arXiv:2603.03819v1 [stat.ME] 4 Mar 2026

用于回归不连续设计中条件平均处理效应的直接贝叶斯加性回归树*

Daisuke Kondo[1] 和 Shonosuke Sugasawa[1]

[1]庆应义塾大学经济学部

摘要

回归不连续设计（断点回归设计）广泛用于因果推断。在许多实证应用中，处理效应会随协变量而显著变化；忽略这种异质性可能导致误导性结论，因此有必要在断点回归设计中灵活地建模异质性处理效应。为此，我们提出一种基于贝叶斯加性回归树（贝叶斯加性回归树）的贝叶斯非参数方法，用于估计异质性处理效应。我们方法的关键特点是采用广义贝叶斯框架，通过定义在损失函数上的伪模型，拟合截断点附近的局部线性模型，从而利用贝叶斯加性回归树直接建模异质性处理效应。局部模型带宽参数的最优选择通过 Hyv¨arinen 得分实现。数值实验表明，所提出的方法能够灵活捕捉异质性处理效应作为协变量函数时的复杂结构。

关键词：广义贝叶斯；贝叶斯加性回归树；局部估计；不确定性量化

*2026年3月5日。通讯作者：Shonosuke Sugasawa（电子邮件：sugasawa@econ.keio.ac.jp）

1 引言

回归不连续设计（断点回归设计）是一种具有代表性的因果推断方法，适用于干预由连续（运行）变量决定的情形，已广泛应用于经济学、营销学、流行病学和临床研究等多个领域。回归不连续设计由 Thistlethwaite 和 Campbell（1960年）提出，Hahn 等人（2001年）对识别条件进行了形式化，并提出了局部线性估计量。后续的方法学发展主要集中于改进截断点处平均处理效应的估计与推断，包括对相关文献的全面综述（Imbens 和 Lemieux，2008年；Lee 和 Lemieux，2010年）、最优带宽选择（Imbens 和 Kalyanaraman，2012年）、稳健置信区间（Calonico 等人，2014年）以及贝叶斯非参数方法（Chib 等人，2023年）。尽管取得了这些方法学进展，大多数回归不连续设计文献仍集中于估计截断点处处理组的平均处理效应（ATE）。然而，在许多实证应用中，处理效应很可能因具有不同处理前协变量的个体而存在显著差异（例如 Becker 等人，2013年；Card 等人，2008年）。忽视异质性可能掩盖处理效应的重要分布特征，并导致具有误导性的结论。

为了在断点回归设计中估计条件平均处理效应（CATE），文献中已提出多种方法。Becker 等人（2013）和 Calonico 等人（2025）采用局部多项式估计，并在参数化设定下估计 CATE。Sugasawa et al. (2023) 和 Tao 等人（2025）提出了分组平均处理效应估计，Sawada 等人（2024）则研究了运行变量不同区域之间的处理效应异质性。然而，这些方法无法将 CATE 表示为协变量的光滑函数。Reguly（2021）提出了一种基于树的方法，利用协变量对数据进行递归划分，并在每个叶节点内进行标准的断点回归设计估计。尽管这种方法为 CATE 提供了一种非参数方法，但单棵树可能缺乏足够的灵活性，尤其是在协变量较多的情况下；此外，对 CATE 估计量进行连贯的不确定性量化也可能颇具挑战。作为与之最为密切的

相关工作，Alcantara 等人（2025）提出使用贝叶斯加性回归树（BART），在考虑阈值的同时，根据协变量和运行变量估计结果变量的条件期望。尽管该方法能够灵活地估计 CATE，但它对整个条件结果响应面进行建模，并通过对比间接得到处理效应。因此，与处理效应无关的干扰成分也必须通过 BART 进行建模，而响应模型可能无法充分控制对 CATE 本身的正则化。在贝叶斯因果推断文献中，人们已经认识到，如果不直接对异质性处理效应进行正则化，而是对整个响应面建模，可能会导致有偏或不稳定的估计；针对 CATE 设计独立的先验结构则可以改善性能（例如 Hahn 等人，2020）。

本文提出一种替代性贝叶斯方法，利用贝叶斯加性回归树对断点回归设计中的条件平均处理效应进行估计与推断。本方法的关键特征在于采用广义贝叶斯框架，通过由损失函数定义的伪模型来拟合截断点附近的局部线性模型。该设定使得能够通过贝叶斯加性回归树直接对条件平均处理效应建模，我们称之为“Direct-BART”，并充分利用贝叶斯加性回归树的灵活性来刻画未知且可能复杂的异质性形式。Direct-BART的后验计算可通过简单的吉布斯抽样完成，其中贝叶斯加性回归树的后向拟合算法可直接纳入，且局部线性回归系数的抽样步骤十分直接。尽管所提方法包含一个决定阈值附近局部权重的调节参数，我们提出一种基于Hyv¨arinen得分的数据驱动带宽选择方法（Hyv¨arinen, 2005; Yonekura and Sugasawa, 2023），该方法可利用后验样本高效评估。我们通过模拟实验展示Direct-BART，并表明其相对于现有方法具有更优表现。

本文其余部分安排如下。第2节介绍所提出的方法，包括后验计算算法，以及基于 Hyv¨arinen 评分的数据驱动带宽选择。第3节介绍模拟研究，以评估所提出方法的

性能，并将其与几种方法进行比较。总结见第4节。

2 用于在断点回归设计中直接推断条件平均处理效应的贝叶斯加性回归树

2.1 用于直接推断条件平均处理效应的伪模型

我们考虑精确断点回归设计的标准设定。令 Y_i 表示结果变量，X_i 表示单位 $i = 1,\ldots,n$ 的运行变量。处理指示变量定义为 $W_i = I(X_i \geq c)$，其中 c 是已知的断点值。此外，我们还观测到一个 d 维协变量向量 Z_i，用于建模处理效应异质性。给定 $Z_i = z$ 时，截断点处的条件平均处理效应定义为

$$\tau(z) = \lim_{x \downarrow c} \mathbb{E}[Y_i \mid X_i = x, Z_i = z] - \lim_{x \uparrow c} \mathbb{E}[Y_i \mid X_i = x, Z_i = z].$$

在假设处理效应同质性的情况下，断点回归设计的一种标准方法是将局部线性（多项式）回归拟合为运行变量的函数。局部回归通常分别拟合处理区域与控制区域，但也可以通过最小化以下损失函数，同时对两个区域拟合加权局部线性回归：

$$\frac{1}{2} \sum_{i=1}^{n} K\left(\frac{|X_i - c|}{h}\right) \left(Y_i - \tau W_i - X_i(c)^\top \beta\right)^2,$$

其中，$X_i(c) = \left(1, (X_i - c)_-, (X_i - c)_+, \ldots, (X_i - c)_-^q, (X_i - c)_+^q\right)^\top$ 是多项式回归的向量，$K(\cdot)$ 是核函数，h 是带宽参数。Sugasawa et al. (2023) 采用上述损失函数来估计分组平均处理效应。

由于我们允许存在异质性处理效应 $\tau(Z_i)$，局部回归框架也必须纳入协变量依赖系数。具体而言，在存在异质性的情况下，局部回归系数 β 自然会成为 Z_i 的函数，我们将其记为 $\beta(Z_i)$。为了灵活捕捉这种依赖关系，同时保持分析上的

