Statistical Theory and Related Fields

ISSN: 2475-4269 (Print) 2475-4277 (Online) Journal homepage: www.tandfonline.com/journals/tstf20

Leveraging density ratio models in a binary instrumental variable inference with a binary outcome: A retrospective approach

Wenli Liu, Jing Qin & Yukun Liu

To cite this article: Wenli Liu, Jing Qin & Yukun Liu (2025) Leveraging density ratio models in a binary instrumental variable inference with a binary outcome: A retrospective approach, Statistical Theory and Related Fields, 9:4, 331-356, DOI: 10.1080/24754269.2025.2537517

To link to this article: https://doi.org/10.1080/24754269.2025.2537517

STATISTICAL THEORY AND RELATED FIELDS
2025, VOL. 9, NO. 4, 331–356
https://doi.org/10.1080/24754269.2025.2537517

Leveraging density ratio models in a binary instrumental variable inference with a binary outcome: A retrospective approach

Wenli Liu[a], Jing Qin[b] and Yukun Liu[a]

[a]KLATASDS-MOE, School of Statistics, East China Normal University, Shanghai, People's Republic of China; [b]National Institute of Allergy and Infectious Diseases, MD, USA

ABSTRACT

Conditional Local Risk Ratio (CLRR) is a widely used metric for assessing heterogeneous treatment effects of binary outcomes in randomized clinical trials involving noncompliance. Existing methods, such as moment-based and likelihood-based approaches, often overlook the inherent mixture structure in data, necessitate stringent parametric assumptions, or yield estimates with implausible values. In this paper, we introduce a novel semiparametric likelihood-based (SPL) method for estimating CLRR. Our method requires only three parametric model assumptions, significantly fewer than the six models needed by existing likelihood-based methods, thereby reducing model complexity and enhancing robustness. This simplicity also results in fewer unknown parameters, further boosting computational efficiency. Unlike moment-based methods, our SPL method fully exploits the mixture structure of the observed data and the principal strata framework. Additionally, our method ensures that the final CLRR estimate always fall within a valid range. We establish the asymptotic normality of our estimator and demonstrate its superiority over existing methods through numerical simulations. We further apply our method to analyze the Oregon Health Insurance Experiment dataset, providing valuable insights into the heterogeneous effects of Medicaid on both physical and mental health.

ARTICLE HISTORY
Received 22 January 2025
Accepted 16 July 2025

KEYWORDS
Conditional local risk ratio; instrumental variable; noncompliance; retrospective approach; semiparametric model

1. Introduction

Randomized controlled trials (RCTs) are universally acknowledged as the gold standard for assessing the efficacy of novel interventions or treatments across diverse scientific disciplines (Kohavi & Thomke, 2017; Piantadosi, 2024; Shadish et al., 2002). This paper specifically focuses on binary outcomes, which are commonly employed to capture crucial endpoints such as depression status (depressed or not) (Taylor et al., 2016), mortality (dead or alive) (Stamler et al., 1986), and disease status (ill or not) (Have et al., 2003), particularly prevalent in global health research. The binary outcome is pivotal in RCTs due to their extensive use in trial design, as evidenced by Charles et al. (2009), who reported that nearly half of the sample

CONTACT Yukun Liu ykliu@sfs.ecnu.edu.cn KLATASDS-MOE, School of Statistics, East China Normal University, Shanghai 200062, People's Republic of China

Supplemental data for this article can be accessed online at http://dx.doi.org/10.1080/24754269.2025.2537517.

size calculations in the trials they reviewed were based on binary outcomes, underscoring their critical role in determining statistical power and facilitating treatment comparisons.

However, in practice, ethical considerations and individual noncompliance frequently lead to deviations from the intended protocols in RCTs, with the reasons for these deviations often remaining unobserved (Baicker et al., 2013; Lois et al., 2008). This phenomenon, known as nonnegligible noncompliance, introduces unmeasured confounders that concurrently influence both the treatments received and the outcomes. It poses a substantial challenge in assessing the effectiveness of experimental treatments in clinical trials and causal inference, as it violates the unconfoundedness assumption required by traditional causal inference methods (Dodd et al., 2012).

The existing methodologies for addressing noncompliance in RCTs can broadly be categorized into two groups: naive approaches and instrumental variable (IV) approaches. Naive approaches encompass intent-to-treat (ITT), as-treated (AT), and per-protocol (PP) analyses. A critical limitation shared by these naive methods is that they fail to provide valid estimates of the target causal effects due to selection bias and confounding bias.

On the contrary, IV approaches leverage the random assignment mechanism as an instrument, enabling the identification of causal effects even in the presence of unmeasured confounders. With IV, the literature has explored both parametric identification of average treatment effect (ATE) (Imbens, 2004; Robins, 1994) and nonparametric identification of local average treatment effect (LATE) (Frölich, 2007; Imbens & Angrist, 1994), which focuses on the average treatment effect among potential compliers. However, for bounded causal estimands induced by binary outcomes, the aforementioned methods may yield misleading or absurd estimates. For instance, in the context of binary outcomes, the additive LATE is theoretically bounded within the range of $[-1, 1]$, yet its estimates may fall outside this range, which is absurd (Abadie, 2002; Hirano et al., 2000; Tan, 2006).

Instead of relying on traditional additive LATE, multiplicative LATE measures such as causal risk ratio (RR) and causal odds ratio (OR) are commonly used to assess the impact of a binary treatment on a binary outcome. It is well-documented that odds ratios are not collapsible, meaning the marginal odds ratio does not fall within the convex hull of stratum-specific odds ratios. Conversely, risk ratios are collapsible, making them a more favorable choice (Rothman et al., 2008). Furthermore, the pervasive nature of individual heterogeneity highlights the critical need to incorporate covariates in analytical frameworks. This integration can improve estimation efficiency, enable identification of distinct subpopulations, and support evidence-based individualized decision-making. These substantive considerations jointly emphasize the methodological importance of conducting inference on conditional local risk ratios (CLRR) – the multiplicative counterpart to the additive LATE.

Recently, there has been a surge of interest in CLRR within the literature. Existing methods for estimating CLRR are semiparametric methods, which can broadly be categorized into moment-based and likelihood-based approaches. Moment-based methods rely on specific moment conditions to obtain estimates. For example, Abadie (2003) specifies a model for the outcome given compliers, treatment, and covariates (such as a logistic model), minimizing a weighted ordinary least squares criterion. However, this approach may suffer from instability and lack interpretability in the estimated parameters, which arises from the use of negative weights for individuals whose received treatment deviates from their assigned treatment. Alternatively, Okui et al. (2012) and Ogburn et al. (2015) proposed doubly robust methods that offer additional flexibility but do not fully leverage compliance information and can be highly sensitive to initial values when solving certain estimating equations. In RCTs

with noncompliance, the data are best viewed through a mixture model framework under the theory of principal strata (Imbens & Rubin, 1997). Unfortunately, moment-based methods often fail to fully address the complexities arising from this mixture.

On the other hand, likelihood-based approaches aim to maximize the likelihood function. While maximum likelihood and Bayesian methods provide a theoretically sound framework under six parametric model specifications (Hirano et al., 2000; Imbens & Rubin, 1997; Little & Yau, 1998), their validity is compromised if the working models are misspecified. Wang et al. (2021) assumes model specifications that are variation-independent from the CLRR and introduces a one-to-one mapping to traditional parametric models, but their method also risks model misspecification. Compared to moment-based methods, likelihood-based methods are computationally more demanding due to the involvement of more unknown parameters. The trade-off between computational efficiency and robustness remains a critical challenge in the estimation of CLRR.

The aforementioned discussions motivate the development of our semiparametric likelihood-based (SPL) method for RCTs with noncompliance and binary outcomes. Firstly, unlike moment-based methods, SPL fully utilizes the mixture structure in the data and the principal strata framework. It is specifically designed for estimating CLRR, effectively addressing binary outcomes and accounting for individual heterogeneity. Secondly, the proposed SPL method imposes parametric assumptions only on the three density ratios, inspired by Anderson (1979). It requires fewer parametric assumptions than existing likelihood-based methods, such as Wang et al. (2021), which rely on six parametric models. This enhances robustness, reduces the number of unknown parameters, and significantly decreases computational time. Thirdly, a key advantage of our approach is that the final CLRR estimate always remains within a valid range, ensuring plausible results.

The remainder of this paper is organized as follows: In Section 2, we introduce the IV setup, provide definitions of potential outcomes and assumptions referred to throughout the paper, and present basic notions for later use. Section 3 details the proposed estimation procedure and its large-sample properties. In Section 4, we evaluate the finite sample performance of the proposed estimators via simulations. Our method is applied to estimate the causal effect of a health insurance program on physical and mental health in OHIE using data from the RCTs conducted by Baicker et al. (2013). Section 6 concludes with a brief discussion. For clarity, all proofs are relegated to the Appendix A.

2. Methodology

2.1. Notations and assumptions

We consider estimating causal effects in randomized controlled trials (RCTs) with noncompliance and a binary outcome Y. Let Z represent the treatment assigned, D represent the actual treatment received. Here, $Z = 1$ and 0 indicate assignments to the treatment and control groups, respectively, while $D = 1$ and 0 indicate whether the treatment was received or not, respectively. Due to potential noncompliance, the treatment received D may differ from the treatment assigned Z, leading to confounding of the effect of D on Y by both observed covariates X and unobserved confounders U.

Suppose that the sample size is n. We adopt the potential outcomes framework (Rubin, 1974), and let $D(z_1, \ldots, z_n) \in \{0, 1\}$ be the potential treatment received under randomization assignment $(Z_1, \ldots, Z_n) = (z_1, \ldots, z_n)$. Similarly, let $Y(z_1, \ldots, z_n, d_1, \ldots, d_n)$

Table 1. Principal stratum describing potential complier status $(D(1), D(0))$.

$D(1)$	$D(0)$	Principal stratum	Abbreviation S
1	1	Always-taker	a
1	0	Complier	c
0	1	Defier	d
0	0	Never-taker	n

represent the potential outcome under treatment assigned $(Z_1, \ldots, Z_n) = (z_1, \ldots, z_n)$ and treatment received $(D_1(z_1, \ldots, z_n), \ldots, D_n(z_1, \ldots, z_n)) = (d_1, \ldots, d_n)$. We make the following commonly-used assumption:

(A0) Stable Unit Treatment Value Assumption (SUTVA): $D_i = D_i(Z_i) = D_i(Z_1, \ldots, Z_n)$, $Y_i = Y_i(Z_i, D_i) = Y_i(Z_1, \ldots, Z_n, D_1, \ldots, D_n)$.

Under (A0), the potential outcome become $D_i(Z_i)$ and $Y_i(Z_i, D_i)$. Suppose that the samples (Z_i, D_i, X_i, Y_i) are independent and identically distributed (i.i.d.) as (Z, D, X, Y).

Based on the value of potential received treatment $(D(1), D(0))$, the population can be divided into four latent groups S, called principal strata (Frangakis & Rubin, 2002),

- Always-takers ($S = \mathrm{a}$): Units with $D(z) \equiv 1$ for all $z \in \{0, 1\}$, who systematically accept treatment regardless of assignment.
- Never-takers ($S = \mathrm{n}$): Units with $D(z) \equiv 0$ for all $z \in \{0, 1\}$, who persistently reject treatment irrespective of assignment.
- Compliers ($S = \mathrm{c}$): Units with $D(z) = z$ for all $z \in \{0, 1\}$, who adhere to assignment.
- Defiers ($S = \mathrm{d}$): Units with $D(z) = 1 - z$ for all $z \in \{0, 1\}$, who counteract assignment.

See Table 1. Note that the principal stratum is unidentifiable. This is because $D(1)$ and $D(0)$ cannot be observed simultaneously, and observed data cannot distinguish the principal strata due to the following overlaps: $(Z = 0, D = 0)$ could be either never-taker or complier, $(Z = 0, D = 1)$ could be either always-taker or defier, $(Z = 1, D = 0)$ could be either never-taker or defier, $(Z = 1, D = 1)$ could be either always-taker or complier. Among these four principal stratum, compliers play a crucial role in RCTs with noncompliance.

The objective of this paper is to make inferences about the conditional local risk ratio (CLRR) for compliers, i.e.,

$$R(x) = \frac{\mathbb{E}\{Y(1,1) \mid X = x, S = \mathrm{c}\}}{\mathbb{E}\{Y(0,0) \mid X = x, S = \mathrm{c}\}}, \tag{1}$$

which falls within $[0, \infty)$ when dealing with the binary outcome. To ensure the CLRR is well defined, in addition to (A0), we make another assumption:

(A1) $\mathbb{E}\{Y(0,0) \mid X = x, S = \mathrm{c}\} \neq 0$ almost surely.

Since the unidentification of compliance strata renders the CLRR unidentifiable, we adopt five additional assumptions from Abadie (2003) to identify it.

(A2) Exclusion Restriction: $Y(Z = 1, D = d) = Y(Z = 0, D = d) = Y(D = d)$ for all d.

