Institute for Economic Studies, Keio University

Keio-IES Discussion Paper Series

CAUSAL INFERENCE WITH AUXILIARY OBSERVATIONS

太田悠太、星野崇宏、大津泰介

2024 年 12 月 2 日
DP2024-022
https://ies.keio.ac.jp/publications/24652/

Institute for Economic Studies, Keio University
2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
ies-office@adst.keio.ac.jp
2 December, 2024

CAUSAL INFERENCE WITH AUXILIARY OBSERVATIONS
太田悠太、星野崇宏、大津泰介
IES Keio DP2024-022
2024 年 12 月 2 日
JEL Classification: C14、C31
キーワード: generalized method of moments, instrumental variables, noncompliance, nonparametric identification, treatment effect

【要旨】
Random assignment of treatment and concurrent data collection on treatment and control groups is often impossible in the evaluation of social programs. A standard method for assessing treatment effects in such infeasible situations is to estimate the local average treatment effect under exclusion restriction and monotonicity assumptions. Recently, several studies have proposed methods to estimate the average treatment effect by additionally assuming treatment effects homogeneity across principal strata or conditional independence of assignment and principal strata. However, these assumptions are often difficult to satisfy. We propose a new strategy for nonparametric identification of causal effects that relaxes these assumptions by using auxiliary observations that are readily available in a wide range of settings. Our strategy identifies the average treatment effect for compliers and average treatment effect on treated under only exclusion restrictions and the assumptions on auxiliary observations. The average treatment effect is then identified under relaxed treatment effects homogeneity. We propose sample analog estimators when the assignment is random and multiply robust estimators when the assignment is non-random. We then present details of the GMM estimation and testing methods which utilize overidentified restrictions. The proposed methods are illustrated by empirical examples which revisit the studies by Thornton (2008), Gerber et al. (2009), and Beam (2016), as well as an experimental data related to marketing in a private sector.

太田悠太
慶應義塾大学経済学部
yuta-ota@keio.jp

星野崇宏
慶應義塾大学経済学部
bayesian@keio.jp

大津泰介
London School of Economics, Department of Economics
T.Otsu@lse.ac.uk

謝辞：This work is supported by JSPS KAKENHI Grant Number JP19KK0322, 23K24809 and 24K22616.

CAUSAL INFERENCE WITH AUXILIARY OBSERVATIONS

YUTA OTA[1], TAKAHIRO HOSHINO[2], AND TAISUKE OTSU[3]

1: Department of Economics, Keio University. Email address: yuta-ota@keio.jp
2: Department of Economics, Keio University and RIKEN AIP. Email address: hoshino@econ.keio.ac.jp
3: Department of Economics, London School of Economics, and Keio Economic Observatory (KEO), Keio University. Email address: t.otsu@lse.ac.uk

Abstract. Random assignment of treatment and concurrent data collection on treatment and control groups is often impossible in the evaluation of social programs. A standard method for assessing treatment effects in such infeasible situations is to estimate the local average treatment effect under exclusion restriction and monotonicity assumptions. Recently, several studies have proposed methods to estimate the average treatment effect by additionally assuming treatment effects homogeneity across principal strata or conditional independence of assignment and principal strata. However, these assumptions are often difficult to satisfy. We propose a new strategy for nonparametric identification of causal effects that relaxes these assumptions by using auxiliary observations that are readily available in a wide range of settings. Our strategy identifies the average treatment effect for compliers and average treatment effect on treated under only exclusion restrictions and the assumptions on auxiliary observations. The average treatment effect is then identified under relaxed treatment effects homogeneity. We propose sample analog estimators when the assignment is random and multiply robust estimators when the assignment is non-random. We then present details of the GMM estimation and testing methods which utilize overidentified restrictions. The proposed methods are illustrated by empirical examples which revisit the studies by Thornton (2008), Gerber et al. (2009), and Beam (2016), as well as an experimental data related to marketing in a private sector.

1. Introduction

Knowledge of causal effects is important for those engaged in policy-making at governmental or non-governmental organization levels, as well as for decision-makers within private sectors (Imbens, 2024). Typically, a causal effect of interest is the average treatment effect (ATE), which represents the average effect over the entire population. One may also be interested in the average treatment effect on treated (ATT), which is the causal effect in th e treated population. Identification and estimation of the treatment effects are typically conducted under the untestable assumption of unconfoundedness (or ignorability), that is, independence between treatment status and potential outcomes of interest (Imbens and Rubin, 2015). The gold standard for achieving unconfoundedness and inferring causal effects is randomized controlled experiments. However, in many cases, such an experiment remains difficult or impossible to implement due to financial, political, or ethical reasons (Athey and Imbens, 2 017). In social program evaluations, it is difficult to archive a perfect randomized controlled experiment because noncompliance with an assigned treatment may occur. When noncompliance occurs, the random assignment of treatment and simultaneous data collection for the treatment and con-trol groups will not be accomplished, and the assumption of unconfoundedness will be violated (Imbens and Angrist, 1994).

In such cases, the local average treatment effect (LATE) can be identified and estimated using the treatment assignment as an instrumental variable under conditions weaker than unconfoundedness (Imbens and Angrist, 1994; Angrist et al., 1996). LATE can consistently estimate causal effects with internal validity in a nonparametric manner, without requiring restrictions on the heterogeneity of causal effects. However, LATE is the average treatment effect only for a subset of the population, the compliers, who react on the assignment as intended by researchers. As compilers constitute a subset of the population, they may not be representative of the overall population. Consequently, it has been argued that LATE may not be a valid parameter for policy policy-making (Robins and Greenland, 1996; Freedman, 2006; Pearl, 2009; Deaton, 2009; Heckman and Urzua, 2010; Aronow and Carnegie, 2013; Swanson and Hernán, 2014; Imbens, 2024). Imbens (2014) argued "If the noncompliance is substantial, we are limited in the questions we can answer credibly and precisely." In addition, identifying LATE requires an exclusion restriction (i.e., assignments do not affect outcomes) and monotonicity (i.e., there is no defiers who oppositely react on the assignment as intended by the researchers) in the population of interest. Of these, with respect to monotonicity, it has been pointed out that this assumption may not hold in many applications, such as when using the assignment of judges with different sentencing rates as an instrumental variable and when conducting randomized controlled trials relying on an encouragement design (Klein, 2010; De Chaisemartin, 2017; Small et al., 2017; Dahl et al., 2023). If there are defiers, the LATE estimator converges to a weighted difference between the effect of the treatment among compliers and defiers (Angrist et al., 1996; De Chaisemartin, 2017).

This paper presents a strategy to identify ATE and ATT by using auxiliary observations when compliance with the assigned treatment is not perfect. Our identification strategies cover the cases with and without monotonicity assumption. (1) In the setup with monotonicity, a variable for pre-assignment outcome of interest is used as an auxiliary observation, in addition to the basic observations including the variable of assignment, treatment, and outcome after assignment. Our identification strategy requires some new assumptions: a parallel trend assumption between always-taker and never-taker in a controlled group (to identify ATT), and additionally homogeneity assumption of causal effects between always-taker and never-taker (to identify ATE). (2) In the setup without monotonicity, a variable for pre-assignment treatment status is used as an additional auxiliary observation. To identify ATE and ATT, we introduce an additional assumption that the compliance status in the controlled group is unchanged between before and after assignment. However, we show that this assumption can be dropped by employing an alternative identification strategy. The assumptions listed in (1) and (2), except for the homogeneity assumption, are assumptions about the relationship between basic and auxiliary observations, and we do not impose restrictions on causal effects of interest. Also since our identification strategy does not require covariate functions for the variables of treatment, outcome of interest, and auxiliary observation, there is no need to bother about misspecification of the functional forms. Furthermore, some assumptions are testable.

There are many examples where such auxiliary observations can be obtained. In the setup with monotonicity, the only auxiliary observation required to apply our method is the variable

of pre-assignment outcome of interest. This is often observed in the baseline survey of a randomized experiment. For example, researchers want to know the effect of job training on wages observe the wages at the baseline survey. In the setup without monotonicity, two auxiliary variables are needed: pre-assignment outcome of interest and pre-assignment treatment status. It may be possible in some applications to obtain them in the baseline survey, and this paper presents several experiment designs to obtain such observations. In addition, if an automatic data collection system (e.g., point-of-sale, marketing platform) is available, the auxiliary variables are easier to obtain. For example, suppose that a manufacturer's decision-makers want to know whether the sale of a new product contributes to the sales of the company's entire product line. One of the fears of them is that even if new product sales are sufficient, cannibalization does not boost sales of their entire lineup. Since it is difficult to force consumers to purchase a new product, suppose that an experiment is conducted in which coupons are assigned to promote the purchase of the new product. In this case, the coupon is the assignment, the purchase of the new product is the treatment, and the sales of the entire product line is the outcome of interest. Each subject's purchase history of new products before the experiment (the period when there are no coupons) is available as the pre-assignment treatment status. And each subject's purchases amount of lineup before the new product is launched can be used as the pre-assignment outcome of interest. These auxiliary variables are usually stored in a database in the system.

Studies on how to identify and estimate ATE when noncompliance occurs are limited. Several studies have proposed methods to estimate ATE under standard LATE assumptions, assuming that ATE is the same as LATE under covariate conditions (Angrist and Fernandez-Val, 2010; Aronow and Carnegie, 2013; Fricke et al., 2020). Wang and Tchetgen Tchetgen (2018) showed that, in addition to the standard LATE assumption excluding monotonicity, ATE and LATE are equal conditional on covariates if either ATE or the difference in treatment proportions between the treatment and control groups is conditionally mean independence from an unobserved factor. Each of these methods essentially requires an assumption of homogeneity whereby the compliers and all other noncompliers (i.e., always-takers, never-takers, and defiers in the case without monotonicity) are equal in their dependence on observational covariates. In other approaches, Heckman and Vytlacil (1999), Heckman and Vytlacil (2005), and Heckman and Vytlacil (2007) found that marginal treatment effects (MTE) can be used to identify ATE under standard LATE assumptions, including monotonicity. However, this method requires an instrumental variable that continuously supports the treatment probability for each value of the covariate, making full nonparametric identification difficult (Brinch et al., 2017). Brinch et al. (2017) showed how to estimate MTE using discrete instrumental variables, but this is only a linear approximation. These approaches can be problematic because they a priori impose restrictions on the effects of interest. Several other approaches based on partial identification have been proposed (Balke and Pearl, 1997; Kennedy et al., 2020), but we omit their details because their focus is different from ours. With the help of auxiliary observations, we provide point identification of ATE under relaxed, only between always-takers and never-takers homogeneity assumptions. Therefore, our method allows the average treatment effect for compliers (LATE) to differ from that of

noncompliers. That relaxed homogeneity may be relatively convincing because always-takers and never-takers have in common that they do not react to assignments.

To the best of our knowledge, the only attempt to identify ATE without monotonicity is Wang and Tchetgen Tchetgen (2018) mentioned above, but several attempts exist in identifying LATE (Small et al., 2017; De Chaisemartin, 2017; van't Hoff et al., 2023; Dahl et al., 2023). These papers provide conditions for identifying LATE in local compliers and conditions for identifying LATE under relaxed, non-global monotonicity. For example, De Chaisemartin (2017) showed that it is possible to identify LATE of a subpopulation of compliers without monotonicity under the assumption that a certain fraction of the total compliers have the same average treatment effect and population as defiers. Unlike these methods, we do not introduce new assumptions related to monotonicity, but allow for global non-monotonicity by introducing assumptions about auxiliary observations.

Several methods for identification of causal effects by using auxiliary observations and extended experiments have been proposed in setups where researchers are interested in the causal effects of assignments when intermediate variables occur between assignments and outcomes (Mealli and Pacini, 2013; Yang and Small, 2016; Jiang et al., 2016; Gabriel and Follmann, 2016; Jiang and Ding, 2021). These causal effects are called principal causal effects (PCEs). The problem setting of noncompliance can also be viewed in this setup (Frangakis and Rubin, 2002). Among them, Jiang and Ding (2021) identified conditions for auxiliary observations that would point identify PCEs when noncompliance occurs and showed that partial identification is possible even in the absence of monotonicity. Mealli and Pacini (2013) also proposed a method for partial identification of PCEs when exclusion restrictions are violated under monotonicity by using secondary outcomes (such as side-effects) when noncompliance occurs. Furthermore, an extended experimentation method using an encouragement design to identify direct and indirect effects in causal mediation analysis has been proposed (Mattei and Mealli, 2011; Imai et al., 2013). The reason why there are several methods to achieve identification and estimation empowered by auxiliary observations or extended experiments in causal inference when intermediate variables occur is probably due to the need to deal with two types of potential variables: intermediate variables and outcome of interest. Nevertheless, there have been no studies attempting to identify ATE under noncompliance with auxiliary observations. Identification of ATE is complicated by the need to address the outcomes of always-takers and never-takers who are nonreactive to assignments, as well as intermediate and outcome variables. In this paper, two types of auxiliary observations, pre-assignment outcome of interest and pre-assignment treatment status, are used to identify ATE.

This paper proceeds as follows. In Section 2, we present two benchmark results for the cases where the researcher can observe an outcome or treatment variable before the assignment. We also consider the case where the conditional ignorability is satisfied and present a multiply robust estimator. Section 3 considers setups without monotonicity setups and presents three experimental designs for auxiliary observations. Section 4 presents some empirical illustrations of the proposed methods.

2. Benchmark results

In this section, we present our benchmark results. Section 2.1 considers the case where the researcher can observe an outcome Y^{pre} before the assignment and there is no defier in the population. Section 2.2 studies the case where the researcher can observe a treatment D^{pre} before the assignment and defiers may exist in the population.

This section employs the following notation. Let $Z \in \{0,1\}$ be an assignment indicator, $D \in \{0,1\}$ be a treatment status indicator, and $Y \in \mathcal{Y} \subset \mathbb{R}$ be an outcome of interest. Then let $D_z \in \{0,1\}$ be the potential treatment variable realized only when $Z = z$, and $Y_{zd} \in \mathcal{Y}$ be the potential outcome realized only when $Z = z$ and $D = d$.

2.1. **Observable outcome before assignment.**

To begin with, we consider the case where the researcher can observe an outcome variable Y^{pre} before the assignment. We impose the following basic assumptions.

Assumption Y.

(i): *It holds* $Y_d = Y_{zd}$ *for each* $z \in \{0,1\}$ *and* $d \in \{0,1\}$*, and*

$$\begin{aligned} D &= ZD_1 + (1-Z)D_0, \\ Y &= ZDY_{11} + Z(1-D)Y_{10} + (1-Z)DY_{01} + (1-Z)(1-D)Y_{00}. \end{aligned}$$

(ii): D_z *is weakly monotone in* z*, i.e.,* $\mathbb{P}(D_1 \geq D_0) = 1$.

Assumption Y (i) is standard in the literature of causal inference using randomized experiments with non-compliance (e.g., Angrist et al., 1996). Note that the assumption $Y_d = Y_{zd}$ rules out direct effects of Z on the potential outcomes. To understand Assumption Y (ii), we introduce a principal strata variable:

$$U = \begin{cases} a & \text{if } D_1 = 1, D_0 = 1, \\ c & \text{if } D_1 = 1, D_0 = 0, \\ d & \text{if } D_1 = 0, D_0 = 1, \\ n & \text{if } D_1 = 0, D_0 = 0. \end{cases} \tag{1}$$

The compliers ($U = c$) react on the assignment as intended by the researcher, and other three strata do not. The always-takers ($U = a$) are always treated, the never-takers ($U = n$) are never treated, and the defiers ($U = d$) react conversely to the assignment. Then Assumption Y (ii) says that there is no defier in the population, i.e., $\mathbb{P}(U = d) = 0$. The following sections present identification results without this monotonicity assumption.

Our causal effects of interest are the average treatment effect ($\text{ATE} = \mathbb{E}[Y_1 - Y_0]$), average effect of treatment on the treated ($\text{ATT} = \mathbb{E}[Y_1 - Y_0 | D_1 = 1]$), and compliers' average treatment effect or local average treatment effect ($\text{ATE}(c) = \mathbb{E}[Y_1 - Y_0 | U = c]$). To describe our identification strategy, it is insightful to express these estimands by using the notation $\mu_d^u = \mathbb{E}[Y_d | U = u]$

U	Y	D	Z	Y^{pre}
c or a	Y_1	1	1	Y^{pre}
n	Y_0	0	1	Y^{pre}
a	Y_1	1	0	Y^{pre}
c or n	Y_0	0	0	Y^{pre}

TABLE 1. Benchmark case: Observable Y^{pre}

and $p^u = \mathbb{P}(U = u)$ for $u \in \{a, c, n\}$ as follows

$$\begin{aligned} \text{ATE}(c) &= \mu_1^c - \mu_0^c, \\ \text{ATT} &= \frac{p^c(\mu_1^c - \mu_0^c) + p^a(\mu_1^a - \mu_0^a)}{p^c + p^a}, \\ \text{ATE} &= p^c(\mu_1^c - \mu_0^c) + p^a(\mu_1^a - \mu_0^a) + p^n(\mu_1^n - \mu_0^n). \end{aligned} \tag{2}$$

It is known that ATE(c) is identified under mild conditions (Angrist et al., 1996). However, ATT and ATE cannot be generally identified in the present setup. This paper provides several strategies to identity μ_d^u's and p^u's when the researcher can access to auxiliary observations in addition to the main observable (Y, D, Z).

This subsection considers the following situation.

Assumption Y.

(iii): *[Observable pre-treatment outcome] An outcome variable $Y^{\text{pre}} \in \mathcal{Y}$ is observable at a time before the treatment D is realized.*

(iv): *[Random assignment] Z is independent from $(Y^{\text{pre}}, D_1, D_0, Y_{11}, Y_{10}, Y_{01}, Y_{00})$.*

This setup should be considered as a benchmark and the following sections present other identification strategies without the monotonicity assumption (Assumption Y (ii)). One of the fundamental challenges of causal inference using randomized experiments with non-compliance is that U is never observed for all subjects because only D can be observed. The observations from randomized experiments can be divided into the four rows in Table 1 according to the values of D and Z. Some rows are mixtures of two principal strata. Another fundamental challenge, especially for identifying ATT and ATE, is that the outcomes for the cases where the always-takers receive no treatment and the never-takers receive treatment are never observed. To address this problem, our key assumption (Assumption Y (iii)) introduces the auxiliary outcome Y^{pre}, which is observed before realization of the treatment variable D.