可处理性和计算上的可行性，我们假设变化系数可以用 Z_i的线性函数近似，即 $\beta(Z_i) = B\tilde{Z}_i$，其中 $\tilde{Z}_i = (1, Z_{i1}, \ldots\ldots, Z_{id})^\top$，且 $B \in \mathbb{R}^{(2q+1)\times(d+1)}$ 是一个系数矩阵。这一设定可视为对传统局部线性断点回归框架的自然扩展；在该框架中，假设 β 在各个单位间保持不变。在此，我们保留标准局部回归方法所依据的线性结构，同时允许回归系数随观测到的协变量系统性变化。因此，我们考虑以下损失函数：

$$L(\tau, B; h) = \frac{1}{2}\sum_{i=1}^{n} K\left(\frac{|X_i - c|}{h}\right)\left(Y_i - \tau(Z_i)W_i - X_i(c)^\top B\tilde{Z}_i\right)^2, \tag{1}$$

其中 $\tilde{Z}_i = (1, Z_{i1}, \ldots\ldots, Z_{id})^\top$。对于核函数，本文始终采用均匀核$K(x) = I(|x| \leq 1)$。

为在贝叶斯框架下实现损失函数 (1)，我们考虑一个关于 $Y = (Y_1, \ldots\ldots, Y_n)$ 的伪统计模型，定义为 $p_h(Y \mid \tau, B, \omega) \propto \exp\{-\omega L(\tau, B; h)\}$，其中 $\omega > 0$ 是通用缩放参数。使用伪模型定义后验分布的方法称为广义贝叶斯方法（例如 Bissiri 等人，2016）。该伪模型也可等价地表示为高斯模型的幂似然：

$$p_h(Y_i \mid \tau, B, \omega) \propto \phi(Y_i; \tau(Z_i)W_i + X_i(c)^\top B\tilde{Z}_i, \omega^{-1})^{k_i}, \tag{2}$$

对 $i = 1, \ldots\ldots, n$ 独立地，其中 $k_i = K(|X_i - c|/h)$。该模型使我们能够直接为 $\tau(Z)$ 指定 BART 先验，这将在下一节中讨论。

对于局部回归系数矩阵和缩放参数，我们采用先验 $B \sim \mathcal{MN}_{2q+1,d+1}(M_0, V_0, U_0)$ 和 $\omega \sim \text{Ga}(\nu_0, \eta_0)$，其中 M_0、U_0、V_0、ν_0 和 η_0 为固定超参数。这里，$\mathcal{MN}_{2q+1,d+1}(M_0, V_0, U_0)$ 表示均值矩阵为 M_0 、行协方差矩阵和列协方差矩阵分别为 V_0 和 U_0 的矩阵正态分布，而 $\text{Ga}(\nu_0, \eta_0)$ 表示形状参数为 ν_0 、率参数为 η_0 的伽马分布。作为默认选择，我们设定 $M_0 = \mathbf{0}_{2q+1,d+1}, V_0 = 100\mathbf{I}_{2q+1}, U_0 = \mathbf{I}_{d+1}$, $\nu_0 = 1, \eta_0 = 1$，并将在数值研究中采用这一设定。

2.2 条件平均处理效应的贝叶斯加性回归树先验

贝叶斯加性回归树（BART）是 Chipman 等人（2010年）提出的一种针对回归函数的非参数先验，能够捕捉协变量的复杂非线性函数。我们在此直接为条件平均处理效应函数引入 BART 先验$\tau(Z)$。因此，条件平均处理效应函数建模为 $\tau(Z_i) = \sum_{j=1}^{m} g(Z_i \mid T_j, M_j)$，其中 m 为树的数量，T_j 表示第 j 棵回归树的结构，包括分裂规则；$M_j = (\mu_{j1}, \mu_{j2}, \ldots, \mu_{jb_j})$ 表示相应的叶节点参数集合。每棵树 T_j 通过其分裂规则将协变量空间划分为一组互不相交的区域 $\{G_{j1}, G_{j2}, \ldots, G_{jb_j}\}$，并为每个区域分配一个叶参数。如果协变量向量 z 落入协变量区域 G_{jk}，函数 $g(z \mid T_j, M_j)$ 就返回相应的叶参数 μ_{jk}，即 $g(z \mid T_j, M_j) = \mu_{jk}$，若 $z \in G_{jk}$ 对于 $k = 1$、2、$\ldots$、b_j 成立。通过汇总所有树，BART 将协变量空间划分为由树特定区域交集构成的精细互不相交区域，例如 $\bigcap_{j=1}^{m} G_{jk_j}$，从而能够灵活地估计条件平均处理效应函数。

BART 中的参数为 $(T_j, M_j)_{j=1}^{m}$，我们为树组件 (T_j, M_j) 指定独立先验。此外，假设每棵树中的叶节点参数先验上相互独立。因此，我们将先验指定为 $p(T_j, M_j) = p(T_j) \prod_{k=1}^{b_j} p(\mu_{jk} \mid T_j)$。树先验 $p(T_j)$由三个部分组成：（i）节点发生分裂的概率；（ii）选择分裂变量的概率；（iii）选定分裂变量后选择阈值的概率。对于部分（i），为鼓励生成较浅的树结构，深度为 $d (= 0, 1, 2, \ldots)$ 的节点发生分裂的先验概率设为 $\alpha(1 + d)^{-\beta}$，其中 $\alpha \in (0, 1)$ 和 $\beta \in [0, \infty)$，遵循标准的 BART 设定（例如，Chipman 等人，2010年；Linero, 2018），默认取值为 $\alpha = 0.95, \beta = 2$。对于部分（ii）和（iii），从各自的候选样本集中均匀抽取分裂变量和阈值。叶参数的先验服从正态分布，具体为 $p(\mu_{jk} \mid T_j) \sim N(\mu_\mu, \sigma_\mu^2)$，且相互独立。误差方差 $p(\sigma^2)$ 的先验服从逆伽马分布。与此不同，我们明确指定叶节点参数的先验均值和标准差，分别记为 μ_μ 和

σ_μ。为设定 μ_μ 和 σ_μ，我们采用与 Chipman 等人（2010年）类似的策略。

鉴于处理效应的定义，可以合理地假设处理效应位于区间（$\tau_{\min}, \tau_{\max}$）内，其中

$$\tau_{\max} = \max_{i:x_i \in (c,c+\delta)} Y_i - \min_{i:x_i \in (c-\delta,c)} Y_i, \quad \tau_{\min} = \min_{i:x_i \in (c,c+\delta)} Y_i - \max_{i:x_i \in (c-\delta,c)} Y_i,$$

对某个 $\delta > 0$而言。此处，$\tau_{\max}$ 表示处理效应的经验上界，定义为截断点的 δ邻域内处理组个体的最大结果变量与对照组个体的最小结果变量之差。类似地，$\tau_{\min}$表示处理效应的经验下界，定义为截断点的 δ邻域内处理组个体的最小结果变量与对照组个体的最大结果变量之差。

由于具有 m 棵树的 BART 先验意味着处理效应服从正态分布 $\mathcal{N}$（$m\mu_\mu, m\sigma_\mu^2$），我们将 μ_μ 和σ_μ 设定为使区间（$\tau_{\min}, \tau_{\max}$）具有较高的先验概率。这通过针对某个合适的常数 $k > 0$设定 $m\mu_\mu + k\sqrt{m}\sigma_\mu = \tau_{\max}$ 和 $m\mu_\mu - k\sqrt{m}\sigma_\mu = \tau_{\min}$ 来实现。据此，我们设定

$$\mu_\mu = \frac{\tau_{\max} + \tau_{\min}}{2m}, \quad \sigma_\mu = \frac{\tau_{\max} - \tau_{\min}}{2k\sqrt{m}}$$

