

处理效应的异质性的识别

——基于因果机器学习的仿真和估计

晏发发^{1,2}, 颜丽金¹, 陈铮浩^{1*}

1. 广州理工学院经济管理学院, 广东 广州 510540

2. 广东特色金融与高质量发展研究中心, 广东 广州 510540

DOI:10.61369/ASDS.2026010010

摘 要 : 既有因果机器学习的相关文献所提出的用于识别异质群体的平均处理效应 (GATE) 的方法未能考虑在协变量变化的情况下解释不同群体间处理效应的异质性。为了解决该问题, 本文提出基于无偏机器学习 (DML) 的平衡组平均处理效应 (BGATE) 来衡量具有预先确定的协变量特定分布的组平均处理效应 (GATE), 通过计算两个 BGATE 之间的差值来比较两个 GATE 的值, 从而更好地识别因果效应的异质性, 最终将由协变量不同分布所导致的差异与由解释变量所导致的差异区分开来。该估计量在标准条件下具有 \sqrt{N} -一致性和渐近正态的性质。通过对比 DML、自动无偏机器学习 (Auto-DML) 和重新加权 (Reweighting Approach) 三种估计方法的仿真结果可知: 如果已知 DML 没有性能问题, 如当倾向得分很极端时, 建议采用 DML 估计量, 其模拟表现最好; 如果已知 DML 有性能问题, 建议采用 Auto-DML 估计量或重新加权方法。

关 键 词 : 调节效应; 无偏机器学习; 重新加权方法; 异质性; 平衡组平均处理效应

Identification of Treatment Effect Heterogeneity — Simulation and Estimation Based on Causal Machine Learning

Yan Fafa^{1,2}, Yan Lijin¹, Chen Zhenghao^{1*}

1. School of Economics and Management, Guangzhou Institute of Science and Technology, Guangzhou, Guangdong 510540

2. Guangdong Research Center of Characteristic Finance and High-Quality Development, Guangzhou, Guangdong 510540

Abstract : Existing studies on causal machine learning proposes methods for identifying the Average Treatment Effect for heterogeneous groups (GATE), failed to account for the heterogeneity of treatment effects across different groups under varying covariates. To address this issue, this paper proposes the Balanced Group Average Treatment Effect (BGATE) based on Double Machine Learning (DML) to measure the Group Average Treatment Effect (GATE) with predetermined specific distributions of covariates. By calculating the difference between two BGATE values, the GATE values of two groups can be compared, thereby better identifying the heterogeneity of causal effects. Ultimately, this approach distinguishes between differences caused by varying distributions of covariates and those caused by explanatory variables. The estimator exhibits \sqrt{N} -consistency and asymptotic normality under standard conditions. A comparison of simulation results from three estimation methods—the DML, the Auto-DML, and the Reweighting approach—shows that if DML is known to have no performance issues (e.g., when the propensity score is not extreme), the DML estimator is recommended, as it performs best in simulations. If DML is known to have performance issues, the Auto-DML estimator or the reweighting approach is recommended.

Keywords : moderation effect; debiased machine learning; reweighting approach; heterogeneity; balanced group average treatment effect

基金项目: 2024年广州理工学校级项目“数字信贷对服务业创业的影响效应和机制——基于广东县域面板数据的经验证据”(2024KY052); 2024-2025年度广州理工学院校级教学质量与教学改革工程项目“多语言程序设计+数字经济专项人才培养计划(实验班)”(2024XZLGC39); 2025年广州理工学校级项目“数字经济对中国经济不平衡的影响效应与影响路径研究”(2025KYQ003); 2025年广州理学院工作室项目铭心博士工作室; 2025年广东省教育科学规划课题(高等教育专项)“产学研赛+混合编程+数字经济跨学科人才培养探究”(2025GXJK0660); 2025年广东省教育科学规划课题(高等教育专项)“AI虚拟助教助力高校国贸专业互动式教学的实践探索”(2025GXJK0664); 2025-2026年度广东省教育评估协会一般课题“学生视角下高校课程思政实施效果评价体系构建研究”(BDPG25100); 2025年广东省普通高校特色创新类项目哲学社会科学类“广东省低空经济产业发展的路径优化: 基于多维系统关联的研究”(2025GXJK0664)。

作者简介:

晏发发, 广州理工学院经济管理学院, 副教授, 研究领域: 计量经济学理论和应用、宏观模型构建和应用、农业经济、数字经济、教育经济、资产定价、机制设置、博弈论等;

颜丽金, 广州理工学院经济管理学院, 本科生在读, 研究领域: 计量经济学理论和应用、宏观模型构建和应用、会计学等;

通讯作者: 陈铮浩(第三作者), 广州理工学院经济管理学院, 本科生在读, 研究领域: 计量经济学理论和应用、宏观模型构建和应用、会计学等。

引言

一个特定的变量是否导致处理效果的差异称为因果调节变量。除了平衡所有协变量混淆组间调节效应的分布外，还必须进行其他假设来解释因果关系的处理效应的差异，只有这样组变量才可以被认为是一个不混杂的调节因素。机器学习方法不仅可以融入传统计量方法来改进因果推断计量方法，还可以采用数据驱动以考虑广泛的模型，从而避免模型设定错误。因此，机器学习在因果推断经济学中的应用有着重要现实意义。

