arXiv:2303.01620v1 [stat.ME] 2 Mar 2023

使用贝叶斯决策树集成估计异质性因果中介效应

AngelaTing 和 Antonio R. Linero德
克萨斯大学奥斯汀分校统计与数据科学系

2023年3月6日

摘要

因果推断文献日益认识到，明确针对治疗效应异质性能够增进科学理解并改善政策建议。出于同样的目的，研究连接处理与结果的因果路径也很有帮助。本文在因果中介分析的背景下探讨这些问题。我们引入一种基于贝叶斯加性回归树的变系数模型，用于识别异质性因果中介效应并对其进行正则化；与线性结构方程模型类似，这些效应对应于协变量依赖的系数乘积。我们表明，即使在协变量较少的大型数据集上，LSEM 也可能产生极不稳定的条件平均直接效应和间接效应估计，而我们的贝叶斯因果中介森林模型能够产生稳定的估计结果。我们发现，我们的方法较为保守，其效应估计会“向同质性收缩”。我们利用医疗支出专题调查数据和基于实证的模拟数据，考察了该方法的重要性质。最后，我们展示了如何将模型与后验汇总策略相结合，以识别具有研究价值的亚组并解释模型拟合结果。

1 引言

利用观察性数据估计异质性因果效应是一项至关重要的研究课题，应用领域包括个性化医疗（Obermeyerand Emanuel，2016年）、政策建议（Athey，2017）和社会科学（Yeager 等人，2019年）。因果推断领域近来广受关注的一个问题是，如何充分利用机器学习领域开发的最先进预测算法来估计

异质性处理效应（K¨unzel 等人，2019年；Nie 和 Wager，2021；Hahn 等，2020年）。这方面的文献大多聚焦于一个问题：如何改进机器学习文献中使用的估计策略，使其适用于推断异质性因果效应。

制定更完善政策建议的一种补充方法，是了解我们关注的处理如何通过影响那些本身与结果存在因果联系的下游变量来影响结果；这称为因果中介分析，而这些中间变量称为中介变量（Robins 和 Greenland，1992年；Pearl，2001年；Rubin，2004年）。除了能更深入地理解其中发挥作用的因果机制外，我们还将看到，因果中介分析在某些情况下可以提高我们检测因果效应的能力。在这一领域，人们也提出了如何有效利用预测算法的类似问题，研究重点多在估计平均而非异质性中介效应（Farbmacher 等，2022年；Linero 和 Zhang，2022年；Zheng 和 van der Laan，2012年；Tchetgen 和 Shpitser，2012年；Kim 等，2017年）。

据我们所知，鲜有研究同时涉及这两种设定，即利用机器学习估计直接和间接因果中介效应层面的治疗效应异质性。在中介分析中估计治疗效应的异质性这一问题称为调节中介（Muller 等，2005年）。这一主题在社会科学文献中备受关注，研究中常采用线性结构方程建模（LSEM）。例如，（Preacher 等，2007年）和（Kershaw 等，2010年）分别将基于 LSEM 的调节中介应用于教育学和健康心理学领域的问题。

以非参数方法估计异质性中介效应是一项艰巨的任务，既依赖于关于混杂的强假设，也需要大量数据才能可靠地估计因果效应。在这一背景下，有一些重要挑战亟待解决，包括：(i) 确定如何恰当地对干扰参数和关注参数进行正则化，以确保结果合理；(ii) 开发

能够以有意义的方式总结黑箱拟合程序结果的方法；以及 (iii) 建立可靠的技术，以识别有证据表明存在调节中介的亚组，并确定哪些变量充当效应修饰因子。

本文提出了贝叶斯因果森林（BCF）算法的一种双层扩展，用于估计异质性中介效应。该方法将针对中介变量的标准 BCF 模型与针对结果的变系数 BART 模型相结合（Hahn 等，2020年；Deshpande 等人，2020年）。我们的方法受到 BCF 在因果推断竞赛和实际应用中的优异表现启发（Dorie 等人，2019年）。我们的方法直接以处理对结果的直接效应和间接效应为参数来构建模型。这使我们能够“向同质性收缩”，从而稳定中介效应的估计。在治疗效应近乎同质的情形下，我们的方法表现极为出色：个体层面的中介效应均方根误差较小，而且对大多数个体而言，可信区间的覆盖率接近名义水平。因此，我们提出的方法为估计异质性中介效应提供了强有力的工具。

1.1 医疗支出面板调查

医疗支出专题调查（MEPS）是由医疗保健研究与质量局实施的一项持续开展的大规模调查，旨在衡量患者、医院和保险公司对医疗保健系统的使用情况。为展示我们提出的方法，我们利用 MEPS 研究吸烟对健康的后果。具体而言，我们旨在回答以下问题：（i）吸烟是否会对整体医疗保健支出产生因果影响？（ii）吸烟对整体健康的影响在多大程度上介导（或未介导）了这种影响？（iii）是否存在影响吸烟与医疗支出之间关联的调节变量？

在第4节中，我们对该数据集进行了分析，并得出了一个令人意外的发现：吸烟对医疗支出的总因果效应，可能会被直接效应估计所产生的不稳定性掩盖

，这种不稳定性源于对吸烟对医疗保健费用的直接效应进行估计。尽管人们可能凭直觉认为，吸烟对医疗支出的影响完全由其对健康的影响所中介，但我们在序贯可忽略性假设下的分析表明，吸烟对支出的估计直接效应为负，并在很大程度上抵消了吸烟对支出的正向间接效应；这一直接效应很可能是由我们尚未纳入分析的其他变量所导致的。此外，我们还识别出多个能够调节吸烟对支出间接效应的变量，其中年龄最为重要。

1.2 文章结构

在第2 节中，我们回顾中介作用的潜在结果框架、序贯可忽略性假设、贝叶斯加性回归树（BART）框架以及贝叶斯因果森林（BCF）。在第3 节中，我们定义贝叶斯因果中介森林模型，并说明如何使用该模型稳定地估计直接效应和间接效应。在第4节中，我们运用所提出的方法分析 MEPS 数据，研究吸烟对健康的影响如何进一步影响医疗保健支出，以及这种中介效应的异质性；此外，我们还开展了一项基于实证设计的模拟研究，以说明该方法在估计平均中介效应和条件平均中介效应时，在覆盖率和估计误差方面表现良好。最后，我们在第5 节中讨论相关问题及可能的扩展。计算细节和更多模拟结果见补充材料。

2 中介分析与 BART 回顾

2.1 中介分析概述

中介作用是指处理（A）通过作用于中间的中介变量（M），从而影响结果（Y）的过程；该过程发生在处理与结果之间，其图示见图1。例如，我们考虑

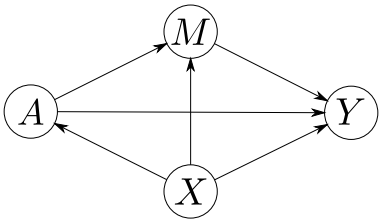

图 1：处理 A、中介变量 M、结果 Y 以及混杂因素/效应修饰因子 X 的示意图。箭头表示因果关系的方向。

吸烟是否会直接影响医疗支出，以及是否会通过影响健康而间接影响医疗支出这一问题。这里，吸烟状态是二元处理（A），所关注的结果是医疗支出的对数（Y）。我们的目标是将吸烟对医疗支出的影响分解为吸烟的直接效应，以及由吸烟对整体健康的影响所中介的间接效应（整体健康以个人自我感知健康质量进行衡量）。一种自然的假设是，吸烟并不会直接导致更高的医疗支出，而是通过降低一个人的整体健康来实现这一影响。健康处于处理（吸烟）与结果（医疗支出）之间的因果路径上，因此是一个中介变量。

中介分析已应用于许多科学领域，包括MacKinnon andDwyer，1993年；Rubin，2004年；MacKinnon，2008年；Albert，2008年；Imai 等人，2010年；VanderWeele，2016年）。大量相关文献聚焦于结构方程模型（SEMs），通过参数模型中系数的乘积来量化中介效应。尤其是，线性结构方程模型（线性结构方程模型）得到了广泛应用（Baron and Kenny，1986年；MacKinnon and Dwyer，1993年；MacKinnon，2008年）。

线性结构方程模型的一个主要局限是，中介效应的识别依赖于特定参数模型的选择及其正确设定，从而限制了其适用性。为解决这一局限，Imai 等人（2010年）提出了一种基于潜在结果（Rubin，2004年，1974年）的非参数方法，在序贯可忽略性假设下识别平均因果中介效应。该假设

指出，给定协变量后，处理与结果和中介变量的所有潜在取值相互独立；给定观测到的处理和协变量后，观测到的中介变量与所有潜在结果相互独立。通过避免参数假设，这一框架提供了一种通用的估计程序，不受结果和中介变量模型选择的限制，因此适用于广泛的设定。

对于个体 $i = 1, ..., n$ 和处理 $a \in \{0, 1\}$，将潜在结果 $M_i(a)$ 定义为：若该个体接受处理 a，本可观察到的中介变量值。请注意，对于每个个体，$M_i(0)$ 和 $M_i(1)$ 中只有一个能够实际观察到。对于接受处理的个体（$A_i = 1$），$M_i(0)$ 是一个反事实，即若该个体未接受处理，本可观察到的中介变量值。类似地，潜在结果 $Y_i(a, m)$ 是指若该个体接受处理 a 且中介变量处于水平 m，本可观察到的结果值。例如，$Y_i\{0, M_i(1)\}$ 是指若该个体未接受处理，且其中介变量值处于其接受处理时本会达到的水平，本可观察到的结果值。我们通过一致性 假设将潜在结果与观测数据联系起来。该假设指出，我们观测到中介变量 $M_i = M_i(A_i)$ 和结果 $Y_i = Y_i\{A_i, M_i(A_i)\}$。由于 Y_i 和 M_i 的取值仅由个体 i 可能接受的处理 a 决定（而不取决于其他个体接受的处理），这种记法也隐含地表明单位间不存在干扰；这称为稳定单位处理值 （SUTVA）假设。

利用这些潜在结果，我们可以定义所关注的因果估计。在因果中介分析中，我们尤其关注自然直接效应和自然间接效应的估计（Pearl，2001年；Robins 和 Greenland，1992年）。自然直接效应定义为

$$\zeta_a = E[Y_i\{1, M_i(a)\} - Y_i\{0, M_i(a)\}] \tag{1}$$

而自然间接效应定义为

$$\delta_a = E[Y_i\{a, M_i(1)\} - Y_i\{a, M_i(0)\}]. \tag{2}$$

自然直接效应将潜在中介变量固定，从而单独考察处理的效应；它可以解释为处理对结果的直接影响 Y_i。相反，自然间接效应将处理固定，单独考察潜在中介变量对不同处理值的响应所产生的效应；它可以解释为处理通过中介变量 M_i间接影响结果的效应 Y_i。处理对结果的总效应是直接效应和间接效应之和，可定义为

$$\tau = \zeta_0 + \delta_1 = \zeta_1 + \delta_0 = E[Y_i\{1, M_i(1)\} - Y_i\{0, M_i(0)\}]. \tag{3}$$

类似地，我们可以定义直接效应和间接效应的条件平均 变体，形式如下：

$$\begin{aligned} \zeta_a(x) &= E[Y_i\{1, M_i(a)\} - Y_i\{0, M_i(a)\} \mid X_i = x] \qquad \text{and} \\ \delta_a(x) &= E[Y_i\{a, M_i(1)\} - Y_i\{a, M_i(0)\} \mid X_i = x]. \end{aligned} \tag{4}$$

我们主要关注式（4）所定义的条件平均直接效应和间接效应。

2.2 假设

令陈述 $[A \perp\!\!\!\perp B \mid C = c]$ 表示 A 在给定 B 的条件下与 $C = c$条件独立；令 $\mathcal{X}$ 表示 X_i的样本空间，并令 $\mathcal{M}$ 表示M_i的样本空间。遵循 Imai 等人 (2010年)，我们始终作出以下序贯可忽略性 假设，从而可以识别直接效应和间接效应。

SI1 $\{Y_i(a',\ m)\ M_i(a)\} \perp\!\!\!\perp A_i \mid X_i = x$ 对于 $a,\ a' = 0,\ 1$ 以及所有 $x \in \mathcal{X}$。

