Identification and estimation of causal effects in the presence of confounded principal strata

Shanshan Luo[*], Wei Li[†], Wang Miao[‡] and Yangbo He[§]

June 20, 2022

Abstract

The principal stratification has become a popular tool to address a broad class of causal inference questions, particularly in dealing with non-compliance and truncation-by-death problems. The causal effects within principal strata which are determined by joint potential values of the intermediate variable, also known as the principal causal effects, are often of interest in these studies. Analyses of principal causal effects from observed data in the literature mostly rely on ignorability of the treatment assignment, which requires practitioners to accurately measure as many as covariates so that all possible confounding sources are captured. However, collecting all potential confounders in observational studies is often difficult and costly, the ignorability assumption may thus be questionable. In this paper, by leveraging available negative controls that have been increasingly used to deal with uncontrolled confounding, we consider identification and estimation of causal effects when the treatment and principal strata are confounded by unobserved variables. Specifically, we show that the principal causal effects can be nonparametrically identified by invoking a pair of negative controls that are both required not to directly affect the outcome. We then relax this assumption and establish identification of principal causal effects under various semiparametric or parametric models. We also propose an estimation method of principal causal effects. Extensive simulation studies show good performance of the proposed approach and a real data application from the National Longitudinal Survey of Young Men is used for illustration.

Keywords: Causal Inference; Negative Control; Non-compliance; Principal Stratification; Unmeasured Confounding.

---
[*]School of Mathematical Sciences, Peking University. Email: shan3_luo@pku.edu.cn
[†]Center for Applied Statistics and School of Statistics, Renmin University of China. Email: weilistat@ruc.edu.cn
[‡]School of Mathematical Sciences, Peking University. Email: mwfy@pku.edu.cn
[§]School of Mathematical Sciences, Peking University. Email: heyb@math.pku.edu.cn

1 Introduction

Many scientific problems are concerned with evaluating the causal effect of a treatment on an outcome in the presence of an intermediate variable. Direct comparisons conditional on observed values of the intermediate variable are usually not causally interpretable. Frangakis and Rubin (2002) propose the principal stratification framework and define principal causal effects that can effectively compare different treatment assignments in such settings. The principal stratification is defined by joint potential values of the intermediate variable under each treatment level being compared, which is not affected by treatment assignment, and hence it can be viewed as a pretreatment covariate to classify individuals into subpopulations. Principal causal effects that are defined as potential outcome contrasts within principal strata thus exhibit clear scientific interpretations in many practical studies (VanderWeele, 2011). For instance, in non-compliance problems, the intermediate variable is the actual treatment received, the principal stratification represents the compliance status, and the treatment assignment plays the role of an instrumental variable in identifying the complier average causal effect (Angrist et al., 1996). In truncation-by-death problems, the intermediate variable denotes survival status, and a meaningful parameter termed survivor average causal effect is defined as the effect among the subgroup who would survive under both treatment levels (Rubin et al., 2006; Zhang et al., 2009; Ding et al., 2011).

Analysis of principal causal effects from observed data is challenging, because principal stratification is often viewed as an unobserved confounder between the intermediate and outcome variables. Most works in the literature rely on the ignorability of treatment assignment, which states that the distributions of potential values of intermediate and outcome variables do not vary across the treatment assignment given observed covariates. This assumption essentially requires that observed covariates account for all possible confounding factors between the treatment and post-treatment variables. Since the principal causal effects are defined on the latent principal strata, one can only establish large sample bounds or conduct sensitivity analysis for principal causal effects under the ignorability assumption (Zhang and Rubin, 2003; Lee, 2009; Long and Hudgens, 2013), but fails to obtain identifiability results without additional assumptions. Previous literature has used an auxiliary variable that satisfies some conditional independence conditions to address the identification issues (Ding et al., 2011; Jiang et al., 2016; Ding and Lu, 2017; Wang et al., 2017; Luo et al., 2021). However, as can happen in observational studies, one may not sufficiently collect the pretreatment covariates. The existence of unmeasured variables may render the ignorability assumption invalid and thus the traditional causal estimates in principal stratification analysis can be biased.

As far as we know, there has not been much discussion on principal causal effects when the ignorability assumption fails. Several authors have considered the setting where the potential values of the intermediate variable are correlated with the treatment assignment even after conditioning on observed covariates. In other words, the treatment and the intermediate variable are confounded by unmeasured factors in this setting. Schwartz et al. (2012) present model-based approaches for assessing the sensitivity of complier average causal effect estimates in non-compliance problems when there exists unmeasured confounding in the treatment arms. Kédagni (2021) discusses similar problems and provides identifiability results by using a proxy for the confounded treatment assignment

under some tail restrictions for the potential outcome distributions. Deng et al. (2021) study truncation-by-death problems and establish identification of the conditional average treatment effects for always-survivors given observed covariates by employing an auxiliary variable whose distribution is informative of principal strata. However, because the conditional distributions of principal strata given covariates are not identified, the survivor average causal effect is generally not identifiable in their setting.

To overcome these limitations, we establish identification of principal causal effects by leveraging a pair of negative control variables. In the absence of intermediate variables, many researchers have employed a negative control exposure and a negative control outcome to identify the average causal effects when unobserved confounders exist (Miao et al., 2018; Shi et al., 2020; Miao et al., 2020; Cui et al., 2020). However, the principal causal effects may be of more interest in the presence of an intermediate variable. For instance, in truncation-by-death problems, individuals may die before their outcome variables are measured, and hence the outcomes for dead individuals are not well defined. Then the survivor average causal effect is more scientifically meaningful in these studies (Rubin et al., 2006; Tchetgen Tchetgen, 2014). While the identification and estimation of average causal effects within the negative control framework have been well studied in the literature, it remains uncultivated in studies where an intermediate variable exists and principal causal effects are of interest.

In this paper, we develop identification and estimation of principal causal effect in the presence of unmeasured confounders. Specifically, we first introduce a confounding bridge function that links negative controls and the intermediate variable to identify proportions of the principal strata. We then establish nonparametric identification of principal causal effects by assuming that the negative controls have no direct effect on the outcome. We next relax this assumption and show alternative identifiability results based on semiparametric and parametric models. Finally, we provide an estimation method and discuss the asymptotic properties. We evaluate the performance of the proposed estimator with simulation studies and a real data application.

2   Notation and assumptions

Assume that there are n individuals who are independent and identically sampled from a superpopulation of interest. Let Z denote a binary treatment assignment with 1 indicating treatment and 0 for control. Let Y denote an outcome of interest, and let S denote a binary intermediate variable. Let X denote a vector of covariates observed at baseline. We use the potential outcomes framework and make the stable unit treatment value assumption; that is, there is only one version of potential outcomes and there is no interference between units (Rubin, 1990). Let S_z and Y_z denote the potential values of the intermediate variable and outcome that would be observed under treatment $Z = z$. The observed values S and Y are deterministic functions of the treatment assignment and their respective potential values: $S = ZS_1 + (1 - Z)S_0$ and $Y = ZY_1 + (1 - Z)Y_0$.

Frangakis and Rubin (2002) define the principal stratification as joint potential values of the intermediate variable under both the treatment and control. We denote the basic principal stratum by G and it can be expressed as $G = (S_0, S_1)$. Since each of the potential values has two levels, there are four different principal strata in total. For simplicity, we

refer to these principal strata, namely, $\{(0,0), (0,1), (1,1), (1,0)\}$ as never-takers ($\bar{s}\bar{s}$), compliers ($s\bar{s}$), always-takers (ss), and defiers ($\bar{s}s$), respectively. The causal estimand of interest is the principal causal effect, i.e.,

$$\Delta_g = E(Y_1 - Y_0 \mid G = g), \ g \in \{\bar{s}\bar{s}, s\bar{s}, \bar{s}s, ss\}.$$

The principal causal effect conditional on a latent variable G is not identifiable without additional assumptions. Here we do not impose the exclusion restriction assumption (Angrist et al., 1996) that requires no individual causal effect on the outcome among the subpopulations $G = ss$ and $G = \bar{s}\bar{s}$, because in many settings with intermediate variables, such as truncation-by-death or surrogate problems (Gilbert and Hudgens, 2008), the very scientific question of interest is to test whether the principal causal effect Δ_{ss} or $\Delta_{\bar{s}\bar{s}}$ is zero. Under this setup, the identification of Δ_g in the literature often relies on the following monotonicity assumption.

Assumption 1 (Monotonicity). $S_1 \geq S_0$.

Monotonicity rules out the existence of the defier group $G = \bar{s}s$. This assumption may be plausible in some observational studies. For example, in studies evaluating the effect of educational attainment on future earnings, a subject living near a college is likely to receive a higher educational level. The second commonly-used assumption is the treatment ignorability assumption: $Z \perp\!\!\!\perp (S_0, S_1, Y_0, Y_1) \mid X$. This assumption entails that the baseline covariates X control for all confounding factors between the treatment and post-treatment variables. However, the ignorability fails in the presence of unmeasured confounding. Let U denote an unobserved variable, which together with observed covariates X, captures all potential confounding sources between the treatment Z and variables (S, Y). We impose the following latent ignorability assumption.

Assumption 2 (Latent ignorability). *(i) $Z \perp\!\!\!\perp (S_0, S_1) \mid (U, X)$; (ii) $Z \perp\!\!\!\perp (Y_0, Y_1) \mid (G, X)$.*

The type of confounding considered in Assumption 2(i) is termed S-confounding by Schwartz et al. (2012). The presence of the unmeasured variable U in this assumption brings about dramatic methodological changes and important technical challenges to principal stratification analysis. For example, when the traditional ignorability assumption holds, the inequality $\mathrm{pr}(S = 1 \mid Z = 1, X) < \mathrm{pr}(S = 1 \mid Z = 0, X)$ can be used to falsify the monotonicity assumption. However, if U exists, it is no longer possible to empirically test monotonicity using this inequality. In addition, if we define principal score $\pi_g(X)$ as the proportion of the principal stratum given observed covariates (Ding and Lu, 2017), namely, $\pi_g(X) = \mathrm{pr}(G = g \mid X)$, the presence of U impedes identification of $\pi_g(X)$. Assumption 2(ii) means that the confounding factors between the treatment and the outcome are fully characterized by the latent principal stratification G and observed covariates X (Wang et al., 2017). Assumption 2 has also been considered by Kédagni (2021) and Deng et al. (2021).

