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

2025 年 9 月 4 日
DP2025-021
https://ies.keio.ac.jp/publications/26824/

Institute for Economic Studies, Keio University
2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
ies-office@adst.keio.ac.jp
4 September, 2025

※ DP2024-022 の改訂版

CAUSAL INFERENCE WITH AUXILIARY OBSERVATIONS

太田悠太、星野崇宏、大津泰介

IES Keio DP2025-021
2025 年 9 月 4 日
JEL Classification: C14、C31
キーワード: study design, auxiliary variable, non-parametric identification, non-compliance, local average treatment effect

【要旨】

In the evaluation of social programs, it is often difficult to conduct randomized controlled experiments due to non-compliance; therefore the local average treatment effect (LATE) is commonly applied. However, LATE identifies the average treatment effect only for a subpopulation known as compliers and requires the monotonicity assumption. Given these limitations of LATE, this paper proposes a study design and strategy to non-parametrically identify the causal effects for larger populations (such as ATT and ATE) and to remove the monotonicity assumption in the cases of non-compliance. Our strategy utilizes two types of auxiliary observations, one is an outcome before assignment and the other is a treatment before assignment. These observations do not require specially designed experiments, and are likely to be observed in baseline surveys of the standard experiment or panel data. We present the results for the random assignment and those of multiply robust representations in the case where the random assignment is violated. We then present details of the GMM estimation and testing methods which utilize overidentified restrictions. The proposed methodology is illustrated by empirical examples which revisit influential studies by Thornton (2008), Gerber et al. (2009), and Beam (2016), as well as the data set from the Oregon Health Insurance Experiment and that from an experimental data on 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, TAKAHIRO HOSHINO, AND TAISUKE OTSU

Abstract. In the evaluation of social programs, it is often difficult to conduct randomized controlled experiments due to non-compliance; therefore the local average treatment effect (LATE) is commonly applied. However, LATE identifies the average treatment effect only for a subpopulation known as compliers and requires the monotonicity assumption. Given these limitations of LATE, this paper proposes a study design and strategy to non-parametrically identify the causal effects for larger populations (such as ATT and ATE) and to remove the monotonicity assumption in the cases of non-compliance. Our strategy utilizes two types of auxiliary observations, one is an outcome before assignment and the other is a treatment before assignment. These observations do not require specially designed experiments, and are likely to be observed in baseline surveys of the standard experiment or panel data. We present the results for the random assignment and those of multiply robust representations in the case where the random assignment is violated. We then present details of the GMM estimation and testing methods which utilize overidentified restrictions. The proposed methodology is illustrated by empirical examples which revisit influential studies by Thornton (2008), Gerber et al. (2009), and Beam (2016), as well as the data set from the Oregon Health Insurance Experiment and that from an experimental data on 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). Identification and estimation of the causal effects are typically carried out under the untestable assumption of unconfoundedness, that is, independence between treatment 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, 2017). In social program evaluations, it is difficult to conduct a randomized controlled experiment because noncompliance with an assigned treatment may occur, and the unconfoundedness will be violated (Imbens and Angrist, 1994).

In such cases, the local average treatment effect (LATE) can be identified using the assignment of treatment as an instrumental variable (IV) under conditions weaker than unconfoundedness (Imbens and Angrist, 1994; Angrist et al., 1996), and it consistently estimates a causal effect in a non-parametric manner without restrictions on the effect heterogeneity. However, LATE identifies the average treatment effect only for a subpopulation called compliers who react on the assignment as intended by the researcher. As compilers may not be representative of the entire population, it has been argued that LATE may not be a valid parameter for 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). For example, Imbens (2014) argued that "If the noncompliance is substantial, we are limited in the questions we can answer credibly and precisely." Therefore, even if LATE can be estimated, the causal effects of larger populations, such as the average treatment effect on treated (ATT) and average treatment effect (ATE) are still of interest. In addition, identifying LATE requires the monotonicity (i.e., there are no defiers who oppositely react on the assignment as intended by the researcher in the population or interest). If there are defiers in the population, LATE will be biased as it converges to a weighted difference between the effect of the treatment among compliers and defiers (Angrist et al., 1996; De Chaisemartin, 2017).

Given these limitations of LATE in the cases where non-compliance occurs, this paper presents a study design and strategy to non-parametrically identify the causal effects for ATT and ATE and to drop the monotonicity assumption. The strategy utilizes two types of auxiliary observations under the standard LATE assumptions consisting with the relevance between the assignment and the treatment, the exclusion restriction, and the monotonicity. One is an outcome before assignment that is commonly observed in baseline surveys of randomized experiments and is also widely used in difference-in-differences (DID) designs. The other is a treatment (or an other but similar treatment) before assignment that can be observed in the situations where the treatment is available before assignment (e.g., the treatment is available on the market before the experiment). Primarily supposing additional assumptions on the auxiliary observations, the outcome before assignment is used to identify ATT and ATE, and the treatment before assignment is used to drop the monotonicity. More specifically, by using the outcome before assignment, ATT can be identified under a parallel trend assumption on subpopulations, similar to that used in DID designs, and ATE can be identified by an extra assumption of another parallel trend or homogeneity between certain subpopulations. Furthermore, by using the treatment before assignment, the monotonicity can be dropped if either it is the same as the treatment status when assigned to the control group or it is a kind of instrument variable for the outcome through the treatment. Either of these strategies with auxiliary observations works alone, and they can also work together. In addition to identification results in the random assignment, this paper provides multiply robust representations in the case where the random assignment is violated such as in observational studies. Throughout the paper, we assume that both the assignment and treatment are binary.

One useful feature of these auxiliary observations is that they do not require specially designed experiments because they are observed without any manipulation or intervention. So they are likely to be observed in baseline surveys in standard experiments or several types of panel data such as administrative data or marketing platform data, and also available in data sets from existing studies. We now list some influential empirical studies, where LATE estimates may suffer from the above limitations and have one or both of the auxiliary observations. We revisit all these examples in our empirical illustrations. Each of Thornton (2008), Gerber et al. (2009) and Beam (2016) reports insignificant causal estimates from encouragement design experiments, and in each case the complier fraction is less than one half.[1] Importantly, each of the outcomes before assignment is available in their baseline surveys so that we can use their original data sets. As another example, based on the Oregon Health Insurance Experiment, which is a natural experiment with a lottery, Finkelstein et al. (2012) and Taubman et al. (2014) estimated the effects of enrolling in a health insurance called Medicaid by LATE. However, the population in the experiment may structurally include defiers because only about 30% of those who won the lottery actually enrolled, and there was another Medicaid program where people who had lost the lottery could also be enrolled (Finkelstein et al., 2012). Indeed, their original data set contains auxiliary data to apply our methodology: (i) a treatment status of an other public assistance program before assignment, which can be used as the similar treatment before assignment, is available in their administrative data, and (ii) the outcomes before assignment are also available in the data set of Taubman et al. (2014). We revisit these existing studies with our methodology in Section 5. This paper also discusses alternative auxiliary observations and augmented experiment designs.

Research on identification of other than the average treatment effect for compliers (hereafter, ATE(c)) with an IV has been limited, and it has discussed only under the use of basic observations that are the assignment, treatment, and outcome. Most studies investigated valid assumptions to establish equivalence between ATE and LATE. Martens et al. (2006) and Hernán and Robins (2006) assumed homogeneity conditions that are guaranteed to hold under the null hypothesis of no treatment effect, e.g., the case where the treatment effect is identical for all subjects (Hernán and Robins, 2006; Tan, 2010). Also several studies assumed another homogeneity condition, where ATE equals to LATE conditional on covariates (Angrist and Fernandez-Val, 2010; Aronow and Carnegie, 2013; Fricke et al., 2020). Instead of homogeneity, some studies imposed restrictions on unobserved confounders with or without covariates (Wang and Tchetgen Tchetgen, 2018; Cui and Tchetgen Tchetgen, 2021; Hartwig et al., 2023). Applicability of these approaches would be limited without covariates, and even with covariates; it would be often difficult to test or argue that sufficient covariates are included to guarantee their conditions. Alternatively, the marginal treatment effects identify general parameters, including ATT and ATE, with a continuous instrument variable (Heckman and Vytlacil, 1999, 2005). However, the assignment is binary in many cases of social experiments and identification of the marginal treatment effects with a binary instrument needs parametric assumptions (Brinch et al., 2017). It is also known that with continuous instruments estimation of ATE requires specification of a parametric marginal treatment effect curve and its extrapolation (Cornelissen et al., 2016; Sigstad, 2024). Unlike previous research, by using the outcome before assignment, our strategy achieves non-parametric identification of ATT and ATE based on indirectly testable or interpretable assumptions with or without covariates.

Toleration on defiers has been the focus of several studies on LATE. Most studies have investigated the relaxation of the monotonicity (Small et al., 2017; Klein, 2010; van't Hoff et al.,

[1]Thornton (2008) employs two-stage least squares estimation, while Beam (2016) adopts the LATE estimator. Gerber et al. (2009) report intention-to-treat estimates; however, as discussed in our empirical illustrations (Section 5), applying the LATE estimator to their data also fails to yield statistically significant effects. The fraction of compliers in each case, which is also detailed in Section 5, is less than half.

2023; Dahl et al., 2023). For example, Dahl et al. (2023) assumed that there are only compliers or defiers in each supported value of the potential outcome. As another approach, De Chaisemartin (2017) assumed a restriction on the characteristics of the target population that a certain fraction of the compliers have the same average effect and population as the defiers. In contrast to these studies, by making use of the treatment before assignment, our strategy drops entirely the monotonicity assumption based on additional conditions on the auxiliary data.

Auxiliary variables and augmented experiment designs have been used in the literature to adjust for post-treatment variables such as non-compliance, censoring by death, or surrogate outcomes (Follmann, 2006; Ding et al., 2011; Mealli and Pacini, 2013; Gabriel and Gilbert, 2014; Jiang et al., 2016; Gabriel and Follmann, 2016; Yang and Small, 2016; Jiang and Ding, 2021). Much of this literature, following the principal stratification framework (Frangakis and Rubin, 2002), focused on identifying principal causal effects (PCEs), which are the effects of the treatment on the outcome conditioning on post-treatment variables between the treatment and outcome. In contrast, we focus on the more policy-relevant effects of the post-assignment variable on the outcome, i.e. the effect of the treatment itself under non-compliance. Although several studies employed an outcome before assignment (Ding et al., 2011; Mealli and Pacini, 2013) or observations of a post-assignment variable at baseline, i.e. the treatment before assignment under non-compliance (Follmann, 2006; Gabriel and Gilbert, 2014; Gabriel and Follmann, 2016), the way we use the auxiliary observations also differs from theirs. Some of this literature proposed additional conditions to identify PCEs without the monotonicity (Ding et al., 2011; Jiang et al., 2016). However, it has been shown that these approaches are limited to achieve local identification even when further constraints are placed on the auxiliary variables. The aspects with partial relevance are covered in Section 2.2.2. Several studies on mediation analysis also used augmented experiment designs (Mattei and Mealli, 2011; Imai et al., 2013), such as directly or indirectly manipulating the mediation variables. These approaches are also different from our objectives and methodology.

This paper proceeds as follows. Section 2 presents identification results for one or both uses of the auxiliary observations. The results for the random assignment and those of multiply robust representations in the case where the random assignment is violated are also provided. Section 3 presents the GMM estimation and testing methods which utilize overidentified restrictions. In Section 4, we present identification results with alternative auxiliary observations or augmented experiment designs. In Section 5, we present empirical illustrations of the proposed methods. Section 6 concludes.

This paper contains supplementary materials. In Appendix A, we discuss extensions of our identification results in Sections 2.1 and 2.2 to the case where the conditional ignorability holds true. Appendix B presents proofs and derivations in the main paper. Appendix C illustrates finite sample performances of the proposed methodology by Monte Carlo simulation. Appendix D provides an additional result for an empirical illustration in Section 5.

2. Identification

In this section, we present our identification results. Section 2.1 provides a strategy for identifying ATT and ATE, in addition to ATE for compliers (denoted by ATE(c)), in the case where the researcher can observe an outcome Y^{pre} before the assignment and maintain a monotonicity assumption to rule out defiers. Section 2.2 presents a strategy for identifying ATE(c) in the case where defiers may exist in the population but the researcher can observe a treatment D^{pre} before the assignment. Furthermore, Section 2.3 presents a strategy for identifying ATE(c), ATT, and ATE in the case where defiers may exist in the population but the researcher can observe both Y^{pre} and D^{pre}. While these sections assume that the assignment is randomized, Section 2.4 investigates the case where the random assignment is violated, such as in observational studies, but ignorability conditional on observed covariates holds true.

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 $D = d$. The basic assumption that underlies in this section is as follows:

Assumption 1.

(i): [Exclusion and consistency] *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): [Relevance] $\text{Cov}(D, Z) \neq 0$.

This assumption is standard in the literature of causal inference 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. We introduce a principal strata variable (Frangakis and Rubin, 2002):

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

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.

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 $\mu_d^u = \mathbb{E}[Y_d | U = u]$

and $p^u = \mathbb{P}(U = u)$ for $d \in \{1, 0\}$ and $u \in \{c, a, n, d\}$ as follows

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

Identification of all parameters on the right-hand side of each estimand leads to identification of the estimands on the left-hand side.

Remark 1. [Alternative definitions of ATT] To the best of our knowledge, the definition of ATT under two-sided noncompliance has received little attention in the econometrics literature. In this paper, we propose and formalize three distinct and meaningful definitions of ATT:

$$\begin{aligned}
\mathrm{ATT}_{D_1} &= \mathbb{E}[Y_1 - Y_0 | D_1 = 1] = \frac{p^c(\mu_1^c - \mu_0^c) + p^a(\mu_1^a - \mu_0^a)}{p^c + p^a}, \\
\mathrm{ATT}_{D_0} &= \mathbb{E}[Y_1 - Y_0 | D_0 = 1] = \frac{p^a(\mu_1^a - \mu_0^a) + p^d(\mu_1^d - \mu_0^d)}{p^a + p^d}, \qquad (3) \\
\mathrm{ATT}_{D} &= \mathbb{E}[Y_1 - Y_0 | D = 1] \\
&= \frac{p^a(\mu_1^a - \mu_0^a) + \mathbb{P}(Z=1)p^{c|Z=1}(\mu_1^{c|Z=1} - \mu_0^{c|Z=1}) + \mathbb{P}(Z=0)p^{d|Z=0}(\mu_1^{d|Z=0} - \mu_0^{d|Z=0})}{p^a + \mathbb{P}(Z=1)p^{c|Z=1} + \mathbb{P}(Z=0)p^{d|Z=0}},
\end{aligned}$$

where $\mu_d^{u|Z=z} = \mathbb{E}[Y_d | U = u, Z = z]$ and $p^{u|Z=z} = \mathbb{P}(U = u | Z = z)$. The object ATT_{D_1} captures the treatment effect for those who receive treatment when incentivized, and corresponds to ATT identified by LATE under one-sided compliance. Under one-sided compliance, the restriction $D_0 = 0$ forces individuals who would otherwise act as always-takers to behave as compliers. As a result, LATE identified under one-sided compliance coincides with a weighted average of the treatment effects for the compliers and always-takers in a hypothetical setting without enforcement—that is, under two-sided compliance. The object ATT_{D_0} captures the treatment effect for individuals who receive treatment in the absence of incentives. This is conceptually close to ATT identified by a DID design in non-experimental settings. In the absence of defiers, this object coincides with the treatment effect of always-takers. The object ATT_D represents the treatment effect for individuals who actually received treatment within the experiment. Under random assignment, this becomes a weighted average of the treatment effects for compliers, always-takers, and defiers, weighted by the proportion $(p^a, \mathbb{P}(Z = 1)p^c, \mathbb{P}(Z = 0)p^d)$ for each group treated.

A key insight is that all these objects for ATT are identifiable once the corresponding components on the right-hand side are identified. Throughout the paper, we focus on the definition $\mathrm{ATT} := \mathrm{ATT}_{D_1}$, while we also comment on identification for the other objects, ATT_{D_0} and ATT_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. In this subsection, we impose the monotonicity assumption.

Assumption 2. [Monotonicity] *D_z is weakly monotone in z, i.e.,* $\mathbb{P}(D_1 \geq D_0) = 1$.

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. Observable Y^{pre}

This assumption says that there is no defiers in the population, i.e., $\mathbb{P}(U = d) = 0$. Assumptions 1 and 2 implies that ATE takes the following form:

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

It is known that ATE(c) is identified under Assumptions 1 and 2 (Angrist et al., 1996), whereas ATT and ATE are not identified in general under the same assumptions. This subsection provides a strategy to identity ATT and ATE when the researcher can access to the outcome variable Y^{pre} before the assignment in addition to the main observable (Y, D, Z).

Assumption Y.

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

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

The observations from randomized experiments can be divided into the four rows in Table 1 according to the values of D and Z. As described below, the parameters in (2) other than μ_0^a and μ_1^n are identified under Assumptions 1 and 2 without using Y^{pre}. On the other hand, the outcomes to compute μ_0^a and μ_1^n are never directly observed because of their definitions (i.e., always-takers always receive treatment and never-takers never receive treatment). To address this problem, we introduce the auxiliary outcome Y^{pre}.

First of all, Table 1 suggests that the following objects are identified:

$$\begin{aligned}
\mu_1^a &= \mathbb{E}[Y|Z = 0, D = 1], \qquad p^a = \mathbb{P}(D = 1|Z = 0),\\
\mu_1^n &= \mathbb{E}[Y|Z = 1, D = 0], \qquad p^n = \mathbb{P}(D = 0|Z = 1),\\
p^c &= \mathbb{P}(D = 1|Z = 1) - p^a,\\
\mu_1^c &= \frac{(p^c + p^a)\mathbb{E}[Y|Z = 1, D = 1] - p^a\mu_1^a}{p^c},\\
\mu_0^c &= \frac{(p^c + p^n)\mathbb{E}[Y|Z = 0, D = 0] - p^n\mu_0^n}{p^c}. \qquad (4)
\end{aligned}$$

Therefore, under Assumptions 1, 2, and Y, we can identify ATE for compliers as ATE$(c) = \mu_1^c - \mu_0^c$, which is an alternative expression for the standard LATE (note that this identification step does not use Y^{pre}).

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

Assumption 3.

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

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

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. It should be noted that this assumption can be tested if at least one additional pre-treatment outcome is available. 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

$$\begin{aligned}
\mu_1^a - \mu_0^a &= \mathbb{E}[Y_1 - Y^{\text{pre}}|U = a] - \mathbb{E}[Y_0 - Y^{\text{pre}}|U = a] \\
&= \mathbb{E}[Y_1 - Y^{\text{pre}}|U = a] - \mathbb{E}[Y_0 - Y^{\text{pre}}|U = n] \\
&= \mu_1^a - \mu_{\text{pre}}^a - \mu_0^n + \mu_{\text{pre}}^n, \qquad (5)
\end{aligned}$$

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 3 (i). Since μ_{pre}^a and μ_{pre}^n are identified by

$$\begin{aligned}
\mu_{\text{pre}}^a &= \mathbb{E}[Y^{\text{pre}}|Z = 0, D = 1], \\
\mu_{\text{pre}}^n &= \mathbb{E}[Y^{\text{pre}}|Z = 1, D = 0], \qquad (6)
\end{aligned}$$

we can identify μ_0^a using (5), which shows that $\mu_0^a = \mu_{\text{pre}}^a + \mu_0^n - \mu_{\text{pre}}^n$, and thus ATT is also identified by (2). 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 (2). Assumption 3 (ii) 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 Assumptions 1, 2,* ATE(c) *is identified.*

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

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

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

Remark 2. [Alternative assumptions] Assumption 3 (i) posits parallel trends between always-takers and never-takers (i.e., between non-reactive strata). Alternatively, μ_0^a can also be identified by assuming parallel trends between different strata. One alternative is to assume parallel trends between always-takers and compliers (i.e., between strata with $D_1 = 1$):

$$\mathbb{E}[Y_0 - Y^{\text{pre}}|U = a] \quad = \quad \mathbb{E}[Y_0 - Y^{\text{pre}}|U = c].$$

Other alternatives can be formulated by assuming parallel trends between strata defined by their potential treatment status. For example, one could assume trends are parallel based on the value of D_0 or D_1, i.e., $\mathbb{E}[Y_0 - Y^{\text{pre}}|D_0 = 1] = \mathbb{E}[Y_0 - Y^{\text{pre}}|D_0 = 0]$ or $\mathbb{E}[Y_0 - Y^{\text{pre}}|D_1 = 1] = \mathbb{E}[Y_0 - Y^{\text{pre}}|D_1 = 0]$. These alternative assumptions are testable if at least one additional pre-treatment outcome is available. See Appendix B.3 for the proofs of identification of μ_0^a under these assumptions.

Remark 3. [Partial identification] If we relax Assumption 3 (ii) and allow for $\mathbb{E}[Y_1 - Y_0|U = a] \neq \mathbb{E}[Y_1 - Y_0|U = n]$, ATE is no longer point identified. Instead, ATE is partially identified within an interval determined by $\mu_1^n = \mathbb{E}[Y_1|U = n]$. If we assume that μ_1^n lies in a known interval $[\mu_1^{n,\min}, \mu_1^{n,\max}]$, then the identified set for ATE is obtained as

$$\text{ATE} \quad \in \quad \left[\sum_{u \in \{c,a\}} p^u \text{ATE}(u) + p^n(\mu_1^{n,\min} - \mu_0^n), \sum_{u \in \{c,a\}} p^u \text{ATE}(u) + p^n(\mu_1^{n,\max} - \mu_0^n) \right],$$

where $\text{ATE}(u) = \mathbb{E}[Y_1 - Y_0|U = u]$. Since $\text{ATE}(c)$ and $\text{ATE}(a)$ are point identified, only $\text{ATE}(n)$ is subject to partial identification. A small proportion of never-takers (i.e., smaller value of p^n) leads to a relatively tight and informative identified set for ATE. In practice, for binary outcomes, we can set $\mu_1^{n,\min} = 0$ and $\mu_1^{n,\max} = 1$, reflecting the logical limits on potential outcomes (see Manski 2003). For continuous outcomes, $\mu_1^{n,\min}$ and $\mu_1^{n,\max}$ may be set based on the theoretical support of Y. For example, if it is theoretically guaranteed that treatment effects are non-negative, setting $\mu_1^{n,\min} = \mu_0^n$ may be reasonable.

Remark 4. [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 two additional restrictions:

$$\begin{aligned}
\mu_{\text{pre}}^c &= \frac{(p^c + p^a)\mathbb{E}[Y^{\text{pre}}|Z = 1, D = 1] - p^a \mu_{\text{pre}}^a}{p^a}, \\
\mu_{\text{pre}}^n &= \frac{(p^c + p^n)\mathbb{E}[Y^{\text{pre}}|Z = 0, D = 0] - p^c \mu_{\text{pre}}^c}{p^n}.
\end{aligned} \tag{7}$$

These additional moment conditions can be incorporated by using the GMM approach.

Remark 5. [Alternative definitions of ATT] Under Assumptions 1, 2, and Y, we can express the objects ATT_{D_0} and ATT_D in (3) as

$$\text{ATT}_{D_0} = \mu_1^a - \mu_0^a, \qquad \text{ATT}_D = \frac{p^a(\mu_1^a - \mu_0^a) + \mathbb{P}(Z = 1)p^c(\mu_1^c - \mu_0^c)}{p^a + \mathbb{P}(Z = 1)p^c}.$$

Since all components on the right-hand side of each expression are identified under the approach described above, Theorem 1 (ii) remains valid regardless of whether ATT is defined by ATT_{D_0} or ATT_D.

2.2. **Observable treatment before assignment.** This subsection focuses exclusively on ATE(c) as the estimand in the presence of defiers (i.e., without requiring the monotonicity in Assumption 2). Clearly ATE(c) cannot be identified under Assumption 1. In this subsection, we assume that the researcher can observe a treatment variable D^{pre} before the assignment in addition to the main observable (Y, D, Z).

Assumption D.

- **(i):** [Observable pre-assignment treatment indicator] *A treatment variable* $D^{\text{pre}} \in \{0, 1\}$ *is observable at a time before the assignment.*
- **(ii):** [Random assignment] Z *is independent from* $(D^{\text{pre}}, D_1, D_0, Y_{11}, Y_{10}, Y_{01}, Y_{00})$.

Assumption D (i) is considered natural when the treatment of interest is available before the assignment (e.g. ever had medical check-up). Assumption D (ii) is a standard assumption for random assignment of Z. Intuitively, D^{pre} is used for untangling the mixtures of principal strata in observations. The observations from randomized experiments can be divided into the four rows in Table 2 (left) according to the values of D and Z. 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 ATE(c) are identified by using the single principal stratum moments from the second and third rows. Otherwise we have to deal with the mixtures, and we employ D^{pre} to overcome this issue. We first present an identification strategy for ATE(c) under a naive assumption on D^{pre} in Section 2.2.1 to grasp an intuition, and then discuss identification under a relaxed assumption in Section 2.2.2.

2.2.1. *Stable case.* To conduct identification analysis for ATE(c) in the presence of defiers, we impose the following assumption.

Assumption 4. [Stable treatment status] $\mathbb{P}(\mathrm{D}_0 = \mathrm{D}^{\text{pre}}) = 1$.