本文聚焦于无混淆设定下组间处理效果差异的估计方法研究。首先，提出平衡组平均处理效果（BGATE），该参数是基于预先确定的协变量特定分布的组平均处理效应（ Δ GATE），通过计算两个 BGATE 的差值（同样记作 BGATE），能够有效实现组间处理效果的对比。本文展示 BGATE 它如何与两个组平均处理效应的差异（ Δ GATE）相关，讨论了它的识别方法，并提出了不同的估计量来描述离散调节因素（为了简单起见，本文将发现异质处理效应的变量称为调节因素）和离散处理。关于无偏机器学习的估计量（Debiased Machine Learning, DML）（Chernozhukov 等，2018）^[1]，本文证明了该估计量在真实值处渐近正态分布，是一个有效的统计量。此外，本文还将 DML 方法与自动无偏机器学习（Auto-DML）和重新加权方法（Reweighting Approach）估计 Δ BGATE^[2] 的结果进行对比。DML 依赖于使用一个基于倾向得分的双重稳健评分函数，如果倾向得分极端，这可能会出现（Lechner 和 Mareckova，2024^[3]；Frolich，2004^[4]）。因为 Auto-DML 使用 Riesz 表示而不是倾向得分，所以 Auto-DML 不会受到极端倾向得分的问题影响（Chernozhukov 等，2021^[5]；Chernozhukov 等，2018^[1]）；重新加权方法允许用户使用任何他们偏好的方法来估计 Δ BGATE，因为 Δ BGATE 等于重新加权数据上的 Δ GATE。这两种估计方法是有效的（郭峰和陶旭辉，2023）^[6]。

本文其余章节安排如下：第一节对相关领域文献进行综述，第二节定义了因果调节框架；第三节给出了 DML 估计量的渐近性质；第四节报告了样本规模为 500 的蒙特卡罗仿真设计和结果；第五节对全文进行总结归纳，得出研究结论。

一、文献综述

既有研究探讨了如何在组水平上检测（低维）异质性，称为“组平均处理效应”（GATEs 是一种条件平均处理效），目前已经开发了几种方法来估计 GATEs。例如，Abrevaya 等（2015）^[7]展示了如何在无混杂条件下非参数识别 GATEs，并使用逆概率加权估计量对其进行估计。Lechner（2018）^[8]改进了一种随机森林算法，以调整混淆因素从而来估计异质性处理效应。Semenova 和 Chernozhukov（2021）^[9]使用 DML 框架来寻找基于线性模型的异质性。Zimmert 和 Lechner（2019）^[10]以及 Fan 等（2022）^[11]开发了一个两步估计量，允许非参数估计 GATEs。第一步是使用机器学习方法估计，第二步是应用非参数局部常数回归。

本文通过提出 Δ BGATE 估计量以刻画不同组间处理效果差异，即，在平衡一些协变量的分布的同时，比较两组的处理效果。若研究目的聚焦于识别处理结果中赢家与输家的特征，该方法具有较高适配性；然而，当核心诉求在于量化某一变量对两组处理效果差异的影响程度时，该方法则更具针对性与有效性。

DML 框架由 Chernozhukov 等（2018）^[1]提出，该框架使用机器学习的方法来进行因果分析。机器学习算法可能会引入两种偏差：正则化偏差和过拟合偏差。DML 的主要思想是，通过使用奈曼正交得分函数（Neyman-orthogonal score functions）以克服正则化偏差。既有文献已经将根据不同的设置调整了通用 DML 框架用于中介分析（Farbmacher 等，2022）^[12]。本文通过使用这个框架来估计因果调节效应来为该领域的文献做出贡献。

二、因果调节框架

（一）定义

本文使用的因果调节框架是基于（Rubin，1974）^[13]的潜在结果框架。因果效应被定义为两个潜在结果之间的差异，而对于一个个体，本文只能观察到两个潜在结果中的一个。因此，找到一个可信的反事实是很棘手的。我们观察到来自独立随机变量 $H_i = (D_i, Y_i, Z_i, X_i)$ 的 N 个独立同分布的样本，它们以未知概率的分布 P 来抽取。重点在于处理变量 D_i 和调节变量 Z_i 。为简单起见，假设它们是二元的（处理变量的结果为 $d \in \{0, 1\}$ ，调节变量的结果为 $z \in \{0, 1\}$ ）。与既有研究中的设定一样，潜在结果由处理变量索引： (Y_i^0, Y_i^1) 。最后，一组 $k \in \{0, \dots, p\}$ 协变量 $X_{i,k}$ 可能同时影响处理和潜在的结果，其中 $X_i = (X_i^0, X_i^1, Z_i^0, Z_i^1)$ 。此外，潜在的协变量和潜在的调节因子被定义为： $X_i = (X_i^0, X_i^1, Z_i^0, Z_i^1)$ 。