SI2 $Y_i(a',\ m) \perp\!\!\!\perp M_i(a) \mid A_i = a,\ X_i = x$ 对于 $a,\ a' = 0,\ 1$ 以及所有 $x \in \mathcal{X}$。

SI3 $\Pr(A_i = a \mid X_i = x) > 0$ 和 $f\{M_i(a) = m \mid A_i = a,\ X_i = x\} > 0$ 对于 $a = 0,\ 1$ 以及所有$x \in \mathcal{X}$ 和 $m \in \mathcal{M}$。

第一个假设指出，给定协变量后，处理分配是可忽略的，即它与潜在结果和潜在中介变量相互独立。当个体被随机分配到处理组和对照组时，该假设会自动满足；但在观察性研究中，并不能保证该假设成立。在这种情况下，研究人员通常会尽可能收集更多处理前混杂因素，从而在处理组之间的协变量差异得到控制后，处理分配的可忽略性是合理的。第二个假设指出，给定观测到的处理和协变量后，中介变量是可忽略的，即它与潜在结果相互独立。然而，即使在随机试验中，也不能保证该假设成立。一般而言，无法直接根据数据检验该假设。第三个假设是关于处理和中介变量的正值性假设，即接受处理和对照的概率应为非零。

在 SI1–SI3 条件下，我们可以非参数地识别任意反事实结果 $Y_i\{a', M_i(a)\}$ 的分布，如下所示

$$\begin{aligned} &f(Y_i\{a, M_i(a')\} = y \mid X_i = x) \\ &= \int_{\mathcal{M}} f(Y_i = y \mid M_i = m, A_i = a, X_i = x)\, f(M_i = m \mid A_i = a, X_i = x)\ dm \end{aligned} \tag{5}$$

对于任意 $x \in \mathcal{X}$和 $a,\ a' = 0,\ 1$（Imai 等人，2010年，定理 1）。这使我们能够利用观测到的结果和中介变量（右侧）推断未观测到的反事实（左侧）。此外，（5）不依赖于特定的参数模型，因此可应用于灵活的（非参数）模型。

2.3 贝叶斯加性回归树综述

我们将采用Chipman 等人（2010年）提出的贝叶斯加性回归树（BART）模型。考虑一个未知函数 r，它使用输入向量 X_i来预测输出 Y_i

$$Y_i = r(X_i) + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2). \tag{6}$$

BART 模型 $r(x)$ 将其表示为 m 回归树 $r(x) = \sum_{j=1}^{m} g(x; T_j, M_j)$ 的加和，其中 T_j 是一棵二叉决策树，由内部节点的决策规则以及一组终端节点构成，而 $M_j = \{\mu_{j1}, \ldots, \mu_{jb_j}\}$ 是与树 T_j 的每个 b_j终端节点相关联的一组参数值。每个 x 对应于 T_j 的一个终端节点 k，并被赋予值 $g(x; T_j, M_j) = \mu_{jk}$。根据（6），$E(Y_i \mid X_i = x)$ 等于分配给 x 的所有终端节点 μ_{jk} 之和，这些节点由 $g(x; T_j, M_j)$ 确定。关于 BART 及其应用的全面综述，请参见 Hill 等人（2020年）。

要应用 BART，必须为树加和模型的所有参数指定先验分布，即 (T_j, M_j)，其中 $j = 1, \ldots, m$。该先验应通过防止单棵树效应产生过大的影响来正则化拟合。先验由两个部分组成：每棵树 T_j 的先验，以及终端节点 $M_j \mid T_j$ 的先验，其中$\pi(T_j, M_j) = \pi_T(T_j)\pi_M(M_j \mid T_j)$。随后，BART模型设定 $(T_j, M_j) \overset{\text{iid}}{\sim} \pi(T, M)$。

先验 π_T 由三个变量决定：(i) 给定节点为内部节点的概率；(ii) 每个内部节点处分裂变量分配的分布；以及 (iii) 在给定分裂变量的条件下，每个内部节点处分裂规则分配的分布。对于 (i)，深度为 d 的节点是内部节点的概率为

$$\alpha(1 + d)^{-\beta}, \quad \alpha \in (0, 1), \beta \in [0, \infty) \tag{7}$$

其中 $\alpha = 0.95$ a 并且 $\beta = 2$ 是一种偏向较小树的默认设置。对于 (ii) 和 (iii), e

每个内部节点的分裂变量分配分布，以及给定分裂变量后每个内部节点的分裂规则分配分布，均采用均匀先验。

对于终端节点的先验，我们假设 $\pi(M_j \mid T_j) = \prod_{k=1}^{b_j} \pi_\mu(\mu_{jk})$。为指定$\sigma_\mu$，本文首先对 Y_i进行平移和缩放，使 Y_i 的均值为 0、方差为 1。然后采用先验

$$\pi_\mu(\mu_{jk}) = N(\mu_{jk} \mid 0, \sigma_\mu^2) \qquad \text{where} \qquad \sigma_\mu = \frac{3}{k\sqrt{m}}$$

其中 k 取适当的值，默认值为 $k = 2$。请注意，该先验会将终端节点值 μ_{jk} 向零收缩，并且随着树的数量 m 增加而施加更强的收缩，从而确保每棵树在树集成中都是一个弱学习器。

2.4 用于因果推断的贝叶斯决策树集成

BART 在因果推断问题中表现尤为出色，可用于推断异质性治疗效应和平均处理效应（Hill，2011年；Wendling 等人，2018；Dorie 等人，2019年）。对于结果 Y_i、二元处理 A_i以及混杂变量/效应修饰变量 X_i，Hill（2011年）提出了如下模型

$$Y_i(a) = \mu(X_i, a) + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2). \tag{8}$$

因此，接受处理的效应为

$$E\{Y_i(1) - Y_i(0) \mid X_i = x\} = \tau(x) = \mu(x, 1) - \mu(x, 0).$$

鉴于 BART 具有出色的预测性能，Hill (2011) 建议对$\mu(\cdot, \cdot)$ 使用 BART 先验，以灵活地建模结果，从而获得灵活的处理效应估计。

Hahn 等 (2020年) 指出，成功的预测建模在很大程度上取决于谨慎的正则化，并在 Hill (2011) 的研究基础上指出了模型（8）的两个缺点：首先，倾向得分与 $\mu(x,a)$ 之间的相关性可能导致 正则化诱发的混杂 (RIC)，进而导致因果估计存在严重偏差；其次，基于参数化（8）的先验编码了治疗效应高度非同质的先验信息。为缓解 RIC，他们提出了一种依赖于倾向得分估计值的先验 $\widehat{\pi}_i$ ，将其作为协变量的一维概括；为解决非同质性问题，他们则对回归模型重新参数化如下：

$$Y_i(a) = \mu(X_i, \widehat{\pi}_i) + a\,\tau(X_i) + \epsilon_i$$

其中， $\mu(x,\widehat{\pi})$ 表示控制变量 X_i 的预后效应，而 $\tau(x)$ 则是处理效应本身。随后，我们分别为 $\mu(\cdot)$ 和 $\tau(\cdot)$ 设定相互独立的 BART 先验，其中针对 $\tau(\cdot)$ 的先验体现了我们对治疗效应异质性程度的先验信念。

Linero 和 Zhang(2022年) 研究了使用 BART 模型估计因果中介中的直接效应和间接效应。结合 RIC 的概念，他们还指出，未经审慎指定的先验可能具有高度的 教条性 (Linero，2021)，也就是说，这些先验编码了这样一种先验信念：平均而言，中介变量和结果的潜在结果之间不存在混杂（因此，对平均中介效应的推断可以假设不存在混杂）。先验教条性会导致正则化诱导的混杂，因为它会强烈倾向于促使模型将结果上的因果效应归因于处理，而非混杂因素。为了解决这个问题，他们将“巧妙协变量” $\widehat{m}_{ai} = E(M_i \mid A_i = a,\ X_i)$ 纳入$a \in \{0,1\}$的结局模型中。这些巧妙协变量类似于 $\widehat{\pi}_i$ BCF 中用于校正 RIC 的倾向得分估计。Linero 和 Zhang(2022年) 随后提出了贝叶斯因果

中介森林(BCMF) 模型

$$
\begin{aligned}
Y_i(a,m) &= \mu_y(m,a,X_i) + \epsilon_i \\
M_i(a) &= \mu_m(a,X_i) + \epsilon_i
\end{aligned}
\tag{9}
$$

其中，函数 $\mu_y(\cdot,\ 0,\ \cdot)$、$\mu_y(\cdot,\ 1,\ \cdot)$、$\mu_m(0,\ \cdot)$ 和 $\mu_m(1,\ \cdot)$ 分别设定为相互独立的 BART 先验，而巧妙协变量 $\widehat{m}_{0i}$ 和 $\widehat{m}_{1i}$ 则作为预测变量纳入结果的 BART 模型。

尽管模型（9）能够很好地实现估计平均中介效应的目标（即解决 RIC 问题），但它无法恰当地控制条件平均中介效应的异质性程度。本研究的一项贡献是借鉴 BCF 的参数化思路，开发出一种分别对直接效应和间接效应进行正则化的模型。

3 异质性中介效应的BART

现在我们介绍因果中介分析模型；补充材料中给出了用于拟合该模型的“贝叶斯反向拟合”算法。与 BCF 类似，所提出的模型能够对以下内容进行直接正则化：$\delta_a(x)$ 和 $\zeta_a(x)$。这种直接正则化已被证明对于在其他情境下生成可靠的异质性因果效应估计至关重要（Hahn 等，2020年；Nie 和 Wager，2021）。

对于数值型结果和中介变量，我们设定模型

$$
Y_i(a,m) = \mu(X_i) + a\,\zeta(X_i) + m\,d(X_i) + \epsilon_i, \tag{10}
$$

$$
M_i(a) = \mu_m(X_i) + a\,\tau_m(X_i) + \nu_i \tag{11}
$$

其中，为 $(\mu, \zeta, d, \mu_m, \tau_m)$ 指定了相互独立的 BART 先验。中介变量模型 (11) 即为 Hahn 等 (2020年) 提出的 BCF 模型，其中 $\tau_m(x)$

对应于处理对结果的异质性因果效应。另一方面，结局模型 (10) 对应于 变化系数 *BART* (VC-BART) 模型，该模型由 Deshpande 等人(2020年) 提出，其中处理变量 A_i 和中介变量 M_i 以线性形式纳入模型。

模型（10）–（11）是常用线性结构方程模型的变系数版本，其中系数被建模为 X_i 的非参数函数。因此，条件平均中介效应也可以表示为系数的乘积：

$$\zeta_a(x) = \zeta(x), \qquad \text{and}$$
$$\delta_a(x) = \tau_m(x)\, d(x).$$

因此，这种参数化使我们能够分离出 $\zeta(x)$ 和 $\delta(x)$ 这两个组成部分，并对它们施加不同程度的正则化。请注意，（10）–（11）假设结果模型中中介变量与处理之间不存在交互作用，因此 $\delta_a(x)$ 和 $\zeta_a(x)$ 不依赖于处理水平 a，即 $\delta_0(x) = \delta_1(x)$，并且 $\zeta_0(x) = \zeta_1(x)$。

对于平均效应，注意， $Y_i\{a, M_i(a')\}$ 的边际分布为

$$f\big(Y_i\{a, M_i(a')\} = y\big) = \int f\big(Y_i\{a, M_i(a')\} = y \mid X_i = x\big)\, f(X_i = x)\, dx. \tag{12}$$

因此，有必要为协变量的分布指定一个模型。通常，在未对该分布进行显式建模时，会改用经验分布作为估计，即 $F_X(dx) = \sum_i \omega_i\, \delta_{X_i}(dx)$，其中 $\delta_x(\cdot)$ 表示在 x和 $\omega_i = n^{-1}$处的点质量分布。经验分布的一种替代方案是贝叶斯自助法（BB；Rubin，1981年），它既能体现我们对 F_X 固有的不确定性，又能避免对协变量分布建模的需要。BB 与经验分布类似，但我们不再设定 $\omega_i = n^{-1}$ ，而是使用非正常先验 $\pi(\omega) = \prod_i \omega_i^{-1}$ ；由此得到权重的后验分布 $\omega \sim$狄利克雷分布$(1, \ldots, 1)$。在 BB 下，平均
效应可识别为 $\delta = \sum_i \omega_i\, \delta(X_i)$ 和 $\zeta = \sum_i \omega_i\, \zeta(X_i)$。