Assumption 4 says that D^{pre} plays the role of D_z when $Z = 0$. Since the treatment D^{pre} occurs before the assignment of Z, this assumption is reasonable if the treatment status is stable. The relationships of the observables and principal strata variable can be summarized as in Table 2. Due to Assumption 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:

$$
\begin{aligned}
\mu_1^c &= \mathbb{E}[Y|Z=1, D=1, D^{\text{pre}}=0], \qquad p^c = \mathbb{P}(D=1, D^{\text{pre}}=0|Z=1), \\
\mu_1^a &= \mathbb{E}[Y|Z=1, D=1, D^{\text{pre}}=1], \qquad p^a = \mathbb{P}(D=1, D^{\text{pre}}=1|Z=1), \\
\mu_0^n &= \mathbb{E}[Y|Z=1, D=0, D^{\text{pre}}=0], \qquad p^n = \mathbb{P}(D=0, D^{\text{pre}}=0|Z=1), \\
\mu_0^d &= \mathbb{E}[Y|Z=1, D=0, D^{\text{pre}}=1], \qquad p^d = \mathbb{P}(D=0, D^{\text{pre}}=1|Z=1). \qquad (8)
\end{aligned}
$$

U	Y	D	Z
c or a	Y_1	1	1
n or d	Y_0	0	1
a or d	Y_1	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_0	0	1	0
d	Y_0	0	1	1
a or d	Y_1	1	0	1
c or n	Y_0	0	0	0

TABLE 2. Case where monotonicity is not assumed with (right) and without (left) D^{pre}

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

$$\mu_1^d = \frac{(p^a + p^d)\mathbb{E}[Y|Z=0, D=1] - p^a \mu_1^a}{p^d},$$
$$\mu_0^c = \frac{(p^c + p^n)\mathbb{E}[Y|Z=0, D=0] - p^n \mu_0^n}{p^c}. \tag{9}$$

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. *Under Assumptions 1, 4, and 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 statistical inference based on standard methods. There are several ways to utilize this theorem for empirical analyses. First, we can estimate the probability of defiers p^d as a diagnostics for the monotonicity 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 $\text{ATE}(c)$ are same for the cases with or without D^{pre}. Finally, although we cannot identify ATT or ATE, this theorem can be utilized to obtain tighter identified sets for these objects compared to the conventional ones.

Remark 6. [Overidentification] The above argument for establishing Theorem 2 shows just identification of the 10 parameters, $\mu_1^c, \mu_0^c, \mu_1^d, \mu_0^d, \mu_1^a, \mu_0^n$ and p^u for $u \in \{c, a, n, d\}$. Indeed the last two rows of Table 2 (right panel) imply the additional moment conditions:

$$\begin{aligned} p^a + p^d &= \mathbb{P}(D=1|Z=0, D^{\text{pre}}=1) = \mathbb{P}(D=1|Z=0), \\ p^c + p^n &= \mathbb{P}(D=0|Z=0, D^{\text{pre}}=0) = \mathbb{P}(D=0|Z=0), \end{aligned} \tag{10}$$

which can be incorporated by using the GMM approach.

2.2.2. *Unstable case* . This subsection relaxes Assumption 4 (i.e., $D_0 = D^{\text{pre}}$ almost surely). Without this assumption, the relationships of the observables and principal strata variable are summarized in Table 3. Since we allow $D_0 \neq D^{\text{pre}}$, each principal strata of the first four rows is not uniquely identified. The fifth and seventh rows for the cases of $D \neq D^{\text{pre}}$ with $Z = 0$ are added to Table 2 (right) . We call this case as the unstable case.

Instead of Assumption 4, we impose the following assumptions.

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

TABLE 3. Unstable case with D^{pre}

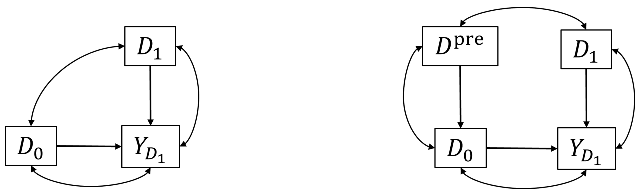

FIGURE 1. Relationship with (right) and without (left) D^{pre}. The bi-directed arrows indicate common causes.

Assumption 5.

- **(i):** [Exclusion restriction on treatment before assignment] $\mathbb{E}[Y_{D_1}|U = u, D^{\text{pre}}] = \mathbb{E}[Y_{D_1}|U = u]$ *for each* $u \in \{c, a, n, d\}$.
- **(ii):** [Exclusion restriction through treatment before assignment] $\mathbb{P}(D_0 = 1|D^{\text{pre}}, D_1) = \mathbb{P}(D_0 = 1|D^{\text{pre}})$.
- **(iii):** [Relevance condition on treatment before assignment] $\text{Cov}(D_0, D^{\text{pre}}) \neq 0$.

As described below, ATE(c), ATE(d), and p^u for all $u \in \{a, c, d, n\}$ are identified under Assumption 5. Assumptions 5 (i) and (iii) together imply D^{pre} works as an instrumental variable for Y_{D_1} through D_0. Assumption 5 (ii) implies that D_1 relates to D_0 only through D^{pre}; in other words, D^{pre} fully mediates the common cause between D_1 and D_0. Figure 1 (left) provides the general relationship without D^{pre} for comparison. Figure 1 (right) illustrates a relationship with D^{pre} satisfying Assumption 5. This relationship arises naturally in certain settings. D^{pre} is expected to account many causes of D_0 and not affect Y_{D_1} when there is no incentive before the assignment. Moreover, as long as D^{pre} satisfies the assumptions, it does not need to correspond to the same treatment as the one eventually assigned. A similar treatment may also satisfy the assumptions. For example, if the goal is to estimate the effect of coupon usage for product A, the coupon usage status for a different product B may serve as D^{pre}, which is likely to be associated with D but unlikely to affect sales of product A.

Our Assumption 5 (i) adopts a similar principle to the auxiliary independence assumption that has been used to identify PCEs (Ding et al., 2011; Jiang et al., 2016; Jiang and Ding, 2021). While this literature has established conditions for identifying PCEs without monotonicity, these results are limited to local identification. In contrast, although our estimand differs, our approach achieves point identification of ATE(c) and ATE(d) without monotonicity by leveraging D^{pre} and additional Assumptions 5 (ii) and (iii).

We now present the identification result under Assumptions 1, D, and 5. First, note that the following objects are identified under Assumptions 1 and D:

$$\begin{aligned} \delta_{(z,d,d')} &= \mathbb{E}[Y_d|D_z=d, D^{\text{pre}}=d'], & \delta_{(z,d)} &= \mathbb{E}[Y_d|D_z=d], \\ \pi_{(z,d,d')} &= \mathbb{P}(D_z=d, D^{\text{pre}}=d'), & \rho_{(z,d,d')} &= \mathbb{P}(D_z=d|D^{\text{pre}}=d'). \end{aligned} \tag{11}$$

Then $\mu_1^a, \mu_1^c, \mu_0^d, \mu_0^n$, and p^u for $u \in \{a, c, d, n\}$ are identified as follows. Observe that

$$\begin{aligned} \mu_b^u &= \frac{\delta_{(1,b,b')}\rho_{(0,1-b',1-b')} - \delta_{(1,b,1-b')}\rho_{(0,1-b',b')}}{\rho_{(0,1-b',1-b')} - \rho_{(0,1-b',b')}}, \\ p^u &= \pi_{(1,b,b')}\rho_{(0,b',b')} + \pi_{(1,b,1-b')}\rho_{(0,b',1-b')}, \end{aligned} \tag{12}$$

where

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

Intuitively, the first four rows in Table 3 provide $\delta_{(z,d,d')}$ and $\pi_{(z,d,d')}$, which are weighted averages of expectations from subjects of $D_0 = D^{\text{pre}}$ and $D_0 \neq D^{\text{pre}}$, and the last four rows in Table 3 provide $\rho_{(z,d,d')}$, which is a fraction of the subjects of $D_0 = D^{\text{pre}}$ and $D_0 \neq D^{\text{pre}}$. μ_b^u is identified by a weighted difference of expectations from subjects of $D_0 = D^{\text{pre}}$ and $D_0 \neq D^{\text{pre}}$. p^u is identified by a weighted average of expectations from subjects of $D_0 = D^{\text{pre}}$ and $D_0 \neq D^{\text{pre}}$.

Identification of μ_1^c is outlined as follows. Note that

$$\begin{aligned} \delta_{(1,1,0)} &= \mathbb{E}[Y_1|D_1=1, D^{\text{pre}}=0] \\ &= \mathbb{E}[Y_1|U=c, D^{\text{pre}}=0]\mathbb{P}(D_0=0|D_1=1, D^{\text{pre}}=0) \\ &\quad + \mathbb{E}[Y_1|U=a, D^{\text{pre}}=0]\mathbb{P}(D_0=1|D_1=1, D^{\text{pre}}=0) \\ &= \mathbb{E}[Y_1|U=c]\mathbb{P}(D_0=0|D^{\text{pre}}=0) + \mathbb{E}[Y_1|U=a]\mathbb{P}(D_0=1|D^{\text{pre}}=0) \\ &= \mu_1^c\rho_{(0,0,0)} + \mu_1^a\rho_{(0,1,0)}, \end{aligned} \tag{13}$$

where the third equality follows from Assumptions 5 (i) and (ii). $\delta_{(1,1,0)}$ consists of the subject with $D_0 = D^{\text{pre}}$ and the subject with $D_0 \neq D^{\text{pre}}$ both have $D_1 = 1$. Since it is conditional on $D^{\text{pre}} = 0$, the subject with $D_0 = D^{\text{pre}}$ is a complier and the subject with $D_0 \neq D^{\text{pre}}$ is an always-taker. Then $\delta_{(1,1,0)}$ will be the sum of $\mu_1^c\rho_{(0,0,0)}$ and $\mu_1^a\rho_{(0,1,0)}$, where $\rho_{(0,0,0)}$ is the fraction of $D_0 = D^{\text{pre}}$ and $\rho_{(0,1,0)}$ is the fraction of $D_0 \neq D^{\text{pre}}$ among those with $D^{\text{pre}} = 0$. Similarly, we have

$$\delta_{(1,1,1)} = \mu_1^a\rho_{(0,1,1)} + \mu_1^c\rho_{(0,0,1)}. \tag{14}$$

Therefore, eliminating the term μ_1^a from these equations yields identification of μ_1^c:

$$\mu_1^c = \frac{\delta_{(1,1,0)}\rho_{(0,1,1)} - \delta_{(1,1,1)}\rho_{(0,1,0)}}{\rho_{(0,1,1)} - \rho_{(0,1,0)}}. \tag{15}$$

Here zero-division is avoided by Assumption 5 (iii).

Furthermore, p^c can be identified as follows:

$$
\begin{aligned}
p^c &= \mathbb{P}(D_1 = 1, D_0 = 0) \\
&= \mathbb{P}(D_1 = 1, D^{\text{pre}} = 0)\mathbb{P}(D_0 = 0|D_1 = 1, D^{\text{pre}} = 0) \\
&\quad + \mathbb{P}(D_1 = 1, D^{\text{pre}} = 1)\mathbb{P}(D_0 = 0|D_1 = 1, D^{\text{pre}} = 1) \\
&= \mathbb{P}(D_1 = 1, D^{\text{pre}} = 0)\mathbb{P}(D_0 = 0|D^{\text{pre}} = 0) \\
&\quad + \mathbb{P}(D_1 = 1, D^{\text{pre}} = 1)\mathbb{P}(D_0 = 0|D^{\text{pre}} = 1) \\
&= \pi_{(1,1,0)}\rho_{(0,0,0)} + \pi_{(1,1,1)}\rho_{(0,0,1)},
\end{aligned} \tag{16}
$$

where the third equality follows from Assumption 5 (ii). Using the identified parameters above, μ_1^d and μ_0^c are identified as

$$
\mu_1^d = \frac{(p^d + p^a)\delta_{(0,1)} - p^a\mu_1^a}{p^d}, \qquad \mu_0^c = \frac{(p^c + p^n)\delta_{(0,0)} - p^n\mu_0^n}{p^c}. \tag{17}
$$

Therefore, $\text{ATE}(c) = \mu_1^c - \mu_0^c$ and $\text{ATE}(d) = \mu_1^d - \mu_0^d$ are identified. Combining these results, our identification results for this case are presented as follows.

Theorem 3. *Under Assumptions 1, D, and 5,* $\text{ATE}(c)$*,* $\text{ATE}(d)$*, and* p^u *for all* $u \in \{a, c, d, n\}$ *are identified.*

Remark 7. [Switched version of Assumption 5] Under Assumption 5' (which switches D_0 and D_1 in Assumption 5), a similar identification approach can be applied. This assumption is natural in settings where encouragement for treatment has been frequently (or extensively) implemented in advance. D^{pre} is expected to account many causes of D_1 and not affect Y_{D_0} when there are incentives before assignment. The identification result under 1, D, and 5' is following. Note that

$$
\begin{aligned}
\mu_{b'}^u &= \frac{\delta_{(0,b',b')}\rho_{(1,1-b,1-b')} - \delta_{(0,b',1-b')}\rho_{(1,1-b,b')}}{\rho_{(1,1-b,1-b')} - \rho_{(1,1-b,b')}}, \\
p^u &= \pi_{(0,b',b')}\rho_{(1,b,b')} + \pi_{(0,b,1-b')}\rho_{(1,b,1-b')},
\end{aligned} \tag{18}
$$

where (b, b') is same as the one in (12). Using the identified parameters above, μ_1^c and μ_0^d are identified as

$$
\mu_1^c = \frac{(p^c + p^a)\delta_{(1,1)} - p^a\mu_1^a}{p^c}, \qquad \mu_0^d = \frac{(p^d + p^n)\delta_{(1,0)} - p^n\mu_0^n}{p^d}. \tag{19}
$$

Remark 8. [Overidentification] The above argument for establishing Theorem 3 shows just identification of the 10 parameters, $\mu_1^c, \mu_0^c, \mu_1^d, \mu_0^d, \mu_1^a, \mu_0^n$ and p^u for $u \in \{c, a, n, d\}$. For identification under Assumption 5, the last four rows of Table 3 imply the additional two moment conditions:

$$
p^a + p^d = \pi_{(0,1,1)} + \pi_{(0,1,0)}, \qquad p^c + p^n = \pi_{(0,0,1)} + \pi_{(0,0,0)}. \tag{20}
$$

For identification under Assumption 5', the first four rows of Table 3 imply the additional two moment conditions:

$$
p^c + p^a = \pi_{(1,1,1)} + \pi_{(1,1,0)}, \qquad p^n + p^d = \pi_{(1,0,1)} + \pi_{(1,0,0)}. \tag{21}
$$

These moment conditions can be incorporated by using the GMM approach.

Remark 9. [Impose both Assumptions 5 and 5'] When both Assumptions 5 and 5' hold, the 10 parameters, $\mu_1^c, \mu_0^c, \mu_1^d, \mu_0^d, \mu_1^a, \mu_0^n$, and p^u for $u \in \{c, a, n, d\}$ are identified by (12) and (18). Among these parameters, the six parameters, μ_1^a, μ_0^n, and p^u for $u \in \{c, a, n, d\}$ are identified in the two ways and can be incorporated by using the GMM approach.

Remark 10. [Special case] Assumption 4 (i.e., $D_0 = D^{\text{pre}}$ almost surely) obviously satisfies Assumption 5. By imposing $D_0 = D^{\text{pre}}$, (12) reduces to (8), and (20) reduces to (10) because of $\rho_{(0,1,1)} = \rho_{(0,0,0)} = 1$, $\rho_{(0,1,0)} = \rho_{(0,0,1)} = 0$, and $\pi_{(0,1,0)} = \pi_{(0,0,1)} = 0$. Since (17) is the same as (9), the identification results under Assumption 5 reduces to those in Section 2.2.1.

Remark 11. [Role of D^{pre}] To provide a more rigorous foundation for our arguments, we can explicitly define the role of D^{pre} within the potential outcomes framework. Let $D_{zd'} \in \{0,1\}$ be the potential treatment status when $Z = z$ and $D^{\text{pre}} = d'$, and let $Y_{dd'} \in \mathcal{Y}$ be the potential outcome when $D = d$ and $D^{\text{pre}} = d'$. With this more granular notation, the foundational assumptions can be restated. In particular, Assumptions 1 (i) and D (ii) can be restated as

Assumption 1 (i)': It holds $Y_{dd'} = Y_{zdd'}$ for each $z \in \{0,1\}$ and $d \in \{0,1\}$ and $d' \in \{0,1\}$ and the observed variables D and Y are constructed from the potential outcomes via the following consistency assumptions:

$$D_z = D^{\text{pre}} D_{z1} + (1 - D^{\text{pre}}) D_{z0}, \qquad D = Z D_1 + (1 - Z) D_0,$$
$$Y_d = D^{\text{pre}} Y_{d1} + (1 - D^{\text{pre}}) Y_{d0}, \qquad Y = D Y_1 + (1 - D) Y_0.$$

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

Assumption D (ii)': Z is independent from $(D^{\text{pre}}, D_{11}, D_{10}, D_{01}, D_{00}, Y_{111}, Y_{101}, Y_{011}, Y_{001}, Y_{110}, Y_{100}, Y_{010}, Y_{000})$.

The identification results in Theorems 2 and 3 hold true even under this more formal setting. This is because Assumptions 4 and 5, in the stable and unstable cases respectively, provide sufficient restrictions to eliminate heterogeneity induced by D^{pre}, ensuring the original identification strategies remain valid. See Appendix B for a detailed discussion.

2.3. **Observe previous treatment and outcome.** By combining the results in Sections 2.1 and 2.2.2, ATE(c), ATT, and ATE are identified without monotonicity. We now consider the situation where both Y^{pre} and D^{pre} are observed.

Assumption YD.

- **(i):** [Observable pre-treatment outcome] *An outcome variable* $Y^{\text{pre}} \in \mathcal{Y}$ *is observable at a time before the treatment* D *is realized.*
- **(ii):** [Observable pre-assignment treatment indicator] *An treatment variable* $D^{\text{pre}} \in \{0,1\}$ *is observable at a time before the assignment.*
- **(iii):** [Random assignment] Z *is independent from* $(Y^{\text{pre}}, D^{\text{pre}}, D_1, D_0, Y_{11}, Y_{10}, Y_{01}, Y_{00})$.

Although both Y^{pre} and D^{pre} are denoted with the superscript "pre" under Assumption YD, this does *not* necessarily indicate that they are observed at the same point in time. The timing

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

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

TABLE 4. Case where monotonicity is not assumed with (right) and without (left) auxiliary data

relationship between Y^{pre} and D^{pre} is discussed in connection with Assumption 6, which is introduced below. The relationships of the observables and principal strata variable U can be summarized as in Table 4.

First, note that the following objects are identified under Assumptions 1 and YD:

$$\begin{aligned} \delta_{(z,d,d')} &= \mathbb{E}[Y_d|D_z = d, D^{\text{pre}} = d'], & \delta_{(z,d)} &= \mathbb{E}[Y_d|D_z = d], \\ \delta^{\text{pre}}_{(z,d,d')} &= \mathbb{E}[Y^{\text{pre}}|D_z = d, D^{\text{pre}} = d'], & \delta^{\text{pre}}_{(z,d)} &= \mathbb{E}[Y^{\text{pre}}|D_z = d], \\ \pi_{(z,d,d')} &= \mathbb{P}(D_z = d, D^{\text{pre}} = d'), & \rho_{(z,d,d')} &= \mathbb{P}(D_z = d|D^{\text{pre}} = d'). \end{aligned} \tag{22}$$

Under Assumption 5, $\mu_1^a, \mu_1^c, \mu_0^d, \mu_0^n$ and p^u's are identified in the same manner as in (12):

$$\begin{aligned} \mu_b^u &= \frac{\delta_{(1,b,b')}\rho_{(0,1-b',1-b')} - \delta_{(1,b,1-b')}\rho_{(0,1-b',b')}}{\rho_{(0,1-b',1-b')} - \rho_{(0,1-b',b')}}, \\ p^u &= \pi_{(1,b,b')}\rho_{(0,b',b')} + \pi_{(1,b,1-b')}\rho_{(0,b',1-b')}, \end{aligned} \tag{23}$$

where

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

Similarly, μ_1^d and μ_0^c are identified in the same manner as in (17):

$$\mu_1^d = \frac{(p^a + p^d)\delta_{(0,1)} - p^a\mu_1^a}{p^d}, \qquad \mu_0^c = \frac{(p^c + p^n)\delta_{(0,0)} - p^n\mu_0^n}{p^c}. \tag{24}$$

These equations are restated here for ease of reference and are exactly identical to those in Section 2.2.2. Then, under Assumptions 1, YD, and 5, we can identify ATE for compliers and defiers as ATE$(c) = \mu_1^c - \mu_0^c$ and ATE$(d) = \mu_1^d - \mu_0^d$, respectively.

Next, as in the result from Section 2.1, μ_0^a is represented under Assumption 3 (i) by (5), which yields

$$\mu_0^a = \mu_{\text{pre}}^a + \mu_0^n - \mu_{\text{pre}}^n. \tag{25}$$

This expression is also reproduced here for ease of reference. Since monotonicity is not assumed here, unlike in Section 2.1, μ^a_{pre} and μ^n_{pre} are not directly identified. To conduct identification analysis for μ^a_{pre} and μ^n_{pre}, we impose the following assumption.

Assumption 6. [Exclusion restriction on treatment before assignment for pre-treatment outcome] $\mathbb{E}[Y^{\text{pre}}|U = u, D^{\text{pre}}] = \mathbb{E}[Y^{\text{pre}}|U = u]$ *for each* $u \in \{c, a, n, d\}$.

Assumption 6 is a counterpart of Assumption 5 (i) with Y_{D_1} replaced by Y^{pre}. Therefore, Assumptions 5 (iii) and 6 together imply D^{pre} works as an instrumental variable for Y^{pre} through D_0. There are two possible types of D^{pre} that satisfy Assumptions 5 and 6. One is when D^{pre} represents a treatment similar to D. For example, if the goal is to estimate the effect of coupon usage for product A, the coupon usage status for a different product B can serve as D^{pre}, which is likely to be related to D but unlikely to affect pre-treatment sales of product A. The other case is when D^{pre} represents the same treatment as D, but is measured at a different time point, after Y^{pre} and before D. In this case, D^{pre} may be strongly associated with D, while Y^{pre} is unaffected by D^{pre}.

Under Assumptions 5 (ii), 5 (iii), and 6, μ^a_{pre} and μ^n_{pre} are identified as follows. Using the same argument for identification of μ^u_b, except that $\delta_{(z,d,d')}$ is replaced by $\delta^{\text{pre}}_{(z,d,d')}$, we have

$$\mu^u_{\text{pre}} = \frac{\delta^{\text{pre}}_{(1,b,b')}\rho_{(0,1-b',1-b')} - \delta^{\text{pre}}_{(1,b,1-b')}\rho_{(0,1-b',b')}}{\rho_{(0,1-b',1-b')} - \rho_{(0,1-b',b')}}. \tag{26}$$

We can identify μ^a_0 by (25) and thus ATT is also identified by (2). Finally, Assumption 3 (ii) guarantees identification of μ^n_1 as $\mu^n_1 = \mu^n_0 + \mu^a_1 - \mu^a_0$ so that ATE is identified by the expression in (2). Combining these results, identification of the causal objects in (2) is established as follows.

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

(i): *Under Assumptions 1, YD, and 5,* ATE(c) *and* ATE(d) *are identified.*

(ii): *Under Assumptions 1, YD, 3 (i), 5, and 6,* ATT *is identified.*

(iii): *Under Assumptions 1, YD, 3, 5, and 6,* ATE *is identified.*

Based on this theorem, we can estimate these causal objects by taking sample counterparts, and conduct statistical inference based on standard methods.

Remark 12. [Alternative assumptions] If defiers exist, then Assumption 3 can 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 (i.e., the group of $D_0 = 1$ including always-takers and defiers) may share common characteristics in that they expect to have worse outcomes if they do not receive the treatment. In this case, Assumption 3 (i) may be replaced with

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

In addition, the group that does not receive treatment even if they receive an external incentive (the group of $D_1 = 0$ 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

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

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)', any homogeneity assumption is not required to identify ATE.

Remark 13. [Partial identification] Similar to Remark 3, we can obtain partial identification of ATE under weaker assumptions. In this case, however, we allow for the existence of defiers. If we assume that μ_1^n lies in a known interval $[\mu_1^{n,\min}, \mu_1^{n,\max}]$, then the identified set for ATE is:

$$\text{ATE} \in \left[\sum_{u\in\{c,a,d\}} p^u \text{ATE}(u) + p^n(\mu_1^{n,\min} - \mu_0^n), \sum_{u\in\{c,a,d\}} p^u \text{ATE}(u) + p^n(\mu_1^{n,\max} - \mu_0^n) \right],$$

where $\text{ATE}(u) = \mathbb{E}[Y_1 - Y_0|U = u]$. Since ATE($c$), ATE($a$), and ATE($d$) are point identified, only ATE(n) is subject to partial identification. Therefore, a small proportion of never-takers (i.e., smaller value of p^n) leads to a relatively tight and informative identified set for ATE.