由于只能观察到一种潜在结果的实现，因此我们永远无法一致地估计个体处理效应（ITE） $\xi_i = Y_i^1 - Y_i^0$ 的实现。然而，在适当的假设下，平均处理效应（ATE） $\theta = E[Y_i^1 - Y_i^0]$ 的识别是可能的（Imbens 和 Wooldridge，2009）^[14]。另外，为了进一步研究 ξ_i 异质性的不同方面，可以通过条件平均处理效应（CATE）来捕捉。CATE 测量了一组（子）协变量 x_i 条件下的平均处理效果。个体化平均处理效果（IATE）和组平均处理效果（GATE）都是特异性的 CATEs。IATE 在最精细的聚合水平上测量处理效果。也就是说，它比较了所有个体的平均处理效果与所有相关协变量的特定值。在形式上，IATE 被定义为

$$\tau(x, z) = E[Y_i^1 - Y_i^0 | Z_i = z, X_i] \quad (1)$$

GATE 测量组间处理效果, 即在比 IATE 更聚合的水平上且比 ATE 更精细。GATE 定义为

$$\theta^G(z) = E[Y^1 - Y^0 | Z_i = z] = E[\tau(X_i, 0) | Z_i = 0] \quad (2)$$

如果仅仅刻画处理效应的异质性, 那么采用 IATEs 和 GATEs 就足够了。如果识别两组之间的处理效果差异, 那么需要采用两组 GATE 之间的差异 (Δ GATE), 即

$$\begin{aligned} \theta^{\Delta G} &= E[Y^1 - Y^0 | Z_i = 1] - E[Y^1 - Y^0 | Z_i = 0] \\ &= [\tau(X_i, 1) | Z_i = 1] - [\tau(X_i, 0) | Z_i = 0] \end{aligned} \quad (3)$$

然而, 因为两组在协变量 x_i 的分布上可能存在差异, 通常该效应难以解释。因此, 本文引入平衡组平均处理效应 (BGATE) 以平衡本文想要相互比较的各 (由不同 Z_i 值定义) 中其他变量的分布。用于平衡 GATEs 的变量表示为 W_i 。 W_i 是 X_i 的一部分。如果 W_i 为空集, 或 W_i 独立于 Z_i , 则 BGATE 减少为 GATE。因此, 新的感兴趣的参数, 即用 $\theta^B(z)$ 表示, 被定义为

$$\theta^B(z) = E[E[Y_i^1 - Y_i^0 | Z_i = z, W_i]] = E[E[\tau(X_i, z) | Z_i = z, W_i]] \quad (4)$$

以及其差异

$$\begin{aligned} \theta^{\Delta B} &= E[E[Y_i^1 - Y_i^0 | Z_i = 1, W_i]] - E[E[Y_i^1 - Y_i^0 | Z_i = 0, W_i]] \\ &= E[E[\tau(X_i, 1) | Z_i = 1, W_i]] - E[E[\tau(X_i, 0) | Z_i = 0, W_i]] \end{aligned} \quad (5)$$

其中, Δ BGATE (θ^B) 表示两组之间的差异, 将两组中协变量 (W_i) 的分布调整为总体人群分布。 Δ BGATE 通常显示出调节效应。平衡组平均处理效应 (Δ BGATE) 作为调节因果效应。

(二) 效应分解

为了明确 Δ GATE 和 Δ BGATE 之间的区别, Δ GATE 被分解为两个组成部分: Δ BGATE, 代表调节变量的直接效应, 以及由各组间 W_i 分布的差异引起的组合效应组, 组合成分根据其相对重要性对两个 Z_i 子样本进行加权, 以捕捉 W_i 分布的差异:

$$\begin{aligned} & \frac{E[Y_i^1 - Y_i^0 | Z_i = 1] - E[Y_i^1 - Y_i^0 | Z_i = 0]}{\Delta \text{GATE}} \\ &= \underbrace{E[E[Y_i^1 - Y_i^0 | W_i, Z_i = 1] - E[Y_i^1 - Y_i^0 | W_i, Z_i = 0] | Z_i = 0]}_{\text{直接效应: } \Delta \text{BGATE}} \\ &+ \underbrace{\frac{P(Z_i = 0)}{P(Z_i = 1)} E[E[Y_i^1 - Y_i^0 | W_i, Z_i = 1]] - E[E[Y_i^1 - Y_i^0 | W_i, Z_i = 1] | Z_i = 0]}_{\text{组合效应(1)}} \\ &- \underbrace{\frac{P(Z_i = 1)}{P(Z_i = 0)} E[E[Y_i^1 - Y_i^0 | W_i, Z_i = 0]] - E[E[Y_i^1 - Y_i^0 | W_i, Z_i = 0] | Z_i = 1]}_{\text{组合效应(2)}} \end{aligned} \quad (6)$$