3.1 通过先验设定控制异质性

模型（10）–（11）相较于 Linero 和 Zhang(2022年) 的模型，一个重要优势在于，我们可以通过审慎选择超参数，使模型拟合结果向同质的中介变量效应和治疗效应收缩；完成这一步后，我们便可以确信，检测到的任何异质性确实有数据充分支持，而非非参数估计器使用不稳定所致。

直接效应的异质性程度可以通过以下对象的先验设定来控制： $\zeta(x)$。具体而言，我们可以通过以下方式将 $\zeta(x)$收缩为仅含少量效应调节变量的常数函数：(i) 将（ α ）中的参数7设为较小值（例如 $\alpha = 0.5$），使得大多数树不包含协变量；(ii) 使用较少数量的树（例如 $m = 20$）。采用同样的策略，我们可以控制 $\tau_m(x)$中的异质性程度，该量表示处理对中介变量的因果效应。

间接效应的考量稍微复杂一些，因为 $\delta(x) = d(x)\,\tau_m(x)$由两个部分组成。请注意，如果$\tau_m(x)$不是常数函数，那么 $\delta(x)$中的异质性不可避免。然而，如果 $\tau_m(x)$是常数，那么通过将 $d(x)$收缩至常数函数，可以使 $\delta(x)$变得同质。因此，对于$d(x)$，我们采用与 $\zeta(x)$和 $\tau_m(x)$相同的策略：使用较少数量的树，并将 α 设为较小值。

3.2 对非数值数据进行建模

还可以通过使用非参数 probit 回归模型，轻松纳入二元中介变量

$$[M_i(a) \mid X_i = x] \sim \text{Bernoulli}[\Phi\{\mu_m(x) + a\,\tau_m(x)\}],$$

并再次对 (μ_m, τ_m) 使用 BART 先验。在满足序贯可忽略性和 (10) 的条件下，我们可以将中介效应识别为

$$\zeta_a(x) = \zeta(x) \qquad \text{and} \qquad \delta_a(x) = d(x)\left[\Phi\{\mu_m(x) + \tau_m(x)\} - \Phi\{\mu_m(x)\}\right]$$

其中，$\Phi(\cdot)$ 是标准正态随机变量的累积分布函数。注意到，当 $[Y_i(a, m) \mid X_i = x] \sim$ 伯努利$[\Phi\{\mu(x) + a\,\zeta(x) + m\,d(x)\}]$ 且 $[M_i(a) \mid X_i = x] \sim N\{\mu_m(x) + a\,\tau_m(x),\ \sigma_m^2\}$ 时，也可以计算直接效应和间接效应的类似表达式。

$$E[Y_i\{a, M_i(a')\} \mid X_i = x] = \Phi\left(\frac{\mu(x) + a\,\zeta(x) + \{\mu_m(x) + a'\,\tau_m(x)\}\,d(x)}{\sqrt{1 + d^2(x)\,\sigma_m^2}}\right),$$

注意到 Probit模型意味着 $E[Y_i\{a, M_i(a')\} \mid X_i = x] = \Pr(\epsilon_i \le \mu(x) + a\,\zeta(x) + M_i(a')\,d(x) \mid X_i = x)$，即可推导出该式，其中 $\epsilon_i \sim N(0, 1)$，且 ϵ_i 与 $M_i(a')$ 相互独立。例如，这意味着当 Y_i 为二元变量且 M_i 为连续变量时，我们有

$$\delta_a(x) = \Phi\left(\frac{\mu(x) + a\,\zeta(x) + d(x)\,\{\mu_m(x) + \tau_m(x)\}}{\sqrt{1 + d^2(x)\,\sigma_m^2}}\right) - \Phi\left(\frac{\mu(x) + a\,\zeta(x) + d(x)\,\mu_m(x)}{\sqrt{1 + d^2(x)\,\sigma_m^2}}\right).$$

3.3 正则化导致的混杂偏差修正 g

如 Linero 和 Zhang（2022年）所示，用于中介作用的贝叶斯非参数模型也会受到与观察性数据模型相同的 RIC 现象影响。他们提出了以下建议来应对这一问题：

1. 在结局模型和中介变量模型中都加入倾向得分 $\hat{\pi}_i = \Pr(A_i = 1 \mid X_i = x)$ 的估计值。

2. 在针对 $a \in \{0, 1\}$ 的结果回归中加入中介变量回归函数 $\hat{m}_{ai} = E(M_i \mid A_i = a, X_i = x)$ 的估计值。

参见 Linero 和 Z hang (2022年) 详细讨论了为什么有必要纳入 e

这些变量，并进行全面的模拟实验。原则上，$(\hat{\pi}_i, \hat{m}_{0i}, \hat{m}_{1i})$ 的获取方式并不重要；唯一需要注意的是，$\hat{\pi}_i$ 应仅取决于 (A_i, X_i)，而$\hat{m}_{ai}$ 应仅取决于 (M_i, A_i, X_i)；我们使用 BART 来估计这些量。

3.4 后验总结

除了（条件）平均中介效应之外，生成对BCMF模型拟合数据的可解释总结也很有意义。这些总结有助于识别对处理反应不同的亚群体，帮助我们解释效应调节变量的影响，并深入了解BCMF的预测过程。

Woody等人（2021）提出了一个后验总结的一般框架，其基础是将复杂模型投影到可解释的替代模型上。例如，我们可以将 δ（x）的样本投影到一个加性函数 γ（x）$= \alpha + \sum_{j=1}^{p} \gamma_j$（$x_j$）上，其思想是，如果 γ（x）能够很好地近似 δ（x），那么我们就可以利用γ（x）的可解释结构来理解 δ（x）如何进行预测。

为简单起见　我们聚焦于间接效应 $\delta(x)$，并考虑两类总结　　：

- 一个加性函数，构造为 $\hat{\gamma} = \arg\min_{\gamma} \sum_i \{\delta(X_i) - \gamma(X_i)\}^2 + q_\lambda(\gamma)$，其中$\gamma(x)$ $= \alpha + \sum_j \gamma_j(x_j)$，且每个 $\gamma_j(x_j)$都是一个单变量样条。这里，$q_\lambda(\gamma) = \sum_j q_\lambda(\gamma_j)$是各个加性组成部分的粗糙度惩罚。该总结可以通过拟合广义加性模型（GAM，参见Wood，2006年的综述）来计算，其中 $\{\delta(X_i)$：$i = 1$、...、$n\}$ 作为结果，$\{X_i$：$i = 1$、...、$n\}$ 作为预测变量。

- 一份决策树总结，其中 $\hat{\gamma}$ 通过运行分类与回归树算法构建（Breiman 等人，1984），将 $\{\delta(X_i)$ 视为结果：$i = 1$，.....，$n\}$ 作为结果，并将 $\{X_i$：$i = 1$，.....，$n\}$ 作为预测变量。

分类与回归树总结有助于识别治疗效果存在显著差异的亚群体，而广义加性模型总结有助于单独了解不同

预测变量对估计效果的影响。关于这一方法在4 MEPS数据集上的示例，请参见第 4 节。

作为衡量摘要质量的总体指标，我们使用 $\delta(x)$ 与 $\gamma(x)$ 之间相关性的平方，其定义为

$$R^2 = 1 - \frac{\sum_i \{\delta(X_i) - \widehat{\gamma}(X_i)\}^2}{\sum_i \{\delta(X_i) - \widehat{\delta}\}^2}$$

其中 $\widehat{\delta} = n^{-1} \sum_i \delta(X_i)$; Woody 等人 (2021) 将 R^2 称为“摘要 R^2”。

4 医疗支出专题调查数据

现在，我们将模型应用于医疗支出专题调查（MEPS）的一个子集。我们关注以下问题：(i) 吸烟是否会对个人预期年度医疗支出产生因果效应；(ii) 是否有证据表明，这一效应完全由吸烟对健康的影响所中介；以及 (iii) 所提出的混杂因素是否也会修饰间接效应。我们将结果 Y_i 设定为个人在2012年调查中报告的年度净医疗支出的对数，将处理 A_i设定为个人是否吸烟（是或否），并将中介变量 M_i 设定为对整体自我感知健康状况的有序测量（1：极好，2：很好，3：好，4：一般，5：差）。

首先，我们注意到，对吸烟者与非吸烟者的医疗支出差异进行简单的双样本 t 检验后发现，有强有力的证据（P 值< 0.0005）表明，非吸烟者的医疗支出更多。因此，在评估任何因果关系时，控制混杂因素十分重要。我们的模型将以下患者属性纳入为可能的混杂因素：

- `age`：年龄（岁）。
- `bmi`：体重指数，可能是健康与医疗支出的处理后混杂因素。

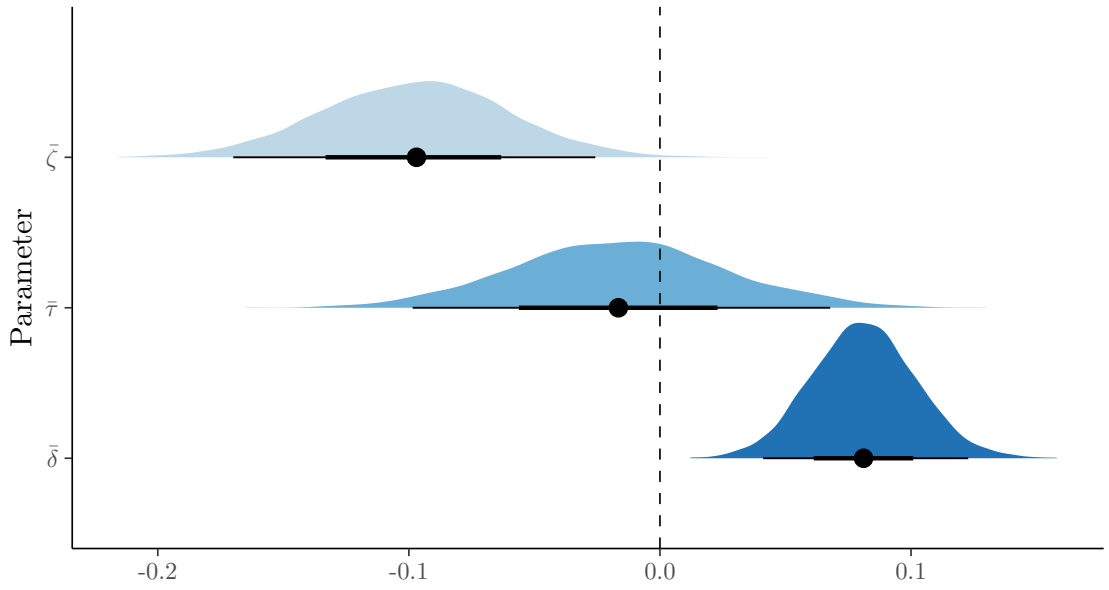

图 2：平均直接效应 $\bar{\zeta}$（顶部）、平均间接效应 $\bar{\delta}$（底部）和平均总效应 $\bar{\tau} = \bar{\zeta} + \bar{\delta}$（中部）的后验分布。

- `education_level`：受教育年限。
- `income`：每年的家庭总收入。
- `poverty_level`：家庭收入占贫困线的百分比。
- `region`：美国东北部、美国西部、美国南部或美国中西部。
- `sex`：男性或女性。
- `marital_status`：已婚、离婚、分居或丧偶。
- `race`：白人、太平洋岛民、原住民、黑人、亚裔或多种族。
- `seatbelt`：个人在车内是否系安全带（总是、几乎总是 ays, 有时、从不、很少、从不驾驶或乘坐汽车）。