Remark 14. [Switched version of Assumption 5] Even if Assumption 5 is replaced by Assumption 5', each estimand is identified in the same manner. Under Assumption 5', $\mu_1^a, \mu_1^d, \mu_0^c, \mu_0^n$ and p^u's are identified in the same manner as in Remark 7:

$$\begin{aligned} \mu_{b'}^u &= \frac{\delta_{(0,b',b')}\rho_{(1,1-b,1-b')} - \delta_{(0,b',1-b')}\rho_{(1,1-b,b')}}{\rho_{(1,1-b,1-b')} - \rho_{(1,1-b,b')}}, \\ p^u &= \pi_{(0,b',b')}\rho_{(1,b,b')} + \pi_{(0,b,1-b')}\rho_{(1,b,1-b')}. \end{aligned} \tag{27}$$

Similarly, μ_1^c and μ_0^d are identified in the same manner as in (19) of Remark 7:

$$\mu_1^c = \frac{(p^c + p^a)\delta_{(1,1)} - p^a\mu_1^a}{p^c}, \qquad \mu_0^d = \frac{(p^d + p^n)\delta_{(1,0)} - p^n\mu_0^n}{p^d}. \tag{28}$$

Also μ_0^a and μ_1^n are identified in the same way as the above discussion. Finally, μ_{pre}^n is identified by

$$\mu_{\text{pre}}^u = \frac{\delta_{(0,b',b')}^{\text{pre}}\rho_{(1,1-b,1-b')} - \delta_{(0,b',1-b')}^{\text{pre}}\rho_{(1,1-b,b')}}{\rho_{(1,1-b,1-b')} - \rho_{(1,1-b,b')}}. \tag{29}$$

Remark 15. [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_{pre}^u for $u \in \{a, n\}$. For the identification under Assumption 5, the last four rows of Table 4 (right panel) suggest the additional two moment conditions followed.

$$p^a + p^d = \pi_{(0,1,1)} + \pi_{(0,1,0)}, \qquad p^c + p^n = \pi_{(0,0,1)} + \pi_{(0,0,0)}. \tag{30}$$

In addition, by introducing two more parameters $(\mu_{\text{pre}}^c, \mu_{\text{pre}}^d)$, we have further four additional restrictions under Assumption 5:

$$\mu_{\text{pre}}^a = \frac{(p^a + p^d)\delta_{(0,1)}^{\text{pre}} - p^d\mu_{\text{pre}}^d}{p^a}, \qquad \mu_{\text{pre}}^n = \frac{(p^c + p^n)\delta_{(0,0)}^{\text{pre}} - p^c\mu_{\text{pre}}^c}{p^n}, \tag{31}$$

where μ^c_{pre} and μ^d_{pre} are identified by (26) and $\delta^{\text{pre}}_{(z,d)} = \mathbb{E}[Y^{\text{pre}}|D_z = d]$. For the identification under Assumption 5', the first four rows of Table 4 (right) suggest the additional two moment conditions followed.

$$p^c + p^a = \pi_{(1,1,1)} + \pi_{(1,1,0)}, \qquad p^n + p^d = \pi_{(1,0,1)} + \pi_{(1,0,0)}. \tag{32}$$

In addition, by introducing two more parameters $(\mu^c_{\text{pre}}, \mu^d_{\text{pre}})$, we have further four additional restrictions under Assumption 5':

$$\mu^a_{\text{pre}} = \frac{(p^c + p^a)\delta^{\text{pre}}_{(1,1)} - p^c\mu^c_{\text{pre}}}{p^a}, \qquad \mu^n_{\text{pre}} = \frac{(p^n + p^d)\delta^{\text{pre}}_{(1,0)} - p^d\mu^d_{\text{pre}}}{p^n}, \tag{33}$$

where μ^c_{pre} and μ^d_{pre} are identified by (29) and $\delta^{\text{pre}}_{(z,d)} = \mathbb{E}[Y^{\text{pre}}|D_z = d]$. These additional six moment conditions can be incorporated by using the generalized method of moments. Equations (30) and (32) are reproduced here for ease of reference and and are exactly identical to (20) and (21) in Remark 8, respectively.

Remark 16. [Impose both Assumptions 5 and 5'] When both Assumptions 5 and 5' hold true, the 14 parameters, $\mu^c_1, \mu^c_0, \mu^d_1, \mu^d_0, \mu^a_1, \mu^n_0, \mu^u_{\text{pre}}$ for $u \in \{c, a, n, d\}$ and p^u for $u \in \{c, a, n, d\}$ are identified by (23), (26), (27), and (29). μ^a_0 and μ^n_1 are identified in the same way as the above discussion. Among these parameters, the eight parameters, $\mu^a_1, \mu^n_0, \mu^a_{\text{pre}}, \mu^n_{\text{pre}}$, and p^u for $u \in \{c, a, n, d\}$ are identified in the two ways. These additional moment conditions can be incorporated by using the GMM approach.

Remark 17. [Stable case] As discussed in Remark 10, Assumption 5 can be replaced with Assumption 4 (i.e., $D_0 = D^{\text{pre}}$ almost surely). Under Assumption 4, (23) becomes $\mu^u_b = \delta_{(1,b,b')}$ and $p^u = \pi_{(1,b,b')}$, and (26) becomes $\mu^u_{\text{pre}} = \delta^{\text{pre}}_{(1,b,b')}$, and (30) becomes $p^a + p^d = \pi_{(0,1,1)}$ and $p^c + p^n = \pi_{(0,0,0)}$ because of $\rho_{(0,1,1)} = \rho_{(0,0,0)} = 1$, $\rho_{(0,1,0)} = \rho_{(0,0,1)} = 0$, and $\pi_{(0,1,0)} = \pi_{(0,0,1)} = 0$. (24) and (31) remain the same.

Remark 18. [Alternative definitions of ATT] Under Assumptions 1 and YD (iii), we can express ATT_{D_0} and ATT_D as

$$\begin{aligned}
\text{ATT}_{D_0} &= \frac{p^a(\mu^a_1 - \mu^a_0) + p^d(\mu^d_1 - \mu^d_0)}{p^a + p^d}, \\
\text{ATT}_D &= \frac{p^a(\mu^a_1 - \mu^a_0) + \mathbb{P}(Z=1)p^c(\mu^c_1 - \mu^c_0) + \mathbb{P}(Z=0)p^d(\mu^d_1 - \mu^d_0)}{p^a + \mathbb{P}(Z=1)p^c + \mathbb{P}(Z=0)p^d}.
\end{aligned} \tag{34}$$

Because all components on the right-hand side of each expression are identified under the approach described above, Theorem 4 (ii) remains valid regardless of whether ATT is defined by ATT_{D_0} or ATT_D.

Remark 19. [Role of D^{pre}] The identification result in Theorem 4 holds true even under the more formal model defined by Assumption 1(i)'. This is because Assumptions 5 and 6 provide sufficient restrictions to eliminate heterogeneity induced by D^{pre}, ensuring the original identification strategies remain valid. See Appendix B for further details.

2.4. **Identification under ignorability condition.** In observational studies, it is often the case that the random assignment is violated. In this subsection we show that our identification argument can be extended to the case where certain ignorability condition is satisfied. Let $X \in \mathcal{X} \subset \mathbb{R}^q$ be a vector of q-dimensional covariates. We focus on the case of Section 2.3 here. The results for the remaining cases in Sections 2.1 and 2.2 are presented in Appendix A.

Consider the setup of Section 2.3. In observational studies, Assumption YD (iii) is replaced with the following.

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

This is a standard ignorability or unconfoundedness condition commonly imposed in the literature of causal inference with observational studies. Based on the discussion in Section 2.3, it is sufficient for identification of the causal estimands in (2) to identify

$$\begin{aligned}
\delta_{(z,d,d')} &= \mathbb{E}[Y_d | D_z = d, D^{\text{pre}} = d'], & \delta_{(z,d)} &= \mathbb{E}[Y_d | D_z = d], \\
\delta^{\text{pre}}_{(z,d,d')} &= \mathbb{E}[Y^{\text{pre}} | D_z = d, D^{\text{pre}} = d'], & \delta^{\text{pre}}_{(z,d)} &= \mathbb{E}[Y^{\text{pre}} | D_z = d], \\
\pi_{(z,d,d')} &= \mathbb{P}(D_z = d, D^{\text{pre}} = d'), & \rho_{(z,d,d')} &= \mathbb{P}(D_z = d | D^{\text{pre}} = d'),
\end{aligned} \tag{35}$$

for each $z \in \{0,1\}$ and $d, d' \in \{0,1\}$. To establish multiply robust representations of $\delta_{(z,d,d')}$, $\delta^{\text{pre}}_{(z,d,d')}$, $\delta_{(z,d)}$, $\delta^{\text{pre}}_{(z,d)}$, $\pi_{(z,d,d')}$, and $\rho_{(z,d,d')}$ under Assumption YD (iii)', we introduce parametric models

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

for each $z \in \{0,1\}$ and $d, d' \in \{0,1\}$, where α, β, γ, γ^{pre}, λ, λ^{pre}, and η are finite dimensional parameters. By using these parametric models, multiply robust representations of the population objects $\delta_{(z,d,d')}$, $\delta^{\text{pre}}_{(z,d,d')}$, $\delta_{(z,d)}$, $\delta^{\text{pre}}_{(z,d)}$, $\pi_{(z,d,d')}$, and $\rho_{(z,d,d')}$ are obtained as follows.

Theorem 5. *Under Assumptions 1 and YD (i)-(ii), and YD (iii)', it holds*

$$\begin{aligned}
\delta_{(z,d,d')} &= \mathbb{E}\left[\frac{\mathbb{I}\{Z = z\}}{e_z(X;\alpha)} \frac{\mathbb{I}\{D = d, D^{\text{pre}} = d'\}}{p_{(z,d,d')}(X;\beta)} Y\right] \\
&\quad - \mathbb{E}\left[\frac{\mathbb{I}\{Z = z\}\mathbb{I}\{D = d, D^{\text{pre}} = d'\} - e_z(X;\alpha) p_{(z,d,d')}(X;\beta)}{e_z(X;\alpha) p_{(z,d,d')}(X;\beta)} m_{(z,d,d')}(X;\gamma)\right],
\end{aligned} \tag{36}$$

$$\begin{aligned}
\delta_{(z,d)} &= \mathbb{E}\left[\frac{\mathbb{I}\{Z = z\}}{e_z(X;\alpha)} \frac{\mathbb{I}\{D = d\}}{\sum_{j \in \{0,1\}} p_{(z,d,j)}(X;\beta)} Y\right] \\
&\quad - \mathbb{E}\left[\frac{\mathbb{I}\{Z = z\}\mathbb{I}\{D = d\} - e_z(X;\alpha) \sum_{j \in \{0,1\}} p_{(z,d,j)}(X;\beta)}{e_z(X;\alpha) \sum_{j \in \{0,1\}} p_{(z,d,j)}(X;\beta)} m_{(z,d)}(X;\lambda)\right],
\end{aligned}$$

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

$$\delta^{\text{pre}}_{(z,d)} = \mathbb{E}\left[\frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)}\frac{\mathbb{I}\{D=d\}}{\sum_{j\in\{0,1\}} p_{(z,d,j)}(X;\beta)}Y^{\text{pre}}\right] - \mathbb{E}\left[\frac{\mathbb{I}\{Z=z\}\mathbb{I}\{D=d\} - e_z(X;\alpha)\sum_{j\in\{0,1\}} p_{(z,d,j)}(X;\beta)}{e_z(X;\alpha)\sum_{j\in\{0,1\}} p_{(z,d,j)}(X;\beta)} m^{\text{pre}}_{(z,d)}(X;\lambda^{\text{pre}})\right],$$

$$\pi_{(z,d,d')} = \mathbb{E}\left[\frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)}\mathbb{I}\{D=d, D^{\text{pre}}=d'\}\right] - \mathbb{E}\left[\frac{\mathbb{I}\{Z=z\} - e_z(X;\alpha)}{e_z(X;\alpha)} p_{(z,d,d')}(X;\beta)\right],$$

$$\rho_{(z,d,d')} = \mathbb{E}\left[\frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)}\frac{\mathbb{I}\{D^{\text{pre}}=d'\}}{\sum_{k\in\{0,1\}} p_{(z,k,d')}(X;\beta)}\mathbb{I}\{D=d\}\right] - \mathbb{E}\left[\frac{\mathbb{I}\{Z=z\}\mathbb{I}\{D^{\text{pre}}=d'\} - e_z(X;\alpha)\sum_{k\in\{0,1\}} p_{(z,k,d')}(X;\beta)}{e_z(X;\alpha)\sum_{k\in\{0,1\}} p_{(z,k,d')}(X;\beta)} p_{(z,d,d')}(X;\eta)\right],$$

for each $z \in \{0,1\}$ *and* $d, d' \in \{0,1\}$.

By taking the sample counterparts of these representations, we can construct multiply robust estimators for $\delta_{(z,d,d')}$, $\delta^{\text{pre}}_{(z,d,d')}$, $\delta_{(z,d)}$, $\delta^{\text{pre}}_{(z,d)}$, $\pi_{(z,d,d')}$, and $\rho_{(z,d,d')}$. Then the 16 parameters $\mu^u_1, \mu^u_0, \mu^u_{\text{pre}}, p^u$ for $u \in \{c,a,n,d\}$ are over-identified under Assumptions 1, YD (i)-(ii), YD (iii)', 3, 5, and 6 by the moment restrictions of (23)-(26) and (30)-(31). Just identification of the 16 parameters is guaranteed by the 16 moments of (23)-(26), and the four moments of (30)-(31) provide overidentifying restrictions.

We close this subsection by summarizing the multiply robust properties of the estimators based on Theorem 5 and the moment restrictions.

Proposition 1. *Suppose Assumptions 1, YD (i)-(ii), YD (iii)', 3, 5, and 6 hold true. Then*

- **(i):** $\delta_{(z,d,d')}$ *can be consistently estimated if either* $\{e_z(X;\alpha), p_{(z,d,d')}(X;\beta)\}$ *or* $m_{(z,d,d')}(X;\gamma)$ *is correctly specified,*
- **(ii):** $\delta_{(z,d)}$ *can be consistently estimated if either* $\{e_z(X;\alpha), p_{(z,d,d')}(X;\beta)\}$ *or* $m_{(z,d)}(X;\lambda)$ *is correctly specified,*
- **(iii):** $\delta^{\text{pre}}_{(z,d,d')}$ *can be consistently estimated if either* $\{e_z(X;\alpha), p_{(z,d,d')}(X;\beta)\}$ *or* $m^{\text{pre}}_{(z,d,d')}(X;\gamma^{\text{pre}})$ *is correctly specified,*
- **(iv):** $\delta^{\text{pre}}_{(z,d)}$ *can be consistently estimated if either* $\{e_z(X;\alpha), p_{(z,d,d')}(X;\beta)\}$ *or* $m^{\text{pre}}_{(z,d)}(X;\lambda^{\text{pre}})$ *is correctly specified,*
- **(v):** $\pi_{(z,d,d')}$ *can be consistently estimated if either* $e_z(X;\alpha)$ *or* $p_{(z,d,d')}(X;\beta)$ *is correctly specified,*
- **(vi):** $\rho_{(z,d,d')}$ *can be consistently estimated if either* $\{e_z(X;\alpha), p_{(z,d,d')}(X;\beta)\}$ *or* $p_{(z,d,d')}(X;\eta)$ *is correctly specified,*

(vii): ATE(c) *and* ATE(d) *can be consistently estimated if either* $\{e_z(X;\alpha), p_{(z,d,d')}(X;\beta)\}$, $\{m_{(z,d,d')}(X;\gamma), m_{(z,d)}(X;\lambda), p_{(z,d,d')}(X;\beta), p_{(z,d,d')}(X;\eta)\}$, *or* $\{m_{(z,d,d')}(X;\gamma), m_{(z,d)}(X;\lambda), e_z(X;\alpha), p_{(z,d,d')}(X;\eta)\}$ *is correctly specified,*

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

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

Remark 20. [Switched version of Assumption 5] Under Assumption 5', the moment restrictions above are replaced with (27)-(29) and (32)-(33). Just identification of the 16 parameters is guaranteed by the 16 moments of (25) and (27)-(29), and the four moments of (32)-(33) provide overidentifying restrictions.

Remark 21. [Impose both Assumptions 5 and 5'] Under both Assumptions 5 and 5', the moment restrictions above are replaced with (23), (25), (26), (27), and (29). Among the 16 parameters, $\mu_1^u, \mu_0^u, \mu_{\text{pre}}^u, p^u$ for $u \in \{c,a,n,d\}$, the eight parameters, μ_1^a, μ_0^n,$\mu_{\text{pre}}^a, \mu_{\text{pre}}^n$, and p^u for $u \in \{c,a,n,d\}$ are identified in the two ways. $\delta_{(z,d)}$ and $\delta^{\text{pre}}_{(z,d)}$ are not used in this identification argument.

Remark 22. [Special case] Under Assumption 4 instead of Assumption 5, the moment restrictions are obtained from the discussion in Remark 17, and the moments $\rho_{(z,d,d')}$ are not used in this identification argument.

Remark 23. [Alternative definitions of ATT] Under Assumptions 1 and YD (iii)', we can express ATT_{D_0} and ATT_D as

$$\begin{aligned}
\text{ATT}_{D_0} &= \frac{p^a(\mu_1^a - \mu_0^a) + p^d(\mu_1^d - \mu_0^d)}{p^a + p^d}, \\
\text{ATT}_D &= \frac{p^a(\mu_1^a - \mu_0^a) + \mathbb{P}(Z=1)p^{c|Z=1}(\mu_1^{c|Z=1} - \mu_0^{c|Z=1}) + \mathbb{P}(Z=0)p^{d|Z=0}(\mu_1^{d|Z=0} - \mu_0^{d|Z=0})}{p^a + \mathbb{P}(Z=1)p^{c|Z=1} + \mathbb{P}(Z=0)p^{d|Z=0}}.
\end{aligned}$$

Since all components on the right-hand side of ATT_{D_0} are identified under the approach described above, Proposition 1 (vi) remains valid even when ATT is defined by ATT_{D_0}. Moreover, while ATT_D contains parameters conditional on Z, extending Assumptions 5 (i)-(iii) and 6 to be conditional on covariates (e.g., replace Assumption 5(i) with $\mathbb{E}[Y_{D_1}|U=u, D^{\text{pre}}, X=x] = \mathbb{E}[Y_{D_1}|U=u, X=x]$) allows for identification of $\mu_1^u(x)$ for $u \in \{c,a,d\}$, $\mu_0^u(x)$ for $u \in \{c,n,d\}$, $\mu_{\text{pre}}^u(x)$ for $u \in \{c,a,n,d\}$, and $p^u(x)$ for $u \in \{c,a,n,d\}$, where $\mu_d^u(x) = \mathbb{E}[Y_d|U=u, X=x]$, $\mu_{\text{pre}}^u(x) = \mathbb{E}[Y^{\text{pre}}|U=u, X=x]$, and $p^u(x) = \mathbb{P}(U=u|X=x)$. The parameters on the right-hand side of ATT_D with the exception of μ_0^a are then obtained by integrating x out as

$$\begin{aligned}
\mu_d^{c|Z=1} &= \int \mu_d^c(x) f(x|Z=1)dx \text{ for } d \in \{0,1\}, \quad p^{c|Z=1} = \int p^c(x) f(x|Z=1)dx, \\
\mu_d^{d|Z=0} &= \int \mu_d^d(x) f(x|Z=0)dx \text{ for } d \in \{0,1\}, \quad p^{d|Z=0} = \int p^d(x) f(x|Z=0)dx, \\
\mu_1^a &= \int \mu_1^a(x) f(x)dx, \quad p^a = \int p^a(x) f(x)dx,
\end{aligned}$$

where $f(x|Z=z)$ and $f(x)$ are the conditional density of $X|Z=z$ and marginal density of X, respectively. Finally, the remaining parameter μ_0^a is identified from (25). The inputs required for (25) (i.e., $\mu_{\text{pre}}^a, \mu_{\text{pre}}^n$, and μ_0^n) are identified by integrating x out for their respective conditional counterparts $\mu_{\text{pre}}^a(x)$, $\mu_{\text{pre}}^n(x)$, and $\mu_0^n(x)$ in the manner shown above.

3. Estimation

In this section, 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. The results for the remaining cases in Section 2.1 and 2.2 are presented in Appendix A.

First, we consider estimation of ATE based on Theorem 4 (iii). Let $\hat{\delta}_{(z,d,d')}$, $\hat{\delta}^{\text{pre}}_{(z,d,d')}$, $\hat{\delta}_{(z,d)}$, $\hat{\delta}^{\text{pre}}_{(z,d)}$, $\hat{\pi}_{(z,d,d')}$, and $\hat{\rho}_{(z,d,d')}$ be the empirical (conditional) moments of $\delta_{(z,d,d')}$, $\delta^{\text{pre}}_{(z,d,d')}$, $\delta_{(z,d)}$, $\delta^{\text{pre}}_{(z,d)}$, $\pi_{(z,d,d')}$, $\rho_{(z,d,d')}$, respectively, and $\hat{\zeta}$ and ζ be their vectorizations. Also let θ be a 14-dimensional vector given by (μ_1^u, μ_0^u, p^u) for $u \in \{c,a,n,d\}$ and μ_{pre}^u for $u \in \{a,n\}$, which provides an identification formula for ATE as

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

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

$$(\hat{\theta}, \hat{\omega}) = \arg\min_{\theta,\omega} g(\hat{\zeta}, \theta, \omega)' \Psi g(\hat{\zeta}, \theta, \omega), \tag{37}$$

where the vector of moment conditions $g(\zeta, \theta, \omega) = 0$ is obtained by stacking the equations (23)-(26) and $\omega = \text{ATE}(\theta)$ (and also (30)-(31)) under Assumptions 1, YD, 3, 5, and 6. 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,\gamma^{\text{pre}},\lambda,\lambda^{\text{pre}},\eta) \\
= & \left[\begin{array}{c}
\left\{\begin{array}{c} \delta_{(z,d,d')} - \frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)}\frac{\mathbb{I}\{D=d,D^{\text{pre}}=d'\}}{p_{(z,d,d')}(X;\beta)}Y \\ +\frac{\mathbb{I}\{Z=z\}\mathbb{I}\{D=d,D^{\text{pre}}=d'\}-e_z(X;\alpha)p_{(z,d,d')}(X;\beta)}{e_z(X;\alpha)p_{(z,d,d')}(X;\beta)}m_{(z,d,d')}(X;\gamma) \end{array}\right\}_{(z,d,d')} \\
\left\{\begin{array}{c} \delta_{(z,d)} - \frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)}\frac{\mathbb{I}\{D=d\}}{\sum_{j\in(1,0)}p_{(z,d,j)}(X;\beta)}Y \\ +\frac{\mathbb{I}\{Z=z\}\mathbb{I}\{D=d\}-e_z(X;\alpha)\sum_{j\in(1,0)}p_{(z,d,j)}(X;\beta)}{e_z(X;\alpha)\sum_{j\in(1,0)}p_{(z,d,j)}(X;\beta)}m_{(z,d)}(X;\lambda) \end{array}\right\}_{(z,d)} \\
\left\{\begin{array}{c} \delta^{\text{pre}}_{(z,d,d')} - \frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)}\frac{\mathbb{I}\{D=d,D^{\text{pre}}=d'\}}{p_{(z,d,d')}(X;\beta)}Y^{\text{pre}} \\ +\frac{\mathbb{I}\{Z=z\}\mathbb{I}\{D=d,D^{\text{pre}}=d'\}-e_z(X;\alpha)p_{(z,d,d')}(X;\beta)}{e_z(X;\alpha)p_{(z,d,d')}(X;\beta)}m^{\text{pre}}_{(z,d,d')}(X;\gamma^{\text{pre}}) \end{array}\right\}_{(z,d,d')} \\
\left\{\begin{array}{c} \delta^{\text{pre}}_{(z,d)} - \frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)}\frac{\mathbb{I}\{D=d\}}{\sum_{j\in(1,0)}p_{(z,d,j)}(X;\beta)}Y^{\text{pre}} \\ +\frac{\mathbb{I}\{Z=z\}\mathbb{I}\{D=d\}-e_z(X;\alpha)\sum_{j\in(1,0)}p_{(z,d,j)}(X;\beta)}{e_z(X;\alpha)\sum_{j\in(1,0)}p_{(z,d,j)}(X;\beta)}m^{\text{pre}}_{(z,d)}(X;\lambda^{\text{pre}}) \end{array}\right\}_{(z,d)} \\
\left\{\pi_{(z,d,d')} - \frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)}\mathbb{I}\{D=d,D^{\text{pre}}=d'\} + \frac{\mathbb{I}\{Z=z\}-e_z(X;\alpha)}{e_z(X;\alpha)}p_{(z,d,d')}(X;\beta)\right\}_{(z,d,d')} \\
\left\{\begin{array}{c} \rho_{(z,d,d')} - \frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)}\frac{\mathbb{I}\{D^{\text{pre}}=d'\}}{\sum_{k\in(1,0)}p_{(z,k,d')}(X;\beta)}\mathbb{I}\{D=d\} \\ +\frac{\mathbb{I}\{Z=z\}\mathbb{I}\{D^{\text{pre}}=d'\}-e_z(X;\alpha)\sum_{k\in(1,0)}p_{(z,k,d')}(X;\beta)}{e_z(X;\alpha)\sum_{k\in(1,0)}p_{(z,k,d')}(X;\beta)}p_{(z,d,d')}(X;\eta) \end{array}\right\}_{(z,d,d')} \\
\xi_1(W,\alpha) \\
\xi_2(W,\beta) \\
\xi_3(W,\gamma) \\
\xi_4(W,\gamma^{\text{pre}}) \\
\xi_5(W,\lambda) \\
\xi_6(W,\lambda^{\text{pre}}) \\
\xi_7(W,\eta)
\end{array}\right],
\end{aligned}
$$

where W means the whole observables, ξ_1, ξ_2, ξ_3, ξ_4, ξ_5, ξ_6 and ξ_7 are estimating equations for the parameters α, β, γ, γ^{pre}, λ, λ^{pre}, and η, respectively. Combining this with the moment conditions $g(\zeta,\theta,\vartheta)=0$, the GMM estimator of ATE is obtained as $\tilde{\omega}$ for

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

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.

Remark 24. [Switched version of Assumption 5] Under Assumptions 1, YD, 3, 5', and 6, the vector of moment conditions $g(\zeta,\theta,\omega)=0$ is obtained by stacking the equations (25), (27)-(29) and (32)-(33).

Remark 25. [Impose both Assumptions 5 and 5'] Under Assumptions 1, YD, 3, 5, 5', and 6, the vector of moment conditions $g(\zeta, \theta, \omega) = 0$ is obtained by stacking the equations (23), (25), (26), (27), and (29).

Remark 26. [Special case] Under Assumption 4 instead of Assumption 5, the vector of moment conditions $g(\zeta, \theta, \omega) = 0$ is obtained by stacking the equations following the discussion in Remark 17.

4. Extensions

In this section, we present three alternative experimental designs to identify and estimate causal objects using auxiliary observations: (I) the case where the treatment status and outcome are observable at the baseline survey (Section 4.1), (II) the case with a two-regime randomization (Section 4.2), and (III) the case with a post treatment variable (Section 4.3).

4.1. Identification with baseline measure of treatment and outcome.

In Section 2.3, we derive identification results under the assumptions that are plausible when D^{pre} represents a different but similar treatment from D, or represents the same treatment but is measured at a different time point from Y^{pre}. This section considers identification in the case where D^{pre} represents the same treatment as D and is measured at the same time as Y^{pre} (e.g., at baseline). While this setting may be more feasible in practice due to the ease of data collection, it makes satisfying the parallel trend assumption of Assumption 3 (i) challenging.