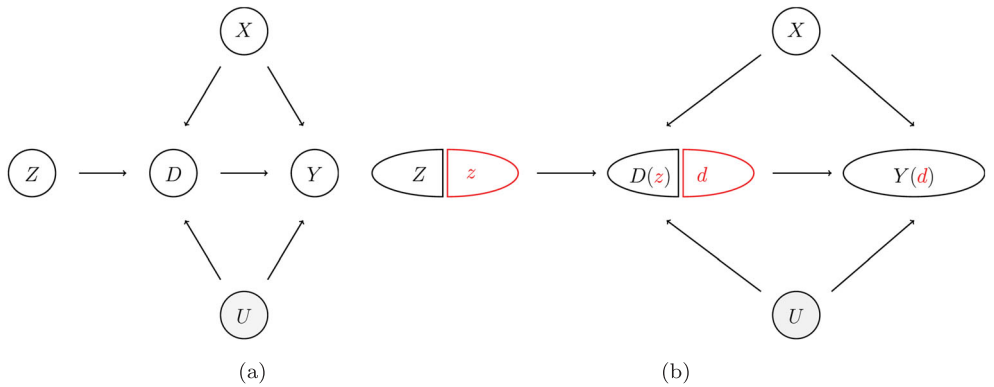


Figure 1. Illustration of an instrumental variable model using a causal graph. Variables *Z*, *D*, *X*, *Y* are observed; *U* is unobserved. The left panel gives a causal Directed Acyclic Graph (Pearl, 2009), and the right panel gives a Single World Intervention Graph (Richardson & Robins, 2013). (a) A Directed Acyclic Graph and (b) A Single World Intervention Graph.

(A3) Exogeneity: Z is independent of $\{D(0), D(1), Y(0), Y(1), X\}$.
(A4) Relevance: $\mathbb{E}\{D(1) - D(0) \mid X\} \neq 0$ almost surely.
(A5) Monotonicity: $\mathrm{p}\{D(1) \geq D(0) \mid X\} = 1$ almost surely.
(A6) Positivity: $0 < \mathrm{p}(Z = 1) < 1$ almost surely.

(A2)–(A4) represent the classical instrumental variable assumptions, and Z is an instrumental variable under these three assumptions. (A2) implies that Z isn't a direct cause of the outcome Y, and (A3) implies ignorability, which allows for direct comparability of D and Y corresponding to different Z. This means that the causal effects of Z on D or Y are identifiable. Note that (A3) naturally holds in randomized experiments. (A4) shows that Z has a non-zero causal effect on D, or equivalenlty $\mathrm{p}(S = \mathrm{c}|X = x) \neq 0$. The relationship among the variables Y, Z, D, X and the unobserved confounder U can be illustrated by a causal directed acyclic graph (Pearl, 2009) and a corresponding Single World Intervention Graph (Richardson & Robins, 2013); see Figure 1. What's more, (A6) naturally holds in RCTs, and (A4) is testable under (A6), since $\mathbb{E}\{D(1) - D(0)|X\}\mathrm{p}(Z = 1)\mathrm{p}(Z = 0) = \mathbb{C}\mathrm{ov}(Z, D|X)$. Next, (A5) is the most important assumption, which rules out the existence of defiers and is proposed to identify causal effects nonparametrically. Under Assumptions (A0)–(A3), the CLRR can be expressed as

$$R(x) = \frac{\mathrm{p}(Y = 1 \mid X = x, Z = 1, S = \mathrm{c})}{\mathrm{p}(Y = 1 \mid X = x, Z = 0, S = \mathrm{c})}, \tag{2}$$

which allows for nonparametric identification.

2.2. Identification and re-expression of CLRR

As disclosed by Abadie (2002), the CLRR for a binary outcome can be identified as

$$R(x) = \frac{\mathrm{p}(Y = 1, D = 1|X = x, Z = 1) - \mathrm{p}(Y = 1, D = 1|X = x, Z = 0)}{\mathrm{p}(Y = 1, D = 0|X = x, Z = 0) - \mathrm{p}(Y = 1, D = 0|X = x, Z = 1)}. \tag{3}$$

Clearly, the four terms $\mathrm{p}(Y = 1, D = d|X = x, Z = z)$ on the right-hand side of (3) are nonparametrically identifiable from observed data, and so is the CLRR. Unfortunately, estimates

of both the numerator and denominator of the right-hand side of (3) may be negative. This leads to negative CLRR estimates, which is inconsistent with the definition of CLRR as a nonnegative measure.

To overcome this problem, we consider an alternative expression of the CLRR in (2). We retrospectively decompose the numerator and the denominator of (2) as

$$\begin{aligned}
&\mathrm{p}(Y = y \mid X = x, Z = 1, S = \mathrm{c}) \\
&\quad = \frac{\mathrm{p}(Y = y|Z = 1, S = \mathrm{c})\mathrm{p}(x|Y = y, Z = 1, S = \mathrm{c})}{\mathrm{p}(x|Z = 1, S = \mathrm{c})}, \\
&\mathrm{p}(Y = y \mid X = x, Z = 0, S = \mathrm{c}) \\
&\quad = \frac{\mathrm{p}(Y = y|Z = 0, S = \mathrm{c})\mathrm{p}(x|Y = y, Z = 0, S = \mathrm{c})}{\mathrm{p}(x|Z = 0, S = \mathrm{c})}.
\end{aligned} \tag{4}$$

Under Assumption (A3), Z is independent of (S, X), and therefore, $\mathrm{p}(x|Z = 0, S = \mathrm{c}) = \mathrm{p}(x|Z = 1, S = \mathrm{c})$. It then follows that

$$R(x) = \frac{\mathrm{p}(x|Y = 1, Z = 1, S = \mathrm{c})}{\mathrm{p}(x|Y = 1, Z = 0, S = \mathrm{c})} \cdot \frac{\mathrm{p}(Y = 1|Z = 1, S = \mathrm{c})}{\mathrm{p}(Y = 1|Z = 0, S = \mathrm{c})}.$$

Let $p_{c1}(y) = \mathrm{p}(Y = y|S = \mathrm{c}, Z = 1)$, $p_{c0}(y) = \mathrm{p}(Y = y|S = \mathrm{c}, Z = 0)$, $g_{c1}(x|y) = \mathrm{p}(X = x|Y = y, S = \mathrm{c}, Z = 1)$ and $g_{c0}(x|y) = \mathrm{p}(X = x|Y = y, S = \mathrm{c}, Z = 0)$. The CLRR can be expressed as

$$R(x) = \frac{g_{c1}(x|1)}{g_{c0}(x|1)} \frac{p_{c1}(1)}{p_{c0}(1)}. \tag{5}$$

Lemma 2.1: *Under Assumptions (A0)–(A6), as functions of (x, y), the parameters $p_{c1}(y), p_{c0}(y), g_{c1}(x|y)$, and $g_{c0}(x|y)$ are all identifiable.*

A proof of Lemma 2.1 is provided in Appendix A.1. As $R(x)$ is uniquely determined by $p_{c1}(1), p_{c0}(1), g_{c1}(x|1)$, and $g_{c0}(x|1)$, Lemma 2.1 implies that $R(x)$ is identifiable. This also suggests that we can estimate $R(x)$ through estimating $p_{c1}(1), p_{c0}(1), g_{c1}(x|1)$, and $g_{c0}(x|1)$, or simply estimating $p_{c1}(1), p_{c0}(1)$, and $g_{c1}(x|1)/g_{c0}(x|1)$.

2.3. Semiparametric likelihood estimation

Recall that $\{(Z_i, D_i, X_i, Y_i)\}_{i=1}^n$ are i.i.d. observations from (Z, D, X, Y). Based on the identification results in the previous subsection, we propose a novel semiparametric likelihood estimation method called SPL. Our estimation procedure consists of two steps, which estimate $\{p_{c1}(1), p_{c0}(1)\}$ and $g_{c1}(x|1)/g_{c0}(x|1)$, respectively.

2.3.1. Step I of SPL

As the parameters $p_{c1}(1)$ and $p_{c0}(1)$ do not involve the covariates x, the first step of our method is built on the data $\{(Z_i, D_i, Y_i), i = 1, \ldots, n\}$. The likelihood function is

$$L_1 = \prod_{i=1}^{n} \Big[\big\{ \mathrm{p}(Z_i = 0) \cdot \mathrm{p}(Y_i, D_i = 1|Z_i = 0) \big\}^{I\{Z_i=0, D_i=1\}}$$

$$\times \left\{\mathrm{p}(Z_i = 1) \cdot \mathrm{p}(Y_i, D_i = 0 | Z_i = 1)\right\}^{I\{Z_i=1, D_i=0\}}$$
$$\times \left\{\mathrm{p}(Z_i = 0) \cdot \mathrm{p}(Y_i, D_i = 0 | Z_i = 0)\right\}^{I\{Z_i=0, D_i=0\}}$$
$$\times \left\{\mathrm{p}(Z_i = 1) \cdot \mathrm{p}(Y_i, D_i = 1 | Z_i = 1)\right\}^{I\{Z_i=1, D_i=1\}}\Big].$$

Let $\delta = \mathrm{p}(Z = 1), \phi_a = \mathrm{p}(S = \mathrm{a}), \phi_n = \mathrm{p}(S = \mathrm{n}), \phi_c = \mathrm{p}(S = \mathrm{c}) = 1 - \phi_a - \phi_n$, $p_a(y) = \mathrm{p}(Y = y | S = \mathrm{a}) = \mathrm{p}(Y = y | S = \mathrm{a}, Z = z)$, and $p_n(y) = \mathrm{p}(Y = y | S = \mathrm{n}) = \mathrm{p}(Y = y | S = \mathrm{n}, Z = z)$. It can be found that

$$\begin{aligned}
&\mathrm{p}(Y_i, D_i = 0 | Z_i = 0) = \phi_n p_n(Y_i) + \phi_c p_{c0}(Y_i),\\
&\mathrm{p}(Y_i, D_i = 1 | Z_i = 0) = \phi_a p_a(Y_i),\\
&\mathrm{p}(Y_i, D_i = 1 | Z_i = 1) = \phi_a p_a(Y_i) + \phi_c p_{c1}(Y_i),\\
&\mathrm{p}(Y_i, D_i = 0 | Z_i = 1) = \phi_n p_n(Y_i).
\end{aligned}$$

Therefore the likelihood L_1 can be expressed as

$$\begin{aligned}
L_1 = \prod_{i=1}^{n} \Big[&\left\{(1-\delta)\phi_a p_a(Y_i)\right\}^{I\{Z_i=0, D_i=1\}}\\
&\times \left\{(1-\delta)[\phi_n p_n(Y_i) + \phi_c p_{c0}(Y_i)]\right\}^{I\{Z_i=0, D_i=0\}} \times \left\{\delta \phi_n p_n(Y_i)\right\}^{I\{Z_i=1, D_i=0\}}\\
&\times \left\{\delta[\phi_a p_a(Y_i) + \phi_c p_{c1}(Y_i)]\right\}^{I\{Z_i=1, D_i=1\}}\Big].
\end{aligned}$$

In the meantime, we summarize the data $\{(Z_i, D_i, Y_i), i = 1, \ldots, n\}$ by defining $I_{z,i} = I\{Z_i = z\}, I_{zd,i} = I\{Z_i = z, D_i = d\}, I_{zdy,i} = I\{Z_i = z, D_i = d, Y_i = y\}, n_z = \sum_{i=1}^{n} I\{Z_i = z\}, n_{zd} = \sum_{i=1}^{n} I\{Z_i = z, D_i = d\}, n_{zdy} = \sum_{i=1}^{n} I\{Z_i = z, D_i = d, Y_i = y\}$. Let $\boldsymbol{\theta} = (\delta, \phi_a, \phi_n, p_a(1), p_n(1), p_{c0}(1), p_{c1}(1))$, which includes the important parameters $p_{c1}(1)$ and $p_{c0}(1)$ as components. The log-likelihood function in this step is

$$\begin{aligned}
\ell_1(\boldsymbol{\theta}) = \; & n_0 \log(1-\delta) + n_1 \log \delta + n_{01} \log \phi_a + n_{011} \log p_a(1) + n_{010} \log\{1 - p_a(1)\}\\
&+ n_{10} \log \phi_n + n_{101} \log p_n(1) + n_{100} \log\{1 - p_n(1)\}\\
&+ n_{001} \log\left[(1 - \phi_a - \phi_n) p_{c0}(1) + \phi_n p_n(1)\right]\\
&+ n_{000} \log\left[(1 - \phi_a - \phi_n)\{1 - p_{c0}(1)\} + \phi_n\{1 - p_n(1)\}\right]\\
&+ n_{111} \log\left[(1 - \phi_a - \phi_n) p_{c1}(1) + \phi_a p_a(1)\right]\\
&+ n_{110} \log\left[(1 - \phi_a - \phi_n)\{1 - p_{c1}(1)\} + \phi_a\{1 - p_a(1)\}\right]. \qquad (6)
\end{aligned}$$

The parameter $\boldsymbol{\theta}$ should belong to

$$\begin{aligned}
\boldsymbol{\Theta} = \{\boldsymbol{\theta} : &\delta \in (0,1), \phi_a, \phi_n \geqslant 0, 1 - \phi_a - \phi_n > 0,\\
&p_{c0}(1) \in (0,1],\ p_a(1), p_n(1), p_{c1}(1) \in [0,1]\Big\},
\end{aligned}$$

where the constraint $1 - \phi_a - \phi_n > 0$ follows directly from Assumption (A5), and $p_{c0}(1) \in (0,1]$ follows from Assumption (A1) and Equation (4). We propose to estimate $\boldsymbol{\theta}$ by its maximum likelihood estimator (MLE) $\widehat{\boldsymbol{\theta}} \equiv (\hat{\delta}, \hat{\phi}_a, \hat{\phi}_n, \hat{p}_a(1), \hat{p}_n(1), \hat{p}_{c0}(1), \hat{p}_{c1}(1)) = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \ell_1(\boldsymbol{\theta})$.

