Difference-in-differences Design with Outcomes Missing Not at Random

Sooahn Shin*†

This version: September 16, 2024

Abstract

This paper addresses one of the most prevalent problems encountered by political scientists working with difference-in-differences (DID) design: missingness in panel data. A common practice for handling missing data, known as complete case analysis, is to drop cases with any missing values over time. A more principled approach involves using nonparametric bounds on causal effects or applying inverse probability weighting based on baseline covariates. Yet, these methods are general remedies that often underutilize the assumptions already imposed on panel structure for causal identification. In this paper, I outline the pitfalls of complete case analysis and propose an alternative identification strategy based on principal strata. To be specific, I impose parallel trends assumption within each latent group that shares the same missingness pattern (e.g., always-respondents, if-treated-respondents) and leverage missingness rates over time to estimate the proportions of these groups. Building on this, I tailor Lee bounds, a well-known nonparametric bounds under selection bias, to partially identify the causal effect within the DID design. Unlike complete case analysis, the proposed method does not require independence between treatment selection and missingness patterns, nor does it assume homogeneous effects across these patterns.

Keywords Difference-in-differences, Causal Inference, Missingness, Panel Data, Principal Strata

*Ph.D. Candidate, Department of Government and Institute for Quantitative Social Science, Harvard University. Email: sooahnshin@g.harvard.edu. URL: https://sooahnshin.com.

†I am grateful to Kosuke Imai, Matthew Blackwell, Naijia Liu, Soichiro Yamauchi, Imai Research Group, Andrew Q. Philips, Anton Strezhnev, and the participants at the 81st MPSA Annual Conference and 120th APSA Annual Meeting and Exhibition for their helpful comments and suggestions. All errors are my own.

1 Introduction

Difference-in-differences (DID) design is a quasi-experimental method in social science widely used to estimate causal effects of a treatment on an outcome variable using repeated observations of units over time. The DID design is particularly useful when the treatment is not randomly assigned, and the researcher is concerned with unobserved time-invariant confounder. Yet, one of the most prevalent problems encountered by researchers working with panel data is missingness. For example, this problem is evident in DID studies that utilize survey data to measure outcome variables, where respondent's non-response is a frequent concern. Chiu et al. (2023), in their extensive review and replication of articles from three leading political science journals using observational panel data with binary treatments, highlighted this issue with unbalanced panels. They especially emphasized that the missingness pattern is appeared to be nonrandom or extremely prevalent in some studies.

Despite the prevalence of missing data in panel studies, most methodological work presumes balanced panels without missing data. When faced with the methodological challenge of addressing potential bias due to missing data, researchers have often resorted to complete case analysis (listwise deletion) or imputation methods. However, these methods are not without their own limitations. Complete case analysis can lead to biased estimates when the missingness is not completely at random (MCAR) or missing at random (MAR). Imputation methods can also be problematic when the missingness is not at random (MNAR) or the covariates are also susceptible to missingness. Alternatively, when the quantity of interest is a causal estimand, a more principled approach to address missing data is to use nonparametric bounds of causal effects (Horowitz and Manski, 2000; Zhang and Rubin, 2003; Imai, 2008; Lee, 2009) or inverse probability weighting using baseline covariates. However, these methods are general remedies that under-utilize the assumptions already imposed on panel structure for causal identification.

This implies that the intersection of causal inference, panel data, and missing values presents a unique challenge in social science studies that remains underexplored. Recently, there have been attempts in disciplines adjacent to social science to address attrition bias by using parallel trends assumptions and the changes-in-changes approach (Ghanem et al., 2022; Dukes, Richardson and Tchetgen Tchetgen, 2022). Ghanem et al. (2022) extends the changes-

in-changes condition so that the distribution of unobserved heterogeniety to be stable across time within treatment-response subpopulations. Using this assumption, they identify the average treatment effect of the treated-respondents and also show that the average treatment effect can be identified with an additional assumption that the distribution is homogeneous across treatment-response subpopulations. Dukes, Richardson and Tchetgen Tchetgen (2022) consider two alternative strategies, one based on the parallel trends assumption and the other based on the 'bespoke instrumental variable' approach, yet their approach is limited to randomized experiments.

In this paper, I provide an alternative approach of addressing missing outcome variable in panel data with DID design for observational studies. My discussion aligns with the recent studies but also differs from them by setting the DID design and parallel trends assumption for *causal identification* at the center and exploring the assumptions and data structure within the DID design to address the *missing data problem*. Specifically, I answer the following questions using the DID design and the principal stratification framework: Under which extension of parallel trends assumptions can we justify the complete case analysis? What are the main identification challenges with missing data in the DID design? How can we identify the average treatment effect for treated (ATT) using auxiliary variables from the panel data that offer additional information yet have not been explored?

What must not be overlooked is that missing indicator in this setup is a post-treatment intermediate variable. In this vein, I proceed to introduce principal stratification (Frangakis and Rubin, 2002)–namely, always-respondents, if-treated-respondents, if-control-respondents, and never-respondents–to induce parallel trends assumptions conditioning on these latent groups that shares the same missingness pattern. I interpret the complete case analysis under this framework and discuss the main identification challenges with missing data: (1) potential dependence between selection into treatment with principal strata and (2) heterogenous effect across principal strata. I then extend the DID design using auxiliary variables (e.g. outcome variables in multiple pretreatment periods), and use its response indicator as an instrumental variable for identifying ATT. The essential intuition behind this approach is that the response indicator of the auxiliary variables can be used as an instrumental variable that affects the time trend of the outcome variable only through the treatment selection and the missingness of the

post-treatment outcome variable. Lastly, I propose alternative approaches with the principal strata specific parallel trends assumption to partially identify the pricipal strata specific ATT. To be specific, I impose parallel trends assumption within each principal stratum and leverage missingness rates over time to estimate the proportions of these groups. Building on this, I tailor Lee bounds (Lee, 2009), a well-known nonparametric bounds under selection bias, to partially identify the causal effect within the DID design.

2 Problems of the Standard Approach

In this section, I review the parallel trends assumptions under which the standard DID design with complete case analysis can be justified. Using the principal stratification framework, I describe the limitations of this standard methodology and discuss the main identification challenges with missing data in the DID design.

2.1 Motivating Applications

To illustrate the problem of missing data in the DID design, I revisit two application studies using two-periods DID design with survey data. In the first application, I revisit Sexton and Zürcher (2024) which studies how small development aid projects affect the public perception and attitudes toward government. Specifically, the authors are interested in the impacts of German development aid during 2017–18 on subsequent political attitudes in northern Afghanistan. The study uses two waves (2016 and 2018) of geocoded public opinion survey data with five different main outcome variables. As shown in Table 1, the missingness of the survey responses in the second wave is substantial, and the pattern of missingness appears to be different across outcome variables. Particularly, the missingness ratio differ in time trend (*before-after*), difference between treated and control groups (*treated-control*), and its interaction (*differece-in-differences*).

In the second study, I revisit Bisgaard and Slothuus (2018) which examines the effects of elite partisan cues on economic perception. The study uses five waves of panel surveys collected from 2010 to 2011 in Denmark to track public opinion on economic issues. After the second wave of survey data, the Center-Right government in Denmark dramatically changed its partisan cue on the severity of the public budget deficit, which led to a change in the

Wave	Treatment Group	Afghanistan Right Direction?	Confidence in President	Local Government Confidence	National Governemnt Good Job	Sympathy for Insurgents
2016	Control	4.52	0.37	18.57	0.47	4.63
2016	Treated	4.74	0.55	69.83	0.14	7.01
2018	Control	8.14	1.83	14.60	0.22	3.12
2018	Treated	8.28	0.96	64.83	1.14	2.34

Table 1: Missingness of Survey Responses in Sexton and Zürcher (2024) in Percentage Points

economic perception of incumbent supporters according the their findings. The missingness of the survey responses in this study is shown in Table 2. Given a non-ignorable amount of missingness, the authors provided additional regression analysis with the second and third waves, and concluded the missingness does *not* appear to be systematically different across treatment groups and time periods, conditioning on prior perceptions of the national economy.

Wave	Period	Treated Group (Incumbent Supporters)	Control Group (Opposition Supporters)
1	Pre-treatment	12.82	16.42
2	Pre-treatment	39.16	42.26
3	Post-treatment	44.87	45.72
4	Post-treatment	47.90	48.40
5	Post-treatment	54.31	54.45

Table 2: Missingness of Survey Responses in Bisgaard and Slothuus (2018) in Percentage Points

Several questions arise from these examples, particularly related to the unique features of panel data. What does the trend of missingness within each treatment group imply for potential bias in the DID estimates? Can we directly compare such trends between different treatment groups, and if they are similar across groups, would the DID estimate from complete case analysis be unbiased? Answering these questions necessitates exploring diverse variants of canonical parallel trends assumptions and their substantive implications. As I will discuss in the following sections, the short answer is no. It requires an additional assumption regarding the parallel trends of the outcome variable between respondents and nonrespondents within each treatment group, which may restrict the heterogeneity of the treatment effect.

Another important aspect of these questions is how we define the groups in terms of miss-

ingness and how the composition of these groups relates to causal identification. In particular, it is crucial to recognize that the missingness of the outcome variable is a post-treatment variable, meaning there exists a latent group of units with different missingness patterns under different treatment statuses. This motivates the need for principal stratification, which examines the potential missingness patterns of the outcome variable under different treatment statuses. Given that we are interested in observational studies where the treatment is not randomly assigned, the composition of these latent groups may vary across treatment groups. This may be due to selection bias, where the treatment selection is correlated with missingness, heterogenous treatment effect across these latent groups, or both.