First of all, Table 1 suggests that the following objects are identified without using Y^{pre}:

$$\begin{aligned} \mu_1^a &= \mathbb{E}[Y|Z = 0, D = 1], \qquad p^a = \mathbb{P}(D = 1|Z = 0), \\ \mu_1^n &= \mathbb{E}[Y|Z = 1, D = 0], \qquad p^n = \mathbb{P}(D = 0|Z = 1), \\ p^c &= \mathbb{P}(D = 1|Z = 1) - p^a, \\ \mu_1^c &= \frac{(p^c + p^a)\mathbb{E}[Y|Z = 1, D = 1] - p^a\mu_1^a}{p^c}, \\ \mu_0^c &= \frac{(p^c + p^n)\mathbb{E}[Y|Z = 0, D = 0] - p^n\mu_0^n}{p^c}. \end{aligned} \tag{3}$$

Therefore, under Assumptions Y (i)-(iv), we can identify ATE for compliers as $\text{ATE}(c) = \mu_1^c - \mu_0^c$ as far as $p^c > 0$.

In order to identify ATT and ATE, it remains to identify μ_0^a and μ_1^n by using the additional data Y^{pre}. To this end, we add the following assumptions.

Assumption Y.

(v): *[Parallel trend of nonreactive strata]* $\mathbb{E}[Y_0 - Y^{\text{pre}}|U = a] = \mathbb{E}[Y_0 - Y^{\text{pre}}|U = n]$.

(vi): *[Homogeneity of nonreactive strata]* $\mathbb{E}[Y_1 - Y_0|U = a] = \mathbb{E}[Y_1 - Y_0|U = n]$.

Assumption Y (v) is an analog of the parallel trend assumption on the types a and n whose participation decisions are not affected by Z. Assumption Y (vi) requires homogeneous treatment effects on the types a and n. Note that in the conventional identification analysis for ATE, we typically impose homogeneity over all types. On the other hand, we only require homogeneity over the types a and n. To see how Assumption Y (v) is utilized to identify μ_0^a, observe that

$$\begin{aligned} \mu_1^a - \mu_0^a &= \mathbb{E}[Y_1 - Y^{\text{pre}}|U = a] - \mathbb{E}[Y_0 - Y^{\text{pre}}|U = a] \\ &= \mathbb{E}[Y_1 - Y^{\text{pre}}|U = a] - \mathbb{E}[Y_0 - Y^{\text{pre}}|U = n] \\ &= \mu_1^a - \mu_{\text{pre}}^a - \mu_0^n + \mu_{\text{pre}}^n, \end{aligned} \tag{4}$$

where $\mu_{\text{pre}}^a = \mathbb{E}[Y^{\text{pre}}|U = a]$, $\mu_{\text{pre}}^n = \mathbb{E}[Y^{\text{pre}}|U = n]$, and the second equality uses Assumption Y (v). Since μ_{pre}^a and μ_{pre}^n are identified by

$$\begin{aligned} \mu_{\text{pre}}^a &= \mathbb{E}[Y^{\text{pre}}|Z = 1, D = 1], \\ \mu_{\text{pre}}^n &= \mathbb{E}[Y^{\text{pre}}|Z = 1, D = 0], \end{aligned} \tag{5}$$

we can identify μ_0^a by (4) and thus ATT is also identified by (2). Finally, Assumption Y (vi) guarantees identification of μ_1^n as $\mu_1^n = \mu_0^n + \mu_1^a - \mu_0^a$ so that ATE is identified by the expression in (2). This assumption is considered natural in this setup because both always-takers and never-takers are units who determine their treatment status D without being influenced by the value of the assignment indicator Z. In other words, unlike compliers and defiers, they are units who are not influenced by the provision of information, incentives, or resistance to coercion due to receiving an assignment, or units who are considered to be influenced to a small extent by these factors. Furthermore, it is thought that units for whom the hidden cost of receiving treatment is smaller than the size of the treatment effect will become always-takers, and units for whom the hidden cost is larger will become never-takers.

Combining these results, identification of the causal objects in (2) is established as follows.

Theorem 1. *Consider the setup of this subsection.*

(i): *Under Assumption Y (i)-(iv),* $\text{ATE}(c)$ *is identified.*

(ii): *Under Assumption Y (i)-(v),* ATT *is identified.*

(iii): *Under Assumption Y (i)-(vi),* ATE *is identified.*

Based on this theorem, we can estimate these causal objects by taking sample counterparts, and conduct inference based on standard methods, such as the delta method and bootstrap.

Remark 1. [Overidentification] The above argument for establishing Theorem 1 (iii) is based on showing just identification of the 11 parameters, (μ_1^u, μ_0^u, p^u) for $u \in \{c, a, n\}$ and μ_{pre}^u for $u \in \{a, n\}$. Indeed by introducing the parameter μ_{pre}^c, we have three additional restrictions:

$$\begin{aligned} p^c &= \mathbb{P}(D=0|Z=0) - p^n, \\ \mu_{\text{pre}}^c &= \frac{(p^c + p^a)\mathbb{E}[Y^{\text{pre}}|Z=1, D=1] - p^a \mu_{\text{pre}}^a}{p^a}, \\ \mu_{\text{pre}}^n &= \frac{(p^c + p^n)\mathbb{E}[Y^{\text{pre}}|Z=0, D=0] - p^c \mu_{\text{pre}}^c}{p^n}. \end{aligned} \tag{6}$$

These additional moment conditions can be incorporated by using the generalized method of moments.

2.2. **Observable treatment before assignment.** This subsection considers the case where we can observe a treatment D^{pre} before the assignment under the following basic assumption.

Assumption D. ***(i)*** *It holds $Y_d = Y_{zd}$ for each $z \in \{0,1\}$ and $d \in \{0,1\}$, and*

$$\begin{aligned} D &= ZD_1 + (1-Z)D_0, \\ Y &= ZDY_{11} + Z(1-D)Y_{10} + (1-Z)DY_{01} + (1-Z)(1-D)Y_{00}. \end{aligned}$$

We employ the principal strata variable U defined in (1). This assumption is identical to Assumption Y (i), but we do not impose the monotonicity assumption on D_z, so the defier can present in the population, i.e., $\mathbb{P}(U=d) > 0$. In this case, the causal effects of interest can be expressed as

$$\begin{aligned} \text{ATE}(c) &= \mu_1^c - \mu_0^c, \\ \text{ATT} &= \frac{p^c(\mu_1^c - \mu_0^c) + p^a(\mu_1^a - \mu_0^a)}{p^c + p^a}, \\ \text{ATE} &= p^c(\mu_1^c - \mu_0^c) + p^a(\mu_1^a - \mu_0^a) + p^n(\mu_1^n - \mu_0^n) + p^d(\mu_1^d - \mu_0^d), \end{aligned} \tag{7}$$

where $\mu_d^d = \mathbb{E}[Y_d|U=d]$ and $p^d = \mathbb{P}(U=d)$. Clearly these objects cannot be identified under the present setup. In this subsection, we assume that in addition to the main observable (Z, D, Y), the researcher observes

$$D^{\text{pre}} \in \{0,1\} : \text{treatment indicator to be observed at the time before the assignment,} \tag{8}$$

Intuitively D^{pre} is used for untangling the mixtures of principal strata in observations. One of the fundamental challenges of causal inference using randomized experiments with non-compliance is that U is never observed for all subjects because only D or D^{pre} can be observed. The observations from randomized experiments can be divided into the four rows in Table 2 (left) according to D and Z values. All rows are mixtures of two principal strata. If we assume absence of defiers, the parameters $\mu_1^c, \mu_0^c, \mu_1^a, \mu_0^n, p^c, p^a, p^n$ and $\text{ATE}(c)$ are identified by using the single principal stratum moments from the second and third rows. But if not, we have to deal with the mixtures in another way, and we employ D^{pre} to overcome this.

To conduct identification analysis for the objects of interest in (7), we impose the following assumptions.

U	Y	D	Z
c or a	Y_1	1	1
n or d	Y_1	0	1
a or d	Y_0	1	0
c or n	Y_0	0	0

U	Y	D	Z	D^{pre}
c	Y_1	1	1	0
a	Y_1	1	1	1
n	Y_1	0	1	0
d	Y_1	0	1	1
a or d	Y_0	1	0	1
c or n	Y_0	0	0	0

TABLE 2. Benchmark case with (right) and without (left) D^{pre}

Assumption D.

(ii): *[Random assignment] Z is independent from $(D^{\text{pre}}, D_1, D_0, Y_{11}, Y_{10}, Y_{01}, Y_{00})$.*

(iii): *[Pre-assignment treatment status as control] $D_0 = D^{\text{pre}}$.*

Assumption D (ii) is a standard assumption for random assignment of Z. Assumption D (iii) says that the auxiliary observation D^{pre} plays the role of D_z with $Z = 0$. Since the treatment D^{pre} occurs before the assignment of Z, this assumption is reasonable. The relationships of the observables and principal strata variable can be summarized as in Table 2. Due to Assumptions D, we do not have rows for the cases of $D \neq D^{\text{pre}}$ with $Z = 0$.

Indeed the first four rows of this table (right panel) suggest that the following objects are identified:

$$\begin{aligned}
\mu_1^c &= \mathbb{E}[Y|Z=1, D=1, D^{\text{pre}}=0], \qquad p^c = \mathbb{P}(D=1, D^{\text{pre}}=0|Z=1),\\
\mu_1^a &= \mathbb{E}[Y|Z=1, D=1, D^{\text{pre}}=1], \qquad p^a = \mathbb{P}(D=1, D^{\text{pre}}=1|Z=1),\\
\mu_0^n &= \mathbb{E}[Y|Z=1, D=0, D^{\text{pre}}=0], \qquad p^n = \mathbb{P}(D=0, D^{\text{pre}}=0|Z=1),\\
\mu_0^d &= \mathbb{E}[Y|Z=1, D=0, D^{\text{pre}}=1], \qquad p^d = \mathbb{P}(D=0, D^{\text{pre}}=1|Z=1). \qquad (9)
\end{aligned}$$

Furthermore, the last two rows of this table (right panel) can be utilized to identify

$$\begin{aligned}
\mu_1^d &= \frac{(p^a+p^d)\mathbb{E}[Y|Z=0, D=1, D^{\text{pre}}=1] - p^a\mu_1^a}{p^d},\\
\mu_0^c &= \frac{(p^c+p^n)\mathbb{E}[Y|Z=0, D=0, D^{\text{pre}}=0] - p^n\mu_0^n}{p^c}. \qquad (10)
\end{aligned}$$

Therefore, under Assumption D, we can identify ATE for compliers and defiers as $\text{ATE}(c) = \mu_1^c - \mu_0^c$ and $\text{ATE}(d) = \mu_1^d - \mu_0^d$, respectively.

Combining these results, our identification results for this case are presented as follows.

Theorem 2. *Consider the setup of this subsection. Under Assumption D,* $\text{ATE}(c)$, $\text{ATE}(d)$, *and p^u for all $u \in \{a, c, d, n\}$ are identified.*

Based on this theorem, we can estimate these causal objects by taking sample counterparts, and conduct inference based on standard methods, such as the delta method and bootstrap. There are several ways to utilize this theorem for empirical analyses. First, we can estimate the probability of compliers p^d as a diagnostics for the monotonicity (or no defier) assumption. Second, we can formally test the validity of the local average treatment analysis by testing the null of $\text{ATE}(c) = \text{ATE}(d)$. Our proof shows that this null is equivalent that the identification

formulae for ATE(c) are same for the cases with or without D^{pre}. Finally, although we cannot identify ATE or ATT, this theorem can be utilized to obtain tighter bounds for these objects compared to the conventional ones.

2.3. **Identification under ignorability condition.** In observational studies, it is often the case that the random assignment of Z (Assumption Y (iv)) is violated. In this subsection we show that our identification argument can be extended to the case where the following ignorability condition is satisfied. Let $X \in \mathcal{X} \subset \mathbb{R}^q$ be a vector of q-dimensional covariates.

Assumption Y. ***(iv)'*** *[Ignorability] Conditionally on X, Z is independent from $(D, Y^{\text{pre}}, D_1, D_0, Y_{11}, Y_{10}, Y_{01}, Y_{00})$.*

This is a standard ignorability or unconfoundedness condition commonly imposed in the literature of causal inference with observational studies. Based on the discussion of the previous subsection, it is sufficient for identification of the causal estimands in (2) to identify

$$\begin{aligned} \delta_{(z,d)} &= \mathbb{E}[Y_d | D_z = d], \qquad \delta^{\text{pre}}_{(z,d)} = \mathbb{E}[Y^{\text{pre}}_d | D_z = d], \\ \pi_{(z,d)} &= \mathbb{P}(D_z = d), \end{aligned} \tag{11}$$

for each $z \in \{0,1\}$ and $d \in \{0,1\}$. To establish multiply robust representations of $\delta_{(z,d)}$ and $\pi_{(z,d)}$ under Assumption Y (iv)', we introduce parametric models

$$\begin{aligned} e_z(X;\alpha) \quad &\text{for} \quad \mathbb{P}(Z = z|X), \\ p_{(z,d)}(X;\beta) \quad &\text{for} \quad \mathbb{P}(D = d|Z = z, X), \\ m(X;\gamma) \quad &\text{for} \quad \mathbb{E}[Y|X], \\ m^{\text{pre}}(X;\gamma^{\text{pre}}) \quad &\text{for} \quad \mathbb{E}[Y^{\text{pre}}|X], \\ p_d(X;\eta) \quad &\text{for} \quad \mathbb{P}(D = d|X), \end{aligned}$$

where α, β, γ, and η are finite dimensional parameters. By using these parametric models, multiply robust representations of the population objects $\delta_{(z,d)}$ and $\pi_{(z,d)}$ are obtained as follows.

Theorem 3. *Consider the setup of this subsection. Under Assumption Y (i)-(iii), (iv)', (v)-(vi), it holds*

$$\begin{aligned} \delta_{(z,d)} &= \mathbb{E}\left[\frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)}\frac{\mathbb{I}\{D=d\}}{p_{(z,d)}(X;\beta)}Y\right] \\ &\quad -\mathbb{E}\left[\frac{\mathbb{I}\{Z=z\}\mathbb{I}\{D=d\} - e_z(X;\alpha)p_{(z,d)}(X;\beta)}{e_z(X;\alpha)p_{(z,d)}(X;\beta)}m(X;\gamma)\right], \end{aligned}$$

$$\begin{aligned} \delta^{\text{pre}}_{(z,d)} &= \mathbb{E}\left[\frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)}\frac{\mathbb{I}\{D=d\}}{p_{(z,d)}(X;\beta)}Y^{\text{pre}}\right] \\ &\quad -\mathbb{E}\left[\frac{\mathbb{I}\{Z=z\}\mathbb{I}\{D=d\} - e_z(X;\alpha)p_{(z,d)}(X;\beta)}{e_z(X;\alpha)p_{(z,d)}(X;\beta)}m^{\text{pre}}(X;\gamma^{\text{pre}})\right], \end{aligned}$$

$$\pi_{(z,d)} = \mathbb{E}\left[\frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)}\mathbb{I}\{D=d\}\right] - \mathbb{E}\left[\frac{\mathbb{I}\{Z=z\} - e_z(X;\alpha)}{e_z(X;\alpha)}p_d(X;\eta)\right].$$

By taking the sample counterparts of these representations, we can construct multiply robust estimators for $\delta_{(z,d)}$, $\delta^{\text{pre}}_{(z,d)}$, and $\pi_{(z,d)}$. Then the 11 parameters μ^u_1, μ^u_0, p^u for $u \in \{c, a, n\}$ and μ^u_{pre} for $u \in \{a, n\}$ are over-identified by the moment restrictions:

$$\begin{aligned}
\mu^a_1 &= \delta_{(0,1)}, \qquad \mu^n_1 = \delta_{(1,0)}, \qquad p^a = \pi_{(0,1)}, \qquad p^n = \pi_{(1,0)}, \qquad p^c = \pi_{(1,1)} - p^a, \\
\mu^c_1 &= \frac{(p^c + p^a)\delta_{(1,1)} - p^a \mu^a_1}{p^c}, \qquad \mu^c_0 = \frac{(p^c + p^n)\delta_{(0,0)} - p^n \mu^n_0}{p^c}, \\
\mu^a_0 &= = \mu^a_{\text{pre}} + \mu^n_0 - \mu^n_{\text{pre}}, \qquad \mu^n_1 = \mu^n_0 + \mu^a_1 - \mu^a_0, \qquad \mu^a_{\text{pre}} = \delta^{\text{pre}}_{(0,1)}, \qquad , \mu^n_{\text{pre}} = \delta^{\text{pre}}_{(1,0)}, \\
p^c &= \pi_{(0,0)} - p^n, \qquad \mu^c_{\text{pre}} = \frac{(p^c + p^a)\delta_{(1,1)} - p^a \mu^a_{\text{pre}}}{p^c}, \qquad \mu^n_{\text{pre}} = \frac{(p^c + p^n)\delta^{\text{pre}}_{(0,0)} - p^c \mu^c_{\text{pre}}}{p^n}, \qquad (12)
\end{aligned}$$

where the first two lines are obtained from the rows in Table 1, the third line is obtained from Assumption Y (v)-(vi) and the second and third rows in Table 1, and the last line follows from the first and last rows in Table 1. Just identification of μ^u_1, μ^u_0, p^u for $u \in \{c, a, n\}$ and μ^u_{pre} for $u \in \{a, n\}$ is guaranteed by the first three lines, and the last line provides overidentifying restrictions.

We close this subsection by summarizing the doubly robust properties of the estimators based on Theorem 3 and (12).