As in Remark 17, the parameters $\mu_1^a, \mu_1^c, \mu_0^d, \mu_0^n$, and p^u are identified under Assumptions 1, YD, and 4 based solely on the observables $(Z, D, Y, D^{\text{pre}})$. On the other hand, Assumption 3 (i), which is required to identify μ_0^a, implicitly assumes the absence of treatment during the observation period of Y^{pre}. This implicit assumption is violated by the relationship between D^{pre} and Y^{pre} in the setting of this section. Therefore, we introduce alternative assumptions to identify μ_0^a.

Assumption YD-a. *The observable Y^{pre} satisfies*

$$Y^{\text{pre}} = D^{\text{pre}} Y_1^{\text{pre}} + (1 - D^{\text{pre}}) Y_0^{\text{pre}},$$

where $Y_{d'}^{\text{pre}}$ is the potential outcome realized only when $D^{\text{pre}} = d'$.

This is a natural assumption in the present setting, where D^{pre} and Y^{pre} are observed at the same time point. The relationships of the observables and principal strata variable are summarized as in Table 5 (left).

As suggested by Assumption YD-a and Table 5 (left), unlike in Remark 17, Y^{pre} is observed after being potentially influenced by D^{pre}. Therefore, identification of μ_0^a based on Assumption 3 is no longer valid in this setting. To address this issue, we introduce Assumption 3a as a replacement for Assumption 3.

Assumption 3a.

- **(i):** *[Parallel trend of strata with $D_0 = 1$]* $\mathbb{E}[Y_0 - Y_1^{\text{pre}} | U = a] = \mathbb{E}[Y_0 - Y_1^{\text{pre}} | U = d]$.
- **(ii):** *[Parallel trend of strata with $D_0 = 0$]* $\mathbb{E}[Y_1 - Y_0^{\text{pre}} | U = c] = \mathbb{E}[Y_1 - Y_0^{\text{pre}} | U = n]$.

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

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

TABLE 5. Case of baseline measure (left) and Two-regime design (right)

Now Assumption 3a (i) implies

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

which can be written as

$$\mu_0^a = \mathbb{E}[Y_1^{\text{pre}}|U = a] + \mu_0^d - \mathbb{E}[Y_1^{\text{pre}}|U = d]. \tag{38}$$

By using (2) and (38), we can identify ATT. Similarly, Assumption 3a (ii) implies

$$\mu_1^n = \mu_1^c - \mathbb{E}[Y_0^{\text{pre}}|U = c] + \mathbb{E}[Y_0^{\text{pre}}|U = n]. \tag{39}$$

Based on (38) and/or (39), 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, YD, YD-a, and 4,* ATE(c) *is identified.*
- **(ii):** *Under Assumptions 1, YD, YD-a, 4, and 3a (i),* ATT *is identified.*
- **(iii):** *Suppose Assumptions 1, YD, YD-a, and 4 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.*

When Assumption 4 is extended to the unstable case (i.e., $D_0 \neq D^{\text{pre}}$), the identification argument for μ_0^a and μ_1^n becomes problematic due to lack of identifiable structure for $\mathbb{E}[Y_d^{\text{pre}}|U = u]$.

4.2. **Two-regime design.** In this subsection, we consider a two-regime setting, where we do not need to observe Y^{pre}. First, subjects are randomly assigned to one of two regimes $R \in \{1, 0\}$. In the group with $R = 1$, we observe $(Y, D, Z, D^{\text{pre}})$. In the group with $R = 0$, $D = 0$ is forced so that we observe $Y = Y_0$. 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 (right panel). In this setup, we impose the following assumptions.

Assumption 3b.

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

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

TABLE 6. Case of treatment after main observations D^{post} instead of D^{pre}

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

Under Assumptions 1, D, and 5, identification of μ_1^c, μ_0^c, μ_1^d, μ_0^d, μ_1^a, μ_0^n, and p^u for $u \in \{c, a, n, d\}$ in (23) and (24) is achieved in the same way as in Section 2.2.2. 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 (right panel)), Assumption 3b implies

$$\mu_0^a = \frac{\mathbb{E}[Y_0] - (p^c\mu_0^c + p^n\mu_0^n + p^d\mu_0^d)}{p^a} = \frac{\mathbb{E}[Y|R = 0] - (p^c\mu_0^c + p^n\mu_0^n + p^d\mu_0^d)}{p^a}. \tag{40}$$

By the same argument in Section 2.3, 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, D, and 5,* ATE(c) *and* ATE(d) *are identified.*
- **(ii):** *Under Assumptions 1, D, 3b, and 5,* ATT *is identified.*
- **(iii):** *Under Assumptions 1, D, 3b, 3 (ii), and 5,* ATE *is identified.*

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

4.3. **Treatment after main observations.** In this subsection, we assume that in addition to the main observations (Z, D, Y) and Y^{pre}, the researcher observes:

$$D^{\text{post}} \in \{0, 1\} \quad : \quad \text{treatment indicator to be observed at the time after the main observations.} \tag{41}$$

The relationships of the observables and principal strata variable are summarized in Table 6.

In this case, we impose the following assumptions.

Assumption YD-c. *Assumption YD holds true with replacement of "D^{post}" with "D^{pre}".*

Assumption 5c. *Assumption 5 holds true with replacement of "D^{post}" with "D^{pre}".*

Assumption 6c. *Assumption 6 holds true with replacement of "D^{post}" with "D^{pre}".*

An example of this setup is when, following an encouragement experiment, the treatment status is subsequently observed to serve as D^{post} during a period in which the encouragement is absent. Also the status of a similar subsequently observed treatment may serve as D^{post}. The parameters μ_1^c, μ_0^c, μ_1^d, μ_0^d, μ_1^a, μ_0^n, and p^u for $u \in \{c, a, n, d\}$ are identified under Assumption 5c. By the same argument in 2.3, Assumptions 3 and 6 guarantee identification of μ_0^a and μ_1^n. Combining these results, we obtain the following identification results.

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

(i): *Under Assumptions 1, YD-c, and 5c,* ATE(c) *and* ATE(d) *are identified.*

(ii): *Under Assumptions 1, YD-c, 3 (i), 5c, and 6c,* ATT *is identified.*

(iii): *Under Assumptions 1, YD-c, 3, 5c, and 6c,* ATE *is identified.*

5. Empirical illustrations

This section presents empirical illustrations for each of the following cases: with Y^{pre} (Section 5.1), with D^{pre} (Section 5.2), and with both Y^{pre} and D^{pre} (Section 5.3). For the case with Y^{pre}, we revisit three important empirical studies that use randomized encouragement designs: Thornton (2008), Gerber et al. (2009), and Beam (2016). For the case with D^{pre} and with both Y^{pre} and D^{pre}, we revisit other influential empirical studies by Finkelstein et al. (2012) and Taubman et al. (2014) both of which examine the effects of enrolling in Medicaid using the data set from the Oregon Health Insurance Experiment (OHIE). Furthermore, Section 5.4 presents an application of the case with both Y^{pre} and D^{pre} in marketing.

5.1. **Randomized encouragement design with Y^{pre}.** We illustrate the identification method of Theorem 1 in Section 2.1 by revisiting three important empirical studies in the literature. Thornton (2008), Gerber et al. (2009), and Beam (2016) used randomized encouragement designs to investigate the causal effects of knowing one's HIV status on contraceptive behavior, the effects of a newspaper subscription on political attitudes, and the effects of job fair participation on the intention to work abroad, respectively. Since enforcing treatment was difficult in these studies, they employed encouragement designs with incentives. Thornton (2008) used two-stage least squares estimation, while Beam (2016) adopted the LATE estimator. Gerber et al. (2009) reported intention-to-treat (ITT) estimates; however, we find that applying the LATE estimator to their data also fails to produce statistically significant effects. Using the identification method proposed in Section 2.1, we revisit their data to estimate ATT and ATE under Assumptions 1, 2, Y, and 3. The outcomes variable from each baseline survey is employed as Y^{pre}. Among the outcomes we analyze, the "Voted" outcome from Gerber et al. (2009) can be used for an assessment of the parallel trend assumption (Assumption 3(i)). This is because its baseline survey provides a history of past voting behavior. A visualization of this pre-treatment data supports our parallel trend assumption (see Appendix D for details). Causal objects are estimated by taking the sample counterparts. Standard errors and p-values are calculated based on 2000 bootstrap resamples in all analyses. The results are shown in Table 7.

Across all three studies we revisit, the fractions of compliers are estimated to be less than half, and ATE(c) is not statistically significant. These findings are consistent with the original studies. In the re-analysis of Thornton (2008), the estimated ATT and ATE for the "Purchase Condom" outcome are negative and statistically significant, with magnitudes larger than the insignificant estimate of ATE(c). This is driven by ATE(a), the effect on those who would have learned their results even without an incentive, which is also estimated to be significantly negative. Given that the sample contains far more HIV-negative subjects than HIV-positive subjects, this may suggest that learning one's HIV-negative status could discourage condom purchases. For the "Having sex" outcome in Thornton (2008), as well as for all outcomes from Gerber et al. (2009) and Beam (2016), the estimates for ATT and ATE are all statistically insignificant, suggesting that the interventions in those studies may have had no discernible effect on the overall population in addition to the complier subpopulation.

	Thornton (2008)		Gerber et al. (2009)	Beam (2016)	
	Purchase condom	Having sex	Voted in 2005	Plan to abroad	Passport
p^c	0.425 ***	0.438 ***	0.243 ***	0.337 ***	0.337 ***
	(0.033)	(0.027)	(0.029)	(0.043)	(0.043)
p^a	0.390 ***	0.376 ***	0.225 ***	0.136 ***	0.136 ***
	(0.032)	(0.027)	(0.020)	(0.014)	(0.014)
p^n	0.185 ***	0.186 ***	0.532 ***	0.527 ***	0.527 ***
	(0.012)	(0.011)	(0.021)	(0.041)	(0.041)
ATE(c)	−0.022	0.014	0.015	−0.061	−0.012
	(0.061)	(0.075)	(0.111)	(0.076)	(0.058)
ATE(a)	−0.127 *	0.006	0.122	−0.087	0.031
	(0.058)	(0.065)	(0.068)	(0.080)	(0.042)
ATT	−0.072 *	0.010	0.067	−0.068	0.000
	(0.035)	(0.035)	(0.056)	(0.056)	(0.037)
ATE	−0.083 **	0.009	0.096	−0.078	0.016
	(0.035)	(0.034)	(0.051)	(0.056)	(0.027)
n	1,008	1,328	1,079	865	865

Note: $ATE(a)$ is defined as $\mu_1^a - \mu_0^a$.

Table 7. Estimates and standard errors. Standard errors in parentheses. $^*p < 0.05, ^{**}p < 0.01, ^{***}p < 0.001$.

5.2. **Oregon Health Insurance Experiment with D^{pre}.** This subsection revisits Finkelstein et al. (2012) to illustrate the identification result in Theorem 10, which is an analogue to Theorem 3 in Section 2.2.2 under the ignorability condition. Using the data set from OHIE, Finkelstein et al. (2012) estimated the effects of enrolling in a health insurance known as Medicaid by LATE, focusing on outcomes related to health care utilization, financial strain, and health. However, the population in the experiment may structurally include defiers because only about 30% of those who won the lottery actually enrolled, and there was another Medicaid program where people who had lost the lottery could also be enrolled (Finkelstein et al., 2012). We revisit the data set to estimate ATE(c) using the pre-treatment variable, D^{pre} , under Assumptions 1, D (i), D (ii)', and 5. Note that this method does not require the monotonicity assumption. TANF (another public assistance program) enrollment status in year prior to OHIE, which is available in the administrative data from the data set, is used as D^{pre}. For this re-analysis, we focus on the 24 outcomes obtained from their mail survey. The analysis uses a total sample of 23,777, excluding observations with missing values for each outcome. The sample size for each specific outcome is reported in the fourth column of Table 8 (for the outcomes related to health care utilization and financial strain) and 9 (for the outcomes related to health). As Assuming ignorability given

	Monotonicity ATE(c)	Without monotonicity ATE(c)	ATE(d)	n
Health care utilization:				
Prescription drugs currently	0.092 *** (0.029)	0.038 (0.046)	−0.151 (0.180)	18,332
Outpatient visit	0.210 *** (0.026)	0.163 *** (0.038)	−0.016 (0.157)	23,528
ER visits	0.026 (0.024)	−0.057 (0.049)	−0.364 * (0.223)	23,550
Inpatient hospital admissions	0.008 (0.014)	0.032 (0.019)	0.125 (0.080)	23,609
Blood cholesterol checked	0.111 *** (0.026)	0.116 * (0.050)	0.140 (0.228)	23,426
Blood tested for high blood sugar/diabetes	0.089 *** (0.027)	0.104 * (0.045)	0.166 (0.198)	23,410
Financial strain:				
Any out of pocket medical expenses	−0.197 *** (0.026)	−0.126 ** (0.051)	0.149 (0.225)	23,462
Owe money for medical expenses currently	−0.180 *** (0.027)	−0.145 ** (0.047)	−0.006 (0.212)	23,487
Borrowed money or skipped other bills to pay medical bills	−0.151 *** (0.026)	−0.085 * (0.044)	0.167 (0.190)	23,446
Refused treatment because of medical debt	−0.036 * (0.014)	−0.033 (0.029)	−0.020 (0.128)	22,605

TABLE 8. Estimates and standard errors. Standard errors in parentheses. $^{*}p < 0.05, ^{**}p < 0.01, ^{***}p < 0.001$.

household number who listed in that lottery, as in Finkelstein et al. (2012), the causal objects are estimated by inverse probability weighting. Standard errors and p-values are calculated based on 800 bootstrap resamples in the analyses. Although the principal strata probabilities differ slightly for each outcome due to different sample compositions, we find no significant differences among them. The average probabilities across outcomes for compliers, always-takers, never-takers, and defiers are (0.37, 0.06, 0.50, and 0.08), respectively, which suggests the presence of defiers. Estimated ATE(c) and ATE(d) are shown in the second and third columns in Table 8 and Table 9. For comparison, the first column of Table 1 shows the ATE(c) estimates assuming monotonicity, which are consistent with to the result from Finkelstein et al. (2012).

Under monotonicity condition, statistically significant estimates of ATE(c) were reported for 21 out of 24 outcomes. However, when we account for the existence of defiers and estimate ATE(c) using our method, five of these outcomes were no longer statistically significant. Notably, no new outcome becomes significant in our analysis. Furthermore, among the 16 outcomes that remain significant, 11 outcomes exhibit weaker ATE(c) estimates compared to those of LATE. These results suggest that failing to account for the existence of defiers may lead to overestimation of the causal effects.

	Monotonicity ATE(c)	Without monotonicity ATE(c)	ATE(d)	n
Health:				
Self-reported health good/very good/excellent	0.134 *** (0.027)	0.069 (0.050)	−0.181 (0.226)	23,397
Self-reported health not poor	0.101 *** (0.018)	0.072 *** (0.029)	−0.041 (0.125)	23,397
Health about the same or gotten better	0.113 *** (0.024)	0.073 * (0.037)	−0.081 (0.163)	23,443
# of days physical health good, past 30 days	1.631 ** (0.592)	0.313 (1.007)	−4.703 (4.548)	21,415
# of days physical or mental health did not impair usual activity, past 30 days	1.322 * (0.574)	0.725 (0.906)	−1.569 (4.029)	21,915
# of days mental health good, past 30 days	2.173 *** (0.641)	2.831 *** (1.052)	5.369 (4.464)	21,632
Did not screen positive for depression, last two weeks	0.080 *** (0.027)	0.117 ** (0.050)	0.258 (0.223)	23,406
Health (Mechanisms):				
Have usual place of clinic-based care	0.340 *** (0.028)	0.265 *** (0.040)	−0.035 (0.177)	21,577
Have personal doctor	0.277 *** (0.028)	0.257 *** (0.046)	0.180 (0.202)	23,537
Got all needed medical care, last six months	0.238 *** (0.024)	0.198 *** (0.043)	0.049 (0.193)	22,940
Got all needed drugs, last six months	0.194 *** (0.020)	0.254 *** (0.049)	0.481 ** (0.228)	22,860
Didn't use ER for nonemergency, last six months	0.003 (0.015)	0.000 (0.025)	−0.009 (0.112)	23,566
Quality of care received last six months good/very good/excellent	0.143 *** (0.027)	0.101 ** (0.042)	−0.044 (0.163)	16,336
Very happy or pretty happy	0.192 *** (0.026)	0.157 *** (0.042)	0.021 (0.180)	23,450

TABLE 9. Estimates and standard errors. Standard errors in parentheses. $^{*}p < 0.05, ^{**}p < 0.01, ^{***}p < 0.001$.

5.3. **Oregon Health Insurance Experiment with Y^{pre} and D^{pre}.** This subsection revisits Taubman et al. (2014) to illustrate Theorem 5 in Section 2.4. Using the data set from OHIE as in Finkelstein et al. (2012), Taubman et al. (2014) estimated the effects of enrolling in Medicaid by LATE, focusing on outcomes of emergency department use. We revisit the data set using Y^{pre} and D^{pre} to estimate ATE(u) for $u \in (c, a, n, d)$, ATT, and ATE without the monotonicity assumption (i.e., in the presence of defiers). The emergency department use in the year prior to OHIE and TANF enrollment status in year prior to OHIE is used as Y^{pre} and D^{pre}, respectively. We suppose Assumptions 1, YD (i)-(ii), YD (iii)', 3 (i), 3 (ii)', 5, and 6. Following Taubman et al. (2014), we analyze two outcomes related to outpatient visits: a binary indicator for "Percent with any visits" and a count measure for the "Number of visits". The total sample size is 24,646. For the analysis of the "Number of visits" outcome, we exclude observations with missing values,

	Monotonicity	Without monotonicity					
		ATE(c)	ATE(a)	ATE(n)	ATE(d)	ATT	ATE
Percent with any visits	0.065 **	0.045	0.174 *	0.147	-0.011	0.066 *	0.101
	(0.025)	(0.037)	(0.076)	(0.092)	(0.128)	(0.026)	(0.050)
Number of visits	0.288 *	0.280 *	0.518	-0.355	0.201	0.319 ***	-0.038
	(0.111)	(0.125)	(0.419)	(0.432)	(0.406)	(0.087)	(0.250)

Note: $ATE(u)$ is defined as $\mu_1^u - \mu_0^u$ for each stratum $u \in \{c, a, n, d\}$.

TABLE 10. Estimates and standard errors. Standard errors in parentheses. $^{*}p < 0.05, ^{**}p < 0.01, ^{***}p < 0.001$.

resulting in a sample size of 24,615. Assuming ignorability given household number who listed in that lottery, as in Taubman et al. (2014), the causal objects are estimated by the multiply robust estimators. Standard errors and p-values are calculated based on 800 bootstrap resamples in the analyses. The estimated probabilities for complier, always-taker, never-taker, and defier are (0.33, 0.07, 0.51, 0.09) for both outcomes, suggesting the presence of defiers. Estimated causal effects are shown in Table 10. For comparison, the first column of Table 1 shows the ATE(c) estimates assuming monotonicity, which are consistent with to the result from Taubman et al. (2014).[2]

Regarding ATE(c), relaxing the monotonicity assumption lowers the point estimate for the "Percent with any visits" outcome, causing it to lose statistical significance. This suggests that ATE(c) estimated under the monotonicity assumption in the original study may be an overestimate. In contrast, ATT is estimated to be significantly positive for both outcomes. Furthermore, the magnitude of ATT estimates slightly exceeds that of ATE(c) estimated under monotonicity. Therefore, when viewed from the perspective of ATT, our results reinforce the conclusion that Medicaid enrollment increases emergency department utilization. Finally, ATE is not statistically significant for either outcome. This result suggests that the inclusion of never-takers and defiers dilutes the increase in ED utilization from Medicaid enrollment.

5.4. **Application in marketing using Y^{pre} and D^{pre}.** Using the proposed methods in Section 2.3, we evaluate the cannibalization effect associated with introduction of a new product. It is difficult to measure the effect because of changes in market structures. We 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 be the coupon assignment and D be whether or not each subject purchased the new product in the week following the coupon assignment. As outcomes of interest, we consider four spending measures per subject from this manufacturer, defined by category scope (beer only vs. beer and RTD) and by whether the new product is included or excluded. These outcomes are measured over a one-week period following the coupon assignment. For each outcome, let Y^{pre} be measured for one week in March 2023,

[2]Our estimate for the "Number of visits" outcome is smaller than the estimate reported in Taubman et al. (2014). This is likely because variables such as the number of ED visits have been censored in the publicly available data to ensure de-identification.

before the new product is released. Let D^{pre} be whether or not each subject purchased the new product during the week in May before the experiment. We suppose Assumptions 1, YD, 3, 5, and 6. Causal objects were estimated by the taking sample counterparts. Standard errors and p-values were calculated based on 2000 bootstrap resamples in the analyses. The results are presented in Table 11.

In all estimates, the total sales of the category including the new product increase significantly, and the change in the total sales of the category excluding the new product is not significant. These results indicate that there is no cannibalization within the category and that the entry of the new product increases the total sales of the category. Comparing the estimated values of LATE, ATT, and ATE shows that LATE underestimates increased sales. In addition, the estimated probabilities for complier, always-taker, never-taker, and defier are (0.015, 0.001, 0.971, 0.013). The large proportion of never-takers indicates that there are few purchasers of the new product, no matter whether consumers have coupons or not. Since getting consumers to buy this new product may 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. Furthermore, the fact that ATE is larger than ATE(c) suggests that the treatment effect can be larger for never-takers than for compliers. Since never-takers might be induced to purchase with a higher incentive, providing coupons with a stronger incentive can be an effective sales promotion strategy.

	Including the new product		Excluding the new product	
	Beer only	Beer and RTD	Beer only	Beer and RTD
ATE(c)	537.8 ***	570.1 *	-10.9	21.5
	(159.3)	(268.6)	(154.0)	(260.7)
ATT	550.3 ***	578.3 *	-21.6	6.3
	(148.0)	(250.6)	(143.7)	(242.5)
ATE	713.4 ***	681.9 **	-165.2	-196.7
	(203.6)	(261.9)	(147.8)	(222.2)

Note: The All figures are in Japanese Yen (JPY).

Table 11. Estimates and standard errors. Standard errors in parentheses. $^{*}p < 0.05, ^{**}p < 0.01, ^{***}p < 0.001$

6. Conclusion

This paper presents a new strategy to overcome the well-known limitations of the LATE framework in settings with non-compliance. We address two central challenges in the LATE literature: identifying causal effects for broader populations (i.e., ATT and ATE) under assumptions that avoid untestable homogeneity or parametric restrictions, and relaxing the monotonicity assumption.

In developing our framework, we propose a study design that employs two auxiliary observations commonly available in baseline surveys or administrative panel data. By invoking a subpopulation parallel-trends condition on pre-treatment outcomes, we establish point identification of ATT without covariates. We then show that under an additional parallel-trends or subgroup-homogeneity assumption, ATE is also identified. Independently, by using a pre-assignment treatment indicator as an instrument for the potential treatment variable, our framework allows for the violation of the monotonicity assumption. Recognizing that practical applications often depart from strict randomization, we extend these identification results through multiply robust representations, guaranteeing consistent estimation even in observational settings. Furthermore, we present the GMM estimators that leverage over-identified moment conditions to achieve efficiency gain. We demonstrate the practical utility of our method through empirical illustrations that revisited Thornton (2008), Gerber et al. (2009), and Beam (2016), and analyze data sets from the Oregon Health Insurance Experiment (Finkelstein et al., 2012; Taubman et al., 2014) and a marketing experiment in the private sector. By broadening the target population and accommodating defiers, our method yields insights that extend beyond the traditional LATE approach.

Several avenues for future research warrant investigation. First, a promising research direction lies in partial identification of treatment effects. A significant body of literature, including a foundational work by Balke and Pearl (1997) and more recent contributions by Machado et al. (2013), has focused on deriving informative bounds on ATE—or simply identifying its sign. In this context, utilizing our key identifying assumptions (such as those in Assumptions 3, 3', 5, and 6) by introducing auxiliary variables, or relaxing those assumptions into inequality constraints, could yield informative bounds for various estimands. While Remarks 3 and 13 illustrate bounds for ATE when Assumption 3 (ii) is excluded, a more comprehensive analysis would be a valuable direction. As explored by Machado et al. (2013), such an approach is particularly powerful for determining the sign of an effect. Extending our framework in this manner would therefore contribute to this important literature by enabling a more refined discussion on identifying the sign of ATE. Second, the key assumptions of this paper, such as the parallel-trends and homogeneity in Assumptions 3 and 3' and the exclusion restrictions in Assumptions 5 and 6, could potentially be relaxed by replacing them with assumptions conditioned on covariates. While Remark 23 briefly touches on conditional identification in relation to identification of ATT_D, a more rigorous discussion of the conditional case is desirable. Third, extending the framework to multi-valued or continuous instruments and treatments would enhance its applicability beyond binary regimes. Finally, integrating modern machine learning-based methods for nuisance parameter estimation—such as double/debiased machine learning (Chernozhukov et al., 2018)—is expected to bring greater efficiency and robustness in high-dimensional settings. Collectively, these extensions would further empower researchers to conduct credible policy evaluations in complex experimental and observational contexts.

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.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters.

Cornelissen, T., Dustmann, C., Raute, A., and Schönberg, U. (2016). From late to mte: Alternative methods for the evaluation of policy interventions. *Labour Economics*, 41:47–60.

Cui, Y. and Tchetgen Tchetgen, E. (2021). A semiparametric instrumental variable approach to optimal treatment regimes under endogeneity. *Journal of the American Statistical Association*, 116(533):162–173.

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.

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

Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J. P., Allen, H., Baicker, K., and Oregon Health Study Group, t. (2012). The oregon health insurance experiment: evidence from the first year. *The Quarterly journal of economics*, 127(3):1057–1106.

Follmann, D. (2006). Augmented designs to assess immune response in vaccine trials. *Biometrics*, 62(4):1161–1169.

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.

Gabriel, E. E. and Gilbert, P. B. (2014). Evaluating principal surrogate endpoints with time-to-event data accounting for time-varying treatment efficacy. *Biostatistics*, 15(2):251–265.

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.