Lastly, it is worth noting that panel data provides additional information that can be crucial for identifying causal effects. At a minimum, researchers always have access to the missingness rate of the outcome variable over time, which can be used in causal identification. In more favorable cases, researchers may also have access to auxiliary variables from the panel data, such as other outcome measures or multiwave data, which can further aid in identifying the causal effect. In this paper, I propose a novel identification strategy that combines the principal strata framework with these auxiliary variables to address the missing data problem in the DID design.

2.2 Setup and Notation

I consider a two-period DID design with binary treatment and missingness in the outcome variable. Let D_i be a binary treatment group indicator of unit i, where $D_i = 1$ for the treated group and 0 for the control group. Let Y_{it} be the outcome variable of unit i at time $t = 1, 2$, where time 1 is the pre-treatment period and 2 is the post-treatment period. I assume the consistency and no anticipation assumptions for the potential outcomes:

$$Y_{it} = D_i Y_{it}(1) + (1 - D_i) Y_{it}(0) \text{ for } t = 1, 2 \qquad \text{(Consistency)}$$

$$Y_{it}(1) = Y_{it}(0) \text{ for } t = 1 \qquad \text{(No anticipation of treatment)}$$

Let R_i denote a binary response indicator at time $t = 2$. $R_i = 1$ if Y_{i2} is observed, and 0 if it is missing. Here, $R_i(d)$ is the potential response indicator if $D_i = d$. I assume the same consistency and no anticipation assumptions for R_i.

One example of this setting is a study using a DID design (or two-way fixed effects) where the data comes from a survey conducted in two waves, with attrition observed in the second wave. Another potential example of missing outcomes is the "don't know" response to survey questions. A common practice in this case is to treat "don't know" as missing, resulting in a similar setup to survey data with attrition. Although I illustrate the problem of missing data in the DID design using survey attrition to facilitate understanding, the proposed methodology can be applied to other types of missing data and is not limited to surveys.

Based on the joint distribution of the response indicator and treatment group indicator we can define the following groups:

$$(R_i, D_i) = \begin{cases} (1,1) & \text{"treated-respondents"} \\ (0,1) & \text{"treated-nonrespondents"} \\ (1,0) & \text{"control-respondents"} \\ (0,0) & \text{"control-nonrespondents"} \end{cases}$$

One critical limitation of such a group is that it does not account for the fact that the response itself is a post-treatment variable. This group is considered crude because it only captures the realized response under a given treatment assignment and does not consider the counterfactual response (i.e., whether these individuals would have responded if they had been selected into the other treatment group). For example, "treated-respondents" in the first motivating example correspond to those people whose districts received aid and responded to the survey question afterward. We do not know whether this group of people would have also responded to the question if their district had not received such aid.

Alternatively, we can introduce the principal strata framework (Frangakis and Rubin, 2002) in this setup based on the joint distribution of the potential outcomes of the response indicator. Let $S_i = (R_i(1), R_i(0))$ denote a principal strata define as below. For example,

$S_i = (1, 1)$ implies that unit i's outcome is always observed, no matter the treatment status.

$$S_i \equiv (R_i(1), R_i(0)) = \begin{cases} (0,0) & \text{“never-respondents”} \\ (1,0) & \text{“if-treated-respondents”} \\ (0,1) & \text{“if-control-respondents”} \\ (1,1) & \text{“always-respondents”} \end{cases}$$

Note that this can be viewed as a latent pre-treatment covariate, distinct from the observed post-treatment response indicator R_i. With this, we can consider the joint distribution of this principal stratum and treatment group (e.g. never-respondents who are treated), as opposed to the realized response and treatment group (e.g. treated-respondents, which is a mix of never-respondents and if-treated-respondents within the treated group). We use the term “respondents” for ease of interpretation, but it may not necessarily refer to survey respondents. In general, this term should be understood as a group of units that exhibit different missingness patterns under two possible treatment regimes.

The principal strata framework is widely utilized in the causal inference literature, particularly for examining issues of truncation by death and noncompliance. In the context of truncation by death in medical studies, for example, “always-respondents” corresponds to “survivors,” who would have always survived no matter what the medical treatment (e.g. an uptake of a medicine) was. In the context of noncompliance, “always-respondents” correspond to “always-taker,” who would have always taken the medicine no matter what the treatment assignment, an encouragement to take the medicine, was.

Similar to these issues, the potential outcome of the response indicator in our context is significant because it defines latent subgroups of units with substantively different characteristics. For example, in the first motivating application, if-treated-respondents represent a group of people who respond to these sensitive survey questions only if their district received aid. In contrast, if-control-respondents are those who respond to the survey questions only if their district did not receive the aid. One can imagine that there might be a systematic reason for these opposing behaviors, which could be related to the impact of aid. In the second motivating application, always-respondents are individuals who consistently respond to survey

questions about economic issues, regardless of partisan cues. In contrast, never-respondents are those who never respond to the survey questions on economic issues. These two groups may possess distinct characteristics, such as varying levels of political engagement or interest in economic matters, which could correlate with partisanship.

Two main issues with different principal strata are that (1) the proportions of these latent groups may vary between the treated and control groups, and (2) the treatment effect may differ across these latent groups. I will formally demonstrate this intuition in this section, where I discuss the identification challenges within this principal strata framework. To be specific, in this paper, I treat the principal strata as one of the conditioning variables when imposing parallel trends assumptions for causal identification, a topic that will be elaborated upon in this section. For simplicity, I assume there is no missing data in the first wave (alternatively, one could assume MCAR for the pre-treatment outcome missingness for now). However, this assumption will be relaxed in the following section with the proposed solution to the missing data problem.

2.3 Complete Case Analysis and Parallel Trends Assumptions

To begin with, we discuss how the general practice of complete case analysis can be justified under the parallel trends assumptions. In the DID design, researchers are interested in identifying the ATT defined as

$$\text{ATT} = \mathbb{E}[Y_{i2}(1) - Y_{i2}(0) \mid D_i = 1]$$

We employ the canonical parallel trends assumption to identify the counterfactual outcome $Y_{i2}(0)$ for treated units.

Assumption 1 (*Parallel trends*).

$$\mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 1] = \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 0]$$

With this assumption and the consistency assumption, we have

$$\text{ATT} = \mathbb{E}[Y_{i2} - Y_{i2}(0) \mid D_i = 1]$$

$$= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1] - \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 1]$$
$$= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 0] \tag{2.1}$$

The final expression is not identified with missing data in the outcome variable. Upon the missingness in panel data with a randomized treatment, Dukes, Richardson and Tchetgen Tchetgen (2022) consider imposing the parallel trends assumption between the respondents and nonrespondents, for the time trend under each treatment status respectively.

Assumption 2 (*Parallel trends of observed outcome between respondents and nonrespondents*).

$$\mathbb{E}[Y_{i2}(d) - Y_{i1} \mid R_i = 1, D_i = d] = \mathbb{E}[Y_{i2}(d) - Y_{i1} \mid R_i = 0, D_i = d]$$

for $d \in \{0, 1\}$.

What this assumption implies is as follows. We first consider the expression under the control group, $d = 0$.

$$\mathbb{E}[\underbrace{Y_{i2}(0) - Y_{i1}}_{\text{time trend}} \mid R_i = 1, D_i = 0] = \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid R_i = 0, D_i = 0]$$

It assumes that the time trend of outcome, $Y_{i2}(0) - Y_{i1}$, is on average parallel between control respondents and control nonrespondents. For example, in the motivating application, within the districts that received aid, the average change in perception over time among respondents is equal to that among nonrespondents. This can be considered a variant of the canonical parallel trends assumption, where we are interested in DID of control-respondent and control-nonrespondents groups. Since the assumption is made on the time trends, not the outcome levels, it allows for the missingness to be correlated with the outcome level. For instance, if respondents who are more positive towards the government are more likely to respond to the survey, yet the change in perception over time remains constant, the assumption may still hold. Additionally, if there are multiple pre-treatment periods without missingness, this assumption can be indirectly tested by examining the parallel trends of the outcome between control-respondent and control-nonrespondents groups.

Next, let's consider Assumption 2 under the treated group, $d = 1$.

$$\mathbb{E}[\underbrace{(Y_{i2}(1) - Y_{i2}(0))}_{\text{causal effect}} + \underbrace{(Y_{i2}(0) - Y_{i1})}_{\text{time trend}} \mid R_i = 1, D_i = 1]$$
$$= \mathbb{E}[(Y_{i2}(1) - Y_{i2}(0)) + (Y_{i2}(0) - Y_{i1}) \mid R_i = 0, D_i = 1]$$

This assumption is more restrictive than the former, as it requires the parallel trends of the *sum* of causal effect and time trend of the outcome between treated-respondents and treated-nonrespondents. For example, within the districts that received the aid, the change in perception pertaining to the aid and a common time shock among the respondents is on average equivalent to that of the nonrespondents. Hypothetically, this can be violated if the missingness is correlated with the treatment effect size, holding the time trend constant. For example, if the respondents who received the aid and became more positive towards the government are more likely to respond to the survey, the assumption may be violated.