在我们的实现中，我们设定 $\delta = 0.1\text{sd}(X)$ 和 $k = 2$，这对应于截断点附近的一个狭窄邻域，并将约 95% 的先验概率质量赋予经验处理效应区间。

我们这一表述的一个关键特征在于，BART 先验被直接置于条件平均处理效应函数 $\tau(Z)$ 之上，而非完整的条件结果回归曲面之上。这意味着由该先验所诱导的正则化直接作用于异质性处理效应，而不会通过与因果估计量无关的干扰成分间接传导。通过显式地对 $\tau(Z)$ 建模，我们对 (μ_μ, σ_μ) 至 $(\tau_{\min}, \tau_{\max})$ 的设定，使得能够依据处理效应的合理量级进行透明的先验 elicitation。

此外，这种定向正则化在高维情形中可能尤为重要：对整个响应面进行建模可能导致处理效应本身出现有偏收缩，这一点在条件平均处理效应估计的相关背景下已得到强调（例如，Hahn 等人，2020年）。

2.3 后验计算

所提出的 Direct-BART 的后验推断采用吉布斯抽样进行，并使用标准贝叶斯加性回归树框架（Chipman 等人，2010年）中的回代方案。˜为方便记号表示，本节中我们用 X_i 代替 $X_i(c)$。$\{B, \omega, (T_j,\ M_j)_{j=1}^m\}$ 的联合后验与以下表达式成正比

$$p(B)p(\omega)\prod_{j=1}^{m} p(T_j, M_j)\prod_{i=1}^{n}\phi\left(Y_i; W_i\sum_{j=1}^{m} g(Z_i \mid T_j, M_j) + \tilde{X}_i^\top B\tilde{Z}_i, \omega^{-1}\right)^{k_i},$$

其中，$p(B)$、$p(\omega)$ 和 $p(T_j,\ M_j)$ 均为先验分布，已在上一节中说明。

给定条件平均处理效应函数 $\tau(Z_i) \equiv \sum_{j=1}^m g(Z_i \mid T_j,\ M_j)$，即给定 $(T_j,\ M_j)_{j=1}^m$ 后，B 和 ω 的全条件分布形式已知，抽样过程也很直接。具体而言，B 的全条件分布为 $\mathcal{MN}(M_n,\ \Sigma_n)$，其中

$$M_n = \Sigma_n\Big[\omega\sum_{i=1}^{n} k_i\Big\{Y_i - W_i\sum_{j=1}^{m} g(Z_i \mid T_j, M_j)\Big\}(\tilde{Z}_i \otimes \tilde{X}_i) + \Sigma_0\,\mathrm{vec}(M_0)\Big]$$
$$\Sigma_n = \Big\{\omega\sum_{i=1}^{n} k_i(\tilde{Z}_i \otimes \tilde{X}_i)(\tilde{Z}_i \otimes \tilde{X}_i)^\top + \Sigma_0\Big\}^{-1},$$

ω 的全条件分布为 $\mathrm{Ga}(\nu_n,\ \eta_n)$，其中 $\nu_n = \nu_0 + \sum_{i=1}^n k_i/2$ 和

$$\eta_n = \eta_0 + \frac{1}{2}\sum_{i=1}^{n} k_i\Big\{Y_i - W_i\sum_{j=1}^{m} g(Z_i \mid T_j, M_j) - \tilde{X}_i^\top B\tilde{Z}_i\Big\}.$$

对于 BART 中的参数 $(T_j,\ M_j)_{j=1}^m$，它们以顺序方式更新，这一过程称为“贝叶斯回代”。具体而言，对于 $j = 1, \ldots, m$，我们使用 Metropolis–Hastings 算法从完全条件分布 $p(T_j,\ M_j \mid (T_h,\ M_h)_{h\neq j},\ B,\ \omega,\ y)$ 中生成 $(T_j,\ M_j)$ 的样本，

其中 $y = (Y_1, \ldots, Y_n)$ 表示 n 个单位的观测响应。为避免因 M_j 维数变化而产生的复杂性，我们首先从边际分布 $p(T_j \mid (T_h, M_h)_{h \neq j}, B, \omega, y)$ 中抽取 T_j，然后从 $p(M_j \mid T_j, (T_h, M_h)_{h \neq j}, B, \omega, y)$ 中抽取 M_j。为从 $p(T_j \mid (T_h, M_h)_{h \neq j}, B, \omega, y)$ 中抽取样本，我们按照 Chipman 等人（2010）提出的策略生成提议 T_j^*，并计算相应的接受概率。注意，T_j 中叶节点 b 的对数似然 $\log l_{j,b}$ 为

$$\log l_{j,b} = -\frac{n_b}{2}\log(2\pi) + \frac{n_b}{2}\log(\omega) - \frac{1}{2}\log(\sigma_\mu^2) - \frac{1}{2}\log\left(\sigma_\mu^{-2} + \omega \sum_{i \in \{i; Z_i \in G_{jb}\}} k_i\right) + \frac{1}{2}\left\{\frac{\left(\sigma_\mu^{-2}\mu_\mu + \omega \sum_{i \in \{i; Z_i \in G_{jb}\}} k_i R_{ji}\right)^2}{\sigma_\mu^{-2} + \omega \sum_{i \in \{i; Z_i \in G_{jb}\}} k_i} - \omega \sum_{i \in \{i; Z_i \in G_{jb}\}} k_i R_{ji}^2 - \frac{\mu_\mu^2}{\sigma_\mu^2}\right\},$$

且 $R_{ji} = Y_i - \sum_{h \neq j} g(Z_i \mid T_h, M_h) - X_i(c)^\top B \tilde{Z}_i$ 是残差，n_b 表示分配到区域 G_{jb} 的观测值数量，即 $n_b = \#\{i : Z_i \in G_{jb}\}$。抽取一个 T_j 样本后，我们从 $\mathcal{N}(m_n, \sigma_n^2)$ 中对包含在 M_j 中的叶参数 μ_{jb} 进行抽样，其中

$$m_n = \frac{\sigma_\mu^{-2}\mu_\mu + \omega \sum_{i \in \{i; Z_i \in G_{jb}\}} k_i R_{ji}}{\sigma_\mu^{-2} + \omega \sum_{i \in \{i; Z_i \in G_{jb}\}} k_i}, \qquad \sigma_n^2 = \frac{1}{\sigma_\mu^{-2} + \omega \sum_{i \in \{i; Z_i \in G_{jb}\}} k_i}.$$

基于后验样本，我们可以近似任意协变量 z 对应的 $\tau(z)$ 的后验分布。通过计算后验均值，可以得到点估计；此外，还可以通过推导 $\tau(z)$ 的可信区间等方式，获得不确定性度量。

2.4 带宽调整

与局部多项式回归一样，带宽参数 h 在断点回归设计中起着至关重要的作用。具体而言，在所提出的模型下，带宽参数不仅决定局部多项式近似的偏差-方差权衡，还决定 BART 先验实际利用的数据量。为选择 h，我们考虑基于伪模型的样本外观测的留一伪预测分布，

其定义为 $p_h(Y_i \mid Y_{-i}) = \int p_h(Y_i \mid \theta) p_h(\theta \mid Y_{-i}) d\theta$，其中 Y_{-i} 表示除 Y_i之外的观测值集合，$\theta \equiv (\tau, B, \omega)$ 是一组参数，$p_h(Y_i \mid \theta)$ 在（2）中定义，而 $p_h(\theta \mid Y_{-i})$ 表示给定 Y_{-i} 和带宽 h时 θ 的后验分布。由于伪预测分布 $p_h(Y_i \mid Y_{-i})$ 不一定经过归一化，我们采用 Hyv¨arinen 评分（Hyv¨arinen，2005；Shao et al.，2019）作为 h的函数来评估预测分布。具体而言，遵循 Sugasawa et al. (2023)，我们提出以下局部 Hyv¨arinen 评分，以便将评估聚焦于截断点附近：

$$H(h) = \sum_{i \in N_s} \left[2 \frac{\partial^2}{\partial Y_i^2} \log p_h(Y_i \mid Y_{-i}) + \left(\frac{\partial}{\partial Y_i} \log p_h(Y_i \mid Y_{-i}) \right)^2 \right],$$