这种分解用于将两组之间结果差异分解为两个部分: 一部分是由于协变量分布的差异所致, 另一部分是由于感兴趣的变量所致。本文关注总体而非特定部分。虽然 Δ BGATE 是使用协变量 W_i 的无条件分布来定义的, 但是 Δ BGATE 也可以使用条件分布 $W_i | Z_i = 1$ 和 $W_i | Z_i = 0$ 来定义。

(三) 识别

为了在不混杂的环境中识别 GATE、BGATE、 Δ GATE 或 Δ BGATE, 通常需要识别假设 (例如, Imbens, 2004)^[15], 具体如下:

假设 1

(1) 条件独立性:

$$(Y_i^1, Y_i^0) \perp D_i | X_i = x, Z_i = z, \forall x \in X, \forall z \in Z$$

(2) 共同支持: $0 < P(D_i = d | X_i = x, Z_i = z) < 1$,

$\forall d \in \{0, 1\}, \forall x \in X, \forall z \in Z$,

(3) 混杂因子的外生性: $X_i^0 = X_i^1, Z_i^0 = Z_i^1, (d)$,

(4) 稳定的单位处理值假设:

$$Y_i = D_i Y_i^1 + (1 - D_i) Y_i^0 \quad (7)$$

定理 1

在假设 1 下, 参数

$$\theta^B(z) = E[E[Y_i^1 - Y_i^0 | Z_i = z, W_i]] \quad (8)$$

被识别为

$$E[E[\mu_1(Z_i - X_i) - \mu_0(Z_i - X_i) | Z_i = z, W_i]] \quad (9)$$

其中

$$\mu_d(z, x) = E[Y_i | D_i = d, Z_i = z, X_i = x] \quad (10)$$

因此

$$\theta^{\Delta B} = E[E[Y_i^1 - Y_i^0 | Z_i = 1, W_i]] - E[E[Y_i^1 - Y_i^0 | Z_i = 0, W_i]] \quad (11)$$

已识别为

$$E[E[\mu_1(Z_i - X_i) - \mu_0(Z_i - X_i) | Z_i = 1, W_i]] - E[E[\mu_1(Z_i - X_i) - \mu_0(Z_i - X_i) | Z_i = 0, W_i]] \quad (12)$$

三、基于 DML 的因果调节效应的估计

由于本研究关注的是处理效应的异质性, 因此本文采用 DML 估计策略侧重于 Δ BGATE。

(一) 估计方法

无偏机器学习方法采用三步估计策略。为了灵活处理潜在高维协变量向量, 可以选用 Chernozhukov 等 (2018)^[1] 提出的 DML 方法。第一步, 估计通常的双鲁棒评分函数。第二步, 评分函数回归到由调节变量 Z_i 的协变量定义的两个指标变量上。最后, 计算由调节变量所划分的两组之间的差异, 并对平衡变量 W_i 求平均值。估计的双鲁棒评分函数由以下表达式:

$$\begin{aligned} \hat{\phi}^{\Delta B}(h; \theta^{\Delta B}, \hat{\eta}) &= \hat{g}_1(w) - g_0(w) + \frac{z(\hat{\delta}(h) - \hat{g}_1(w))}{\hat{\lambda}_1(w)} \\ &- \frac{(1-z)(\hat{\delta}(h) - \hat{g}_0(w))}{1 - \hat{\lambda}_1(w)} - \theta^{\Delta B} \end{aligned} \quad (13)$$

$$\begin{aligned} \delta(h) &= \mu_1(z, x) - \mu_0(z, x) + \frac{d(y - \mu_1(z, x))}{\pi_1(z, x)} \\ &- \frac{(1-d)(y - \mu_0(z, x))}{1 - \pi_1(z, x)} \end{aligned} \quad (14)$$

$$\lambda_z(w) = P(Z_i = z | W_i = w) \quad (15)$$

$$g_z(w) = E[\delta(H_i) | Z_i = z, W_i = w] \quad (16)$$

$$\mu_d(z, x) = E[Y_i | D_i = d, Z_i = z, X_i = x] \quad (17)$$

$$\text{和 } \pi_d(z, x) = P(D_i = d | Z_i = z, X_i = x) \quad (18)$$

其中, $\hat{\delta}(h)$, $\hat{\mu}_d(z, x)$, $\hat{\lambda}_z(w)$, $\hat{\pi}_d(z, x)$ 分别代表了 $\delta(h)$, $\mu_d(z, x)$, $\lambda_z(w)$, $\pi_d(z, x)$ 的估计值。另外, 需注意 $\hat{\delta}(h)$ 对 Z_i 和 W_i , $\hat{g}_z(w) = \hat{E}[\delta(H_i) | Z_i = z, W_i = w]$ 的回归是 $\hat{\delta}(h)$ 对 Z_i 和 W_i 的相应回归 $\hat{g}_z(w) = \hat{E}[\delta(H_i) | Z_i = z, W_i = w]$ 。因此, $E[\dots]$ 表示一个一般的回归估计量, 它可以是线性的, 也可以是非线性的, 这取决于假定的数据生成过程。最后, 估计的干扰参数为 $\hat{\eta} = (\hat{\mu}_d(z, x), \hat{\lambda}_z(w), \hat{g}_z(w))$ 。

