分析国际会议（*DSAA*）689–696. IEEE. BIBAUT, A. F. and VAN DER LAAN, M. J. (2019). 有界截面变差范数的 c\adl\ag 函数上的经验风险最小化快速收敛率。*arXiv* 预印本 *arXiv:1907.09244*。 BUSE, J. B., BAIN, S. C., MANN, J. F., NAUCK, M. A., NISSEN, S. E., POCOCK, S., POULTER, N. R., PRATLEY, R. E., LINDER, M., FRIES, T. M. et al. (2020). 利拉鲁肽降低心血管风险：LEADER 试验的探索性中介分析。《糖尿病护理》 **43** 1546–1552. COLE, S. R. and FRANGAKIS, C. E. (2009). 因果推断中的一致性陈述：定义还是假设？ 《流行病学》 **20**3–5. D'AGOSTINO JR, R. B. (1998). 在比较治疗与非随机对照组时，采用倾向评分方法降低偏差。《医学统计学》 **17** 2265–2281.

DÍAZ, I. and HEJAZI, N. S. (2020). 随机干预的因果中介分析。皇家统计学会会刊：B辑（统计方法学）82 661–683. DÍAZ, I. and VAN DER LAAN, M. J. (2017). 随机试验中针对缺失结局数据的目标最小损失估计的双重稳健推断。《医学统计学》36 3807–3819. DÍAZ MUÑOZ, I. and VAN DER LAAN, M. J. (2011). 基于 Super Learner 的条件密度估计及其在边际结构模型中的应用。《国际生物统计学杂志》7 000010220215574679135 6. DIDELEZ, V., DAWID, A. and GENELETTI, S. (2006). 序贯治疗的直接效应和间接效应。载于第23届人工智能不确定性年会。 DUDOIT, S. and VAN DER LAAN, M. J. (2008). 多重检验程序及其在基因组学中的应用。Springer. HEJAZI, N. S., BENKESER, D. C. and VAN DER LAAN, M. J. (2022). haldensify：高度自适应套索条件密度估计。https://github.com/nhejazi/haldensify。R 软件包版本 0.2.4。 HEJAZI, N. S., COYLE, J. R. and VAN DER LAAN, M. J. (2020). hal9001：R 中可扩展的高度自适应套索回归。《开源软件杂志》5 2526. HUBBARD, A. E., KENNEDY, C. J. and VAN DER LAAN, M. J. (2018). 数据自适应目标参数。载于《数据科学中的目标学习》125–142. Springer. JU, C., SCHWAB, J. and VAN DER LAAN, M. J. (2019). 因果推断中的自适应倾向评分截断。《医学研究统计方法》28 1741–1760. LAI, E. T., SCHLÜTER, D. K., LANGE, T., STRAATMANN, V., ANDERSEN, A.-M. N., STRANDBERG-LARSEN, K. and TAYLOR-ROBINSON, D. (2020). 理解儿童心理健康不平等的形成路径：对英国和丹麦两个全国出生队列进行的反事实中介分析。《BMJ开放》10 e040056. LUEDTKE, A. R., SOFRYGIN, O., VAN DER LAAN, M. J. and CARONE, M. (2017). 右删失纵向模型中的序贯双重稳健性。arXiv 预印本 arXiv:1705.02459. MOLINA, J., ROTNITZKY, A., SUED, M. and ROBINS, J. (2017). 因子化似然模型中的多重稳健性。《生物计量学》104 561–581. PEARL, J. (2009). 统计学中的因果推断：概述。《统计调查》3 96–146. PETERSEN, M. L., PORTER, K. E., GRUBER, S., WANG, Y. and VAN DER LAAN, M. J. (2011). 正值性。载于《目标学习》161–184. Springer. ROBINS, J. M. and RICHARDSON, T. S. (2010). 替代图形因果模型与直接效应的识别。《因果关系与精神病理学：寻找疾病及其疗法的决定因素》103–158. ROBINS, J., SUED, M., LEI-GOMEZ, Q. and ROTNITZKY, A. (2007). 评论：当“逆概率”权重高度可变时双重稳健估计量的表现。《统计科学》22 544–559. ROSE, S. and VAN DER LAAN, M. J. (2018). LTMLE。载于《数据科学中的目标学习》35–47. Springer.