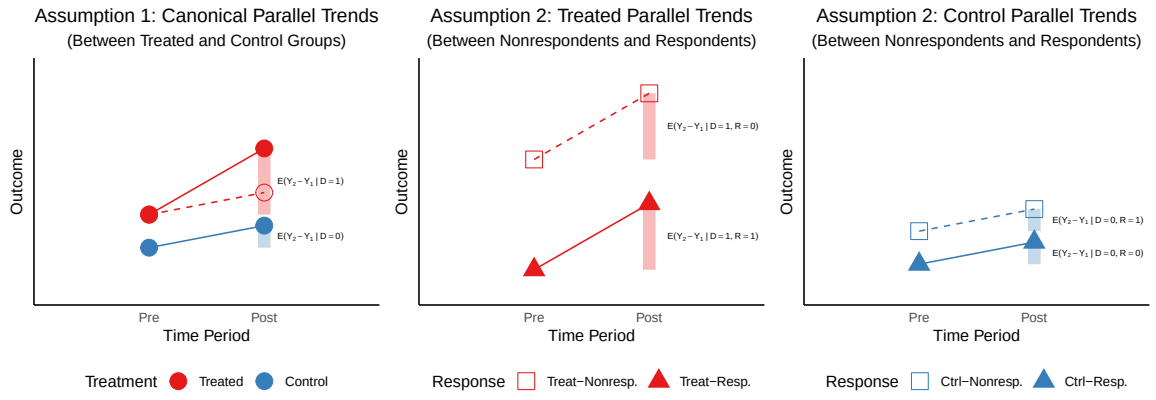

Figure 1: Illustration of the Parallel Trends Assumption 1 and Assumption 2. (Left): Canonical parallel trends assumption between treated and control groups. Each dot represents the expected mean outcome for the treated (red) and control (blue) groups. Since these groups are a mix of respondents and nonrespondents, each quantity in post-treatment period is *not* directly observed. (Middle): Parallel trends assumption between respondents and nonrespondents within the treated group. Each dot represents the expected mean outcome for the treated-respondent (red square) and treated-nonrespondent (red triangle) groups. The difference in the outcome of treated-respondents at the bottom is observed. (Right): Parallel trends assumption between respondents and nonrespondents within the control group. Each dot represents the expected mean outcome for the control-respondent (blue square) and control-nonrespondent (blue triangle) groups. The difference in the outcome of control-respondents at the bottom is observed.

Under Assumption 1 and Assumption 2, the ATT can be identified by the DID estimand using complete case analysis. The intuition is as follows: As illustrated in the left panel of Figure 1, the canonical parallel trends assumption (Assumption 1) allows us to identify the ATT with the DID estimand. Here, the red vertical line represents the difference in outcome for the treated group ($\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1]$), and the blue vertical line represents the difference in outcome for the control group ($\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 0]$). By taking the difference between these two lines, we can identify the ATT as shown in Equation 2.1. Yet, due to the missingness in the outcome variable, these quantities are not directly observed. Under Assumption 2, however, we can identify each of these quantities by the difference in outcome of the treated-respondents and control-respondents, respectively, as shown in the middle and right panels of Figure 1. For example, two red vertical lines in the middle panel represent the difference in outcome for the treated-respondents ($\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, R_i = 1]$) and treated-nonrespondents ($\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, R_i = 0]$), respectively, and are identical to the red vertical line in the left panel by Assumption 2. Since the red vertical line at the bottom is observed ($\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, R_i = 1]$), we can identify the difference in outcome for the treated group ($\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1]$). Similarly, we can identify that of the control group as well. I formally state this identification result in the following proposition.

Proposition 1 (*Identification of ATT with complete case analysis*). *Under Assumption 1 (Parallel trends) and Assumption 2 (Parallel trends of observed outcome between respondents and nonrespondents), we have*

$$ATT = \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, R_i = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 0, R_i = 1]$$

Proof. See Appendix A.1 for the proof. □

In simpler terms, if we adopt the canonical parallel trends assumption (Assumption 1), and additionally assume that there is no selection bias in the missingness of the before-after difference in the outcome variable within each treatment group (Assumption 2), then the ATT can be identified using DID estimator with complete case analysis. As previously discussed, the parallel trends assumption between treated-respondents and treated-nonrespondents is more restrictive than the canonical parallel trends assumption. This is particularly the case

when heterogeneous treatment effects related to missingness are present. In the subsequent section, we will delve further into the main identification challenges associated with missing data in the DID design.

2.4 Identification Challenges with Missing Data

In a general setup, when the parallel trends assumption is less plausible, the researcher may alternatively consider conditional parallel trends assumption where we assume parallel trends for each subgroup defined by pretreatment covariates. The rationale behind this is that the parallel trends assumption will become more plausible when the baseline covariates are balanced between the treated and control groups. Given that the missingness is a post-treatment binary variable, we can consider the parallel trends assumptions conditioning on the principal strata, which can be viewed as a conditional parallel trends assumption, conditioning on a latent subgroup: never-respondents, if-treated-respondents, if-control-respondents, and always-respondents.

Assumption 3 (*Principal strata parallel trends*).

$$\mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 0, S_i = s] = \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 1, S_i = s]$$

for $s \in \{(0,0),(0,1),(1,0),(1,1)\}$.

This can be understood as a weaker assumption than the canonical parallel trends assumption (Assumption 1). It is worth pointing out that Assumption 3 does not imply Assumption 1 in general, and vice versa.

Remark 1 (*Principal strata parallel trends does not imply canonical parallel trends*). *Assumption 3 does not imply Assumption 1, and vice versa. One special case where Assumption 3 does imply Assumption 1 is when the principal strata is independent of treatment selection, i.e.* $S_i \perp\!\!\!\perp D_i$. *Under Assumption 3, by the law of iterated expectation*

$$\begin{aligned}\mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 1] &= \sum_s \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 1, S_i = s] \Pr(S_i = s \mid D_i = 1) \\ &= \sum_s \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 0, S_i = s] \Pr(S_i = s \mid D_i = 1)\end{aligned}$$

If $\Pr(S_i = s \mid D_i = 1) = \Pr(S_i = s \mid D_i = 0)$ *then the last line is equal to* $\mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 0]$*, which implies Assumption 1.*

Intuitively speaking, even if the time trend of outcome is parallel between treated and control groups within a specific principal strata (e.g. always-respondents), unless the distribution of principal strata is equivalent between treated and control groups, the canonical parallel trends assumption may not hold. For example, in the second motivating application, if the incumbent supporters are more likely to respond to the questions on economic issues regardless of the partisan cues, the proportion of always-respondents may be higher in the treated group than the control group, which may potentially violate the canonical parallel trends assumption.

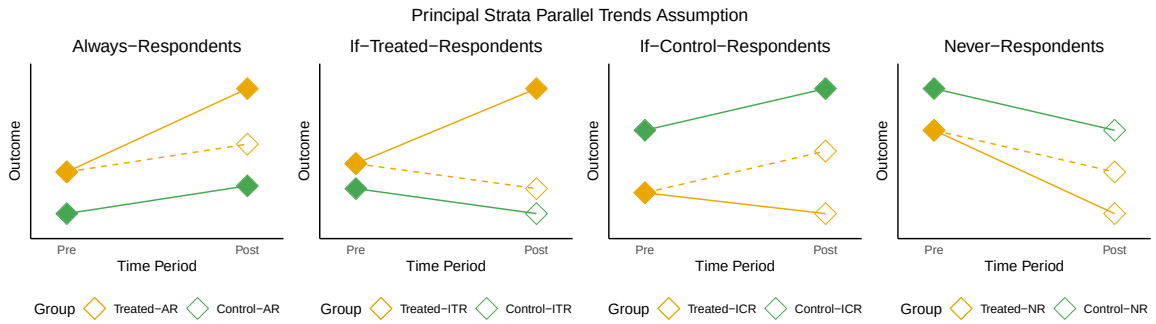

Figure 2: Toy Example of the Principal Strata Parallel Trends Assumption 3. Each dot represents the expected mean outcome for the treated (yellow) and control (green) groups within each principal strata. Filled dots represent the observed quantity and empty dots represent the unobserved quantity. Dashed lines represent the outcome trends under no administration of treatment for the treated group, i.e. line connecting $\mathbb{E}[Y_{i1} \mid D_i = 1, S_i = s]$ and $\mathbb{E}[Y_{i2}(0) \mid D_i = 1, S_i = s]$. Note that the time trend and causal effect are allowed to be heterogenous across principal strata.

Now, we discuss and clarify the main identification challenges with outcome MNAR using this principal strata parallel trends assumption.

Proposition 2 (*Decomposition of ATT with outcome MNAR*). *Under Assumption 3, we have*

$$\begin{aligned}
&\mathbb{E}[Y_{i2}(1) - Y_{i2}(0) \mid D_i = 1] \\
&= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, R_i = 1] \Pr(R_i = 1 \mid D_i = 1)
\end{aligned}$$

$$
\begin{aligned}
&- \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 0, R_i(1) = 1, R_i = 1] \Pr(R_i = 1, R_i(0) = 1 \mid D_i = 1) \\
&- \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 0, R_i(1) = 1, R_i = 0] \Pr(R_i = 1, R_i(0) = 0 \mid D_i = 1) \\
&+ \mathbb{E}[Y_{i2} - Y_{i2}(0) \mid D_i = 1, R_i = 0, R_i(0) = 0] \Pr(R_i = 0, R_i(0) = 0 \mid D_i = 1) \\
&+ \mathbb{E}[Y_{i2} - Y_{i2}(0) \mid D_i = 1, R_i = 0, R_i(0) = 1] \Pr(R_i = 0, R_i(0) = 1 \mid D_i = 1)
\end{aligned}
$$

where the term in blue is not identified since principal strata is latent, and the term in red is not identified due to the missingness. This implies that the identification of ATT is not possible without further assumptions on:

1. *Dependence of selection into treatment with principal strata*
2. *Heterogeneous effect across principal strata*

Proof. See Appendix A.2 for the proof. □

This decomposition is intuitive in the sense that it follows a natural logic starting from the respondents data and then correcting the potential bias due to the missingness. First, suppose that we naively compute $\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, R_i = 1] \Pr(R_i = 1 \mid D_i = 1)$, which is before-after difference of treated-respondents weighted by the proportion of respondents. Then, we need to adjust for the time trend for the respondents: always-respondents and if-treated-respondents. This corresponds to the second and third lines, using the principal strata parallel trends assumption. Also, we need to incorporate the effect for the nonrespondents: never-respondents and if-control-respondents. This corresponds to the last two lines, which cannot be identified with the principal strata parallel trends assumption since the outcome is missing.