如上所述, 必须估计分数函数 $\delta(h)$ 。 $\delta(h)$ 的干扰函数误差的乘积必须收敛速度快于或等于 \sqrt{N} , 并且必须使用与 K 折交叉验证 (K-folds) (K > 1) 的交叉拟合。在第二步估计中, 是干扰项的乘积。数误差必须以 \sqrt{N} 的速度收敛, 并必须使用与 J 折交叉验证 (J-folds) (J > 1) 交叉拟合。如果满足某些条件, 则估计量是 \sqrt{N} 一致且渐近正态的 (见前文所述)。因为 $E[\phi^{AB}(H_i; \theta^{AB}, \eta)] = 0$, 所以 θ^{AB} 的方差为

$$\text{Var}(\hat{\theta}^{AB}) = \text{Var}(\phi^{AB}(H_i; \theta^{AB}, \eta)) = E[\phi^{AB}(H_i; \theta^{AB}, \eta)^2] - E[\phi^{AB}(H_i; \theta^{AB}, \eta)]^2 = E[\phi^{AB}(H_i; \theta^{AB}, \eta)^2] \quad (19)$$

其中, S_J 是第二次估计步骤中的一个随机折叠, 具体步骤参考 (Chernozhukov 等, 2018) [1] 的研究。

(二) 渐近性质

在以下假设条件下, 本文研究了估计量的渐近性质:

假设 2

(交叠性): 倾向得分 $\lambda_z(w)$ 和 $\pi_d(z, x)$ 都不趋于 0 和 1:

$\kappa < \lambda_z(w), \pi_d(z, x) < 1 - \kappa \quad \forall x \in \mathcal{X}, z \in \mathcal{Z}$, 对于某些常数 $\kappa > 0$ 。

假设 3

(一致性): 扰动函数估计量是上确界范数一致的:

$$\begin{aligned} (1) \quad & \sup_{w \in \mathcal{W}} |\hat{\lambda}_z(w) - \lambda_z(w)| \xrightarrow{p} 0, \\ (2) \quad & \sup_{w \in \mathcal{W}} |\hat{g}_z(w) - g_z(w)| \xrightarrow{p} 0, \\ (3) \quad & \sup_{x \in \mathcal{X}, z \in \mathcal{Z}} |\hat{\mu}_d(z, x) - \mu_d(z, x)| \xrightarrow{p} 0, \\ (4) \quad & \sup_{x \in \mathcal{X}, z \in \mathcal{Z}} |\hat{\pi}_d(z, x) - \pi_d(z, x)| \xrightarrow{p} 0. \end{aligned}$$

假设 4

(风险衰减性): 结果和倾向得分函数的估计误差的乘积衰减为:

$$E[(\hat{g}_z(W_i) - g_z(W_i))^2] E[(\hat{\lambda}_z(W_i) - \lambda_z(W_i))^2] = o_p\left(\frac{1}{N}\right) \quad (20)$$

$$E[(\hat{\mu}_d(Z_i, X_i) - \mu_d(Z_i, X_i))^2] E[(\hat{\pi}_d(Z_i, X_i) - \pi_d(Z_i, X_i))^2] = o_p\left(\frac{1}{N}\right) \quad (21)$$

其中, 期望是指 X_i 和 Z_i 或 W_i 取 $\hat{\mu}_d$, $\hat{\pi}_d$, \hat{g}_z 和 $\hat{\lambda}_z$ 。

如果两个扰动参数都以 \sqrt{N} 一致速率的估计, 则误差的乘积是 $o_p\left(\frac{1}{N}\right)$ 同阶项。因此, 参数的估计是 $N^{\frac{1}{2}}$ 一致就足够了。

假设 5

(条件方差的平稳性): 结果和倾向得分函数的条件方差是

有界的:

$$(1) \quad \sup_{w \in \mathcal{W}} \text{Var}(\delta(H_i) | Z_i = z, W_i = w) < \hat{\Delta}_{z1} < \infty,$$

$$(2) \quad \sup_{w \in \mathcal{W}} \text{Var}(\delta(H_i) | Z_i = z, W_i = w) < \hat{\Delta}_{z0} < \infty,$$

(3) $\sup_{x \in \mathcal{X}, z \in \mathcal{Z}} \text{Var}(Y_i | D_i = d, Z_i = z, X_i = x) < \hat{\Delta}_d < \infty$ 对于某些常数 $\hat{\Delta}_{z1}, \hat{\Delta}_{z0}, \hat{\Delta}_d > 0$ 。