Remark 2.1: *The constraints in* $\boldsymbol{\Theta}$ *are essential to ensure the validity of the MLE. If we ignore these constraints, the MLE of* $\boldsymbol{\theta}$ *is* $\widetilde{\boldsymbol{\theta}} \equiv (\tilde{\delta}, \tilde{\phi}_a, \tilde{\phi}_n, \hat{p}_a(1), \tilde{p}_n(1), \tilde{p}_{c0}(1), \tilde{p}_{c1}(1))$, *where* $\tilde{\delta} = n_1/n, \tilde{\phi}_a = n_{01}/n_0, \tilde{\phi}_n = n_{10}/n_1, \tilde{\phi}_c = 1 - (n_{01}/n_0) - (n_{10}/n_1)$, *and*

$$\tilde{p}_a(y) = \frac{n_{01y}}{n_{01}}, \quad \tilde{p}_n(y) = \frac{n_{10y}}{n_{10}}, \quad \tilde{p}_{c1}(y) = \frac{(n_{11y}/n_1) - (n_{01y}/n_0)}{1 - (n_{01}/n_0) - (n_{10}/n_1)},$$
$$\tilde{p}_{c0}(y) = \frac{(n_{00y}/n_0) - (n_{10y}/n_1)}{1 - (n_{01}/n_0) - (n_{10}/n_1)}.$$

This corresponds to the moment estimator in Appendix A.1. However, the moment estimator $\widetilde{\boldsymbol{\theta}}$ *may be misleading because the estimates* $\tilde{\phi}_c$, $\tilde{p}_{c0}(y)$ *and* $\tilde{p}_{c1}(y)$ *may be negative or greater than 1, falling outside their ranges.*

2.3.2. Step II of SPL

It remains to estimate $g_{c1}(x|1)/g_{c0}(x|1)$. Define two conditional density functions

$$g_a(x|y) = \mathrm{p}(X = x|S = \mathrm{a}, Y = y) = \mathrm{p}(X = x|S = \mathrm{a}, Y = y, Z = z),$$
$$g_n(x|y) = \mathrm{p}(X = x|S = \mathrm{n}, Y = y) = \mathrm{p}(X = x|S = \mathrm{n}, Y = y, Z = z).$$

In Step II, we postulate three density ratio models

$$\frac{g_a(x|1)}{g_{c0}(x|1)} = \mathrm{e}^{\eta_1(x;\boldsymbol{\beta}_1)}, \quad \frac{g_n(x|1)}{g_{c0}(x|1)} = \mathrm{e}^{\eta_2(x;\boldsymbol{\beta}_2)}, \quad \frac{g_{c1}(x|1)}{g_{c0}(x|1)} = \mathrm{e}^{\eta_3(x;\boldsymbol{\beta}_3)}, \tag{7}$$

where $\eta_1(\cdot)$, $\eta_2(\cdot)$ and $\eta_3(\cdot)$ are user-specified functions. For example, we may choose them to be linear functions to facilitate computation and interpretation.

Step II of our SPL is built on the conditional likelihood

$$L_2 = \prod_{i:Y_i=1} \mathrm{p}(Z_i, D_i, X_i, Y_i = 1).$$

The following lemma plays an important role in expressing the conditional likelihood L_2 in terms of $\boldsymbol{\theta}$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \boldsymbol{\beta}_3^\top)^\top$.

Lemma 2.2: *Under Assumptions (A0)–(A6) and models (7), we have*

$$\mathrm{p}(Z = 0, D = 0, X = x, Y = 1) = (1 - \delta) g_{c0}(x|1) \left\{ \phi_c p_{c0}(1) + \phi_n p_n(1) \mathrm{e}^{\eta_2(x;\boldsymbol{\beta}_2)} \right\},$$
$$\mathrm{p}(Z = 0, D = 1, X = x, Y = 1) = (1 - \delta) g_{c0}(x|1) \phi_a p_a(1) \mathrm{e}^{\eta_1(x;\boldsymbol{\beta}_1)},$$
$$\mathrm{p}(Z = 1, D = 0, X = x, Y = 1) = \delta g_{c0}(x|1) \phi_n p_n(1) \mathrm{e}^{\eta_2(x;\boldsymbol{\beta}_2)},$$
$$\mathrm{p}(Z = 1, D = 1, X = x, Y = 1) = \delta g_{c0}(x|1) \left\{ \phi_c p_{c1}(1) \mathrm{e}^{\eta_3(x;\boldsymbol{\beta}_3)} + \phi_a p_a(1) \mathrm{e}^{\eta_1(x;\boldsymbol{\beta}_1)} \right\}.$$

With the preparations in Lemma 2.2, we can express the logarithm of L_2 as

$$\ell_2(\boldsymbol{\beta}, g_{c0}|\boldsymbol{\theta}) = \ell_{21}(\boldsymbol{\beta}, g_{c0}|\boldsymbol{\theta}) + \ell_{22}(\boldsymbol{\theta}),$$

where

$$\ell_{21}(\boldsymbol{\beta}, g_{c0}|\boldsymbol{\theta}) = \sum_{i=1}^{n} I\{Y_i = 1\} \Big[\log\{g_{c0}(X_i|1) dX_i\}$$

$$+ I_{01,i}\eta_1(X_i;\boldsymbol{\beta}_1) + I_{10,i}\eta_2(X_i;\boldsymbol{\beta}_2)$$
$$+ I_{00,i}\log\left\{\phi_c p_{c0}(1) + \phi_n p_n(1)\mathrm{e}^{\eta_2(X_i;\boldsymbol{\beta}_2)}\right\}$$
$$+ I_{11,i}\log\left\{\phi_c p_{c1}(1)\mathrm{e}^{\eta_3(X_i;\boldsymbol{\beta}_3)} + \phi_a p_a(1)\mathrm{e}^{\eta_1(X_i;\boldsymbol{\beta}_1)}\right\}\Big],$$

$$\ell_{22}(\boldsymbol{\theta}) = \sum_{i=1}^{n} I\{Y_i = 1\}\big[-\log(dX_i) + I_{01,i}\log(1-\delta)\phi_a p_a(1)$$
$$+ I_{10,i}\log\{\delta\phi_n p_n(1)\} + I_{00,i}\log\{(1-\delta)\} + I_{11,i}\log\delta\big].$$

In this step, we propose to replace $\boldsymbol{\theta}$ in $\ell_2(\boldsymbol{\beta}, g_{c0}|\boldsymbol{\theta})$ by $\widehat{\boldsymbol{\theta}}$ and estimate $\boldsymbol{\beta}$ by maximizing $\ell_2(\boldsymbol{\beta}, g_{c0}|\widehat{\boldsymbol{\theta}})$ or equivalently maximizing $\ell_{21}(\boldsymbol{\beta}, g_{c0}|\widehat{\boldsymbol{\theta}})$ because $\ell_{22}(\widehat{\boldsymbol{\theta}})$ does not depend on $\boldsymbol{\beta}$. Directly maximizing $\ell_{21}(\boldsymbol{\beta}, g_{c0}|\widehat{\boldsymbol{\theta}})$ with respect to $(\boldsymbol{\beta}, g_{c0})$ is questionable because g_{c0} is an infinite-dimensional parameter.

In the expression of $\ell_{21}(\boldsymbol{\beta}|\widehat{\boldsymbol{\theta}})$, g_{c0} is related with all observations X_i such that $Y_i = 1$. As g_{c0} is a density function, we handle it with the empirical likelihood method. Let $G_{c0}(x|1)$ denote the cumulative probability distribution function corresponding to $g_{c0}(x|1)$. Then $dG_{c0}(X_i|1) = g_{c0}(X_i|1)dX_i$. In the principle of empirical likelihood, we model $G_{c0}(x|1)$ by a step function $G_{c0}(x|1) = \sum_{i=1}^{n} I(Y_i = 1)w_i I(X_i \le x)$. Because $G_{c0}(x|1)$ is a distribution function, this together with (7) implies that $\boldsymbol{w} = (w_1, \ldots, w_n)$ should belong to

$$\mathcal{W}(\boldsymbol{\beta}) = \left\{ \boldsymbol{w} \,\middle|\, w_i \ge 0, \sum_{i=1}^{n} I(Y_i = 1)w_i = 1, \right.$$
$$\left. \sum_{i=1}^{n} I(Y_i = 1)w_i \mathrm{e}^{\eta_j(X_i;\boldsymbol{\beta}_j)} = 1, j = 1,2,3 \right\}.$$

Given $\boldsymbol{\beta}$, after $g_{c0}(X_i|1)dX_i = dG_{c0}(X_i|1)$ are replaced with w_i, the likelihood $\ell_{21}(\boldsymbol{\beta}, g_{c0}|\boldsymbol{\theta})$ is maximized with respect to $\boldsymbol{w}$ in $\mathcal{W}(\boldsymbol{\beta})$ when $w_i = I\{Y_i = 1\}/[m\{1 + \sum_{j=1}^{3}\lambda_j(\mathrm{e}^{\eta_j(X_i;\boldsymbol{\beta}_j)} - 1)\}]$, where $m = \sum_{i=1}^{n} Y_i$ and $(\lambda_1, \lambda_2, \lambda_3) = (\lambda_1(\boldsymbol{\beta}), \lambda_2(\boldsymbol{\beta}), \lambda_3(\boldsymbol{\beta}))$ is the solution to the following equation arrays

$$\frac{1}{m}\sum_{i=1}^{n}\frac{I\{Y_i = 1\}\{\mathrm{e}^{\eta_k(X_i;\boldsymbol{\beta}_k)} - 1\}}{1 + \sum_{j=1}^{3}\lambda_j\{\mathrm{e}^{\eta_j(X_i;\boldsymbol{\beta}_j)} - 1\}} = 0, \quad k = 1,2,3. \tag{8}$$

Therefore, the profiled log-likelihood (up to a constant not depending on $\boldsymbol{\beta}$) of $\boldsymbol{\beta}$ is

$$\ell_{21}(\boldsymbol{\beta}|\boldsymbol{\theta}) = \sum_{i=1}^{n} I\{Y_i = 1\}\left[-\log\left\{1 + \sum_{j=1}^{3}\lambda_j(\mathrm{e}^{\eta_j(X_i;\boldsymbol{\beta}_j)} - 1)\right\} + I_{01,i}\eta_1(X_i;\boldsymbol{\beta}_1)\right.$$
$$+ I_{10,i}\eta_2(X_i;\boldsymbol{\beta}_2) + I_{00,i}\log\left(\phi_c p_{c0} + \phi_n p_n \mathrm{e}^{\eta_2(X_i;\boldsymbol{\beta}_2)}\right)$$
$$\left. + I_{11,i}\log\left(\phi_c p_{c1}\mathrm{e}^{\eta_3(X_i;\boldsymbol{\beta}_3)} + \phi_a p_a \mathrm{e}^{\eta_1(X_i;\boldsymbol{\beta}_1)}\right)\right]. \tag{9}$$

We propose to estimate $\boldsymbol{\beta}$ by its MLE $\widehat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \ell_{21}(\boldsymbol{\beta}|\widehat{\boldsymbol{\theta}})$. Finally, by substituting the quantities in (5) with the estimators $(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\beta}})$, the proposed estimator of CLRR $R(x)$ is $\widehat{R}(x) = \exp\{\eta_3(x;\widehat{\boldsymbol{\beta}}_3)\} \cdot \hat{p}_{c1}(1)/\hat{p}_{c0}(1)$. Obviously, this estimator always lies in the range $[0, \infty)$.

Because of the presence of a two-component mixture structure in the likelihood $\ell_{21}(\boldsymbol{\beta}|\widehat{\boldsymbol{\theta}})$, its maximization is challenging. We overcome this challenge via an EM algorithm.

2.3.3. EM algorithm

Let O denote the observed data and S_i denote the principle strata that individual i belongs to. In our EM algorithm, we take O as missing data and take $O \cup \{S_1, \ldots, S_n\}$ as complete data. Based on the complete data, the profile log-likelihood is ℓ_{21}, that is, Equation (9).

To account for latent principal strata, we propose an EM algorithm for optimization. Assume that $\eta_k(x)$, for $k = 1, 2, 3$, each contains a constant term. This is a mild assumption that is commonly met in practice and allows for a closed-form expression of $\boldsymbol{\lambda}$. Let $\boldsymbol{\beta}^{(0)}$ be an initial value of $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^{(k)} = (\boldsymbol{\beta}_1^{(k)}, \boldsymbol{\beta}_2^{(k)}, \boldsymbol{\beta}_3^{(k)})$, be the parameter value in the kth round of EM algorithm.