We next discuss identification of Δ_g under Assumptions 1 and 2. For simplicity, we define $\mu_{z,g} = E(Y_z \mid G = g)$, and hence $\Delta_g = \mu_{1,g} - \mu_{0,g}$. It suffices to identify $\mu_{z,g}$ for the identification of Δ_g. Let $\mu_{z,g}(X) = E(Y \mid Z = z, G = g, X)$. Then under Assumption 2(ii), we have

$$\mu_{z,g} = E\{\mu_{z,g}(X)\pi_g(X)\}/E\{\pi_g(X)\}. \tag{1}$$

4

It can be seen that the identification of $\mu_{z,g}$ depends on that of $\pi_g(X)$ and $\mu_{z,g}(X)$. Under Assumptions 1 and 2(i), we have that

$$\pi_{ss}(X) = E\left\{p_0(X,U) \mid X\right\}, \quad \pi_{\bar{s}\bar{s}}(X) = 1 - E\left\{p_1(X,U) \mid X\right\},$$
$$\pi_{s\bar{s}}(X) = E\left\{p_1(X,U) - p_0(X,U) \mid X\right\},$$

where $p_z(X,U) = \mathrm{pr}(S = 1 \mid Z = z, X, U)$. Because U is unobserved, the principal scores in the above equations cannot be identified without additional assumptions. As for the conditional outcome means $\{\mu_{z,g}(X) : z = 0, 1; g = ss, \bar{s}\bar{s}, s\bar{s}\}$, only $\mu_{0,ss}(X)$ and $\mu_{1,\bar{s}\bar{s}}(X)$ can be identified under Assumptions 1 and 2(ii) by $\mu_{0,ss}(X) = E(Y \mid Z = 0, S = 1, X)$ and $\mu_{1,\bar{s}\bar{s}}(X) = E(Y \mid Z = 1, S = 0, X)$. However, the identifiability of other conditional outcome means is not guaranteed, because the observed data $(Z = 1, S = 1, X)$ and $(Z = 0, S = 0, X)$ are mixtures of two principal strata:

$$
\begin{aligned}
E(Y \mid Z = 1, S = 1, X) &= \sum_{g=ss,s\bar{s}} \eta_g(1, X)\mu_{1,g}(X), \\
E(Y \mid Z = 0, S = 0, X) &= \sum_{g=s\bar{s},\bar{s}\bar{s}} \eta_g(0, X)\mu_{0,g}(X),
\end{aligned}
\tag{2}
$$

where $\eta_g(1, X) = \omega_g(1, X)/\{\omega_{ss}(1, X) + \omega_{s\bar{s}}(1, X)\}$, $\eta_g(0, X) = \omega_g(0, X)/\{\omega_{\bar{s}\bar{s}}(0, X) + \omega_{s\bar{s}}(0, X)\}$, and $\omega_g(z, X) = \mathrm{pr}(G = g \mid Z = z, X)$. In later sections, the conditional probabilities of principal strata given only a subset V of covarites X may be of interest, and we simply denote them by replacing X with V in the original notations. For example, $\pi_g(V) = \mathrm{pr}(G = g \mid V)$. Other notations, such as $\omega_g(z, V)$ and $\eta_g(z, V)$, can be similarly interpreted. Due to the presence of unobserved confounders U, the weights $\eta_g(z, X)$ in (2) are no longer identifiable, which complicates the identification and differs from most of existing results in the literature. In such a case, the large sample bounds or sensitivity analysis for these conditional outcome means cannot be easily obtained without further assumptions and it would be even more difficult to obtain their identifiability results. In the following section, we discuss how to establish the identifiability of principal causal effects based on auxiliary variables.

3   Identification

3.1   Nonparametric identification using a pair of negative controls

In this section, we establish a nonparametric identification result for principal causal effects through a pair of negative control variables when the ignorability assumption fails. Motivated from the proximal causal inference framework for identifying average treatment effects (Miao et al., 2018; Shi et al., 2020; Miao et al., 2020; Cui et al., 2020), we assume that the covariates X can be decomposed into (A, W, C^{T}) such that A serves as a negative control exposure, W serves as a negative control intermediate variable and C accounts for the remaining observed confounders. For convenience, we may use the notation X and (A, W, C^{T}) interchangeably below.

Assumption 3 (Negative control). $(Z, A) \perp\!\!\!\perp (S_0, S_1, W) \mid (C, U)$
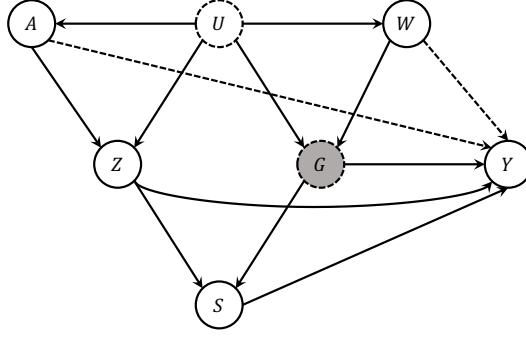
Figure 1: A causal diagram illustrating treatment and intermediate variable confounding proxies when ignorability assumption fails. Dashed arrows indicate edges can exist when semiparametric or parametric models are considered. Observed covariates are omitted for simplicity.

Assumption 4 (Confounding bridge). *There exists a function $h(z, W, C)$ such that $\mathrm{pr}(S = 1 \mid Z = z, C, U) = E\{h(z, W, C) \mid C, U\}$ almost surely for all z.*

Assumption 3 implies that the variables (C, U) are sufficient to account for the confounding between (Z, A) and (S_0, S_1, W). The negative control exposure A does not directly affect either the intermediate variable S or the negative control intermediate variable W. Assumption 3 imposes no restrictions on Z-A association or G-W association, and allows the two negative controls A and W to be confounded by the unmeasured variable U. See Fig. 1 for a graphic illustration. The confounding bridge function in Assumption 4 establishes the connection between the negative control W and the intermediate variable S. Assumption 4 defines an inverse problem known as the Fredholm integral equation of the first kind. The technical conditions for the existence of a solution are provided in Carrasco et al. (2007). Since the principal stratum G is a latent variable, Assumptions 3 and 4, which are used to control for unobserved confounding between treatment and intermediate variables, are not sufficient to nonparametrically identify principal causal effects. We thus impose the following conditional independence condition between the negative controls and potential outcomes given the latent variable G and observed covariates C.

Assumption 5. $(Z, A, W) \perp\!\!\!\perp (Y_0, Y_1) \mid (G, C)$.

Under Assumption 5, we can view the observed variables A and W as proxies of G, the role of which resembles the usual instrumental variables that preclude direct effects on the outcome Y (Angrist et al., 1996); see Fig. 1 for an illustration. Similar assumptions have been widely used in principal stratification literature (Ding et al., 2011; Jiang et al., 2016; Wang et al., 2017; Luo et al., 2021). In the next subsection, we shall consider to relax this assumption based on semiparametric or parameric models.

Theorem 1. *Suppose that Assumptions 1, 2(i), 3 and 4 hold. Then the conditional probabilities of principal strata are identified by*

$$
\begin{aligned}
\omega_{ss}(Z, A, C) &= E\{h(0, W, C) \mid Z, A, C\}, \\
\omega_{\bar{s}\bar{s}}(Z, A, C) &= 1 - E\{h(1, W, C) \mid Z, A, C\}, \\
\omega_{s\bar{s}}(Z, A, C) &= 1 - \omega_{ss}(Z, A, C) - \omega_{\bar{s}\bar{s}}(Z, A, C).
\end{aligned}
\tag{3}
$$

6

Under additional Assumptions 2(ii) and 5, the principal causal effects are identifiable if for any $C = c$, the functions in the following two vectors

$$\{\eta_{s\bar{s}}(0, A, c), \eta_{\bar{s}\bar{s}}(0, A, c)\}^{\mathrm{T}} \quad and \quad \{\eta_{s\bar{s}}(1, A, c), \eta_{ss}(1, A, c)\}^{\mathrm{T}} \tag{4}$$

are respectively linearly independent.

The identifiability result (3) in Theorem 1 links the confounding bridge function with the conditional probabilities of principal strata given observed variables (Z, A, C). With an additional completeness condition, the bridge function in Assumption 4 can be equivalently characterized by a solution to an equation based on observed variables (see Lemma S2 in the supplementary materials). Note that Assumption 4 only requires the existence of solutions to the integral equation. Theorem 1 implies that even if $h(z, W, C)$ is not unique, all solutions to Assumption 4 must result in an identical value of each conditional proportion of the principal stratification.

In the absence of unmeasured confounding, Ding et al. (2011) and Wang et al. (2017) use only one proxy variable whose distribution is informative of principal stratum G to establish nonparametric identification. When the principal strata are confounded by the unmeasured variable U, Theorem 1 shows that principal causal effects can also be identified with two proxy variables. The conditions in (4) are similar to the relevance assumption in instrumental variable analyses (Angrist et al., 1996), which requires the association between the negative control exposure A and principal stratum G. Because the weights $\omega_g(z, A, c)$'s are identified according to (3), the linear independence conditions among functions in each vector of (4) are in principle testable based on observed data.

3.2   Identification under semiparametric or parametric models

In this section, we relax Assumption 5 to some extent and discuss the identifiability of principal causal effects under semiparametric or parametric models.

Assumption 6. $(Z, W) \perp\!\!\!\perp (Y_0, Y_1) \mid (G, C, A)$,

Assumption 6 is notably weaker than Assumption 5 by requiring only one substitutional variable W for the principal stratum G, which allows negative control exposure A to directly affect the outcome. This is in parallel to the usual assumption for the identifiability of principal causal effects when unmeasured confounding is absent (Ding et al., 2011; Jiang et al., 2016; Wang et al., 2017; Luo et al., 2021). We consider a semiparametric linear model for $\mu_{z,g}(X)$ to facilitate identification of principal causal effects under Assumption 6.

Theorem 2. *Suppose that Assumptions 1–4 and 6 hold. We further assume $\mu_{z,g}(X)$ follows a linear model:*

$$\mu_{z,g}(X) = \theta_{z,g,0} + \theta_c C + \theta_a A. \tag{5}$$

Then the principal causal effects are identified if the functions in these two vectors

$$\{\eta_{s\bar{s}}(0, A, C), \eta_{\bar{s}\bar{s}}(0, A, C), A, C\}^{\mathrm{T}} \quad and \quad \{\eta_{s\bar{s}}(1, A, C), \eta_{ss}(1, A, C), A, C\}^{\mathrm{T}}$$

are respectively linearly independent.

One may further relax Assumption 6 by considering the following model,

$$\mu_{z,g}(X) = \theta_{z,g,0} + \theta_c C + \theta_a A + \theta_w W, \tag{6}$$

which allows the outcome to be affected by all observed covariates X, including the negative control intermediate variable W. The above semiparametric linear model (6) has also been considered in Ding et al. (2011) and Luo et al. (2021). As shown in the supplementary material, the parameters in (6) are identifiable under some regularity conditions. This means that we can identify the conditional outcome mean $\mu_{z,g}(X)$. However, since the proportions of principal strata $\pi_g(X)$ are not identifiable under the assumptions in Theorem 2, the parameter $\mu_{z,g}$ expressed in (1) cannot be identified unless additional conditions exist. Below we consider parametric models that would make it possible to identify the principal causal effects even if Assumption 6 were violated.

Proposition 1. *Suppose that Assumptions 1, 2(i), 3 and 4 hold. The principal stratum G follows an ordered probit model, namely,*

$$G = \begin{cases} \bar{s}\bar{s}, & \text{if } G^* + \varepsilon \le 0, \\ s\bar{s}, & \text{if } 0 < G^* + \varepsilon \le \exp(\psi_1), \\ ss, & \text{if } \exp(\psi_1) < G^* + \varepsilon, \end{cases} \tag{7}$$

where $G^* = \psi_0 + \psi_z Z + \psi_w W + \psi_a A + \psi_c C$ and $\varepsilon \sim N(0,1)$. We further assume that $W \mid Z, A, C \sim N\{m(Z,A,C), \sigma_w^2\}$ and the functions $\{1, Z, A, C, m(Z,A,C)\}^{\mathrm{T}}$ are linearly independent. Then the proportions of principal strata $\omega_g(Z,X)$ are identified for all g.