平均直接效应和间接效应的后验分布见图2。我们看到，在序贯可忽略性假设下，有证据表明存在直接效应

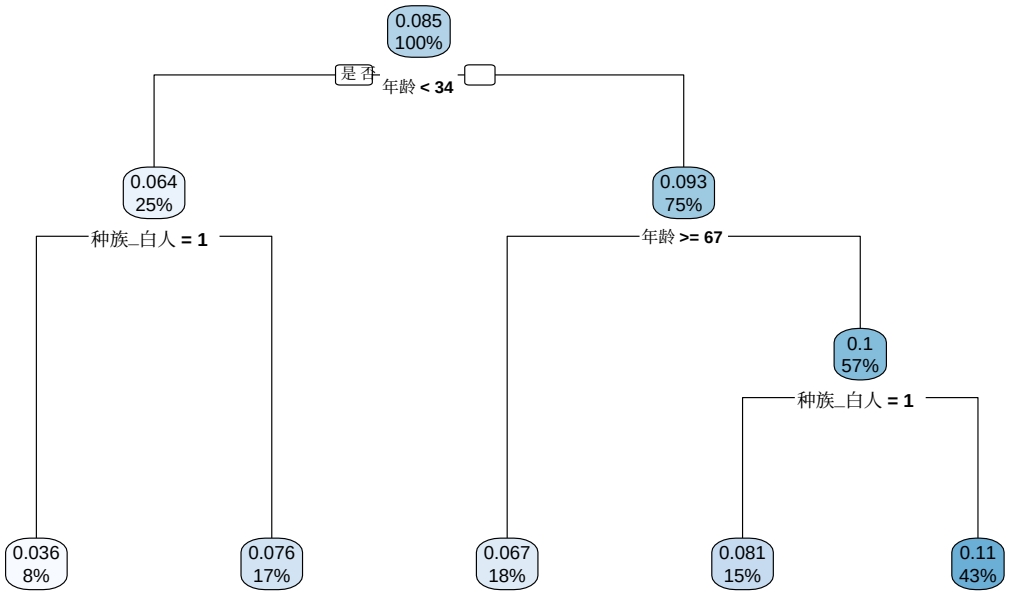

图3：P　　使用单棵回归树对间接效应进行后验总结。

和吸烟对支出的间接效应。有趣的是，这些效应的方向相反，并在很大程度上相互抵消。因此，总效应的符号并不确定。这说明了中介分析的一项重要潜在优势：我们能够确立吸烟与医疗支出之间的因果关系，而如果我们只关注总效应，就无法做到这一点。

4.1 后验总结

我们采用第 3.4节概述的总结策略来解读模型拟合，并更好地理解导致间接效应异质性的协变量和交互作用；具体而言，我们将间接效应函数 $\delta(x)$ 投影到一棵回归树和一个加性函数上。

首先，我们考虑在对 MEPS数据集进行初步模型拟合后得到的 $\delta(x$后验均值的分类与回归树（CART）摘要。根据图3中的回归树摘要，种族、年龄和性别是间接效应最显著的效应修饰因子。

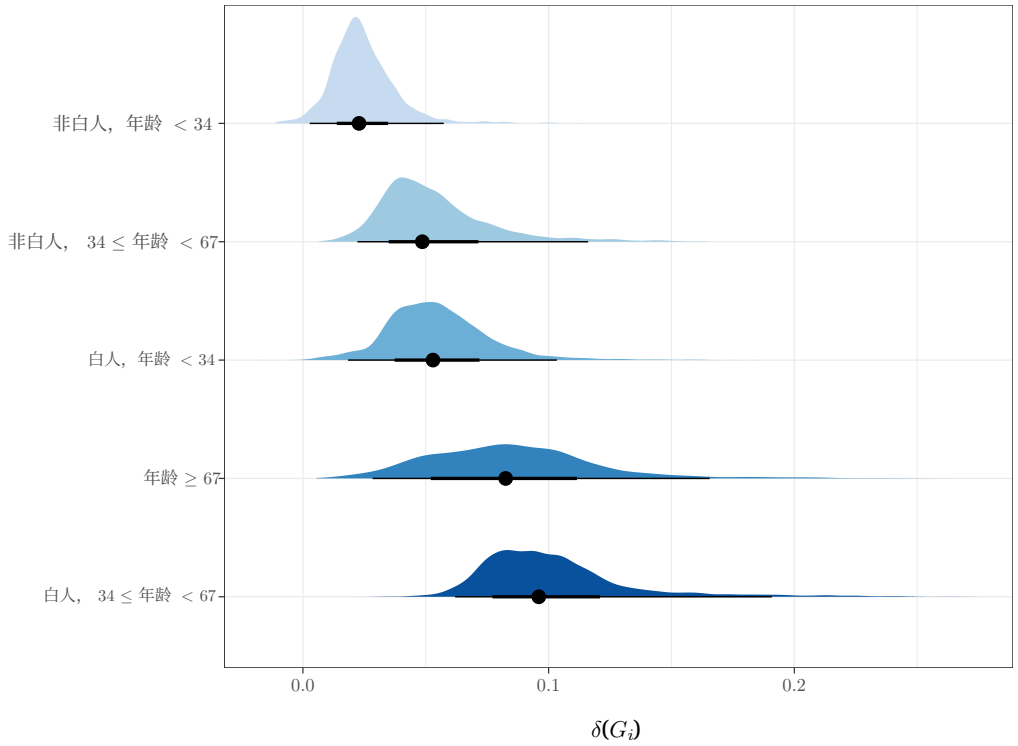

图4：各亚组内平均间接效应的后验密度 G_1、 G_2、..... G_5来自图3中的终端节点。

受图3中亚组划分的启发，我们在图4 展示了按年龄和种族划分的不同亚组中的平均间接效应。我们发现，白人中年人的间接效应最大，而非白人年轻成年人的效应最小。

图5和图6分别展示了根据连续变量和离散变量的广义加性模型摘要得出的结果。这些图再次凸显了年龄和种族作为效应修饰因子的重要性，表明年龄较大且为白人的个体，其通过感知健康状况影响医疗支出的间接效应更高。

总之，分类与回归树和广义加性模型总结均表明，年龄和种族是重要的效应修饰因子。为衡量总结函数近似的充分性，图7展示了通过将总结拟合到 δ的每个后验样本而得到的 R^2 $(\cdot)$ 的后验分布，以及通过将总结拟合到 $\delta(\cdot$的后验均值而得到的单个 R^2 。我们的分析表明，回归树作为总结略优于

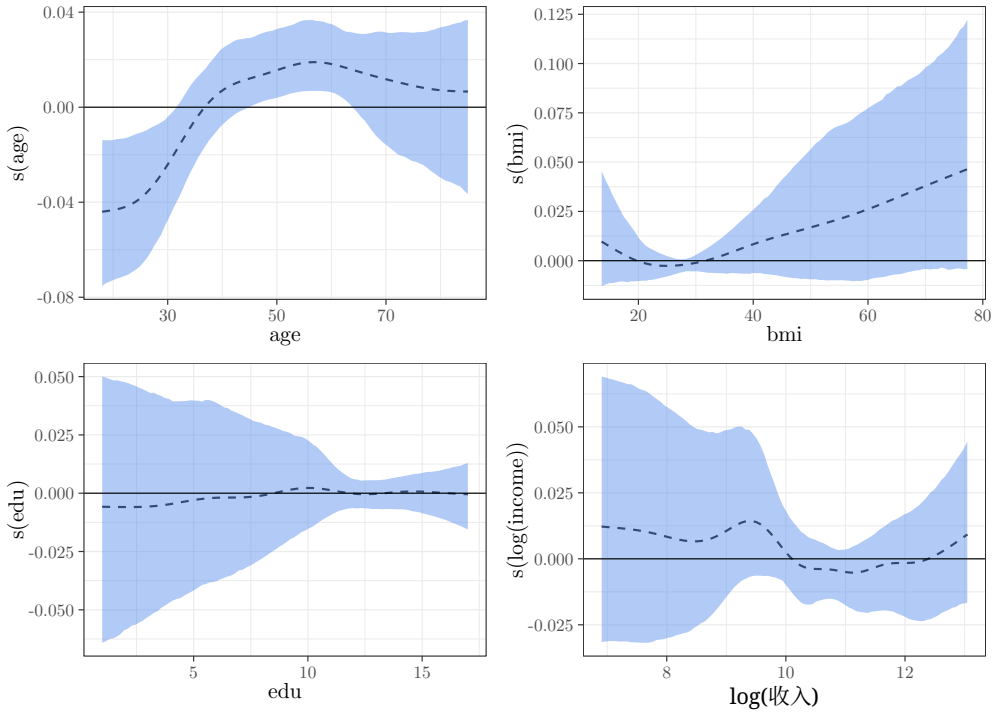

图5：使用广义加性模型对连续变量的间接效应进行后验总结。后验均值的投影以虚线表示，阴影区域则表示投影的95%可信带。

广义加性模型，这表明图3 中检测到的交互作用为理解模型对 $\delta(\cdot)$的预测过程提供了重要见解。

4.2 BCMF与LSEM的比较

为评估BCMF模型(10)–(11)，我们将其预测性能与处理、中介变量和协变量之间包含交互作用的LSEM进行比较，

$$\begin{aligned} Y_i(a,m) &= \beta_{0Y} + X_i^T\beta_Y + a(\gamma_{0Y} + X_i^T\gamma_Y) + m(\xi_0 + X_i^T\xi) + \epsilon_i \\ M_i(a) &= \beta_{0M} + X_i^T\beta_M + a(\gamma_{0M} + X_i^T\gamma_M) + \nu_i. \end{aligned} \tag{13}$$