In the E-step of the kth iteration of our EM algorithm, we calculate the conditional expectation of ℓ_{21} given the observed data O with $Y = 1$ and the parameter value $\boldsymbol{\beta}^{(k-1)}$ in the previous iteration. It can be seen that

$$
\begin{aligned}
&\mathbb{E}(\ell_{21}|O_i, Y_i = 1; \boldsymbol{\beta}^{(k-1)}) \\
&= \sum_{i=1}^{n} I\{Y_i = 1\} \left[-\log \left\{ 1 + \sum_{j=1}^{3} \lambda_j^{(k)} (\mathrm{e}^{\eta_j(X_i;\boldsymbol{\beta}_j)} - 1) \right\} \right. \\
&\quad + I_{01,i}\eta_1(X_i;\boldsymbol{\beta}_1) + I_{10,i}\eta_2(X_i;\boldsymbol{\beta}_2) \\
&\quad + I_{00,i} \left\{ w_{1,i}^{(k)} \log \hat{\phi}_c \hat{p}_{c0}(1) + \left(1 - w_{1,i}^{(k)}\right) \log \hat{\phi}_n \hat{p}_n(1) \mathrm{e}^{\eta_2(X_i;\boldsymbol{\beta}_2)} \right\} \\
&\quad \left. + I_{11,i} \left\{ w_{2,i}^{(k)} \log \hat{\phi}_c \hat{p}_{c1}(1) \mathrm{e}^{\eta_3(X_i;\boldsymbol{\beta}_3)} + \left(1 - w_{2,i}^{(k)}\right) \log \hat{\phi}_a \hat{p}_a(1) \mathrm{e}^{\eta_1(X_i;\boldsymbol{\beta}_1)} \right\} \right] \\
&\propto \sum_{i=1}^{n} I\{Y_i = 1\} \left[-\log \left\{ 1 + \sum_{j=1}^{3} \lambda_j^{(k)} (\mathrm{e}^{\eta_j(X_i;\boldsymbol{\beta}_j)} - 1) \right\} \right. \\
&\quad + I_{11,i} w_2^{(k)} \eta_3(X_i;\boldsymbol{\beta}_3) + \left\{ I_{01,i} + I_{11,i} \left(1 - w_{2,i}^{(k)}\right) \right\} \eta_1(X_i;\boldsymbol{\beta}_1) \\
&\quad \left. + \left\{ I_{10,i} + I_{00,i} \left(1 - w_{1,i}^{(k)}\right) \right\} \eta_2(X_i;\boldsymbol{\beta}_2) \right], \qquad (10)
\end{aligned}
$$

where

$$
\begin{aligned}
w_{1,i}^{(k)} &= \mathrm{p}(S_i = c | Z_i = 0, D_i = 0, Y_i = 1, X_i) \\
&= \frac{\hat{\phi}_c \hat{p}_{c0}(1)}{\hat{\phi}_c \hat{p}_{c0}(1) + \hat{\phi}_n \hat{p}_n(1) \mathrm{e}^{\eta_2(X_i;\boldsymbol{\beta}_2^{(k-1)})}}, \\
w_{2,i}^{(k)} &= \mathrm{p}(S_i = c | Z_i = 1, D_i = 1, Y_i = 1, X_i) \\
&= \frac{\hat{\phi}_c \hat{p}_{c1}(1) \mathrm{e}^{\eta_3(X_i;\boldsymbol{\beta}_3^{(k-1)})}}{\hat{\phi}_c \hat{p}_{c1}(1) \mathrm{e}^{\eta_3(X_i;\boldsymbol{\beta}_3^{(k-1)})} + \hat{\phi}_a \hat{p}_a(1) \mathrm{e}^{\eta_1(X_i;\boldsymbol{\beta}_1^{(k-1)})}}.
\end{aligned}
$$

and

$$\lambda_1^{(k)} = \frac{\sum_{i=1}^n \left\{ I_{011,i} + I_{111,i} \left(1 - w_{2,i}^{(k)}\right)\right\}}{m},$$
$$\lambda_2^{(k)} = \frac{\sum_{i=1}^n \left\{ I_{101,i} + I_{001,i} \left(1 - w_{1,i}^{(k)}\right)\right\}}{m},$$
$$\lambda_3^{(k)} = \frac{\sum_{i=1}^n I_{111,i} w_{2,i}^{(k)}}{m},$$

with $m = \sum_{i=1}^n I\{Y_i = 1\}$.

Given the initial value of $\boldsymbol{\beta}^{(k-1)}$, we can obtain $\boldsymbol{w}_1^{(k)} = (w_{1,1}^{(k)}, \ldots, w_{1,n}^{(k)})$, $\boldsymbol{w}_2^{(k)} = (w_{2,1}^{(k)}, \ldots, w_{2,n}^{(k)})$, and $\boldsymbol{\lambda}^{(k)} = (\lambda_1^{(k)}, \lambda_2^{(k)}, \lambda_3^{(k)})$. This allows us to compute the conditional expectation (10). The M-step in the kth iteration of our EM algorithm is to calculate

$$\boldsymbol{\beta}^{(k)} = \arg\max_{\boldsymbol{\beta}} \mathbb{E}\left\{ \ell_{21} | O, S_i, Y_i = 1; \boldsymbol{\beta}^{(k-1)} \right\}, \tag{11}$$

and thus we get the kth estimator $\boldsymbol{\beta}^{(k)}$. The process is iterated until convergence.

Algorithm 1: *EM algorithm to calculate the maximizer of* ℓ_{21}

Data: $\{(Z_i, D_i, X_i, Y_i = 1) : 1, \ldots, n\}$

Input: Working models (7), an estimate $\widehat{\boldsymbol{\theta}}$, and an initial value $\boldsymbol{\beta}^{(0)}$ for $\boldsymbol{\beta}$

- E-step: Calculate the conditional expectation in (10);
- M-step: Calculate (11).
- Repeat the above two steps until convergence.

Output: The final parameter value $\boldsymbol{\beta}^{(k)}$ is the MLE of $\boldsymbol{\beta}$.

3. Large-sample properties

We assume that $\eta_3(x; \boldsymbol{\beta}_3)$ is smooth enough in $\boldsymbol{\beta}_3$. Then the proposed CLRR estimator $\widehat{R}(x)$ is a smooth function of the MLEs $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\beta}}$. To study its large-sample properties, it suffices to investigate the large-sample properties of $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\beta}}$.

For an observation $O = (Z, D, X, Y)$, let $I_z = I(Z = z)$, $I_{zd} = I(Z = z, D = d)$ and $I_{zdy} = I(Z = z, D = d, Y = y)$. Define

$$\begin{aligned} f(O; \boldsymbol{\theta}) = {} & I_0 \log(1 - \delta) + I_1 \log \delta \\ & + I_{01} \log \phi_a + I_{011} \log p_a(1) + I_{010} \log\{1 - p_a(1)\} \\ & + I_{10} \log \phi_n + I_{101} \log p_n(1) + I_{100} \log\{1 - p_n(1)\} \\ & + I_{000} \log\left[(1 - \phi_a - \phi_n)\{1 - p_{c0}(1)\} + \phi_n\{1 - p_n(1)\}\right] \\ & + I_{001} \log\left[(1 - \phi_a - \phi_n) p_{c0}(1) + \phi_n p_n(1)\right] \\ & + I_{110} \log\left[(1 - \phi_a - \phi_n)\{1 - p_{c1}(1)\} + \phi_a\{1 - p_a(1)\}\right] \\ & + I_{111} \log\left[(1 - \phi_a - \phi_n) p_{c1}(1) + \phi_a p_a(1)\right]. \end{aligned}$$

The true parameter value of $\boldsymbol{\theta}$ is $\boldsymbol{\theta}_0 = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Pf(O;\boldsymbol{\theta})$, where P is the probability measure of O. Let $O_i = (Z_i, D_i, X_i, Y_i)$, $i = 1, \ldots, n$, and $\mathbb{P}_n$ denote their empirical measure. Then the MLE $\widehat{\boldsymbol{\theta}}$ can be expressed as $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathbb{P}_n f(O;\boldsymbol{\theta})$.

Theorem 3.1: *Suppose that $\boldsymbol{\theta}_0$ is an interior of the parameter space $\boldsymbol{\Theta}$, and Assumptions* (A0)–(A6) *are satisfied. Then as $n \to \infty$,*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, V^{-1}),$$

where $V = -\mathbb{E}\{\ddot{f}(O;\boldsymbol{\theta}_0)\}$ and $\ddot{f}(O;\boldsymbol{\theta})$ denotes the second-order partial derivative of $f(O;\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$.

Then we examine the large-sample properties of $\widehat{\boldsymbol{\beta}} = (\widehat{\boldsymbol{\beta}}_1^\top, \widehat{\boldsymbol{\beta}}_2^\top, \widehat{\boldsymbol{\beta}}_3^\top)^\top$ and $\widehat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3)$. Let $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)$ and define

$$\begin{aligned} h(O;\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{\theta}) = {} & I_{01}\eta_1(X;\boldsymbol{\beta}_1) + I_{10}\eta_2(X;\boldsymbol{\beta}_2) \\ & + I_{11}\log\left\{\phi_c p_{c1} e^{\eta_3(X;\boldsymbol{\beta}_3)} + \phi_a p_a e^{\eta_1(X;\boldsymbol{\beta}_1)}\right\} \\ & + I_{00}\log\left\{\phi_c p_{c0} + \phi_n p_n e^{\eta_2(X;\boldsymbol{\beta}_2)}\right\} \\ & - \log\left\{1 + \sum_{j=1}^{3} \lambda_j (e^{\eta_j(X;\boldsymbol{\beta}_j)} - 1)\right\}. \end{aligned}$$

We use $\dot{h}_{\boldsymbol{\beta}}(O;\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{\theta})$ to denote the partial derivative of h with respect to $\boldsymbol{\beta}$ and define $\dot{h}_{\boldsymbol{\theta}}(O;\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{\theta})$ and $\dot{h}_{\boldsymbol{\lambda}}(O;\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{\theta})$ similarly. Define $\dot{h} = (\dot{h}_{\boldsymbol{\beta}}^\top, \dot{h}_{\boldsymbol{\lambda}}^\top)^\top$. Then $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\lambda}})$ is the solution to $\mathbb{P}_n I\{Y = 1\}\dot{h}(O;\boldsymbol{\beta},\boldsymbol{\lambda},\widehat{\boldsymbol{\theta}}) = 0$, and the true value $(\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0)$ is the solution to $PI\{Y = 1\}\dot{h}(O;\boldsymbol{\beta},\boldsymbol{\lambda},\widehat{\boldsymbol{\theta}}) = 0$.

Let $\ddot{h}_{\boldsymbol{\beta\theta}}(O;\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{\theta})$ denote the second-order partial derivative of h with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, and define other second-order partial derivatives in a similar way. We make the following assumptions.

(C1) (i)$(\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0)$ is the unique solution to $PI\{Y = 1\}\dot{h}(O;\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{\theta}_0) = 0$. (ii)The parameter space $\boldsymbol{B}$ of $\boldsymbol{\beta}$ is compact. Take $\boldsymbol{\Lambda} = [0,1] \times [0,1] \times [0,1]$ to be the range of $\boldsymbol{\lambda}$.

(C2) (i) $G_{c0}(x|1)$ is non-degenerate, (ii) η_1, η_2 and η_3 have continuous second-order derivatives with respect to $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ and $\boldsymbol{\beta}_3$, respectively. (iii) There exists a positive function $K_1(x)$ satisfying $\mathbb{E}\{K_1(X)\} < \infty$ such that $\|\dot{\eta}_1(X;\boldsymbol{\beta}_1)\|_2^2$, $\|\dot{\eta}_2(X;\boldsymbol{\beta}_2)\|_2^2$, $\|\dot{\eta}_3(X;\boldsymbol{\beta}_3)\|_2^2$, $\|\ddot{\eta}_1(X;\boldsymbol{\beta}_1)|$, $\|\ddot{\eta}_2(X;\boldsymbol{\beta}_2)\|$ and $\|\ddot{\eta}_3(X;\boldsymbol{\beta}_3)\|$ are all controlled by $K_1(X)$ for all $\boldsymbol{\beta} \in \boldsymbol{B}$.

(C3) The matrix $\Sigma_{11} = -\mathbb{E}\{I(Y = 1)\ddot{h}_{\boldsymbol{\beta\beta}}(O;\boldsymbol{\beta}_0,\boldsymbol{\lambda}_0,\boldsymbol{\theta}_0)\}$ is positive definite.

Define

$$Q_1 = P\begin{pmatrix} I(Y=1)\ddot{h}_{\boldsymbol{\beta\beta}}(O;\boldsymbol{\beta}_0,\boldsymbol{\lambda}_0,\boldsymbol{\theta}_0) & I(Y=1)\ddot{h}_{\boldsymbol{\beta\lambda}}(O;\boldsymbol{\beta}_0,\boldsymbol{\lambda}_0,\boldsymbol{\theta}_0) \\ I(Y=1)\ddot{h}_{\boldsymbol{\lambda\beta}}(O;\boldsymbol{\beta}_0,\boldsymbol{\lambda}_0,\boldsymbol{\theta}_0) & I(Y=1)\ddot{h}_{\boldsymbol{\lambda\lambda}}(O;\boldsymbol{\beta}_0,\boldsymbol{\lambda}_0,\boldsymbol{\theta}_0) \end{pmatrix},$$

$$Q_2 = P\begin{pmatrix} I(Y=1)\ddot{h}_{\boldsymbol{\beta\theta}}(O;\boldsymbol{\beta}_0,\boldsymbol{\lambda}_0,\boldsymbol{\theta}_0) \\ I(Y=1)\ddot{h}_{\boldsymbol{\lambda\theta}}(O;\boldsymbol{\beta}_0,\boldsymbol{\lambda}_0,\boldsymbol{\theta}_0) \end{pmatrix}.$$

We use d_k to denote the dimension of $\boldsymbol{\beta}_k$ for $k = 1, 2, 3$, and define $I_5 = (0, 0, 0, 0, 0, 1, 0)$, $I_6 = (0, 0, 0, 0, 0, 0, 1)$, and

$$I_{d_3} = (\underbrace{0, \ldots, 0}_{d_1}, \underbrace{0, \ldots, 0}_{d_2}, \underbrace{1, \ldots, 1}_{d_3}, 0, 0, 0).$$

Theorem 3.2: *Suppose that $\boldsymbol{\theta}_0$ is an interior of the parameter space $\boldsymbol{\Theta}$, and Assumptions (A0)–(A6) and (C1)–(C3) are satisfied. Then as $n \to \infty$, we have*

$$\sqrt{n}\begin{pmatrix}\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_0\end{pmatrix} \xrightarrow{d} N(0, \Sigma),$$

where $\Sigma = Q_1^{-1} W (Q_1^{-1})^\top$ and $W = \mathbb{V}\mathrm{ar}\{I(Y = 1)\dot{h}(O; \boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \boldsymbol{\theta}_0) + Q_2 V^{-1} \dot{f}(O; \boldsymbol{\theta}_0)\}$.