Proposition 1. *Consider the setup of this subsection. Suppose Assumption Y (i)-(iii), (iv)', (v)-(vi) hold true. Then*

- **(i):** $\delta_{(z,d)}$ *can be consistently estimated if either* $\{e_z(X;\alpha), p_{(z,d)}(X;\beta)\}$ *or* $m(X;\gamma)$ *is correctly specified, and* $\delta^{\text{pre}}_{(z,d)}$ *can be consistently estimated if either* $\{e_z(X;\alpha), p_{(z,d)}(X;\beta)\}$ *or* $m^{\text{pre}}(X;\gamma^{\text{pre}})$ *is correctly specified,*
- **(ii):** $\pi_{(z,d)}$ *can be consistently estimated if either* $e_z(X;\alpha)$ *or* $p_d(X;\eta)$ *is correctly specified,*
- **(iii):** ATE(c) *can be consistently estimated if either* $\{e_z(X;\alpha), p_{(z,d)}(X;\beta)\}$, $\{m(X;\gamma), p_d(X;\eta)\}$, *or* $\{e_z(X;\alpha), m(X;\gamma)\}$ *is correctly specified,*
- **(iv):** ATT *and* ATE *can be consistently estimated if either* $\{e_z(X;\alpha), p_{(z,d)}(X;\beta)\}$, $\{m(X;\gamma), m^{\text{pre}}(X;\gamma^{\text{pre}}), p_d(X;\eta)\}$, *or* $\{e_z(X;\alpha), m(X;\gamma), m^{\text{pre}}(X;\gamma^{\text{pre}})\}$ *is correctly specified.*

Furthermore, the multiply robust estimator for ATE(c) is asymptotically locally efficient if $\{e_z(X;\alpha), p_{(z,d)}(X;\beta), m(X;\gamma), p_d(X;\eta)\}$ are correctly specified, and also the multiply robust estimators for ATT and ATE are asymptotically locally efficient if $\{e_z(X;\alpha), p_{(z,d)}(X;\beta), m(X;\gamma), m^{\text{pre}}(X;\gamma^{\text{pre}})\}$ are correctly specified.

2.4. **Estimation and testing.** In this subsection, we briefly discuss estimation and testing methods for ATE identified by Theorems 1 and 3 above. The methods for ATE(c) and ATT can be obtained in the same manner.

First, we consider estimation of ATE based on Theorem 1 (iii). Let $\hat{\delta}_{(z,d)}$, $\hat{\delta}^{\text{pre}}_{(z,d)}$, and $\hat{\pi}_{(z,d)}$ be the empirical (conditional) moments of $\delta_{(z,d)} = \mathbb{E}[Y_d | D_z = d]$, $\delta^{\text{pre}}_{(z,d)} = \mathbb{E}[Y^{\text{pre}} | D_z = d]$, and $\pi_{(z,d)} = \mathbb{P}(D_z = d)$, respectively, and $\hat{\zeta}$ and ζ be their vectorizations. Also let θ be a 11-dimensional vector given by (μ^u_1, μ^u_0, p^u) for $u \in \{c, a, n\}$ and μ^u_{pre} for $u \in \{a, n\}$, which provides

a formula for ATE as

$$\mathrm{ATE}(\theta) = p^c(\mu_1^c - \mu_0^c) + p^a(\mu_1^a - \mu_0^a) + p^n(\mu_1^n - \mu_0^n).$$

Then the minimum distance estimator for ATE is obtained as $\hat{\omega}$ for

$$(\hat{\theta}, \hat{\omega}) = \arg\min_{\theta,\omega} g(\hat{\zeta}, \theta, \omega)' \Psi g(\hat{\zeta}, \theta, \omega), \tag{13}$$

where the vector of moment conditions $g(\zeta, \theta, \omega) = 0$ is obtained by stacking the equations (3)-(5) and $\omega = \mathrm{ATE}(\theta)$ (and also (6)). The weight matrix Ψ may be chosen to achieve the asymptotic efficiency. Statistical inference on ω can be conducted by the Wald statistic, likelihood ratio-type statistic, or bootstrap method.

Next, if the parameters ζ are identified by the ignorability condition as in Theorem 3, their estimating equations are given by

$$\begin{aligned} & g_1(W, \zeta, \alpha, \beta, \gamma, \gamma^{\mathrm{pre}}) \\ = & \begin{bmatrix} \left\{ \delta_{(z,d)} - \frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)} \frac{\mathbb{I}\{D=d\}}{p_{(z,d)}(X;\beta)} Y + \frac{\mathbb{I}\{Z=z\}\mathbb{I}\{D=d\} - e_z(X;\alpha)p_{(z,d)}(X;\beta)}{e_z(X;\alpha)p_{(z,d)}(X;\beta)} m(X;\gamma) \right\}_{(z,d)} \\ \left\{ \delta^{\mathrm{pre}}_{(z,d)} - \frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)} \frac{\mathbb{I}\{D=d\}}{p_{(z,d)}(X;\beta)} Y^{\mathrm{pre}} + \frac{\mathbb{I}\{Z=z\}\mathbb{I}\{D=d\} - e_z(X;\alpha)p_{(z,d)}(X;\beta)}{e_z(X;\alpha)p_{(z,d)}(X;\beta)} m^{\mathrm{pre}}(X;\gamma^{\mathrm{pre}}) \right\}_{(z,d)} \\ \left\{ \pi_{(z,d)} - \frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)} \mathbb{I}\{D=d\} + \frac{\mathbb{I}\{Z=z\} - e_z(X;\alpha)}{e_z(X;\alpha)} p_d(X;\eta) \right\}_{(z,d)} \\ \xi_1(W, \alpha) \\ \xi_2(W, \beta) \\ \xi_3(W, \gamma) \\ \xi_3^{\mathrm{pre}}(W, \gamma^{\mathrm{pre}}) \end{bmatrix}, \end{aligned}$$

where W mean the whole observables, ξ_1, ξ_2, ξ_3, and ξ_3^{pre} are estimating equations for the parameters α, β, γ, and γ^{pre}, respectively. Combining this with the moment conditions $g(\zeta, \theta, \vartheta) = 0$, the GMM estimator of ATE is obtained as $\tilde{\omega}$ for

$$\begin{aligned} & (\tilde{\zeta}, \tilde{\theta}, \tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}, \tilde{\gamma}^{\mathrm{pre}}, \tilde{\omega}) \\ = & \arg\min_{\zeta,\theta,\alpha,\beta,\gamma,\gamma^{\mathrm{pre}},\omega} \left[g(\zeta, \theta, \omega)', \frac{1}{n}\sum_{i=1}^n g_1(W_i, \zeta, \alpha, \beta, \gamma, \gamma^{\mathrm{pre}})' \right] \Psi_1 \begin{bmatrix} g(\zeta, \theta, \omega) \\ \frac{1}{n}\sum_{i=1}^n g_1(W_i, \zeta, \alpha, \beta, \gamma, \gamma^{\mathrm{pre}}) \end{bmatrix}, \end{aligned}$$

where Ψ_1 is a weighting matrix. The conventional GMM theory applies to obtain the asymptotic properties of the estimator and statistical inference on ω.

3. Generalizations

In this section, we present three experimental designs to identify and estimate causal objects by auxiliary data: (I) the case where previous treatment and outcome are observable at different time points (Section 3.1), (II) the case where previous treatment and outcome are observable at the same time point (Section 3.2), and (III) the case with a two-regime randomization (Section 3.3).

3.1. **Case I: Observe previous treatment and outcome.** Hereafter we use the following notation. Let $Z^{(1)} \in \{0, 1\}$ be an assignment indicator, $D^{(1)} \in \{0, 1\}$ be a treatment status indicator, and $Y^{(1)} \in \mathcal{Y} \subset \mathbb{R}$ be an outcome of interest, where the superscript "(1)" indicates

the variables associated with the main dataset. Then let $D_z^{(1)} \in \{0,1\}$ be the potential treatment variable realized only when $Z^{(1)} = z$, and $Y_{zd}^{(1)} \in \mathcal{Y}$ be the potential outcome realized only when $Z^{(1)} = z$ and $D^{(1)} = d$. We impose the following basic assumptions.

Assumption 1. *It holds* $Y_d^{(1)} = Y_{zd}^{(1)}$ *for each* $z \in \{0,1\}$ *and* $d \in \{0,1\}$*, and*

$$\begin{aligned} D^{(1)} &= Z^{(1)} D_1^{(1)} + (1 - Z^{(1)}) D_0^{(1)}, \\ Y^{(1)} &= Z^{(1)} D^{(1)} Y_{11}^{(1)} + Z^{(1)}(1 - D^{(1)}) Y_{10}^{(1)} + (1 - Z^{(1)}) D^{(1)} Y_{01}^{(1)} + (1 - Z^{(1)})(1 - D^{(1)}) Y_{00}^{(1)}. \end{aligned}$$

In this case, the principal stratum variable is defined as

$$U = \begin{cases} a & \text{if } D_1^{(1)} = 1, D_0^{(1)} = 1, \\ c & \text{if } D_1^{(1)} = 1, D_0^{(1)} = 0, \\ d & \text{if } D_1^{(1)} = 0, D_0^{(1)} = 1, \\ n & \text{if } D_1^{(1)} = 0, D_0^{(1)} = 0. \end{cases}$$

Note that we do not impose the monotonicity assumption on $D_z^{(1)}$, so the defier can present in the population, i.e., $\mathbb{P}(U = d) > 0$. As in the last section, we are interested in the average treatment effect (ATE $= \mathbb{E}[Y_1^{(1)} - Y_0^{(1)}]$), average effect of treatment on the treated (ATT $= \mathbb{E}[Y_1^{(1)} - Y_0^{(1)} | D_1^{(1)} = 1]$), and compliers' average treatment effect or local average treatment effect (ATE$(c) = \mathbb{E}[Y_1^{(1)} - Y_0^{(1)} | U = c]$). Letting $\mu_d^u = \mathbb{E}[Y_d^{(1)} | U = u]$ and $p^u = \mathbb{P}(U = u)$ for $u \in \{a, c, d, n\}$, these objects can be written as

$$\begin{aligned} \text{ATE}(c) &= \mu_1^c - \mu_0^c, \\ \text{ATT} &= \frac{p^c(\mu_1^c - \mu_0^c) + p^a(\mu_1^a - \mu_0^a)}{p^c + p^a}, \\ \text{ATE} &= p^c(\mu_1^c - \mu_0^c) + p^a(\mu_1^a - \mu_0^a) + p^n(\mu_1^n - \mu_0^n) + p^d(\mu_1^d - \mu_0^d). \end{aligned} \tag{14}$$

Clearly these objects cannot be identified under the present setup. In this subsection, we assume that in addition to the main observations $(Z^{(1)}, D^{(1)}, Y^{(1)})$, the researcher observes:

$$\begin{aligned} D^{(0)} \in \{0,1\} &: \text{ treatment indicator to be observed at the time before the assignment,} \\ Y^{(*)} \in \mathcal{Y} &: \text{ outcome variable to be observed at a time before } D^{(0)}. \end{aligned} \tag{15}$$

Intuitively $D^{(0)}$ is used for untangling the mixtures of principal strata in observations. One of the fundamental challenges of causal inference using randomized experiments with non-compliance is that U is never observed for all subjects because only $D_1^{(1)}$ or $D_0^{(1)}$ can be observed. The observations from randomized experiments can be divided into the four rows in Table 3 (left) according to $D^{(1)}$ and $Z^{(1)}$ values. All of the row are mixtures of two principal stratum, not single principal strata. If one assume absence of defier, parameters $\mu_1^c, \mu_0^c, \mu_1^a, \mu_0^n, p^c, p^a, p^n$ andATE(c) are identified by using the single principal strata moments from 2nd and 3rd rows. But if not, one have to deal with the mixtures in the another way. $D^{(0)}$ is used to overcome this. Even if the mixture problem can be addressed, ATT and ATE cannot be identified. Another fundamental challenge, especially for identifying ATT and ATE, is that the outcome when always-taker receive

U	$Y^{(1)}$	$D^{(1)}$	$Z^{(1)}$
c or a	$Y_1^{(1)}$	1	1
n or d	$Y_1^{(1)}$	0	1
a or d	$Y_0^{(1)}$	1	0
c or n	$Y_0^{(1)}$	0	0

U	$Y^{(1)}$	$D^{(1)}$	$Z^{(1)}$	$Y^{(*)}$	$D^{(0)}$
c	$Y_1^{(1)}$	1	1	$Y^{(*)}$	0
a	$Y_1^{(1)}$	1	1	$Y^{(*)}$	1
n	$Y_1^{(1)}$	0	1	$Y^{(*)}$	0
d	$Y_1^{(1)}$	0	1	$Y^{(*)}$	1
a or d	$Y_0^{(1)}$	1	0	$Y^{(*)}$	1
c or n	$Y_0^{(1)}$	0	0	$Y^{(*)}$	0

TABLE 3. Benchmark case with (right) and without (left) auxiliary data

no treatment and the outcome when never-taker receive treatment is never observed. $Y^{(*)}$ is used to address this problem.

To conduct identification analysis for the objects of interest in (14), we impose the following assumptions.

Assumption 2.

(i): *[Random assignment]* $Z^{(1)}$ *is independent from* $(D^{(0)}, Y^{(*)}, D_1^{(1)}, D_0^{(1)}, Y_{11}^{(1)}, Y_{10}^{(1)}, Y_{01}^{(1)}, Y_{00}^{(1)})$.

(ii): *[Pre-assignment treatment status as control]* $D_0^{(1)} = D^{(0)}$.

Assumption 2 (i) is a standard assumption for random assignment of $Z^{(1)}$. Assumption 2 (ii) says that the auxiliary observation $D^{(0)}$ plays the role of $D_z^{(1)}$ with $Z^{(1)} = 0$. Since the treatment $D^{(0)}$ occurs before the assignment of $Z^{(1)}$, this assumption is reasonable. the relationships of the observables and principal strata variable U can be summarized as in Table 3. Due to Assumptions 2, we do not have rows for the cases of $D^{(1)} \neq D^{(0)}$ with $Z^{(1)} = 0$.

Indeed the first four rows of this table (right panel) suggests that the following objects are identified:

$$\begin{aligned}
\mu_1^c &= \mathbb{E}[Y^{(1)}|Z^{(1)}=1, D^{(1)}=1, D^{(0)}=0], \qquad p^c = \mathbb{P}(D^{(1)}=1, D^{(0)}=0|Z^{(1)}=1),\\
\mu_1^a &= \mathbb{E}[Y^{(1)}|Z^{(1)}=1, D^{(1)}=1, D^{(0)}=1], \qquad p^a = \mathbb{P}(D^{(1)}=1, D^{(0)}=1|Z^{(1)}=1),\\
\mu_0^n &= \mathbb{E}[Y^{(1)}|Z^{(1)}=1, D^{(1)}=0, D^{(0)}=0], \qquad p^n = \mathbb{P}(D^{(1)}=0, D^{(0)}=0|Z^{(1)}=1),\\
\mu_0^d &= \mathbb{E}[Y^{(1)}|Z^{(1)}=1, D^{(1)}=0, D^{(0)}=1], \qquad p^d = \mathbb{P}(D^{(1)}=0, D^{(0)}=1|Z^{(1)}=1).
\end{aligned} \tag{16}$$

Furthermore, the last two rows of this table (right panel) can be utilized to identify

$$\begin{aligned}
\mu_1^d &= \frac{(p^a+p^d)\mathbb{E}[Y^{(1)}|Z^{(1)}=0, D^{(1)}=1, D^{(0)}=1] - p^a\mu_1^a}{p^d},\\
\mu_0^c &= \frac{(p^c+p^n)\mathbb{E}[Y^{(1)}|Z^{(1)}=0, D^{(1)}=0, D^{(0)}=0] - p^n\mu_0^n}{p^c}.
\end{aligned} \tag{17}$$

Therefore, under Assumptions 1-2, we can identify ATE for compliers and defiers as $\text{ATE}(c) = \mu_1^c - \mu_0^c$ and $\text{ATE}(d) = \mu_1^d - \mu_0^d$, respectively.

In order to identify ATT and ATE, it remains to identify μ_0^a and μ_1^n by using the additional data $Y^{(*)}$. To this end, we add the following assumptions.

Assumption 3.

(i): *[Parallel trend of nonreactive strata]* $\mathbb{E}[Y_0^{(1)} - Y^{(*)}|U=a] = \mathbb{E}[Y_0^{(1)} - Y^{(*)}|U=n]$.

(ii): *[Homogeneity of nonreactive strata]* $\mathbb{E}[Y_1^{(1)} - Y_0^{(1)}|U = a] = \mathbb{E}[Y_1^{(1)} - Y_0^{(1)}|U = n]$.

Assumption 3 (i) is an analog of the parallel trend assumption on the types a and n whose participation decisions are not affected by $Z^{(1)}$. Assumption 3 (ii) requires homogeneous treatment effects on the types a and n. Note that in the conventional identification analysis for ATE, we typically impose homogeneity over all types. On the other hand, we only require homogeneity over the types a and n. To see how Assumption 3 (i) is utilized to identify μ_0^a, observe that

$$\begin{aligned}\mu_1^a - \mu_0^a &= \mathbb{E}[Y_1^{(1)} - Y^{(*)}|U = a] - \mathbb{E}[Y_0^{(1)} - Y^{(*)}|U = a] \\ &= \mathbb{E}[Y_1^{(1)} - Y^{(*)}|U = a] - \mathbb{E}[Y_0^{(1)} - Y^{(*)}|U = n] \\ &= \mu_1^a - \mu_*^a - \mu_0^n + \mu_*^n, \end{aligned} \tag{18}$$

where $\mu_*^a = \mathbb{E}[Y^{(*)}|U = a]$, $\mu_*^n = \mathbb{E}[Y^{(*)}|U = n]$, and the second equality uses Assumption 3 (i). Since μ_*^a and μ_*^n are identified by

$$\begin{aligned}\mu_*^a &= \mathbb{E}[Y^{(*)}|Z^{(1)} = 1, D^{(1)} = 1, D^{(0)} = 1], \\ \mu_*^n &= \mathbb{E}[Y^{(*)}|Z^{(1)} = 1, D^{(1)} = 0, D^{(0)} = 0], \end{aligned} \tag{19}$$

we can identify μ_0^a by (18) and thus ATT is also identified by (14). Finally, Assumption 3 (ii) guarantees identification of μ_1^n as $\mu_1^n = \mu_0^n + \mu_1^a - \mu_0^a$ so that ATE is identified by the expression in (14).