As implied by the latent ignorability assumption, the association $Z \not\perp\!\!\!\perp G \mid X$ may occur in the presence of unobserved confounder U, so the coefficient ψ_z in the model for G after (7) may not be zero. The ordinal model in (7) is compatible with monotonicity assumption and can be rewritten in the following form under this assumption:

$$\begin{aligned} \mathrm{pr}\left(S_0 = 1 \mid Z, X; \psi\right) &= \Phi\{\psi_0 - \exp(\psi_1) + \psi_z Z + \psi_w W + \psi_a A + \psi_c C\}, \\ \mathrm{pr}\left(S_1 = 1 \mid Z, X; \psi\right) &= \Phi(\psi_0 + \psi_z Z + \psi_w W + \psi_a A + \psi_c C), \end{aligned} \tag{8}$$

where $\psi = (\psi_0, \psi_1, \psi_z, \psi_w, \psi_a, \psi_c)^{\mathrm{T}}$. In fact, we model the distribution of potential values of the intermediate variable using a generalized linear model, which is similar in spirit to the marginal and nested structural mean models proposed by Robins et al. (2000). Under such parametric models, Proposition 1 shows that we can identify the conditional proportions of the principal strata $\omega_g(Z,X)$ given all observed covariates. This is a stronger result than that in Theorem 1, where only the conditional proportions of principal strata given covariates (A,C) are identifiable. With this result, we can consider another weaker version of Assumption 5, which is in parallel to Assumption 6.

Assumption 7. $(Z,A) \perp\!\!\!\perp (Y_0, Y_1) \mid (G, C, W)$.

This condition is similar to the "selection on types" assumption considered in Kédagni (2021), which entails that the negative control exposure A has no direct effect on the outcome Y. We next consider identification of principal causal effects under the ordinal model (7) and other various conditions.

Theorem 3. *Under Assumptions 1–4 and the model parameterization in Proposition 1, the following statements hold:*

(i) *with additional Assumption 5, the principal causal effects are identified if for any $C = c$, the functions in the vectors $\{\eta_{s\bar{s}}(0, A, W, c), \eta_{\bar{s}\bar{s}}(0, A, W, c)\}^{\mathrm{T}}$ and $\{\eta_{s\bar{s}}(1, A, W, c),$ $\eta_{ss}(1, A, W, c)\}^{\mathrm{T}}$ are respectively linearly independent.*

(ii) *with additional Assumption 6, the principal causal effects are identified if for any $(A, C) = (a, c)$, the functions in the vectors $\{\eta_{s\bar{s}}(0, a, W, c), \eta_{\bar{s}\bar{s}}(0, a, W, c)\}^{\mathrm{T}}$ and $\{\eta_{s\bar{s}}(1, a, W, c), \eta_{ss}(1, a, W, c)\}^{\mathrm{T}}$ are respectively linearly independent.*

(iii) *with additional Assumption 7, the principal causal effects are identified if for any $(W, C) = (w, c)$, the functions in the vectors $\{\eta_{s\bar{s}}(0, A, w, c), \eta_{\bar{s}\bar{s}}(0, A, w, c)\}^{\mathrm{T}}$ and $\{\eta_{s\bar{s}}(1, A, w, c), \eta_{ss}(1, A, w, c)\}^{\mathrm{T}}$ are respectively linearly independent.*

(iv) *with the additional model (6), the principal causal effects are identified if the functions in the vectors $\{\eta_{s\bar{s}}(0, X), \eta_{\bar{s}\bar{s}}(0, X), X\}^{\mathrm{T}}$ and $\{\eta_{s\bar{s}}(1, X), \eta_{ss}(1, X), X\}^{\mathrm{T}}$ are respectively linearly independent.*

In contrast to Theorem 1, Theorem 3 shows that under the parametric models in Proposition 1, the principal causal effects are always identifiable as long as certain linear independence conditions are satisfied, albeit Assumption 5 may partially or completely fails. Besides, since the functions $\{\eta_g(z, X); z = 0, 1; g = ss, \bar{s}\bar{s}, s\bar{s}\}$ are identifiable based on Proposition 1, those linear independence conditions in Theorem 3 are testable from observed data.

4 Estimation

While the nonparametric identification results provide useful insight, nonparametric estimation, however, is often not practical especially when the number of covariates is large due to the curse of dimensionality. We consider parametric working models for estimation of principal causal effects in this section.

Model 1 (Bridge function). *The bridge function $h(Z, W, C; \alpha)$ is known up to a finite-dimensional parameter α.*

Model 2 (Treatment and negative control intermediate variable). *The treatment model $\mathrm{pr}(Z \mid A, C; \beta)$ is known up to a finite-dimensional parameter β, and the negative control intermediate variable model $f(W \mid Z, A, C; \gamma)$ is known up to a finite-dimensional parameter γ.*

Model 3 (Outcome). *The conditional outcome mean function $\mu_{z,g}(X; \theta_{z,g})$ is known up to a finite-dimensional parameter $\theta_{z,g}$.*

Note that $\mu_{z,g}(X; \theta_{z,g})$ in Model 3 should be compatible with the requirements in Theorems 1-3. For example, under the conditions in Theorem 1 or 3(i), we consider a parametric form for $\mu_{z,g}(X; \theta_{z,g})$ that is only related to the covariate C, but should not be dependent on A and W. Given the above parameterizations in Models 1–3, we are now ready to provide a three-step procedure for estimation of the principal causal effects.

In the first step, we aim to estimate the conditional probabilities $\omega_g(Z, A, C)$ considered in Theorem 1. The expression in (3) implies that for estimation of $\omega_g(Z, A, C)$, we only need to estimate the parameters α and γ that are in the bridge function $h(Z, W, C; \alpha)$ and negative control intermediate variable model $f(W \mid Z, A, C; \gamma)$, respectively. Under Assumptions 3, 4, and the completeness condition, we have the following equation (see Lemma S2 in the supplementary material for details):

$$\mathrm{pr}(S = 1 \mid Z, A, C) = E\{h(Z, W, C; \alpha) \mid Z, A, C\}.$$

We then obtain an estimator $\widehat{\alpha}$ by solving the following estimating equations

$$\mathbb{P}_n[\{S - h(Z, W, C; \alpha)\}B(Z, A, C)] = 0, \tag{9}$$

where $\mathbb{P}_n(\xi) = \sum_{i=1}^n \xi_i / n$ for some generic variable ξ, and $B(Z, A, C)$ is an arbitrary vector of functions with dimension no smaller than that of α. If the dimension of the user-specified function $B(Z, A, C)$ is larger than that of α, we may adopt the generalized method of moments (Hansen, 1982) to estimate α. We next obtain the estimators $\widehat{\beta}$ and $\widehat{\gamma}$ in Model 2 by maximum likelihood estimation. With the parameter estimates $\widehat{\alpha}$ and $\widehat{\gamma}$, we can finally obtain the estimators $\widehat{\omega}_g(Z, A, C)$ based on (3). The calculation of the estimated probabilities involves integral equations with respect to the distribution $f(W \mid Z, A, C; \widehat{\gamma})$, which may be numerically approximated to circumvent computational difficulties. Consequently, we have a plug-in estimator $\widehat{\eta}_g(z, A, C)$ of $\eta_g(z, A, C)$ defined in (2) and an estimator of $\pi_g(A, C)$ as follows:

$$\widehat{\pi}_g(A, C) = \sum_{z=0}^1 \widehat{\omega}_g(z, A, C)\mathrm{pr}(z \mid A, C; \widehat{\gamma}).$$

In addition, if the model assumptions in Proposition 1 hold, we can further estimate the conditional probabilities given fully observed covariates $\omega_g(Z, X)$ and $\pi_g(X)$; the estimation details are relegated to the supplementary material for space cosiderations. As noted by Theorems 1–3, the estimation of principal causal effects requires different conditional probabilities of principal strata, depending on which assumptions are imposed. For example, Theorem 1 requires $\omega_g(Z, A, C)$, whereas Theorem 3 requires $\omega_g(Z, X)$. For simplicity, we denote these conditional probabilities by a unified notation $\omega_g(Z, V)$ with $V = (A, C^\mathrm{T})^\mathrm{T}$ in Theorems 1–2 and $V = X$ in Theorem 3. The notations $\eta_g(z, V)$ and $\pi_g(V)$ are equipped with similar meanings.

In the second step, we aim to estimate the parameters $\theta_{z,g}$ for $z = 0, 1$ and $g = ss, \bar{s}\bar{s}, s\bar{s}$ in the outcome Model 3. To derive an estimator for $\theta_{0,g}$, we observe the following moment constraints by invoking the monotonicity assumption:

$$
\begin{aligned}
&E\{Y - \mu_{0,ss}(X; \theta_{0,ss}) \mid Z = 0, S = 1, X\} = 0, \\
&E\{Y - \textstyle\sum_{g=s\bar{s},\bar{s}\bar{s}} \widehat{\eta}_g(0, V)\mu_{0,g}(X; \theta_{0,g}) \mid Z = 0, S = 0, X\} = 0.
\end{aligned} \tag{10}
$$

We emphasize here that the specifications of $\mu_{z,g}(X; \theta_{z,g})$ in Model 3 may not always depend on all the observed covariates X due to identifiability concerns; see also the discussions below Model 3. With the above moment constraints, we can apply the generalized method of moments again to obtain a consistent estimator for $\theta_{0,g}$. The estimation of $\theta_{1,g}$ is similar, because we have another pair of moment constraints:

$$
\begin{aligned}
&E\{Y - \mu_{1,\bar{s}\bar{s}}(X; \theta_{1,\bar{s}\bar{s}}) \mid Z = 1, S = 0, X\} = 0, \\
&E\{Y - \textstyle\sum_{g=ss,s\bar{s}} \widehat{\eta}_g(1, V)\mu_{1,g}(X; \theta_{1,g}) \mid Z = 1, S = 1, X\} = 0.
\end{aligned} \tag{11}
$$

Finally, in view of (1), we can obtain our proposed estimator for the principal causal effect as follows:
$$\widehat{\Delta}_g = \frac{\mathbb{P}_n\{\mu_{1,g}(X;\widehat{\theta}_{1,g})\widehat{\pi}_g(V)\}}{\mathbb{P}_n\{\widehat{\pi}_g(V)\}} - \frac{\mathbb{P}_n\{\mu_{0,g}(X;\widehat{\theta}_{0,g})\widehat{\pi}_g(V)\}}{\mathbb{P}_n\{\widehat{\pi}_g(V)\}}.$$

Using empirical process theories, one can show that the resulting estimator $\widehat{\Delta}_g$ is consistent and asymptotically normally distributed.

5   Simulation studies

We conduct simulation studies to investigate the finite sample performance of the proposed estimators in this section. We consider the following data-generating mechanism:

(a). We generate covariates (A, C) from $\quad (A,C)^{\mathrm{T}} \sim N\left\{(\delta_a, \delta_c)^{\mathrm{T}}, \begin{pmatrix} \sigma_a^2 & \rho_1\sigma_a\sigma_c \\ \rho_1\sigma_a\sigma_c & \sigma_c^2 \end{pmatrix}\right\}$.

(b). We generate the binary treatment Z from a Bernoulli distribution with $\mathrm{pr}(Z = 1 \mid A, C) = \Phi\left(\beta_0 + \beta_a A + \beta_c C\right)$.