With the above results, we show that our final estimator $\widehat{R}(x)$ is also asymptotically normal.

Theorem 3.3: *Suppose that $\boldsymbol{\theta}_0$ is an interior of the parameter space $\boldsymbol{\Theta}$, and Assumptions (A0)–(A6) and (C1)–(C3) are satisfied. Then for each fixed x, as $n \to \infty$, it holds that $\sqrt{n}\{\widehat{R}(x) - R_0(x)\} \to N(0, \mathbb{V}\mathrm{ar}(M(x))$, where*

$$\begin{aligned} M(x) = & \frac{\exp\{\eta_3(x; \boldsymbol{\beta}_{3,0})\}}{p_{c0,0}(1)} \left(I_6 - \frac{p_{c1,0}(1)}{p_{c0,0}(1)} I_5\right)^\top V^{-1} \dot{f}(O; \boldsymbol{\theta}_0) \\ & - \dot{\eta}_3(x; \boldsymbol{\beta}_{3,0}) \exp\{\eta_3(x; \boldsymbol{\beta}_{3,0})\} \\ & \times \frac{p_{c1,0}(1)}{p_{c0,0}(1)} I_{d_3}^\top Q_1^{-1} \left\{I(Y = 1)\dot{h}(O; \boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \boldsymbol{\theta}_0) + Q_2 V^{-1} \dot{f}(O; \boldsymbol{\theta}_0)\right\}. \end{aligned}$$

In addition, if the support $\mathcal{X}$ of X is compact, then as $n \to \infty$, $\sup_{x \in \mathcal{X}} |\widehat{R}(x) - R_0(x)| = o_p(1)$.

To make inference about $\widehat{\boldsymbol{\beta}}$ and $\widehat{R}(x)$ based on Theorems 3.2 and 3.3, it is necessary to construct consistent estimates for the asymptotic variances in the two theorems. However, the asymptotic variance of $\widehat{\boldsymbol{\beta}}$ and $\widehat{R}(x)$ both exhibit very complex structures, making their direct estimation formidable. To overcome this problem, we recommend using the usual nonparametric bootstrap method to estimate the asymptotic variance of $\widehat{\boldsymbol{\beta}}$ and $\widehat{R}(x)$.

4. Simulation

4.1. Settings

We conduct simulations to examine the finite-sample performance of the proposed SPL estimation procedure. For comparison, we also consider the following competitors: Abadie (2003)'s least square estimation method (LSE), Wang et al. (2021)'s maximum likelihood method (MLE), Wang et al. (2021)'s doubly robust estimator with optimal weighting function (DRW), Wang et al. (2021)'s doubly robust estimator with identity weighting function (DRU), and Richardson et al. (2017)'s maximum likelihood method ignoring the information of D (ITT).

While the SPL framework was originally developed through retrospective analysis, our simulation study adopts a prospective approach to data generation. This method not only

aligns with the inherent causal mechanisms of the system but also enables more straightforward implementation. We generate a univariate covariate $\widetilde{X}$ from the standard normal distribution and set $X = (1, \widetilde{X})$. In our data generating process (DGP), we choose

$$\begin{aligned}
\phi_c(x) &= \mathrm{p}(S = \mathrm{c}|X = x) = \{1 + \exp(\boldsymbol{\alpha}_a^\top x) + \exp(\boldsymbol{\alpha}_n^\top x)\}^{-1},\\
\phi_a(x) &= \mathrm{p}(S = \mathrm{a}|X = x) = \exp(\boldsymbol{\alpha}_a^\top x)\phi_c(x),\\
\phi_n(x) &= \mathrm{p}(S = \mathrm{n}|X = x) = \exp(\boldsymbol{\alpha}_n^\top x)\phi_c(x),\\
p_{c0}(1|x) &= \mathrm{p}(Y = 1|S = \mathrm{c}, Z = 0, X = x) = \pi(-\boldsymbol{\gamma}_c^\top x),\\
p_{c1}(1|x) &= \mathrm{p}(Y = 1|S = \mathrm{c}, Z = 1, X = x) = \pi(\boldsymbol{\gamma}_c^\top x),\\
p_a(1|x) &= \mathrm{p}(Y = 1|S = \mathrm{a}, X = x) = \pi(\boldsymbol{\gamma}_a^\top x),\\
p_n(1|x) &= \mathrm{p}(Y = 1|S = \mathrm{n}, X = x) = \pi(\boldsymbol{\gamma}_n^\top x),
\end{aligned}$$

where $\pi(t) = \mathrm{e}^t/(1 + \mathrm{e}^t)$. We set $\delta = 0.5$, and cosider two settings

- $(\boldsymbol{\alpha}_a, \boldsymbol{\alpha}_n) = (-3, -0.5, -2, 0.1)$, and $\boldsymbol{\gamma}_a = \boldsymbol{\gamma}_n = \boldsymbol{\gamma}_c = (-3, 1)$,
- $(\boldsymbol{\alpha}_a, \boldsymbol{\alpha}_n) = (-1, -1, -0.8, 0.2)$, and $\boldsymbol{\gamma}_a = \boldsymbol{\gamma}_n = \boldsymbol{\gamma}_c = (1, -1)$.

Under this DGP,

$$\begin{aligned}
g_a(x|1)/g_{c0}(x|1) &= \exp\left\{(\boldsymbol{\alpha}_a - \boldsymbol{\gamma}_a)^\top x + \log[\{\phi_c p_{c0}(1)\}/\{\phi_a p_a(1)\}]\right\},\\
g_n(x|1)/g_{c0}(x|1) &= \exp\left\{(\boldsymbol{\alpha}_n - \boldsymbol{\gamma}_n)^\top x + \log[\{\phi_c p_{c0}(1)\}/\{\phi_n p_n(1)\}]\right\},\\
g_{c1}(x|1)/g_{c0}(x|1) &= \exp\left[\boldsymbol{\gamma}_c^\top x + \log\{p_{c0}(1)/p_{c1}(1)\}\right].
\end{aligned}$$

Because $\boldsymbol{\gamma}_a = \boldsymbol{\gamma}_n = \boldsymbol{\gamma}_c$, it follows that $\eta_1(x; \boldsymbol{\beta}_1) = (\boldsymbol{\alpha}_a - \boldsymbol{\gamma}_a)^\top x + \log(\phi_c/\phi_a) - \boldsymbol{\gamma}_a^\top x$, $\eta_2(x; \boldsymbol{\beta}_1) = (\boldsymbol{\alpha}_n - \boldsymbol{\gamma}_a)^\top x - \boldsymbol{\gamma}_a^\top x$, and $\eta_3(x; \boldsymbol{\beta}_1) = \boldsymbol{\gamma}_c^\top x + \log(p_{c0}(1)/p_{c1}(1))$. The quantity of interest is the CLRR

$$R(x) = \frac{g_{c1}(x|1)}{g_{c0}(x|1)}\frac{p_{c1}(1)}{p_{c0}(1)} = \exp\left\{\boldsymbol{\beta}_3^\top x + \log\frac{p_{c1}(1)}{p_{c0}(1)}\right\} = \exp(\boldsymbol{\gamma}_c^\top x),$$

where $\boldsymbol{\beta}_3 = (\boldsymbol{\beta}_{31}, \boldsymbol{\beta}_{32})^\top$ and $\boldsymbol{\gamma}_c = (\gamma_1, \gamma_2) = (\log(p_{c1}(1)/p_{c0}(1)) + \boldsymbol{\beta}_{31}, \boldsymbol{\beta}_{32})^\top$. We choose $\eta_k(x, \boldsymbol{\beta}_k)$ $(k = 1, 2, 3)$ to be linear functions in the proposed SPL method, and compare it with its competitors through their finite-sample performances in the estimation of $R(x)$ and $\boldsymbol{\gamma}_c$.

We conducted $L = 500$ Monte Carlo simulations, each with sample sizes $n = 500$ and 1000, across different simulation settings. First, we take simulated biases (Bias), scaled biases ($\sqrt{n}|\mathrm{Bias}|$), and root mean squared errors (RMSE) based on 500 Monte Carlo as criteria to evaluate an estimator of $\boldsymbol{\gamma}_c = (\gamma_1, \gamma_2)$. Second, we also investigate the estimation results for the CLRR curve $R(x)$. For a generic estimator $\widehat{R}(x)$, we evaluate its accuracy using the integrated absolute error (IAE),

$$\mathrm{IAE}(\widehat{R}) = \frac{1}{1001}\sum_{i=1}^{1000} |\log\widehat{R}(t_i) - \log R(t_i)|,$$

Table 2. Simulation results for the estimation of $\boldsymbol{\gamma}_c = (\gamma_1, \gamma_2)$ when $X \sim N(0,1)$.

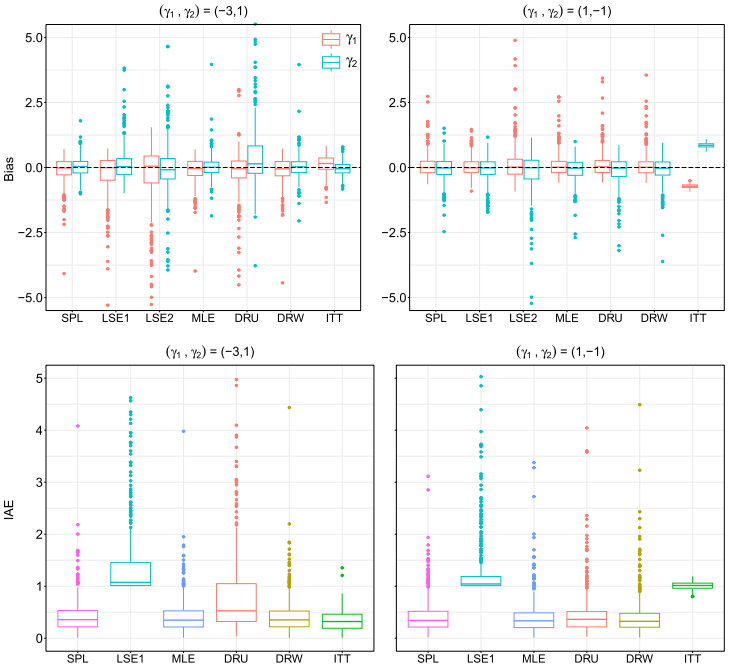
		$\widehat{\gamma}_1$			$\widehat{\gamma}_2$			
$\boldsymbol{\gamma}_c$	Methods	Bias	$\sqrt{n}\lvert\text{Bias}\rvert$	RMSE	Bias	$\sqrt{n}\lvert\text{Bias}\rvert$	RMSE	IAE
$n = 500$								
(−3,1)	SPL	−0.082	2.601	0.473	0.040	1.260	0.361	0.427
	LSE1	−3.730	117.957	36.317	1.641	51.896	16.167	8.044
	LSE2	−10.362	327.662	88.226	1.961	62.022	34.801	–
	MLE	−0.101	3.178	0.554	0.024	0.771	0.390	0.431
	DRU	−0.493	15.598	3.522	3.256	102.959	22.385	3.748
	DRW	−0.107	3.386	0.567	0.033	1.054	0.404	0.438
	ITT	0.132	4.168	0.353	−0.042	1.319	0.249	0.346
(1,−1)	SPL	0.064	2.023	0.404	−0.061	1.915	0.443	0.426
	LSE1	0.028	0.876	0.347	−0.064	2.018	0.417	3.164
	LSE2	0.695	21.992	7.702	−0.784	24.808	7.314	–
	MLE	0.078	2.473	0.424	−0.086	2.708	0.420	0.404
	DRU	0.121	3.842	0.695	−0.129	4.074	0.572	0.479
	DRW	0.077	2.442	0.443	−0.084	2.668	0.451	0.413
	ITT	−0.722	22.847	0.726	0.851	26.916	0.856	1.009
$n = 1000$								
(−3,1)	SPL	−0.049	1.563	0.295	0.032	1.010	0.254	0.290
	LSE1	−1.155	36.513	23.279	0.532	16.813	10.667	3.436
	LSE2	−6.003	189.822	86.644	3.073	97.165	44.980	–
	MLE	−0.045	1.420	0.283	0.021	0.680	0.215	0.269
	DRU	−0.083	2.614	0.372	0.318	10.045	1.580	0.583
	DRW	−0.049	1.558	0.293	0.026	0.814	0.222	0.273
	ITT	0.149	4.703	0.267	−0.044	1.384	0.175	0.256
(1,−1)	SPL	0.032	1.015	0.259	−0.038	1.199	0.273	0.291
	LSE1	0.015	0.470	0.236	−0.035	1.092	0.257	1.133
	LSE2	0.038	1.213	0.325	−0.047	1.484	0.389	–
	MLE	0.035	1.118	0.253	−0.034	1.076	0.251	0.264
	DRU	0.053	1.677	0.304	−0.067	2.106	0.327	0.309
	DRW	0.030	0.960	0.257	−0.035	1.107	0.262	0.270
	ITT	−0.721	22.786	0.722	0.848	26.824	0.851	1.004

where $t_i = x_l + (i-1)\cdot(x_u - x_l)/1000$, x_l and x_u denote the 0.05 and 0.95 quantiles of the distribution of $\widetilde{X}$.