Combining these results, identification of the causal objects in (14) is established as follows.

Theorem 4. *Consider the setup of this subsection.*

(i): *Under Assumptions 1-2,* ATE(c) *is identified.*

(ii): *Under Assumptions 1-3 (i),* ATT *is identified.*

(iii): *Under Assumptions 1-3,* ATE *is identified.*

Based on this theorem, we can estimate these causal objects by taking sample counterparts, and conduct inference based on standard methods, such as the delta method and bootstrap.

Remark 2. [Alternative assumptions] If defiers exist, then Assumption 3 could be replaced with another reasonable assumption on the targeted outcome and situation. The group that receives treatment on their own initiative without any external incentives (the group of $D_0^{(1)} = 1$ including always-takers and defiers) may share common characteristics in that they expect to have worse outcomes if they do not receive treatment. In this case, Assumption 3 (i) may be replaced with

$$\text{Assumption 3 (i)':} \quad \mathbb{E}[Y_0 - Y^*|U = a] = \mathbb{E}[Y_0 - Y^*|U = d].$$

In addition, the group that does not receive treatment even if they receive an external incentive (the group of $D_1^{(1)} = 0$ including never-takers and defiers) may have common characteristics in that they expect the outcomes do not change much even if they receive treatment. In this case, Assumption 3 (ii) may be replaced with

$$\text{Assumption 3 (ii)':} \quad \mathbb{E}[Y_1 - Y^*|U = n] = \mathbb{E}[Y_1 - Y^*|U = d].$$

Assumption 3 (i)' can be used to identify μ_0^a, and Assumption 3 (ii)' can be used to identify μ_1^n. ATT and ATE are identified in analogous ways. When using Assumption 3 (ii)', it is not necessary to assume the same average treatment effect as in Assumption 3 (ii).

Remark 3. [Overidentification] The above argument for establishing Theorem 4 (iii) is based on showing just identification of the 14 parameters, (μ_1^u, μ_0^u, p^u) for $u \in \{c, a, n, d\}$ and μ_*^u for $u \in \{a, n\}$. Indeed by introducing two more parameters (μ_*^c, μ_*^d), we have four additional restrictions:

$$\begin{aligned} p^a + p^d &= \mathbb{E}[D^{(1)} = 1, D^{(0)} = 1 | Z^{(1)} = 0], \\ p^c + p^n &= \mathbb{E}[D^{(1)} = 0, D^{(0)} = 0 | Z^{(1)} = 0], \\ \mu_*^a &= \frac{(p^a + p^d)\mathbb{E}[Y^{(*)} | Z^{(1)} = 0, D^{(1)} = 1, D^{(0)} = 1] - p^d \mu_*^d}{p^a}, \\ \mu_*^n &= \frac{(p^c + p^n)\mathbb{E}[Y^{(*)} | Z^{(1)} = 0, D^{(1)} = 0, D^{(0)} = 0] - p^c \mu_*^c}{p^n}. \end{aligned} \tag{20}$$

Here $\mu_*^c = \mathbb{E}[Y^{(*)}|U = c]$ and $\mu_*^d = \mathbb{E}[Y^{(*)}|U = d]$ are identified by

$$\begin{aligned} \mu_*^c &= \mathbb{E}[Y^{(*)} | Z^{(1)} = 1, D^{(1)} = 1, D^{(0)} = 0], \\ \mu_*^d &= \mathbb{E}[Y^{(*)} | Z^{(1)} = 1, D^{(1)} = 0, D^{(0)} = 1]. \end{aligned} \tag{21}$$

These additional moment conditions can be incorporated by using the generalized method of moments.

3.1.1. *Identification under ignorability condition.* In observational studies, it is often the case that the random assignment of $Z^{(1)}$ (Assumption 2 (i)) is violated. In this subsection we show that our identification argument can be extended to the case where the following ignorability condition is satisfied. Let $X \in \mathcal{X} \subset \mathbb{R}^q$ be a vector of q-dimensional covariates.

Assumption 2. ***(i)'*** *[Ignorability] Conditionally on X, $Z^{(1)}$ is independent from $(D^{(0)}, Y^{(*)}, D_1^{(1)}, D_0^{(1)}, Y_{11}^{(1)}, Y_{10}^{(1)}, Y_{01}^{(1)}, Y_{00}^{(1)})$.*

This is a standard ignorability or unconfoundedness condition commonly imposed in the literature of causal inference with observational studies. Based on the discussion of the previous subsection, it is sufficient for identification of the causal estimands in (14) to identify

$$\begin{aligned} \delta_{(z,d,d')}^{(t)} &= \mathbb{E}[Y_d^{(t)} | D_z^{(1)} = d, D^{(0)} = d'], \\ \pi_{(z,d,d')} &= \mathbb{P}(D_z^{(1)} = d, D^{(0)} = d'), \end{aligned} \tag{22}$$

for each $t \in \{1, *\}$, $z \in \{0, 1\}$, and $d, d' \in \{0, 1\}$. To establish multiply robust representations of $\delta_{(z,d,d')}^{(t)}$ and $\pi_{(z,d,d')}$ under Assumption 2 (i)', we introduce parametric models

$$\begin{aligned} e_z(X; \alpha) &\quad \text{for} \quad \mathbb{P}(Z^{(1)} = z | X), \\ p_{(z,d,d')}(X; \beta) &\quad \text{for} \quad \mathbb{P}(D^{(1)} = d, D^{(0)} = d' | Z^{(1)} = z, X), \\ m^{(t)}(X; \gamma^{(t)}) &\quad \text{for} \quad \mathbb{E}[Y^{(t)} | X], \\ p_{(d,d')}(X; \eta) &\quad \text{for} \quad \mathbb{P}(D^{(1)} = d, D^{(0)} = d' | X), \end{aligned}$$

where α, β, $\gamma^{(t)}$, and η are finite dimensional parameters. By using these parametric models, multiply robust representations of the population objects $\delta^{(t)}_{(z,d,d')}$ and $\pi_{(z,d,d')}$ are obtained as follows.

Theorem 5. *Consider the setup of this subsection. Under Assumptions 1 and 2 (i)', (ii), and (iii), it holds*

$$\begin{aligned}\delta^{(t)}_{(z,d,d')} &= \mathbb{E}\left[\frac{\mathbb{I}\{Z^{(1)}=z\}}{e_z(X;\alpha)}\frac{\mathbb{I}\{D^{(1)}=d,D^{(0)}=d'\}}{p_{(z,d,d')}(X;\beta)}Y^{(t)}\right]\\ &\quad -\mathbb{E}\left[\frac{\mathbb{I}\{Z^{(1)}=z\}\mathbb{I}\{D^{(1)}=d,D^{(0)}=d'\}-e_z(X;\alpha)p_{(z,d,d')}(X;\beta)}{e_z(X;\alpha)p_{(z,d,d')}(X;\beta)}m^{(t)}(X;\gamma^{(t)})\right],\\ \pi_{(z,d,d')} &= \mathbb{E}\left[\frac{\mathbb{I}\{Z^{(1)}=z\}}{e_z(X;\alpha)}\mathbb{I}\{D^{(1)}=d,D^{(0)}=d'\}\right]-\mathbb{E}\left[\frac{\mathbb{I}\{Z^{(1)}=z\}-e_z(X;\alpha)}{e_z(X;\alpha)}p_{(d,d')}(X;\eta)\right].\end{aligned}$$

By taking the sample counterparts of these representations, we can construct multiply robust estimators for $\delta^{(t)}_{(z,d,d')}$ and $\pi_{(z,d,d')}$. Then the 16 parameters $\mu^u_1,\mu^u_0,\mu^u_*,p^u$ for $u\in\{c,a,n,d\}$ are over-identified by the moment restrictions:

$$\begin{aligned}&\mu^c_1=\delta^{(1)}_{(1,1,0)},\quad \mu^a_1=\delta^{(1)}_{(1,1,1)},\quad \mu^n_0=\delta^{(1)}_{(1,0,0)},\quad \mu^d_0=\delta^{(1)}_{(1,0,1)},\\ &p^c=\pi_{(1,1,0)},\quad p^a=\pi_{(1,1,1)},\quad p^n=\pi_{(1,0,0)},\quad p^d=\pi_{(1,0,1)},\\ &\mu^d_1=\frac{(p^a+p^d)\delta^{(1)}_{(0,1,1)}-p^a\mu^a_1}{p^d},\quad \mu^c_0=\frac{(p^c+p^n)\delta^{(1)}_{(0,0,0)}-p^n\mu^n_0}{p^c},\\ &\mu^a_0=\mu^a_*+\mu^n_0-\mu^n_*,\quad \mu^n_1=\mu^n_0+\mu^a_1-\mu^a_0,\quad \mu^a_*=\delta^{(*)}_{(1,1,1)},\quad ,\mu^n_*=\delta^{(*)}_{(1,0,0)}\\ &p^a+p^d=\pi_{(0,1,1)},\quad p^c+p^n=\pi_{(0,0,0)},\quad \mu^c_*=\delta^{(*)}_{(1,1,0)},\quad \mu^d_*=\delta^{(*)}_{(1,0,1)},\\ &\mu^a_*=\frac{(p^a+p^d)\delta^{(*)}_{(0,1,1)}-p^d\mu^d_*}{p^a},\quad \mu^n_*=\frac{(p^c+p^n)\delta^{(*)}_{(0,0,0)}-p^c\mu^c_*}{p^n},\end{aligned}\tag{23}$$

where the first three lines are obtained from the rows in Table 3, the fourth line is obtained from Assumption 3 and the second and third rows in Table 3, and the last two lines follow from the first and last three rows in Table 3. Just identification of $\mu^u_1,\mu^u_0,\mu^u_*,p^u$ for $u\in\{c,a,n,d\}$ is guaranteed by the first four lines, and the last two lines provide overidentifying restrictions.

We close this subsection by summarizing the doubly robust properties of the estimators based on Theorem 5 and (23).

Proposition 2. *Consider the setup of this subsection. Suppose Assumptions 1, 2 (i)', (ii), and (iii), and 3 hold true. Then*

- **(i):** $\delta^{(t)}_{(z,d,d')}$ *can be consistently estimated if either* $\{e_z(X;\alpha),p_{(z,d,d')}(X;\beta)\}$ *or* $m^{(t)}(X;\gamma^{(t)})$ *is correctly specified,*
- **(ii):** $\pi_{(z,d,d')}$ *can be consistently estimated if either* $e_z(X;\alpha)$ *or* $p_{(d,d')}(X;\eta)$ *is correctly specified,*
- **(iii):** ATE(c) *can be consistently estimated if either* $\{e_z(X;\alpha),p_{(z,d,d')}(X;\beta)\}$, $\{m^{(1)}(X;\gamma^{(1)}),p_{(d,d')}(X;\eta)\}$, *or* $\{e_z(X;\alpha),m^{(1)}(X;\gamma^{(1)})\}$ *is correctly specified,*

(iv): ATT *and* ATE *can be consistently estimated if either* $\{e_z(X;\alpha), p_{(z,d,d')}(X;\beta)\}$, $\{m^{(1)}(X;\gamma^{(1)}), m^{(*)}(X;\gamma^{(*)}), p_{(d,d')}(X;\eta)\}$, *or* $\{e_z(X;\alpha), m^{(1)}(X;\gamma^{(1)}), m^{(*)}(X;\gamma^{(*)})\}$ *is correctly specified.*

Furthermore, the multiply robust estimator for ATE(c) is asymptotically locally efficient if $\{e_z(X;\alpha), p_{(z,d,d')}(X;\beta), m^{(1)}(X;\gamma^{(1)}), p_{(d,d')}(X;\eta)\}$ are correctly specified, and also the multiply robust estimators for ATT and ATE are asymptotically locally efficient if $\{e_z(X;\alpha), p_{(z,d,d')}(X;\beta), m^{(1)}(X;\gamma^{(1)}), m^{(*)}(X;\gamma^{(*)})\}$ are correctly specified.

3.1.2. *Estimation and testing.* In this subsection, we briefly discuss estimation and testing methods for ATE identified by Theorems 4 and 5 above. The methods for ATE(c) and ATT can be obtained in the same manner.

First, we consider estimation of ATE based on Theorem 4 (iii). Let $\hat{\delta}^{(t)}_{(z,d,d')}$ and $\hat{\pi}_{(z,d,d')}$ be the empirical (conditional) moments of $\delta^{(t)}_{(z,d,d')} = \mathbb{E}[Y^{(t)}_d | D^{(1)}_z = d, D^{(0)} = d']$ and $\pi_{(z,d,d')} = \mathbb{P}(D^{(1)}_z = d, D^{(0)} = d')$, respectively, and $\hat{\zeta}$ and ζ be their vectorizations. Also let θ be a 14-dimensional vector given by (μ^u_1, μ^u_0, p^u) for $u \in \{c, a, n, d\}$ and μ^u_* for $u \in \{a, n\}$, which provides a formula for ATE as

$$\mathrm{ATE}(\theta) = p^c(\mu^c_1 - \mu^c_0) + p^a(\mu^a_1 - \mu^a_0) + p^n(\mu^n_1 - \mu^n_0) + p^d(\mu^d_1 - \mu^d_0).$$

Then the minimum distance estimator for ATE is obtained as $\hat{\omega}$ for

$$(\hat{\theta}, \hat{\omega}) = \arg\min_{\theta,\omega} g(\hat{\zeta}, \theta, \omega)' \Psi g(\hat{\zeta}, \theta, \omega), \tag{24}$$

where the vector of moment conditions $g(\zeta, \theta, \omega) = 0$ is obtained by stacking the equations (16)-(19) and $\omega = \mathrm{ATE}(\theta)$ (and also (20)-(21)). The weight matrix Ψ may be chosen to achieve the asymptotic efficiency (see, e.g., Newey and McFadden, 1994). Statistical inference on ω can be conducted by the Wald statistic, likelihood ratio-type statistic, or bootstrap method.

Next, if the parameters ζ are identified by the ignorability condition as in Theorem 5, their estimating equations are given by

$$\begin{aligned} & g_1(W, \zeta, \alpha, \beta, \{\gamma^{(t)}\}_t) \\ = & \begin{bmatrix} \left\{ \begin{array}{c} \delta^{(t)}_{(z,d,d')} - \frac{\mathbb{I}\{Z^{(1)}=z\}}{e_z(X;\alpha)} \frac{\mathbb{I}\{D^{(1)}=d, D^{(0)}=d'\}}{p_{(z,d,d')}(X;\beta)} Y^{(t)} \\ + \frac{\mathbb{I}\{Z^{(1)}=z\}\mathbb{I}\{D^{(1)}=d, D^{(0)}=d'\} - e_z(X;\alpha) p_{(z,d,d')}(X;\beta)}{e_z(X;\alpha) p_{(z,d,d')}(X;\beta)} m^{(t)}(X;\gamma^{(t)}) \end{array} \right\}_{(t,z,d,d')} \\ \left\{ \pi_{(z,d,d')} - \frac{\mathbb{I}\{Z^{(1)}=z\}}{e_z(X;\alpha)} \mathbb{I}\{D^{(1)} = d, D^{(0)} = d'\} + \frac{\mathbb{I}\{Z^{(1)}=z\} - e_z(X;\alpha)}{e_z(X;\alpha)} p_{(d,d')}(X;\eta) \right\}_{(z,d,d')} \\ \xi_1(W, \alpha) \\ \xi_2(W, \beta) \\ \{\xi^{(t)}_3(W, \gamma^{(t)})\}_t \end{bmatrix}, \end{aligned}$$

where W mean the whole observables, ξ_1, ξ_2, and $\xi^{(t)}_3$ are estimating equations for the parameters α, β, and $\gamma^{(t)}$, respectively. Combining this with the moment conditions $g(\zeta, \theta, \vartheta) = 0$, the GMM

U	$Y^{(1)}$	$D^{(1)}$	$Z^{(1)}$	$Y^{(0)}$	$D^{(0)}$
c	$Y_1^{(1)}$	1	1	$Y_0^{(0)}$	0
a	$Y_1^{(1)}$	1	1	$Y_1^{(0)}$	1
n	$Y_1^{(1)}$	0	1	$Y_0^{(0)}$	0
d	$Y_1^{(1)}$	0	1	$Y_1^{(0)}$	1
a or d	$Y_0^{(1)}$	1	0	$Y_1^{(0)}$	1
c or n	$Y_0^{(1)}$	0	0	$Y_0^{(0)}$	0

TABLE 4. Case of observable $Y^{(0)}$ instead of $Y^{(*)}$

estimator of ATE is obtained as $\tilde{\omega}$ for

$$\begin{aligned}&(\tilde{\zeta}, \tilde{\theta}, \tilde{\alpha}, \tilde{\beta}, \{\tilde{\gamma}^{(t)}\}_t, \tilde{\omega}) \\ = \ &\arg \min_{\zeta,\theta,\alpha,\beta,\{\gamma^{(t)}\}_t,\omega} \left[g(\zeta,\theta,\omega)', \frac{1}{n}\sum_{i=1}^n g_1(W_i,\zeta,\alpha,\beta,\{\gamma^{(t)}\}_t)' \right] \Psi_1 \left[\begin{array}{c} g(\zeta,\theta,\omega) \\ \frac{1}{n}\sum_{i=1}^n g_1(W_i,\zeta,\alpha,\beta,\{\gamma^{(t)}\}_t) \end{array} \right],\end{aligned}$$

where Ψ_1 is a weighting matrix. The conventional GMM theory applies to obtain the asymptotic properties of the estimator and statistical inference on ω.

3.2. **Case II: Identification with $Y^{(0)}$ instead of $Y^{(*)}$.** In the benchmark case considered in the last section, a critical requirement is availability of the outcome $Y^{(*)}$ that is observed at a point in time before $D^{(0)}$ so that $Y^{(*)}$ is not affected by $D^{(0)}$. This subsection considers the situation where we observe the outcome $Y^{(0)}$ instead of $Y^{(*)}$ and treatment $D^{(0)}$ at the same time so that $Y^{(0)}$ is affected by $D^{(0)}$. In this case, the relationships of the observables and principal strata variable are summarized as in Table 4.