(c). Given $(Z, A, C)^{\mathrm{T}}$, we generate (U, W) from the following joint normal distribution
$$(U, W)^{\mathrm{T}} \mid Z, A, C \sim N\left\{\begin{pmatrix} \iota_0 + \iota_z Z + \iota_a A + \iota_{c1}C + \iota_{c2}C^2 \\ \gamma_0 + \gamma_z Z + \gamma_a A + \gamma_{c1}C + \gamma_{c2}C^2 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \rho_2\sigma_u\sigma_w \\ \rho_2\sigma_u\sigma_w & \sigma_w^2 \end{pmatrix}\right\}.$$

To guarantee $W \perp\!\!\!\perp (Z, A) \mid (U, C)$, we set $\gamma_a = \iota_a\sigma_w\rho_2/\sigma_u$ and $\gamma_z = \iota_z\sigma_w\rho_2/\sigma_u$. For simplicity, we assume that $E(U \mid Z, A, W, C)$ is linear in C by setting $\gamma_{c2} = \iota_{c2}\sigma_w/\sigma_u\rho_2$.

(d). Define $G^\dagger = \zeta_0 + \zeta_w W + \zeta_u U + \zeta_c C$, and we generate the principal stratum G from the following ordered probit model:
$$\mathrm{pr}(G = \bar{s}\bar{s} \mid U, A, W, C) = \Phi(-G^\dagger),$$
$$\mathrm{pr}(G = s\bar{s} \mid U, A, W, C) = \Phi\{\exp(\zeta_1) - G^\dagger\} - \Phi(-G^\dagger),$$
$$\mathrm{pr}(G = ss \mid U, A, W, C) = \Phi\{G^\dagger - \exp(\zeta_1)\}.$$

(e). The outcome Y is finally generated from the following conditional normal distribution:
$$Y \mid (Z = z, G = g, A, W, C) \sim N(\theta_{z,g,0} + \theta_a A + \theta_w W + \theta_c C, \sigma_y^2).$$

The true values of parameters are set as follows:

(a). $\delta_a = 0$, $\delta_c = 0$, $\sigma_a = 0.5$, $\rho_1 = 0.5$, $\sigma_c = 0.5$.

(b). $\beta_0 = 0$, $\beta_a = 1$, $\beta_c = 1$.

(c). $\iota_0 = 1$, $\iota_z = 1$, $\iota_a = 1.5$, $\iota_{c1} = 1.5$, $\iota_{c2} = -0.75$, $\gamma_0 = 1$, $\gamma_z = 0.5$, $\gamma_a = 0.75$, $\gamma_{c1} = 1.5$, $\gamma_{c2} = -1.5$, $\sigma_u = 0.5$, $\rho_2 = 0.5$, $\sigma_w = 0.5$.

Table 1: Simulation studies with bias ($\times 100$), standard error ($\times 100$) and 95% coverage probability ($\times 100$) for various settings and sample sizes.

n	Case	(θ_a, θ_w)	Δ_{ss}			$\Delta_{s\bar{s}}$			$\Delta_{\bar{s}\bar{s}}$		
						$\zeta_u = 0.2$					
			Bias	Sd	CP	Bias	Sd	CP	Bias	Sd	CP
1000	(i)	$(0,0)$	−0.6	7.5	96.2	2.4	39.8	95.8	−2.8	28.6	95.0
	(ii)	$(1,0)$	−0.5	8.0	95.8	0.1	46.3	96.4	−0.7	21.9	95.8
	(iii)	$(0,1)$	−1.0	12.5	95.6	7.8	48.5	94.0	−3.3	21.6	94.4
	(iv)	$(1,1)$	−0.2	12.6	95.8	1.1	50.3	96.8	−0.9	23.4	95.2
			Bias	Sd	CP	Bias	Sd	CP	Bias	Sd	CP
5000	(i)	$(0,0)$	−0.4	3.5	94.4	1.6	18.9	96.6	0.7	12.9	94.8
	(ii)	$(1,0)$	−0.4	3.8	94.6	1.0	21.9	96.4	0.3	9.6	95.8
	(iii)	$(0,1)$	−1.2	5.3	95.8	4.9	20.8	96.0	−0.1	9.5	95.6
	(iv)	$(1,1)$	−1.0	5.3	96.4	3.7	21.8	96.0	0.3	10.1	96.6
						$\zeta_u = 0.5$					
			Bias	Sd	CP	Bias	Sd	CP	Bias	Sd	CP
1000	(i)	$(0,0)$	−0.5	5.5	95.2	8.2	44.8	93.6	−6.6	25.0	94.4
	(ii)	$(1,0)$	−0.5	5.9	95.8	8.3	51.0	94.8	−3.2	19.0	94.8
	(iii)	$(0,1)$	−0.6	6.8	96.6	11.1	47.0	93.8	−3.5	20.9	93.4
	(iv)	$(1,1)$	−0.5	6.8	96.4	6.4	51.0	96.2	−2.2	22.4	95.4
			Bias	Sd	CP	Bias	Sd	CP	Bias	Sd	CP
5000	(i)	$(0,0)$	−0.4	2.6	95.0	5.1	21.8	94.4	−0.1	11.8	95.4
	(ii)	$(1,0)$	−0.4	2.8	94.6	4.8	24.6	94.4	−0.2	8.6	95.6
	(iii)	$(0,1)$	−0.7	3.2	95.2	7.1	21.6	94.0	−0.4	8.8	94.4
	(iv)	$(1,1)$	−0.6	3.2	95.2	6.0	23.2	94.2	0.0	9.3	95.4

Sd: empirical standard error.   CP: 95% coverage probability.

(d). $\zeta_0 = 0.5$, $\zeta_1 = 0$, $\zeta_w = 0.5$, $\zeta_c = 1$. Since ζ_u controls for magnitude of unobserved confounding, we consider 6 different values, i.e., $\zeta_u \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5\}$.

(e). $\theta_{0,\bar{s}\bar{s},0} = 0$, $\theta_{0,s\bar{s},0} = 1$, $\theta_{0,ss,0} = 2$, $\theta_{1,\bar{s}\bar{s},0} = 2$, $\theta_{1,s\bar{s},0} = 3$, $\theta_{1,ss,0} = 4$, $\theta_c = 1$, $\sigma_y = 0.5$. We consider 4 settings for (θ_a, θ_w): $(0,0)$, $(1,0)$, $(0,1)$ and $(1,1)$, which corresponds to different identifying assumptions.

Under the above data generating mechanism, the bridge function with the following form is compatible with Assumption 4 (see supplementary materials for details):

$$h(Z, W, C) = \Phi\left\{\alpha_0 + \exp(\alpha_1)Z + \alpha_w W + \alpha_{c1}C + \alpha_{c2}C^2\right\}.$$

We thus model the bridge function in Model 1 with this parametric form and specify all correct parametric models in Model 2. We estimate the bridge function by solving the estimating equation (9) with the user-specified functions $B(Z, A, C) = \{1, A, Z, C, C^2\}^{\mathrm{T}}$. It is worth pointing out that our data-generating mechanism also satisfies the model assumptions in Proposition 1, and we can consistently estimate the probabilities $\omega_g(Z, X)$

and $\pi_g(X)$ using the method given in the supplementary material. We investigate the performance of the proposed estimators under various values of (θ_a, θ_w), which represent different conditional independence conditions between (A, W) and the outcome Y. For the four different settings in (e), we consider estimation of principal causal effects with four different correct parametric forms in Model 3, respectively. For example, the setting $(\theta_a, \theta_w) = (0, 0)$ implies that Assumption 5 holds, and we specify the working model $\mu_{z,g}(X; \theta_{zg}) = \theta_{z,g,0} + \theta_c C$; if $(\theta_a, \theta_w) = (1, 1)$, then the outcome can be affected by all covariates, and we employ the linear model given in (6). For simplicity, we refer to these four different estimation procedures as cases (i)–(iv), respectively.

For each value of ζ_u, we consider sample size $n = 1000$ and $n = 5000$. Table 1 reports the bias, standard error and coverage probabilities of 95% confidence intervals averaged across 500 replications with $\zeta_u = 0.2$ and 0.5. The corresponding results for other values of ζ_u are provided in the supplementary material. The results in all the settings are similar. It can be found that our method has negligible biases with smaller variances as the sample size increases. Estimators of Δ_{ss} and $\Delta_{\bar{s}\bar{s}}$ are more stable than that of $\Delta_{s\bar{s}}$. This may be because the estimation of $\Delta_{s\bar{s}}$ requires solving the joint estimating equations (10) and (11) rather than only one of them. The proposed estimators have coverage probabilities close to the nominal level in all scenarios. All these results demonstrate the consistency of our proposed estimators.

6   Application to Return to Schooling

We illustrate our approach by reanalyzing the dataset from the National Longitudinal Survey of Young Men (Card, 1993; Tan, 2006). This cohort study includes 3,010 men who were aged 14-24 when first interviewed in 1966, with follow-up surveys continued until 1976. We are interested in estimating the causal effect of education on earnings, which might be confounded by unobserved preferences for students' abilities and family costs (Kédagni, 2021).

The treatment Z is an indicator of living near a four-year college. Following Tan (2006), we choose the educational experience beyond high school as the intermediate variable S. The outcome Y is the log wage in the year 1976, ranging from 4.6 to 7.8. We consider the average parental education years as the negative control exposure A, because parents' education years are highly correlated with whether their children have the chance to live close to a college. We use the intelligence quotient (IQ) scores as the negative control intermediate variable W, because IQ is related to students' learning abilities, and students with higher IQ are more likely to enter college. The data set also includes the following covariates C: race, age, scores on the Knowledge of the World of Work test, a categorical variable indicating whether children living with both parents, single mom, or step-parents, and several geographic variables summarizing living areas in the past. The missing covariates are imputed via the k-Nearest Neighbor algorithm with $k = 10$ (Franzin et al., 2017).

Monotonicity is plausible because living near a college would make an individual more likely to receive higher education. Following Jiang et al. (2022), we do not invoke the exclusion restriction assumption that living near a college can affect the earnings only through education. In fact, we can evaluate the validity of this assumption by applying

Table 2: Analysis of the National Longitudinal Survey of Young Men.

Case	Δ_{ss}			$\Delta_{s\bar{s}}$			$\Delta_{\bar{s}\bar{s}}$		
(i)	0.07	(−0.07,	0.81)	−0.68	(−1.85,	−0.25)	−0.86	(−2.83,	−0.32)
(ii)	0.06	(−0.09,	0.77)	0.13	(−0.45,	0.59)	0.02	(−0.56,	0.16)
(iii)	−0.18	(−0.55,	0.80)	0.87	(0.05,	1.93)	−0.07	(−0.91,	0.23)
(iv)	−0.18	(−0.56,	0.78)	0.85	(0.00,	1.88)	−0.07	(−0.89,	0.26)

the proposed approach in this paper. We employ similar model parameterizations as used in simulation studies, and our analyses here are also conducted under the cases (i)–(iv) that represent different conditional independence assumptions for the outcome model.

Table 2 shows the point estimates and their associated 95% confidence intervals obtained via the nonparametric bootstrap method. We first observe that the results in cases (iii) and (iv) are very close. Compared with them, the corresponding results in cases (i) and (ii) are completely different. Because the outcome model in cases (i) and (ii) do not include the proxy variable W as a predictor, the empirical findings may indicate misspecifications of outcome models in these two cases. Thus, the results in cases (iii) and (iv), where the IQ score W is allowed to directly affect the wage Y, are more credible. Based on these results, we find that both the 95% confidence intervals for Δ_{ss} and $\Delta_{\bar{s}\bar{s}}$ cover zero, which implies no significant evidence of violating the exclusion restriction. The estimate of $\Delta_{s\bar{s}}$ is positive and its corresponding confidence interval does not cover zero. This implies that education has a significantly positive effect on earnings, which is consistent with previous analyses (Tan, 2006; Jiang et al., 2022; Kédagni, 2021).