该模型允许存在异质性中介效应，但将其限制为混杂因素的线性函数。为了量化LSEM估计的不确定性，我们使用残差

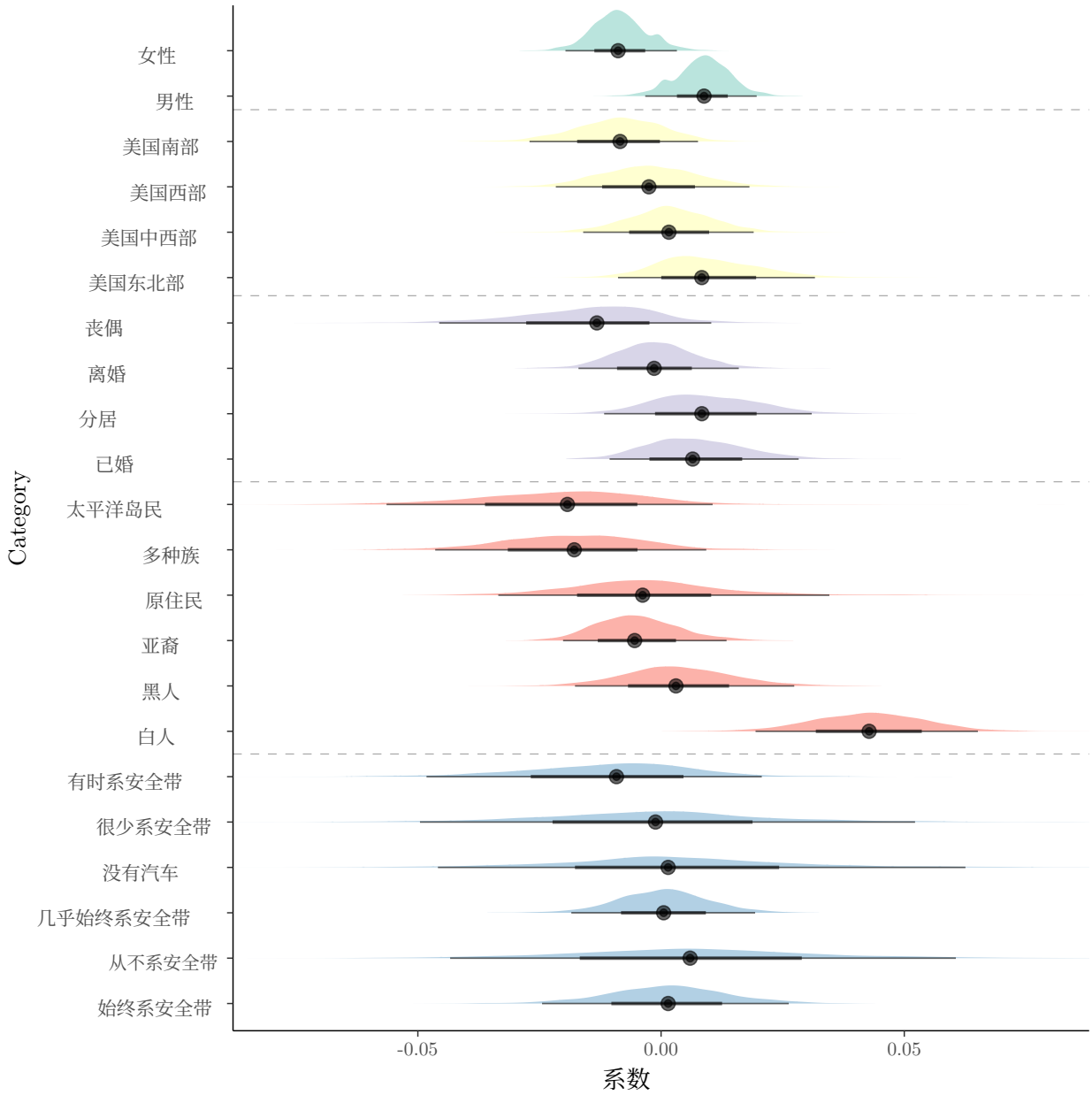

图6：使用GAM投影得到的分类变量对间接效应影响的后验分布。实线表示GAM模型中相关系数的后验均值，粗线表示66%的可信带，细线表示95%的可信带。

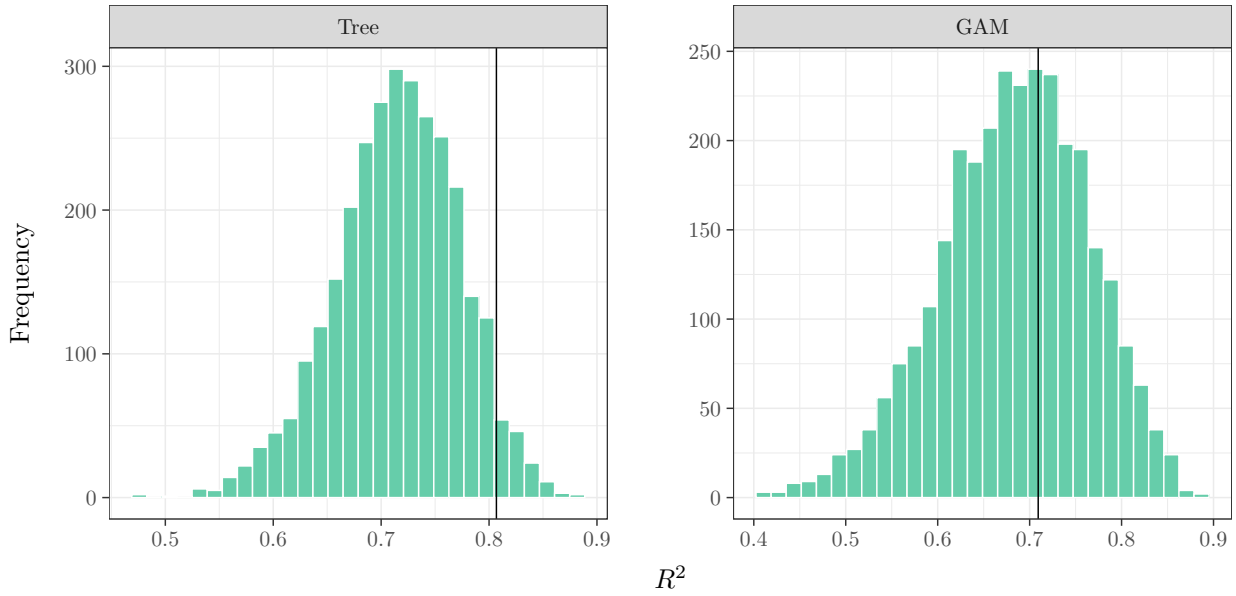

图7：回归树和广义加性模型对 $\delta(x)$ 的摘要 R^2 的后验分布。黑线表示 $\delta(x)$ 后验均值的摘要 R^2 。

自助法。通过比较这两个模型的预测性能，我们可以评估 BCMF模型增加的复杂性是否合理。

为了了解 BCMF 和 LSEM 所作预测之间的显著差异，我们在图 $\delta(X_i)$ 和 $\zeta(X_i$中比较每个个体的估计值，并同时考察总体情况及按种族分层的情况。图8展示了每个个体的 $\delta($ $X_i)$ 和 $\zeta(X_i$估计值，包括总体估计和按种族分层的估计。BCMF 的效应估计异质性显著低于 LSEM；对于 $\delta(X_i)$ 和 $\zeta(X_i)$，LSEM 均估计出大量正效应和负效应。此外，我们观察到不同种族之间的异质性显著较低；例如，LSEM 对太平洋岛民群体中吸烟的直接效应和间接效应都作出了违背直觉的预测。MEPS数据集规模庞大，但其中太平洋岛民人数相对较少；在我们分析的子集中，吸烟者仅有13人。通过应用正则化，BCMF 将这一亚群体中的直接效应和间接效应收缩至更接近其他种族的水平。

接下来，我们将 BCMF 和 LSEM 模型拟合到同一个 $n = 8056$ indi-训练集上

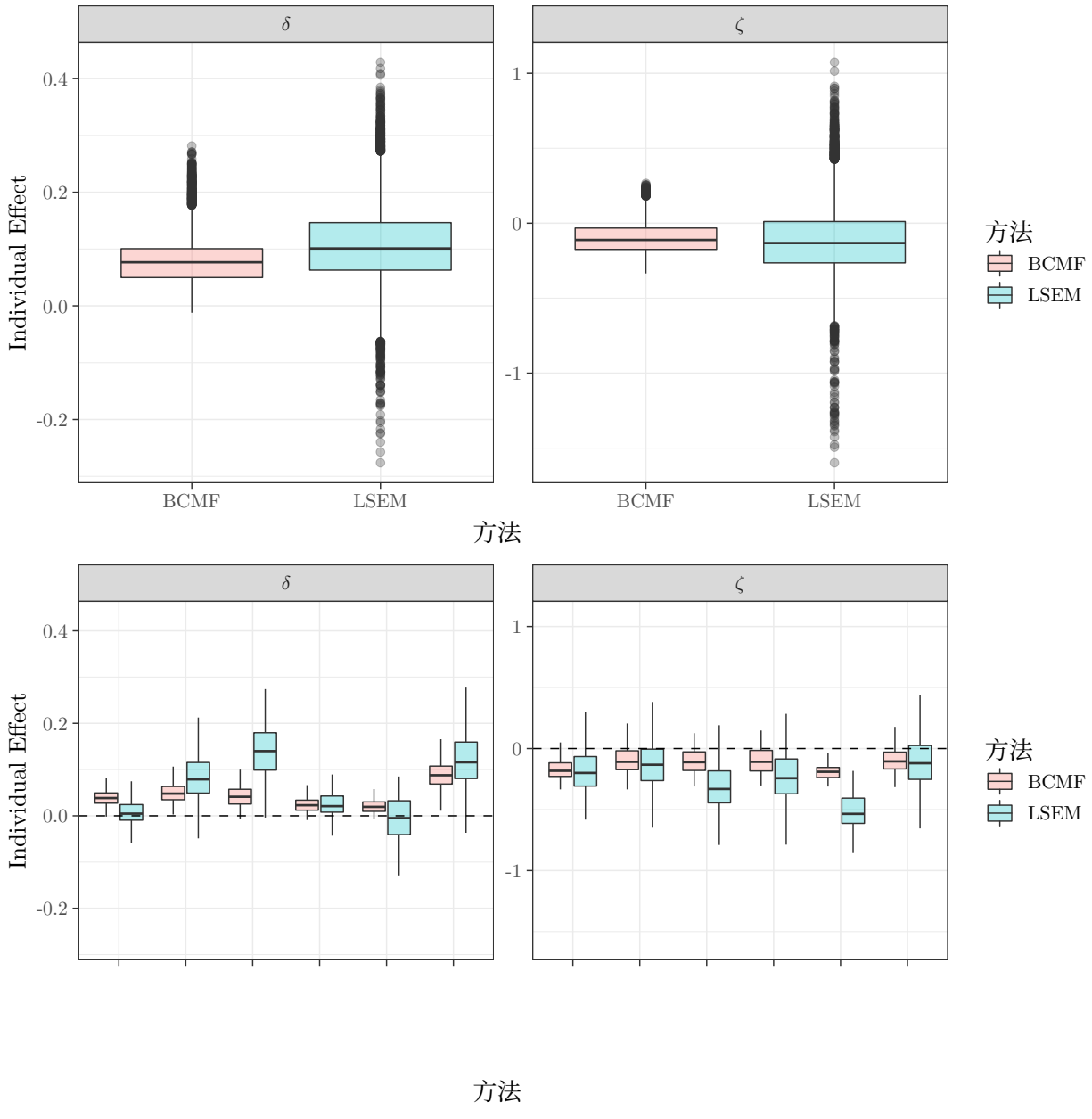

图8：上：箱线图展示了对原始MEPS数据集拟合的BCMF和LSEM模型所估计的个体直接效应和间接效应的分布。下：箱线图展示了BCMF和LSEM估计的个体效应分布，并按种族分层。

变量	R_{test} (BART)	R_{test} (LSEM)	P值
`phealth`	0.445	0.419	0.0001
`log(Y)`	0.454	0.431	0.0005

表1：BART 和 LSEM 拟合在所有个体中的中介变量（`phealth`）和结果（`logY`）的留出相关性，以及用于比较这两个模型在留出数据上的预测性能的配对 Wilcoxon 匹配对符号秩检验的 p值。

Term	估计	标准误差	统计量	P值
$\widehat{M}_{\text{lsem}}$	0.1561	0.0803	1.9447	0.0518
$\widehat{M}_{\text{bcmf}}$	0.8477	0.0809	10.4835	< 0.0001
$\widehat{Y}_{\text{lsem}}$	0.1509	0.0725	2.0824	0.0373
$\widehat{Y}_{\text{bcmf}}$	0.9030	0.0747	12.0840	< 0.0001

表2：汇总 MEPS 测试数据上线性模型与 BART 拟合结果的线性模型的系数估计。

$\widehat{\ }$ $\widehat{\ }$ $\widehat{\ }$ $\widehat{\ }$个体并计算预测值（$Y_{i,\ \text{lsem}}$，$M_{i,\ \text{lsem}}$，$M_{i,\ \text{bcmf}}$，$Y_{i,\ \text{bcmf}}$），使用拟合模型对测试集中的 $n = 8057$个体进行预测。我们使用这些测试集上的预测值，从三个方面评估模型的性能。首先，考察（M_i，Y_i）与其在测试集上的预测值之间的相关性。其次，我们进行配对威尔科克森符号秩检验，比较平方差（$Y_i - \widehat{Y}_{i,\ \text{lsem}}$）^{2} 与（$Y_i - \widehat{Y}_{i,\ \text{bcmf}}$）^{2}（$M_i$也采用相同方法）。结果见表1。我们发现，BCMF模型的相关性略高于 LSEM 模型，并且根据符号秩检验，两者的性能差异具有高度统计显著性。

我们的第三项比较考察了 堆叠（Wolpert，1992年）BCMF$\widehat{\ }$ $\widehat{\ }$和 LSEM 的预测结果，方法是拟合线性模型$Y_i = \beta_0 + \beta_1 Y_{i,\ \text{lsem}} + \beta_2 Y_{i,\ \text{bcmf}} + \epsilon_i$（对 M_i也采用类似方法）。堆叠过程的结果见表 2。从此次拟合结果可以看出，与线性模型相比，该线性模型更加依赖 BCMF 的预测结果；此外，BCMF 预测结果在统计上远比 LSEM 的预测结果显著（也就是说，有强有力的证据表明 BCMF 预测变量优于 LSEM 预测变量，而只有微弱证据支持

相反的结论）。有趣的是，LSEM 的预测结果被发现具有统计显著性，这表明，对混杂因素进行线性调整（即在函数中纳入线性项 $x^\top b$ （μ（x） ζ（x） d（x） μ_m（x） τ_m（x））的 BCMF 改进版本可能会提升模型的拟合效果。

4.3 模拟研究

我们现在开展一项模拟研究，以更好地了解BCMF模型的运行特性。本研究旨在回答以下问题：（i）在估计中介效应方面，BCMF模型的预测准确性是否更高？（ii） δ(X_i)和 ζ(X_i)的可信区间的覆盖率是否接近其名义水平？（iii）BCMF模型能否在CART汇总所识别出的数据亚组内准确估计效应?