Hartwig, F. P., Wang, L., Smith, G. D., and Davies, N. M. (2023). Average causal effect estimation via instrumental variables: the no simultaneous heterogeneity assumption. *Epidemiology*, 34(3):325–332.

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.

Hernán, M. A. and Robins, J. M. (2006). Instruments for causal inference: an epidemiologist's dream? *Epidemiology*, 17(4):360–372.

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.

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

Machado, C., Shaikh, A., Vytlacil, E., and Lunch, C. (2013). Instrumental variables and the sign of the average treatment effect. *Unpublished Manuscript, Getulio Vargas Foundation, University of Chicago, and New York University.[2049].*

Manski, C. F. (2003). *Partial identification of probability distributions.* Springer Science & Business Media.

Martens, E. P., Pestman, W. R., de Boer, A., Belitser, S. V., and Klungel, O. H. (2006). Instrumental variables: application and limitations. *Epidemiology*, 17(3):260–267.

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.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245.

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.

Sigstad, H. (2024). Marginal treatment effects and monotonicity. *arXiv preprint arXiv:2404.03235.*

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.

Tan, Z. (2010). Marginal and nested structural models using instrumental variables. *Journal of the American Statistical Association*, 105(489):157–169.

Taubman, S. L., Allen, H. L., Wright, B. J., Baicker, K., and Finkelstein, A. N. (2014). Medicaid increases emergency-department use: evidence from oregon's health insurance experiment. *Science*, 343(6168):263–268.

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.

Graduate School of Economics, Keio University.
Email address: yuta-ota@keio.jp

Faculty of Economics, Keio University and RIKEN AIP.
Email address: hoshino@econ.keio.ac.jp

Department of Economics, London School of Economics.
Email address: t.otsu@lse.ac.uk

Appendix A. Identification and estimation

A.1. **Identification under ignorability condition.** In this section we show that our identification argument in Sections 2.1 and 2.2 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.

A.1.1. *Observable outcome before assignment.* Consider the setup in Section 2.1. In observational studies, Assumption Y (ii) is replaced with

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

This is a standard ignorability or unconfoundedness condition commonly imposed in the literature of causal inference with observational studies. Based on the discussion in Section 2.1, it is sufficient for identification of the causal estimands in (2) to identify

$$\delta_{(z,d)} = \mathbb{E}[Y_d | D_z = d], \qquad \delta^{\text{pre}}_{(z,d)} = \mathbb{E}[Y^{\text{pre}} | D_z = d], \qquad \pi_{(z,d)} = \mathbb{P}(D_z = d),$$

for each $z \in \{0,1\}$ and $d \in \{0,1\}$. To derive multiply robust representations of $\delta_{(z,d)}$, $\delta^{\text{pre}}_{(z,d)}$, and $\pi_{(z,d)}$ under Assumption Y (ii)', we introduce parametric models

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

for each $z \in \{0,1\}$ and $d \in \{0,1\}$, where α, β, λ, λ^{pre} and η are finite dimensional parameters. By using these parametric models, multiply robust representations of the population objects $\delta_{(z,d)}$, $\delta^{\text{pre}}_{(z,d)}$, and $\pi_{(z,d)}$ are obtained as follows.

Theorem 9. *Under Assumptions 1, Y (i), and (ii)', it holds*

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

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

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

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 12 parameters $\mu^u_1, \mu^u_0, \mu^u_{\text{pre}}, p^u$ for $u \in \{c, a, n\}$ are over-identified under Assumptions 1, 2, Y (i), Y(ii)', and 3 by the moment restrictions of

(4)-(7). Just identification of the 12 parameters is guaranteed by the first 12 moments of (4)-(6), and the three moments of (7) provide overidentifying restrictions. By summarizing the multiply robust properties of the estimators based on Theorem 1 and the moment restrictions, we obtain the following results.

Proposition 2. *Consider the setup of Section 2.1. Suppose Assumptions 1, 2, Y (i), Y(ii)', and 3 hold true. Then*

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

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

A.1.2. *Observable treatment before assignment.* Consider the setup in Section 2.2. In observational studies, Assumption D (ii) is replaced with

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

Based on the discussion of the previous subsection, it is sufficient for identification of the causal estimands in (2) to identify

$$\delta_{(z,d,d')} = \mathbb{E}[Y_d | D_z = d, D^{\text{pre}} = d'], \qquad \delta_{(z,d)} = \mathbb{E}[Y_d | D_z = d],$$
$$\pi_{(z,d,d')} = \mathbb{P}(D_z = d, D^{\text{pre}} = d'), \qquad \rho_{(z,d,d')} = \mathbb{P}(D_z = d | D^{\text{pre}} = d').$$

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

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

for each $z \in \{0,1\}$ and $d, d' \in \{0,1\}$, where α, β, γ, λ and η are finite dimensional parameters. By using these parametric models, multiply robust representations of the population objects $\delta_{(z,d,d')}$, $\delta_{(z,d)}$, $\pi_{(z,d,d')}$, and $\rho_{(z,d,d')}$ are obtained as follows.

Theorem 10. *Consider the setup of this subsection. Under Assumptions D (i) and (ii)', it holds*

$$\begin{aligned}
\delta_{(z,d,d')} &= \mathbb{E}\left[\frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)}\frac{\mathbb{I}\{D=d,D^{\text{pre}}=d'\}}{p_{(z,d,d')}(X;\beta)}Y\right] \\
&\quad -\mathbb{E}\left[\frac{\mathbb{I}\{Z=z\}\mathbb{I}\{D=d,D^{\text{pre}}=d'\}-e_z(X;\alpha)p_{(z,d,d')}(X;\beta)}{e_z(X;\alpha)p_{(z,d,d')}(X;\beta)}m_{(z,d,d')}(X;\gamma)\right], \\
\delta_{(z,d)} &= \mathbb{E}\left[\frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)}\frac{\mathbb{I}\{D=d\}}{\sum_{j\in(1,0)}p_{(z,d,j)}(X;\beta)}Y\right] \\
&\quad -\mathbb{E}\left[\frac{\mathbb{I}\{Z=z\}\mathbb{I}\{D=d\}-e_z(X;\alpha)\sum_{j\in(1,0)}p_{(z,d,j)}(X;\beta)}{e_z(X;\alpha)\sum_{j\in(1,0)}p_{(z,d,j)}(X;\beta)}m_{(z,d)}(X;\lambda)\right], \\
\pi_{(z,d,d')} &= \mathbb{E}\left[\frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)}\mathbb{I}\{D=d,D^{\text{pre}}=d'\}\right]-\mathbb{E}\left[\frac{\mathbb{I}\{Z=z\}-e_z(X;\alpha)}{e_z(X;\alpha)}p_{(z,d,d')}(X;\beta)\right], \\
\rho_{(z,d,d')} &= \mathbb{E}\left[\frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)}\frac{\mathbb{I}\{D^{\text{pre}}=d'\}}{\sum_{k\in(1,0)}p_{(z,k,d')}(X;\beta)}\mathbb{I}\{D=d\}\right] \\
&\quad -\mathbb{E}\left[\frac{\mathbb{I}\{Z=z\}\mathbb{I}\{D^{\text{pre}}=d'\}-e_z(X;\alpha)\sum_{k\in(1,0)}p_{(z,k,d')}(X;\beta)}{e_z(X;\alpha)\sum_{k\in(1,0)}p_{(z,k,d')}(X;\beta)}p_{(z,d,d')}(X;\eta)\right].
\end{aligned}$$

By taking the sample counterparts of these representations, we can construct multiply robust estimators for $\delta_{(z,d,d')}$, $\delta_{(z,d)}$, $\pi_{(z,d,d')}$, and $\rho_{(z,d,d')}$. Then the 10 parameters $\mu_1^c, \mu_0^c, \mu_1^d, \mu_0^d, \mu_1^a, \mu_0^n$, and p^u for $u \in \{c,a,n,d\}$ are over-identified under Assumptions 1, D (i), D (ii)', and 5 by the moment restrictions of (12), (17), and (20). Just identification of the 10 parameters is guaranteed by the 10 moments of (12) and (17), and the two moments of (20) provide overidentifying restrictions.

We close this subsection by summarizing the multiply robust properties of the estimators based on Theorem 3 and the moment restrictions.

Proposition 3. *Consider the setup of this subsection. Suppose Assumptions 1, D (i), D (ii)', and 5 hold true. Then*

(i): $\delta_{(z,d,d')}$ *can be consistently estimated if either* $\{e_z(X;\alpha), p_{(z,d,d')}(X;\beta)\}$ *or* $m_{(z,d,d')}(X;\gamma)$ *is correctly specified,*

(ii): $\delta_{(z,d)}$ *can be consistently estimated if either* $\{e_z(X;\alpha), p_{(z,d,d')}(X;\beta)\}$ *or* $m_{(z,d)}(X;\lambda)$ *is correctly specified,*

(iii): $\pi_{(z,d,d')}$ *can be consistently estimated if either* $e_z(X;\alpha)$ *or* $p_{(z,d,d')}(X;\beta)$ *is correctly specified,*

(iv): $\rho_{(z,d,d')}$ *can be consistently estimated if either* $\{e_z(X;\alpha), p_{(z,d,d')}(X;\beta)\}$ *or* $p_{(z,d,d')}(X;\eta)$ *is correctly specified,*

(v): $\mathrm{ATE}(c)$ *and* $\mathrm{ATE}(d)$ *can be consistently estimated if either* $\{e_z(X;\alpha), p_{(z,d,d')}(X;\beta)\}$, $\{m_{(z,d,d')}(X;\gamma), m_{(z,d)}(X;\lambda), p_{(z,d,d')}(X;\beta), p_{(z,d,d')}(X;\eta)\}$ *or* $\{m_{(z,d,d')}(X;\gamma), m_{(z,d)}(X;\lambda), e_z(X;\alpha), p_{(z,d,d')}(X;\eta)\}$ *is correctly specified.*

Furthermore, the multiply robust estimator for ATE(c) and ATE(d) are asymptotically locally efficient if $\{e_z(X;\alpha), p_{(z,d,d')}(X;\beta), m_{(z,d,d')}(X;\gamma), m_{(z,d)}(X;\lambda), p_{(z,d,d')}(X;\eta)\}$ are correctly specified.

Remark 27. [Switched version of Assumption 5] Under Assumption 5' (which switches D_0 and D_1 in Assumption 5), the moment restrictions above are replaced with (18), (19), and (21). Just identification of the 10 parameters is guaranteed by the 10 moments of (18), (19), and the two moments of (21) provide overidentifying restrictions.

Remark 28. [Impose both Assumptions 5 and 5'] Under both Assumptions 5 and 5', the moment restrictions above are replaced with (12) and (18). Among the 10 parameters, $\mu_1^c, \mu_0^c, \mu_1^d, \mu_0^d, \mu_1^a, \mu_0^n$, and p^u for $u \in \{c,a,n,d\}$, the six parameters, μ_1^a, μ_0^n, and p^u for $u \in \{c,a,n,d\}$ are identified in the two ways. The moments of $\delta_{(z,d)}$ is not used for identification.

Remark 29. [Special case] Under Assumption 4 (i.e., $D_0 = D^{\mathrm{pre}}$) instead of Assumption 5, the moment restrictions follow the discussion in Remark 10, and the moments of $\rho_{(z,d,d')}$ are not used for identification.

A.2. **Estimation.**

A.2.1. *Observable outcome before assignment.* In this subsection, we briefly discuss estimation and testing methods for ATE identified by Theorems 1 and 9 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 $\hat{\delta}_{(z,d)}$, $\hat{\delta}^{\mathrm{pre}}_{(z,d)}$, and $\hat{\pi}_{(z,d)}$ be the empirical (conditional) moments of $\delta_{(z,d)} = \mathbb{E}[Y_d|D_z = d]$, $\delta^{\mathrm{pre}}_{(z,d)} = \mathbb{E}[Y^{\mathrm{pre}}|D_z = d]$, 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_1^u, μ_0^u, 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 GMM estimator for ATE is obtained as $\hat{\omega}$ for

$$(\hat{\theta}, \hat{\omega}) = \arg\min_{\theta,\omega} g(\hat{\zeta}, \theta, \omega)' \Psi g(\hat{\zeta}, \theta, \omega), \tag{42}$$

where the vector of moment conditions $g(\zeta, \theta, \omega) = 0$ is obtained by stacking the equations (4)-(6) and $\omega = \text{ATE}(\theta)$ (and also (7)). 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 9, their estimating equations are given by

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

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

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

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.

A.2.2. *Observable treatment before assignment.* In this subsection, we briefly discuss estimation and testing methods for ATE(c) identified by Theorems 3 and 10 above.

First, we consider estimation of ATE(c) based on Theorem 3. Let $\hat{\delta}_{(z,d,d')}$, $\hat{\delta}_{(z,d)}$, $\hat{\pi}_{(z,d,d')}$, $\hat{\rho}_{(z,d,d')}$ be the empirical (conditional) moments of $\delta_{(z,d,d')}$, $\delta_{(z,d)}$, $\pi_{(z,d,d')}$, $\rho_{(z,d,d')}$, respectively, and $\hat{\zeta}$ and ζ be their vectorizations. Also let θ be a 10-dimensional vector given by $\mu_1^c, \mu_0^c, \mu_1^d, \mu_0^d, \mu_1^a, \mu_0^n$, and p^u for $u \in \{c, a, n, d\}$, which provides a formula for ATE(c) as

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

Then the GMM estimator for ATE(c) is obtained as $\hat{\omega}$ for

$$(\hat{\theta}, \hat{\omega}) = \arg\min_{\theta,\omega} g(\hat{\zeta}, \theta, \omega)' \Psi g(\hat{\zeta}, \theta, \omega), \tag{43}$$

where the vector of moment conditions $g(\zeta, \theta, \omega) = 0$ is obtained by stacking the equations (12), (17), and $\omega = \text{ATE(c)}(\theta)$ (and also (20)). 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

$$g_1(W, \zeta, \alpha, \beta, \gamma, \eta) = \begin{bmatrix} \left\{ \begin{array}{c} \delta_{(z,d,d')} - \frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)} \frac{\mathbb{I}\{D=d, D^{\text{pre}}=d'\}}{p_{(z,d,d')}(X;\beta)} Y \\ + \frac{\mathbb{I}\{Z=z\}\mathbb{I}\{D=d, D^{\text{pre}}=d'\} - e_z(X;\alpha) p_{(z,d,d')}(X;\beta)}{e_z(X;\alpha) p_{(z,d,d')}(X;\beta)} m_{(z,d,d')}(X;\gamma) \end{array} \right\}_{(z,d,d')} \\ \left\{ \begin{array}{c} \delta_{(z,d)} - \frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)} \frac{\mathbb{I}\{D=d\}}{\sum_{j\in\{0,1\}} p_{(z,d,j)}(X;\beta)} Y \\ + \frac{\mathbb{I}\{Z=z\}\mathbb{I}\{D=d\} - e_z(X;\alpha) \sum_{j\in\{0,1\}} p_{(z,d,j)}(X;\beta)}{e_z(X;\alpha) \sum_{j\in\{0,1\}} p_{(z,d,j)}(X;\beta)} m_{(z,d)}(X;\lambda) \end{array} \right\}_{(z,d)} \\ \left\{ \pi_{(z,d,d')} - \frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)} \mathbb{I}\{D=d, D^{\text{pre}}=d'\} + \frac{\mathbb{I}\{Z=z\} - e_z(X;\alpha)}{e_z(X;\alpha)} p_{(z,d,d')}(X;\beta) \right\}_{(z,d,d')} \\ \left\{ \begin{array}{c} \rho_{(z,d,d')} - \frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)} \frac{\mathbb{I}\{D^{\text{pre}}=d'\}}{\sum_{k\in\{0,1\}} p_{(z,k,d')}(X;\beta)} \mathbb{I}\{D=d\} \\ + \frac{\mathbb{I}\{Z=z\}\mathbb{I}\{D^{\text{pre}}=d'\} - e_z(X;\alpha) \sum_{k\in\{0,1\}} p_{(z,k,d')}(X;\beta)}{e_z(X;\alpha) \sum_{k\in\{0,1\}} p_{(z,k,d')}(X;\beta)} p_{(z,d,d')}(X;\eta) \end{array} \right\}_{(z,d,d')} \\ \xi_1(W,\alpha) \\ \xi_2(W,\beta) \\ \xi_3(W,\gamma) \\ \xi_4(W,\lambda) \\ \xi_5(W,\eta) \end{bmatrix},$$

where W mean the whole observables, ξ_1, ξ_2, ξ_3, ξ_4, and ξ_5 are estimating equations for the parameters α, β, γ, λ and η, respectively. Combining this with the moment conditions $g(\zeta, \theta, \vartheta) = 0$, the GMM estimator of ATE(c) is obtained as $\tilde{\omega}$ for

$$\begin{aligned} &(\tilde{\zeta}, \tilde{\theta}, \tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}, \tilde{\lambda}, \tilde{\eta}, \tilde{\omega}) \\ = \; &\arg\min_{\zeta,\theta,\alpha,\beta,\gamma,\lambda,\eta,\omega} \begin{bmatrix} g(\zeta,\theta,\omega) \\ \frac{1}{n}\sum_{i=1}^n g_1(W_i, \zeta, \alpha, \beta, \gamma, \lambda, \eta) \end{bmatrix}' \Psi_1 \\ &\times \begin{bmatrix} g(\zeta,\theta,\omega) \\ \frac{1}{n}\sum_{i=1}^n g_1(W_i, \zeta, \alpha, \beta, \gamma, \lambda, \eta) \end{bmatrix}, \end{aligned}$$

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.

Remark 30. [Switched version of Assumption 5] Under Assumption 5' (which switches D_0 and D_1 in Assumption 5), the vector of moment conditions $g(\zeta, \theta, \omega) = 0$ is obtained by stacking the equations (18), (19), and (21).

Remark 31. [Supposing both Assumptions 5 and 5'] Under both Assumptions 5 and 5', the vector of moment conditions $g(\zeta, \theta, \omega) = 0$ is obtained by stacking the equations (12) and (18).

Remark 32. [Special case] Under Assumption 4 (i.e., $D_0 = D^{\text{pre}}$) instead of Assumption 5, the vector of moment conditions $g(\zeta, \theta, \omega) = 0$ is obtained by stacking the equations following the discussion in Remark 10.

Appendix B. Mathematical appendix

B.1. **Derivation of expressions in Remark 1.** Suppose Assumption 1 holds. For clarity and reader's convenience, we first restate the key notation used in this section. For $u \in \{c, a, n, d\}$, $d \in \{1, 0\}$, and $z \in \{1, 0\}$, we define

$$\begin{aligned} \mu_d^u &:= \mathbb{E}[Y_d | U = u], \qquad p^u := \mathbb{P}(U = u), \\ \mu_d^{u|Z=z} &:= \mathbb{E}[Y_d | U = u, Z = z], \qquad p^{u|Z=z} := \mathbb{P}(U = u | Z = z). \end{aligned}$$

Using this notation, the estimands ATT_{D_1}, ATT_{D_0}, and ATT_D can expressed as

$$\begin{aligned} \text{ATT}_{D_1} &:= \mathbb{E}[Y_1 - Y_0 | D_1 = 1] \\ &= \mathbb{E}[Y_1 - Y_0 | D_1 = 1, D_0 = 0]\mathbb{P}(D_0 = 0 | D_1 = 1) + \mathbb{E}[Y_1 - Y_0 | D_1 = 1, D_0 = 1]\mathbb{P}(D_0 = 1 | D_1 = 1) \\ &= \frac{p^c(\mu_1^c - \mu_0^c) + p^a(\mu_1^a - \mu_0^a)}{p^c + p^a}, \end{aligned}$$

$$\begin{aligned} \text{ATT}_{D_0} &:= \mathbb{E}[Y_1 - Y_0 | D_0 = 1] \\ &= \mathbb{E}[Y_1 - Y_0 | D_1 = 1, D_0 = 1]\mathbb{P}(D_1 = 1 | D_0 = 1) + \mathbb{E}[Y_1 - Y_0 | D_1 = 0, D_0 = 1]\mathbb{P}(D_1 = 0 | D_0 = 1) \\ &= \frac{p^a(\mu_1^a - \mu_0^a) + p^d(\mu_1^d - \mu_0^d)}{p^a + p^d}, \end{aligned}$$

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

B.2. **Proof of Theorem 1.** First, under Assumptions 1, 2, and Y, the parameters $\mu_1^a, \mu_0^n, p^a, p^n, p^c, \mu_1^c$, and μ_0^c are identified as

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

In the expressions above, the second equality holds by Assumption 2, and the third equality holds by Assumptions 1 and Y. The parameters for compliers are then identified as

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

Thus, ATE for compliers is identified as $\text{ATE}(c) = \mu_1^c - \mu_0^c$.

Next, by adding Assumption 3 (i), we identify μ_0^a. The derivation is as follows:

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

where

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

The fifth equality holds by Assumption 3 (i). By rearranging the terms, we obtain $\mu_0^a = \mu_{\text{pre}}^a + \mu_0^n - \mu_{\text{pre}}^n$. Since all components are identified, 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, Assumption 3 (ii) implies $\mu_1^n - \mu_0^n = \mu_1^a - \mu_0^a$, which allows us to identify μ_1^n as $\mu_1^n = \mu_0^n + (\mu_1^a - \mu_0^a)$. Therefore, ATE is identified as

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

B.3. **Proof of Remark 2.** Under the alternative parallel trends assumption between always-takers and compliers, $\mathbb{E}[Y_0 - Y^{\text{pre}}|U = a] = \mathbb{E}[Y_0 - Y^{\text{pre}}|U = c]$, we show that the parameter μ_0^a is identified. Note that

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

where $\mu_{\text{pre}}^a = \mathbb{E}[Y^{\text{pre}}|U = a]$ and $\mu_{\text{pre}}^c = \mathbb{E}[Y^{\text{pre}}|U = c]$. Note that μ_{pre}^a and μ_{pre}^c are identified by

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

where $\mu_{\text{pre}}^n = \mathbb{E}[Y^{\text{pre}}|Z = 1, D = 0]$. Therefore, solving the initial identity for μ_0^a yields final identification result

$$\mu_0^a = \mu_{\text{pre}}^a + \mu_0^c - \mu_{\text{pre}}^c.$$

The second alternative is based on strata defined by D_0. Under the assumption that $\mathbb{E}[Y_0 - Y^{\text{pre}}|D_0 = 1] = \mathbb{E}[Y_0 - Y^{\text{pre}}|D_0 = 0]$, the parameter μ_0^a is also identified. First, this assumption allows us to express $\mathbb{E}[Y_0|D_0 = 1]$ in terms of observable quantities. By applying the assumption, we obtain

$$\begin{aligned} \mathbb{E}[Y_1|D_0 = 1] - \mathbb{E}[Y_0|D_0 = 1] &= \mathbb{E}[Y_1 - Y^{\text{pre}}|D_0 = 1] - \mathbb{E}[Y_0 - Y^{\text{pre}}|D_0 = 1] \\ &= \mathbb{E}[Y_1 - Y^{\text{pre}}|D_0 = 1] - \mathbb{E}[Y_0 - Y^{\text{pre}}|D_0 = 0] \\ &= \mathbb{E}[Y_1|D_0 = 1] - \mathbb{E}[Y^{\text{pre}}|D_0 = 1] - \mathbb{E}[Y_0|D_0 = 0] + \mathbb{E}[Y^{\text{pre}}|D_0 = 0]. \end{aligned}$$

By definition, $\mathbb{E}[Y_0|D_0 = 1]$ is also a weighted average of the outcomes for always-takers and compliers

$$\mathbb{E}[Y_0|D_0 = 1] = \frac{p^a \mu_0^a + p^d \mu_0^d}{p^a + p^d}.$$

Equating these these two expressions for $\mathbb{E}[Y_0|D_0 = 1]$ and solving for μ_0^a imply the identification result

$$\mu_0^a = \frac{\mathbb{P}(D_0 = 1)\{\mathbb{E}[Y^{\text{pre}}|D_0 = 1] + \mathbb{E}[Y_0|D_0 = 0] - \mathbb{E}[Y^{\text{pre}}|D_0 = 0]\} - p^d \mu_0^d}{p^a}.$$

Note that under the monotonicity assumption (Assumption 2), this expression simplifies as the term $p^d \mu_0^d$ drops out since there are no defiers.