7   Discussion

With the aid of a pair of negative controls, we have established identification and estimation of principal causal effects when the treatment and principal strata are confounded by unmeasured variables. The availability of negative control variables is crucial for the proposed approach. Although it is generally not possible to test the negative control assumptions via observed data without additional assumptions, the existence of such variables is practically reasonable in the empirical example presented in this paper and similar situations where two or more proxies of unmeasured variables may be available (Miao et al., 2018; Shi et al., 2020; Miao et al., 2020; Cui et al., 2020).

The proposed methods may be improved or extended in several directions. First, we consider parametric methods to solve integral equations involved in our estimation procedure. One may also consider nonparametric estimation techniques to obtain the solutions (Newey and Powell, 2003; Chen and Pouzo, 2012; Li et al., 2021). Second, we relax the commonly-used ignorability assumption by allowing unmeasured confounders between the treatment and principal strata, and it is possible to further relax this assumption and consider the setting where Assumption 2(ii) fails. Third, our identifiability results rely on the monotonicity assumption which may not hold in some real applications. In principle, one can conduct sensitivity analysis to assess the principal causal effects of violations of monotonicity assumption (Ding and Lu, 2017). Finally, it is also of interest to develop doubly robust estimators for the principal causal effects as provided by Cui et al. (2020)

for average treatment effects. The study of these issues is beyond the scope of this paper and we leave them as future research topics.

References

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455.

Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling. Technical report, National Bureau of Economic Research.

Carrasco, M., Florens, J.-P., and Renault, E. (2007). Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. *Handbook of Econometrics*, 6:5633–5751.

Chen, X. and Pouzo, D. (2012). Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Econometrica*, 80(1):277–321.

Cui, Y., Pu, H., Shi, X., Miao, W., and Tchetgen Tchetgen, E. J. (2020). Semiparametric proximal causal inference. *arXiv*: 2011.08411.

Deng, Y., Guo, Y., Chang, Y., and Zhou, X.-H. (2021). Identification and estimation of the heterogeneous survivor average causal effect in observational studies. *arXiv*: 2109.13623v3.

Ding, P., Geng, Z., Yan, W., and Zhou, X.-H. (2011). Identifiability and estimation of causal effects by principal stratification with outcomes truncated by death. *Journal of the American Statistical Association*, 106(496):1578–1591.

Ding, P. and Lu, J. (2017). Principal stratification analysis using principal scores. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3):757–777.

Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, 58(1):21–29.

Franzin, A., Sambo, F., and Di Camillo, B. (2017). Bnstruct: an r package for bayesian network structure learning in the presence of missing data. *Bioinformatics*, 33(8):1250–1252.

Gilbert, P. B. and Hudgens, M. G. (2008). Evaluating candidate principal surrogate endpoints. *Biometrics*, 64(4):1146–1154.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50:1029–1054.

Jiang, Z., Ding, P., and Geng, Z. (2016). Principal causal effect identification and surrogate end point evaluation by multiple trials. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(4):829–848.

Jiang, Z., Yang, S., and Ding, P. (2022). Multiply robust estimation of causal effects under principal ignorability. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. doi:10.1111/rssb.12538.

Kédagni, D. (2021). Identifying treatment effects in the presence of confounded types. *Journal of Econometrics*, doi: 10.1016/j.jeconom.2021.01.012.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, 76(3):1071–1102.

Li, W., Miao, W., and Tchetgen Tchetgen, E. J. (2021). Nonparametric inference about mean functionals of nonignorable nonresponse data without identifying the joint distribution. *arXiv*: 2110.05776v2.

Long, D. M. and Hudgens, M. G. (2013). Sharpening bounds on principal effects with covariates. *Biometrics*, 69(4):812–819.

Luo, S., Li, W., and He, Y. (2021). Causal inference with outcomes truncated by death in multiarm studies. *Biometrics*. doi: 10.1111/biom.13554.

Miao, W., Geng, Z., and Tchetgen Tchetgen, E. J. (2018). Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, 105(4):987–993.

Miao, W., Shi, X., and Tchetgen Tchetgen, E. J. (2020). A confounding bridge approach for double negative control inference on causal effects. *arXiv*: 1808.04945.

Newey, W. K. and Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5):1565–1578.

Robins, J. M., Hernan, M. A., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560.

Rubin, D. B. (1990). Comment: Neyman (1923) and causal inference in experiments and observational studies. *Statistical Science*, 5(4):472–480.

Rubin, D. B. et al. (2006). Causal inference through potential outcomes and principal stratification: application to studies with "censoring" due to death. *Statistical Science*, 21(3):299–309.

Schwartz, S., Li, F., and Reiter, J. P. (2012). Sensitivity analysis for unmeasured confounding in principal stratification settings with binary variables. *Statistics in Medicine*, 31(10):949–962.

Shi, X., Miao, W., Nelson, J. C., and Tchetgen Tchetgen, E. J. (2020). Multiply robust causal inference with double-negative control adjustment for categorical unmeasured confounding. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(2):521–540.

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

Tchetgen Tchetgen, E. J. (2014). Identification and estimation of survivor average causal effects. *Statistics in Medicine*, 33(21):3601–3628.

VanderWeele, T. J. (2011). Principal stratification–uses and limitations. *The International Journal of Biostatistics*, 7(1):1–14.

Wang, L., Zhou, X.-H., and Richardson, T. S. (2017). Identification and estimation of causal effects with outcomes truncated by death. *Biometrika*, 104(3):597–612.

Zhang, J. L. and Rubin, D. B. (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by "death". *Journal of Educational and Behavioral Statistics*, 28(4):353–368.

Zhang, J. L., Rubin, D. B., and Mealli, F. (2009). Likelihood-based analysis of causal effects of job-training programs using principal stratification. *Journal of the American Statistical Association*, 104(485):166–176.

Supplementary Material

In the supplementary material, we provide proofs of theorems and claims in the main paper. We also provide additional details for the estimation and simulation studies.

S1 Proofs of propositions and theorems

S1.1 The proof of expression (1)

Proof. By the Law of Iterated Expectation (LIE), we have

$$
\begin{aligned}
E(Y_z \mid G = g) &= E\left\{E(Y_z \mid G = g, X) \mid G = g\right\} \\
&= \int E(Y_z \mid G = g, X) f(x \mid G = g)\, \mathrm{d}\,x \\
&= \int E(Y_z \mid G = g, X)\mathrm{pr}(G = g \mid X = x) f(x)\, \mathrm{d}\,x / \mathrm{pr}(G = g) \\
&= E\{E(Y_z \mid G = g, X)\pi_g(X)\}/E\{\pi_g(X)\}.
\end{aligned}
\tag{S12}
$$

Given Assumption 2(ii), we have

$$
E(Y_z \mid G = g, X) = E(Y_z \mid Z = z, G = g, X) = E(Y \mid Z = z, G = g, X) = \mu_{z,g}(X).
$$

Combining these two pieces, we have

$$
E(Y_z \mid G = g) = E\{\mu_{z,g}(X)\pi_g(X)\}/E\{\pi_g(X)\}.
$$

$\square$

S1.2 Lemmas

We first prove the first conclusion in Theorem 1, which is summarized as the following lemma.

Lemma S1. *Suppose that Assumptions 1, 2(i), 3 and 4 hold. Then the conditional probabilities of principal strata are identified by*

$$
\begin{aligned}
\omega_{ss}(Z, A, C) &= E\{h(0, W, C) \mid Z, A, C\}, \\
\omega_{\bar{s}\bar{s}}(Z, A, C) &= 1 - E\{h(1, W, C) \mid Z, A, C\}, \\
\omega_{s\bar{s}}(Z, A, C) &= 1 - \omega_{ss}(Z, A, C) - \omega_{\bar{s}\bar{s}}(Z, A, C).
\end{aligned}
$$

Proof. Given the equality $\mathrm{pr}(S = 1 \mid Z = z, C, U) = E\{h(z, W, C) \mid Z = z, C, U)\}$, for any $Z = z'$, we have that

$$
\begin{aligned}
\mathrm{pr}(S_z = 1 \mid Z = z', A, C) &= E\left\{E(S_z \mid Z = z', A, C, U) \mid Z = z', A, C\right\} \\
&= E\left\{E(S_z \mid Z = z, C, U) \mid Z = z', A, C\right\} \\
&= E\left\{E(S \mid Z = z, C, U) \mid Z = z', A, C\right\} \\
&= E\left\{E\{h(z, W, C) \mid C, U\} \mid Z = z', A, C\right\} \\
&= E\left\{E\{h(z, W, C) \mid Z = z', A, C, U\} \mid Z = z', A, C\right\} \\
&= E\{h(z, W, C) \mid Z = z', A, C\}.
\end{aligned}
\tag{S13}
$$

where the first and final equalities are due to LIE, the second and fifth equalities are due to Assumption 3, the third equality is due to consistency, the forth equality is due to Assumption 4.

Given monotonicity assumption 1, we have the equivalence of $\{G = ss\}$ and $\{S_0 = 1\}$ as well as $\{G = \bar{s}\bar{s}\}$ and $\{S_1 = 0\}$, namely

$$\text{pr}(G = ss \mid Z, A, C) = \text{pr}(S_0 = 1 \mid Z, A, C) = E\{h(0, W, C) \mid Z, A, C\},$$
$$\text{pr}(G = \bar{s}\bar{s} \mid Z, A, C) = \text{pr}(S_1 = 0 \mid Z, A, C) = 1 - E\{h(1, W, C) \mid Z, A, C\},$$
$$\text{pr}(G = s\bar{s} \mid Z, A, C) = E\{h(1, W, C) \mid Z, A, C\} - E\{h(0, W, C) \mid Z, A, C\}.$$

$\square$

Lemma S2. *Suppose that Assumptions 2(i), 3 and 4 hold. We also assume that for any c, z and square-integrable function g, $E\{g(U) \mid Z = z, A, C = c\} = 0$ almost surely if and only if $g(U) = 0$ almost surely. Then any function h satisfying*

$$\text{pr}(S = 1 \mid Z, A, C) = E\{h(Z, W, C) \mid Z, A, C\} \tag{S14}$$

is also a valid outcome bridge function in Assumption 4.

Proof. Given (S14), for any c, z, we have that

$$
\begin{aligned}
0 &= E\{S - h(z, W, c) \mid Z = z, A, C = c\} \\
&= E\left[E\{S - h(z, W, c) \mid Z = z, A, C = c, U\} \mid Z = z, A, C = c\right] \\
&= E\left[E\{S - h(z, W, c) \mid Z = z, C = c, U\} \mid Z = z, A, C = c\right],
\end{aligned}
$$

where the second equality is due to the LIE, and the last equality is due to Assumption 3. Given the completeness condition, we have that

$$E\{S - h(z, W, c) \mid Z = z, C = c, U\} = 0,$$

which indicates that any function $h(Z, W, C)$ that solves equation (S14) also satisfies Assumption 4. $\square$

S1.3 The proof of Theorems 1-2

We next prove the second conclusion in Theorem 1 and Theorem 2.