In other words, if we assume that the missingness pattern is independent of selection into treatment (i.e., $S_i \perp\!\!\!\perp D_i$) *and* that the treatment effect is homogeneous across principal strata within the treated group (i.e., $\mathbb{E}[Y_{i2}(1) - Y_{i2}(0) \mid D_i = 1, S_i = s]$ is constant for all s), then the ATT can be identified using the principal strata parallel trends assumption and complete case analysis as before. While these assumptions may be plausible in some applications, they are generally strong and may not hold. In particular, the assumption of a homogeneous treatment effect across principal strata within the treated group is restrictive and may be violated under missing not at random (MNAR).

In the subsequent section, we will propose an alternative approach to identify the ATT using auxiliary variables from the panel data without imposing the homogeneous treatment effect assumption across principal strata. Note that this can be viewed as MNAR in the missing data literature, where the missingness is dependent on the unobserved outcome variable. Specifically, I first adapt an instrumental variable approach from Dukes, Richardson and Tchetgen Tchetgen (2022), where the response indicator of the auxiliary variables is used as an instrumental variable for the missingness of the outcome variable and thus point identification of ATT is possible. I also propose a partial identification approach based on Lee (2009) using pre-treatment missingness trend. This allows us to partially identify the ATT for always-respondents, in a more general setup of the dependence between the missingness and the treatment selection.

3 The Proposed Methodology

In this section, I propose two alternative approaches that do not require the homogeneous treatment effect assumption across principal strata. The first approach is based on the instrumental variable (IV) method, where the key idea is to utilize an IV that is associated with baseline missingness probability but not with the magnitude of the bias. I motivate this IV approach using randomized incentives for participation in the survey and also consider the response indicator of the auxiliary variables. The second approach is based on the partial identification method, where I tailor Lee bounds (Lee, 2009) to address the missing data problem in the DID design. Based on the parallel trends assumptions of response rate over time, I show that the ATT for always-respondents can be partially identified without requiring the homogeneous treatment effect assumption across principal strata or the independence assumption between missingness and treatment selection.

3.1 Identification of ATT with Baseline Response Indicator

We first consider the point identification of the ATT using an IV that captures the baseline probability of response under a canonical parallel trends assumption. This approach is motivated by 'bespoke instrument variable (IV)' from Dukes, Richardson and Tchetgen Tchetgen (2022), in which they introduce a special type of IV that can be leveraged to identify the

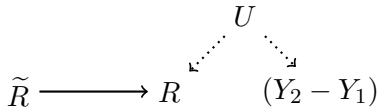

Figure 3: A DAG example of Randomized Incentive ($\widetilde{R}$) and Post-treatment Response Indicator (R). Omitted D for simplicity.

selection bias due to missingness (also see Tchetgen Tchetgen and Wirth, 2017; Richardson and Tchetgen Tchetgen, 2022). Here, I adapt the basic idea of (1) introducing an IV that satisfies a certain exclusion restriction and (2) assuming that the bias is homogeneous across the subgroups defined by the IV. In Dukes, Richardson and Tchetgen Tchetgen (2022), the focus was on a general framework for identifying selection bias due to missingness, specifically the difference $\mathbb{E}[Y_{i2} - Y_{i1} \mid R_i = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid R_i = 0]$, under the assumption of a randomized experiment. However, this paper considers observational studies, where we employ parallel trends assumptions to identify the ATT. Our primary focus is on identifying selection bias due to missingness within each treatment group, i.e. $\delta_d \equiv \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, R_i = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, R_i = 0]$ for $d = 0, 1$.

To illustrate this IV approach, hypothetically assume that we offer a randomized incentive for survey participation. Let $\widetilde{R}_i$ denote the binary indicator of whether unit i received the incentive. We consider the following set of assumptions on this indicator variable.

Assumption 4 (*IV approach*). *An indicator variable $\widetilde{R}_i \in \{0, 1\}$ satisfies the following set of assumptions:*

1. ***Relevance to missingness**: $\widetilde{R}_i$ is relevant to the missingness of the post-treatment outcome.*

$$\Pr(R_i = 0 \mid D_i = d, \widetilde{R}_i = 0) \neq \Pr(R_i = 0 \mid D_i = d, \widetilde{R}_i = 1)$$

   for $d = 0, 1$.

2. ***Parallel trends of observed outcome**: The observed trend of the outcome is parallel between two groups defined by $\widetilde{R}_i$.*

$$\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, \widetilde{R}_i = 0] = \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, \widetilde{R}_i = 1]$$

   for $d = 0, 1$.

3. ***Bias homogeneity****: The bias due to post-treatment missingness in the time trend of the outcome is homogeneous between two groups defined by* $\widetilde{R}_i$*.*

$$\begin{aligned}\delta_d &= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, \widetilde{R}_i = 0, R_i = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, \widetilde{R}_i = 0, R_i = 0] \\ &= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, \widetilde{R}_i = 1, R_i = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, \widetilde{R}_i = 1, R_i = 0]\end{aligned}$$

where

$$\delta_d \equiv \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, R_i = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, R_i = 0]$$

for $d = 0, 1$*.*

The first assumption, **relevance to missingness** applies if $\widetilde{R}_i$ is correlated with the missingness of the post-treatment outcome. For example, if the incentive encourages the respondents to participate in the survey, then the missingness of the outcome may be negatively correlated with the receipt of the incentive. Intuitively speaking, if we consider the post-treatment missingness to be a combination of MAR (in terms of baseline response probability) and MNAR mechanisms, $\widetilde{R}_i$ helps us control for the MAR component of the missingness. Unlike others, this assumption can be empirically tested.

The second assumption, **parallel trends of observed outcome**, corresponds to Assumption 2 (Parallel trends of observed outcome between respondents and nonrespondents) with regards to $\widetilde{R}_i$ instead of R_i. As discussed in a previous section, this assumption can be violated if $\widetilde{R}_i$ is correlated with the treatment effect size while the time trend remains constant. This suggests that a crucial criterion for $\widetilde{R}_i$ as an IV is the expectation of homogeneous effects between groups defined by this indicator. Under the incentive scenario, this assumption is plausible since the incentive is randomized.

The last assumption, **bias homogeneity**, is the core assumption for using $\widetilde{R}_i$ to correct the bias. This assumption holds if the bias resulting from post-treatment missingness is consistent across individuals who received the incentive and those who did not, within each treatment group. This assumption might be violated if there is an unmeasured confounder, denoted as U_i, that interacts with the outcome and $\widetilde{R}_i$. For instance, in an additive outcome model, the presence of an interaction term $U_i \times \widetilde{R}_i$ suggests that the bias in post-treatment outcomes between two groups defined by $\widetilde{R}_i$ could be heterogeneous.

Intuitively speaking, when we consider all these assumptions together, we can deduce that any observed differences between the $\{\widetilde{R}_i = 0\}$ and $\{\widetilde{R}_i = 1\}$ groups are due to their association with post-treatment missingness, which induces bias. The identification result is formally stated in the following theorem.

Theorem 1 (*Identification of ATT using IV*). *Under Assumption 1 and 4, the ATT can be identified using* $\widetilde{R}_i$ *as an IV.*

$$
\begin{aligned}
&ATT \\
&= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, R_i = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 0, R_i = 1] \\
&\quad + \frac{\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 1, R_i = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 0, R_i = 1]}{\Pr(R_i = 0 \mid D_i = 1, \widetilde{R}_i = 0) - \Pr(R_i = 0 \mid D_i = 1, \widetilde{R}_i = 1)} \\
&\qquad \times \Pr(R_i = 0 \mid D_i = 1) \\
&\quad - \frac{\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 0, \widetilde{R}_i = 1, R_i = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 0, \widetilde{R}_i = 0, R_i = 1]}{\Pr(R_i = 0 \mid D_i = 0, \widetilde{R}_i = 0) - \Pr(R_i = 0 \mid D_i = 0, \widetilde{R}_i = 1)} \\
&\qquad \times \Pr(R_i = 0 \mid D_i = 0)
\end{aligned}
$$

Proof. Proof is straightforward by Lemma 1 in Appendix B.1. □

So far, I have used a randomized incentive for survey participation as a potential choice of an IV. Alternatively, one might consider using the response indicator of auxiliary variables as an IV. These auxiliary variables can be constructed from other pre-treatment variables, such as those survey questions unrelated to the main outcome of interest. For instance, in the first motivating example, we could use other outcome variables (e.g., "Afghanistan right direction?") from the 2016 wave as auxiliary variables. However, it is crucial to ensure that the auxiliary variable satisfies the assumptions in Assumption 4. For example, in the initial motivating application, if "local government confidence" is the primary outcome of interest and there is concern that the missingness of the auxiliary variable correlates with the treatment effect size, it would be advisable to choose another outcome variable that is not directly associated with local governance.

In Appendix B.1, I generalize the setup to allow for missingness in pre-treatment periods and discuss the identification of the ATT using the IV approach. In Appendix C, I also present

a variant of the IV method discussed above, where the assumption of parallel trends in the observed outcome for the auxiliary variable is relaxed. Instead, this variant employs multiple auxiliary variables that exhibit consistent bias between respondents and nonrespondents. This approach offers a more flexible framework, accommodating scenarios where the missingness of the auxiliary variables may correlate with the treatment effect size.

Despite its potential, the proposed IV approach has several limitations. First and foremost, the bias homogeneity assumption is crucial for the identification of the ATT, yet its validity is difficult to verify. Tchetgen Tchetgen and Wirth (2017) provides an example of semiparametric shared parameter model that satisfies the bias homogeneity assumption. A future research direction could be to show which types of models in our setup satisfy this assumption. For example, a separable model of the post-treatment missingness in terms of $\widetilde{R}_i$ and an unmeasured confounder U_i could be a potential candidate [1]. Alternatively, as mentioned in Tchetgen Tchetgen and Wirth (2017), one may consider a hypothesis test of $\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, \widetilde{R}_i = 1, R_i = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, \widetilde{R}_i = 1, R_i = 0] = 0$ without making such bias homogeneity assumption.