其中，N_s 是定义为距离阈值最近的 s个观测值的评估集。我们将 s 默认设为 $\max(0.02n, 5)$，并在数值研究中采用这一设定。正如 Yonekura 和 Sugasawa (2023) 所指出的，上述准则可以表示为

$$H(h) = \sum_{i \in N_s} \left\{ 2\mathbb{E}\left[\ell_h^{(2)}(\theta) + \left\{ \ell_h^{(1)}(\theta) \right\}^2 \right] - \left(\mathbb{E}\left[\ell_h^{(1)}(\theta) \right] \right)^2 \right\}, \tag{3}$$

其中，期望是相对于全样本后验分布 $p_h(\theta \mid Y)$ 和 $\ell_h^{(k)}(\theta) \equiv \partial^k \log p_h(Y_i \mid \theta) / \partial Y_i^k$（$k = 1, 2$）而言的。由 (2) 可得

$$\ell_h^{(1)}(\theta) = -\omega k_i (Y_i - \tau(Z_i) W_i - X_i(c)^\top B \tilde{Z}_i), \quad \ell_h^{(2)}(\theta) = -\omega k_i.$$

为了基于 Hyv¨arinen 得分（3）选择带宽参数 h，我们考虑一组有限的候选带宽值 $\{a_1, \ldots, a_L\}$。对于每个候选值 a_ℓ，我们固定 $h = a_\ell$并运行马尔可夫链蒙特卡罗算法进行一小批迭代，以计算相应的 ˆ Hyv¨arinen 得分 $H(a_\ell)$。然后，我们通过 $h = \arg\min_{1 \le \ell \le L} H(a_\ell)$ 选择带宽。

3 模拟研究

我们通过基于两种数据生成过程情景的模拟研究，展示所提出的 Direct-BART 与一些现有方法的表现。

3.1 数据生成过程

两个场景的结果变量均采用以下共同的结构形式：

$$Y_i = \mu(X_i, Z_i) + W_i\tau(Z_i) + \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{N}(0, \sigma^2),$$

其中，$W_i = I(X_i \geq c)$表示断点值为 $c = 0$的处理指示变量。两个场景在运行变量 X_i 和协变量 Z_i的分布，以及基线函数 μ 和处理效应函数 τ的设定方面有所不同。

（场景1）我们设定 $n = 1200$，样本的前半部分和后半部分分别对应对照组个体和处理组个体。对照组个体的运行变量 X_i 服从U$(-1, 0)$分布，处理组个体的运行变量服从U$(0, 1)$分布。五维协变量向量 $Z_i = (Z_{i1}, \ldots, Z_{i5})^\top$ 由四个连续变量和一个分类变量组成，其中$(Z_{i1}, \ldots, Z_{i4})^\top \sim \mathcal{N}_4((\gamma_0 + \gamma_1 X_i)1_4, \Sigma)$，且 1_4 是一个由1构成的四维向量。我们设定$(\gamma_0, \gamma_1) = (-1, 0.55)$，得到Cor$(X_i, Z_{i1}) \approx 0.3$。协方差矩阵 Σ 的(j, k)元素为 $\Sigma_{jk} = \{1 + |j - k|\}^{-1}$，其中 $j, k \in \{1, 2, 3, 4\}$，表示中等相关性。分类协变量 $Z_{i5} \in \{1, 2, 3\}$ 生成如下：对于 $k = 1$，2，3，Pr$(Z_{i5} = k)$ $=$ exp$(\eta_{ik})/\sum_{k'=1}^{3}$ exp$(\eta_{ik'})$，其中 $\eta_{i1} = 0.8X_i + 0.5Z_{i1} - 0.3Z_{i2}$，$\eta_{i2} = -0.4X_i + 0.2Z_{i1} + 0.4Z_{i2}$，且 $\eta_{i3} = 0$。

基线函数 μ 被指定为 $\mu(x, z) = g_{\text{int}}(z) + g_{\text{slope}}(z)\, f(x)$，其中函数 f 的构造满足 $f(0) = 0$。因此，截断点处的基线水平

由 $\mu(0, z) = g_{\text{int}}(z)$ 给出。分量函数定义为

$$g_{\text{int}}(z) = \alpha_\mu \Big[1 + \Phi\left(\frac{z_1 + 1}{2}\right) + 0.1 \sin(\pi z_1) + \frac{1}{\pi} \tan^{-1}(z_2 - 1) + 0.5\delta_1(z_5) + \delta_3(z_5)\Big]$$

$$g_{\text{slope}}(z) = 2 + \frac{\exp(z_1)}{1 + \exp(z_1)} + I(z_3 < 0)\sqrt{|z_3|} + \max\{0, z_4\} + \delta_3(z_5),$$

以及 $f(x) = x + \sin(2\pi x)$，其中 Φ 是标准正态分布的累积分布函数，且 $\delta_a(z_5) = I(z_5 = a)$，其中 I 为示性函数。参数 α_μ 通过 $\text{Var}(\mu(0, Z) \mid X = 0)$ 直接控制截断点处的基线变异幅度。处理效应函数 τ 表示为

$$\tau(z) = \alpha_\tau \Big\{1 + 0.5 \cos(2\pi z_1) + 0.6 z_2 z_3 + 0.4 z_4 - 0.5 I(z_5 = 2)\Big\}.$$

缩放参数 α_τ 的选取使得 $\text{Var}\big(\tau(Z) \mid X = 0\big) = 0.5$。我们考虑基线变异幅度不同的两种情形：(a) $\text{Var}(\mu(0, Z) \mid X = 0) = 1$ （小基线变异性）；(b) $\text{Var}(\mu(0, Z) \mid X = 0) = 15$ （大基线变异性）。对于每种情形，我们将噪声方差设为不同取值 $\sigma^2 \in \{0.25, 0.5, 1\}$。在小基线变异性的情况下，处理效应的变异占结果变量总变异的很大一部分，因此条件平均处理效应估计较为容易。相比之下，在困难情形中，基线变异主导结果变量的变异，导致处理效应估计的信噪比较低，使条件平均处理效应估计变得困难。

(情景2) 我们采用 Alcantara 等人（2025）最初提出的数据生成过程。设定 $n = 600$，并从多元正态分布中生成四维协变量向量 $Z_i = (Z_{i1}, \ldots, Z_{i4})^\top$， $Z_i \sim \mathcal{N}_4(0, \Sigma_Z)$，其中 Σ_Z 为托普利兹协方差矩阵，其元素介于 0 和 2 之间。运行变量 X_i 由正态分布生成。$X_i \mid Z_i \sim \mathcal{N}(\gamma_0 + \gamma^\top Z_i, \nu)$，其中 γ 为各分量相同的系数向量。通过加上常数项 $\gamma_0 = 1$使均值向处理区域偏移，得到的运行变量分布便会围绕 0 处的截断点呈现不对称。关键参数为 ρ，它控制运行变量与协变量之间相关性