Proof. Given the conditions in Lemma S1, we know that the weights $\omega_g(Z, A, C)$ and $\eta_g(Z, A, C)$ are identifiable for all g. Under monotonicity assumption, the causal estimands $\mu_{1,\bar{s}\bar{s}}(X)$ and $\mu_{0,ss}(X)$ can be identified by

$$\mu_{0,ss}(X) = E(Y \mid Z = 0, S = 1, X), \quad \mu_{1,\bar{s}\bar{s}}(X) = E(Y \mid Z = 1, S = 0, X).$$

We next show that $\mu_{1,ss}(X)$ and $\mu_{1,s\bar{s}}(X)$ are also identifiable. For simplicity, we omit the proof of $\mu_{0,\bar{s}\bar{s}}(X)$ and $\mu_{0,s\bar{s}}(X)$. Applying LIE to get

$$E(Y \mid Z = 1, S = 1, A, C) = \eta_{ss}(1, A, C)\mu_{1,ss}(A, C) + \eta_{s\bar{s}}(1, A, C)\mu_{1,s\bar{s}}(A, C). \tag{S15}$$

1. Given additional Assumptions 2(ii) and 5, we have

$$\mu_{1,g}(A,C) = E(Y \mid Z = 1, G = g, A, C) = E(Y \mid Z = 1, G = g, C) = \mu_{1,g}(C).$$

Therefore, (S15) can be simplified as

$$E(Y \mid Z = 1, S = 1, A, C) = \eta_{ss}(1, A, C)\mu_{1,ss}(C) + \eta_{s\bar{s}}(1, A, C)\mu_{1,s\bar{s}}(C).$$

For any $C = c$, if $\{\eta_{ss}(1, A, c), \eta_{s\bar{s}}(1, A, c)\}^{\mathrm{T}}$ is linearly independent, we can identify $\mu_{1,ss}(C)$ and $\mu_{1,s\bar{s}}(C)$. Thus, $\mu_{1,ss}(A, C)$ and $\mu_{1,s\bar{s}}(A, C)$ are also identifiable.

2. Given the conditions in Theorem 2, (S15) can be simplified as

$$E(Y \mid Z = 1, S = 1, A, C) = \theta_{1,ss,0}\eta_{ss}(1, A, C) + \theta_{1,s\bar{s},0}\eta_{s\bar{s}}(1, A, C) + \theta_a A + \theta_c C,$$

If the functions $\{\eta_{ss}(1, A, C), \eta_{s\bar{s}}(1, A, C), A, C\}^{\mathrm{T}}$ is linearly independent, we then can identify $\theta_{1,s\bar{s}}$, $\theta_{1,ss}$, θ_a and θ_c. Thus, $\mu_{1,ss}(A, C)$ and $\mu_{1,s\bar{s}}(A, C)$ are also identifiable.

Given the identifiability of $\mu_{1,ss}(X)$ and $\mu_{1,s\bar{s}}(X)$, we show that $\mu_{1,ss}$ and $\mu_{1,s\bar{s}}$ can be identified from the view of (S12). Specifically, we have

$$
\begin{aligned}
E(Y_1 \mid G = g) &= E\left\{E(Y_1 \mid G = g, A, C) \mid G = g\right\} \\
&= \int E(Y_1 \mid G = g, A, C)f(a, c \mid G = g)\,\mathrm{d}\,x \\
&= \int E(Y_1 \mid G = g, a, c)\mathrm{pr}(G = g \mid A = a, C = c)f(a, c)\,\mathrm{d}\,a\,\mathrm{d}\,c/\mathrm{pr}(G = g) \\
&= E\{E(Y_1 \mid G = g, A, C)\pi_g(A, C)\}/E\{\pi_g(A, C)\}
\end{aligned}
$$

for $g \in \{ss, s\bar{s}\}$. Moreover, given Assumptions 2(ii) and 5 or 6, we have

$$E(Y_1 \mid G = g, A, C) = E(Y \mid Z = 1, G = g, A, C) = \mu_{1,g}(A, C)$$

Therefore, for $g \in \{ss, s\bar{s}\}$, we have

$$\mu_{1,g} = \frac{E\{\mu_{1,g}(A, C)\pi_g(A, C)\}}{E\{\pi_g(A, C)\}}.$$

Similarly, $\mu_{0,\bar{s}\bar{s}}$ and $\mu_{0,s\bar{s}}$ are also identifiable. □

S1.4 The identifiability of (6)

In this section, we show the identifiability of (6).

Proof. By the LIE, we have

$$
\begin{aligned}
E(Y \mid Z &= 1, S = 1, A, C) \\
&= \eta_{ss}(1, A, C)\mu_{1,ss}(A, C) + \eta_{s\bar{s}}(1, A, C)\mu_{1,s\bar{s}}(X) \\
&= \eta_{ss}(1, A, C)E\{\mu_{1,ss}(X) \mid Z = 1, G = ss, A, C\} \\
&\quad + \eta_{s\bar{s}}(1, A, C)E\{\mu_{1,s\bar{s}}(X) \mid Z = 1, G = s\bar{s}, A, C\} \\
&= \eta_{ss}(1, A, C)\theta_{1,ss,0} + \eta_{s\bar{s}}(1, A, C)\theta_{1,s\bar{s},0} + \theta_c C + \theta_a A \\
&\quad + \theta_w E(W \mid Z = 1, G = ss, A, C)\eta_{ss}(1, A, C) \\
&\quad + \theta_w E(W \mid Z = 1, G = s\bar{s}, A, C)\eta_{s\bar{s}}(1, A, C) \\
&= \eta_{ss}(1, A, C)\theta_{1,ss,0} + \eta_{s\bar{s}}(1, A, C)\theta_{1,s\bar{s},0} + \theta_c C + \theta_a A \\
&\quad + \theta_w E(W \mid Z = 1, S = 1, A, C).
\end{aligned}
$$

20

Given the conditions in Lemma S1, we know that the proportions of principal strata $\eta_g(Z, A, C)$ for all g are identifiable. If the functions

$$\{\eta_{ss}(1, A, C), \eta_{s\bar{s}}(1, A, C), A, C, E(W \mid Z = 1, S = 1, A, C)\}^{\mathrm{T}}$$

is linearly independen, we then can identify $\theta_{1,s\bar{s},0}$, $\theta_{1,ss,0}$, θ_a, θ_c and θ_w. Subsequently, $\mu_{1,ss}(X)$ and $\mu_{1,s\bar{s}}(X)$ are also identifiable. Similarly, $\mu_{0,\bar{s}\bar{s}}(X)$ and $\mu_{0,s\bar{s}}(X)$ are also identifiable. □

S1.5   The proof of Proposition 1

Proof. First, we can identify $m(Z, A, C)$ and σ_w^2 from the observed data based on normal distribution. Next, we consider the identifiablity of ψ_1, ψ_0, ψ_a, ψ_c and ψ_w. Given equation (7) in main text, we have that

$$\begin{aligned}
\mathrm{pr}(G = \bar{s}\bar{s} \mid Z, X) &= \Phi\left(-G^*\right), \\
\mathrm{pr}(G = s\bar{s} \mid Z, X) &= \Phi\left\{\exp(\psi_1) - G^*\right\} - \Phi\left(-G^*\right), \\
\mathrm{pr}(G = ss \mid Z, X) &= \Phi\left\{G^* - \exp(\psi_1)\right\},
\end{aligned}$$

where $G^* = \psi_0 + \psi_z Z + \psi_w W + \psi_a A + \psi_c C$. The above equalities indicate (8) holds. According to Lemma S1, we can nonparametriclly identify the distribution $\mathrm{pr}(S_t = 1 \mid Z, A, C)$. Also,

$$\int \mathrm{pr}(S_t = 1 \mid Z, X) f(W \mid Z, A, C) \, \mathrm{d}W$$

$$= \int \Phi\left\{\psi_0 + \exp(\psi_1)(t - 1) + \psi_z Z + \psi_a A + \psi_c C + \psi_w W\right\} f(W \mid Z, A, C) \, \mathrm{d}W$$

$$= \Phi\left\{\frac{\psi_0 + \exp(\psi_1)(t - 1) + \psi_z Z + \psi_a A + \psi_c C + \psi_w E(W \mid Z, A, C)}{\sqrt{1 + \psi_w^2 \sigma_w^2}}\right\}$$

$$= \Phi\left\{\frac{\psi_0 + \exp(\psi_1)(t - 1) + \psi_z Z + \psi_a A + \psi_c C + \psi_w m(Z, A, C)}{\sqrt{1 + \psi_w^2 \sigma_w^2}}\right\}$$

$$= \mathrm{pr}(S_t = 1 \mid Z, A, C).$$

Since $\{1, Z, A, C, E(W \mid Z, A, C)\}^{\mathrm{T}}$ are linearly independent, then we can use probit regression to identify all the parameters. □

S1.6   The proof of Theorem 3

Proof. Given the conditions in Proposition 1, we know that the proportions of principal strata $\omega_g(Z, X)$ for all g are identifiable. Also, under Assumption 1, we know that the causal estimands $\mu_{0,ss}(X)$ or $\mu_{0,\bar{s}\bar{s}}(X)$ can be identified by

$$\mu_{0,ss}(X) = E(Y \mid Z = 0, S = 1, X), \quad \mu_{1,\bar{s}\bar{s}}(X) = E(Y \mid Z = 1, S = 0, X).$$

We then show that $\mu_{1,ss}(X)$ and $\mu_{1,s\bar{s}}(X)$ are also identifiable. We omit the proof of $\mu_{0,\bar{s}\bar{s}}(X)$ and $\mu_{0,s\bar{s}}(X)$ due to the similarity. Applying LIE to get

$$E(Y \mid Z = 1, S = 1, X) = \eta_{ss}(1, X)\mu_{1,ss}(X) + \eta_{s\bar{s}}(1, X)\mu_{1,s\bar{s}}(X). \tag{S16}$$

1. Given the conditions in Theorem 3(i), we have

$$\mu_{1,g}(X) = E(Y \mid Z = 1, G = g, X) = E(Y \mid Z = 1, G = g, C) = \mu_{1,g}(C).$$

Therefore, equation (S16) can be simplified as

$$E(Y \mid Z = 1, S = 1, X) = \eta_{ss}(1, X)\mu_{1,ss}(C) + \eta_{s\bar{s}}(1, X)\mu_{1,s\bar{s}}(C).$$

For any $C = c$, the vector function $\{\eta_{ss}(1, A, W, c), \eta_{s\bar{s}}(1, A, W, c)\}^{\mathrm{T}}$ is linearly independent, we then can identify $\mu_{1,ss}(C)$ and $\mu_{1,s\bar{s}}(C)$. Thus, $\mu_{1,ss}(X)$ and $\mu_{1,s\bar{s}}(X)$ are also identifiable with additional Assumption 5.

2. Given the conditions in Theorem 3(ii), we have

$$\mu_{1,g}(X) = E(Y \mid Z = 1, G = g, X) = E(Y \mid Z = 1, G = g, A, C) = \mu_{1,g}(A, C).$$

Therefore, equation (S16) can be simplified as

$$E(Y \mid Z = 1, S = 1, X) = \eta_{ss}(1, X)\mu_{1,ss}(A, C) + \eta_{s\bar{s}}(1, X)\mu_{1,s\bar{s}}(A, C).$$

Given any $(A, C) = (a, c)$, the function $\{\eta_{ss}(1, a, W, c), \eta_{s\bar{s}}(1, a, W, c)\}^{\mathrm{T}}$ are linearly independent, we then can identify $\mu_{1,ss}(A, C)$ and $\mu_{1,s\bar{s}}(A, C)$. Thus, $\mu_{1,ss}(X)$ and $\mu_{1,s\bar{s}}(X)$ are also identifiable with additional Assumption 6.