ROSENBAUM, P. R. 和 RUBIN, D. B.（1983）。倾向评分在观察性研究因果效应分析中的核心作用。*Biometrika***70** 41–55。RYTGAARD, H. C.、GERDS, T. A. 和 VAN DER LAAN, M. J.（2021）。干预特异性平均结局的连续时间靶向最小损失估计。*arXiv preprint arXiv:2105.02088*。RYTGAARD, H. C. 和 VAN DER LAAN, M. J.（2022）。生存分析与竞争风险分析中用于因果推断的靶向最大似然估计。*Lifetime Data Analysis* 1–30。TCHETGEN, E. J. T.（2009）。关于 G. Molenberghs 对缺失数据方法的综述的评论。*Drug Infor-mation Journal***43** 433–435。TCHETGEN, E. J. T. 和 SHPITSER, I.（2012）。因果中介分析的半参数理论：效率界、多重稳健性与敏感性分析。*Annals of statistics* **40** 1816。VAN DER LAAN, M. J.（2010a）。基于靶向最大似然的因果推断：第一部分。*The internationaljournal of biostatistics***6**。VAN DER LAAN, M.（2010b）。基于靶向最大似然的因果推断：第二部分。*The InternationalJournal of Biostatistics***6** 文章–3。VAN DER LAAN, M. J.（2018）。有效影响曲线的 HAL 估计量。载于数据科学中的目标学习 103–123。Springer。VAN DER LAAN, M. J. 和 DUDOIT, S.（2003）。用于在估计量之间进行选择的统一交叉验证方法，以及一般的交叉验证自适应 epsilon 网估计量：有限样本预言机不等式与实例。VAN DER LAAN, M. J.、DUDOIT, S. 和 KELES, S.（2004）。基于似然的交叉验证的渐近最优性。*Statistical Applications in Genetics and Molecular Biology***3**。VAN DER LAAN, M. J. 和 GRUBER, S.（2010）。协作式双重稳健靶向最大似然估计。*The international journal of biostatistics***6**。VAN DER LAAN, M. 和 GRUBER, S.（2016）。基于通用最有利一维子模型的一步靶向最小损失估计。*The international journal of biostatistics***12** 351–378。VAN DER LAAN, M. J. 和 PETERSEN, M. L.（2008）。直接效应模型。*The internationaljournal of bio-statistics* **4**。VAN DER LAAN, M. J.、POLLEY, E. C. 和 HUBBARD, A. E.（2007）。超级学习器。*Statisticalapplications ingenetics and molecular biology***6**。VAN DER LAAN, M. J. 和 ROSE, S.（2011）。目标学习：观察性和实验性数据的因果推断。Springer Science & Business Media。VAN DER LAAN, M. J.、ROSE, S. 和 VAN DER LAAN, M. J.（2018）。有效影响曲线的 HAL 估计量。数据科学中的目标学习：复杂纵向研究的因果推断103–123。VAN DER LAAN, M. J. 和 RUBIN, D.（2006）。靶向最大似然学习。*The international journalof biostatistics***2**。VAN DER LAAN, M.、WANG, Z. 和 VAN DER LAAN, L.（2021）。高阶靶向最大似然估计。*arXiv preprint arXiv:2101.06290*。

VAN DER LAAN, M. J.、ROSE, S.、VAN DER LAAN, M. J. 和 BENKESER, D.（2018）。高度自适应套索（HAL）。数据科学中的目标学习：复杂纵向研究的因果推断77–94。VANDERWEELE, T. J. 和 TCHETGEN, E. J. T.（2017）。时变暴露和中介变量的中介分析。*Journal of the Royal Statistical Society. Series B, Statistical Methodology***79** 917。VANDERWEELE, T. J. 和 VANSTEELANDT, S.（2009）。关于中介、干预与组合的概念问题。*Statistics and its Interface***2** 457–468。VANSTEELANDT, S.、LINDER, M.、VANDENBERGHE, S.、STEEN, J. 和 MADSEN, J.（2019）。对受时变混杂影响的重复测量中介变量进行校正的事件发生时间终点中介分析。*Statistics in medicine***38** 4828–4840。ZHENG, W. 和 VAN DER LAAN, M. J.（2011）。交叉验证的目标最小损失估计。载于目标学习459–474。Springer。ZHENG, W. 和 VAN DER LAAN, M. J.（2012）。具有时依中介变量的生存分析情境中的因果中介。ZHENG, W. 和 VAN DER LAAN, M.（2017）。具有时变中介变量和暴露的纵向中介分析及其在生存结局中的应用。*Journal of causal inference* **5**。