的强度。给定 ρ的指定值，确定参数 γ 和 ν，使得 $\mathrm{Var}(X_i) = 1$ 且 $\mathrm{Cor}(X_i, \gamma^\top Z_i) = \rho$。基线函数 μ 构造为 $\mu(x, z) = \beta_\mu \mu^\star(x, z)$，其中缩放常数 β_μ 的取值使得 $\mathrm{Var}(\mu(0, Z) \mid X = 0) = 1$。未缩放函数 $\mu^\star(x, z)$ 定义为

$$\mu^\star(x, z) = (x+1)^3 + (z^\star + 2)^2 \Big(\mathrm{sign}(x+1)\sqrt{|x+1|} \Big),$$

其中 $z^\star = \sum_{j=1}^4 z_j/\sqrt{4}$。处理效应函数 τ定义为 $\tau(z) = \alpha_\tau + \beta_\tau \tau^\star(z)$，其中平移参数 α_τ和缩放参数 β_τ固定为满足

$$\sqrt{\mathrm{Var}(\tau(Z) \mid X = 0)} = 1 \text{ and } \min_{Z|X=0} \tau(Z) = 0.$$

未缩放函数 $\tau^\star(z)$ 定义为

$$\tau^\star(z) = \frac{1}{2}\Phi(2z_1 + 3) + \phi(z_1),$$

其中，Φ 和 ϕ 分别是标准正态分布的累积分布函数和概率密度函数。我们考虑 ρ 的三种取值：(a) $\rho = 0$ （无相关性）；(b) $\rho = 0.25$（低相关性）；以及 (c) $\rho = 0.5$（高相关性）。在所有设置中，我们将噪声方差设为 $\sigma^2 \in \{0.5, 1\}$。

3.2 比较方法与评估准则

对于模拟数据集，我们采用了以下方法：

- **Direct-BART** （提出的方法）：通过截断点附近的局部损失函数，将 BART（包含 20 棵树）直接应用于条件平均处理效应（CATE）。根据对结果变量与运行变量之间关系的探索性考察，场景1采用 $q = 2$ 阶局部多项式，情景2采用 $q = 1$ 阶局部多项式。对于带宽调整，我们在从 0 到局部多项式估计量所选带宽两倍的区间内取六个等间距值，构造候选带宽集合。为计算 Hyv¨arinen 得分（3），我们在舍弃前 500 个样本作为预烧期后，使用 500 个后验样本。

- **BARDDT**：该方法是 Alcantara 等人（2025）提出的一种基于 BART 的方法，其中使用运行变量和协变量对数据进行递归划分。在每个叶节点内，拟合一个允许在截断点处发生跳跃的分段线性函数。该方法通过使用 R 软件包 ”stochtree” 实现，采用 20 棵树和 50 棵树。

- **S-BART**：该方法对应于 S-学习器的实现，使用 BART（含 20 棵树）对结果变量的条件期望建模，并将处理指示变量 W_i 作为额外的协变量纳入模型。该方法通过 R 软件包“stochtree”实现。

- **LP**：使用 R 软件包“rdrobust”实现的局部多项式估计量，不含协变量。由于该方法仅提供平均处理效应，因此所得的条件平均处理效应估计为常数。

对于 Direct-BART，在两个场景中，我们均丢弃前 500 个样本作为预烧期，并生成 4500 个后验样本。对于其他基于 BART 的方法，在场景1中，我们按照 Starling 等人（2020）的马尔可夫链蒙特卡罗设置，丢弃前 1000 个样本作为预烧期后，生成 10000 个后验样本。在场景2中，我们首先进行了 30 次从根生长（GFR）迭代，以获得初始树集，然后按照 Alcantara 等人（2025）的马尔可夫链蒙特卡罗设置生成 1000 个后验样本。请注意，不含协变量的局部多项式估计量可提供平均处理效应估计量，并作为基准，用以说明显式建模条件处理效应异质性所带来的提升。我们的比较中不包括基于 BART 的 T-学习器实现（T-BART）。T-BART 分别对处理组和对照组拟合两个独立的 BART 模型，并将截断点处预测结果之差作为处理效应估计。由于截断点处的预测必然涉及外推，而基于树的方法通常在外推任务中表现不佳，因此我们不将 T-BART 纳入比较。

我们使用以下指标评估每种方法的样本内和样本外表现。

均方根误差（RMSE）和95%可信区间的覆盖概率，评估各方法的样本内和样本外表现。对于样本外评估，我们根据条件分布 $Z \mid X = 0$ 新生成200个目标单位。相比之下，样本内评估使用原始样本中的观测单位进行。遵循 Alcantara 等人（2025）的做法，我们考虑如下定义的条件平均处理效应函数： $\tau(Z_i) \equiv \mathbb{E}[Y_i(1) - Y_i(0) \mid Z_i]$。由于远离截断点 $X = c$ 的观测值对处理效应的识别贡献甚微，我们将关注范围限定在截断点附近的单位。具体而言，对于$i \in \{i : |X_i - c| \leq 0.1 \text{ sd}(X)\}$，我们评估 $\tau(Z_i)$，其中 sd(X) 是 X_i 的标准差。所有结果均基于两种情景下各设置的15次重复。

3.3 结果

我们首先展示场景1中的结果。图 1 展示了四种方法在三个噪声方差水平下、针对样本内单位计算的均方根误差箱线图，并分别呈现小基线变异性和大基线变异性两种情形。图 2 展示了针对样本外单位计算的相应结果。在小变异性情形下，除情形 (c) 外，所提方法在所有设置中的均方根误差均低于竞争方法。在高噪声设置 (c) 中，所提方法的表现与局部多项式估计量和 S-BART 相当。在大变异性情形下，无论噪声方差如何，所提方法的均方根误差通常都低于竞争方法。总体而言，这些结果表明，在场景1下，就均方根误差而言，所提方法能够实现准确估计。表 1 报告了四种方法在三个噪声方差水平下点估计的平均均方根误差和 95% 可信（置信）区间的平均覆盖概率，并分别呈现小基线变异性和大基线变异性两种情形。在区间估计方面，与其他基于 BART 的替代方法相比，Direct-BART 提高了覆盖率，但仍存在一定程度的覆盖不足，尤其是在大变异性设置下。总体而言，这些结果表明，在场景1下，与竞争的基于 BART 的方法相比，Direct-BART 能够提供更可靠的区间估计。