More importantly, we apply the canonical parallel trends assumption (Assumption 1) in this approach instead of the principal strata parallel trends assumption (Assumption 3). It's crucial to highlight that, should we choose to adopt assumption Assumption 3 in lieu of Assumption 1, it may be required to introduce an additional assumption such as the selection into treatment is independent of the principal strata (see Remark 2 for more details). In the subsequent section, I modify this assumption by allowing the selection into treatment to depend on the principal strata, and propose a partial identification approach to estimate the ATT for the subgroup of always-respondents.

[1] For instance, consider the following semiparametric model:

$$
\begin{aligned}
\mathbb{E}[Y_2 - Y_1 \mid U, D] &= f(D) + U \\
\operatorname{logit} \Pr(R = 1 \mid \widetilde{R}, U, D) &= \underbrace{g_1(D, \widetilde{R}) + g_2(D)U}_{\text{A separable model}} \\
\operatorname{logit} \Pr(\widetilde{R} = 1 \mid \widetilde{U}, D) &= h(D)\widetilde{U} \\
U \mid R = 0, \widetilde{R}, D &\sim m(D, \widetilde{R}) + \zeta
\end{aligned}
$$

3.2 Partial Identification of ATT for Always-Respondents with Response Parallel Trends

In this section, I introduce a partial identification of the ATT specifically for the subgroup of always-respondents, leveraging the trend observed in pre-treatment missingness. The partial identification follows the 'trimming bounds' approach of Zhang and Rubin (2003) and Lee (2009), where the bounds are derived by considering the extreme cases based on the observed distribution. Here, we are interested in the following quantity:

$$\text{ATT-AR} = \mathbb{E}[Y_{i2}(1) - Y_{i2}(0) \mid D_i = 1, R_{i2}(1) = 1, R_{i2}(0) = 1]$$

ATT-AR is an average treatment effect among always-respondents who are selected into the treatment. As discussed in the previous section, if researchers believe that the missingness is MNAR, but are reluctant to make additional assumptions about effect heterogeneity across principal strata, then the ATT-AR is the only quantity that can be partially identified with minimal assumptions including Assumption 3 (Principal strata parallel trends). When the primary interest is in the ATT, this approach serves as a valuable starting point to understand the potential biases introduced by missingness and to develop a corresponding sensitivity analysis.

Since we utilize pre-treatment missingness in this approach, we slightly modify the notation to distinguish the missingness indicator at different time points. Specifically, we use R_{it} instead of R_i to denote the response indicator of the outcome variable at time t. That is, $R_{it} = 1$ if Y_{it} is observed, and 0 if it is missing, for $t = 1, 2$. We also assume that we are interested in the ATT for the population who responded to the survey at time 1, while omitting $R_{i1} = 1$ in the conditioning set for simplicity of notation. Alternatively, we could consider an exclusion restriction type of assumption for pre-treatment missingness and time trends of the outcome, as described in Assumption 8 in Appendix B.1.

Trimming bounds. We first review the main idea of the trimming bounds approach. Let $\pi_{r_1,r_0}(d) \equiv \Pr(R_{i2}(1) = r_1, R_{i2}(0) = r_0 \mid D_i = d)$. For now, suppose that we identified the proportion of principal strata in the treated group. We will discuss more about this in the later part.

Under Assumption 3 (Principal strata parallel trends), we have

$$\text{ATT-AR} = \mathbb{E}[Y_{i2}-Y_{i1} \mid D_i = 1, R_{i2}(1) = 1, R_{i2}(0) = 1]+\mathbb{E}[Y_{i2}-Y_{i1} \mid D_i = 0, R_{i2}(1) = 1, R_{i2}(0) = 1]$$

where the right hand side is the DID among always respondents. Our goal here is to bound this quantity using the observed data:

$$\begin{aligned} f_1(y) &\equiv \Pr(Y_{i2} - Y_{i1} \leq y \mid D_i = 1, R_{i2} = 1) \\ &= \frac{\pi_{11}(1)}{\pi_{11}(1) + \pi_{10}(1)} \Pr(Y_{i2} - Y_{i1} \leq y \mid D_i = 1, R_{i2}(1) = 1, R_{i2}(0) = 1) \\ &\quad + \frac{\pi_{10}(1)}{\pi_{11}(1) + \pi_{10}(1)} \Pr(Y_{i2} - Y_{i1} \leq y \mid D_i = 1, R_{i2}(1) = 1, R_{i2}(0) = 0) \end{aligned}$$

Suppose $\frac{\pi_{11}(1)}{\pi_{11}(1)+\pi_{10}(1)} = p/100$. Following 'trimming bounds' approach of Zhang and Rubin (2003) and Lee (2009), we can get the lower bound of $\mathbb{E}[Y_{i2}-Y_{i1} \mid D_i = 1, R_{i2}(1) = 1, R_{i2}(0) = 1]$ by 'trimming' lower $p\%$ of treated-respondent group and taking average. That is,

$$\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, R_{i2}(1) = 1, R_{i2}(0) = 1] \geq \int_{-\infty}^{q_1^{\text{low}}(f_1)} y \frac{f_1(y)}{\int_{-\infty}^{q_1^{\text{low}}(f_1)} f_1(y)dy} dy$$

where $q_1^{\text{low}}(f_1)$ is the $\frac{\pi_{10}(1)}{\pi_{11}(1)+\pi_{10}(1)}$ quantile of $f_1(\cdot)$. Similarly, we can get the upper bound as follows:

$$\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, R_{i2}(1) = 1, R_{i2}(0) = 1] \leq \int_{q_1^{\text{high}}(f_1)}^{\infty} y \frac{f_1(y)}{\int_{q_1^{\text{high}}(f_1)}^{\infty} f_1(y)dy} dy$$

where $q_1^{\text{high}}(f_1)$ is the $\frac{\pi_{11}(1)}{\pi_{11}(1)+\pi_{10}(1)}$ quantile of $f_1(\cdot)$.

Using the same strategy, we can bound $\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 0, R_{i2}(1) = 1, R_{i2}(0) = 1]$:

$$\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 0, R_{i2}(1) = 1, R_{i2}(0) = 1] \geq \int_{-\infty}^{q_0^{\text{low}}(f_0)} y \frac{f_0(y)}{\int_{-\infty}^{q_0^{\text{low}}(f_0)} f_0(y)dy} dy$$

$$\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 0, R_{i2}(1) = 1, R_{i2}(0) = 1] \leq \int_{q_0^{\text{high}}(f_0)}^{\infty} y \frac{f_0(y)}{\int_{q_0^{\text{high}}(f_0)}^{\infty} f_0(y)dy} dy$$

where $f_0 \equiv \Pr(Y_{i2} - Y_{i1} \leq y \mid D_i = 0, R_{i2} = 1)$, $q_0^{\text{low}}(f_0)$ is the $\frac{\pi_{01}(0)}{\pi_{11}(0)+\pi_{01}(0)}$ quantile of $f_0(\cdot)$,

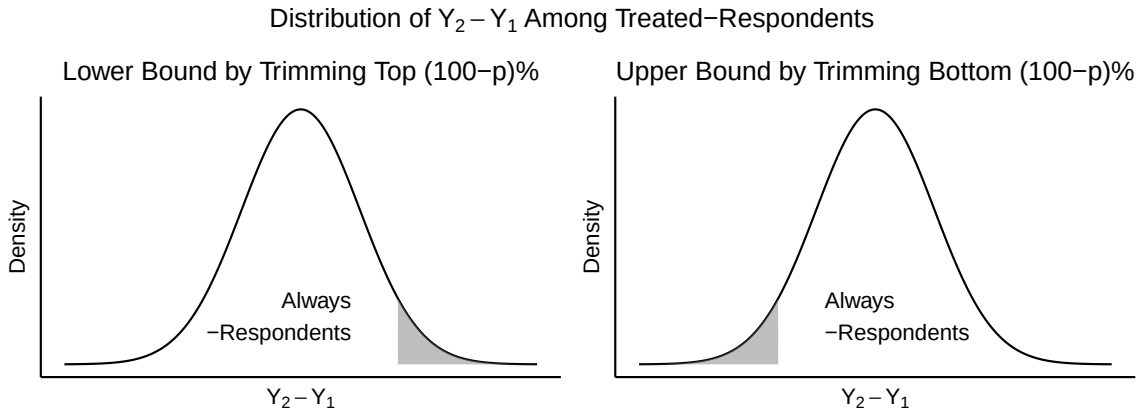

Figure 4: Illustration of the trimming bounds approach for the ATT-AR. The un-shaded area represents the lower and upper bounds of $\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, R_{i2}(1) = 1, R_{i2}(0) = 1]$.

$q_0^{\text{high}}(f_0)$ is the $\frac{\pi_{11}(0)}{\pi_{11}(0)+\pi_{01}(0)}$ quantile of $f_0(\cdot)$. Consequently, our remaining task involves identifying the proportion of principal strata within the treated group. This analysis diverges from the standard principal strata analysis, which typically focuses on randomized experiments. Here, we are delving into observational studies and employing a DID design instead of relying on the assumption of ignorability (refer to Wang, Zhou and Richardson, 2017). In this paper, we introduce the parallel trends assumption in the missingness of outcomes as a strategy to address the challenges associated with identifying the proportion of principal strata in the treated group.

Principal strata proportion. In this paper, we consider two approaches for identifying the principal strata proportion within the treated group: one that incorporates a monotonicity assumption and another that does not. We first start with the method that assumes monotonicity alongside parallel trends in the response indicator.