3. Given the conditions in Theorem 3(iii), we have

$$\mu_{1,g}(X) = E(Y \mid Z = 1, G = g, X) = E(Y \mid Z = 1, G = g, W, C) = \mu_{1,g}(W, C).$$

Therefore, equation (S16) can be simplified as

$$E(Y \mid Z = 1, S = 1, X) = \eta_{ss}(1, X)\mu_{1,ss}(W, C) + \eta_{s\bar{s}}(1, X)\mu_{1,s\bar{s}}(W, C).$$

Given any $(W, C) = (w, c)$, the function $\{\eta_{s\bar{s}}(1, A, w, c), \eta_{ss}(1, A, w, c)\}^{\mathrm{T}}$ is linearly independent, we then can identify $\mu_{1,ss}(W, C)$ and $\mu_{1,s\bar{s}}(W, C)$. Thus, $\mu_{1,ss}(X)$ and $\mu_{1,s\bar{s}}(X)$ are also identifiable with additional Assumption 7.

4. Given the conditions in Theorem 3(iv), equation (S16) can be simplified as

$$\begin{aligned} E(Y \mid Z = 1, S = 1, X) \\ = \theta_{1,ss,0}\eta_{ss}(1, X) + \theta_{1,s\bar{s},0}\eta_{s\bar{s}}(1, X) + \theta_a A + \theta_c C + \theta_w W, \end{aligned}$$

if the function $\{\eta_{s\bar{s}}(1, X), \eta_{ss}(1, X), X\}^{\mathrm{T}}$ is linearly independent, we then can identify $\theta_{1,s\bar{s},0}$, $\theta_{1,ss,0}$, θ_a, θ_w and θ_c. Thus, $\mu_{1,ss}(X)$ and $\mu_{1,s\bar{s}}(X)$ are also identifiable.

Given the identifiability of $\mu_{1,ss}(X)$ and $\mu_{1,s\bar{s}}(X)$, we can identify $\mu_{1,ss}$ and $\mu_{1,s\bar{s}}$ from (S12). Similarly, we can identify $\mu_{0,\bar{s}\bar{s}}$ and $\mu_{0,s\bar{s}}$.

$\square$

S2 Estimation details

S2.1 Estimation details about Model 1

From (S13), we know that

$$
\begin{aligned}
\mathrm{pr}(S_1 = 1 \mid Z = z', A, C) &= E\{h(1, W, C) \mid Z = z', A, C\} \\
&\geq E\{h(0, W, C) \mid Z = z', A, C\} \\
&= \mathrm{pr}(S_0 = 1 \mid Z = z', A, C),
\end{aligned}
$$

which is compatible with the monotonicity assumption.

S2.2 Estimation details about Proposition 1

If we assume the conditions in Proposition 1 hold, that is, the conditional distribution $f(W \mid Z, A, C)$ in Model 2 is normal distributed and equation (8) holds, we can further estimate the weights $\omega_g(Z, X)$ and $\pi_g(X)$. In order to ensure the linearly independent condition in Proposition 1, we suggest adding higher-order polynomial, square, or interaction terms to the conditional expectation $m(Z, A, C)$ of the conditional distribution $f(W \mid Z, A, C)$, especially in the case of linear regression. After obtaining $\widehat{\omega}_g(Z, A, C)$ as shown in the main text, there are some approaches to estimate the parameters in equation (8). For example, we can use GMM again to solve $\widehat{\psi}$ through the following moment constraints,

$$
\begin{aligned}
\widehat{\omega}_{ss}(Z, A, C) &= E\{\mathrm{pr}\,(S_0 = 1 \mid Z, X; \psi) \mid Z, A, C; \widehat{\gamma}\}, \\
\widehat{\omega}_{s\bar{s}}(Z, A, C) + \widehat{\omega}_{ss}(Z, A, C) &= E\{\mathrm{pr}\,(S_1 = 1 \mid Z, X; \psi) \mid Z, A, C; \widehat{\gamma}\},
\end{aligned}
$$

or we can directly derive the specific form of the right hand of the above estimating equation, and then use Probit regression to solve for $\widehat{\psi}$. Under monotonicity assumption 1, we can estimate $\omega_g(Z, X)$ by plugging $\widehat{\psi}$ into equation (8):

$$
\begin{aligned}
\widehat{\omega}_{ss}(Z, X) &= \mathrm{pr}(S_0 = 1 \mid Z, X; \widehat{\psi}), \quad \widehat{\omega}_{\bar{s}\bar{s}}(Z, X) = 1 - \mathrm{pr}(S_1 = 1 \mid Z, X; \widehat{\psi}), \\
\widehat{\omega}_{s\bar{s}}(Z, X) &= \mathrm{pr}(S_1 = 1 \mid Z, X; \widehat{\psi}) - \mathrm{pr}(S_0 = 1 \mid Z, X; \widehat{\psi}).
\end{aligned}
$$

We then find the estimate of $\pi_g(X)$ as follows:

$$
\widehat{\pi}_g(X) = \frac{\sum_{z=0}^{1} \widehat{\omega}_g(z, X)\mathrm{pr}(z \mid A, C; \widehat{\beta})f(W \mid Z = z, A, C; \widehat{\gamma})}{\sum_{z=0}^{1} \mathrm{pr}(z \mid A, C; \widehat{\beta})f(W \mid Z = z, A, C; \widehat{\gamma})}.
$$

S3 Simulation details

S3.1 Simulation details about bridge function in Section 5

1. We first present the specific form of the conditional density function $f(W \mid U, C)$. Since $W \mid Z, A, C, U \sim N\{q_w(Z, A, C, U), \Sigma_w^2\}$, where $\Sigma_w^2 = \sigma_w^2(1 - \rho_2^2)$ and the specific form of $q_w(Z, A, C, U)$ is

$$
q_w(Z, A, C, U) = \tau_0 + \tau_z Z + \tau_a A + \tau_{c1} C + \tau_{c2} C^2 + \tau_u U,
$$

$$\tau_0 = \frac{\gamma_0 \sigma_u - \iota_0 \sigma_w \rho_2}{\sigma_u}, \quad \tau_z = \frac{\gamma_z \sigma_u - \iota_z \sigma_w \rho_2}{\sigma_u} = 0, \quad \tau_a = \frac{\gamma_a \sigma_u - \iota_a \sigma_w \rho_2}{\sigma_u} = 0,$$

$$\tau_{c1} = \frac{\gamma_{c1} \sigma_u - \iota_{c1} \sigma_w \rho_2}{\sigma_u}, \quad \tau_{c2} = \frac{\gamma_{c2} \sigma_u - \iota_{c2} \sigma_w \rho_2}{\sigma_u}, \quad \tau_u = \frac{\sigma_w \rho_2}{\sigma_u}.$$

2. We next present the specific form of the latent distribution $\mathrm{pr}(S_z = 1 \mid U, C)$ is as follows,

$$
\begin{aligned}
&\mathrm{pr}(S = 1 \mid Z = 1, U, C) \\
&= \mathrm{pr}(S_1 = 1 \mid U, C) \\
&= E\{\mathrm{pr}(S_1 = 1 \mid U, A, W, C) \mid U, C\} \\
&= E\{\mathrm{pr}(S_1 = 1 \mid U, W, C) \mid U, C\} \\
&= E\{\Phi(\zeta_0 + \zeta_w W + \zeta_u U + \zeta_c C) \mid U, C\} \\
&= \Phi\left\{ \frac{\zeta_0 + \zeta_w E(W \mid U, C) + \zeta_u U + \zeta_c C}{\sqrt{1 + \zeta_w^2 \Sigma_w^2}} \right\} \\
&= \Phi\left\{ \frac{\zeta_0 + \zeta_w (\tau_0 + \tau_u U + \tau_{c1} C + \tau_{c2} C^2) + \zeta_u U + \zeta_c C}{\sqrt{1 + \zeta_w^2 \Sigma_w^2}} \right\} \\
&= \Phi\left\{ \frac{\zeta_0 + \zeta_w \tau_0 + \left(\zeta_u + \zeta_w \tau_u\right) U + \left(\zeta_c + \zeta_w \tau_{c1}\right) C + \zeta_w \tau_{c2} C^2}{\sqrt{1 + \zeta_w^2 \Sigma_w^2}} \right\}.
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{pr}(S = 1 \mid Z = 0, U, C) \\
&= \mathrm{pr}(S_0 = 1 \mid U, C) \\
&= E\{\mathrm{pr}(G = \mathrm{AT} \mid U, A, W, C) \mid U, C\} \\
&= E\{\mathrm{pr}(G = \mathrm{AT} \mid U, W, C) \mid U, C\} \\
&= E\{\Phi(\zeta_0 - \exp(\zeta_1) + \zeta_w W + \zeta_u U + \zeta_c C) \mid U, C\} \\
&= \Phi\left\{ \frac{\zeta_0 - \exp(\zeta_1) + \zeta_w E(W \mid U, C) + \zeta_u U + \zeta_c C}{\sqrt{1 + \zeta_w^2 \Sigma_w^2}} \right\} \\
&= \Phi\left\{ \frac{\zeta_0 + \zeta_w \tau_0 - \exp(\zeta_1) + \left(\zeta_u + \zeta_w \tau_u\right) U + \left(\zeta_c + \zeta_w \tau_{c1}\right) C + \zeta_w \tau_{c2} C^2}{\sqrt{1 + \zeta_w^2 \Sigma_w^2}} \right\}.
\end{aligned}
$$

And thus,

$$
\mathrm{pr}(S_t = 1 \mid U, C) = \Phi\left\{ \frac{\zeta_0 + \zeta_w \tau_0 - \exp(\zeta_1) + \exp(\zeta_1) t + \left(\zeta_u + \zeta_w \tau_u\right) U}{\sqrt{1 + \zeta_w^2 \Sigma_w^2}} \right.
$$
$$
\left. + \frac{\left(\zeta_c + \zeta_w \tau_{c1}\right) C + \zeta_w \tau_{c2} C^2}{\sqrt{1 + \zeta_w^2 \Sigma_w^2}} \right\}. \tag{S17}
$$

3. We finally verify the Assumption 4, that is, our data generating mechanism is compatible with the condition $\mathrm{pr}(S = 1 \mid Z = z, C, U) = E\{h(z, W, C) \mid C, U\}$ holds for all z, where

$$
h(t, W, C) = \Phi\left\{ \alpha_0 + \exp(\alpha_1) t + \alpha_w W + \alpha_{c1} C + \alpha_{c2} C^2 \right\}.
$$

We verify as follows:

$$E\{h(t,W,C) \mid U,C\}$$
$$=E\left[\Phi\left\{\alpha_0 + \exp(\alpha_1)t + \alpha_w W + \alpha_{c1}C + \alpha_{c2}C^2\right\} \mid U,C\right]$$
$$=\Phi\left\{\frac{\alpha_0 + \exp(\alpha_1)t + \alpha_w E(W \mid U,C) + \alpha_{c1}C + \alpha_{c2}C^2}{\sqrt{1+\alpha_w^2\Sigma_w^2}}\right\}$$
$$=\Phi\left\{\frac{\alpha_0 + \alpha_w\tau_0 + \exp(\alpha_1)t + \alpha_w\tau_u U + (\alpha_w\tau_{c1} + \alpha_{c1})\,C + (\alpha_w\tau_{c2} + \alpha_{c2})\,C^2}{\sqrt{1+\alpha_w^2\Sigma_w^2}}\right\}.$$

(S18)

Comparing (S17) and (S18), we observe that $\mathrm{pr}(S_t = 1 \mid U,C)$ and $E\{h(t,W,C) \mid U,C\}$ have the same parametric form.