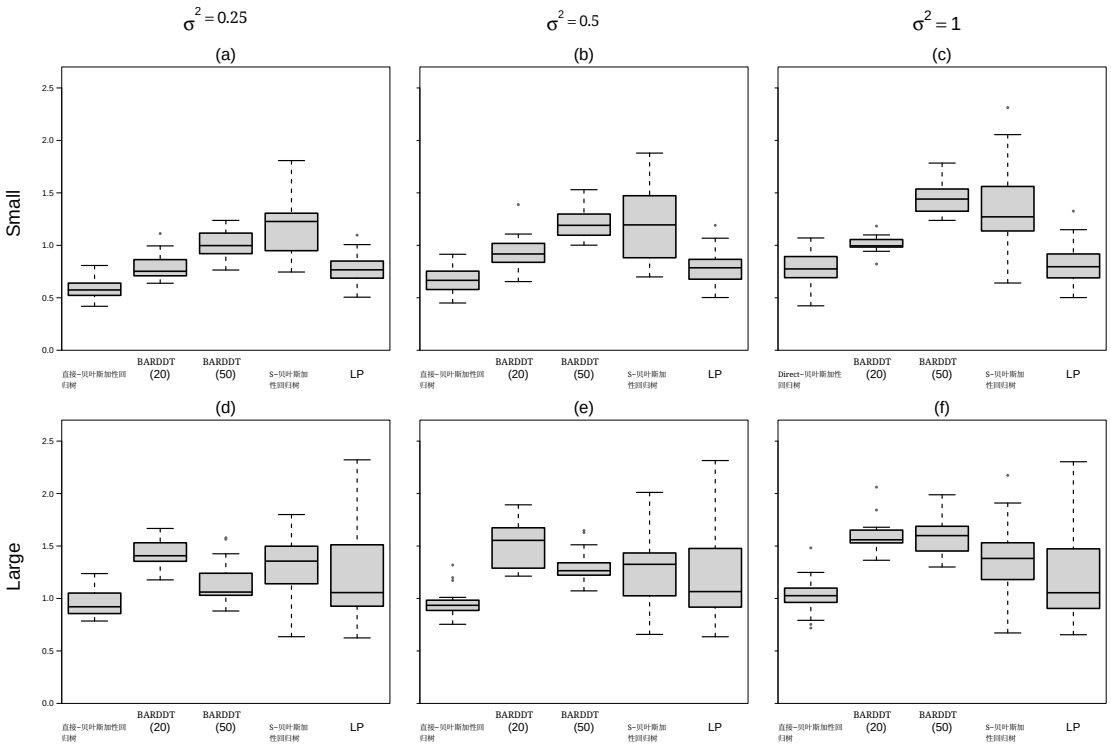

图 1：场景1下四种方法基于15次重复得到的样本内单位均方根误差箱线图。上排对应小基线变异性情形，下排对应大基线变异性情形。左列、中列和右列分别报告噪声方差为 $\sigma^2 = 0.25$、0.5和1时的结果。

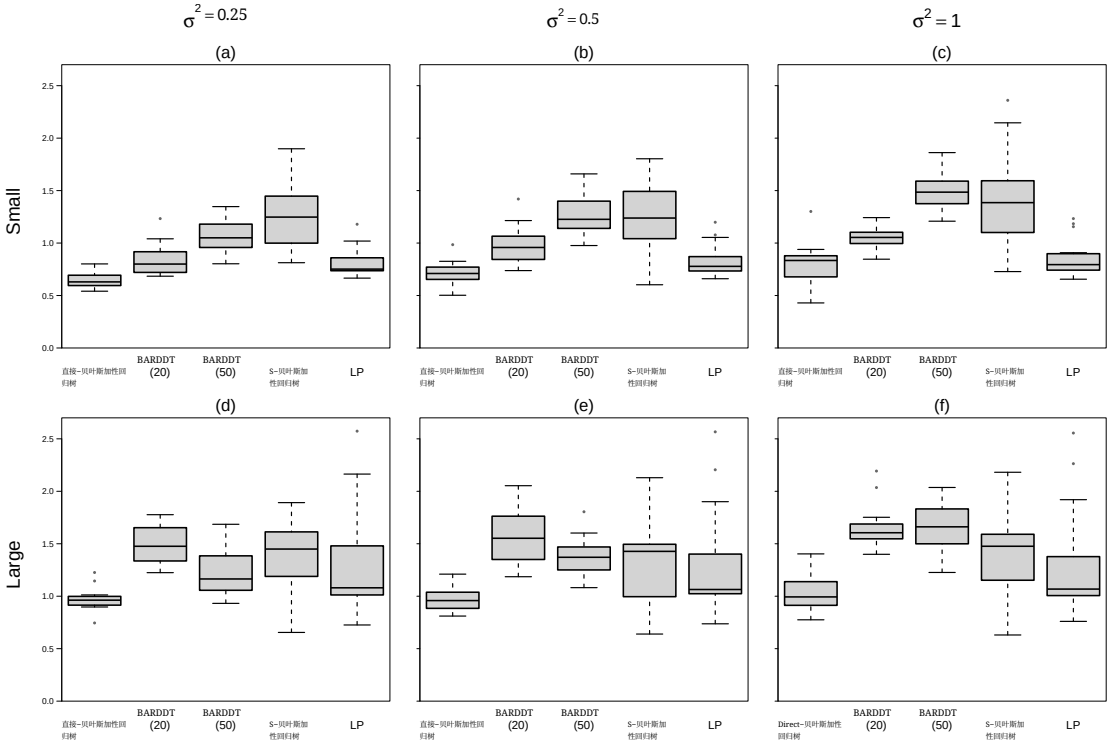

图 2：场景1下四种方法基于15次重复所得的样本外单位均方根误差箱线图。上排对应小基线变异性情形，下排对应大基线变异性情形。左、中、右列分别报告噪声方差为 $\sigma^2 = 0.25$、0.5和1时的结果。

表 1：在场景1下，基于15次蒙特卡洛重复计算的点估计均方根误差及95%可信（置信）区间覆盖率的平均值。

	Case	小变异性			大变异性		
	σ^2	0.25	0.5	1.0	0.25	0.5	1.0
均方根误差（样本内）	Direct-BART	0.59	0.66	0.78	0.95	0.97	1.03
	BARDDT（20）	0.80	0.94	1.01	1.43	1.51	1.61
	BARDDT（50）	1.01	1.22	1.47	1.15	1.31	1.60
	S-BART	1.18	1.21	1.37	1.32	1.26	1.37
	LP	0.78	0.79	0.83	1.28	1.27	1.27
均方根误差（样本外）	Direct-BART	0.64	0.72	0.80	0.97	0.98	1.03
	BARDDT（20）	0.85	0.97	1.04	1.48	1.56	1.66
	BARDDT（50）	1.06	1.26	1.50	1.22	1.38	1.66
	S-BART	1.24	1.27	1.40	1.39	1.31	1.41
	LP	0.82	0.83	0.86	1.31	1.30	1.30
覆盖概率 (%) (样本内)	Direct-BART	88.9	91.1	93.9	74.9	83.2	90.2
	BARDDT (20)	50.4	42.5	42.1	34.7	36.5	33.4
	BARDDT (50)	41.6	35.6	28.6	48.9	44.1	38.8
	S-BART	41.5	51.5	51.7	36.5	43.9	49.7
	LP	75.8	77.0	79.2	93.4	93.1	93.0
覆盖概率（%）（样本外）	Direct-BART	88.3	91.6	94.2	75.6	83.7	91.7
	BARDDT (20)	48.9	43.1	40.6	34.8	34.9	34.4
	BARDDT (50)	41.4	34.8	26.8	47.0	43.0	37.5
	S-BART	39.1	50.3	51.6	34.6	43.4	47.9
	LP	74.7	76.6	79.0	92.9	92.9	92.8

接下来，我们展示情景2下的结果。图 3 展示了四种方法在三个相关性水平下对样本内单位进行评估时的均方根误差箱线图，并分别报告了两个噪声方差水平下的结果。图 4 展示了对样本外单位进行评估时的相应结果。在该设置下 $\rho = 0$，就样本内和样本外评估而言，Direct-BART 的均方根误差均低于其他方法。随着 ρ 增加，Direct-BART 在均方根误差方面的优势有所减弱；然而，在所有设置下，其表现仍具有竞争力。这些结果表明，在情景2下，Direct-BART 保持了稳健的点估计性能。表 2 报告了三个相关性水平下点估计的平均均方根误差和 95% 可信（置信）区间的平均覆盖概率，并分别列出了两个噪声方差水平下的结果。关于区间估计，Direct-BART 在所有设置下均保持了相对稳定的覆盖率，但仍存在一定程度的覆盖不足，尤其是在ρ 较大时。总体而言，这些结果表明，在情景2下，与竞争性的基于 BART 的方法相比，Direct-BART 提供了更加稳定的区间估计。结合点估计和区间估计方面的评估结果来看，在情景2下，Direct-BART 在点估计准确性与区间估计稳定性之间实现了较为理想的平衡。