The third alternative is to assume parallel trends between strata defined by D_0, i.e.,

$$\mathbb{E}[Y_0 - Y^{\text{pre}}|D_1 = 1] = \mathbb{E}[Y_0 - Y^{\text{pre}}|D_1 = 0]$$

Following a similar derivation as above, the parameter μ_0^a is identified as

$$\mu_0^a = \frac{\mathbb{P}(D_1 = 1)\{\mathbb{E}[Y^{\text{pre}}|D_1 = 1] + \mathbb{E}[Y_0|D_1 = 0] - \mathbb{E}[Y^{\text{pre}}|D_1 = 0]\} - p^c \mu_0^c}{p^a}.$$

B.4. **Proof of Theorem 2.** Under Assumptions 1, D, and 4, the parameters $\mu_1^c, \mu_1^a, \mu_0^n, \mu_0^d, p^c, p^a, p^n$, and p^d are identified as

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

In each expression above, the second equality holds by Assumption 4, and the third equality holds by Assumptions 1 and D. Also μ_1^d and μ_0^c are identified as

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

Therefore, ATE for compliers and ATE for defiers are identified as

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

B.5. **Proof of Theorem 3.** For clarity, we first recall the notation relevant to this subsection:

$$\begin{aligned}
\delta_{(z,d,d')} &:= \mathbb{E}[Y_d|D_z=d, D^{\text{pre}}=d'], \qquad \delta_{(z,d)} := \mathbb{E}[Y_d|D_z=d],\\
\pi_{(z,d,d')} &:= \mathbb{P}(D_z=d, D^{\text{pre}}=d'), \qquad \rho_{(z,d,d')} := \mathbb{P}(D_z=d|D^{\text{pre}}=d').
\end{aligned}$$

Under Assumptions 1 and D, these quantities are identified from the observed data as

$$\begin{aligned}
\delta_{(z,d,d')} &= \mathbb{E}[Y|Z=z, D=d, D^{\text{pre}}=d'],\\
\delta_{(z,d)} &= \mathbb{E}[Y|Z=z, D=d],\\
\pi_{(z,d,d')} &= \mathbb{P}(Z=z, D=d, D^{\text{pre}}=d'),\\
\rho_{(z,d,d')} &= \mathbb{P}(D=d|Z=z, D^{\text{pre}}=d'),
\end{aligned}$$

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

First, we present the identification result for $\mu_1^c, \mu_1^a, \mu_0^n$, and μ_0^d. Using Assumptions 5 (i)-(ii), we express $\delta_{(1,d,d')}$ for $d, d' \in \{0,1\}$ as

$$\begin{aligned}
\delta_{(1,d,d')} &= \mathbb{E}[Y_d | D_1 = d, D^{\text{pre}} = d'] \\
&= \mathbb{E}[Y_d | D_1 = d, D_0 = 1-d, D^{\text{pre}} = d'] \mathbb{P}(D_0 = 1-d | D_1 = d, D^{\text{pre}} = d') \\
&\quad + \mathbb{E}[Y_d | D_1 = d, D_0 = d, D^{\text{pre}} = d'] \mathbb{P}(D_0 = d | D_1 = d, D^{\text{pre}} = d') \\
&= \mathbb{E}[Y_d | D_1 = d, D_0 = 1-d] \mathbb{P}(D_0 = 1-d | D^{\text{pre}} = d') \\
&\quad + \mathbb{E}[Y_d | D_1 = d, D_0 = d] \mathbb{P}(D_0 = d | D^{\text{pre}} = d') \\
&= \mu_d^{(d,1-d)} \rho_{(0,1-d,d')} + \mu_d^{(d,d)} \rho_{(0,d,d')},
\end{aligned}$$

where the third equality holds by Assumptions 5 (i)-(ii), and in the last line, we use the shorthand notations $\mu_d^{(d,1-d)} := \mathbb{E}[Y_d | D_1 = d, D_0 = 1-d]$ and $\mu_d^{(d,d)} := \mathbb{E}[Y_d | D_1 = d, D_0 = d]$ for $d \in \{0,1\}$. By setting $d' = 0$ and $d' = 1$, respectively, we obtain the following system of equations

$$\begin{aligned}
\delta_{(1,d,0)} &= \mu_d^{(d,1-d)} \rho_{(0,1-d,0)} + \mu_d^{(d,d)} \rho_{(0,d,0)}, \\
\delta_{(1,d,1)} &= \mu_d^{(d,1-d)} \rho_{(0,1-d,1)} + \mu_d^{(d,d)} \rho_{(0,d,1)}.
\end{aligned} \tag{44}$$

Next, we solve this system for the μ terms. To find $\mu_d^{(d,1-d)}$, we eliminate $\mu_d^{(d,d)}$. Multiplying the first equation in the system by $\rho_{(0,d,1)}$ and the second by $\rho_{(0,d,0)}$ yields

$$\begin{aligned}
\delta_{(1,d,0)} \rho_{(0,d,1)} &= \mu_d^{(d,1-d)} \rho_{(0,1-d,0)} \rho_{(0,d,1)} + \mu_d^{(d,d)} \rho_{(0,d,0)} \rho_{(0,d,1)}, \\
\delta_{(1,d,1)} \rho_{(0,d,0)} &= \mu_d^{(d,1-d)} \rho_{(0,1-d,1)} \rho_{(0,d,0)} + \mu_d^{(d,d)} \rho_{(0,d,0)} \rho_{(0,d,1)}.
\end{aligned}$$

Subtracting the second new equation from the first eliminates the term with $\mu_d^{(d,d)}$, we have

$$\begin{aligned}
\delta_{(1,d,0)} \rho_{(0,d,1)} - \delta_{(1,d,1)} \rho_{(0,d,0)} &= \mu_d^{(d,1-d)} \rho_{(0,1-d,0)} \rho_{(0,d,1)} - \mu_d^{(d,1-d)} \rho_{(0,1-d,1)} \rho_{(0,d,0)} \\
&= \{\rho_{(0,1-d,0)} \rho_{(0,d,1)} - \rho_{(0,1-d,1)} \rho_{(0,d,0)}\} \mu_d^{(d,1-d)}.
\end{aligned}$$

Then solving for $\mu_d^{(d,1-d)}$, we obtain

$$\mu_d^{(d,1-d)} = \frac{\delta_{(1,d,0)} \rho_{(0,d,1)} - \delta_{(1,d,1)} \rho_{(0,d,0)}}{\rho_{(0,1-d,0)} \rho_{(0,d,1)} - \rho_{(0,1-d,1)} \rho_{(0,d,0)}}.$$

The denominator in this expression can be simplified. Since $\rho_{(0,1-d,d')} = 1 - \rho_{(0,d,d')}$, we have

$$\begin{aligned}
\rho_{(0,1-d,0)} \rho_{(0,d,1)} - \rho_{(0,1-d,1)} \rho_{(0,d,0)} &= (1 - \rho_{(0,d,0)}) \rho_{(0,d,1)} - (1 - \rho_{(0,d,1)}) \rho_{(0,d,0)} \\
&= \rho_{(0,d,1)} - \rho_{(0,d,0)} \rho_{(0,d,1)} - \rho_{(0,d,0)} + \rho_{(0,d,1)} \rho_{(0,d,0)} \\
&= \rho_{(0,d,1)} - \rho_{(0,d,0)}.
\end{aligned}$$

Substituting this result back yields

$$\mu_d^{(d,1-d)} = \frac{\delta_{(1,d,0)} \rho_{(0,d,1)} - \delta_{(1,d,1)} \rho_{(0,d,0)}}{\rho_{(0,d,1)} - \rho_{(0,d,0)}}.$$

To avoid division by zero, Assumption 5 (iii) is required. Similarly, to solve for $\mu_d^{(d,1-d)}$, we follow a parallel procedure. Multiplying the first equation of the original system (44) by $\rho_{(0,1-d,1)}$

and the second by $\rho_{(0,1-d,0)}$, and then subtract the former from the latter to obtain

$$\mu_d^{(d,d)} \quad = \quad \frac{\delta_{(1,d,1)}\rho_{(0,1-d,0)} - \delta_{(1,d,0)}\rho_{(0,1-d,1)}}{\rho_{(0,1-d,0)} - \rho_{(0,1-d,1)}}.$$

Finally, we obtain the four quantities of interest by setting $d = 1$ and $d = 0$ in the expressions for $\mu_d^{(d,1-d)}$ and $\mu_d^{(d,d)}$ as

$$\begin{aligned}
\mu_1^c &= \frac{\delta_{(1,1,0)}\rho_{(0,1,1)} - \delta_{(1,1,1)}\rho_{(0,1,0)}}{\rho_{(0,1,1)} - \rho_{(0,1,0)}}, \qquad \mu_0^d = \frac{\delta_{(1,0,0)}\rho_{(0,0,1)} - \delta_{(1,0,1)}\rho_{(0,0,0)}}{\rho_{(0,0,1)} - \rho_{(0,0,0)}}, \\
\mu_1^a &= \frac{\delta_{(1,1,1)}\rho_{(0,0,0)} - \delta_{(1,1,0)}\rho_{(0,0,1)}}{\rho_{(0,0,0)} - \rho_{(0,0,1)}}, \qquad \mu_0^n = \frac{\delta_{(1,0,1)}\rho_{(0,1,0)} - \delta_{(1,0,0)}\rho_{(0,1,1)}}{\rho_{(0,1,0)} - \rho_{(0,1,1)}}.
\end{aligned} \tag{45}$$

Thus, the probabilities p^u for $u \in \{c, a, n, d\}$ are identified under Assumption 5 (ii) as

$$\begin{aligned}
\mathbb{P}(D_1 = d, D_0 = d') &= \mathbb{P}(D_1 = d, D_0 = d', D^{\text{pre}} = 0) + \mathbb{P}(D_1 = d, D_0 = d', D^{\text{pre}} = 1) \\
&= \mathbb{P}(D_1 = d, D^{\text{pre}} = 0)\mathbb{P}(D_0 = d'|D_1 = d, D^{\text{pre}} = 0) \\
&\quad + \mathbb{P}(D_1 = d, D^{\text{pre}} = 1)\mathbb{P}(D_0 = d'|D_1 = d, D^{\text{pre}} = 1) \\
&= \mathbb{P}(D_1 = d, D^{\text{pre}} = 0)\mathbb{P}(D_0 = d'|D^{\text{pre}} = 0) \\
&\quad + \mathbb{P}(D_1 = d, D^{\text{pre}} = 1)\mathbb{P}(D_0 = d'|D^{\text{pre}} = 1) \\
&= \pi_{(1,d,0)}\rho_{(0,d',0)} + \pi_{(1,d,1)}\rho_{(0,d',1)},
\end{aligned}$$

where the third equality follows from Assumption 5 (ii). By setting $(d, d') = (1, 1)$, $(d, d') = (1, 0)$, $(d, d') = (0, 1)$ and $(d, d') = (0, 0)$, respectively, we obtain the probabilities for always-takers (p^a), compliers (p^c), defiers (p^d), and never-takers (p^n):

$$\begin{aligned}
p^a &= \pi_{(1,1,1)}\rho_{(0,1,1)} + \pi_{(1,1,0)}\rho_{(0,1,0)}, \qquad p^c = \pi_{(1,1,0)}\rho_{(0,0,0)} + \pi_{(1,1,1)}\rho_{(0,0,1)}, \\
p^d &= \pi_{(1,0,1)}\rho_{(0,1,1)} + \pi_{(1,0,0)}\rho_{(0,1,0)}, \qquad p^n = \pi_{(1,0,0)}\rho_{(0,0,0)} + \pi_{(1,0,1)}\rho_{(0,0,1)}.
\end{aligned} \tag{46}$$

Rewriting the results, such as those in equations (45) and (46), yields

$$\begin{aligned}
\mu_b^u &= \frac{\delta_{(1,b,b')}\rho_{(0,1-b',1-b')} - \delta_{(1,b,1-b')}\rho_{(0,1-b',b')}}{\rho_{(0,1-b',1-b')} - \rho_{(0,1-b',b')}}, \\
p^u &= \pi_{(1,b,b')}\rho_{(0,b',b')} + \pi_{(1,b,1-b')}\rho_{(0,b',1-b')},
\end{aligned} \tag{47}$$

where the indices (b, b') correspond to type u as

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

Furthermore, μ_1^d and μ_0^c are also identified. These expressions are derived by rearranging the law of total expectation; for example, $\delta_{(0,1)} = \mathbb{E}[Y_1|D_0 = 1]$ is the weighted average of μ_1^a and μ_1^d. This gives

$$\mu_1^d = \frac{(p^a + p^d)\delta_{(0,1)} - p^a\mu_1^a}{p^d}, \qquad \mu_0^c = \frac{(p^c + p^n)\delta_{(0,0)} - p^n\mu_0^n}{p^c}.$$

Therefore, the ATE for compliers and defiers are identified as

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

B.6. **Proof of Remark 7.** The proof is analogous to that of Theorem 3. The difference is that identification relies on Assumption 5' instead of Assumption 5.

First, we express $\delta_{(0,d,d')}$ by conditioning on the value of D_1. Following the same algebraic steps as in Appendix B.5 (solving the system of equations for $d' = 0, 1$), we obtain the identifiable expressions for $\mu_d^{(1-d,d)}$ and $\mu_d^{(d,d)}$. This requires Assumptions 5' (i)-(iii). The results are

$$\mu_d^{(1-d,d)} = \frac{\delta_{(0,d,1)}\rho_{(1,d,0)} - \delta_{(0,d,0)}\rho_{(1,d,1)}}{\rho_{(1,d,0)} - \rho_{(1,d,1)}}, \qquad \mu_d^{(d,d)} = \frac{\delta_{(0,d,0)}\rho_{(1,1-d,1)} - \delta_{(0,d,1)}\rho_{(1,1-d,0)}}{\rho_{(1,1-d,1)} - \rho_{(1,1-d,0)}}.$$

By setting $d = 1$ and $d = 0$ in these equations, we can identify the following four quantities

$$\mu_1^d = \frac{\delta_{(0,1,1)}\rho_{(1,1,0)} - \delta_{(0,1,0)}\rho_{(1,1,1)}}{\rho_{(1,1,0)} - \rho_{(1,1,1)}}, \qquad \mu_0^c = \frac{\delta_{(0,0,1)}\rho_{(1,0,0)} - \delta_{(0,0,0)}\rho_{(1,0,1)}}{\rho_{(1,0,0)} - \rho_{(1,0,1)}},$$
$$\mu_1^a = \frac{\delta_{(0,1,0)}\rho_{(1,0,1)} - \delta_{(0,1,1)}\rho_{(1,0,0)}}{\rho_{(1,0,1)} - \rho_{(1,0,0)}}, \qquad \mu_0^n = \frac{\delta_{(0,0,0)}\rho_{(1,1,1)} - \delta_{(0,0,1)}\rho_{(1,1,0)}}{\rho_{(1,1,1)} - \rho_{(1,1,0)}}.$$

Next, the probabilities p^u for $u \in \{c, a, n, d\}$ are identified under Assumption 5' (ii). The derivation follows a similar logic to that in B.5 , which yields a general expression

$$\mathbb{P}(D_1 = d, D_0 = d') \;=\; \pi_{(0,d',0)}\rho_{(1,d,0)} + \pi_{(0,d',1)}\rho_{(1,d,1)}.$$

By setting $(d, d') = (1, 1)$, $(d, d') = (1, 0)$, $(d, d') = (0, 1)$, and $(d, d') = (0, 0)$, respectively, we obtain the specific probabilities as

$$p^a = \pi_{(0,1,0)}\rho_{(1,1,0)} + \pi_{(0,1,1)}\rho_{(1,1,1)}, \qquad p^c = \pi_{(0,0,0)}\rho_{(1,1,0)} + \pi_{(0,0,1)}\rho_{(1,1,1)},$$
$$p^d = \pi_{(0,1,0)}\rho_{(1,0,0)} + \pi_{(0,1,1)}\rho_{(1,0,1)}, \qquad p^n = \pi_{(0,0,0)}\rho_{(1,0,0)} + \pi_{(0,0,1)}\rho_{(1,0,1)}.$$

Rewriting (45) and (46) yields

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

where (b, b') is same as the one in (47). Finally, μ_1^c and μ_0^d are identified using the law of total expectation, analogous to the final step in Appendix B.5, that is

$$\mu_1^c = \frac{(p^c + p^a)\delta_{(1,1)} - p^a\mu_1^a}{p^c}, \qquad \mu_0^d = \frac{(p^n + p^d)\delta_{(1,0)} - p^n\mu_0^n}{p^d}.$$

This implies identification of ATE(c) and ATE(d) as defined previously.

B.7. **Proof of Remark 11.**

B.7.1. *Stable case.* We demonstrate that Assumptions 1 (i)', 1 (ii), D (i), D (ii)', and 4 are sufficient to identify ATE(c) and ATE(d). First, note that the parameters $\mu_1^c, \mu_1^a, \mu_0^n, \mu_0^d$ and

p^c, p^a, p^n, p^d are identified as

$$\begin{aligned}
\mu_1^c &= \mathbb{E}[Y_1|D_1=1, D_0=0] = \mathbb{E}[Y_{10}|D_{10}=1, D^{\text{pre}}=0] = \mathbb{E}[Y|Z=1, D=1, D^{\text{pre}}=0],\\
\mu_1^a &= \mathbb{E}[Y_1|D_1=1, D_0=1] = \mathbb{E}[Y_{11}|D_{11}=1, D^{\text{pre}}=1] = \mathbb{E}[Y|Z=1, D=1, D^{\text{pre}}=1],\\
\mu_0^n &= \mathbb{E}[Y_0|D_1=0, D_0=0] = \mathbb{E}[Y_{00}|D_{10}=0, D^{\text{pre}}=0] = \mathbb{E}[Y|Z=1, D=0, D^{\text{pre}}=0],\\
\mu_0^d &= \mathbb{E}[Y_0|D_1=0, D_0=1] = \mathbb{E}[Y_{01}|D_{11}=0, D^{\text{pre}}=1] = \mathbb{E}[Y|Z=1, D=0, D^{\text{pre}}=1],\\
p^c &= \mathbb{P}(D_1=1, D_0=0) = \mathbb{P}(D_{10}=1, D^{\text{pre}}=0) = \mathbb{P}(D=1, D^{\text{pre}}=0|Z=1),\\
p^a &= \mathbb{P}(D_1=1, D_0=1) = \mathbb{P}(D_{11}=1, D^{\text{pre}}=1) = \mathbb{P}(D=1, D^{\text{pre}}=1|Z=1),\\
p^n &= \mathbb{P}(D_1=0, D_0=0) = \mathbb{P}(D_{10}=0, D^{\text{pre}}=0) = \mathbb{P}(D=0, D^{\text{pre}}=0|Z=1),\\
p^d &= \mathbb{P}(D_1=0, D_0=1) = \mathbb{P}(D_{11}=0, D^{\text{pre}}=1) = \mathbb{P}(D=0, D^{\text{pre}}=1|Z=1),
\end{aligned}$$

where the second equality in each expression holds due to Assumption 4. Under this assumption, an individual's principal stratum uniquely determines their value of D^{pre}: for example, compliers ($D_1=1, D_0=0$) must have $D^{\text{pre}}=0$, while always-takers ($D_1=1, D_0=1$) must have $D^{\text{pre}}=1$. This fixed relationship ensures that the potential heterogeneity due to D^{pre} does not affect the parameters defined for each principal stratum. The third equality in each line then follows from Assumptions 1 (i)' and D (ii)'.

Next, the remaining parameters, μ_1^d and μ_0^c, are identified as

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

Thus, we identify $\text{ATE}(c) = \mu_1^c - \mu_0^c$ and $\text{ATE}(d) = \mu_1^d - \mu_0^d$.

B.7.2. *Unstable case* . We demonstrate that Assumptions 1 (i)', 1 (ii), D (i), D (ii)', and 5, are sufficient to identify ATE(c) and ATE(d). Under Assumption 1 (i)', Assumption 5 implies the following conditions hold:

- **(i):** [Exclusion restriction on treatment before assignment] $\mathbb{E}[Y_{dD^{\text{pre}}}|D_{1D^{\text{pre}}}=d, D_{0D^{\text{pre}}}=d', D^{\text{pre}}] = \mathbb{E}[Y_d|D_1=d, D_0=d']$ for each $u \in \{c, a, n, d\}$.
- **(ii):** [Exclusion restriction through treatment before assignment] $\mathbb{P}(D_{0D^{\text{pre}}}=1|D^{\text{pre}}, D_{1D^{\text{pre}}}) = \mathbb{P}(D_{0D^{\text{pre}}}=1|D^{\text{pre}})$.
- **(iii):** [Relevance condition on treatment before assignment] $\text{Cov}(D_{0D^{\text{pre}}}, D^{\text{pre}}) \neq 0$.

We use the following notation

$$\begin{aligned}
\delta_{(z,d,d')} &= \mathbb{E}[Y_{dd'}|D_{zd'}=d, D^{\text{pre}}=d'], & \delta_{(z,d)} &= \mathbb{E}[Y_d|D_z=d],\\
\pi_{(z,d,d')} &= \mathbb{P}(D_{zd'}=d, D^{\text{pre}}=d'), & \rho_{(z,d,d')} &= \mathbb{P}(D_{zd'}=d|D^{\text{pre}}=d').
\end{aligned}$$

These quantities are identified from the observed data as

$$\begin{aligned}\delta_{(z,d,d')} &= \mathbb{E}[Y|Z=z,D=d,D^{\text{pre}}=d'],\\ \delta_{(z,d)} &= \mathbb{E}[Y|Z=z,D=d],\\ \pi_{(z,d,d')} &= \mathbb{P}(Z=z,D=d,D^{\text{pre}}=d'),\\ \rho_{(z,d,d')} &= \mathbb{P}(D=d|Z=z,D^{\text{pre}}=d').\end{aligned}$$

for each $z \in \{0,1\}$ and $d, d' \in \{0,1\}$. First, we present the identification result for $\mu_1^c, \mu_1^a, \mu_0^n$, and μ_0^d. As a first step, using Assumption 5 (i)-(ii), we express $\delta_{(1,d,d')}$ for $d, d' \in \{0,1\}$ as

$$\begin{aligned}\delta_{(1,d,d')} &= \mathbb{E}[Y_{dd'}|D_{1d'}=d,D^{\text{pre}}=d']\\ &= \mathbb{E}[Y_{dd'}|D_{1d'}=d,D_{0d'}=1-d,D^{\text{pre}}=d']\mathbb{P}(D_{0d'}=1-d|D_{1d'}=d,D^{\text{pre}}=d')\\ &\quad+\mathbb{E}[Y_{dd'}|D_{1d'}=d,D_{0d'}=d,D^{\text{pre}}=d']\mathbb{P}(D_{0d'}=d|D_{1d'}=d,D^{\text{pre}}=d')\\ &= \mathbb{E}[Y_d|D_1=d,D_0=1-d]\mathbb{P}(D_{0d'}=1-d|D^{\text{pre}}=d')\\ &\quad+\mathbb{E}[Y_d|D_1=d,D_0=d]\mathbb{P}(D_{0d'}=d|D^{\text{pre}}=d')\\ &= \mu_d^{(d,1-d)}\rho_{(0,1-d,d')}+\mu_d^{(d,d)}\rho_{(0,d,d')},\end{aligned}$$

where the third equality holds by Assumptions 5 (i)-(ii), and in the last line, we use the shorthand notation $\mu_d^{(d,1-d)} := \mathbb{E}[Y_d|D_1=d,D_0=1-d]$ and $\mu_d^{(d,d)} := \mathbb{E}[Y_d|D_1=d,D_0=d]$ for $d \in \{0,1\}$. Following the same procedure as in Appendix B.5, we obtain

$$\mu_d^{(d,1-d)} = \frac{\delta_{(1,d,0)}\rho_{(0,d,1)}-\delta_{(1,d,1)}\rho_{(0,d,0)}}{\rho_{(0,d,1)}-\rho_{(0,d,0)}}.$$

To avoid division by zero, Assumption 5 (iii) is required. Applying a similar procedure to eliminate $\mu_d^{(d,1-d)}$ instead of $\mu_d^{(d,d)}$, we obtain

$$\mu_d^{(d,d)} = \frac{\delta_{(1,d,1)}\rho_{(0,1-d,0)}-\delta_{(1,d,0)}\rho_{(0,1-d,1)}}{\rho_{(0,1-d,0)}-\rho_{(0,1-d,1)}}.$$

Thus, the probabilities p^u for $u \in \{c,a,n,d\}$ are identified under Assumption 5 (ii) as

$$\begin{aligned}\mathbb{P}(D_1=d,D_0=d') &= \mathbb{P}(D_1=d,D_0=d')\\ &= \mathbb{P}(D_{10}=d,D_{00}=d',D^{\text{pre}}=0)+\mathbb{P}(D_{11}=d,D_{01}=d',D^{\text{pre}}=1)\\ &= \mathbb{P}(D_{10}=d,D^{\text{pre}}=0)\mathbb{P}(D_{00}=d'|D_{10}=d,D^{\text{pre}}=0)\\ &\quad+\mathbb{P}(D_{10}=d,D^{\text{pre}}=1)\mathbb{P}(D_{00}=d'|D_{10}=d,D^{\text{pre}}=1)\\ &= \mathbb{P}(D_{10}=d,D^{\text{pre}}=0)\mathbb{P}(D_{00}=d'|D^{\text{pre}}=0)\\ &\quad+\mathbb{P}(D_{11}=d,D^{\text{pre}}=1)\mathbb{P}(D_{01}=d'|D^{\text{pre}}=1)\\ &= \pi_{(1,d,0)}\rho_{(0,d',0)}+\pi_{(1,d,1)}\rho_{(0,d',1)},\end{aligned}$$

where the forth equality follows from Assumption 5 (ii). The rest follows by applying the same procedure as in Appendix B.5 so that ATE(c) and ATE(d) can be identified.

We note that the discussion above also holds when replacing Assumption 5 with Assumption 5'.

B.8. **Proof of Theorem 4 .** This proof combines the results from the preceding sections (Appendices B.2 and B.5). We first recall the relevant notation:

$$\delta_{(z,d,d')} = \mathbb{E}[Y_d|D_z = d, D^{\text{pre}} = d'], \qquad \delta_{(z,d)} = \mathbb{E}[Y_d|D_z = d],$$
$$\delta^{\text{pre}}_{(z,d,d')} = \mathbb{E}[Y^{\text{pre}}|D_z = d, D^{\text{pre}} = d'], \qquad \delta^{\text{pre}}_{(z,d)} = \mathbb{E}[Y^{\text{pre}}|D_z = d],$$
$$\pi_{(z,d,d')} = \mathbb{P}(D_z = d, D^{\text{pre}} = d'), \qquad \rho_{(z,d,d')} = \mathbb{P}(D_z = d|D^{\text{pre}} = d').$$

These quantities are identified from the observed data under Assumptions 1 and YD. First, by following the same identification strategy detailed in Appendix B.5, the parameters $\mu_1^c, \mu_1^a, \mu_0^n, \mu_0^d$, and p^u for $u \in \{c, a, n, d\}$ are identified under Assumptions 1, YD, and 5. This yields the general solutions

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

where the pair (b, b') corresponds to each type u as defined in Appendix B.5. Furthermore, the parameters μ_1^d and μ_0^c are subsequently identified via the law of total expectation, using the quantities already identified above:

$$\mu_1^d = \frac{(p^a + p^d)\delta_{(0,1)} - p^a\mu_1^a}{p^d}, \qquad \mu_0^c = \frac{(p^c + p^n)\delta_{(0,0)} - p^n\mu_0^n}{p^c}.$$

This allows for identification of $\text{ATE}(c) = \mu_1^c - \mu_0^c$ and $\text{ATE}(d) = \mu_1^d - \mu_0^d$.

Next, we identify the remaining parameters required for ATT and ATE. Under Assumption 3 (i), μ_0^a is identified as

$$\mu_0^a = \mu_{\text{pre}}^a + \mu_0^n - \mu_{\text{pre}}^n,$$

where $\mu_{\text{pre}}^a = \mathbb{E}[Y^{\text{pre}}|D_1 = 1, D_0 = 1]$ and $\mu_{\text{pre}}^n = \mathbb{E}[Y^{\text{pre}}|D_1 = 0, D_0 = 0]$.