假设 6

(收敛性): 第二步回归估计量 $E[\dots]$ 在定义域内是平稳的。假设 2 至 5 是 DML 的标准假设 (Chernozhukov 等, 2018) [1], 唯一的区别是这些假设只适用于第一个和第二个估计步骤。假设 4 是必要的, 以确保扰动项函数误差的乘积收敛速度比 \sqrt{N} 更快或与 \sqrt{N} 相等。当稀疏性条件得到满足时, 各种机器学习方法的 L_2 收敛速率符合这些性质。Belloni 和 Chernozhukov (2013) [10] 表明, 在近似稀疏的情况下, 这意味着排序系数的绝对值迅速衰减, Lasso 估计量的误差是阶数

$$O\left(\sqrt{\frac{\log(\max(p, N))}{N}}\right)$$

其中, p 是回归变量, s 是真实系数的未知数量。根据这些假设, 本文可以得出主要的理论结果 (见定理 1):

定理 1

在假设 2—假设 6 下, 所提出的 Δ BGATE 估计策略满足 $\sqrt{N}(\hat{\theta}^{AB} - \theta^{AB}) \xrightarrow{d} N(0, V^*)$ 和 $V^* = E[\phi^{AB}(H_i; \theta^{AB}, \eta)^2]$ 。该估计量是 \sqrt{N} 一致且渐近正态的。

四、因果调节效应的仿真

(一) 数据生成过程

本文从用 $p = 10$ 开始模拟一个 p 维协变量矩阵 $X_{i,p}$ 。前两个协变量来自均匀分布的 $X_{i,0}$, $X_{i,1} \sim U[0, 1]$, 其余的协变量来自正态分布的 $X_{i,2}, \dots, X_{i,p-1} \sim N(0.5, \sqrt{1/12})$ 。所有协变量的平均值均为 0.5, 标准差为 $\sqrt{12}$ 。调节变量 Z_i 来自于一个有概率的伯努利分布

$$P(Z_i = 1 | X_{i,0}, X_{i,1}) = (0.1 + 0.8\beta_{2,4}(X_{i,0} \times X_{i,1})) \quad (22)$$

其中, $\beta_{2,4}(X_{i,0} \times X_{i,1})$ 表示形似参数为 2 和 4 的 Beta 分布的累积分布函数。处理变量 D_i 是从下面伯努利分布中抽取的

$$P(D_i = 1 | X_{i,0}, X_{i,1}, X_{i,2}, X_{i,3}, Z_i) = (0.2 + 0.6\beta_{2,4}(\frac{X_{i,0} \times X_{i,1} \times X_{i,2} \times X_{i,3} \times Z_i}{5})) \quad (23)$$

本文据此指定了处理和非处理下的反应函数, 以及调节变量的两种状态。高度非线性的非处理响应函数由

$$\mu_0(X_i) = \sin(\pi \times X_{i,0} \times X_{i,1}) + (X_{i,2} - 0.5)^2 + 0.1X_{i,3} + 0.3X_{i,5}$$

处理下的反应函数依赖于 Z_i , 并被定义为

$$\mu_1(0, X_i) = \mu_0(X_i) + \sin(1.4X_{i,0}) + \sin(6X_{i,1}) + 0.6X_{i,2}^2 + 0.3X_{i,5} \text{ 和}$$

$\mu_1(1, X_i) = \mu_0(X_i) + \sin(4.9X_{i,0}) + \sin(2X_{i,1}) + 0.7X_{i,2}^4 + 0.4X_{i,3} + 0.2$, 它们的选择使得 Δ BGATE 与 Δ GATE 不同。最后, 本文将潜在的结果仿真为 $Y_i^d(z) = \mu_d(z, X_i) + e_{i,d,z} \quad \forall z \in \{0, 1\}$, 综上所述, 这些数据由一个可观测的四维数据组成 $(y_{i,r}, d_{i,r}, z_{i,r}, x_{i,2})$ 。

（二）仿真和估计

本文对 Δ GATE 和 Δ BGATE 进行仿真和估计，二者的计算公式如下：

$$\theta^{\Delta GATE} = E[E[Y_i^1 - Y_i^0 | Z_i = 1] - E[Y_i^1 - Y_i^0 | Z_i = 0]] \quad (24)$$

$$\theta_{x_{i,0}}^{\Delta BGATE} = E[E[Y_i^1 - Y_i^0 | Z_i = 1, X_{i,0}] - E[Y_i^1 - Y_i^0 | Z_i = 0, X_{i,0}]] \quad (25)$$

$$\theta_{x_{i,2}}^{\Delta BGATE} = E[E[Y_i^1 - Y_i^0 | Z_i = 1, X_{i,2}] - E[Y_i^1 - Y_i^0 | Z_i = 0, X_{i,2}]] \quad (26)$$

其中， $X_{i,0}$ 在两个调节组之间不平衡，而 $X_{i,2}$ 是平衡的。因此， $\theta_{x_{i,0}}^{\Delta B}$ 不同，而 $\theta_{x_{i,2}}^{\Delta B}$ 等于 $\theta^{\Delta G}$ ，本文生成了 $N=500$ 的仿真样本。