Assumption 5 (*Monotonicity*).

$$R_{i2}(1) \geq R_{i2}(0), \quad \forall i$$

The monotonicity assumption suggests that the treatment effect on the response indicator for any individual is always non-negative. This assumption is reasonable in certain contexts. For example, consider our first motivating study where we are interested in the effect of aid

on the public perception toward the government. If this aid positively affects individuals' perceptions, it is plausible to expect a higher response rate in the post-treatment survey under treatment compared to the control. This is based on the premise that individuals with negative perceptions are more hesitant to participate in the survey. Thus, a positive treatment effect on perception would likely increase the likelihood of response, reflecting the monotonicity stated above. With this assumption, we have

$$\Pr(R_{i2}(1) = 1, R_{i2}(0) = 1 \mid D_i = 1) = \Pr(R_{i2}(0) = 1 \mid D_i = 1)$$
$$\Pr(R_{i2}(1) = 1, R_{i2}(0) = 1 \mid D_i = 0) = \Pr(R_{i2}(0) = 1 \mid D_i = 0) = \Pr(R_{i2} = 1 \mid D_i = 0)$$

Thus, we only need to identify $\Pr(R_{i2}(0) = 1 \mid D_i = 1)$ for always respondents in the treated group. To this end, we introduce the following assumption.

Assumption 6 (*Parallel trends of missingness*).

$$\mathbb{E}[R_{i2}(0) - R_{i1}(0) \mid D_i = 1] = \mathbb{E}[R_{i2}(0) - R_{i1}(0) \mid D_i = 0]$$

With this assumption, we can identify $\Pr(R_{i2}(0) = 1 \mid D_i = 1) = \Pr(R_{i2} = 1 \mid D_i = 0) - \Pr(R_{i1} = 1 \mid D_i = 0) + \Pr(R_{i1} = 1 \mid D_i = 1)$.

Proposition 3 (*Identification of principal strata proportion with monotonicity*). *Under Assumption 5 and 6, we can identify the following:*

$$\begin{aligned}
\pi_{11}(1) &= \Pr(R_{i2} = 1 \mid D_i = 0) - \Pr(R_{i1} = 1 \mid D_i = 0) \\
&\quad + \Pr(R_{i1} = 1 \mid D_i = 1) \\
\pi_{10}(1) &= \Pr(R_{i2} = 1 \mid D_i = 1) - \Pr(R_{i2} = 1 \mid D_i = 0) \\
&\quad - \{\Pr(R_{i1} = 1 \mid D_i = 1) - \Pr(R_{i1} = 1 \mid D_i = 0)\} \\
\pi_{11}(0) &= \Pr(R_{i2} = 1 \mid D_i = 0) \\
\pi_{01}(0) &= 0
\end{aligned}$$

In certain scenarios, the monotonicity assumption might not be applicable. Take, for instance, our second motivating example where the treatment effect—altering cues may not

uniformly result in an increased response rate in the post-treatment survey. In such instances, one may consider the following assumption.

Assumption 7 (*Equivalence of ATT and ATC on missingness*).

$$\mathbb{E}[R_{i2}(1) - R_{i2}(0) \mid D_i = 1] = \mathbb{E}[R_{i2}(1) - R_{i2}(0) \mid D_i = 0]$$

Proposition 4 (*Identification of counterfactual response without monotonicity*). *Under Assumption 6 and 7 we have:*

$$\Pr(R_{i2}(0) = 1 \mid D_i = 1) = \Pr(R_{i2} = 1 \mid D_i = 0) - \Pr(R_{i1} = 1 \mid D_i = 0) + \Pr(R_{i1} = 1 \mid D_i = 1)$$
$$\Pr(R_{i2}(1) = 1 \mid D_i = 0) = \Pr(R_{i2} = 1 \mid D_i = 1) - \Pr(R_{i1} = 1 \mid D_i = 1) + \Pr(R_{i1} = 1 \mid D_i = 0)$$

Partial identification of ATT-AR. Combining these identification results, we can bound ATT-AR.

Theorem 2 (*Partial identification of ATT-AR*). *Under Assumption 3 (Principal strata parallel trends) and Assumption 6 (Parallel trends of missingness), ATT-AR is partially identified:*

1. *With Assumption 5 (Monotonicity),*

$$\mathsf{LB}_{y_2-y_1,1} - \int_{-\infty}^{\infty} y f_0(y) dy \leq \textit{ATT-AR} \leq \mathsf{UB}_{y_2-y_1,1} - \int_{-\infty}^{\infty} y f_0(y) dy$$

2. *Without Assumption 5 (Monotonicity), but instead with Assumption 7 (Equivalence of ATT and ATC on missingness)*

$$\mathsf{LB}_{y_2-y_1,1} - \mathsf{UB}_{y_2-y_1,0} \leq \textit{ATT-AR} \leq \mathsf{UB}_{y_2-y_1,1} - \mathsf{LB}_{y_2-y_1,0}$$

where $\mathsf{LB}_{y_2-y_1,d}$ *and* $\mathsf{UB}_{y_2-y_1,d}$ *are the lower and upper bounds of* $\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, R_{i2}(1) = 1, R_{i2}(0) = 1]$.

$$\mathsf{LB}_{y_2-y_1,d} = \int_{-\infty}^{q_d^{low}(f_d)} y \frac{f_d(y)}{\int_{-\infty}^{q_d^{low}(f_d)} f_d(y) dy} dy$$
$$\mathsf{UB}_{y_2-y_1,d} = \int_{q_d^{high}(f_d)}^{\infty} y \frac{f_d(y)}{\int_{q_d^{high}(f_d)}^{\infty} f_d(y) dy} dy$$

where $f_d(y) \equiv \Pr(Y_{i2} - Y_{i1} \leq y \mid D_i = d, R_{i2} = 1)$, $q_d^{low}(f_d)$ *is bottom* $\frac{\pi_{11}(d)}{\Pr(R_{i2}=1|D_i=d)}$ *quantile of* $f_d(\cdot)$ *and* $q_d^{high}(f_d)$ *is top* $\frac{\pi_{11}(d)}{\Pr(R_{i2}=1|D_i=d)}$ *quantile of* $f_d(\cdot)$.

As noted earlier, the results in Theorem 2 provide bounds on the ATT-AR, accommodating both the dependence between treatment selection and principal strata, as well as heterogeneous treatment effects across principal strata. This approach is particularly useful when the missingness rate is moderate, raising concerns about potential bias in the ATT while the proportion of always-respondents remains substantial.

4 Empirical Application

In this section, I revisit the study by Sexton and Zürcher (2024) to illustrate the proposed method. The paper examines the impact of small development aid projects on public perception and attitudes toward the government. The study employs a DID design, aggregating data at the village cluster level to address inferential challenges ("spatial spillovers and incorrect standard errors"). The aim here is to demonstrate the application of the proposed method rather than to replicate the original study in its entirety. Therefore, for the purpose of this illustration, I simplify the data structure and employ a canonical two-period DID estimator.[2]

	$\pi_{11}(1)$		$\pi_{11}(0)$		ATT-AR			
Outcome	LB	UB	LB	UB	LB	UB	DID^1	DID^2
Afghanistan Right Direction?	0.83	0.92	0.84	0.92	-0.24	0.15	0.05 (0.04)	-0.04
Confidence in President	0.97	0.98	0.97	0.98	0.03	0.09	-0.18 (0.13)	0.05
National Gov. Good Job	0.99	0.99	0.98	0.99	0.03	0.05	-0.2 (0.12)	0.04
Local Gov. Confidence	0.00	0.34	0.72	0.85	-1	1	-0.29 (0.12)	-0.04
Sympathy for Insurgents	0.92	0.94	0.97	0.97	-0.02	0.10	0.11 (0.04)	0.04

Table 3: Estimates of Bounds of ATT-AR following Theorem 2. LB = lower bound, UB = upper bound, $\pi_{11}(1)$ = proportion of always-respondents in the treated group, $\pi_{11}(0)$ = proportion of always-respondents in the control group, DID^1 = the DID estimates (clustered by village) from the original study with standard errors in parentheses, DID^2 = two-period DID estimates.

Table 3 shows the partial identification of ATT-AR for the five outcome variables considered in the original study. The table presents the lower and upper bounds of ATT-AR, along

[2]The results presented in this section are provided for illustrative purposes only and should not be interpreted as supporting substantive conclusions.

with the DID estimates. There are two main takeaways here: across the outcome variables, except for "Local Government Confidence," the proportion of always-respondents is about 0.8 to 0.9 in both treatment and control groups. This result is valuable for researchers as ATT-AR provides a useful reference for ATT with the correction of bias, given a large proportion of this latent subgroup. Particularly for "Confidence in President" and "National Government Good Job," where the missingness ratio was relatively small (see Table 1), the proportion of always-respondents is close to 1. In contrast, for "Local Government Confidence," the proportion of always-respondents in the control group is about 0.8, while it is close to 0 in the treatment group. This is intuitive, given the fact that the missingness rate is about 70% and 65% for the treated group in pre- and post-treatment surveys, respectively. Secondly, the bounds of ATT-AR are comparatively tight, except for "Local Government Confidence," where the bounds range from -1 to 1, given the large gap between the treated and control groups.

Outcome	Bias Corrected DID	DID^1	DID^2
Confidence in President	0.57	-0.18 (0.13)	-0.04
National Gov. Good Job	0.24	-0.2 (0.12)	0.00
Sympathy for Insurgents	0.94	0.11 (0.04)	-0.05

Table 4: Estimates of ATT following Theorem 1. Bias Corrected DID = estimates using response indicator of "Employment opportunity (self-reported)" from 2016 (pre-treatment) wave as IV, DID^1 = the DID estimates (clustered by village) from the original study with standard errors in parentheses, DID^2 = two-period DID estimates.