Under Assumptions 5 (ii)-(iii) and 6, μ_{pre}^a and $\mu_{\text{pre}}^n$are identified using a procedure identical to the one used to identify μ_d^u terms, with the sole modification of replacing Y_d with Y^{pre}. This yields the corresponding solution:

$$\mu_{\text{pre}}^u = \frac{\delta^{\text{pre}}_{(1,b,b')}\rho_{(0,1-b',1-b')} - \delta^{\text{pre}}_{(1,b,1-b')}\rho_{(0,1-b',b')}}{\rho_{(0,1-b',1-b')} - \rho_{(0,1-b',b')}},$$

where the pair (b, b') corresponds to each type u as defined in Appendix B.5. With μ_0^a identified, ATT is also 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, Assumption 3 (ii) guarantees identification of μ_1^n as $\mu_1^n = \mu_0^n + \mu_1^a - \mu_0^a$, which allows the identification of ATE:

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

B.9. **Proof of Remark 14.** The proof presented here is nearly identical to Appendix B.8, with the difference being the use of Assumption 5' instead of Assumption 5.

First, the parameters $\mu_0^c, \mu_1^a, \mu_0^n, \mu_1^d$, and p^u for $u \in \{c, a, n, d\}$ are identified under Assumptions 1, YD, and 5'. This yields the general solutions:

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

where the pair (b, b') corresponds to each type u as defined in Appendix B.5. Furthermore, the parameters μ_0^c and μ_1^d are subsequently identified via the law of total expectation, using the quantities already identified above:

$$\mu_1^c = \frac{(p^c + p^a)\delta_{(1,1)} - p^a \mu_1^a}{p^c}, \qquad \mu_0^d = \frac{(p^n + p^d)\delta_{(1,0)} - p^n \mu_0^n}{p^d}.$$

Therefore, ATE for compliers and ATE for defiers are identified. Under Assumption 3 (i), μ_0^a is identified as

$$\mu_0^a = \mu_{\text{pre}}^a + \mu_0^n - \mu_{\text{pre}}^n.$$

Under Assumption 5'(ii)-(iii) and 6, the μ_{pre}^a and μ_{pre}^n are identified as

$$\mu_{\text{pre}}^u = \frac{\delta_{(1,b,b')}^{\text{pre}}\rho_{(0,1-b',1-b')} - \delta_{(1,b,1-b')}^{\text{pre}}\rho_{(0,1-b',b')}}{\rho_{(0,1-b',1-b')} - \rho_{(0,1-b',b')}},$$

where the pair (b, b') corresponds to each type u as defined in Appendix B.5. Then ATT is also identified. Finally, Assumption 3 (ii) guarantees identification of μ_1^n as $\mu_1^n = \mu_0^n + \mu_1^a - \mu_0^a$, which allows identification of ATE.

B.10. **Proof of Remark 12.** We provide a proof of identification of μ_0^a, μ_1^n, ATT, and ATE. Other parameters (i.e., $\mu_1^c, \mu_0^c, \mu_1^a, \mu_0^n, \mu_1^d, \mu_0^d, \mu_{\text{pre}}^c, \mu_{\text{pre}}^a, \mu_{\text{pre}}^n, \mu_{\text{pre}}^d, p^c, p^a, p^n, p^d$, and ATE$(c)$ and ATE(d)) are identified in the same way as the proof of Theorem 4.

Proof under Assumptions 3 (i)' and 3 (ii). Under Assumption 3 (i)' in addition to 1, YD, 5, and 6, μ_0^a and ATT are identified as

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

where the third equality uses assumption 3 (i)'. ATT is identified as

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

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 3 (i) and 3 (ii)'. Under Assumption 3 (i) in addition to 1, YD, 5, and 6, μ_0^a and ATT are identified in the same way as the proof of Theorem 4. Next, under Assumption 3 (ii)' in addition to 1, YD, 5, and 6, μ_1^n and ATE are identified under Assumption 3 (ii)' as

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

where the third equality uses assumption 3 (ii)'. 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 3 (i)' and 3 (ii)'. Under Assumption 3 (i)' in addition to 1, YD, 5, and 6, μ_0^a and ATT are identified in the same way as the proof under Assumptions 3 (i)' and 3 (ii). Then under Assumption 3 (ii)' in addition to 1, YD, 5, and 6, μ_1^n and ATE are identified in the same way as the proof under Assumptions 3 (i) and 3 (ii)'.

We note that the discussion above also holds when replacing Assumption 5 with Assumption 5'.

B.11. **Proof of Remark 17.** This proof demonstrates identification under the stable case, where Assumption 4 holds. 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, YD, and 4 as

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

where each of the second equality holds by Assumption 4. The parameters μ_1^d and μ_0^c are identified via the law of total expectation as

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

This allows for identification of $\text{ATE}(c) = \mu_1^c - \mu_0^c$ and $\text{ATE}(d) = \mu_1^d - \mu_0^d$. Next, we identify the parameters required for ATT and the ATE. Under Assumption 3 (i), μ_0^a is identified as

$$\mu_0^a = \mu_{\text{pre}}^a + \mu_0^n - \mu_{\text{pre}}^n.$$

This requires the identification of μ_{pre}^a and μ_{pre}^n. These terms are identified from the data as

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

In each line above, the second equality holds by Assumption 4. With μ_0^a now identified, ATT is also identified. Finally, Assumption 3 (ii) guarantees identification of μ_1^n as $\mu_1^n = \mu_0^n + \mu_1^a - \mu_0^a$, which allows identification of ATE.

The above argument 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_{pre}^u for $u \in \{a, n\}$. Indeed by introducing two more parameters $(\mu_{\text{pre}}^c, \mu_{\text{pre}}^d)$, we have four additional restrictions:

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

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

$$\mu_{\text{pre}}^c = \mathbb{E}[Y^{\text{pre}}|Z = 1, D = 1, D^{\text{pre}} = 0], \qquad \mu_{\text{pre}}^d = \mathbb{E}[Y^{\text{pre}}|Z = 1, D = 0, D^{\text{pre}} = 1].$$

These additional moment conditions can be incorporated by using the GMM approach.

B.12. **Proof of Remark 19.** Under Assumption 1 (i)', Assumption YD (iii) is replaced with the following condition:

Assumption D (ii)': Z is independent from
$(D^{\text{pre}}, D_{11}, D_{10}, D_{01}, D_{00}, Y_{111}, Y_{101}, Y_{011}, Y_{001}, Y_{110}, Y_{100}, Y_{010}, Y_{000})$.

We demonstrate that Assumptions 1 (i)', 1 (ii), YD (i), YD (ii), YD (iii)', 3, 5, and 6 are sufficient to identify ATE(c), ATT, and ATE. Based on the discussion in Appendix B.7.2, the parameters $\mu_1^c, \mu_0^c, \mu_1^a, \mu_0^n, \mu_1^d, \mu_0^d$ and p^c, p^a, p^n, p^d are identified under Assumptions 1 (i)', 1 (ii), YD (i), YD (ii), YD (iii)', and 5. Next, we identify the parameters required for ATT and ATE. Assumption 3 (i) guarantees identification of μ_0^a as $\mu_0^a = \mu_{\text{pre}}^a + \mu_0^n - \mu_{\text{pre}}^n$. This requires the identification of μ_{pre}^a and μ_{pre}^n. Under Assumption 1 (i)', Assumption 6 implies the following conditions hold:

[Exclusion restriction on treatment before assignment for pre-treatment outcome] $\mathbb{E}[Y^{\text{pre}}|D_{1D^{\text{pre}}} = d, D_{0D^{\text{pre}}} = d', D^{\text{pre}}] = \mathbb{E}[Y^{\text{pre}}|D_1 = d, D_0 = d']$ for each $u \in \{c, a, n, d\}$.

Under Assumption 5 (ii)-(iii) and 6, μ_{pre}^a and $\mu_{\text{pre}}^n$are identified using a procedure identical to the one used to identify μ_d^u terms in Appendix B.7.2, with the sole modification of replacing Y_d with Y^{pre}. With μ_0^a identified, ATT is also identified. Finally, Assumption 3 (ii) guarantees identification of μ_1^n as $\mu_1^n = \mu_0^n + \mu_1^a - \mu_0^a$, which allows the identification of ATE.

We note that the discussion above also holds when replacing Assumption 5 with Assumption 5'.

B.13. **Proof of Theorem 5.** Since the proofs are similar, we only provide a proof of the multiply robust representation of $\delta_{(z,d,d')}$. Observe that the right hand side of (36) is written as

$$
\begin{aligned}
& \mathbb{E}\left[\frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)}\frac{\mathbb{I}\{D=d,D^{\text{pre}}=d'\}}{p_{(z,d,d')}(X;\beta)}Y\right] \\
& -\mathbb{E}\left[\frac{\mathbb{I}\{Z=z\}\mathbb{I}\{D=d,D^{\text{pre}}=d'\}-e_z(X;\alpha)p_{(z,d,d')}(X;\beta)}{e_z(X;\alpha)p_{(z,d,d')}(X;\beta)}m_{(z,d,d')}(X;\gamma)\right], \\
= \; & \mathbb{E}\left[\frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)}\frac{\mathbb{I}\{D=d,D^{\text{pre}}=d'\}}{p_{(z,d,d')}(X;\beta)}\{Y-m_{(z,d,d')}(X;\gamma)\}\right]+\mathbb{E}[m_{(z,d,d')}(X;\gamma)]. \qquad (48)
\end{aligned}
$$

First, consider the case where $e_z(X;\alpha)$ and $p_{(z,d,d')}(X;\beta)$ are correctly specified. In this case, the the first term of (48) is written as

$$
\begin{aligned}
& \mathbb{E}\left[\frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)}\frac{\mathbb{I}\{D=d,D^{\text{pre}}=d'\}}{p_{(z,d,d')}(X;\beta)}\{Y-m_{(z,d,d')}(X;\gamma)\}\right] \\
= \; & \mathbb{E}_X\left[\mathbb{E}\left[\frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)}\frac{\mathbb{I}\{D=d,D^{\text{pre}}=d'\}}{p_{(z,d,d')}(X;\beta)}\{Y-m_{(z,d,d')}(X;\gamma)\}\middle|X\right]\right] \\
= \; & \mathbb{E}_X\left[\frac{\mathbb{P}(Z=z|X)}{e_z(X;\alpha)}\frac{\mathbb{P}(D=d,D^{\text{pre}}=d^{'}|Z=z,X)}{p_{(z,d,d')}(X;\beta)}\mathbb{E}[\{Y-m_{(z,d,d')}(X;\gamma)\}|Z=z,D_z=d,D^{\text{pre}}=d',X]\right] \\
= \; & \mathbb{E}_X\left[\mathbb{E}[\{Y-m_{(z,d,d')}(X;\gamma)\}|Z=z,D_z=d,D^{\text{pre}}=d',X]\right] \\
= \; & \mathbb{E}_X\left[\mathbb{E}[Y_d|Z=z,D_z=d,D^{\text{pre}}=d',X]-m_{(z,d,d')}(X;\gamma)\right] \\
= \; & \mathbb{E}_X\left[\mathbb{E}[Y_d|D_z=d,D^{\text{pre}}=d',X]-m_{(z,d,d')}(X;\gamma)\right] \\
= \; & \delta_{(z,d,d')}-\mathbb{E}[m_{(z,d,d')}(X;\gamma)],
\end{aligned}
$$

where the third equality holds by $e_z(X;\alpha)=\mathbb{P}(Z=z|X)$ and $p_{(z,d,d')}(X;\beta)=\mathbb{P}(D=d,D^{\text{pre}}=d'|Z=z,X)$. Thus we obtain the conclusion.

Next, consider the case where $m_{(z,d,d')}(X;\gamma)=\mathbb{E}[Y|Z=z,D=d,D^{\text{pre}}=d',X]$ is correctly specified. In this case, the the first term of (48) is written as

$$
\begin{aligned}
& \mathbb{E}\left[\frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)}\frac{\mathbb{I}\{D=d,D^{\text{pre}}=d'\}}{p_{(z,d,d')}(X;\beta)}\{Y-m_{(z,d,d')}(X;\gamma)\}\right] \\
= \; & \mathbb{E}_X\left[\mathbb{E}\left[\frac{\mathbb{I}\{Z=z\}}{e_z(X;\alpha)}\frac{\mathbb{I}\{D=d,D^{\text{pre}}=d'\}}{p_{(z,d,d')}(X;\beta)}\{Y-m_{(z,d,d')}(X;\gamma)\}\middle|X\right]\right] \\
= \; & \mathbb{E}_X\left[\frac{\mathbb{P}(Z=z|X)}{e_z(X;\alpha)}\frac{\mathbb{P}(D=d,D^{\text{pre}}=d^{'}|Z=z,X)}{p_{(z,d,d')}(X;\beta)}\mathbb{E}[\{Y-m_{(z,d,d')}(X;\gamma)|Z=z,D_z=d,D^{\text{pre}}=d',X]\right] \\
= \; & \mathbb{E}_X\left[\frac{\mathbb{P}(Z=z|X)}{e_z(X;\alpha)}\frac{\mathbb{P}(D=d,D^{\text{pre}}=d^{'}|Z=z,X)}{p_{(z,d,d')}(X;\beta)}\{\mathbb{E}[Y|Z=z,D_z=d,D^{\text{pre}}=d',X]-m_{(z,d,d')}(X;\gamma)\}\right] \\
= \; & 0,
\end{aligned}
$$

where the fourth equality holds by $m_{(z,d,d')}(X;\gamma)=\mathbb{E}[Y|Z=z,D=d,D^{\text{pre}}=d',X]$. Thus, the conclusion follows by $\delta_{(z,d,d')}=\mathbb{E}[m_{(z,d,d')}(X;\gamma)]$.

B.14. **Proof of Theorem 6.** We provide a proof of identification of μ_0^a, μ_1^n, ATT, and ATE. Other parameters (i.e., $\mu_1^c, \mu_0^c, \mu_1^a, \mu_0^n, \mu_1^d, \mu_0^d, p^c, p^a, p^n, p^d$,ATE($c$) and ATE($d$)) are identified in the same way as the proof of Remark 17.

Proof under Assumptions 3a (i) and 3 (ii). Under Assumption 3a (i) in addition to 1 ,YD, YD-a, and 4, μ_0^a and ATT are identified as

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

where

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

In the equation of $\mu_1^a - \mu_0^a$, the third equality uses assumption 3a (i). Each equality in the equations of $\mathbb{E}[Y_1^{\text{pre}}|U=a]$ and $\mathbb{E}[Y_1^{\text{pre}}|U=d]$ uses Assumptions 1, YD, YD-a and 4. 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

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

where

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

In the equation of $\mu_1^n - \mu_0^n$, the third equality uses Assumption 3a (ii). Each equality in the equations of $\mathbb{E}[Y_0^{\text{pre}}|U=c]$ and $\mathbb{E}[Y_0^{\text{pre}}|U=n]$ uses Assumptions 1, YD, YD-a and 4. 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, YD, YD-a 4, and 3a (i), μ_0^a and ATT are identified as

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

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

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

B.15. **Proof of Theorem 8.** The identification strategy from Theorem 4 applies directly, substituting D^{post} for D^{pre} and relabeling Assumptions YD, 5, and 6 as YD-c, 5c, and 6c.

B.16. **Proof of Theorem 9 and 10.** The proofs for these theorems are analogous to the proof of Theorem 5 and are therefore omitted.

Appendix C. Simulation

Using numerical simulations, we evaluate the finite sample properties of our estimators for ATE(c), ATT, and ATE under six different setups. These setups vary based on the identifying assumptions (with monotonicity vs. without monotonicity) and the assignment mechanism (random assignment vs. ignorability). The setups without monotonicity are further divided: we refer to setups based on Assumption 4 as the "stable case" and those based on Assumption 5 as the "unstable case." The six cases are:

(1) Monotonicity with random assignment: Estimation using Y^{pre} (Theorem 1)
(2) Monotonicity with ignorability: Estimation using Y^{pre} (Theorem 9)
(3) Stable case with random assignment: Estimation using Y^{pre} and D^{pre} (Remark 17)
(4) Stable case with ignorability: Estimation using Y^{pre} and D^{pre} (Remark 22)
(5) Unstable case with random assignment: Estimation using Y^{pre} and D^{pre} (Theorem 4)
(6) Unstable case with ignorability: Estimation using Y^{pre} and D^{pre} (Theorem 5)

For each setup, we evaluate four scenarios by varying the sample sizes ($n = 2000$ or $n = 10000$) and the distribution of the outcome variable (Normal and Bernoulli).

C.1. **Data generating process.** The data generation process in this simulation study is as follows. For unit i, covariates are generated as $X_{i1}, X_{i2} \sim_{\text{iid}} N(1, 0.3)$, $X_{i3}, X_{i4} \sim_{\text{iid}} N(-1, 0.3)$,

$W_{i1}, W_{i2}, W_{i3} \sim_{\text{iid}} N(0, 0.3)$, and $V_{i1}, V_{i2}, V_{i3}, V_{i4} \sim_{\text{iid}} N(0, 0.3)$. Add intercepts and put them together into vectors $\mathbf{X}_i = (1, X_{i1}, \ldots, X_{i4})'$, $\mathbf{W}_i = (1, W_{i1}, \ldots, W_{i3})'$, and $\mathbf{V}_i = (1, V_{i1}, \ldots, V_{i4})'$.

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 $\phi_n = (0.15, 0.2, -0.2, -0.1)'$.

On the other side, in the setup without monotonicity, D_1 and D_0 are generated by the following logistic models:

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

where $\zeta_1 = (0.2, 0.3, -0.1)'$ and $\zeta_0 = (-0.2, 0.3, -0.1)'$. In the stable case, D_i^{pre} is generated such that $D_i^{\text{pre}} = D_{i0}$. In the unstable case, D_i^{pre} is generated to be unequal to D_{i0} for a random 30% of the subjects (i.e., $D_i^{\text{pre}} \neq D_{i0}$), while for the remaining 70%, $D_i^{\text{pre}} = D_{i0}$.

Then, the principal stratum is generated as

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

The outcomes following the normal distribution are generated as

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

where $\alpha = 2$, $\gamma_{X,a} = (1, 1, -1, 1)'$, $\gamma_{X,c} = (3, 1, -1, 1)'$, $\gamma_{X,n} = (-2, 1, -1, 1)'$, $\gamma_{X,d} = (-1, 1, -1, 1)'$, $\gamma_V = (2, -1, 2, -2)'$, $\beta_a = 1$, $\beta_c = 2$, $\beta_n = 1$, $\beta_d = 3$, and $\sigma^2 = 0.5$. The outcomes following the Bernoulli distribution are generated as

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

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^{pre} or Y^{pre} is generated as

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

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

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

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 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 is generated based on the following model:

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

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.

C.2. **Simulation result.** For the ignorability setups, we assume that both the propensity score and outcome models are correctly specified. For each of the six setups, we conduct 1000 simulations and report the average estimate, standard error, and coverage rate. Standard errors are calculated using 200 bootstrap resamples. The results are presented in Tables 12, 13, and 14.

For comparison, the first row of each table reports the LATE estimated using a standard Wald-type estimator (for the random assignment scenarios only). The second and subsequent rows report the results for our proposed method. Our method provides estimates not only for the main estimands (ATE(c), ATT, and ATE), but also for the underlying parameters for each principal stratum, such as μ_1^u, μ_0^u, p^u, and ATE(u). The ability to obtain accurate estimates for these underlying parameters is particularly valuable for policymakers, as they can be used to inform the design and targeting of interventions.

Across all scenarios, our proposed method performs well. The estimates for each estimand and parameter are nearly unbiased, and the empirical coverage rates are close to the 95% level. As expected, the standard errors for the $n = 2000$ scenarios were two to three times larger than for the $n = 10000$ scenarios. However, even with $n = 2000$, the estimates are sufficiently precise to allow for a meaningful interpretation of the causal effects. For example, in the monotonicity setup with a normal distribution and random assignment, the ATE is 1.34, with a standard error of 0.19 for $n = 2000$.Furthermore, the performance of our estimator is consistent across different assignment mechanisms, showing no significant degradation when moving from the random assignment scenarios to the ignorability scenarios. As is theoretically expected, the LATE is equivalent to the ATE(c) in the monotonicity setups. In contrast, in the setups without monotonicity (both stable and unstable), the standard LATE is biased for the ATE(c). Our proposed method provides reliable estimates even in these situations where the LATE is heavily biased or imprecise. This highlights the utility of our proposed method.

	Random assignment								Ignorability							
	$n = 2000$				$n = 10000$				$n = 2000$				$n = 10000$			
	θ	$\hat{\theta}$	s.e.	cover	θ	$\hat{\theta}$	s.e.	cover	θ	$\hat{\theta}$	s.e.	cover	θ	$\hat{\theta}$	s.e.	cover
								Normal distribution								
$LATE$	2.00	1.99	0.53	0.94	2.00	2.01	0.24	0.95	-	-	-	-	-	-	-	-
$ATE(c)$	2.00	1.99	0.53	0.94	2.00	2.01	0.24	0.95	2.00	1.99	0.72	0.94	2.00	1.99	0.32	0.94
$ATE(a)$	1.00	1.00	0.06	0.94	1.00	1.00	0.03	0.95	1.00	1.01	0.10	0.94	1.00	1.00	0.04	0.94
$ATE(n)$	1.00	1.00	0.06	0.94	1.00	1.00	0.03	0.95	1.00	1.01	0.10	0.94	1.00	1.00	0.04	0.94
ATT	1.51	1.51	0.27	0.94	1.51	1.51	0.12	0.96	1.51	1.51	0.36	0.94	1.51	1.51	0.16	0.94
ATE	1.34	1.34	0.19	0.93	1.34	1.34	0.08	0.96	1.34	1.35	0.25	0.94	1.34	1.34	0.11	0.94
μ_1^c	8.00	8.00	0.37	0.95	8.00	8.01	0.16	0.95	8.00	8.01	0.47	0.92	8.00	8.00	0.21	0.93
μ_1^a	5.00	5.00	0.17	0.94	5.00	5.00	0.08	0.95	5.00	5.02	0.27	0.94	5.00	5.01	0.12	0.94
μ_1^n	2.00	2.00	0.19	0.94	2.00	2.00	0.09	0.95	2.00	2.01	0.19	0.93	2.00	2.00	0.09	0.95
μ_0^c	6.00	6.01	0.46	0.95	6.00	6.00	0.20	0.96	6.00	6.01	0.67	0.95	6.00	6.01	0.30	0.94
μ_0^a	4.00	4.00	0.18	0.94	4.00	4.00	0.08	0.95	4.00	4.01	0.25	0.95	4.00	4.00	0.11	0.94
μ_0^n	1.00	1.00	0.19	0.95	1.00	1.00	0.08	0.95	1.00	1.00	0.17	0.93	1.00	1.00	0.07	0.94
p^c	0.34	0.34	0.02	0.94	0.34	0.34	0.01	0.95	0.34	0.34	0.03	0.94	0.34	0.34	0.01	0.95
p^a	0.33	0.33	0.02	0.94	0.33	0.33	0.01	0.95	0.33	0.33	0.02	0.94	0.33	0.33	0.01	0.94
p^n	0.33	0.33	0.02	0.94	0.33	0.33	0.01	0.95	0.33	0.33	0.02	0.93	0.33	0.33	0.01	0.95
								Bernoulli distribution								
$LATE$	0.16	0.16	0.06	0.94	0.16	0.16	0.03	0.94	-	-	-	-	-	-	-	-
$ATE(c)$	0.16	0.16	0.06	0.94	0.16	0.16	0.03	0.94	0.16	0.16	0.09	0.94	0.16	0.16	0.04	0.94
$ATE(a)$	0.11	0.11	0.04	0.94	0.11	0.11	0.02	0.95	0.11	0.11	0.06	0.95	0.11	0.11	0.03	0.93
$ATE(n)$	0.10	0.11	0.04	0.94	0.10	0.11	0.02	0.90	0.10	0.11	0.06	0.94	0.10	0.11	0.03	0.91
ATT	0.14	0.14	0.04	0.94	0.14	0.14	0.02	0.94	0.14	0.14	0.05	0.93	0.14	0.14	0.02	0.94
ATE	0.12	0.13	0.03	0.93	0.12	0.13	0.01	0.94	0.12	0.13	0.04	0.95	0.12	0.13	0.02	0.93
μ_1^c	0.80	0.80	0.04	0.93	0.80	0.80	0.02	0.95	0.80	0.80	0.06	0.96	0.80	0.80	0.03	0.94
μ_1^a	0.67	0.67	0.03	0.94	0.67	0.67	0.01	0.93	0.67	0.67	0.04	0.93	0.67	0.67	0.02	0.95
μ_1^n	0.39	0.40	0.05	0.94	0.39	0.40	0.02	0.93	0.39	0.41	0.06	0.95	0.39	0.40	0.03	0.93
μ_0^c	0.64	0.64	0.05	0.95	0.64	0.64	0.02	0.94	0.64	0.64	0.07	0.93	0.64	0.64	0.03	0.93
μ_0^a	0.56	0.56	0.04	0.95	0.56	0.56	0.02	0.94	0.56	0.56	0.06	0.95	0.56	0.56	0.03	0.94
μ_0^n	0.29	0.29	0.03	0.94	0.29	0.29	0.01	0.95	0.29	0.29	0.03	0.94	0.29	0.29	0.01	0.95
p^c	0.34	0.34	0.02	0.94	0.34	0.34	0.01	0.95	0.34	0.34	0.03	0.95	0.34	0.34	0.01	0.94
p^a	0.33	0.33	0.02	0.95	0.33	0.33	0.01	0.95	0.33	0.33	0.02	0.95	0.33	0.33	0.01	0.93
p^n	0.33	0.33	0.02	0.94	0.33	0.33	0.01	0.94	0.33	0.33	0.02	0.94	0.33	0.33	0.01	0.95

Note: The values in the θ column are the true values calculated from $1,000,000$ 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. $ATE(u)$ is defined as $\mu_1^u - \mu_0^u$ for each stratum $u \in \{c, a, n, d\}$. LATE is calculated by $\frac{\mathbb{E}[Y|Z=1]-\mathbb{E}[Y|Z=0]}{\mathbb{E}[D|Z=1]-\mathbb{E}[D|Z=0]}$.