在两个估计步骤中均使用了两折 ($K=2, J=2$)。DML 估计策略中的所有扰动函数都是使用随机森林来估计的，调整用于估计扰动参数的学习算法是至关重要的 (Bach 等, 2024)^[17]。本文在两个估计步骤中再次使用了两折叠 ($K=2, J=2$)。该算法是利用 Chernozhukov 等 (2021)^[5] 提出的神经网络实现的。最后，通过估计效应重新加权数据，使平衡变量的分布与调节变量定义的组相同。重新加权算法平衡数据后，用 DML 估计量来估计 Δ GATE。

（三）仿真结果

表1给出了样本量为500时的效应和估计量的仿真结果。从不同估计量的表现来看，DML 在几乎所有情况下的均方根误差 (RMSE) 都最小，其次是 Auto-DML。与预期一致，重新加权策略通常会导致标准差 (STD) 略微升高，从而导致 RMSE 升高，它的覆盖率有时较差，原因在于其效应偏差或者标准误差过大。此外，使用神经网络的样本量应该足够大，而这对于500个样本量可能不是这样的情况，这或许就是 Auto-DML 的覆盖率对 Δ GATE 的估计较差的一个原因。最后，尽管 DML 估计量在模拟中表现最好，但这一发现可能不适用于已知 DML 有性能问题的场景，例如当倾向得分很极端时，建议采用 Auto-DML 估计量或重新加权方法。

表1：不同效应和估计量的仿真结果 (N=500)

效应	估计值	偏差	标准差	均方根误差
$\theta_{x_0}^{\Delta B}$	DML	$2.442491e^{-17}$	0.166694	0.166527
$\theta_{x_0}^{\Delta B}$	Auto-DML	$1.154632e^{-17}$	0.178052	0.177874
$\theta_{x_0}^{\Delta B}$	Reweighting	$-5.329071e^{-18}$	0.168738	0.168569
$\theta_{x_2}^{\Delta B}$	DML	$5.585810e^{-18}$	0.003011	0.003008
$\theta_{x_2}^{\Delta B}$	Auto-DML	$-4.847226e^{-19}$	0.003829	0.003825
$\theta_{x_2}^{\Delta B}$	Reweighting	$-3.157197e^{-18}$	0.002978	0.002975

注：数据由作者自行整理。

接下来，本文比较了不同参数下三种放的结果，其中，图1—图3描述了为中的误差的分布 $\hat{\theta}_{x_0}^{\Delta B}$ 和 $\hat{\theta}^{\Delta G}$ 如果利息的影响是 $\theta_{x_0}^{\Delta B}$ ，则估计 $\hat{\theta}^{\Delta G}$ 导致不同的结果，因为变量 $x_{i,0}$ 在两组 Z_i 中不平衡。图中展示了与标准差为 $\theta_{x_0}^{\Delta G}$ 与 $\theta^{\Delta G}$ 对比的正态分布的标准差。图1显示了 DML 的结果，图2显示了 Auto-DML 的结果，图3显示了 Reweighting 方法的结果。

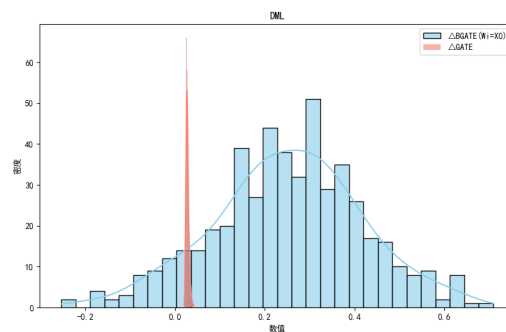


图1：DML 仿真结果

注：横坐标的数值是指 $(\hat{\theta}^{\Delta G} - \theta^{\Delta G})$ or $(\hat{\theta}^{\Delta B} - \theta^{\Delta B})$ 。

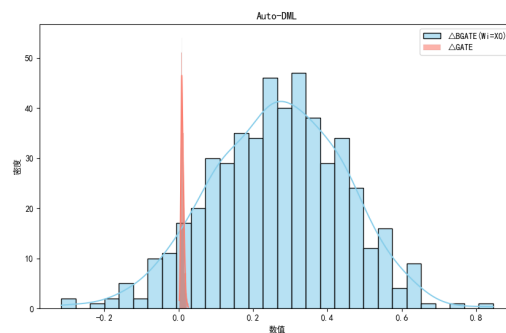


图2：Auto-DML 仿真结果

注：横坐标的数值是指 $(\hat{\theta}^{\Delta G} - \theta^{\Delta G})$ or $(\hat{\theta}^{\Delta B} - \theta^{\Delta B})$ 。