In this case, we introduce alternative assumptions to identify μ_0^a.

Assumption 2a. *The observable $Y^{(0)}$ satisfies*

$$Y^{(0)} = D^{(0)}Y_1^{(0)} + (1 - D^{(0)})Y_0^{(0)},$$

where $Y_d^{(0)}$ is the potential outcome realized only when $D^{(0)} = d$, and Assumption 2 holds true with replacement of "$Y^{(*)}$" with "$Y^{(0)}$".

Assumption 3a.

- **(i):** *[Parallel trend for always-takers and defiers]* $\mathbb{E}[Y_0^{(1)} - Y_1^{(0)}|U = a] = \mathbb{E}[Y_0^{(1)} - Y_1^{(0)}|U = d]$.
- **(ii):** *[Parallel trend for compliers and never-takers]* $\mathbb{E}[Y_1^{(1)} - Y_0^{(0)}|U = c] = \mathbb{E}[Y_1^{(1)} - Y_0^{(0)}|U = n]$.

Assumptions 2a is analogous to Assumption 2 for the benchmark case. Up to the argument in (17) for identification of μ_1^d and μ_0^c, we can proceed in the same way as the benchmark case. Thus, to identify ATT and ATE, it remains to identify μ_0^a and μ_1^n. Now Assumptions 2a and 3a (i) imply

$$\begin{aligned}\mu_1^a - \mu_0^a &= \mathbb{E}[Y_1^{(1)} - Y_1^{(0)}|U = a] - \mathbb{E}[Y_0^{(1)} - Y_1^{(0)}|U = a] \\ &= \mathbb{E}[Y_1^{(1)} - Y_1^{(0)}|U = a] - \mathbb{E}[Y_0^{(1)} - Y_1^{(0)}|U = d],\end{aligned}$$

R	U	$Y^{(1)}$	$D^{(1)}$	$Z^{(1)}$	$D^{(0)}$
1	c	$Y_1^{(1)}$	1	1	0
1	a	$Y_1^{(1)}$	1	1	1
1	n	$Y_1^{(1)}$	0	1	0
1	d	$Y_1^{(1)}$	0	1	1
1	a or d	$Y_0^{(1)}$	1	0	1
1	c or n	$Y_0^{(1)}$	0	0	0
0	all	$Y_0^{(1)}$	0	-	-

TABLE 5. Two-regime case

which can be written as

$$\mu_0^a = \mathbb{E}[Y_1^{(0)}|U=a] + \mu_0^d - \mathbb{E}[Y_1^{(0)}|U=d]. \tag{25}$$

By using (2) and (25), we can identify ATT. Similarly, Assumptions 2a and 3a (ii) imply

$$\mu_1^n = \mu_1^c - \mathbb{E}[Y_0^{(0)}|U=c] + \mathbb{E}[Y_0^{(0)}|U=n]. \tag{26}$$

Based on (25) and/or (26), we can identify ATE under three scenarios. The identification results for this case are summarized as follows.

Theorem 6. *Consider the setup of this subsection.*

(i): *Under Assumptions 1 and 2a,* ATE(c) *is identified.*

(ii): *Under Assumptions 1, 2a, and 3a (i),* ATT *is identified.*

(iii): *Suppose Assumptions 1 and 2a hold true. If either (a) Assumptions 3a (i) and 3 (ii); (b) Assumptions 3a (ii), and 3 (ii); or (c) Assumptions 3a (i) and 3a (ii) holds true, then* ATE *is identified.*

3.3. **Case III: Two-regime design.** In this subsection, we consider a two-regime setting, where we do not need to observe neither $Y^{(0)}$ or $Y^{(*)}$. First, subjects are randomly assigned to one of two regimes $R \in \{1,0\}$. In the group with $R=1$, we observe $(Y^{(1)}, D^{(1)}, Z^{(1)}, D^{(0)})$. In the group with $R=0$, $D^{(1)}=0$ is forced so that we observe $Y^{(1)} = Y_0^{(1)}$. In other words, we block access to the treatment for a randomly selected subgroup. In this case, the relationships of the observables and principal strata variable are summarized in Table 5.

In this setup, we impose the following assumptions.

Assumption 2b. *Assumption 2 holds true with removal of* $Y^{(*)}$.

Assumption 3b.

(i): *[Random regime assignment]* R *is independent from* $(D^{(0)}, D_1^{(1)}, D_0^{(1)}, Y_{11}^{(1)}, Y_{10}^{(1)}, Y_{01}^{(1)}, Y_{00}^{(1)})$.

(ii): *[Block to treatment for* $R=0$*]* $\mathbb{E}[Y^{(1)}|R=0] = \mathbb{E}[Y_0^{(1)}|R=0]$.

(iii): *Assumption 3 (ii) holds true.*

Up to the argument in (17) for identification of μ_1^d and μ_0^c, we can proceed in the same way as the benchmark case. So it remains to identify μ_0^a and μ_1^n for identification of ATT and ATE.

Now for the data with $R = 0$ (i.e., the last row of Table 5), Assumption 3b (i) and (ii) imply

$$\begin{aligned}\mu_0^a &= \frac{\mathbb{E}[Y_0^{(1)}] - (p^c\mu_0^c + p^n\mu_0^n + p^d\mu_0^d)}{p^a} \\ &= \frac{\mathbb{E}[Y_0^{(1)}|R=0] - (p^c\mu_0^c + p^n\mu_0^n + p^d\mu_1^d)}{p^a}.\end{aligned}$$

Also as in the benchmark case, Assumption 3b (iii) guarantees identification of μ_1^n as $\mu_1^n = \mu_0^n + \mu_1^a - \mu_0^a$. Combining these results, we obtain the following identification results.

Theorem 7. *Consider the setup of this subsection.*

(i): *Under Assumptions 1 and 2b,* ATE(c) *is identified.*

(ii): *Under Assumptions 1, 2b and 3b (i)-(ii),* ATT *is identified.*

(iii): *Under Assumptions 1, 2b and 3b,* ATE *is identified.*

When we additionally observe the treatment $D^{(0)}$ for the group with $R = 0$, our identification analysis can be modified by splitting the last row of Table 5 into two rows depending on the value of $D^{(0)}$.

4. Empirical illustrations

4.1. **Case with monotonicity.** This section illustrates the proposed identification and estimation methods by revisiting three important empirical studies in the literature. Thornton (2008), Gerber et al. (2009), and Beam (2016) attempted to understand the following causal effects through randomized comparison experiments: the effect of knowing one's HIV status on promoting contraceptive behavior, the effect of subscribing to a particular newspaper on political attitudes, and the effect of participating in a job fair on increasing one's intention to work abroad, respectively. In these studies, the treatments are difficult to enforce on the subjects, so they adopted encouragement designs with incentives. We revisited their data using the methods in this paper. In Thornton (2008) data, we adopted the indicator whether each of subjects purchased condoms at the time of the follow-up survey and the indicator whether each of subjects reported having sex between the time of the baseline and the time of the follow-up survey as outcomes of interest. And the indicator whether each of subjects reported using a condom during the last year at the baseline and the indicator whether each of subjects reported having sex in the past year were used as auxiliary observation, respectively. In Gerber et al. (2009) data, we adopt the indicator whether each of subjects voted in the 2005 election after the experiment and the indicator whether each of subjects preferred the Democratic Party. The indicator whether each of subjects voted in the 2024 election before the experiment and the indicator whether each of subjects preferred the Democratic Party at the baseline were used as the auxiliary outcomes, respectively. In Beam (2016) data, the indicator whether each of subjects planned to work abroad at follow-up and the indicator whether each of subjects held a passport. The same questions from the baseline survey were used as the auxiliary outcomes. For all analyses, subjects who had completed follow-up surveys and for whom all variables used in each of the analysis were available were included. And for analysis of Thornton (2008) data

with purchasing condom as a outcome, we included subjects who reported on the baseline survey that they had sex in the past 12 months. Standard errors and p-values were calculated based on 200 bootstraps in all analyses. Estimated results from this study are shown in Table 6.

In Thornton (2008) data, higher negative effects were estimated for ATT and ATE than for LATE, with ATE being significant. This result indicates that in the entire population, knowledge of infection status may inhibit contraceptive behavior. The data include a much larger number of HIV-negative subjects than among HIV-positive subjects. Since the data included more negative subjects than HIV-positive subjects, the results may suggest that knowledge of infection status may inhibit contraceptive behavior. In Gerber et al. (2009) data, higher positive effects were estimated for ATT and ATE than for LATE, with ATE being significant. This result suggests that newspaper subscriptions may be more effective for noncompliers than for compliers. In Beam (2016) data, non-significant results were estimated for LATE, ATT, and ATE, suggesting that there may be no effect either for compliers or for the entire population. Our method also allows us to estimate the average treatment effect for each principal stratum. The results are presented in Table 7. In the analysis of Purchase Condom in Thornton (2008), where ATE is estimated that yields a different perspective from LATE, non-compliers have a larger negative treatment effect than compliers. Similarly, in the analysis of Voted 2005 in Gerber et al. (2009), non-compliers have higher treatment effects than compliers. It is not easy to look at the background context in which these differences arise, but inferring the profile of each stratum (Marbach and Hangartner, 2020) would provide more insight.

As described above, the analysis conducted in this study produced results that suggest different implications from the average treatment effects of compliers for some of the outcomes. This result is due to differences between compliers and noncompliers, and it is important to conduct an analysis that does not assume homogeneity between compliers and noncompliers, as our method does. On the other hand, since the parallel trend and homogeneity assumptions between always-takers and never-takers used in our analysis are not testable, it is important to discuss further whether violations of these assumptions occur based on domain knowledge.

	Thornton (2008)		Gerber et al. (2009)		Beam (2016)	
	Purchase	Having sex	Voted in 2005	Democratic	Plan to abroad	Passport
ITT	-0.010	-0.003	0.004	-0.014	-0.020	-0.004
	(0.022)	(0.032)	(0.028)	(0.018)	(0.022)	(0.019)
LATE	-0.024	-0.007	0.015	-0.057	-0.061	-0.012
	(0.053)	(0.073)	(0.116)	(0.077)	(0.067)	(0.056)
ATT	-0.074	0.000	0.067	-0.019	-0.068	0.000
	(0.038)	(0.040)	(0.058)	(0.038)	(0.048)	(0.038)
ATE	**-0.084**	0.001	**0.096**	0.003	-0.078	0.016
	(0.041)	(0.040)	(0.047)	(0.040)	(0.051)	(0.029)
n	1,006	1,301	1,079	1,081	865	865

TABLE 6. Estimates and bootstrap standard error

	Outcome	Parameters	Complier	Always-taker	Never-taker
Thornton (2008)	Purchase Condom	p^u	0.425	0.390	0.185
		μ_1^u	0.092	0.084	-0.039
		μ_0^u	0.116	0.212	0.088
		$\mu_1^u - \mu_0^u$	-0.024	-0.127	-0.127
	Having Sex	p^u	0.436	0.380	0.185
		μ_1^u	0.721	0.660	0.628
		μ_0^u	0.728	0.653	0.621
		$\mu_1^u - \mu_0^u$	-0.007	0.007	0.007
Gerber et al. (2009)	Voted in 2005	p^u	0.243	0.225	0.532
		μ_1^u	0.788	0.775	0.806
		μ_0^u	0.773	0.653	0.685
		$\mu_1^u - \mu_0^u$	0.016	0.122	0.122
	Prefers Democratic	p^u	0.243	0.225	0.533
		μ_1^u	0.033	0.128	0.131
		μ_0^u	0.090	0.105	0.108
		$\mu_1^u - \mu_0^u$	-0.057	0.023	0.023
Beam (2016)	Plan to abroad	p^u	0.337	0.136	0.527
		μ_1^u	0.084	0.275	-0.025
		μ_0^u	0.144	0.363	0.062
		$\mu_1^u - \mu_0^u$	-0.061	-0.088	-0.088
	Passport	p^u	0.337	0.136	0.527
		μ_1^u	0.038	0.175	0.086
		μ_0^u	0.050	0.144	0.055
		$\mu_1^u - \mu_0^u$	-0.012	0.031	0.031

TABLE 7. Probabilities and effects

4.2. **Case without monotonicity.** Alcohol beverage manufacturers regularly introduce new products to the market in response to changing consumer needs, and often one manufacturer will handle multiple products within the same category (beer, RTD, etc.). Therefore, they are interested in whether the new products they have introduced have succeeded in increasing the total sales of their own category without causing cannibalization of their own products. Because the structure of such markets tends to change, it is difficult to judge success or failure by comparing a company's total category sales before and after a product launch. Even if we tried to randomly assign purchases of new products, it would be practically difficult to assign purchases or non-purchases. Therefore, it is reasonable to conduct an experiment to encourage purchases with coupons and to conduct an analysis that allows for non-compliance. The response rate to coupons (the likelihood of purchasing the target product when given a coupon) is generally not high, so the percentage of compliers is not large and it is difficult to say that they are representative of the population. We will therefore attempt to estimate ATE using the proposed method. This kind of randomized encouragement design (Imbens and Rubin, 2015) and the estimation proposed here allows for the existence of consumers who are averse to intentional sales promotions by firms and does not assume monotony. We will use data from a randomized encouragement design experiment conducted by a Japanese alcoholic beverage

manufacturer on a new product in the beer category. The experiment was conducted in May 2023 at stores of a major retail chain. There are 133,733 subjects in the experiment, 80,000 in the treatment group, and 53,733 in the control group. Let $Z^{(1)}$ be the coupon assignment and $D^{(1)}$ be whether or not each subject purchased the new product in the week following the coupon assignment. There are four outcomes of interest $Y^{(1)}$: the amount spent by each subject on products in the beer category from this manufacturer; the amount spent by each subject on products in the beer and RTD category from this manufacturer; the amount spent by each subject on products excluding the new product in the beer category from this manufacturer; and the amount spent by each subject on products excluding the new product in the beer and RTD category from this manufacturer. These outcomes are measured over a one-week period following coupon assignment. For each $Y^{(1)}$, let $Y^{(*)}$ be measured for one week in March 2023, before the new product is released. Let $D^{(0)}$ be whether or not each subject purchased the new product during the week in May before the experiment. During this period, no coupons for this new product were distributed. We estimate LATE, ATT, and ATE with the assumptions of unstable.

The results of the estimation are presented in Table 8. In all estimates (LATE, ATT, ATE), the total sales of the category including the new product increased significantly, and the change in the total sales of the category excluding the new product was not significant. These results indicate that there was no cannibalization within the category and that the entry of the new product increased the total sales of the category. Comparing the estimated values for LATE, ATT, and ATE shows that LATE underestimates increased sales. In addition, it was estimated as $(p^a, p^c, p^d, p^n) = (0.001, 0.015, 0.013, 0.97)$. The large proportion of never-takers indicates that there are few purchasers of new products, whether consumers have coupons or not. Since getting consumers to buy this new product could lead to an increase in total sales for the category, it would probably be worth spending more on sales promotion to get more new purchasers.

	Including the new product		Excluding the new product	
	Beer category	Beer and RTD category	Beer category	Beer and RTD category
LATE	**537.8**	**570.1**	-10.9	21.5
	(145.1)	(259.0)	(161.1)	(243.5)
ATT	**550.3**	**578.3**	-21.6	6.3
	(133.3)	(242.6)	(150.2)	(226.6)
ATE	**713.4**	**681.9**	-165.2	-196.7
	(197.4)	(252.2)	(133.0)	(215.2)

Note: Significant estimates are denoted in boldface type. Parentheses refer to standard deviations with 200 bootstraps. The unit is yen.