TABLE 12. Monotonicity setups

	Random assignment								ignorability							
	$n = 2000$				$n = 10000$				$n = 2000$				$n = 10000$			
	θ	$\hat{\theta}$	s.e.	cover	θ	$\hat{\theta}$	s.e.	cover	θ	$\hat{\theta}$	s.e.	cover	θ	$\hat{\theta}$	s.e.	cover
								Normal distribution								
$LATE$	2.00	-0.28	8.64	0.77	2.00	-0.06	0.88	0.27	-	-	-	-	-	-	-	-
$ATE(c)$	2.00	1.98	0.49	0.94	2.00	2.00	0.22	0.95	2.00	1.98	0.66	0.93	2.00	2.00	0.30	0.95
$ATE(a)$	1.00	1.00	0.06	0.95	1.00	1.00	0.03	0.94	1.00	1.00	0.08	0.94	1.00	1.00	0.04	0.94
$ATE(n)$	1.00	1.00	0.06	0.95	1.00	1.00	0.03	0.94	1.00	1.00	0.08	0.94	1.00	1.00	0.04	0.94
$ATE(d)$	3.00	2.99	0.47	0.95	3.00	3.00	0.21	0.93	3.00	3.04	0.65	0.95	3.00	3.00	0.29	0.94
ATT	1.55	1.54	0.27	0.94	1.55	1.55	0.12	0.95	1.55	1.53	0.36	0.93	1.55	1.54	0.16	0.95
ATE	1.70	1.69	0.18	0.95	1.70	1.70	0.08	0.95	1.70	1.70	0.24	0.94	1.70	1.70	0.11	0.95
μ_1^c	8.01	8.00	0.22	0.94	8.01	8.00	0.10	0.94	8.01	8.02	0.23	0.94	8.01	8.01	0.10	0.95
μ_1^a	4.99	5.00	0.20	0.93	4.99	5.00	0.09	0.95	4.99	5.00	0.21	0.94	4.99	5.01	0.09	0.93
μ_1^n	2.00	2.00	0.22	0.96	2.00	1.99	0.10	0.94	2.00	1.99	0.20	0.94	2.00	2.00	0.09	0.94
μ_1^d	5.00	5.00	0.42	0.96	5.00	5.00	0.19	0.94	5.00	5.05	0.61	0.94	5.00	4.99	0.27	0.93
μ_0^c	6.01	6.02	0.44	0.95	6.01	6.00	0.19	0.95	6.01	6.04	0.63	0.93	6.01	6.01	0.29	0.95
μ_0^a	3.99	4.00	0.21	0.95	3.99	4.00	0.09	0.95	3.99	4.00	0.21	0.95	3.99	4.01	0.09	0.93
μ_0^n	1.00	1.00	0.21	0.94	1.00	0.99	0.10	0.93	1.00	0.99	0.19	0.95	1.00	1.00	0.08	0.93
μ_0^d	2.00	2.01	0.22	0.95	2.00	2.00	0.10	0.95	2.00	2.01	0.21	0.95	2.00	2.00	0.10	0.94
p^c	0.30	0.30	0.01	0.95	0.30	0.30	0.01	0.93	0.30	0.30	0.02	0.95	0.30	0.30	0.01	0.94
p^a	0.25	0.25	0.01	0.93	0.25	0.25	0.01	0.93	0.25	0.25	0.02	0.94	0.25	0.25	0.01	0.95
p^n	0.25	0.25	0.01	0.94	0.25	0.25	0.01	0.93	0.25	0.25	0.01	0.94	0.25	0.25	0.01	0.94
p^d	0.20	0.20	0.01	0.94	0.20	0.20	0.01	0.95	0.20	0.20	0.02	0.94	0.20	0.20	0.01	0.95
								Bernoulli distribution								
$LATE$	0.16	-0.23	2.64	0.63	0.16	-0.18	0.11	0.07	-	-	-	-	-	-	-	-
$ATE(c)$	0.16	0.16	0.05	0.95	0.16	0.16	0.02	0.94	0.16	0.16	0.07	0.94	0.16	0.16	0.03	0.95
$ATE(a)$	0.11	0.11	0.05	0.95	0.11	0.11	0.02	0.94	0.11	0.11	0.05	0.95	0.11	0.11	0.02	0.95
$ATE(n)$	0.10	0.11	0.05	0.94	0.10	0.11	0.02	0.92	0.10	0.11	0.05	0.95	0.10	0.11	0.02	0.93
$ATE(d)$	0.33	0.33	0.07	0.93	0.33	0.33	0.03	0.94	0.33	0.33	0.10	0.95	0.33	0.33	0.04	0.94
ATT	0.14	0.14	0.04	0.95	0.14	0.14	0.02	0.94	0.14	0.14	0.05	0.95	0.14	0.14	0.02	0.94
ATE	0.17	0.17	0.03	0.94	0.17	0.17	0.01	0.93	0.17	0.17	0.04	0.95	0.17	0.17	0.02	0.94
μ_1^c	0.80	0.80	0.02	0.95	0.80	0.80	0.01	0.93	0.80	0.80	0.03	0.96	0.80	0.80	0.01	0.94
μ_1^a	0.67	0.67	0.03	0.94	0.67	0.67	0.01	0.95	0.67	0.67	0.03	0.95	0.67	0.67	0.02	0.93
μ_1^n	0.39	0.40	0.06	0.94	0.39	0.40	0.02	0.93	0.39	0.40	0.06	0.96	0.39	0.40	0.03	0.94
μ_1^d	0.61	0.61	0.06	0.95	0.61	0.61	0.03	0.95	0.61	0.62	0.09	0.94	0.61	0.61	0.04	0.93
μ_0^c	0.64	0.64	0.05	0.94	0.64	0.64	0.02	0.94	0.64	0.64	0.07	0.95	0.64	0.64	0.03	0.94
μ_0^a	0.56	0.56	0.05	0.95	0.56	0.56	0.02	0.94	0.56	0.56	0.05	0.94	0.56	0.56	0.02	0.95
μ_0^n	0.29	0.29	0.03	0.94	0.29	0.29	0.01	0.93	0.29	0.29	0.03	0.94	0.29	0.29	0.01	0.94
μ_0^d	0.28	0.29	0.03	0.94	0.28	0.29	0.01	0.95	0.28	0.29	0.03	0.94	0.28	0.29	0.02	0.95
p^c	0.30	0.30	0.01	0.94	0.30	0.30	0.01	0.94	0.30	0.30	0.02	0.95	0.30	0.30	0.01	0.94
p^a	0.25	0.25	0.01	0.96	0.25	0.25	0.01	0.96	0.25	0.25	0.01	0.95	0.25	0.25	0.01	0.95
p^n	0.25	0.25	0.01	0.94	0.25	0.25	0.01	0.94	0.25	0.25	0.02	0.95	0.25	0.25	0.01	0.93
p^d	0.20	0.20	0.01	0.94	0.20	0.20	0.01	0.96	0.20	0.20	0.02	0.93	0.20	0.20	0.01	0.95

Note: The values in the θ column are the true values calculated from $1,000,000$ 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. $ATE(u)$ is defined as $\mu_1^u - \mu_0^u$ for each stratum $u \in \{c, a, n, d\}$. LATE is calculated by $\frac{\mathbb{E}[Y|Z=1]-\mathbb{E}[Y|Z=0]}{\mathbb{E}[D|Z=1]-\mathbb{E}[D|Z=0]}$.

Table 13. Stable setups

	Random assignment								Ignorability							
	$n = 2000$				$n = 10000$				$n = 2000$				$n = 10000$			
	θ	$\hat{\theta}$	s.e.	cover	θ	$\hat{\theta}$	s.e.	cover	θ	$\hat{\theta}$	s.e.	cover	θ	$\hat{\theta}$	s.e.	cover
	Normal distribution															
$LATE$	2.00	-0.25	16.46	0.78	2.00	-0.08	0.88	0.26	-	-	-	-	-	-	-	-
$ATE(c)$	2.00	2.02	0.65	0.95	2.00	2.03	0.29	0.94	2.00	2.01	0.80	0.95	2.00	2.04	0.35	0.95
$ATE(a)$	1.00	0.98	0.25	0.94	1.00	0.99	0.11	0.95	1.00	0.97	0.39	0.95	1.00	0.98	0.15	0.95
$ATE(n)$	1.00	0.98	0.25	0.94	1.00	0.99	0.11	0.95	1.00	0.97	0.39	0.95	1.00	0.98	0.15	0.95
$ATE(d)$	3.00	3.02	0.82	0.95	3.00	3.03	0.36	0.94	3.00	3.07	1.00	0.94	3.00	3.04	0.43	0.95
ATT	1.54	1.55	0.35	0.95	1.54	1.56	0.16	0.94	1.54	1.54	0.44	0.96	1.54	1.56	0.19	0.94
ATE	1.70	1.71	0.31	0.95	1.70	1.72	0.14	0.96	1.70	1.70	0.35	0.95	1.70	1.71	0.15	0.95
μ_1^c	8.00	7.98	0.43	0.94	8.00	7.99	0.19	0.95	8.00	8.01	0.49	0.96	8.00	8.00	0.21	0.95
μ_1^a	5.00	4.97	0.49	0.95	5.00	4.98	0.22	0.93	5.00	4.98	0.58	0.96	5.00	4.97	0.24	0.95
μ_1^n	2.00	1.97	0.47	0.94	2.00	1.98	0.21	0.93	2.00	1.96	0.54	0.94	2.00	1.97	0.22	0.95
μ_1^d	5.01	5.04	0.69	0.94	5.01	5.03	0.30	0.94	5.01	5.07	0.93	0.95	5.01	5.05	0.39	0.94
μ_0^c	6.00	5.96	0.50	0.94	6.00	5.97	0.22	0.92	6.00	6.00	0.67	0.94	6.00	5.97	0.30	0.93
μ_0^a	4.00	3.98	0.46	0.94	4.00	3.99	0.20	0.94	4.00	4.00	0.46	0.95	4.00	3.99	0.19	0.95
μ_0^n	1.00	0.99	0.40	0.94	1.00	0.99	0.18	0.94	1.00	0.99	0.37	0.94	1.00	0.99	0.16	0.95
μ_0^d	2.01	2.02	0.47	0.95	2.01	1.99	0.20	0.93	2.01	2.00	0.46	0.95	2.01	2.00	0.19	0.96
p^c	0.30	0.30	0.01	0.94	0.30	0.30	0.01	0.92	0.30	0.30	0.02	0.94	0.30	0.30	0.01	0.93
p^a	0.25	0.25	0.01	0.93	0.25	0.25	0.01	0.88	0.25	0.25	0.02	0.94	0.25	0.25	0.01	0.93
p^n	0.25	0.25	0.01	0.94	0.25	0.25	0.01	0.91	0.25	0.25	0.01	0.94	0.25	0.25	0.01	0.92
p^d	0.20	0.20	0.01	0.94	0.20	0.20	0.01	0.91	0.20	0.20	0.02	0.94	0.20	0.20	0.01	0.95
	Bernoulli distribution															
$LATE$	0.16	-0.22	1.27	0.64	0.16	-0.19	0.11	0.05	-	-	-	-	-	-	-	-
$ATE(c)$	0.16	0.16	0.08	0.95	0.16	0.16	0.03	0.95	0.16	0.16	0.10	0.96	0.16	0.16	0.04	0.94
$ATE(a)$	0.11	0.11	0.09	0.94	0.11	0.11	0.04	0.94	0.11	0.11	0.10	0.95	0.11	0.11	0.04	0.95
$ATE(n)$	0.10	0.11	0.09	0.94	0.10	0.11	0.04	0.94	0.10	0.11	0.10	0.94	0.10	0.11	0.04	0.95
$ATE(d)$	0.33	0.33	0.11	0.94	0.33	0.33	0.05	0.94	0.33	0.33	0.14	0.94	0.33	0.33	0.06	0.94
ATT	0.14	0.14	0.06	0.96	0.14	0.14	0.02	0.95	0.14	0.14	0.07	0.94	0.14	0.14	0.03	0.95
ATE	0.17	0.17	0.06	0.96	0.17	0.17	0.02	0.95	0.17	0.17	0.06	0.94	0.17	0.17	0.03	0.95
μ_1^c	0.80	0.80	0.05	0.95	0.80	0.80	0.02	0.94	0.80	0.80	0.06	0.94	0.80	0.80	0.03	0.95
μ_1^a	0.67	0.67	0.06	0.94	0.67	0.67	0.03	0.94	0.67	0.67	0.07	0.93	0.67	0.67	0.03	0.94
μ_1^n	0.39	0.40	0.11	0.95	0.39	0.40	0.05	0.94	0.39	0.40	0.12	0.95	0.39	0.40	0.05	0.94
μ_1^d	0.61	0.62	0.09	0.94	0.61	0.61	0.04	0.95	0.61	0.62	0.12	0.93	0.61	0.62	0.05	0.94
μ_0^c	0.64	0.64	0.06	0.95	0.64	0.64	0.03	0.94	0.64	0.64	0.08	0.95	0.64	0.64	0.04	0.95
μ_0^a	0.56	0.56	0.09	0.94	0.56	0.56	0.04	0.96	0.56	0.56	0.10	0.94	0.56	0.56	0.04	0.94
μ_0^n	0.29	0.29	0.05	0.96	0.29	0.29	0.02	0.95	0.29	0.29	0.06	0.94	0.29	0.29	0.03	0.96
μ_0^d	0.29	0.29	0.06	0.94	0.29	0.28	0.03	0.95	0.29	0.29	0.07	0.95	0.29	0.29	0.03	0.95
p^c	0.30	0.30	0.01	0.95	0.30	0.30	0.01	0.93	0.30	0.30	0.02	0.94	0.30	0.30	0.01	0.94
p^a	0.25	0.25	0.01	0.94	0.25	0.25	0.01	0.90	0.25	0.25	0.02	0.94	0.25	0.25	0.01	0.92
p^n	0.25	0.25	0.01	0.94	0.25	0.25	0.01	0.91	0.25	0.25	0.01	0.94	0.25	0.25	0.01	0.93
p^d	0.20	0.20	0.01	0.94	0.20	0.20	0.01	0.92	0.20	0.20	0.02	0.93	0.20	0.20	0.01	0.94

Note: The values in the θ column are the true values calculated from $1,000,000$ 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. $ATE(u)$ is defined as $\mu_1^u - \mu_0^u$ for each stratum $u \in \{c, a, n, d\}$. $LATE$ is calculated by $\frac{\mathbb{E}[Y|Z=1]-\mathbb{E}[Y|Z=0]}{\mathbb{E}[D|Z=1]-\mathbb{E}[D|Z=0]}$.

TABLE 14. Unstable setups

Appendix D. Detail of empirical illustrations

D.1. Randomized encouragement design with Y^{pre}.

Thornton (2008). We define the assignment indicator Z based on the monetary incentives offered: subjects who received no incentive are assigned to $Z = 0$, while those who received any positive-valued incentive are assigned to $Z = 1$. The treatment status D is an indicator for whether an individual learned their HIV results. We focus on two outcomes Y measured approximately two months after the results were available: an indicator for reported condom purchases and an indicator for reported sexual activity. The corresponding pre-treatment outcomes Y^{pre} are defined using baseline survey data as indicators for reported condom use and sexual activity in the year prior to the study.

Following Thornton (2008), the analysis for the "Purchase Condom" outcome is restricted to the sample of respondents in the Balaka and Rumphi districts who tested for HIV, had age data, were reinterviewed in 2005, and reported having sex in 2004. For the "Having Sex" outcome, the sample is defined similarly but excludes the restriction of having had sex in 2004, as this variable is used as Y^{pre} for this analysis. For both analyses, the sample is restricted to individuals with no missing values for Z, D, Y, and Y^{pre}. Bootstrap resampling is clustered by villages.

Gerber et al. (2009). The assignment indicator Z is defined as 1 for subjects in the treatment groups (free subscription to either the Washington Post or Washington Times) and 0 for the control group. The treatment status D is an indicator for whether an individual reported receiving either newspaper at the time of the follow-up survey. The outcome Y is an indicator for self-reported voter turnout in the 2005 election, and the pre-treatment outcome Y^{pre} is an indicator for voter turnout in the 2004 election.

The sample is restricted to individuals who completed the follow-up survey and have no missing values for Z, D, Y, and Y^{pre}.

A limitation in this setup is that the treatment variable D was measured in a follow-up survey conducted one week after the outcome, Y (the election), occurred. Consequently, this measure may not perfectly capture the newspaper readership that actually influenced voting behavior. Furthermore, as Gerber et al. (2009) note, "the wording of the question" may have caused variance in how respondents interpreted it. Therefore, while the validity and interpretation of our estimate under this setup warrant further discussion, this re-analysis serves primarily as a methodological illustration of our proposed technique.

The baseline survey, which provides a history of past voting behavior, allows us to assess the parallel trends assumption (Assumption 3 (i)). Figure 2 visualizes these pre-treatment turnout trends for always-takers and never-takers. As the figure illustrates, the two groups exhibit clearly parallel trends in their past voting behavior, providing strong support for the parallel trends assumption. The steep increase in self-reported turnout shown in the figure reflects the electoral cycle in Virginia: an off-year election (2001), a midterm election (2002), and a presidential election (2004).

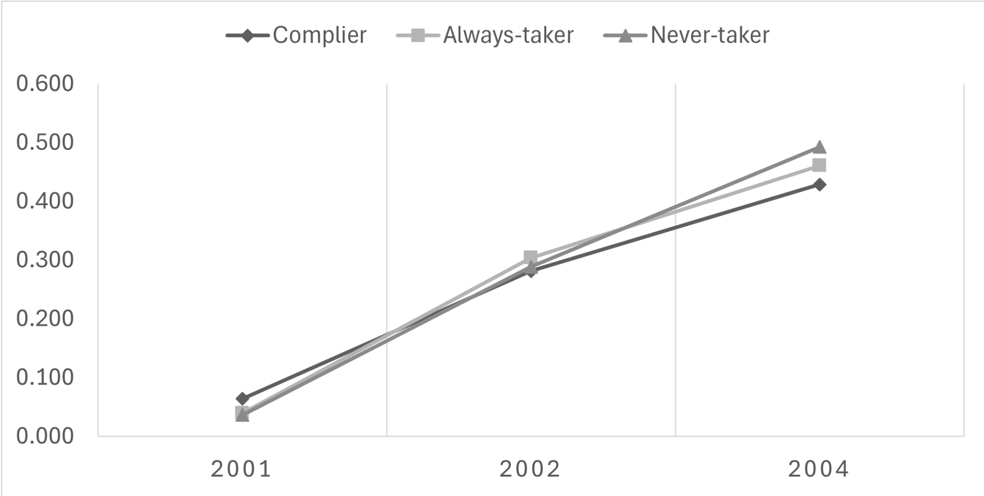

FIGURE 2. Pre-Treatment Trend of the Outcome Variable 'Voted'

Beam (2016). In our re-analysis, the assignment indicator Z is defined as 1 for subjects who received a voucher to encourage job fair attendance and 0 otherwise. The treatment status D is an indicator for whether an individual actually attended the job fair. We focus on outcomes related to migration preparation. The first outcome Y is an indicator for whether the respondent plans to look for work abroad in the next six months, with the corresponding pre-treatment outcome Y^{pre} being an indicator for planning to apply abroad at baseline. The second outcome is an indicator for having a current passport at follow-up, with the corresponding Y^{pre} being an indicator for currently having a passport at baseline.

The sample consists of individuals who completed the follow-up survey. There are no missing values for Z, D, Y, and Y^{pre}. Bootstrap resampling is clustered at the neighborhood level.

D.2. **Oregon Health Insurance Experiment with D^{pre}.** Following Finkelstein et al. (2012), our inverse probability weighting estimation includes dummy variables for the interaction between survey wave and the number of household members listed in the lottery as covariates. We do not employ a multiply robust estimator as the sample size is insufficient to reliably construct an outcome model. Also in line with Finkelstein et al. (2012), we utilize the provided survey weights in our analysis. Bootstrap resampling is clustered by household unit, as defined by the lottery list.

The estimated principal strata probabilities are shown in Table 15 and 16.

D.3. **Oregon Health Insurance Experiment with Y^{pre} and D^{pre}.** In our multiply robust estimation, we follow Taubman et al. (2014) and include covariates for the number of household members listed in the lottery and the corresponding pre-treatment outcome for each main outcome. For stability in the outcome model estimation, the number of household members is included as a continuous variable rather than as a set of dummy variables. The outcome model $m^{\text{pre}}_{(z,d,d')}(X; \gamma^{\text{pre}})$, which predicts the pre-treatment outcome itself, excludes the pre-treatment outcome as a covariate. Bootstrap resampling is clustered by household unit, as defined by the lottery list.

	p^c	p^a	p^n	p^d
Health care utilization:				
Prescription drugs currently	0.366 *** (0.006)	0.065 *** (0.002)	0.489 *** (0.006)	0.081 *** (0.003)
Outpatient visit	0.364 *** (0.005)	0.060 *** (0.002)	0.500 *** (0.005)	0.076 *** (0.002)
ER visits	0.364 *** (0.005)	0.060 *** (0.002)	0.500 *** (0.005)	0.076 *** (0.002)
Inpatient hospital admissions	0.365 *** (0.005)	0.060 *** (0.002)	0.500 *** (0.005)	0.076 *** (0.002)
Blood cholesterol checked	0.365 *** (0.005)	0.059 *** (0.002)	0.501 *** (0.005)	0.075 *** (0.002)
Blood tested for high blood sugar/diabetes	0.364 *** (0.005)	0.060 *** (0.002)	0.500 *** (0.005)	0.076 *** (0.002)
Financial strain:				
Any out of pocket medical expenses	0.364 *** (0.005)	0.060 *** (0.002)	0.500 *** (0.005)	0.076 *** (0.002)
Owe money for medical expenses currently	0.365 *** (0.005)	0.060 *** (0.002)	0.499 *** (0.005)	0.076 *** (0.002)
Borrowed money or skipped other bills to pay medical bills	0.364 *** (0.005)	0.060 *** (0.002)	0.500 *** (0.005)	0.076 *** (0.002)
Refused treatment because of medical debt	0.364 *** (0.005)	0.060 *** (0.002)	0.499 *** (0.005)	0.077 *** (0.002)

TABLE 15. Estimates and standard errors. Standard errors in parentheses. $^{*}p < 0.05, ^{**}p < 0.01, ^{***}p < 0.001$.

The estimated principal strata probabilities are shown in Table 17.

D.4. **Application in marketing using Y^{pre} and D^{pre}.** The estimated principal strata probabilities are shown in Table 18.

	p^c	p^a	p^n	p^d
Health:				
Self-reported health good/very good/excellent	0.365 ***	0.060 ***	0.500 ***	0.076 ***
	(0.005)	(0.002)	(0.005)	(0.002)
Self-reported health not poor	0.365 ***	0.060 ***	0.500 ***	0.076 ***
	(0.005)	(0.002)	(0.005)	(0.002)
Health about the same or gotten better	0.365 ***	0.060 ***	0.500 ***	0.076 ***
	(0.005)	(0.002)	(0.005)	(0.002)
# of days physical health good, past 30 days	0.364 ***	0.060 ***	0.501 ***	0.076 ***
	(0.005)	(0.002)	(0.005)	(0.002)
# of days physical or mental health did not impair usual activity, past 30 days	0.363 ***	0.059 ***	0.502 ***	0.076 ***
	(0.005)	(0.002)	(0.005)	(0.002)
# of days mental health good, past 30 days	0.364 ***	0.060 ***	0.501 ***	0.076 ***
	(0.005)	(0.002)	(0.005)	(0.002)
Did not screen positive for depression, last two weeks	0.364 ***	0.060 ***	0.500 ***	0.076 ***
	(0.005)	(0.002)	(0.005)	(0.002)
Health (Mechanisms):				
Have usual place of clinic-based care	0.363 ***	0.057 ***	0.508 ***	0.073 ***
	(0.005)	(0.002)	(0.006)	(0.002)
Have personal doctor	0.364 ***	0.060 ***	0.500 ***	0.076 ***
	(0.005)	(0.002)	(0.005)	(0.002)
Got all needed medical care, last six months	0.364 ***	0.060 ***	0.501 ***	0.076 ***
	(0.005)	(0.002)	(0.005)	(0.002)
Got all needed drugs, last six months	0.363 ***	0.059 ***	0.502 ***	0.076 ***
	(0.005)	(0.002)	(0.005)	(0.002)
Didn't use ER for nonemergency, last six months	0.364 ***	0.060 ***	0.500 ***	0.076 ***
	(0.005)	(0.002)	(0.005)	(0.002)
Quality of care received last six months good/very good/excellent	0.391 ***	0.086 ***	0.434 ***	0.089 ***
	(0.006)	(0.003)	(0.006)	(0.003)
Very happy or pretty happy	0.365 ***	0.060 ***	0.500 ***	0.076 ***
	(0.005)	(0.002)	(0.005)	(0.002)

TABLE 16. Estimates and standard errors. Standard errors in parentheses. $^{*}p < 0.05, ^{**}p < 0.01, ^{***}p < 0.001$.

	p^c	p^a	p^n	p^d
Percent with any visits	0.333 ***	0.065 ***	0.514 ***	0.088 ***
	(0.005)	(0.002)	(0.005)	(0.002)
Number of visits	0.333 ***	0.065 ***	0.515 ***	0.087 ***
	(0.005)	(0.002)	(0.005)	(0.002)

TABLE 17. Estimates and standard errors. Standard errors in parentheses. $^{*}p < 0.05, ^{**}p < 0.01, ^{***}p < 0.001$.

p^c	p^a	p^n	p^d
0.015 ***	0.001 ***	0.971 ***	0.013 ***
(0.000)	(0.000)	(0.001)	(0.000)

TABLE 18. Estimates and standard errors. Standard errors in parentheses. $^{*}p < 0.05, ^{**}p < 0.01, ^{***}p < 0.001$.