4 总结

本文提出了一种基于贝叶斯加性回归树的贝叶斯非参数方法，用于估计断点回归设计中的条件平均处理效应。通过针对新设计和先前研究过的数据生成过程开展大量模拟研究，我们发现，与现有方法相比，所提出的方法始终具有较优的均方根误差表现，因为它能够捕捉异质性处理效应的复杂结构。

尽管本文采用了标准的 BART 先验，但文献中已提出多种扩展方法和计算改进。例如，为处理高维协变量（Linero, 2018）和大型数据集（He 等，2019），研究者提出了贝叶斯加性回归树的可扩展变体和正则化形式。纳入-

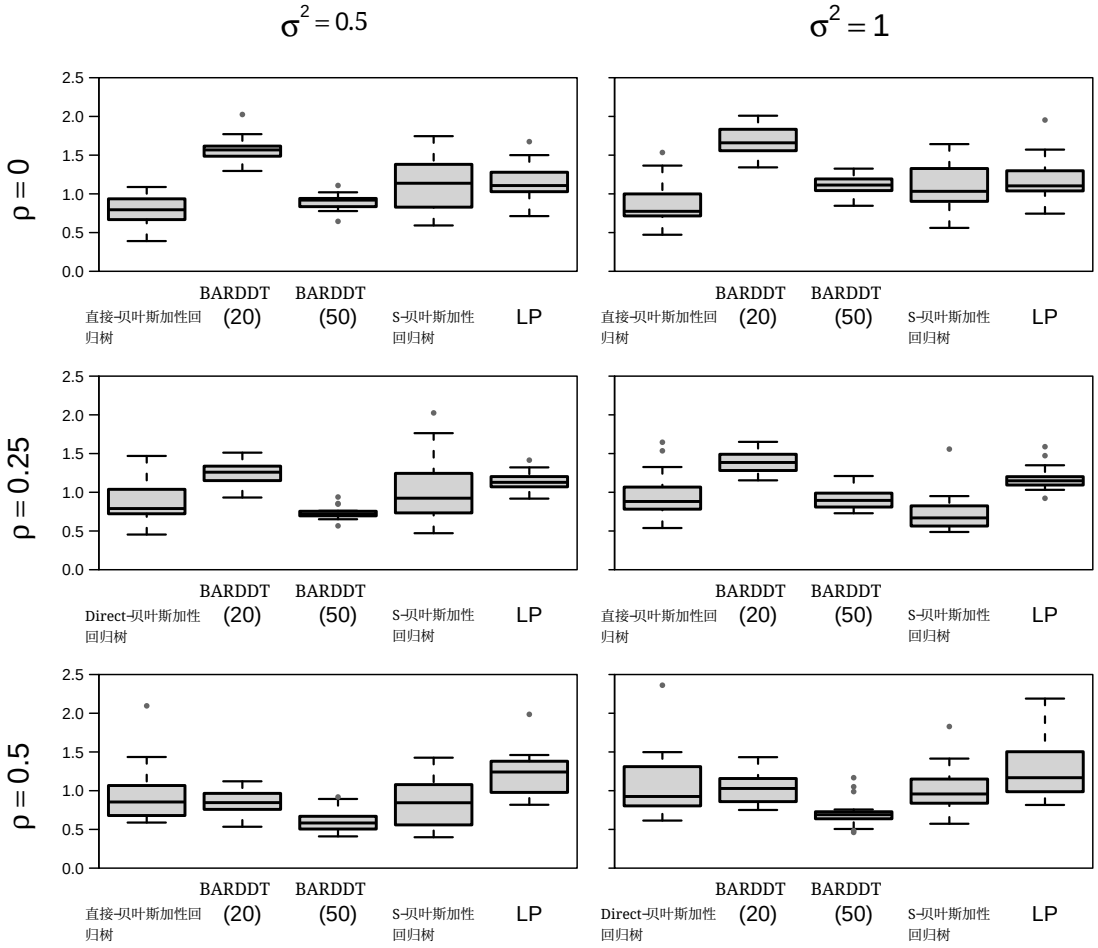

图 3：情景2下，基于15次重复对四种方法的样本内单位均方根误差所绘制的箱线图。顶部、中部和底部面板分别对应设置 $\rho = 0$、$\rho = 0.25$和 $\rho = 0.5$。对于每一行，分别报告噪声方差为 $\sigma^2 = 0.5$和 $\sigma^2 = 1$时的结果。

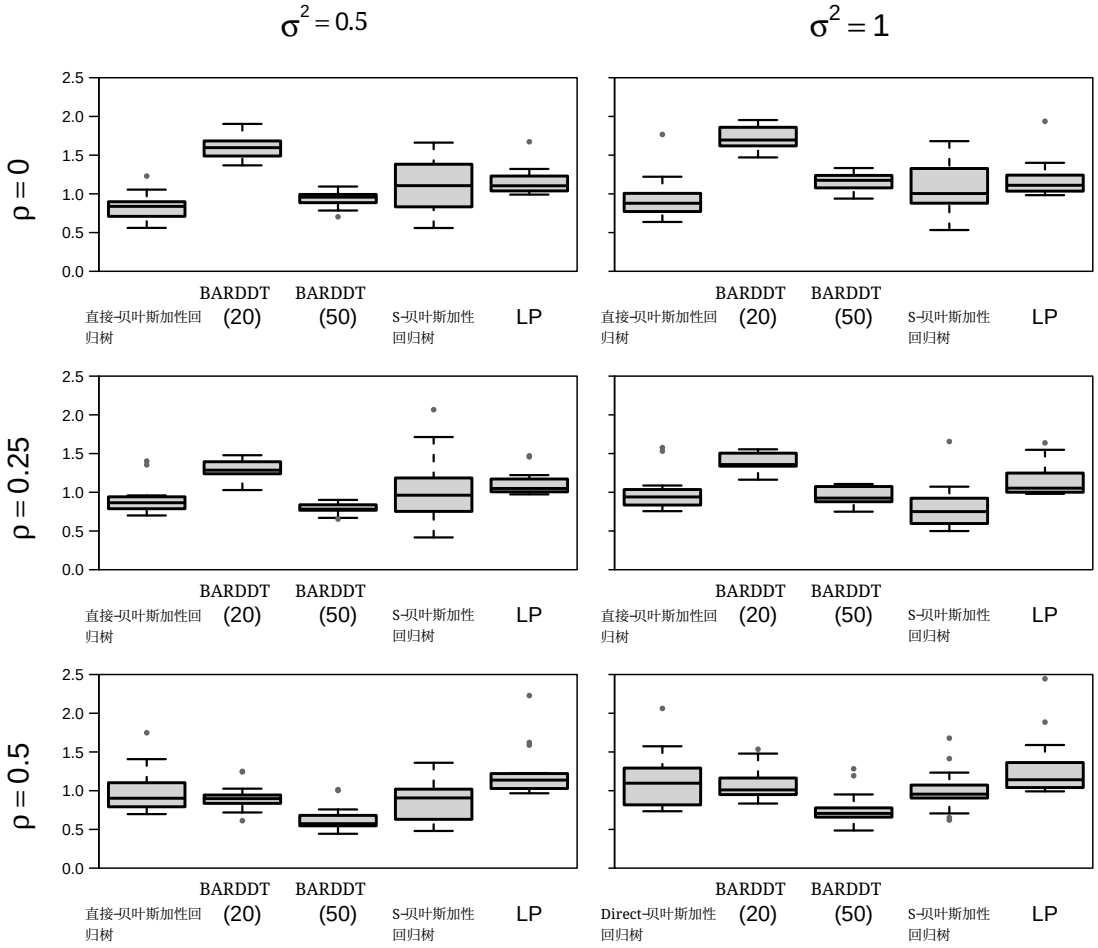

图 4：情景2下，四种方法基于15次重复得到的样本外单位均方根误差箱线图。顶部、中部和底部面板分别对应设置$\rho = 0$、 $\rho = 0.25$和 $\rho = 0.5$。每一行的结果分别针对噪声方差 $\sigma^2 = 0.5$和 $\sigma^2 = 1$进行报告。