TABLE 8. Estimates and bootstrap standard deviations of new product effects

Appendix A. Mathematical appendix

A.1. **Proof of Theorem 1.** First, note that $\mu_1^a, \mu_0^n, p^a, p^n, p^c, \mu_1^c, \mu_0^c$ are identified under Assumptions Y (i), (ii), and (iv) as

$$\begin{aligned}
\mu_1^a &= \mathbb{E}[Y_1^{(1)}|D_1^{(1)}=1, D_0^{(1)}=1] = \mathbb{E}[Y_1^{(1)}|D_0^{(1)}=1] = \mathbb{E}[Y^{(1)}|Z^{(1)}=0, D^{(1)}=1],\\
\mu_0^n &= \mathbb{E}[Y_0^{(1)}|D_1^{(1)}=0, D_0^{(1)}=0] = \mathbb{E}[Y_0^{(1)}|D_1^{(1)}=0] = \mathbb{E}[Y^{(1)}|Z^{(1)}=1, D^{(1)}=0],\\
p^a &= \mathbb{P}(D_1^{(1)}=1, D_0^{(1)}=1) = \mathbb{P}(D_0^{(1)}=1) = \mathbb{P}(D^{(1)}=1|Z^{(1)}=0),\\
p^n &= \mathbb{P}(D_1^{(1)}=0, D_0^{(1)}=0) = \mathbb{P}(D_1^{(1)}=0) = \mathbb{P}(D^{(1)}=0|Z^{(1)}=1),\\
p^c &= \mathbb{P}(D_1^{(1)}=1) - p^a = \mathbb{P}(D^{(1)}=1|Z^{(1)}=1) - p^a,\\
\mu_1^c &= \frac{(p^c+p^a)\mathbb{E}[Y_1^{(1)}|D_1^{(1)}=1] - p^a\mu_1^a}{p^c} = \frac{(p^c+p^a)\mathbb{E}[Y^{(1)}|D^{(1)}=1, Z^{(1)}=1] - p^a\mu_1^a}{p^c},\\
\mu_0^c &= \frac{(p^c+p^n)\mathbb{E}[Y_0^{(1)}|D_0^{(1)}=0] - p^n\mu_0^n}{p^c} = \frac{(p^c+p^n)\mathbb{E}[Y^{(1)}|D^{(1)}=0, Z^{(1)}=0] - p^n\mu_0^n}{p^c}.
\end{aligned}$$

Each of the first equality of the first four equations uses Assumption Y (ii) and the second equality of all equations uses Assumptions Y (i) and (iv). Thus, ATE for compliers is identified as

$$\text{ATE}(c) = \mu_1^c - \mu_0^c.$$

Second, μ_0^a is identified with Assumption Y (v) in addition to Y (i), (ii), (iii), and (iv) as follows:

$$\begin{aligned}
\mu_1^a - \mu_0^a &= \mathbb{E}[Y_1^{(1)}|U=a] - \mathbb{E}[Y_0^{(1)}|U=a]\\
&= \mathbb{E}[Y_1^{(1)}|U=a] - \mathbb{E}[Y_0^{(1)}|U=a] - \mathbb{E}[Y^{(*)}|U=a] + \mathbb{E}[Y^{(*)}|U=a]\\
&= \{\mathbb{E}[Y_1^{(1)}|U=a] - \mathbb{E}[Y^{(*)}|U=a]\} - \{\mathbb{E}[Y_0^{(1)}|U=a] - \mathbb{E}[Y^{(*)}|U=a]\}\\
&= \mathbb{E}[Y_1^{(1)} - Y^{(*)}|U=a] - \mathbb{E}[Y_0^{(1)} - Y^{(*)}|U=a]\\
&= \mathbb{E}[Y_1^{(1)} - Y^{(*)}|U=a] - \mathbb{E}[Y_0^{(1)} - Y^{(*)}|U=n]\\
&= \mathbb{E}[Y_1^{(1)}|U=a] - \mathbb{E}[Y^{(*)}|U=a] - \mathbb{E}[Y_0^{(1)}|U=n] + \mathbb{E}[Y^{(*)}|U=n]\\
&= \mu_1^a - \mathbb{E}[Y^{(*)}|U=a] - \mu_0^n + \mathbb{E}[Y^{(*)}|U=n]\\
&= \mu_1^a - \mu_{pre}^a - \mu_0^n + \mu_{pre}^n,
\end{aligned}$$

where

$$\begin{aligned}
\mu_{pre}^a &= \mathbb{E}[Y^{(0)}|D_1^{(1)}=1, D_0^{(1)}=1] = \mathbb{E}[Y^{(0)}|D_0^{(1)}=1] = \mathbb{E}[Y^{(0)}|Z^{(1)}=0, D^{(1)}=1],\\
\mu_{pre}^n &= \mathbb{E}[Y^{(0)}|D_1^{(1)}=0, D_0^{(1)}=0] = \mathbb{E}[Y^{(0)}|D_1^{(1)}=0] = \mathbb{E}[Y^{(0)}|Z^{(1)}=1, D^{(1)}=0].
\end{aligned}$$

In the equation of $\mu_1^a - \mu_0^a$, the second equality uses Assumption Y (iii) and the fifth equality uses Assumption Y (v). In the equations of μ_{pre}^a and μ_{pre}^n, each of the first equality uses Assumption Y (ii) and the second equality uses Assumptions Y (i) and (iv). Also, ATT is identified as

$$\text{ATT} = \frac{p^c(\mu_1^c - \mu_0^c) - p^a(\mu_1^a - \mu_0^a)}{p^c + p^a}.$$

Finally μ_1^n is identified with Assumption Y (vi) as

$$\mu_1^n = \mu_0^n + \mu_1^a - \mu_0^a$$

Therefore, ATE is identified as

$$\text{ATE} = p^c(\mu_1^c - \mu_0^c) + p^a(\mu_1^a - \mu_0^a) + p^n(\mu_1^n - \mu_0^n) + p^d(\mu_1^d - \mu_0^d).$$

A.2. **Proof of Theorem 4.** First, note that $\mu_1^c, \mu_1^a, \mu_0^n, \mu_0^d, p^c, p^a, p^n, p^d$ are identified under Assumptions 1, 2 (i) and 2 (ii) as

$$\begin{aligned}
\mu_1^c &= \mathbb{E}[Y_1^{(1)}|D_1^{(1)}=1, D_0^{(1)}=0] = \mathbb{E}[Y_1^{(1)}|D_1^{(1)}=1, D^{(0)}=0] = \mathbb{E}[Y^{(1)}|Z^{(1)}=1, D^{(1)}=1, D^{(0)}=0],\\
\mu_1^a &= \mathbb{E}[Y_1^{(1)}|D_1^{(1)}=1, D_0^{(1)}=1] = \mathbb{E}[Y_1^{(1)}|D_1^{(1)}=1, D^{(0)}=1] = \mathbb{E}[Y^{(1)}|Z^{(1)}=1, D^{(1)}=1, D^{(0)}=1],\\
\mu_0^n &= \mathbb{E}[Y_0^{(1)}|D_1^{(1)}=0, D_0^{(1)}=0] = \mathbb{E}[Y_0^{(1)}|D_1^{(1)}=0, D^{(0)}=0] = \mathbb{E}[Y^{(1)}|Z^{(1)}=1, D^{(1)}=0, D^{(0)}=0],\\
\mu_0^d &= \mathbb{E}[Y_0^{(1)}|D_1^{(1)}=0, D_0^{(1)}=1] = \mathbb{E}[Y_0^{(1)}|D_1^{(1)}=0, D^{(0)}=1] = \mathbb{E}[Y^{(1)}|Z^{(1)}=1, D^{(1)}=0, D^{(0)}=1],\\
p^c &= \mathbb{P}(D_1^{(1)}=1, D_0^{(1)}=0) = \mathbb{P}(D_1^{(1)}=1, D^{(0)}=0) = \mathbb{P}(D^{(1)}=1, D^{(0)}=0|Z^{(1)}=1),\\
p^a &= \mathbb{P}(D_1^{(1)}=1, D_0^{(1)}=1) = \mathbb{P}(D_1^{(1)}=1, D^{(0)}=1) = \mathbb{P}(D^{(1)}=1, D^{(0)}=1|Z^{(1)}=1),\\
p^n &= \mathbb{P}(D_1^{(1)}=0, D_0^{(1)}=0) = \mathbb{P}(D_1^{(1)}=0, D^{(0)}=0) = \mathbb{P}(D^{(1)}=0, D^{(0)}=0|Z^{(1)}=1),\\
p^d &= \mathbb{P}(D_1^{(1)}=0, D_0^{(1)}=1) = \mathbb{P}(D_1^{(1)}=0, D^{(0)}=1) = \mathbb{P}(D^{(1)}=0, D^{(0)}=1|Z^{(1)}=1).
\end{aligned}$$

Each of the first equality uses Assumption 2 (ii) and the second equality uses Assumptions 1 and 2 (i). Second, μ_1^d, μ_0^c are also identified with Assumption 1, 2 (i) and (ii) as

$$\begin{aligned}
\mu_1^d &= \frac{(p^a+p^d)\mathbb{E}[Y_1^{(1)}|D_0^{(1)}=1] - p^a\mu_1^a}{p^d}\\
&= \frac{(p^a+p^d)\mathbb{E}[Y^{(1)}|Z^{(1)}=0, D^{(1)}=1] - p^a\mu_1^a}{p^d}\\
&= \frac{(p^a+p^d)\mathbb{E}[Y^{(1)}|Z^{(1)}=0, D^{(1)}=1, D^{(0)}=1] - p^a\mu_1^a}{p^d},\\
\mu_0^c &= \frac{(p^c+p^n)\mathbb{E}[Y_0^{(1)}|D_0^{(1)}=0] - p^n\mu_0^n}{p^c}\\
&= \frac{(p^c+p^n)\mathbb{E}[Y^{(1)}|Z^{(1)}=0, D^{(1)}=0] - p^n\mu_0^n}{p^c}\\
&= \frac{(p^c+p^n)\mathbb{E}[Y^{(1)}|Z^{(1)}=0, D^{(1)}=0, D^{(0)}=0] - p^n\mu_0^n}{p^c}.
\end{aligned}$$

Each of the second equality uses Assumptions 1 and 2 (i) and the second equality uses Assumption 2 (ii). ATE for compliers is identified as

$$\text{ATE}(c) = \mu_1^c - \mu_0^c.$$

Second, μ_0^a is identified with Assumption 3 (i) in addition to 1, 2 (i) and (ii) as

$$\begin{aligned}
\mu_1^a - \mu_0^a &= \mathbb{E}[Y_1^{(1)}|U=a] - \mathbb{E}[Y_0^{(1)}|U=a] \\
&= \mathbb{E}[Y_1^{(1)}|U=a] - \mathbb{E}[Y_0^{(1)}|U=a] - \mathbb{E}[Y^{(*)}|U=a] + \mathbb{E}[Y^{(*)}|U=a] \\
&= \{\mathbb{E}[Y_1^{(1)}|U=a] - \mathbb{E}[Y^{(*)}|U=a]\} - \{\mathbb{E}[Y_0^{(1)}|U=a] - \mathbb{E}[Y^{(*)}|U=a]\} \\
&= \mathbb{E}[Y_1^{(1)} - Y^{(*)}|U=a] - \mathbb{E}[Y_0^{(1)} - Y^{(*)}|U=a] \\
&= \mathbb{E}[Y_1^{(1)} - Y^{(*)}|U=a] - \mathbb{E}[Y_0^{(1)} - Y^{(*)}|U=n] \\
&= \mathbb{E}[Y_1^{(1)}|U=a] - \mathbb{E}[Y^{(*)}|U=a] - \mathbb{E}[Y_0^{(1)}|U=n] + \mathbb{E}[Y^{(*)}|U=n] \\
&= \mu_1^a - \mathbb{E}[Y^{(*)}|U=a] - \mu_0^n + \mathbb{E}[Y^{(*)}|U=n] \\
&= \mu_1^a - \mu_{pre}^a - \mu_0^n + \mu_{pre}^n,
\end{aligned}$$

where

$$\begin{aligned}
\mu_{pre}^a &= \mathbb{E}[Y^{(*)}|D_1^{(1)}=1, D_0^{(1)}=1] = \mathbb{E}[Y^{(*)}|D_1^{(1)}=1, D^{(0)}=1] = \mathbb{E}[Y^{(*)}|Z^{(1)}=1, D^{(1)}=1, D^{(0)}=1], \\
\mu_{pre}^n &= \mathbb{E}[Y^{(*)}|D_1^{(1)}=0, D_0^{(1)}=0] = \mathbb{E}[Y^{(*)}|D_1^{(1)}=0, D^{(0)}=0] = \mathbb{E}[Y^{(*)}|Z^{(1)}=1, D^{(1)}=0, D^{(0)}=0].
\end{aligned}$$

In the equation of $\mu_1^a - \mu_0^a$, the second equality uses Assumption B (iii) and the fifth equality uses Assumption 3 (i). In the equations of μ_{pre}^a and μ_{pre}^n, each of the first equality uses Assumption 2 (ii) and the second equality uses Assumptions 1 and 2 (i). ATT is identified as

$$\text{ATT} = \frac{p^c(\mu_1^c - \mu_0^c) - p^a(\mu_1^a - \mu_0^a)}{p^c + p^a}.$$

Finally μ_1^n is identified with Assumption 3 (ii) as

$$\mu_1^n = \mu_0^n + \mu_1^a - \mu_0^a.$$

Therefore, ATE is identified as

$$\text{ATE} = p^c(\mu_1^c - \mu_0^c) + p^a(\mu_1^a - \mu_0^a) + p^n(\mu_1^n - \mu_0^n) + p^d(\mu_1^d - \mu_0^d).$$

A.3. **Proof of Theorem 6.** We provide a proof of identification of μ_0^a, μ_1^n, ATT, and ATE. Other parameters (i.e., $\mu_1^c, \mu_0^c, \mu_1^a, \mu_0^n, \mu_1^d, \mu_0^d, p^c, p^a, p^n, p^d$, and ATE($c$)) are identified in the same way as the proof of Theorem 4.

Proof under Assumptions 3a (i) and 3 (ii). Under Assumption 3a (i) in addition to 1, 2a (i), and 2a (ii), μ_0^a and ATT are identified as

$$\begin{aligned}
\mu_1^a - \mu_0^a &= \mathbb{E}[Y_1^{(1)}|U=a] - \mathbb{E}[Y_0^{(1)}|U=a] + \mathbb{E}[Y_1^{(0)}|U=a] - \mathbb{E}[Y_1^{(0)}|U=a] \\
&= \mathbb{E}[Y_1^{(1)} - Y_1^{(0)}|U=a] - \mathbb{E}[Y_0^{(1)} - Y_1^{(0)}|U=a] \\
&= \mathbb{E}[Y_1^{(1)} - Y_1^{(0)}|U=a] - \mathbb{E}[Y_0^{(1)} - Y_1^{(0)}|U=d] \\
&= \mu_1^a - \mathbb{E}[Y_1^{(0)}|U=a] - \mu_0^d + \mathbb{E}[Y_1^{(0)}|U=d],
\end{aligned}$$

where

$$\begin{aligned}\mathbb{E}[Y_1^{(0)}|U=a] &= \mathbb{E}[Y^{(0)}|Z^{(1)}=1, D^{(1)}=1, D^{(0)}=1],\\ \mathbb{E}[Y_1^{(0)}|U=d] &= \mathbb{E}[Y^{(0)}|Z^{(1)}=1, D^{(1)}=0, D^{(0)}=1].\end{aligned}$$

In the equation of $\mu_1^a - \mu_0^a$, the third equality uses assumption 3a(i). Each equality in the equations of $\mathbb{E}[Y_1^{(0)}|U=a]$ and $\mathbb{E}[Y_1^{(0)}|U=d]$ uses Assumptions 1, 2a (i), and 2a (ii). ATT is identified as

$$\text{ATT} = \frac{p^c(\mu_1^c - \mu_0^c) - p^a(\mu_1^a - \mu_0^a)}{p^c + p^a}.$$

Next, μ_1^n and ATE are identified under Assumption 3 (ii) as

$$\mu_1^n = \mu_0^n + \mu_1^a - \mu_0^a.$$

Therefore, ATE is identified as

$$\text{ATE} = p^c(\mu_1^c - \mu_0^c) + p^a(\mu_1^a - \mu_0^a) + p^n(\mu_1^n - \mu_0^n) + p^d(\mu_1^d - \mu_0^d).$$

Proof under Assumptions 3a (ii) and 3 (ii). Under Assumption 3a (ii) instead of 3a (i), μ_0^a, ATT, μ_1^n, and ATE are also identified as

$$\begin{aligned}\mu_1^n - \mu_0^n &= \mathbb{E}[Y_1^{(1)}|U=n] - \mathbb{E}[Y_0^{(1)}|U=n] + \mathbb{E}[Y_0^{(0)}|U=n] - \mathbb{E}[Y_0^{(0)}|U=n]\\ &= \mathbb{E}[Y_1^{(1)} - Y_0^{(0)}|U=n] - \mathbb{E}[Y_0^{(1)} - Y_0^{(0)}|U=n]\\ &= \mathbb{E}[Y_1^{(1)} - Y_0^{(0)}|U=c] - \mathbb{E}[Y_0^{(1)} - Y_0^{(0)}|U=n]\\ &= \mu_1^c - \mathbb{E}[Y_0^{(0)}|U=c] - \mu_0^n + \mathbb{E}[Y_0^{(0)}|U=n],\end{aligned}$$

where

$$\begin{aligned}\mathbb{E}[Y_0^{(0)}|U=c] &= \mathbb{E}[Y^{(0)}|Z^{(1)}=1, D^{(1)}=1, D^{(0)}=0],\\ \mathbb{E}[Y_0^{(0)}|U=n] &= \mathbb{E}[Y^{(0)}|Z^{(1)}=1, D^{(1)}=0, D^{(0)}=0].\end{aligned}$$

In the equation of $\mu_1^n - \mu_0^n$, the third equality uses Assumption 3a (ii). Each equality in the equations of $\mathbb{E}[Y_0^{(0)}|U=c]$ and $\mathbb{E}[Y_0^{(0)}|U=n]$ uses Assumptions 1, 2a (i), and 2a (ii). ATT is not yet identified. Next, μ_0^a is identified under Assumption 3 (ii) as

$$\mu_0^a = \mu_1^a - \mu_1^n + \mu_0^n.$$

Then, ATT and ATE are identified as

$$\begin{aligned}\text{ATT} &= \frac{p^c(\mu_1^c - \mu_0^c) - p^a(\mu_1^a - \mu_0^a)}{p^c + p^a},\\ \text{ATE} &= p^c(\mu_1^c - \mu_0^c) + p^a(\mu_1^a - \mu_0^a) + p^n(\mu_1^n - \mu_0^n) + p^d(\mu_1^d - \mu_0^d).\end{aligned}$$

Proof under Assumptions 3a (i) and 3a (ii). As mentioned above, under Assumptions 1, 2 (i), 2 (ii), and 3a (i), μ_0^a and ATT are identified as

$$\begin{aligned}\mu_0^a &= \mu_1^a - \mathbb{E}[Y_1^{(0)}|U=a] - \mu_0^d + \mathbb{E}[Y_1^{(0)}|U=d],\\ \text{ATT} &= \frac{p^c(\mu_1^c - \mu_0^c) - p^a(\mu_1^a - \mu_0^a)}{p^c + p^a}.\end{aligned}$$

Next, μ_1^n and ATE are identified under Assumption 3a (ii) as

$$\begin{aligned}\mu_1^n &= \mu_0^n + \mu_1^a - \mu_0^a, \\ \text{ATE} &= p^c(\mu_1^c - \mu_0^c) + p^a(\mu_1^a - \mu_0^a) + p^n(\mu_1^n - \mu_0^n) + p^d(\mu_1^d - \mu_0^d).\end{aligned}$$