数据生成机制

我们采用一种数据生成机制，其中混杂因素和处理分配直接从MEPS数据集中抽样，而中介变量和结果的真实值则通过分别将我们的模型和LSEM拟合到数据上获得。为评估这两种方法的性能，我们对每种模拟设定重复模拟200次，其中训练集包含8056个观测值，测试集包含8057个观测值。在所有模拟数据集中，我们使用相同的训练集/测试集划分，以评估每种方法生成的置信区间/可信区间的覆盖概率。

LSEM 和 BCMF 模型的一个关键区别在于，LSEM 不会对中介效应进行正则化。因此，LSEM 得出的中介效应真实值比实际情况预期的更加异质，尤其是在样本量较小的人群亚组中。例如，由于 MEPS 数据集中的太平洋岛民人数很少，LSEM 对种族作为效应修饰因子的作用所作的估计在这一群体中并不稳定。为此，我们还考虑了第三种真实值，它同样是

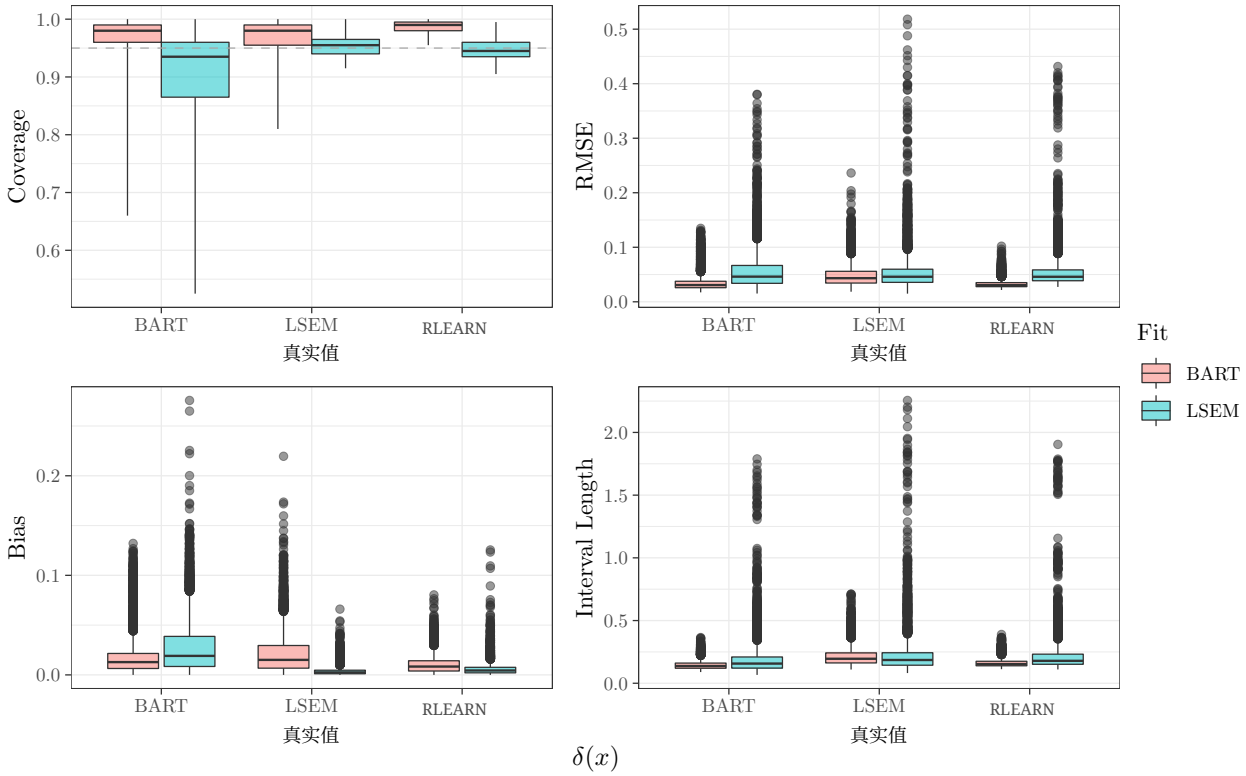

图9：在所有 BART/LSEM/RLEARN 真实值下，采用 BART/LSEM 模型的各种拟合组合时， $\delta(x)$ 的个体模拟结果。左上图显示所有个体的名义 95% 可信区间的覆盖概率，右上图显示均方根误差，左下图显示绝对偏差，右下图显示平均区间长度。

一个 LSEM，只是其中参数 (13) 改为采用 Nie 和 Wager (2021) 提出的 R 学习器方法进行估计，该方法使用套索回归来降低异质性。

结果：个体效应

我们对每个模拟数据集分别拟合模型和 LSEM，并测量效应的点估计、95% 可信区间的上下限及宽度，以及每次重复中区间是否包含真实参数。随后，我们利用 200 次重复计算均方根误差、绝对偏差、区间平均宽度和覆盖概率。

我们在图 9 和图 10中展示模拟研究的结果，在不同真实值组合下比较 BCMF 和 LSEM 模型

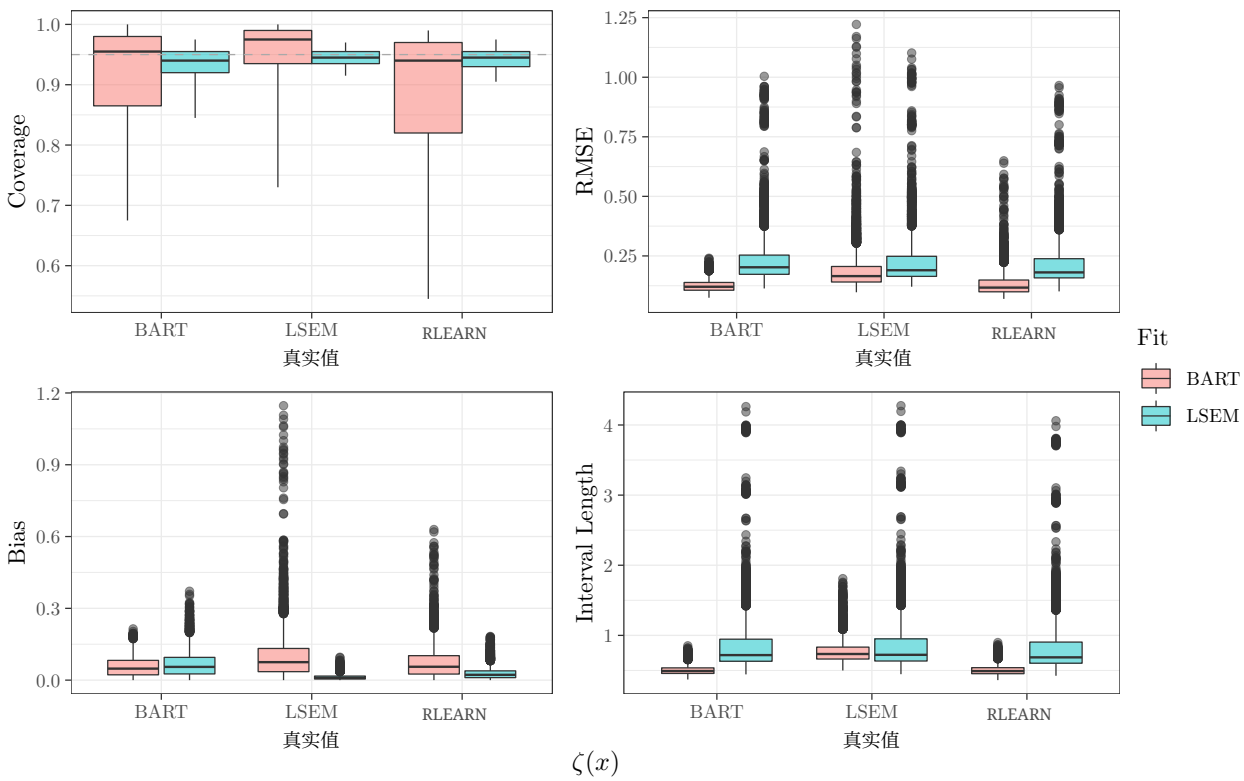

图10：在 BART/LSEM/RLEARN 真实值下，采用 BART/LSEM 模型拟合的所有组合中，$\zeta(x)$ 的个体模拟结果。左上图显示所有个体中名义 95% 可信区间的覆盖概率，右上图显示均方根误差，左下图显示绝对偏差，右下图显示平均区间长度。

设定	方法	覆盖率	RMSE	Bias	长度
BART	BART	0.91	0.12	0.06	0.50
BART	LSEM	0.93	0.23	0.07	0.83
LSEM	BART	0.94	0.19	0.10	0.77
LSEM	LSEM	0.94	0.22	0.01	0.84
RLEARN	BART	0.88	0.13	0.07	0.50
RLEARN	LSEM	0.94	0.21	0.03	0.80

表 3：测试集中所有个体的 $\zeta(X_i)$覆盖概率、均方根误差和绝对偏差。

设定	方法	覆盖率	RMSE	偏差区间长度	
BART	BART	0.95	0.03	0.02	0.14
BART	LSEM	0.89	0.06	0.03	0.19
LSEM	BART	0.96	0.05	0.02	0.21
LSEM	LSEM	0.95	0.05	0.00	0.22
RLEARN	BART	0.99	0.03	0.01	0.16
RLEARN	LSEM	0.95	0.05	0.01	0.21

表 4：测试集中所有个体的 $\delta(X_i)$ 的覆盖概率、均方根误差和绝对偏差。

拟合模型。表 3 和表4 分别汇总了图 9和图10中测试集所有个体的结果。当将 BCMF 拟合到 BCMF真实情况时，与 LSEM 相比，它在区间估计方面更接近名义覆盖率，同时区间长度、均方根误差和绝对偏差显著更低。有趣的是，我们还观察到，当使用 LSEM 生成数据时，BCMF模型在均方根误差方面也具有竞争力。我们推测，这是因为根据原始数据拟合的 LSEM 所估计的数据生成机制虽然仍具有相当强的异质性，但其同质性足够高（且效应足够小），因此 BCMF 的正则化优势超过了 LSEM 被正确设定这一因素。我们观察到类似的现象；当使用 R 学习器生成真实值时，这种现象甚至更加明显。

LSEM和BCMF模型的覆盖率结果揭示了一些有趣的现象。令人意外的是，尽管LSEM产生的区间比BCMF大得多，但其覆盖率似乎具有稳健性。然而，对于个体层面的直接效应，BCMF在部分个体上的名义覆盖率表现不佳。我们在补充材料中进一步考察了这一现象，发现对于异质性效应很强的个体，BCMF未能达到名义覆盖率；也就是说，这些个体的条件平均中介效应（$\delta(X_i)$ $\zeta(X_i)$）与平均¯ ¯效应（δ， ζ）相差很大。这一现象在预期之中，因为BCMF模型的设计明确倾向于收缩至异质性程度较小的结构。尽管这降低了检测强异质性效应的能力，

组	间接效应	直接效应
age ≥ 67	0.99	0.92
non-white, $34 \leq$ age < 67	0.96	0.88
non-white, age < 34	0.86	0.94
white, $34 \leq$ age < 67	0.88	0.92
white, age < 34	0.96	0.96
平均	0.93	0.93
动态	0.95	0.95

表5：使用图 ζ_a（A_i）和 δ_a（A_i）中的亚组所得的亚组覆盖概率3。

但它不会增大检测异质性的第一类错误率，因此BCMF模型在检测异质性时较为保守。

结果：平均效应和亚组平均效应

BCMF还能对亚人群内的平均中介效应给出可靠的估计。这里，我们同时考虑在BCMF真实情况下获得的固定子组和数据依赖子组。固定子组是由图3中的终端节点识别出的组：年龄 $\geq$ 67、非白人，以及 $34 \leq$ 年龄 $<$ 67、非白人和年龄 $<$ 34、白人，以及$34 \leq$ 年龄 $<$ 67、白人和年龄 $<$ 34。数据确定的亚组通过后验投影汇总确定，即对每个模拟数据集拟合一棵树，并识别终端节点组。各模拟情景下平均效应推断的比较见补充材料。