Table 4 presents the ATT estimates obtained using the IV method outlined in Theorem 1. These estimates were derived using the self-reported "Employment opportunity" response indicator from the 2016 (pre-treatment) wave as an instrumental variable. The findings indicate that the difference between the bias-corrected DID estimates and the standard DID estimates is proportional to the missingness ratio observed in the post-treatment survey.

5 Concluding Remarks

Missingness in panel data is a prevalent issue that has not received sufficient attention in the context of DID studies. In this paper, I address the challenges posed by missing data

in DID analyses by exploring different variants of the parallel trends assumption and additional sources of information, such as the trend of missingness rates. Employing the principal strata framework, I identify two primary challenges when outcomes are Missing Not At Random (MNAR): (1) the potential dependence of treatment selection on principal strata and (2) heterogeneous effects across principal strata. To overcome these challenges, I propose two alternative strategies with weaker assumptions: (1) partial identification of the ATT for always-respondents and (2) an IV method that employs baseline missingness indicators as instruments.

The paper also suggests several avenues for extending these methods. For example, the nonparametric identification of ATT-AR could be achieved by adopting the assumptions from Wang, Zhou and Richardson (2017). Additionally, integrating the principal strata framework into the IV approach could provide a way to account for the dependency between treatment selection and principal strata. Exploring partial identification through the bracketing relationship among multiple auxiliary variables, as inspired by Ye et al. (2023), offers another potential extension. Moreover, generalizing the approach to accommodate staggered adoption and multiple treatment groups represents a promising direction for future research. Lastly, addressing missingness in covariates within DID studies is also worth investigating (see Appendix D for a relevant discussion).

References

Bisgaard, Martin and Rune Slothuus. 2018. "Partisan Elites as Culprits? How Party Cues Shape Partisan Perceptual Gaps." *American Journal of Political Science* 62(2):456–469.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12349*

Chiu, Albert, Xingchen Lan, Ziyi Liu and Yiqing Xu. 2023. "What to do (and not to do) with causal panel analysis under parallel trends: Lessons from a large reanalysis study." *arXiv preprint arXiv:2309.15983* .

Ding, Peng and Jiannan Lu. 2016. "Principal Stratification Analysis Using Principal Scores." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 79(3):757–777.
URL: *https://doi.org/10.1111/rssb.12191*

Dukes, Oliver, David Richardson and Eric Tchetgen Tchetgen. 2022. "Alternative approaches for analysing repeated measures data that are missing not at random." *arXiv preprint arXiv:2207.11561* .

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

Ghanem, Dalia, Sarojini Hirshleifer, Desire Kedagni and Karen Ortiz-Becerra. 2022. "Correcting Attrition Bias using Changes-in-Changes." *arXiv preprint arXiv:2203.12740* .

Horowitz, Joel L and Charles F Manski. 2000. "Nonparametric analysis of randomized experiments with missing covariate and outcome data." *Journal of the American statistical Association* 95(449):77–84.

Imai, Kosuke. 2008. "Sharp bounds on the causal effects in randomized experiments with "truncation-by-death"." *Statistics & Probability Letters* 78(2):144–149.
URL: *https://www.sciencedirect.com/science/article/pii/S0167715207002052*

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

Richardson, David B and Eric J Tchetgen Tchetgen. 2022. "Bespoke instruments: a new tool for addressing unmeasured confounders." *American journal of epidemiology* 191(5):939–947.

Sexton, Renard and Christoph Zürcher. 2024. "Aid, Attitudes, and Insurgency: Evidence from Development Projects in Northern Afghanistan." *American Journal of Political Science* 68(3):1168–1182.

Tchetgen Tchetgen, Eric J and Kathleen E Wirth. 2017. "A general instrumental variable framework for regression analysis with outcome missing not at random." *Biometrics* 73(4):1123–1131.

Wang, Linbo, Xiao-Hua Zhou and Thomas S. Richardson. 2017. "Identification and estimation of causal effects with outcomes truncated by death." *Biometrika* 104(3):597–612.
URL: *https://doi.org/10.1093/biomet/asx034*

Ye, Ting, Luke Keele, Raiden Hasegawa and Dylan S Small. 2023. "A negative correlation strategy for bracketing in difference-in-differences." *Journal of the American Statistical Association* pp. 1–13.

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

A Proofs of Section 2

A.1 Proof of Proposition 1

Proof. Under consistency assumption and parallel trends assumption, we have

$$
\begin{aligned}
&\mathbb{E}[Y_{i2}(1) - Y_{i2}(0) \mid D_i = 1] \\
&= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 0] \\
&= \sum_{r=0}^{1} \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, R_i = r] \Pr(R_i = r \mid D_i = 1) \\
&\quad - \sum_{r=0}^{1} \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 0, R_i = r] \Pr(R_i = r \mid D_i = 0) \\
&= \sum_{r=0}^{1} \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, R_i = 1] \Pr(R_i = r \mid D_i = 1) \\
&\quad - \sum_{r=0}^{1} \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 0, R_i = 1] \Pr(R_i = r \mid D_i = 0) \\
&= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, R_i = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 0, R_i = 1]
\end{aligned}
$$

where the second equality uses the law of total expectation and the third equality uses Assumption 2. □

A.2 Proof of Proposition 2

Proof. By the law of total expectation,

$$
\begin{aligned}
&\mathbb{E}[Y_{i2}(1) - Y_{i2}(0) \mid D_i = 1] \\
&= \mathbb{E}[Y_{i2}(1) - Y_{i2}(0) \mid D_i = 1, R_i = 1] \Pr(R_i = 1 \mid D_i = 1) \\
&+ \mathbb{E}[Y_{i2}(1) - Y_{i2}(0) \mid D_i = 1, R_i = 0] \Pr(R_i = 0 \mid D_i = 1)
\end{aligned}
$$

Introducing the pre-treatment period outcome Y_{i1}, we can rewrite the above as

$$
\begin{aligned}
&= \mathbb{E}[Y_{i2}(1) - Y_{i1} \mid D_i = 1, R_i = 1] \Pr(R_i = 1 \mid D_i = 1) \\
&+ \mathbb{E}[Y_{i2}(1) - Y_{i1} \mid D_i = 1, R_i = 0] \Pr(R_i = 0 \mid D_i = 1)
\end{aligned}
$$

$$- \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 1, R_i = 1] \Pr(R_i = 1 \mid D_i = 1)$$
$$- \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 1, R_i = 0] \Pr(R_i = 0 \mid D_i = 1)$$

By the consistency assumption, we have

$$= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, R_i = 1] \Pr(R_i = 1 \mid D_i = 1) \tag{A.1}$$
$$+ \underbrace{\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, R_i = 0] \Pr(R_i = 0 \mid D_i = 1)}_{\text{Part 2}} \tag{A.2}$$
$$- \underbrace{\mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 1, R_i = 1] \Pr(R_i = 1 \mid D_i = 1)}_{\text{Part 1}} \tag{A.3}$$
$$- \underbrace{\mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 1, R_i = 0] \Pr(R_i = 0 \mid D_i = 1)}_{\text{Part 1}} \tag{A.4}$$

where the term in red is not identified due to the missingness.

Part 1 Again, by the law of total expectation and the principal strata parallel trends assumption,

$$\begin{aligned}
&\mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 1, R_i = 1] \Pr(R_i = 1 \mid D_i = 1) \\
&\quad + \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 1, R_i = 0] \Pr(R_i = 0 \mid D_i = 1) \\
&= \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 0, R_i(1) = 1, R_i(0) = 1] \Pr(R_i(1) = 1, R_i(0) = 1 \mid D_i = 1) && \text{(A.5)} \\
&\quad + \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 0, R_i(1) = 1, R_i(0) = 0] \Pr(R_i(1) = 1, R_i(0) = 0 \mid D_i = 1) && \text{(A.6)} \\
&\quad + \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 0, R_i(1) = 0, R_i(0) = 1] \Pr(R_i(1) = 0, R_i(0) = 1 \mid D_i = 1) && \text{(A.7)} \\
&\quad + \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 0, R_i(1) = 0, R_i(0) = 0] \Pr(R_i(1) = 0, R_i(0) = 0 \mid D_i = 1) && \text{(A.8)}
\end{aligned}$$

We can further simplify this with two different strategies:

- Assume missing independent of selection into treatment (See Remark 2).
- Further extend the parallel trend assumption (e.g., parallel trend across principal strata; See Remark 3).

Part 2 Using the law of total expectation and the principal strata parallel trends assumption,

$$
\begin{aligned}
&\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, R_i = 0] \Pr(R_i = 0 \mid D_i = 1) \\
&= \mathbb{E}[Y_{i2}(1) - Y_{i2}(0) \mid D_i = 1, R_i(1) = 0, R_i(0) = 0] \Pr(R_i(1) = 0, R_i(0) = 0 \mid D_i = 1) \\
&\quad + \mathbb{E}[Y_{i2}(1) - Y_{i2}(0) \mid D_i = 1, R_i(1) = 0, R_i(0) = 1] \Pr(R_i(1) = 0, R_i(0) = 1 \mid D_i = 1) \\
&\quad + \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 0, R_i(1) = 0, R_i(0) = 1] \Pr(R_i(1) = 0, R_i(0) = 1 \mid D_i = 1) \\
&\quad + \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 0, R_i(1) = 0, R_i(0) = 0] \Pr(R_i(1) = 0, R_i(0) = 0 \mid D_i = 1)
\end{aligned}
$$

To further simplify this, it requires assumptions on the effect heterogeniety across principal strata. □

A.3 Remarks on Section 2

Remark 2 (*Identification attempt under missing independent of selection into treatment*). *If we assume $S_i \perp\!\!\!\perp D_i$ (missing independent of selection into treatment), the quantity can be further simplified as*

$$
\begin{aligned}
&\mathbb{E}[Y_{i2}(1) - Y_{i2}(0) \mid D_i = 1] \\
&= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, R_i = 1] \Pr(R_i = 1 \mid D_i = 1) \\
&+ \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, R_i = 0] \Pr(R_i = 0 \mid D_i = 1) \\
&- \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 0, R_i = 1] \Pr(R_i = 1 \mid D_i = 0) \\
&- \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 0, R_i = 0] \Pr(R_i = 0 \mid D_i = 0)
\end{aligned}
$$

where the terms in red are not identified due to the missingness.