Appendix B. Simulation

B.1. **Data generating process.** The data generation process in this simulation study is as follows. For unit i, covariates are generated as $X_{i1}, X_{i2} \sim_{\text{iid}} N(1, 0.5)$, $X_{i3}, X_{i4} \sim_{\text{iid}} N(-1, 0.5)$, $W_{i1}, W_{i2}, W_{i3} \sim_{\text{iid}} N(0, 0.3)$, and $V_{i1}, V_{i2}, V_{i3}, V_{i4} \sim_{\text{iid}} N(0, 0.3)$. Add intercepts and put them together into vectors $\mathbf{X}_i = (1, X_{i1}, \ldots, X_{i4})'$, $\mathbf{W}_i = (W_{i1}, \ldots, W_{i4})'$, and $\mathbf{V}_i = (V_{i1}, \ldots, V_{i4})'$. In the setup with monotonicity, the principal strata are generated by the logistic model:

$$\text{logit}(\mathbb{P}(U_i = u|\mathbf{W}_i)) = \frac{\exp(\phi_u'\mathbf{W}_i)}{\sum_v \exp(\phi_v'\mathbf{W}_i)},$$

where $u \in \{c, a, n\}$, $\phi_c = (0.2, 0.1, 0.1, -0.1)'$, $\phi_a = (0.15, -0.2, 0.2, -0.1)'$, and $\phi_n = (0.15, 0.2, -0.2, -0.1)'$. In the setup without monotonicity, $D_1^{(0)}$ and $D_0^{(0)}$ are generated by the following logistic models:

$$\begin{aligned}\text{logit}(\mathbb{P}(D_{i1}^{(0)} = 1|W_{i1}, W_{i2})) &= \frac{\exp(\zeta_1'(1, W_{i1}, W_{i2})')}{1 + \exp(\zeta_1'(1, W_{i1}, W_{i2})')}, \\ \text{logit}(\mathbb{P}(D_{i0}^{(0)} = 1|W_{i1}, W_{i3})) &= \frac{\exp(\zeta_0'(1, W_{i1}, W_{i3}))}{1 + \exp(\zeta_0'(1, W_{i1}, W_{i3}))},\end{aligned}$$

where $\zeta_1 = (0.2, 0.3, -0.1)'$ and $\zeta_0 = (-0.2, 0.3, -0.1)'$. In the setup without monotonicity, the treatment status before and after assignment is allowed to be different so $D_{i1}^{(1)}$ and $D_{i0}^{(1)}$ are generated as

$$D_{i1}^{(1)} = D_{i1}^{(0)}, \qquad D_{i0}^{(1)} = D_{i0}^{(0)}.$$

In the all setups, principal stratum is generated as

$$U_i = \begin{cases} a & \text{if} \quad (D_{i1}^{(1)}, D_{i0}^{(1)}) = (1, 1) \\ c & \text{if} \quad (D_{i1}^{(1)}, D_{i0}^{(1)}) = (1, 0) \\ d & \text{if} \quad (D_{i1}^{(1)}, D_{i0}^{(1)}) = (0, 1) \\ n & \text{if} \quad (D_{i1}^{(1)}, D_{i0}^{(1)}) = (0, 0) \end{cases}.$$

The outcomes following the normal distribution are generated as

$$Y_{iD_i^{(1)}}^{(1)}|\mathbf{X}_i, \mathbf{V}_i, W_{i2}, D_i^{(1)} \sim N(\alpha + \gamma_{X,U_i}'\mathbf{X}_i + \gamma_V'\mathbf{V}_i + \gamma_W W_{i2} + \beta_{U_i}, \sigma^2),$$

where $\alpha = 2$, $\gamma_{X,a} = (1, 1, -1, 1)'$, $\gamma_{X,c} = (3, 1, -1, 1)'$, $\gamma_{X,n} = (-2, 1, -1, 1)'$, $\gamma_{X,d} = (-1, 1, -1, 1)'$, $\gamma_V = (2, -1, 2, -2)'$, $\gamma_W = 1$, $\beta_a = 1$, $\beta_c = 2$, $\beta_n = 1$, $\beta_d = 3$, and $\sigma^2 = 0.3$. The outcomes following the Bernoulli distribution are generated as

$$Y_{iD_i^{(1)}}^{(1)}|\mathbf{X}_i, \mathbf{V}_i, W_{i2}, D_i^{(1)} \sim \text{Ber}\left(\frac{\exp(\alpha + \gamma_{X,U_i}'\mathbf{X}_i + \gamma_V'\mathbf{V}_i + \gamma_W W_{i2} + \beta_{U_i})}{1 + \exp(\alpha + \gamma_{X,U_i}'\mathbf{X}_i + \gamma_V'\mathbf{V}_i + \gamma_W W_{i2} + \beta_{U_i})}\right),$$

where $\gamma_{X,a} = (1, -1, 1, 0.5)'$, $\gamma_{X,c} = (3, -1, 1.5, 1)'$, $\gamma_{X,n} = (-2, -1, 0.5, 0.5)'$, $\gamma_{X,d} = (-1, -1, 1, 1)'$, and the values of the other parameters are the same as those used to generate from the normal

distribution. The pre-assignment outcome $Y^{(*)}$ or Y^{pre} is generated as

$$Y_i^{(*)} = Y_{0i}^{(1)} - \Delta_i, \qquad \Delta_i \sim N(\delta_{U_i}, 0.5),$$

for the normal distribution outcome with $\delta_c = 3$, $\delta_a = 1$, $\delta_n = 1$, and $\delta_d = 2$, and also

$$Y_i^{(*)} = Y_{0i}^{(1)} - \Delta_i, \qquad \Delta_i \sim \text{Ber}(\delta_{U_i}),$$

for the Bernoulli distribution outcome with $\delta_c = 0.3$, $\delta_a = 0.2$, $\delta_n = 0.2$, and $\delta_d = 0.1$.

In the case of the random assignment, the assignment variable $Z^{(1)}$ is randomly generated so that half of the values are 1 and the other half are 0. In the cause of the conditional ignorability, $Z^{(1)}$ is generated based on the following model:

$$\text{logit}(\mathbb{P}(Z_i^{(1)} = 1|\mathbf{X}_i)) = \frac{\exp(\kappa'\mathbf{X}_i)}{1 + \exp(\kappa'\mathbf{X}_i)},$$

where $\kappa = (0.5, 1, 0.5, 0.5, 1)'$. In the case of the random assignment, none of the covariates are observed, and in the case of the conditional ignorability, only $\mathbf{X}$ of the covariates is observed.

B.2. **Simulation result.** Using numerical simulations, we evaluate the properties of estimators based on finite samples in two setups: monotonicity and stable. For each setup, we evaluated the following eight scenarios: two sample sizes (1000 and 10000), two distributions of the outcome variable (normal and Bernoulli), and two assignment assumptions (random assignment and conditional ignorability). The data generation process in the last subsection is based on the assumption that all assumptions hold. In the conditional ignorability scenario, all models are correctly specified. We evaluate the average estimates, the average biases and coverage rates over 1000 repeated drawings from the data generating process. To calculate standard errors, we conduct 200 bootstraps. Table 9 shows the results for the normal distribution scenarios, and Table 10 shows the results for the Bernoulli distribution scenarios. In both tables, the results of the LATE using the typical wald type estimator are shown in the first row for comparison with the proposed method (only for the random assignment scenario), and the results of the estimation using the proposed method are shown in the second row and subsequent rows. In our method, the parameters μ_1^u, μ_0^u, and p^u are estimated for each principal strata, so the tables show the results of the estimation of each parameter. Since there is no difference in the evaluation of the results between the normal distribution and the Bernoulli distribution scenarios, the results of the normal distribution will be primarily evaluated below.

In all scenarios, for each estimand and each parameter, the proposed method provides estimates that are sufficiently close to the true values. The coverage rates are around 0.95. It is important for decision-makers to be able to obtain appropriate estimation results not only for ATE(c), ATT and ATE, but also for μ_1^u, μ_0^u, and p^u for each principal strata, since parameters are useful for determining the content and target of interventions. The standard deviations for the $n = 1000$ scenarios was about three times larger than for the $n = 10000$ scenarios, but even with $n = 1000$ it is not enough to have a significant impact on the interpretation of the causal effect. (For example, in the scenarios of normal distribution and random assignment in the monotonicity setup, ATE is 1.34, with a standard deviation of 0.21 for $n = 1000$ and a

standard deviation of 0.07 for $n = 10000$). In addition, there is no significant difference between the scenarios of random assignment and the scenarios of conditional ignorability (For example, in the scenarios of normal distribution and $n = 1000$ in the stable setup, ATE is 1.70, with a standard deviation of 0.20 for the scenario of random assignment and a standard deviation of 0.23 for the scenario of conditional ignorability). In the monotonicity setup, LATE and ATE(c) are the same. On the other hand, in the stable setting, which do not assume monotonicity, LATE will naturally be a result that deviates from the true value (For example, in the scenarios of normal distribution and $n = 1000$ (the true value is 2.00), the estimated value is -0.52 and the standard deviation is 58.36 for the stable setup). The proposed method can obtain reasonable estimation results even in situations where the bias and variance of LATE are large.

	Random assignment								Conditional ignorability							
	$n = 1000$				$n = 10000$				$n = 1000$				$n = 10000$			
	θ	$\hat{\theta}$	sd	cover	θ	$\hat{\theta}$	sd	cover	θ	$\hat{\theta}$	sd	cover	θ	$\hat{\theta}$	sd	cover
With monotonicity																
LATE	2.00	1.99	0.59	0.94	2.00	2.00	0.18	0.95	2.00	-	-	-	2.00	-	-	-
ATE(c)	2.00	1.99	0.59	0.94	2.00	2.00	0.18	0.95	2.00	1.98	0.70	0.95	2.00	2.00	0.21	0.94
ATT	1.51	1.51	0.30	0.95	1.51	1.51	0.10	0.95	1.51	1.51	0.35	0.94	1.51	1.51	0.11	0.94
ATE	1.34	1.34	0.21	0.95	1.35	1.34	0.07	0.96	1.34	1.35	0.24	0.94	1.35	1.34	0.08	0.94
$\mu_1^c - \mu_0^c$	2.00	1.99	0.59	0.94	2.00	2.00	0.18	0.95	2.00	1.98	0.70	0.95	2.00	2.00	0.21	0.94
$\mu_1^a - \mu_0^a$	1.00	1.00	0.06	0.94	1.00	1.00	0.02	0.94	1.00	1.01	0.08	0.93	1.00	1.00	0.02	0.94
$\mu_1^n - \mu_0^n$	1.00	1.00	0.06	0.94	1.00	1.00	0.02	0.94	1.00	1.01	0.08	0.93	1.00	1.00	0.02	0.94
μ_1^c	8.02	8.03	0.41	0.94	8.00	8.02	0.13	0.93	8.02	8.07	0.47	0.94	8.00	8.02	0.14	0.93
μ_1^a	5.01	5.04	0.17	0.93	5.02	5.04	0.05	0.92	5.01	5.08	0.24	0.94	5.02	5.05	0.07	0.92
μ_1^n	1.95	1.95	0.19	0.95	1.94	1.94	0.06	0.95	1.95	1.95	0.17	0.93	1.94	1.94	0.05	0.95
μ_0^c	6.02	6.04	0.56	0.96	6.00	6.02	0.17	0.95	6.02	6.09	0.70	0.95	6.00	6.02	0.21	0.95
μ_0^a	4.01	4.04	0.17	0.93	4.02	4.04	0.06	0.93	4.01	4.07	0.22	0.93	4.02	4.04	0.07	0.92
μ_0^n	0.95	0.95	0.18	0.94	0.94	0.94	0.06	0.96	0.95	0.94	0.15	0.94	0.94	0.94	0.05	0.95
p^c	0.34	0.34	0.03	0.94	0.35	0.34	0.01	0.94	0.34	0.34	0.03	0.95	0.35	0.34	0.01	0.93
p^a	0.33	0.33	0.02	0.94	0.33	0.33	0.01	0.95	0.33	0.33	0.03	0.93	0.33	0.33	0.01	0.94
p^n	0.33	0.33	0.02	0.94	0.33	0.33	0.01	0.94	0.33	0.33	0.02	0.95	0.33	0.33	0.01	0.95
Without monotonicity																
LATE	2.00	-0.52	58.36	0.86	2.00	-0.06	0.70	0.09	2.00	-	-	-	2.00	-	-	-
ATE(c)	2.00	1.98	0.55	0.95	2.00	2.00	0.17	0.95	2.00	1.96	0.64	0.94	2.00	1.99	0.20	0.95
ATT	1.55	1.54	0.31	0.95	1.55	1.55	0.10	0.95	1.55	1.53	0.36	0.94	1.55	1.54	0.11	0.95
ATE	1.70	1.70	0.20	0.95	1.71	1.71	0.06	0.95	1.70	1.70	0.23	0.94	1.71	1.70	0.07	0.95
$\mu_1^c - \mu_0^c$	2.00	1.98	0.55	0.95	2.00	2.00	0.17	0.95	2.00	1.96	0.64	0.94	2.00	1.99	0.20	0.95
$\mu_1^a - \mu_0^a$	1.00	1.00	0.06	0.94	1.00	1.00	0.02	0.94	1.00	1.00	0.07	0.95	1.00	1.00	0.02	0.94
$\mu_1^n - \mu_0^n$	1.00	1.00	0.06	0.94	1.00	1.00	0.02	0.94	1.00	1.00	0.07	0.95	1.00	1.00	0.02	0.94
$\mu_1^d - \mu_0^d$	3.00	2.98	0.46	0.93	3.00	3.00	0.14	0.94	3.00	3.05	0.55	0.93	3.00	3.00	0.16	0.94
μ_1^c	7.98	7.98	0.21	0.94	8.00	7.99	0.07	0.94	7.98	8.01	0.19	0.94	8.00	7.99	0.06	0.95
μ_1^a	4.99	4.99	0.19	0.94	4.99	4.99	0.06	0.96	4.99	5.01	0.19	0.94	4.99	4.99	0.06	0.94
μ_1^n	2.01	2.01	0.22	0.95	2.01	2.01	0.07	0.94	2.01	2.02	0.18	0.95	2.01	2.01	0.06	0.95
μ_1^d	5.04	5.02	0.41	0.95	4.99	5.02	0.13	0.94	5.04	5.07	0.52	0.93	4.99	5.02	0.15	0.94
μ_0^c	5.98	5.99	0.51	0.95	6.00	5.99	0.16	0.95	5.98	6.05	0.62	0.95	6.00	6.00	0.19	0.95
μ_0^a	3.99	3.98	0.20	0.95	3.99	3.99	0.06	0.96	3.99	4.00	0.18	0.94	3.99	3.99	0.06	0.94
μ_0^n	1.01	1.01	0.21	0.95	1.01	1.01	0.07	0.94	1.01	1.02	0.17	0.95	1.01	1.01	0.05	0.94
μ_0^d	2.04	2.03	0.21	0.96	1.99	2.01	0.07	0.93	2.04	2.02	0.19	0.95	1.99	2.01	0.06	0.93
p^c	0.30	0.30	0.02	0.95	0.30	0.30	0.01	0.94	0.30	0.30	0.02	0.95	0.30	0.30	0.01	0.95
p^a	0.25	0.25	0.02	0.95	0.25	0.25	0.01	0.92	0.25	0.25	0.02	0.95	0.25	0.25	0.01	0.93
p^n	0.25	0.25	0.02	0.94	0.25	0.25	0.01	0.94	0.25	0.25	0.02	0.95	0.25	0.25	0.01	0.94
p^d	0.20	0.20	0.02	0.95	0.20	0.20	0.01	0.96	0.20	0.20	0.02	0.94	0.20	0.20	0.01	0.95

Note: The values in the θ column are the true values calculated from $100,000$ samples from the data generation process. Columns $\hat{\theta}$, sd, and cover contain the average estimates, the average biases and coverage rates over 1000 repeated drawings from the data generating process, respectively. ATE(c) is the same definition as $\mu_1^c - \mu_0^c$. LATE is calculated by $\frac{\mathbb{E}[Y^{(1)}|Z^{(1)}=1]-\mathbb{E}[Y^{(1)}|Z^{(1)}=0]}{\mathbb{E}[D^{(1)}|Z^{(1)}=1]-\mathbb{E}[D^{(1)}|Z^{(1)}=0]}$.