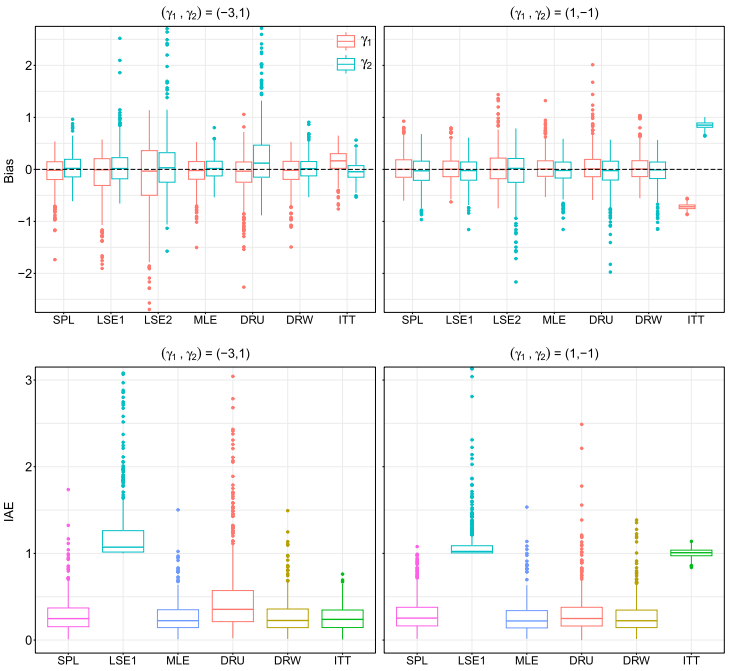
Note that in the LSE method, two different linear logistic models, $\pi(\overline{\boldsymbol{\gamma}}_{c1}^{\top}X)$ and $\pi(-\overline{\boldsymbol{\gamma}}_{c0}^{\top}X)$, are used to model $p_{c1}(1|X)$ and $p_{c0}(1|X)$, respectively. The true values of $\overline{\boldsymbol{\gamma}}_{c1}$ and $\overline{\boldsymbol{\gamma}}_{c0}$ are both equal to $\boldsymbol{\gamma}_{c0}$. Each of them can be taken as an LSE estimator of $\boldsymbol{\gamma}_{c0}$, and we use LSE1 and LSE2 to differentiate them. Despite leading to two different estimates of $\boldsymbol{\gamma}_c$, they produce only one estimate $\widehat{R}(X) = \pi(\widehat{\overline{\boldsymbol{\gamma}}}_{c1}^{\top}X)/\pi(-\widehat{\overline{\boldsymbol{\gamma}}}_{c0}^{\top}X)$ for $R(X)$, and thus we use '-' to replace the repeated items. For $n = 500$ and 1000, $B = 200$ nonparametric bootstrap samples were generated to compute the standard errors (SE) of the estimators and the corresponding Wald confidence interval coverage probabilities (CP) at the 95% level.

4.2. Results

Table 2 presents the simulation results for the estimates of $\boldsymbol{\gamma}_c$ and IAE when $\boldsymbol{\gamma}_c = (-3,1)$ and $(1,-1)$, and $n = 500$ and 1000. And Figure 2 displays the boxplots of their Biases and IAEs. It is worth noting that LSE1 and LSE2 may yield unstable results due to negative weights, DRU and DRW exhibit extreme values when initialized poorly, and ITT estimator is biased under the presence of noncompliance. For better displays in Figure 2, we excludes extremely abnormal estimates (i.e., cases where estimates are NA or the absolute bias exceeds 100).



(a) $N = 500$.



(b) $N = 1000$.

Figure 2. Boxplots for Bias of $\widehat{\boldsymbol{\gamma}}_c$ and IAE of $\widehat{R}(x)$ across different sample sizes. The top panel (a) shows results for $N = 500$ with two columns representing different $\boldsymbol{\gamma}_c$ values, while the bottom panel (b) shows results for $N = 1000$ under the same conditions. Horizontally, the first row compares Bias, and the second row compares IAE.

In Table 2, as the reported biases and RMSEs for all methods are calculated after excluding extremely abnormal estimates, we note that there are no abnormal estimates in all scenarios for SPL, and the biases, RMSEs, and IAEs are generally small, yielding highly accurate estimates. In contrast, for LSE1 and LSE2, when $\boldsymbol{\gamma}_c = (-3, 1)$ and $n = 500$, excluding 4 and 9 abnormal estimates respectively, the resulting estimates show non-negligible biases. The same pattern is observed for $n = 1000$, where 1 and 2 abnormal estimates are excluded for LSE1 and LSE2, respectively. However, for $\boldsymbol{\gamma} = (1, -1)$, LSE2 exhibits bias after excluding 2 abnormal estimates for $n = 500$, with no exclusions needed for $n = 1000$. These findings underscore the inherent instability of the LSE methods, particularly in the context of latent mixture models, as evidenced by the bias boxplots in Figure 2(a). The LSE1 and LSE2 also exhibit the largest IAEs among all approaches.

For MLE, there is a slight bias when $\boldsymbol{\gamma}_c = (-3, 1)$ and $n = 500$, which may be attributed to a few extreme estimates. Nevertheless, MLE performs comparably to SPL, and even slightly better with larger sample sizes, as anticipated given the correct specification of the parametric model. Despite this, MLE's computational cost is heavy, with runtime exceeding five times that of SPL. DRU and DRW, being doubly robust methods, reveal that DRU exhibits bias when $n = 500$, and both its RMSE and IAE are larger than those of SPL. On the other hand, DRW, leveraging optimal weights, produces more stable estimates in terms of bias and RMSE, and performs similarly to SPL. However, both DRU and DRW inherit the computational burden of MLE, since they use MLE's estimates as initial values for optimization.

Lastly, the ITT approach, which ignores actual treatment received, remains one of the most commonly employed methods in clinical trials. Despite its widespread use and sample size of $n = 1000$, ITT exhibits significant bias. Although Figure 2(a,b) demonstrate that ITT has the smallest variance, these results collectively indicate that ITT methods are suboptimal for estimating causal effects in randomized experiments with noncompliance.

The boostrap results are shown in Table 2. To mitigate the influence of outliers, we trim the top 2% of absolute bias values. The refined results, along with standard deviations (SD) calculated from 200 simulation replications, are presented in Table 3. The results demonstrate that, with the exception of ITT, all other methods achieved CPs close to the nominal 95% level. Moreover, a detailed comparison of SD and SE shows that, apart from ITT, SPL has the most consistent SD and SE values, indicating that SPL provides accurate uncertainty quantification and further supports its robustness.

4.3. Analysis under model misspecification

As we recommend linear models for $\eta_k(x; \boldsymbol{\beta}_k)$, $k = 1, 2, 3$, this subsection investigates the scenario where these models are misspecified. In the previous simulations, we set $\boldsymbol{\gamma}_a = \boldsymbol{\gamma}_n = \boldsymbol{\gamma}_c$ to ensure that the true models for $\eta_k(x; \boldsymbol{\beta}_k)$, $k = 1, 2, 3$, were linear. In contrast, in this subsection, we set $\boldsymbol{\gamma}_a = \boldsymbol{\gamma}_n \neq \boldsymbol{\gamma}_c$, so that the true models for η_1 and η_2 are no longer linear.

$$\eta_1 = \log \frac{g_a(x|1)}{g_{c0}(x|1)} = \boldsymbol{\alpha}_a^\top x + \log \frac{1 + \exp(-\boldsymbol{\gamma}_a^\top x)}{1 + \exp(\boldsymbol{\gamma}_c^\top x)} + \log \frac{\phi_c p_{c0}(1)}{\phi_a p_a(1)},$$

$$\eta_2 = \log \frac{g_n(x|1)}{g_{c0}(x|1)} = \boldsymbol{\alpha}_n^\top x + \log \frac{1 + \exp(-\boldsymbol{\gamma}_n^\top x)}{1 + \exp(\boldsymbol{\gamma}_c^\top x)} + \log \frac{\phi_c p_{c0}(1)}{\phi_n p_n(1)},$$

$$\eta_3 = \log \frac{g_{c1}(x|1)}{g_{c0}(x|1)} = \boldsymbol{\gamma}_c^\top x + \log \frac{p_{c0}(1)}{p_{c1}(1)}.$$

Table 3. Bootstrap results for the estimation of $\boldsymbol{\gamma}_c = (\gamma_1, \gamma_2)$ when $X \sim N(0, 1)$.

	$(\gamma_1, \gamma_2) = (-3, 1)$						$(\gamma_1, \gamma_2) = (1, -1)$					
	$\widehat{\gamma}_1$			$\widehat{\gamma}_2$			$\widehat{\gamma}_1$			$\widehat{\gamma}_2$		
Methods	SD	SE	CP	SD	SE	CP	SD	SE	CP	SD	SE	CP
$n = 500$												
SPL	0.358	0.441	0.965	0.322	0.376	0.955	0.294	0.525	0.970	0.340	0.488	0.990
LSE1	1.435	179.201	0.945	0.855	79.646	0.955	0.332	9.578	0.960	0.375	13.332	0.985
LSE2	9.188	195.765	0.955	3.838	94.047	0.960	0.478	100.294	0.970	0.719	139.556	0.965
MLE	0.370	0.720	0.975	0.289	0.482	0.975	0.284	0.397	0.935	0.314	0.431	0.970
DRU	0.957	1.222	0.960	3.449	3.849	0.955	0.312	0.524	0.940	0.350	0.501	0.955
DRW	0.366	0.508	0.975	0.296	0.402	0.970	0.276	0.444	0.955	0.311	0.463	0.955
ITT	0.296	0.321	0.910	0.224	0.241	0.940	0.072	0.072	0.000	0.081	0.086	0.000
$n = 1000$												
SPL	0.247	0.263	0.930	0.232	0.230	0.940	0.225	0.288	0.980	0.251	0.276	0.955
LSE1	0.390	43.724	0.960	0.319	20.307	0.940	0.206	0.240	0.950	0.223	0.265	0.950
LSE2	0.699	143.123	0.955	0.485	68.301	0.950	0.260	1.290	0.970	0.320	1.249	0.965
MLE	0.254	0.282	0.965	0.205	0.224	0.965	0.204	0.234	0.955	0.221	0.242	0.930
DRU	0.332	0.360	0.965	0.541	0.523	0.955	0.238	0.290	0.955	0.267	0.292	0.930
DRW	0.265	0.284	0.965	0.211	0.226	0.960	0.212	0.244	0.955	0.233	0.254	0.950
ITT	0.208	0.217	0.840	0.160	0.165	0.935	0.048	0.050	0.000	0.061	0.061	0.000

Table 4. The simulated results for the estimation of $\boldsymbol{\gamma}_c = (\gamma_1, \gamma_2)$ when models are specified incorrectly.

	$\widehat{\gamma}_1$					$\widehat{\gamma}_2$					
Methods	Bias	RMSE	SD	SE	CP	Bias	RMSE	SD	SE	CP	IAE
SPL	−0.012	0.189	0.189	0.201	0.960	0.082	0.355	0.345	0.346	0.960	0.316
LSE1	−0.010	0.249	0.249	27.297	0.975	0.054	0.421	0.417	59.955	0.955	0.379
LSE2	−0.069	0.394	0.388	111.842	0.960	0.164	0.627	0.605	247.602	0.975	–
MLE	−0.007	0.172	0.172	0.190	0.965	0.065	0.299	0.292	0.306	0.980	0.279
DRU	1.358	9.912	9.818	3.601	0.980	4.180	28.852	28.548	5.289	0.985	4.586
DRW	−0.010	0.173	0.173	0.194	0.970	0.072	0.312	0.303	0.328	0.985	0.278
ITT	0.071	0.161	0.145	0.159	0.930	−0.333	0.392	0.207	0.198	0.550	0.371

Nevertheless, we continue to fit linear models for estimating the CLRR. We set $\delta = 0.5$, and consider $(\boldsymbol{\alpha}_a, \boldsymbol{\alpha}_n) = (-3, -0.5, -2, 0.1)$, with $\boldsymbol{\gamma}_a = \boldsymbol{\gamma}_n = (-3, 1)$ and $\boldsymbol{\gamma}_c = (-1, 2)$. The sample size is set to $n = 500$, and the results are based on $L = 200$ Monte Carlo simulations, with 200 bootstrap samples drawn in each simulation. The results are summarized in Table 4.

Analysis of these results reveals that the Bias, SD, and RMSE of SPL are all negligible. MLE and DRW, which also rely on misspecified models, yield similarly small values for these metrics. The coverage probability (CP) of SPL is approximately 96%, indicating that the use of linear models to approximate the true underlying functions is reasonably adequate in this setting. In contrast, LSE1 and LSE2 exhibit relatively small SDs but large biases, while DRU suffers from both large bias and large SD, suggesting that these estimators are unstable. The ITT estimator remains biased, resulting in a substantially low CP of only 55%.