表5 展示了固定子组和数据依赖子组（标记为“动态”）的模拟结果。可以看到，BCMF 得出的区间覆盖率接近名义水平，但在非白人群体中的结果略差。值得注意的是，数据依赖子组的可信区间覆盖率更高，实际上，直接效应和间接效应¯ ¯的覆盖率均恰好达到 95%。平均效应 δ 和 ζ 的区间覆盖率也接近名义水平。

5 讨论

本文介绍了一种贝叶斯因果中介森林（BCMF）模型，该模型利用变系数模型，分别识别并正则化条件平均自然直接效应和间接效应。我们的方法与 LSEM 相似，因此可以轻松地将这些效应识别为变系数的乘积。此外，我们证明，与可比的 LSEM 相比，我们的模型在真实和模拟的 MEPS 数据上均能取得更低的预测误差。我们还指出，由于假设中介效应较小且大多同质，该模型在估计异质性时较为保守。我们还提供了后验汇总方法，用于解读模型拟合和亚组检测结果。

为了改进我们的方法和分析，可以从几个方向着手。首先，我们可以改进结果和中介变量的模型。例如，医疗支出的对数表现出异方差性，其中 Y_i 和 X_i 之间存在复杂的关系，正如 Linero 等人 (2020年所示。此外，由于本问题中的中介变量是有序的，并且从经验上看可以用四舍五入后的正态分布很好地近似，因此我们可以对 M_i 使用累积概率单位模型，而不是正态模型。对于 M_i 使用连续模型而非有序模型的影响尚不明确，值得进一步研究。

将没有医疗支出的个体排除在分析之外（如我们在本文中所做的）存在问题，因为发生医疗支出的可能性很可能与吸烟状态有关。作为进一步改进分析的措施，更好的做法是采用主分层（Frangakis 和 Rubin，2002）。这种方法可以估计在会发生医疗支出的个体亚组中，吸烟对医疗支出的因果效应，而不论其吸烟状态如何。在此类分析中，我们假设，如果某些个体不吸烟时会发生医疗支出，那么他们吸烟时也会发生医疗支出。这将使我们能够更为审慎地评估吸烟对医疗支出的因果效应。

最后，尽管 我们的模型在均方根误差方面表现良好，但对于某些 e

个体而言，其可信区间未能完全达到名义覆盖率。在补充材料中，我们展示了，对于其 $\bar{\delta}$ 条件中介效应与平均效应存在显著差异的个体，我们的 ζBCMF www.github.com/vcbcmf/vcbcmf会出现覆盖不足。究竟这是一个可以修正的问题，还是使用向同质效应收缩的模型所带来的必然结果，仍有待进一步研究。可在 www.github.com/vcbcmf/vcbcmf获取用于复现我们分析和模拟结果的代码。

参考文献

Albert, J. M.（2008年）。基于潜在结果模型的中介分析。《医学统计学》，27(8):1282–1304。

Athey, S.（2017）。超越预测：利用大数据解决政策问题。科学，355(6324):483–485。

Baron, R. M. 和 Kenny, D. A.（1986年）。社会心理学研究中调节变量与中介变量的区分：概念、策略与统计考量。《人格与社会心理学杂志》，51(6):1173。

Breiman, L., Friedman, J. H., Olshen, R. A. 和 Stone, C. J.（1984）。分类与回归树。Chapman and Hall，纽约，第1版。

Chipman, H. A., George, E. I. 和 McCulloch, R. E.（2010年）。BART：贝叶斯加性回归树。《应用统计年鉴》，4(1):266–298。

Deshpande、S. K.、Bai、R.、Balocchi、C.、Starling、J. E. 和 Weiss、J.（2020年）。VCBART：用于变系数的贝叶斯树。

Dorie, V., Hill, J., Shalit, U., Scott, M. 和 Cervone, D.（2019年）。自动化与亲自操作的因果推断方法：从数据分析竞赛中汲取的经验。《统计科学》，34(1)：43–68。

Farbmacher, H., Huber, M., Laff´ers, L., Langen, H. 和 Spindler, M.（2022年）。运用双重机器学习进行因果中介分析。《计量经济学期刊》，25(2)：277–300。

Frangakis, C. E. 和 Rubin, D. B.（2002）。因果推断中的主分层。生物统计学，58(1)：21–29。

Hahn, P. R.、Murray, J. S. 和 Carvalho, C. M.（2020）。用于因果推断的贝叶斯回归树模型：正则化、混杂与异质性效应（附讨论）。《贝叶斯分析》，15(3)：965–1056。

Hill, J.、Linero, A. 和 Murray, J.（2020）。贝叶斯加性回归树：综述与展望。《统计学及其应用年度评论》，7(1)：251–278。

Hill, J. L.（2011）。用于因果推断的贝叶斯非参数建模。《计算与图形统计学期刊》，20(1)：217–240。

Imai, K.、Keele, L. 和 Tingley, D.（2010）。中介分析的一般方法。《心理学方法》，15(4)：309。

Kershaw, K. N., Mezuk, B., Abdou, C. M., Rafferty, J. A., 和 Jackson, J. S.（2010年）。社会经济地位、健康行为与C反应蛋白：一项调节中介分析。《健康心理学》，29(3):307。

Kim, C., Daniels, M. J., Marcus, B. H., 和 Roy, J. A.（2017）。用于推断中介因果效应的贝叶斯非参数推断框架。《生物统计学》，73(2):401–409。

K¨unzel, S. R., Sekhon, J. S., Bickel, P. J., 和 Yu, B.（2019年）。利用机器学习估计异质性处理效应的元学习器。《美国国家科学院院刊》，116(10):4156–4165。

Linero, A. R.（2021）。在非参数和高维模型中，贝叶斯可忽略性是一种信息先验。

Linero, A.R.、Sinha, D. 和 Lipsitz, S.R.（2020）。使用共享贝叶斯森林的半参数混合尺度模型。《生物统计学》，76(1)：131–144。

Linero, A. R. 和 Zhang, Q.（2022年）。使用贝叶斯树集成进行中介分析。《心理学方法》。

MacKinnon, D. P.（2008年）。统计中介分析导论。

MacKinnon, D. P. 和 Dwyer, J. H.（1993年）。预防研究中的中介效应估计。《评估评论》，17(2):144–158。

Muller, D.、Judd, C. M. 和 Yzerbyt, V. Y.（2005年）。当调节作用被中介时，以及中介作用受到调节时。《人格与社会心理学杂志》，89(6):852。

Nie, X. 和 Wager, S.（2021）。异质性处理效应的准神谕估计。《生物统计学》，108(2):299–319。

Obermeyer, Z.和Emanuel, E.J.（2016年）。预测未来——大数据、机器学习与临床医学。《新英格兰医学杂志》，375(13):1216。

Pearl, J.（2001）。直接效应和间接效应。载于《第十七届人工智能不确定性会议论文集》，第411–420页。摩根考夫曼出版社。

Preacher, K. J.、Rucker, D. D. 和 Hayes, A. F.（2007）。探讨调节中介假设：理论、方法与建议。《多变量行为研究》，42(1)：185–227。

Robins, J. M. 和 Greenland, S.（1992）。直接效应和间接效应的可识别性与可交换性。《流行病学》，第143–155页。

Rubin, D. B.（1974年）。估计随机和非随机研究中处理的因果效应。《教育心理学杂志》，66(5)：688–701。

Rubin, D. B.（1981年）。贝叶斯自助法。《统计年鉴》，第130–134页。

Rubin, D.B.（2004年）。基于潜在结果的直接和间接因果效应。《斯堪的纳维亚统计学杂志》，31(2):161–170。

Tchetgen, E. J. T. 和 Shpitser, I.（2012）。因果中介分析的半参数理论：效率界、多重稳健性和敏感性分析。《统计年鉴》，40(3):1816。

VanderWeele, T. J.（2016）。中介分析：从业者指南。《公共卫生年度评论》，37:17–32。

Wendling, T.、Jung, K.、Callahan, A.、Schuler, A.、Shah, N. H. 和 Gallego, B.（2018）。比较利用医疗保健数据库观察性数据估计异质性治疗效应的方法。《医学统计学》，37(23):3309–3324。

Wolpert, D. H.（1992年）。堆叠泛化。《神经网络》，5(2):241–259。

Wood, S. N.（2006）。《广义加性模型：*R*语言入门》。Chapman and Hall/CRC，纽约，第一版。

Woody, S., Carvalho, C. M., and Murray, J. S. (2021). 通过低维后验总结进行模型解释。《计算与图形统计学期刊》，30(1):144–161。

Yeager, D.S., Hanselman, P., Walton, G.M., Murray, J.S., Crosnoe, R., Muller, C., Tipton, E., Schneider, B., Hulleman, C. S., Hinojosa, C. P., Paunesku, D., Romero, C., Flint, K., Roberts, A., Trott, J., Iachan, R., Buontempo, J., Yang, S. M., Carvalho, C. M., Hahn, P. R., Gopalan, M., Mhatre, P., Ferguson, R., Duckworth, A. L., and Dweck, C. S. (2019年). 一项全国性实验揭示了成长型思维模式在哪些方面能够提升成就。《自然》，573(7774):364–369。

Zheng, W. 和 van der Laan, M. J.（2012）。自然直接效应的目标最大似然估计。《国际生物统计学杂志》，8(1):1–40。

《使用贝叶斯决策树集成估计异质性因果中介效应》的补充材料

AngelaTing 和 Antonio R. Linero统
计与数据科学系 德克萨斯大学奥斯汀分校

arXiv:2303.01620v1 [stat.ME] 2 Mar 2023

S.1 贝叶斯回代算法

现在，我们介绍一种用于拟合第 ?? 节 BCMF 模型的贝叶斯回代算法。根据第 ?? 节，BART 会在 (T_1, M_1) ..., (T_m, M_m) 上诱导一个后验分布。我们使用贝叶斯回代算法（Chipman 等人，2010）更新每个 (T_j, M_j)，该算法将排除 j^{th} 树后的部分残差作为响应变量。令 $T_{(j)}$ 表示除树 T_j 以外的所有树的集合。类似地，令 $M_{(j)}$ 表示除 M_j 以外的所有终端节点的集合。那么，对于树加和模型 $Y_i = \sum_{j=1}^{m} g(X_i; T_j, M_j) + \epsilon_i$，排除第 j 棵树后的部分残差向量为

$$R_{ij} \equiv Y_i - \sum_{k \neq j} g(X_i; T_k, M_k).$$

R_{ij} 可用于使用 T_j、M_j 所描述的单树 Metropolis-Hastings算法更新 (Chipman 等人（1998）提出的算法，其中 R_{ij} 现在取代了数据 Y_i，因为$R_{ij} = g（X_i；T_j，M_j）+ \epsilon$ 对应于单树模型。

在我们的模型中，μ、ζ、d、μ_m和 τ_m 均采用 BART 先验。中介模型参数μ_m 和 τ_m 可以使用Hahn 等中给出的贝叶斯回代算法进行更新。

（2020年）。为处理结局模型参数，令

$$\mu(x) = \sum_j g(x; T_j^{\mu}, M_j^{\mu})$$
$$\zeta(x) = \sum_j g(x; T_j^{\zeta}, M_j^{\zeta})$$
$$d(x) = \sum_j g(x; T_j^{d}, M_j^{d}).$$

然后，我们将 μ 的部分残差定义为

$$\begin{aligned} R_i^{\mu} &= Y_i - A_i\,\zeta(X_i) - M_i\,d(X_i) - \sum_{k \neq j} g(X_i; T_k^{\mu}, M_k^{\mu}) \\ &= g(X_i; T_j^{\mu}, M_j^{\mu}) + \epsilon_i \end{aligned}$$

其中 $\epsilon_i \sim N(0,\ \sigma^2)$。因此，与 Chipman 等人 (1998) 一样，$(T_j^{\mu_Y},\ M_j^{\mu_Y})$ 可以使用上述残差作为新的响应变量，通过 BART 进行更新。同样，$(T_j^{\zeta},\ M_j^{\zeta})$ 可以使用以下部分残差，通过 BART 进行更新

$$\begin{aligned} R_i^{\zeta} &= Y_i - \mu(X_i) - M_i\,d(X_i) - A_i \sum_{k \neq j} g(X_i; T_k^{\zeta}, M_k^{\zeta}) \\ &= A_i\,g(X_i; T_j^{\zeta}, M_j^{\zeta}) + \epsilon_i. \end{aligned}$$