S3.2 Simulation details about other models

1. We now present the specific form of $f(U \mid Z,A,W,C)$. Since $U \mid Z,A,W,C \sim N\{q_u(Z,A,W,C),\Sigma_u^2\}$, where $\Sigma_u^2 = \sigma_u^2(1-\rho_2^2)$ and the specific form of $q_w(Z,A,C,U)$ is

$$q_w(Z,A,W,C) = \nu_0 + \nu_z Z + \nu_a A + \nu_{c1}C + \nu_{c2}C^2 + \nu_w W,$$
$$\nu_0 = \frac{\iota_0\sigma_w - \gamma_0\sigma_u\rho_2}{\sigma_w},\ \nu_z = \frac{\iota_z\sigma_w - \gamma_z\sigma_u\rho_2}{\sigma_w},\ \nu_a = \frac{\iota_a\sigma_w - \gamma_a\sigma_u\rho_2}{\sigma_w},$$
$$\nu_{c1} = \frac{\iota_{c1}\sigma_w - \gamma_{c1}\sigma_u\rho_2}{\sigma_w},\ \nu_{c2} = \frac{\iota_{c2}\sigma_w - \gamma_{c2}\sigma_u\rho_2}{\sigma_w} = 0,\ \nu_u = \frac{\sigma_u\rho_2}{\sigma_w}.$$

2. The specific form of $\mathrm{pr}(S_t \mid Z,A,W,C)$:

$$\mathrm{pr}(S_1 = 1 \mid Z,A,W,C)$$
$$= E\left\{\mathrm{pr}(S_1 = 1 \mid Z,U,A,W,C) \mid Z,A,W,C\right\}$$
$$= E\left\{\mathrm{pr}(S_1 = 1 \mid U,W,C) \mid Z,A,W,C\right\}$$
$$= E\left\{\Phi(\zeta_0 + \zeta_w W + \zeta_u U + \zeta_c C) \mid Z,A,W,C\right\}$$
$$= \Phi\left\{\frac{\zeta_0 + \zeta_w W + \zeta_u E(U \mid Z,W,A,C) + \zeta_c C}{\sqrt{1+\zeta_u^2\Sigma_u^2}}\right\}$$
$$= \Phi\left\{\frac{\zeta_0 + \zeta_w W + \zeta_u(\nu_0 + \nu_z Z + \nu_a A + \nu_w W) + \zeta_c C}{\sqrt{1+\zeta_u^2\Sigma_u^2}}\right\}$$
$$= \Phi\left\{\frac{\zeta_0 + \zeta_u\nu_0 + \zeta_u\nu_z Z + \zeta_u\nu_a A + (\zeta_w + \zeta_u\nu_w)W + (\zeta_u\nu_c + \zeta_c)C}{\sqrt{1+\zeta_u^2\Sigma_u^2}}\right\}.$$

$$\mathrm{pr}(S_0 = 1 \mid Z,A,W,C)$$
$$= E\{\mathrm{pr}(G = \mathrm{AT} \mid U,Z,A,W,C) \mid Z,A,W,C\}$$
$$= E\{\mathrm{pr}(G = \mathrm{AT} \mid U,W,C) \mid Z,A,W,C\}$$

$$= E\{\Phi(\zeta_0 - \exp(\zeta_1) + \zeta_w W + \zeta_u U + \zeta_c C) \mid Z, A, W, C\}$$

$$= \Phi \left\{ \frac{\zeta_0 - \exp(\zeta_1) + \zeta_w W + \zeta_u E(U \mid Z, W, A, C) + \zeta_c C}{\sqrt{1 + \zeta_u^2 \Sigma_u^2}} \right\}$$

$$= \Phi \left\{ \frac{\zeta_0 + \zeta_u \nu_0 - \exp(\zeta_1) + \zeta_u \nu_z Z + \zeta_u \nu_a A + (\zeta_w + \zeta_u \nu_w)W + (\zeta_u \nu_c + \zeta_c)C}{\sqrt{1 + \zeta_u^2 \Sigma_u^2}} \right\}.$$

For simplicity, we let $\mathrm{pr}(S_t = 1 \mid Z, A, W, C) = \Phi\{\psi_0 + \exp(\psi_1)(t-1) + \psi_z Z + \psi_a A + \psi_w W + \psi_c C\}$, where

$$\psi_0 = \frac{\zeta_0 + \zeta_u \nu_0}{\sqrt{1 + \zeta_u^2 \Sigma_u^2}}, \quad \psi_1 = \log \left\{ \frac{\exp(\zeta_1)}{\sqrt{1 + \zeta_u^2 \Sigma_u^2}} \right\}, \quad \psi_z = \frac{\zeta_u \nu_z}{\sqrt{1 + \zeta_u^2 \Sigma_u^2}},$$

$$\psi_a = \frac{\zeta_u \nu_a}{\sqrt{1 + \zeta_u^2 \Sigma_u^2}}, \quad \psi_w = \frac{\zeta_w + \zeta_u \nu_w}{\sqrt{1 + \zeta_u^2 \Sigma_u^2}}, \quad \psi_c = \frac{\zeta_u \nu_c + \zeta_c}{\sqrt{1 + \zeta_u^2 \Sigma_u^2}}.$$

3. The specific form of $\mathrm{pr}(S_t \mid Z, A, C)$,

$$\mathrm{pr}(S_t = 1 \mid Z, A, C)$$
$$= E\{\mathrm{pr}(S_t = 1 \mid Z, A, W, C) \mid Z, A, C\}$$
$$= E[\Phi\{\psi_0 + \exp(\psi_1)(t-1) + \psi_z Z + \psi_a A + \psi_w W + \psi_c C\} \mid Z, A, C]$$
$$= \Phi \left\{ \frac{\psi_0 + \exp(\psi_1)(t-1) + \psi_z Z + \psi_a A + \psi_w E(W \mid Z, A, C) + \psi_c C}{\sqrt{1 + \psi_w^2 \sigma_w^2}} \right\}.$$

S3.3 Additional simulation results

Table 1: Simulation studies with bias ($\times 100$), standard error ($\times 100$) and 95% coverage probability ($\times 100$) for various settings and sample sizes.

n	Case	(θ_a, θ_w)	Δ_{ss}			$\Delta_{s\bar{s}}$			$\Delta_{\bar{s}\bar{s}}$		
			$\zeta_u = 0$								
			Bias	Sd	CP	Bias	Sd	CP	Bias	Sd	CP
1000	(i)	$(0,0)$	−0.9	12.4	94.8	1.5	42.7	96.2	−1.0	35.3	96.0
	(ii)	$(1,0)$	−0.5	13.7	95.8	−3.1	51.4	97.8	2.1	29.3	97.6
	(iii)	$(0,1)$	−7.2	28.8	97.0	18.9	67.4	96.0	−4.1	28.0	94.6
	(iv)	$(1,1)$	−3.4	33.1	95.8	5.9	71.8	97.4	0.3	31.3	95.6
			Bias	Sd	CP	Bias	Sd	CP	Bias	Sd	CP
5000	(i)	$(0,0)$	−0.2	5.5	95.6	−0.2	21.0	96.0	0.6	14.9	95.0
	(ii)	$(1,0)$	−0.1	6.0	95.4	−1.1	24.1	96.0	0.5	11.5	95.4
	(iii)	$(0,1)$	−3.2	13.5	96.6	7.1	31.9	97.2	−0.2	12.3	94.6
	(iv)	$(1,1)$	−2.1	13.9	96.2	4.0	32.2	96.2	0.6	12.6	95.0
			$\zeta_u = 0.1$								
			Bias	Sd	CP	Bias	Sd	CP	Bias	Sd	CP
1000	(i)	$(0,0)$	−0.8	9.7	95.0	2.1	40.9	96.4	−2.0	30.2	95.8
	(ii)	$(1,0)$	−0.6	10.2	95.2	−0.6	47.4	96.8	0.0	24.2	96.8
	(iii)	$(0,1)$	−2.6	17.9	95.4	10.9	54.1	95.2	−3.2	22.9	94.8
	(iv)	$(1,1)$	−1.1	18.7	96.6	3.0	56.5	97.6	−0.4	26.5	96.2
			Bias	Sd	CP	Bias	Sd	CP	Bias	Sd	CP
5000	(i)	$(0,0)$	−0.3	4.2	94.0	0.2	19.4	95.8	0.9	13.7	94.0
	(ii)	$(1,0)$	−0.2	4.7	95.2	−0.6	22.5	95.6	0.6	9.9	96.6
	(iii)	$(0,1)$	−1.7	7.8	96.0	4.8	24.0	95.4	0.1	10.3	95.8
	(iv)	$(1,1)$	−1.2	8.0	95.8	3.0	24.6	95.6	0.6	10.6	96.6

Table 2: Simulation studies with bias ($\times100$), standard error ($\times100$) and 95% coverage probability ($\times100$) for various settings and sample sizes.

n	Case	(θ_a, θ_w)	Δ_{ss}			$\Delta_{s\bar{s}}$			$\Delta_{\bar{s}\bar{s}}$		
			$\zeta_u = 0.3$								
			Bias	Sd	CP	Bias	Sd	CP	Bias	Sd	CP
1000	(i)	$(0,0)$	-0.6	6.5	95.0	3.5	41.4	95.8	-3.8	26.6	93.8
	(ii)	$(1,0)$	-0.5	7.0	96.0	1.7	47.5	96.0	-1.4	20.2	96.0
	(iii)	$(0,1)$	-0.6	9.6	96.2	7.5	45.5	95.8	-2.8	20.9	94.2
	(iv)	$(1,1)$	-0.3	9.6	95.8	1.9	48.8	97.0	-1.0	22.1	96.0
			Bias	Sd	CP	Bias	Sd	CP	Bias	Sd	CP
5000	(i)	$(0,0)$	-0.3	3.2	94.8	1.7	19.3	95.8	0.8	12.4	95.4
	(ii)	$(1,0)$	-0.3	3.5	93.6	1.2	22.2	96.0	0.2	8.8	95.4
	(iii)	$(0,1)$	-0.7	4.3	94.4	4.0	20.1	93.4	0.1	9.1	95.8
	(iv)	$(1,1)$	-0.6	4.3	94.0	3.2	21.3	94.4	0.3	9.5	96.4
			$\zeta_u = 0.4$								
			Bias	Sd	CP	Bias	Sd	CP	Bias	Sd	CP
1000	(i)	$(0,0)$	-0.5	6.0	95.0	4.6	41.4	94.8	-5.2	25.5	94.8
	(ii)	$(1,0)$	-0.4	6.5	95.4	4.2	47.2	95.0	-2.2	19.4	96.0
	(iii)	$(0,1)$	-0.5	8.0	97.0	8.2	44.7	95.0	-3.2	20.2	94.2
	(iv)	$(1,1)$	-0.4	8.0	96.6	2.9	47.9	96.8	-1.4	21.8	96.4
			Bias	Sd	CP	Bias	Sd	CP	Bias	Sd	CP
5000	(i)	$(0,0)$	-0.3	2.8	94.6	3.0	21.0	94.4	0.3	12.3	94.0
	(ii)	$(1,0)$	-0.3	3.1	94.2	2.7	24.2	95.0	0.0	8.7	95.0
	(iii)	$(0,1)$	-0.7	3.6	94.4	5.2	21.2	92.6	-0.1	9.0	95.0
	(iv)	$(1,1)$	-0.6	3.7	94.6	4.0	22.4	93.6	0.3	9.5	95.4