5. Application to OHIE data

In this section, we apply the proposed SPL method to analyze a real dataset from the Oregon Health Insurance Experiment (OHIE) (Baicker et al., 2013). The OHIE data are available at https://www.nber.org/oregon/. In January 2008, in response to the expansion of Medicaid

under the Affordable Care Act, Oregon reopened its Medicaid-based health insurance program for eligible residents, allowing a limited number of individuals to enroll within a short period. This constituted a natural experiment: individuals who met the OHIE eligibility criteria were given the opportunity to apply for Medicaid only if they were randomly selected through a lottery, while those not selected had no such opportunity. Approximately two years after the experiment ended, researchers conducted follow-up surveys and collected interview data from 12229 adults. The OHIE dataset provides a valuable framework for evaluating the effects of Medicaid coverage on various health outcomes within the previously uninsured population. Although numerous studies have used the OHIE data to assess the impact of expanded health coverage on a range of outcomes, most have overlooked the issue of noncompliance between individuals who were selected for the program and those who actually enrolled (Baicker et al., 2013; Finkelstein et al., 2012; Hattab et al., 2024), with the exception of Qiu et al. (2021).

The treatment variable D and the instrumental variable Z are binary indicators of Medicaid coverage and lottery selection, respectively. If an individual was selected in the lottery, then $Z = 1$, and if they decided to complete the application process and eventually enrolled in Medicaid after selection, then $D = 1$. Notably, those selected in the lottery could choose not to enroll in Medicaid by not applying. Theoretically, individuals not selected in the lottery had no opportunity to receive Medicaid, but in practice, some had already enrolled in Medicaid before the lottery results were announced. These individuals are classified as 'always-takers.' The outcome variables of interest are the Mental Component Summary (MCS) and the Physical Component Summary (PCS), which are measured on a 0–100 scale, with higher scores indicating better health. In the medical community for MCS and PCS, a score of 50 is widely recognized as the average level of health; therefore, we code scores greater than or equal to 50 as 1 and those less than 50 as 0 (Jenkinson et al., 1993; Ware & Sherbourne, 1992). Due to the absence of MCS and PCS results for 25 individuals in the follow-up data, we opted for a straightforward approach by excluding these individuals from the analysis. Therefore, the size of OHIE data is 12204.

Suppose there is no interference between individuals, Assumption (A0) is satisfied. For compliers who are not selected, it is still possible to observe favorable MCS and PCS outcomes, which supports Assumption (A1). Because the lottery influences individuals health only indirectly through its effect on Medicaid enrollment, this lends support to Assumption (A2). Assumptions (A3) and (A4) are justified by the random nature of the lottery and the experimental design aimed at expanding healthcare coverage. In addition, given that $(1/n)\sum_{i=1}^{n} Z_i 0.522$, Assumption (A6) appears reasonable. The observed correlation between lottery assignment and Medicaid enrollment is 0.066 (p-value $\leq$ 2.2e-16). Following Strobl et al. (2019), we test the conditional independence assumption and obtain a p-value of 1.3e-15, which provides evidence for Assumption (A4) (i.e., $\mathbb{C}\text{ov}(Z, D \mid X) \neq 0$). Finally, since individuals selected by the lottery are more likely to enroll in Medicaid, Assumption (A5) is also plausible.

Based on the preliminary results from Hattab et al. (2024), we choose four variables as covariates: Gender, Age, whether the individual is classified as high-risk (based on pre-randomization diagnoses of diabetes, hypertension, hyperlipidemia, myocardial infarction, or congestive heart failure, denoted as Risk), and whether the individual had a pre-existing diagnosis of depression (Dep). Following the methodology outlined in Baicker et al. (2013), we categorize Age into three groups: 19–34 (0), 35–49 (1), and 50–64 (2). Table 5 presents the characteristics of the OHIE data. The columns represent four covariates and two outcome

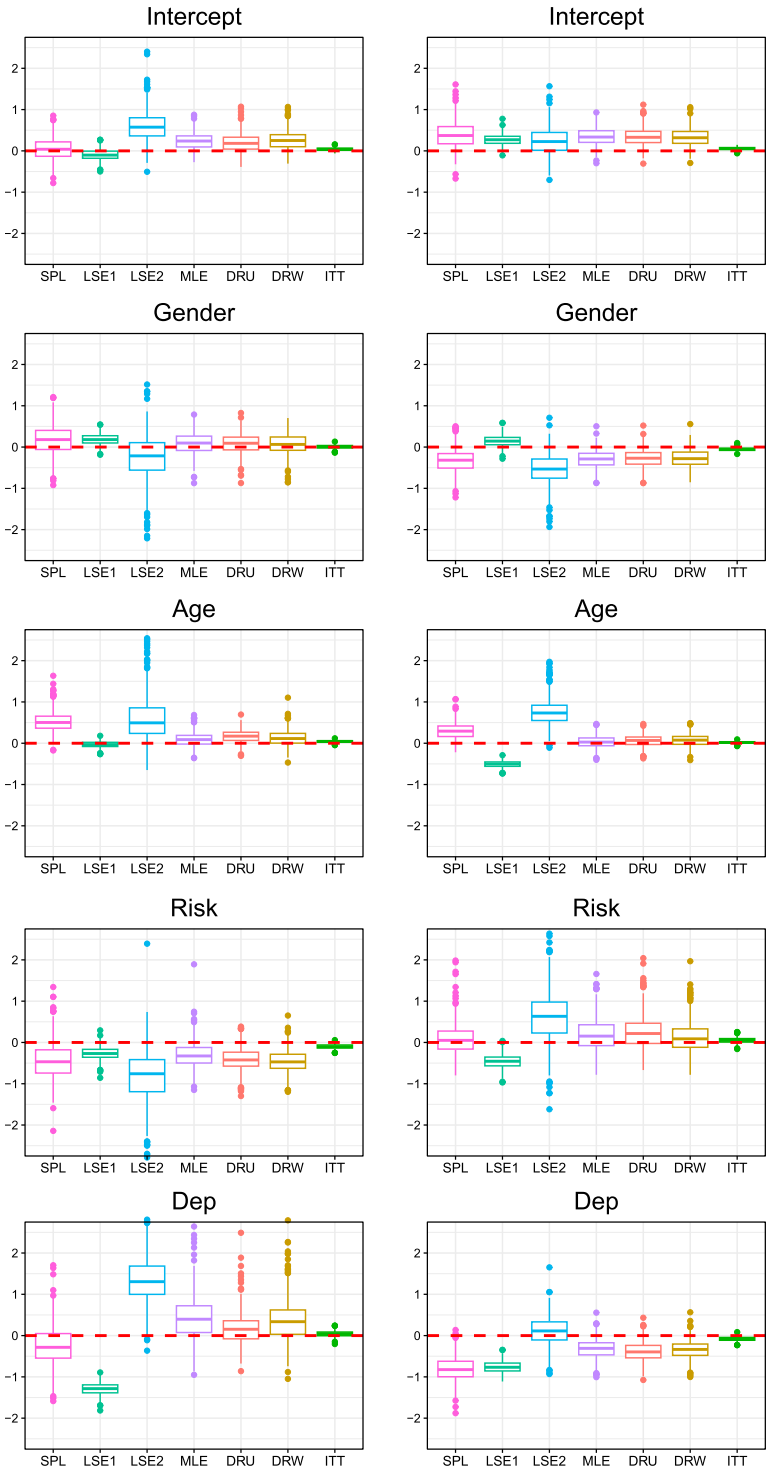
Table 5. Characteristics of the OHIE data.

	$Z=0,D=0$ ($n_{00}=5635$)	$Z=0,D=1$ ($n_{01}=195$)	$Z=1,D=0$ ($n_{10}=4473$)	$Z=1,D=1$ ($n_{11}=901$)	p-values
Gender					1.21e−05
Female	45.9	2.1	36.4	15.6	< 2.2e−16
Male	46.5	0.9	36.9	15.6	< 2.2e−16
Age					5.81e−06
19–34	47.0	1.8	37.8	13.5	< 2.2e−16
35–49	46.0	1.7	36.6	15.6	< 2.2e−16
50–64	45.3	1.3	35.2	18.2	< 2.2e−16
Risk					0.238
Low	46.1	1.6	37.1	15.2	< 2.2e−16
High	46.2	1.6	35.6	16.5	< 2.2e−16
Dep					6.53e−07
No	45.7	1.4	38.2	14.8	< 2.2e−16
Yes	47.1	2.0	33.7	17.2	< 2.2e−16
MCS					9.49e−06
Bad	47.5	1.6	34.9	16.1	< 2.2e−16
Good	44.2	1.7	39.4	14.8	< 2.2e−16
PCS					1.18e−12
Bad	46.5	1.8	34.4	17.2	< 2.2e−16
Good	45.7	1.3	39.7	13.3	< 2.2e−16

variables, with cell values showing the percentage(%) of each variable across the four (Z, D) groups. p-values associated with each covariate category reflect the heterogeneity testing results between these groups. Notably, with the exception of Risk (p-value $= 0.238$), both Gender (p-value $= 1.21\text{e}{-05}$), Age (p-value $= 5.81\text{e}{-06}$), and Dep (p-value $= 6.53\text{e}{-07}$) demonstrate significant heterogeneity. Specifically, the p-values corresponding to different values of covariates and outcomes reflect the significance levels across the various (Z, D) groups. All p-values are below 2.2e−16, indicating significant heterogeneity across the different latent principal strata to some extent. Combined with the estimated proportion of noncompliers being 0.735, these findings suggest that naive methods (e.g., AT, PP, ITT) that ignore principal stratification structures are inadequate for accurately capturing the underlying complexities. When targeting conditional local treatment effects, instrumental variable methods are essential for proper identification and estimation.

Finkelstein et al. (2012) found that Medicaid coverage improves self-reported health as measured by mental and physical component scores. To investigate heterogeneity in treatment effects across subgroups, we estimate CLRR for MCS and PCS using the four baseline covariates. Here we focus on the hetergeneity to find which subgroup benefits from Medicaid coverage. The analysis employs $B = 500$ nonparametric bootstrap samples to construct 95% confidence intervals, with results presented in Tables 6 and 7. The corresponding bootstrap results are visualized using boxplots in Figure 3(a,b).

As shown in Table 6, our method yields a point estimate $\widehat{\boldsymbol{\gamma}}_c = (0.021, 0.183, 0.491, -0.470, -0.214)$. Notably, the 95% confidence interval for Age remains entirely positive, providing moderately strong evidence that older individuals benefit more from Medicaid in terms of mental health. Figure 3(a) reveals negative bootstrap estimates for the risk covariate in our SPL method, suggesting (though not statistically significant) that low-risk individuals and males may derive greater mental health benefits. In contrast, LSE1 and LSE2 produce markedly different results, calling their reliability into question. While MLE, DRU, and DRW are consistent with SPL, they do not provide strong evidence for any covariate. The ITT



(a) MCS. (b) PCS.

Figure 3. Boxplots of bootstrap results for parameters in CLRR, representing the causal effect among compliers for different factors on MCS and PCS. (a) MCS and (b) PCS.

Table 6. Estimates of parameters in CLRR and their 95% Confidence Intervals on MCS.

	Intercept	Gender	Age	Risk	Dep
SPL	0.021	0.183	0.491	−0.470	−0.214
	(−0.391, 0.575)	(−0.574, 0.966)	(0.082, 1.080)	(−1.240, 0.572)	(−1.099, 0.793)
LSE1	−0.097	0.189	−0.033	−0.256	−1.274
	(−0.345, 0.147)	(−0.056, 0.425)	(−0.178, 0.108)	(−0.561, 0.033)	(−1.560, −1.022)
LSE2	0.546	−0.229	0.479	−0.731	1.218
	(0.004, 1.501)	(−1.515, 0.626)	(−0.211, 2.486)	(−3.590, 0.369)	(0.397, 2.972)
MLE	0.229	0.183	0.302	−0.304	−0.115
	(−0.147, 0.627)	(−0.442, 0.565)	(−0.215, 0.470)	(−0.850, 0.347)	(−0.474, 1.527)
DRU	0.179	0.082	0.164	−0.406	0.154
	(−0.197, 0.705)	(−0.398, 0.557)	(−0.112, 0.465)	(−1.022, 0.134)	(−0.462, 0.932)
DRW	0.209	0.027	0.167	−0.394	0.159
	(−0.155, 0.762)	(−0.439, 0.535)	(−0.240, 0.503)	(−1.018, 0.167)	(−0.460, 1.610)
ITT	0.039	0.008	0.042	−0.100	0.045
	(−0.031, 0.116)	(−0.082, 0.089)	(−0.012, 0.097)	(−0.222, 0.004)	(−0.097, 0.176)

Table 7. Estimates of parameters in CLRR and their 95% Confidence Intervals on PCS.