由于 A_i是二元变量，这意味着可以使用 T_j^{ζ}, M_j^{ζ}，并应用 Chipman 等人 (1998) 提出的贝叶斯 CART 算法，对观测值 R_i^{ζ} 进行更新，使得$A_i = 1$。最后，令 d 的部分残差为

$$\begin{aligned} R_i^{d} &= \frac{Y_i - \mu(X_i) - A_i\zeta(X_i)}{M_i} - \sum_{k \neq j} g(X_i; T_k^{d}, M_k^{d}) \\ &= g(X_i; T_j^{d}, M_j^{d}) + \eta_i \end{aligned}$$

其中 $\eta_i = \frac{\epsilon_i}{M_i} \sim N\left(0, \frac{\sigma^2}{M_i^2}\right)$。因此，可以使用已知权重的异方差贝叶斯 CART 更新方法，对 T_j^d, M_j^d进行更新；例如，Pratola 等人 (2020年) 给出了该模型的更新算法。

S.2 其他模拟结果

本节提供了第 ?? 节模拟研究的更多结果。在表 S.1 和表 S.2 中，我们分别给出了在真实值（BART、LSEM 和 R 学习器）与估计方法（BART 或 LSEM）的所有组合下，参数 δ 和 ζ 的覆盖率。从表S.1 可以看出，BCMF 和 LSEM 模型的表现基本相同；BART 的覆盖率略高，均方根误差、绝对偏差和区间长度则相当。从表S.2 可以看出，BCMF 对直接效应的估计偏差略大，导致整体覆盖率较低。

在图 S.1 和图 S.2 中，我们验证了手稿中的论述：对于条件平均效应与平均效应差异较大的个体，BCMF 的覆盖率往往较差。由于 LSEM 不会向同质性收缩，我们发现，直接效应和间接效应的覆盖率随 $|\zeta(X_i) - \zeta|$ 和 $|\delta(X_i) - \delta|$的变化相对稳定，而随着这些量增大，BCMF 的覆盖率往往会变差。

参考文献

Chipman, H. A., George, E. I. 和 McCulloch, R. E. (1998). 贝叶斯分类与回归树模型搜索。美国统计学会会刊，93(443):935–948。

Chipman, H. A., George, E. I. 和 McCulloch, R. E. (2010). BART：贝叶斯加性回归树。《应用统计年鉴》，4(1):266–298。

Hahn, P. R., Murray, J. S. 和 Carvalho, C. M. (2020). 贝叶斯回归树模型 for

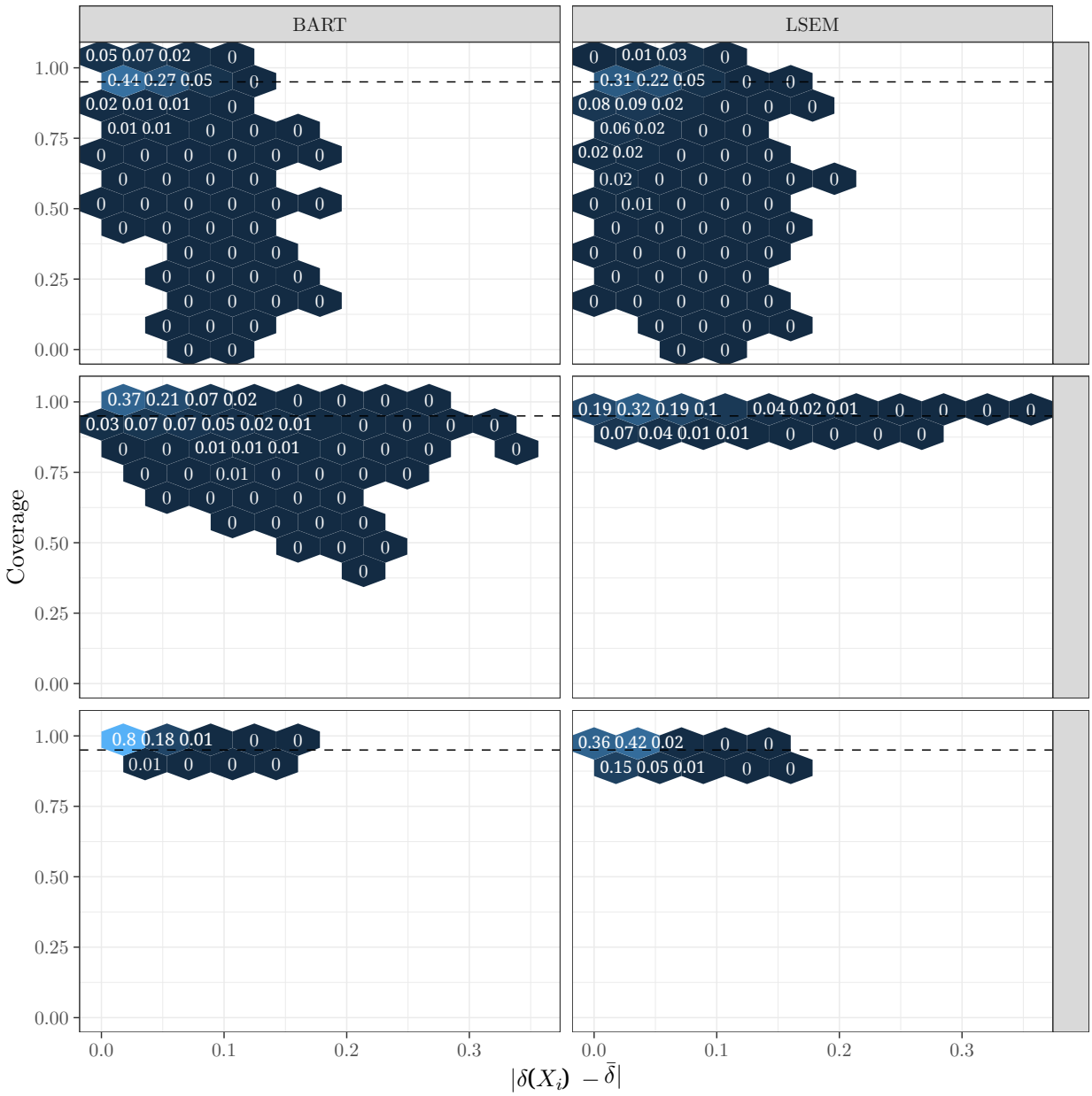

图 S.1：BCMF（左）和 LSEM（右）模型对 $\delta(X_i)$的覆盖率随 $|\delta(X_i) - \bar{\delta}|$的变化，以二维六边形直方图显示。每个分箱中个体所占的比例由分箱标签给出。

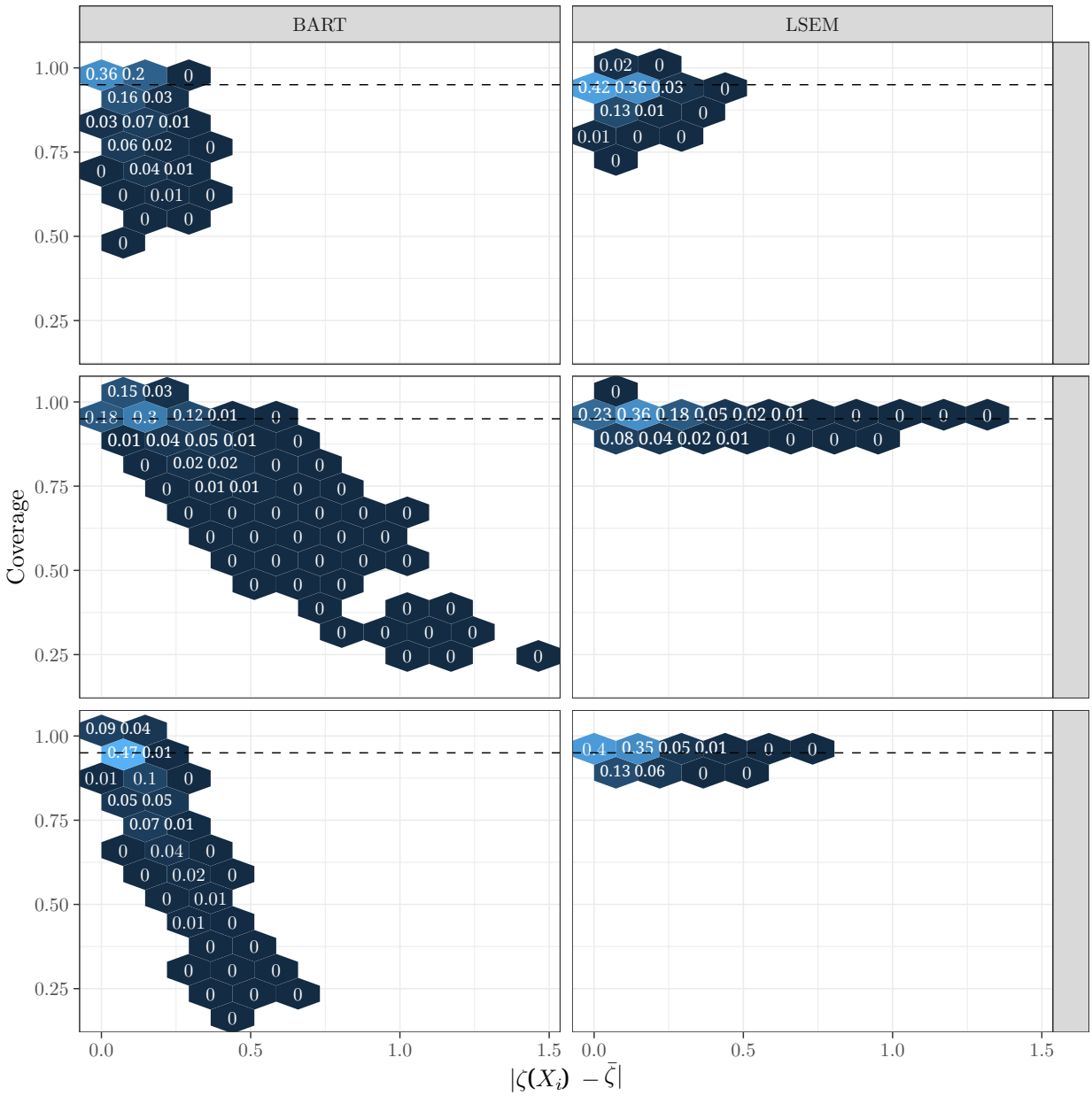

图 S.2：BCMF（左）和 LSEM（右）模型对 $\zeta(X_i)$的覆盖率，作为 $|\zeta(X_i) - \bar{\zeta}|$的函数，以二维六边形直方图显示。落入每个分箱的个体比例由分箱标签给出。

因果推断：正则化、混杂和异质性效应（附讨论）。《贝叶斯分析》，15(3)：965–1056。

Pratola, M. T.、Chipman, H. A.、George, E. I. 和 McCulloch, R. E.（2020年）。通过乘法回归树实现异方差 BART。《计算与图形统计学期刊》，29(2)：405–417。

设定	方法	覆盖率	RMSE	Bias	长度
BART	BART	0.96	0.01	0.00	0.05
BART	LSEM	0.96	0.01	0.01	0.05
LSEM	BART	0.98	0.01	0.00	0.06
LSEM	LSEM	0.96	0.01	0.00	0.06
RLEARN	BART	0.98	0.01	0.00	0.06
RLEARN	LSEM	0.96	0.01	0.00	0.05

表 S.1：δ 的覆盖概率、均方根误差和绝对偏差

设定	方法	覆盖率	RMSE	Bias	长度
BART	BART	0.90	0.06	0.02	0.21
BART	LSEM	0.96	0.06	0.01	0.21
LSEM	BART	0.97	0.06	0.02	0.24
LSEM	LSEM	0.96	0.05	0.00	0.22
RLEARN	BART	0.90	0.05	0.02	0.20
RLEARN	LSEM	0.94	0.05	0.01	0.20

表 S.2：ζ 的覆盖概率、均方根误差和偏差