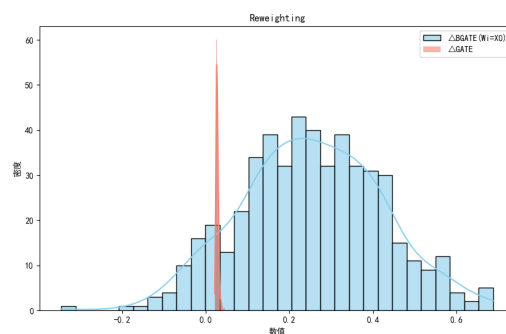


图3：Reweighting Approach 仿真结果

注：横坐标的数值是指 $(\hat{\theta}^{\Delta G} - \theta^{\Delta G})$ or $(\hat{\theta}^{\Delta B} - \theta^{\Delta B})$ 。

在图1—图3中，两种效应的重叠分布体现在两种颜色的混合的效果。与预期一致，估计 $\hat{\theta}^{\Delta G}$ 得出的结果与估计 $\hat{\theta}_{x_2}^{\Delta B}$ 相同，因为 $x_{i,2}$ 已经平衡。因此，如果协变量（或多个协变量）不平衡，则区分这两种效应非常重要。

五、结论

本文提出了 Δ BGATE 估计量以衡量不同组间处理效果差异，并考虑协变量差异，该方法适用于无混杂环境的处理效应的异质性。同时，提出 DML 估计量，并证明该估计量在标准条

件下的一致性和渐近正态性，同时提出两种替代估计策略：一是 Auto-DML，可避免 DML 估计量对极端估计倾向分数的非稳健性问题；二是 Reweighting 方法，可结合适用于重新加权数据 GATE 的一致估计量使用。仿真研究验证了这些估计策略的有效性，为理解政策决策的差异化影响提供支持。

未来研究可从多方面拓展：将估计方法拓展至连续处理和调

节场景；探索在处理变量、调节变量或两者均存在工具变量的设置下应用该方法；深入研究 RieszNet 的调整方式以及 Auto-DML 和重新加权估计量的渐近性质；开展更广泛的模拟研究，以全面了解估计量在有限样本条件下的特性。

参考文献

- [1] Chernozhukov V, Chetverikov D, Demirer M, et al. Double/debiased machine learning for treatment and structural parameters[J]. The Econometrics Journal, 2018, 21(01): C1–C68.
- [2] 钱浩祺, 龚嫣然, 吴力波. 更精确的因果效应识别：基于机器学习的视角[J]. 计量经济学报, 2021, 1(04): 867–891.
- [3] Lechner M, Mareckova J. Comprehensive Causal Machine Learning, arXiv preprint arXiv:2405.2024:10198.
- [4] Frölich M. Finite-Sample Properties of Propensity-Score Matching and Weighting Estimators[J]. The Review of Economics and Statistics, 2004, 86(01): 77–90.
- [5] Chernozhukov V, Newey W, Quintas-Martinez V M, et al. Riesznet and forestriesz: Automatic debiased machine learning with neural nets and random forests[J]. In International Conference on Machine Learning, 2022, : 3901–3914. PMLR.
- [6] 郭峰, 陶旭辉. 机器学习与社会科学中的因果关系：一个文献综述[J]. 经济学(季刊), 2023, 23(01): 1–17.
- [7] Abrevaya J, Hsu Y C, Lieli R P. Estimating Conditional Average Treatment Effects[J]. Journal of Business and Economic Statistics, 2015, 33(4): 00–00. DOI: 10.1080/07350015.2014.975555.
- [8] Lechner M. Modified Causal Forests for Estimating Heterogeneous Causal Effects[J]. Michael Lechner, 2018. DOI: 10.13140/RG.2.2.13540.22405.
- [9] Semenova V, Chernozhukov V. Debiased Machine Learning of Conditional Average Treatment Effects and Other Causal Functions[J]. The Econometrics Journal, 2021, 24(02): 264–289.
- [10] Zimmert M, Lechner M. Nonparametric estimation of causal heterogeneity under high-dimensional confounding[J]. Papers, 2019. DOI: 10.48550/arXiv.1908.08779.
- [11] Fan Q, Hsu Y C, Lieli R P, et al. Estimation of Conditional Average Treatment Effects With High-Dimensional Data[J]. Journal of Business And Economic Statistics, 2022, 40(01): 313–327.
- [12] Farbmacher H, Huber M, Langen H, et al. Causal mediation analysis with double machine learning[J]. The Econometrics Journal, 2022, 25, 277–300.
- [13] Rubin D B. Estimating causal effects of treatments in randomized and nonrandomized studies[J]. Journal of Educational Psychology, 1974, 66, 688.
- [14] Imbens W G, Wooldridge M J. Recent Developments in the Econometrics of Program Evaluation[J]. Journal of Economic Literature, 2009, 47(01): 5–86.
- [15] Imbens W G. Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review[J]. The Review of Economics and Statistics, 2004, 86(01): 4–29.
- [16] Belloni A, Chernozhukov V. Least squares after model selection in high-dimensional sparse models[J]. Bernoulli, 2013, 19(02): 521–547.
- [17] Bach P, Schacht O, Chernozhukov V, et al. Hyperparameter tuning for causal inference with double machine learning: A simulation study[J]. In Causal learning and reasoning, 2024: 1065–1117. PMLR.