TABLE 9. Normal distribution case

	Random assignment								Conditional ignorability							
	$n = 1000$				$n = 10000$				$n = 1000$				$n = 10000$			
	θ	$\hat{\theta}$	sd	cover	θ	$\hat{\theta}$	sd	cover	θ	$\hat{\theta}$	sd	cover	θ	$\hat{\theta}$	sd	cover
With monotonicity																
LATE	0.18	0.18	0.09	0.94	0.18	0.18	0.03	0.96	0.18	-	-	-	0.18	-	-	-
ATE(c)	0.18	0.18	0.09	0.94	0.18	0.18	0.03	0.96	0.18	0.17	0.11	0.95	0.18	0.18	0.03	0.95
ATT	0.16	0.16	0.05	0.94	0.16	0.16	0.02	0.95	0.16	0.15	0.06	0.93	0.16	0.16	0.02	0.95
ATE	0.15	0.15	0.05	0.94	0.15	0.15	0.01	0.94	0.15	0.15	0.05	0.94	0.15	0.15	0.02	0.94
$\mu_1^c - \mu_0^c$	0.18	0.18	0.09	0.94	0.18	0.18	0.03	0.96	0.18	0.17	0.11	0.95	0.18	0.18	0.03	0.95
$\mu_1^a - \mu_0^a$	0.14	0.14	0.06	0.95	0.14	0.14	0.02	0.95	0.14	0.13	0.07	0.95	0.14	0.14	0.02	0.95
$\mu_1^n - \mu_0^n$	0.12	0.14	0.06	0.95	0.12	0.14	0.02	0.86	0.12	0.13	0.07	0.94	0.12	0.14	0.02	0.89
μ_1^c	0.87	0.87	0.06	0.95	0.87	0.87	0.02	0.94	0.87	0.87	0.07	0.94	0.87	0.87	0.02	0.94
μ_1^a	0.72	0.72	0.04	0.93	0.71	0.72	0.01	0.95	0.72	0.72	0.05	0.94	0.71	0.72	0.01	0.94
μ_1^n	0.35	0.37	0.07	0.95	0.35	0.37	0.02	0.88	0.35	0.37	0.08	0.94	0.35	0.37	0.02	0.89
μ_0^c	0.68	0.69	0.07	0.95	0.68	0.68	0.02	0.94	0.68	0.70	0.09	0.94	0.68	0.69	0.03	0.96
μ_0^a	0.58	0.58	0.06	0.94	0.58	0.58	0.02	0.95	0.58	0.59	0.07	0.93	0.58	0.58	0.02	0.94
μ_0^n	0.23	0.23	0.03	0.95	0.23	0.23	0.01	0.94	0.23	0.23	0.03	0.94	0.23	0.23	0.01	0.94
p^c	0.34	0.34	0.03	0.95	0.35	0.34	0.01	0.94	0.34	0.34	0.03	0.95	0.35	0.34	0.01	0.94
p^a	0.33	0.33	0.02	0.94	0.33	0.33	0.01	0.91	0.33	0.33	0.03	0.93	0.33	0.33	0.01	0.93
p^n	0.33	0.33	0.02	0.94	0.33	0.33	0.01	0.95	0.33	0.33	0.02	0.95	0.33	0.33	0.01	0.95
Without monotonicity																
LATE	0.18	-0.39	9.10	0.73	0.18	-0.28	0.12	0.00	0.18	-	-	-	0.18	-	-	-
ATE(c)	0.18	0.18	0.07	0.94	0.18	0.18	0.02	0.93	0.18	0.18	0.09	0.95	0.18	0.18	0.03	0.93
ATT	0.16	0.16	0.05	0.94	0.16	0.16	0.02	0.93	0.16	0.16	0.06	0.95	0.16	0.16	0.02	0.94
ATE	0.20	0.21	0.04	0.93	0.20	0.21	0.01	0.92	0.20	0.21	0.05	0.94	0.20	0.21	0.01	0.94
$\mu_1^c - \mu_0^c$	0.18	0.18	0.07	0.94	0.18	0.18	0.02	0.93	0.18	0.18	0.09	0.95	0.18	0.18	0.03	0.93
$\mu_1^a - \mu_0^a$	0.14	0.14	0.07	0.94	0.14	0.14	0.02	0.95	0.14	0.14	0.07	0.95	0.14	0.14	0.02	0.95
$\mu_1^n - \mu_0^n$	0.12	0.14	0.07	0.93	0.13	0.14	0.02	0.90	0.12	0.14	0.07	0.94	0.13	0.14	0.02	0.90
$\mu_1^d - \mu_0^d$	0.42	0.41	0.10	0.94	0.41	0.41	0.03	0.96	0.42	0.42	0.11	0.95	0.41	0.41	0.03	0.96
μ_1^c	0.86	0.87	0.03	0.93	0.86	0.87	0.01	0.93	0.86	0.87	0.03	0.94	0.86	0.87	0.01	0.92
μ_1^a	0.71	0.71	0.04	0.93	0.71	0.71	0.01	0.94	0.71	0.71	0.04	0.94	0.71	0.71	0.01	0.94
μ_1^n	0.37	0.37	0.08	0.93	0.36	0.38	0.03	0.90	0.37	0.38	0.08	0.94	0.36	0.38	0.02	0.89
μ_1^d	0.65	0.65	0.09	0.94	0.64	0.64	0.03	0.95	0.65	0.65	0.11	0.95	0.64	0.64	0.03	0.95
μ_0^c	0.68	0.68	0.07	0.94	0.68	0.68	0.02	0.94	0.68	0.69	0.08	0.94	0.68	0.68	0.03	0.96
μ_0^a	0.57	0.57	0.07	0.94	0.57	0.57	0.02	0.95	0.57	0.57	0.07	0.94	0.57	0.57	0.02	0.95
μ_0^n	0.25	0.24	0.04	0.94	0.24	0.24	0.01	0.92	0.25	0.24	0.04	0.94	0.24	0.24	0.01	0.92
μ_0^d	0.24	0.23	0.04	0.95	0.24	0.23	0.01	0.94	0.24	0.24	0.04	0.95	0.24	0.23	0.01	0.95
p^c	0.30	0.30	0.02	0.93	0.30	0.30	0.01	0.93	0.30	0.30	0.02	0.94	0.30	0.30	0.01	0.95
p^a	0.25	0.25	0.02	0.94	0.25	0.25	0.01	0.94	0.25	0.25	0.02	0.95	0.25	0.25	0.01	0.95
p^n	0.25	0.25	0.02	0.95	0.25	0.25	0.01	0.93	0.25	0.25	0.02	0.94	0.25	0.25	0.01	0.93
p^d	0.20	0.20	0.02	0.95	0.20	0.20	0.01	0.95	0.20	0.20	0.02	0.95	0.20	0.20	0.01	0.94

Note: The values in the θ column are the true values calculated from $100,000$ samples from the data generation process. Columns $\hat{\theta}$, sd, and cover contain the average estimates, the average biases and coverage rates over 1000 repeated drawings from the data generating process, respectively. ATE(c) is the same definition as $\mu_1^c - \mu_0^c$. LATE is calculated by $\frac{\mathbb{E}[Y^{(1)}|Z^{(1)}=1]-\mathbb{E}[Y^{(1)}|Z^{(1)}=0]}{\mathbb{E}[D^{(1)}|Z^{(1)}=1]-\mathbb{E}[D^{(1)}|Z^{(1)}=0]}$.

TABLE 10. Bernoulli distribution case

U	$Y^{(1)}$	$D^{(1)}$	$Z^{(1)}$	$Y^{(*)}$	$D^{(0)}$
c or a	$Y_1^{(1)}$	1	1	$Y^{(*)}$	0
c or a	$Y_1^{(1)}$	1	1	$Y^{(*)}$	1
n or d	$Y_0^{(1)}$	0	1	$Y^{(*)}$	0
n or d	$Y_0^{(1)}$	0	1	$Y^{(*)}$	1
a or d	$Y_1^{(1)}$	1	0	$Y^{(*)}$	0
a or d	$Y_1^{(1)}$	1	0	$Y^{(*)}$	1
c or n	$Y_0^{(1)}$	0	0	$Y^{(*)}$	0
c or n	$Y_0^{(1)}$	0	0	$Y^{(*)}$	1

TABLE 11. Unstable case with auxiliary data

APPENDIX C. EXTENSION: UNSTABLE TREATMENT STATUS

In this subsection, we relax Assumption 2 (ii) on Case I in Section 3 (i.e., $D_0^{(1)} = D^{(0)}$). Without this assumption, the relationships of the observables and principal strata variable are summarized in Table 11. We call this case as unstable case.

Instead of Assumption 2 (ii), we impose the following assumptions.

Assumption 4.

(i): *[Mean independence from pre-assignment status] It holds*

$$\mathbb{E}[Y_d^{(1)}|U=u, D_0^{(0)}=1] = \mathbb{E}[Y_d^{(1)}|U=u, D_0^{(0)}=0] = \mathbb{E}[Y_d^{(1)}|U=u],$$

for each $u \in \{c,a,n,d\}$ and $d \in \{0,1\}$ but if $u=a$ then $d=1$ and if $u=n$ then $d=0$. If one of the equals is true, the rest are also true.

(ii): *[Independence from opposite treatment status conditional on pre-assignment] It holds*

$$\mathbb{P}(D_z^{(1)}=1|D_0^{(0)}=d, D_{1-z}^{(1)}=1) = \mathbb{P}(D_z^{(1)}=1|D_0^{(0)}=d, D_{1-z}^{(1)}=0) = \mathbb{P}(D_z^{(1)}=1|D_0^{(0)}=d),$$

for each $z \in \{0,1\}$ and $d \in \{0,1\}$. One of the equals guarantees the rest.

(iii): *[Pre-assignment treatment status is relevant to main] It holds*

$$\mathbb{P}(D_z^{(1)}=1|D_0^{(0)}=1) \neq \mathbb{P}(D_z^{(1)}=1|D_0^{(0)}=0),$$

for each $z \in \{0,1\}$.

(iv): *[Mean independence for $Y^{(*)}$] It holds*

$$\mathbb{E}[Y^{(*)}|U=u, D_0^{(0)}=1] = \mathbb{E}[Y^{(*)}|U=u, D_0^{(0)}=0] = \mathbb{E}[Y^{(*)}|U=u],$$

for each $u \in \{c,a,n,d\}$. If one of the equals is true, the rest are also true.

Under these assumptions, we obtain the following identification results.

Theorem 8. *Consider the setup of this subsection.*

(i): *Under Assumptions 1, 2 (i)-(ii), and 4 (i)-(iii),* ATE(c) *is identified.*

(ii): *Under Assumptions 1, 2 (i)-(ii), 3 (i), and 4,* ATT *is identified.*

(iii): *Under Assumptions 1, 2 (i)-(ii), 3, and 4,* ATE *is identified.*

This theorem can be shown as follows. Using the notation $\rho_{(z,d,d')} = \mathbb{E}[D_z^{(1)} = d | D_0^{(0)} = d']$, μ_1^u's and μ_0^u's are identified as follows. Note that

$$\begin{aligned}
\mu_b^u &= \frac{\delta^{(1)}_{(1,b,b')}\rho_{(0,1-b',1-b')} - \delta^{(1)}_{(1,b,1-b')}\rho_{(0,1-b',b')}}{\rho_{(0,1-b',1-b')} - \rho_{(0,1-b',b')}}, \\
\mu_{b'}^u &= \frac{\delta^{(1)}_{(0,b',b')}\rho_{(1,1-b,1-b')} - \delta^{(1)}_{(0,b',1-b')}\rho_{(1,1-b,b')}}{\rho_{(1,1-b,1-b')} - \rho_{(1,1-b,b')}},
\end{aligned}$$

where

$$(b, b') = \begin{cases} (1,1) & \text{for} \quad u = a \\ (1,0) & \text{for} \quad u = c \\ (0,1) & \text{for} \quad u = d \\ (0,0) & \text{for} \quad u = n \end{cases}.$$

The outline for identification of μ_1^c is as follows. Note that

$$\begin{aligned}
\delta^{(1)}_{(1,1,0)} &= \mathbb{E}[Y_1^{(1)} | D_1^{(1)} = 1, D_0^{(0)} = 0] \\
&= \mathbb{E}[Y_1^{(1)} | U = c, D_0^{(0)} = 0]\mathbb{E}[D_0^{(1)} = 0 | D_1^{(1)} = 1, D_0^{(0)} = 0] \\
&\quad + \mathbb{E}[Y_1^{(1)} | U = a, D_0^{(0)} = 0]\mathbb{E}[D_0^{(1)} = 1 | D_1^{(1)} = 1, D_0^{(0)} = 0] \\
&= \mathbb{E}[Y_1^{(1)} | U = c]\mathbb{E}[D_0^{(1)} = 0 | D_0^{(0)} = 0] + \mathbb{E}[Y_1^{(1)} | U = a]\mathbb{E}[D_0^{(1)} = 1 | D_0^{(0)} = 0] \\
&= \mu_1^c \rho_{(0,0,0)} + \mu_1^a \rho_{(0,1,0)},
\end{aligned}$$

where the third equality follows from Assumptions 4 (i)-(ii). Similarly we have

$$\delta^{(1)}_{(1,1,1)} = \mu_1^a \rho_{(0,1,1)} + \mu_1^c \rho_{(0,0,1)}.$$

From these equations, eliminating the term with μ_1^a yields

$$\mu_1^c = \frac{\delta^{(1)}_{(1,1,0)}\rho^{(1)}_{(0,1,1)} - \delta^{(1)}_{(1,1,1)}\rho^{(1)}_{(0,1,0)}}{\rho^{(1)}_{(0,1,1)} - \rho^{(1)}_{(0,1,0)}}.$$

To avoid zero-division, Assumption 4 (iii) is imposed. Using Assumption 4 (ii), p^u's are identified as follows. Note that

$$p^u = \pi_{(1,b,b')}\rho_{(0,b',b')} + \pi_{(1,b,1-b')}\rho_{(0,b',1-b')},$$

where

$$(b, b') = \begin{cases} (1,1) & \text{for} \quad u = a \\ (1,0) & \text{for} \quad u = c \\ (0,1) & \text{for} \quad u = d \\ (0,0) & \text{for} \quad u = n \end{cases}.$$

Using Assumptions 4 (ii)-(iv), μ_*^a is identified as

$$\mu_*^a = \frac{\delta^{(*)}_{(1,1,1)}\rho_{(0,0,0)} - \delta^{(*)}_{(1,1,0)}\rho_{(0,0,1)}}{\rho_{(0,0,0)} - \rho_{(0,0,1)}} = \frac{\delta^{(*)}_{(0,1,1)}\rho_{(1,0,0)} - \delta^{(*)}_{(0,1,0)}\rho_{(1,0,1)}}{\rho_{(1,0,0)} - \rho_{(1,0,1)}}.$$

Using Assumptions 4 (ii)-(iv), μ_*^n is identified as

$$\mu_*^n = \frac{\delta^{(*)}_{(1,0,0)}\rho_{(0,1,1)} - \delta^{(*)}_{(1,0,1)}\rho_{(0,1,0)}}{\rho_{(0,1,1)} - \rho_{(0,1,0)}} = \frac{\delta^{(*)}_{(0,0,0)}\rho_{(1,1,1)} - \delta^{(*)}_{(0,0,1)}\rho_{(1,1,0)}}{\rho_{(1,1,1)} - \rho_{(1,1,0)}}.$$

Then μ_0^a and μ_1^n are identified in the same way as in Case I in Section 3.

For identification under the ignorability condition, we consider the following models for doubly robust estimator of $\rho_{(z,d,d')}$.

$$P_d^{(1)}(\mathbf{X};\zeta^{(1)}) = \mathbb{E}[D^{(1)} = d|\mathbf{X}], \qquad P_{d'}^{(0)}(\mathbf{X};\zeta^{(0)}) = \mathbb{E}[D^{(0)} = d'|\mathbf{X}],$$

for $z, d, d' \in \{0, 1\}$. Considering the setup of this section, we obtain

The other required terms are identified in the same way as in Case I in Section 3.

References

Angrist, J. and Fernandez-Val, I. (2010). Extrapolate-ing: External validity and overidentification in the late framework. Technical report, National Bureau of Economic Research.

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.

Aronow, P. M. and Carnegie, A. (2013). Beyond late: Estimation of the average treatment effect with an instrumental variable. *Political Analysis*, 21(4):492–506.

Athey, S. and Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic perspectives*, 31(2):3–32.

Balke, A. and Pearl, J. (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American statistical Association*, 92(439):1171–1176.

Beam, E. A. (2016). Do job fairs matter? experimental evidence on the impact of job-fair attendance. *Journal of Development Economics*, 120:32–40.

Brinch, C. N., Mogstad, M., and Wiswall, M. (2017). Beyond late with a discrete instrument. *Journal of Political Economy*, 125(4):985–1039.

Dahl, C. M., Huber, M., and Mellace, G. (2023). It is never too late: a new look at local average treatment effects with or without defiers. *The Econometrics Journal*, 26(3):378–404.

De Chaisemartin, C. (2017). Tolerating defiance? local average treatment effects without monotonicity. *Quantitative Economics*, 8(2):367–396.

Deaton, A. S. (2009). Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development. Technical report, National bureau of economic research.

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

Freedman, D. A. (2006). Statistical models for causation: what inferential leverage do they provide? *Evaluation review*, 30(6):691–713.

Fricke, H., Frölich, M., Huber, M., and Lechner, M. (2020). Endogeneity and non-response bias in treatment evaluation–nonparametric identification of causal effects by instruments. *Journal of Applied Econometrics*, 35(5):481–504.

Gabriel, E. E. and Follmann, D. (2016). Augmented trial designs for evaluation of principal surrogates. *Biostatistics*, 17(3):453–467.

Gerber, A. S., Karlan, D., and Bergan, D. (2009). Does the media matter? a field experiment measuring the effect of newspapers on voting behavior and political opinions. *American Economic Journal: Applied Economics*, 1(2):35–52.

Heckman, J. J. and Urzua, S. (2010). Comparing iv with structural models: What simple iv can and cannot identify. *Journal of Econometrics*, 156(1):27–37.

Heckman, J. J. and Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation 1. *Econometrica*, 73(3):669–738.

Heckman, J. J. and Vytlacil, E. J. (1999). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the national Academy of Sciences*, 96(8):4730–4734.

Heckman, J. J. and Vytlacil, E. J. (2007). Econometric evaluation of social programs, part ii: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. *Handbook of econometrics*, 6:4875–5143.

Imai, K., Tingley, D., and Yamamoto, T. (2013). Experimental designs for identifying causal mechanisms. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 176(1):5–51.

Imbens, G. (2014). Instrumental variables: An econometrician's perspective. Technical report, National Bureau of Economic Research.

Imbens, G. W. (2024). Causal inference in the social sciences. *Annual Review of Statistics and Its Application*, 11.

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475.

Imbens, G. W. and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Jiang, Z. and Ding, P. (2021). Identification of causal effects within principal strata using auxiliary variables. *Statistical Science*, 36(4):493–508.

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.

Kennedy, E. H., Balakrishnan, S., and G'sell, M. (2020). Sharp instruments for classifying compliers and generalizing causal effects.

Klein, T. J. (2010). Heterogeneous treatment effects: Instrumental variables without monotonicity? *Journal of Econometrics*, 155(2):99–116.

Marbach, M. and Hangartner, D. (2020). Profiling compliers and noncompliers for instrumental-variable analysis. *Political Analysis*, 28(3):435–444.

Mattei, A. and Mealli, F. (2011). Augmented designs to assess principal strata direct effects. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73(5):729–752.

Mealli, F. and Pacini, B. (2013). Using secondary outcomes to sharpen inference in randomized experiments with noncompliance. *Journal of the American Statistical Association*, 108(503):1120–1131.

Pearl, J. (2009). *Causality*. Cambridge university press.

Robins, J. M. and Greenland, S. (1996). Identification of causal effects using instrumental variables: comment. *Journal of the American Statistical Association*, 91(434):456–458.

Small, D. S., Tan, Z., Ramsahai, R. R., Lorch, S. A., and Brookhart, M. A. (2017). Instrumental variable estimation with a stochastic monotonicity assumption.

Swanson, S. A. and Hernán, M. A. (2014). Think globally, act globally: an epidemiologist's perspective on instrumental variable estimation. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 29(3):371.

Thornton, R. L. (2008). The demand for, and impact of, learning hiv status. *American Economic Review*, 98(5):1829–1863.

van't Hoff, N., Lewbel, A., and Mellace, G. (2023). *Limited Monotonicity and the Combined Compliers LATE*. University of Southern Denmark, Faculty of Business and Social Sciences ….

Wang, L. and Tchetgen Tchetgen, E. (2018). Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 80(3):531–550.

Yang, F. and Small, D. S. (2016). Using post-outcome measurement information in censoring-by-death problems. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 78(1):299–318.