表 2：情景2下，15次蒙特卡洛重复试验中点估计的均方根误差以及95%可信（置信）区间的覆盖概率的平均值。

	ρ	0		0.25		0.5	
	σ^2	0.5	1	0.5	1	0.5	1
均方根误差（样本内）	Direct-BART	0.79	0.89	0.86	0.97	0.96	1.11
	BARDDT（20）	1.57	1.68	1.25	1.38	0.86	1.04
	BARDDT（50）	0.90	1.11	0.73	0.91	0.61	0.72
	S-BART	1.12	1.09	1.02	0.73	0.86	1.03
	LP	1.15	1.18	1.14	1.18	1.21	1.28
均方根误差（样本外）	Direct-BART	0.84	0.93	0.91	1.00	1.00	1.13
	BARDDT (20)	1.61	1.72	1.29	1.40	0.91	1.09
	BARDDT (50)	0.93	1.15	0.79	0.96	0.64	0.76
	S-BART	1.11	1.09	1.03	0.80	0.86	1.00
	LP	1.16	1.18	1.12	1.15	1.23	1.29
覆盖概率（%）（样本内）	Direct-BART	91.0	92.8	90.1	93.1	85.2	84.0
	BARDDT（20）	15.3	14.7	23.0	19.4	44.9	35.4
	BARDDT（50）	34.9	30.8	49.9	45.2	72.6	71.1
	S-BART	66.1	73.0	79.1	96.3	94.9	96.6
	LP	65.6	70.1	72.7	76.5	66.1	66.7
覆盖概率（%）（样本外）	Direct-BART	91.8	94.1	91.5	92.7	86.1	87.5
	BARDDT（20）	14.9	13.1	23.7	22.5	41.0	34.6
	BARDDT（50）	35.5	31.7	47.3	42.8	71.2	67.0
	S-BART	67.1	74.0	79.6	94.8	96.5	94.4
	LP	67.6	73.4	75.0	78.0	65.6	65.9

将此类先进的贝叶斯加性回归树变体纳入所提出的框架，将是未来研究中一个有前景的方向，尤其是在具有高维协变量或复杂处理效应异质性的情形中。另一项重要扩展是考虑模糊回归不连续设计情形（例如 Imbens 和 Lemieux，2008），其中处理分配并非完全由运行变量决定。将所提出的框架扩展以容纳模糊设计，例如通过贝叶斯加性回归树框架内的工具变量形式，将拓宽该方法的适用性，并留待未来研究。

致谢

本研究部分得到日本学术振兴会（日本学术振兴会科学研究费助成事业）资助，项目编号为24K21420和25H00546。

参考文献

Alcantara, R.、P. R. Hahn、C. Carvalho 和 H. Lopes（2025）。利用贝叶斯加性回归树学习回归不连续设计中的条件平均处理效应。*arXiv*预印本 *arXiv:2503.00326*。

Becker, S. O.、P. H. Egger 和 M. Von Ehrlich（2013）。吸收能力与区域转移支付对增长和投资的影响：采用异质性处理效应的回归不连续设计。《美国经济学杂志：经济政策》*5*（4），29–77。

Bissiri, P. G.、C. C. Holmes 和 S. G. Walker（2016）。信念分布更新的一般框架。《皇家统计学会杂志：*B*辑（统计方法论）》*78*（5），1103–1130。

Calonico, S., M. D. Cattaneo, M. H. Farrell, F. Palomba 和 R. Titiunik（2025）。

回归不连续设计中的处理效应异质性。*arXiv* 预印本 *arXiv:2503.13696*。

Calonico, S., M. D. Cattaneo 和 R. Titiunik（2014）。回归不连续设计的稳健非参数置信区间。《计量经济学》*82*(6), 2295–2326。

Card, D., C. Dobkin 和 N. Maestas（2008）。近乎全民保险覆盖对医疗服务利用的影响：来自美国医疗保险的证据。《美国经济评论》*98*(5), 2242–2258。

Chib, S., E. Greenberg 和 A. Simoni（2023）。锐断点与模糊回归不连续设计的非参数贝叶斯分析。《计量经济学理论》*39*(3), 481–533。

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

Hahn, J.、P. Todd 与 W. Van der Klaauw（2001）。基于回归断点设计的处理效应识别与估计。《计量经济学》*69*(1)，201–209。

Hahn, P. R.、J. S. Murray 与 C. M. Carvalho（2020年）。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。

He, J.、S. Yalov 与 P. R. Hahn（2019）。加速贝叶斯加性回归树。载于第*22*届国际人工智能与统计会议，第1130–1138页。机器学习研究论文集。

Hyv¨arinen, A.（2005）。基于得分匹配的非归一化统计模型估计。《机器学习研究杂志》*6*(4)，695–709。

Imbens, G. 和 K. Kalyanaraman（2012）。回归不连续估计量的最优带宽选择。《经济研究评论》*79*(3), 933–959。

Imbens, G. W. 和 T. Lemieux（2008）。回归不连续设计：实践指南。计量经济学杂志 *142*(2), 615–635。

Lee, D. S. 和 T. Lemieux（2010年）。经济学中的回归不连续设计。《经济文献杂志》*48*(2),281–355。

Linero, A. R.（2018）。用于高维预测和变量选择的贝叶斯回归树。《美国统计学会会刊》*113*(522), 626–636。

´Reguly, A.（2021）。在回归不连续设计中发现异质性处理效应。*arXiv* 预印本 *arXiv:2106.11640*。

Sawada, M., T. Ishihara, D. Kurisu 和 Y. Matsuda（2024）。多变量回归不连续设计的局部多项式估计。*arXiv* 预印本 *arXiv:2402.08941*。

Shao, S., P. E. Jacob, J. Ding 和 V. Tarokh（2019）。使用 Hyv¨arinen 得分进行贝叶斯模型比较：计算与一致性。《美国统计学会会刊》*114*(528), 1826–1837。

Starling, J. E., J. S. Murray, C. M. Carvalho, R. K. Bukowski, 和 J. G. Scott（2020年）。采用目标平滑的贝叶斯加性回归树：患者特异性死产风险分析。《应用统计年鉴》。

Sugasawa, S., T. Ishihara, 和 D. Kurisu（2023）。层级回归不连续设计：探究亚组处理效应。*arXiv*预印本 *arXiv:2309.01404*。

Tao, K., Y. S. Wang, 和 D. Ruppert（2025）。异质性处理效应下回归不连续设计的贝叶斯分析。*arXiv* 预印本 *arXiv:2504.10652*。

Thistlethwaite, D. L. 和 D. T. Campbell（1960）。回归不连续分析：事后实验的替代方案。《教育心理学杂志》*51*（6），309–317。

Yonekura, S. 和 S. Sugasawa（2023）。广义贝叶斯推断中稳健散度调节参数的自适应。《统计与计算》*33*(2)，39。