	Intercept	Gender	Age	Risk	Dep
SPL	0.314	−0.275	0.300	0.055	−0.783
	(−0.194, 1.051)	(−0.903, 0.201)	(−0.074, 0.667)	(−0.526, 0.894)	(−1.419, −0.253)
LSE1	0.250	0.150	−0.493	−0.462	−0.757
	(0.020, 0.530)	(−0.119, 0.378)	(−0.655, −0.355)	(−0.768, −0.124)	(−1.045, −0.487)
LSE2	0.214	−0.459	0.695	0.573	0.103
	(−0.339, 0.905)	(−1.321, 0.138)	(0.258, 1.651)	(−0.550, 18.642)	(−0.547, 0.784)
MLE	0.285	−0.103	0.065	0.033	−0.395
	(−0.041, 0.777)	(−0.686, 0.066)	(−0.249, 0.320)	(−0.409, 1.062)	(−0.764, 0.108)
DRU	0.299	−0.239	0.064	0.203	−0.385
	(−0.012, 0.804)	(−0.679, 0.135)	(−0.232, 0.370)	(−0.392, 1.148)	(−0.833, 0.051)
DRW	0.295	−0.249	0.071	0.083	−0.345
	(−0.054, 0.771)	(−0.698, 0.111)	(−0.221, 0.384)	(−0.494, 0.979)	(−0.800, 0.109)
ITT	0.052	−0.048	0.018	0.047	−0.077
	(−0.008, 0.123)	(−0.129, 0.025)	(−0.045, 0.078)	(−0.071, 0.186)	(−0.190, 0.014)

estimates are close to zero, and the 95% confidence intervals further suggest a lack of heterogeneity. This is consistent with the findings in Hattab et al. (2024) that most subgroups exhibited no statistically significant impacts on MCS despite substantial overall effects.

When investigating the conditional effect on PCS, heterogeneity becomes more pronounced. In Table 7, our method yields $\widehat{\boldsymbol{\gamma}}_c = (0.314, -0.275, 0.300, 0.055, -0.783)$. Unlike the MCS results, the 95% confidence interval for Dep is less than 0, suggesting that individuals with depression benefit less in terms of physical health from Medicaid. Figure 3(b) demonstrates that the bootstrap estimates of our SPL method for both the Intercept and Age are positive, while those for both Gender and Dep are negative. This indicates that older, female beneficiaries without depression derive greater physical health benefits, a finding corroborated by Hattab et al. (2024). Similar to the MCS case, LSE1 and LSE2 yield disparate results, further undermining their credibility.

Overall, this study reveals heterogeneous patterns of Medicaid effects across Age, Gender, Risk and Dep through different subgroup analyses, although the heterogeneity in all considered outcomes is relatively low. Interestingly, the effects of gender on MCS and PCS appear to be in opposite directions, which is new to the literature.

6. Conclusion

CLRR is a widely used and highly regarded measure of causal effect in causal analysis involving binary outcomes, particularly in health and medical research. In this paper, we introduce a novel semiparametric two-step likelihood-based estimation procedure, termed SPL, for evaluating CLRR in RCTs with a binary outcome and noncompliance. As the traditional identification of CLRR in (3) is unsuitable for binary outcomes, our SPL approach subtly re-expresses CLRR in terms of two constants, $p_{c0}(1)$ and $p_{c1}(1)$, and two functions, $g_{c0}(x|1)$ and $g_{c1}(x|1)$. By reducing the number of required posited models from six to three density ratio models, the SPL approach enhances robustness in estimation and significantly improves computational efficiency. This method is especially advantageous in scenarios characterized by limited pre-existing data information, as it aims to capture the underlying treatment effects with greater reliability and robustness.

The data structure considered in this paper aligns with at least three common scenarios in the literature: (1) randomized trials in which noncompliance is either permitted or unavoidable, e.g., the FILMS trial (Lois et al., 2008) and the COPERS trial (Taylor et al., 2016), (2) natural experiments, e.g., the OHIE in Section 5, and (3) encouragement experiments, e.g., the moving to opportunity for Fair Housing Demonstration Project from the Department of Housing and Urban Development (Matsouaka & Tchetgen Tchetgen, 2017), where treatment uptake is influenced by randomized encouragement rather than direct assignment. In the first two scenarios, noncompliance in randomized experiments arises due to unavoidable constraints, such as ethical or moral considerations. Natural experiments are not originally designed for causal inference, but happen to contain a naturally occurring instrumental variable in the data. Under both settings, it is more appropriate to adopt a retrospective approach.

This paper focuses solely on causal effect for compliers, which represents a genuine cause-and-effect relationship, restricted to the complier population. As the most commonly used method in clinical studies, ITT ignores the actual treatment and is influenced by the causal effect on noncompliers, i.e., selection bias. It is worth noting the causal effect for compliers is essential and aligns with real-world needs. For instance, in the realm of marketing, compliers refer to the target audience members who are particularly sensitive to the marketing campaign and, consequently, garner focussed attention. However, identifying compliers is challenging in practice, and is an interesting direction worth further exploring (Kennedy et al., 2020).

A fundamental assumption in this paper is the independence assumption, i.e., $Z \perp \{D(0), D(1), Y(0), Y(1), X\}$. As an anonymous referee, the proposed SPL method may be extended to the case where on the conditional independence (unconfoundedness) assumption, i.e., $Z \perp \{D(0), D(1), Y(0), Y(1)\} \mid X$, holds. This extension generalizes our SPL method for estimating CLRR in randomized controlled trials to stratified randomized designs and observational studies. However, under the unconfoundedness assumption,

$$\begin{aligned} &\mathrm{p}(x|Z=z, S=s) \neq \mathrm{p}(x|S=s), \quad s=\mathrm{a},\mathrm{n},\mathrm{c}, \\ &\mathrm{p}(x|Z=z, S=s, Y=y) \neq \mathrm{p}(x|S=s, Y=y), \quad s=\mathrm{a},\mathrm{n},\mathrm{c}, \end{aligned}$$

which makes the estimations of immediate parameters and CLRR more challenging. We leave this topic for future research.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research is supported by the National Natural Science Foundation of China [Grant numbers 12371293, 32030063, 12171157].

References

Abadie, A. (2002). Bootstrap tests for distributional treatment effects in instrumental variable models. *Journal of the American Statistical Association*, *97*(457), 284–292. https://doi.org/10.1198/016214502753479419

Abadie, A. (2003). Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics*, *113*(2), 231–263. https://doi.org/10.1016/S0304-4076(02)00201-4

Anderson, J. E. (1979). A theoretical foundation for the gravity equation. *The American Economic Review*, *69*(1), 106–116.

Baicker, K., Taubman, S. L., Allen, H. L., Bernstein, M., Gruber, J. H., J. P. Newhouse, Schneider, E. C., Wright, B. J., Zaslavsky, A. M., & A. N. Finkelstein (2013). The Oregon experiment—Effects of Medicaid on clinical outcomes. *New England Journal of Medicine*, *368*(18), 1713–1722. https://doi.org/10.1056/NEJMsa1212321

Charles, P., Giraudeau, B., Dechartres, A., Baron, G., & Ravaud, P. (2009). Reporting of sample size calculation in randomised controlled trials: Review. *British Medical Journal*, *338:b1732*. https://doi.org/10.1136/bmj.b1732

Dodd, S., White, I. R., & Williamson, P. (2012). Nonadherence to treatment protocol in published randomised controlled trials: A review. *Trials*, *13*(1), Article 84. https://doi.org/10.1186/1745-6215-13-84

Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J. P., Allen, H., & Baicker, K. & Oregon Health Study Group, t. (2012). The Oregon health insurance experiment: Evidence from the first year. *The Quarterly Journal of Economics*, *127*(3), 1057–1106. https://doi.org/10.1093/qje/qjs020

Frangakis, C. E., & Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, *58*(1), 21–29. https://doi.org/10.1111/biom.2002.58.issue-1

Frölich, M. (2007). Nonparametric iv estimation of local average treatment effects with covariates. *Journal of Econometrics*, *139*(1), 35–75. https://doi.org/10.1016/j.jeconom.2006.06.004

Hattab, Z., Doherty, E., Ryan, A. M., & O'Neill, S. (2024). Heterogeneity within the oregon health insurance experiment: An application of causal forests. *PloS one*, *19*(1), 1–21. https://doi.org/10.1371/journal.pone.0297205

Have, T. R. T., Joffe, M., & Cary, M. (2003). Causal logistic models for non-compliance under randomized treatment with univariate binary response. *Statistics in Medicine*, *22*(8), 1255–1283. https://doi.org/10.1002/sim.v22:8

Hirano, K., Imbens, G. W., Rubin, D. B., & Zhou, X.-H. (2000). Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics*, *1*(1), 69–88. https://doi.org/10.1093/biostatistics/1.1.69

Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics*, *86*(1), 4–29. https://doi.org/10.1162/003465304323023651

Imbens, G. W., & Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, *62*(2), 467–475. https://doi.org/10.2307/2951620

Imbens, G. W., & Rubin, D. B. (1997). Bayesian inference for causal effects in randomized experiments with noncompliance. *The Annals of Statistics*, *25*(1), 305–327. https://doi.org/10.1214/aos/1034276631

Jenkinson, C., Coulter, A., & Wright, L. (1993). Short form 36 (SF36) health survey questionnaire: Normative data for adults of working age. *British Medical Journal*, *306*(6890), 1437–1440. https://doi.org/10.1136/bmj.306.6890.1437

Kennedy, E. H., Balakrishnan, S., & G'Sell, M. (2020). Sharp instruments for classifying compliers and generalizing causal effects. *The Annals of Statistics*, *48*(4), 2008–2030. https://doi.org/10.1214/19-AOS1874

Kohavi, R., & Thomke, S. (2017). The surprising power of online experiments. *Harvard Business Review*, *95*(5), 74–82.

Little, R. J., & Yau, L. H. (1998). Statistical techniques for analyzing data from prevention trials: Treatment of no-shows using rubin's causal model. *Psychological Methods*, *3*(2), 147–159. https://doi.org/10.1037/1082-989X.3.2.147

Lois, N., Burr, J., Norrie, J., Vale, L., Cook, J., & McDonald, A. (2008). Clinical and cost-effectiveness of internal limiting membrane peeling for patients with idiopathic full thickness macular hole. protocol for a randomised controlled trial: Films (full-thickness macular hole and internal limiting membrane peeling study). *Trials*, *9*(1), Article 61. https://doi.org/10.1186/1745-6215-9-61

Matsouaka, R. A., & Tchetgen Tchetgen, E. J. (2017). Instrumental variable estimation of causal odds ratios using structural nested mean models. *Biostatistics*, *18*(3), 465–476. https://doi.org/10.1093/biostatistics/kxw059

Ogburn, E. L., Rotnitzky, A., & Robins, J. M. (2015). Doubly robust estimation of the local average treatment effect curve. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *77*(2), 373–396. https://doi.org/10.1111/rssb.12078

Okui, R., Small, D. S., Tan, Z., & Robins, J. M. (2012). Doubly robust instrumental variable regression. *Statistica Sinica*, *22*(1), 173–205. https://doi.org/10.5705/ss.2011.v22n1a

Pearl, J. (2009). *Causality* (2nd ed.). Cambridge University Press.

Piantadosi, S. (2024). *Clinical Trials: A Methodologic Perspective*. John Wiley & Sons.

Qiu, Y., Tao, J., & Zhou, X. (2021). Inference of heterogeneous treatment effects using observational data with high-dimensional covariates. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *83*(5), 1016–1043. https://doi.org/10.1111/rssb.12469

Richardson, T. S., & Robins, J. M. (2013). *Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality* (Working Paper, 128). Center for the Statistics and the Social Sciences, University of Washington.

Richardson, T. S., Robins, J. M., & Wang, L. (2017). On modeling and estimation for the relative risk and risk difference. *Journal of the American Statistical Association*, *112*(519), 1121–1130. https://doi.org/10.1080/01621459.2016.1192546

Robins, J. M. (1994). Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics-Theory and Methods*, *23*(8), 2379–2412. https://doi.org/10.1080/03610929408831393

Rothman, K. J., Greenland, S., & Lash, T. L. (2008). *Modern Epidemiology* (3rd ed.). Lippincott Williams & Wilkins.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701. https://doi.org/10.1037/h0037350

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin.

Stamler, J., Wentworth, D., & Neaton, J. D. (1986). Is relationship between serum cholesterol and risk of premature death from coronary heart disease continuous and graded? Findings in 356 222 primary screenees of the multiple risk factor intervention trial (MRFIT). *Journal of the American Medical Association*, *256*(20), 2823–2828. https://doi.org/10.1001/jama.1986.03380200061022

Strobl, E. V., Zhang, K., & Visweswaran, S. (2019). Approximate kernel-based conditional independence tests for fast non-parametric causal discovery. *Journal of Causal Inference*, *7*(1), 20180017. https://doi.org/10.1515/jci-2018-0017

Tan, Z. (2006). Regression and weighting methods for causal inference using instrumental variables. *Journal of the American Statistical Association*, *101*(476), 1607–1618. https://doi.org/10.1198/016214505000001366

Taylor, S. J., Carnes, D., Homer, K., Kahan, B. C., Hounsome, N., Eldridge, S., Spencer, A., Pincus, T., Rahman, A., & Underwood, M. (2016). Novel three-day, community-based, nonpharmacological

group intervention for chronic musculoskeletal pain (copers): A randomised clinical trial. *PLOS Medicine*, *13*(6), 1–18. https://doi.org/10.1371/journal.pmed.1002040

Wang, L., Zhang, Y., Richardson, T. S., & Robins, J. M. (2021). Estimation of local treatment effects under the binary instrumental variable model. *Biometrika*, *108*(4), 881–894. https://doi.org/10.1093/biomet/asab003

Ware, J. E. J., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. conceptual framework and item selection. *Medical Care*, *30*(6), 473–483. https://doi.org/10.1097/00005650-199206000-00002