Proof. From Eq. (A.1) – (A.4), we have

$$
\begin{aligned}
&\mathbb{E}[Y_{i2}(1) - Y_{i2}(0) \mid D_i = 1] \\
&= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, R_i = 1] \Pr(R_i = 1 \mid D_i = 1) \\
&\quad + \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, R_i = 0] \Pr(R_i = 0 \mid D_i = 1)
\end{aligned}
$$

$$
\begin{aligned}
&- \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 1, R_i = 1] \Pr(R_i = 1 \mid D_i = 1) \\
&- \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 1, R_i = 0] \Pr(R_i = 0 \mid D_i = 1)
\end{aligned}
$$

where the last two lines are equal to Eq. (A.5) – (A.8) respectively. Substituting $\Pr(S_i = s \mid D_i = 1)$ with $\Pr(S_i = s \mid D_i = 0)$, we have

$$
\begin{aligned}
&\mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 0, R_i(1) = 1, R_i(0) = 1] \Pr(R_i(1) = 1, R_i(0) = 1 \mid D_i = 0) \\
&\quad + \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 0, R_i(1) = 1, R_i(0) = 0] \Pr(R_i(1) = 1, R_i(0) = 0 \mid D_i = 0) \\
&\quad + \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 0, R_i(1) = 0, R_i(0) = 1] \Pr(R_i(1) = 0, R_i(0) = 1 \mid D_i = 0) \\
&\quad + \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 0, R_i(1) = 0, R_i(0) = 0] \Pr(R_i(1) = 0, R_i(0) = 0 \mid D_i = 0) \\
&= \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 0, R_i(0) = 1] \Pr(R_i(0) = 1 \mid D_i = 0) \\
&\quad + \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 0, R_i(0) = 0] \Pr(R_i(0) = 0 \mid D_i = 0) \\
&= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 0, R_i = 1] \Pr(R_i = 1 \mid D_i = 0) \\
&\quad + \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 0, R_i = 0] \Pr(R_i = 0 \mid D_i = 0)
\end{aligned}
$$

Putting these together, we have

$$
\begin{aligned}
&\mathbb{E}[Y_{i2}(1) - Y_{i2}(0) \mid D_i = 1] \\
&= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, R_i = 1] \Pr(R_i = 1 \mid D_i = 1) \\
&\quad + \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, R_i = 0] \Pr(R_i = 0 \mid D_i = 1) \\
&\quad - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 0, R_i = 1] \Pr(R_i = 1 \mid D_i = 0) \\
&\quad - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 0, R_i = 0] \Pr(R_i = 0 \mid D_i = 0)
\end{aligned}
$$

This implies that missing independent of selection into treatment is not sufficient for identification. □

Remark 3 (*Identification attempt under parallel trends across principal strata*). *Assume the following extended parallel trends assumption:*

$$
\mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = d, S_i = s] = \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = d', S_i = s']
$$

for $d, d' \in \{0,1\}$ *and* $s, s' \in \{(0,0),(0,1),(1,0),(1,1)\}$. *Somewhat trivial result is that, we can identify* $\mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 1]$ *by* $\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 0, R_i = 1]$. *However, we cannot identify ATT without further assumptions such as homogeneous effect across principal strata.*

Proof.

$$
\begin{aligned}
&\mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 1] \\
&= \sum_{r_1, r_0} \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 1, R_i(1) = r_1, R_i(0) = r_0] \Pr(R_i(1) = r_i, R_i(0) = r_0 \mid D_i = 1) \\
&= \sum_{r_1, r_0} \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 0, R_i(1) = r_1, R_i(0) = r_0] \Pr(R_i(1) = r_i, R_i(0) = r_0 \mid D_i = 1) \\
&= \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 0, R_i(1) = r_1', R_i(0) = r_0'] \sum_{r_1, r_0} \Pr(R_i(1) = r_i, R_i(0) = r_0 \mid D_i = 1) \\
&= \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 0, R_i(1) = 1, R_i = 1] \\
&= \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 0, R_i(1) = 1, R_i = 1] \Pr(R_i(1) = 1 \mid R_i = 1, D_i = 0) \\
&\quad + \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 0, R_i(1) = 1, R_i = 1] \Pr(R_i(1) = 0 \mid R_i = 1, D_i = 0) \\
&= \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 0, R_i(1) = 1, R_i = 1] \Pr(R_i(1) = 1 \mid R_i = 1, D_i = 0) \\
&\quad + \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 0, R_i(1) = 0, R_i = 1] \Pr(R_i(1) = 0 \mid R_i = 1, D_i = 0) \\
&= \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 0, R_i = 1] \\
&= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 0, R_i = 1]
\end{aligned}
$$

□

B Proofs of Section 3

B.1 Proofs of Section 3.1

Setup. Suppose we have multiple waves of panel data where there exists missing data in the pre-treatment periods. From now on, we use R_{it} instead of R_i to denote the response indicator of the outcome variable at time t. That is, $R_{it} = 1$ if Y_{it} is observed, and 0 if it is missing, for $t = 1, 2$. Furthermore, we introduce an auxiliary variable, W_i and its response indicators, $\widetilde{R}_i$. The auxiliary variable can be constructed from other pre-treatment variables. For instance, in the first motivating example, we may use other outcome variables (e.g. "Afghanistan right direction?", etc) at 2016 wave as auxiliary variables.

ID	W_i	Y_{i1}	D_i	Y_{i2}
1	0	0	0	1
2	1	NA	0	2
3	1	NA	1	2
4	NA	2	0	3
5	1	2	1	4
6	2	3	1	NA
7	NA	0	1	1
8	NA	1	1	NA
9	NA	2	1	NA

ID	$\widetilde{R}_i$	R_{i1}	D_i	R_{i2}
1	1	1	0	1
2	1	0	0	1
3	1	0	1	1
4	0	1	0	1
5	1	1	1	1
6	1	1	1	0
7	0	1	1	1
8	0	1	1	0
9	0	1	1	0

Table 5: Toy Example of Observed Data with Three Waves.

Note that we now allow for the missingness in the pre-treatment periods, which may make DID estimator infeasible unless we assume further about the missingness mechanism of the pre-treatment outcome. Here, we assume an exclusion restriction type of assumption for the pre-treatment outcome missingness and time trend of outcome.

Assumption 8 (*Exclusion restriction for pre-treatment outcome missingness*). *The pre-treatment outcome missingness may affect the time trend of the outcome only through the treatment selection and the post-treatment outcome missingness.*

$$Y_{i2} - Y_{i1} \perp\!\!\!\perp R_{i1} \mid D_i, R_{i2}$$

One may consider this assumption as a variant of the missing at random (MAR) assump-

tion for the pre-treatment outcome missingness, where the missingness of the pre-treatment outcome is independent of the missing values of the time trend of the outcome given other information including the treatment selection and the post-treatment outcome missingness. To be precise, this is not to say that the pre-treatment outcome is MAR, since the assumption is not about the missing values of Y_{i1} itself but about the missingness of the time trend of the outcome $Y_{i2} - Y_{i1}$. Based on this assumption, we introduce a set of assumptions for using the baseline response indicator as IV for the time trend of the outcome, conditioning on the treatment and the pre-treatment outcome missingness. Note that a weaker version of the assumption is required for the identification ($\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, R_{i2} = r, R_{i1} = 1] = \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, R_{i2} = r, R_{i1} = 0]$), yet we consider the stronger version for the sake of clarity. Alternatively, we can also define ATT for the population who responded to the pre-treatment survey, i.e. $R_{i1} = 1$.

Lemma 1 *Under Assumption 4 and Assumption 8, we have*

$$
\begin{aligned}
&\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, R_{i2} = 0] \\
&= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, R_{i1} = 1, R_{i2} = 1] \\
&\quad + \left\{\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, \widetilde{R}_i = 1, R_{i1} = 1, R_{i2} = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, \widetilde{R}_i = 0, R_{i1} = 1, R_{i2} = 1]\right\} \\
&\qquad \times \left\{ \Pr(R_{i2} = 0 \mid D_i = d, \widetilde{R}_i = 0, R_{i1} = 1) - \Pr(R_{i2} = 0 \mid D_i = d, \widetilde{R}_i = 1, R_{i1} = 1)\right\}^{-1}
\end{aligned}
$$

for $d = 0, 1$.

Proof.

$$
\begin{aligned}
&\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1] \\
&= \sum_{r=0,1} \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1, R_{i2} = r] \Pr(R_{i2} = r \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1) \\
&= \sum_{r=0,1} \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1, R_{i2} = r] \Pr(R_{i2} = r \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1) \\
&\quad + \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1, R_{i2} = 1] \Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1) \\
&\quad - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1, R_{i2} = 1] \Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1) \\
&= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, \widetilde{R}_i = 1, R_{i1} = 1, R_{i2} = 1]
\end{aligned}
$$

$$
\begin{aligned}
&+ \{\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1, R_{i2} = 0] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1, R_{i2} = 1]\} \\
&\quad \times \Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1)
\end{aligned}
$$

With similar steps,

$$
\begin{aligned}
&\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 0, R_{i1} = 1] \\
&= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 0, R_{i1} = 1, R_{i2} = 1] \\
&\quad + \{\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 0, R_{i1} = 1, R_{i2} = 0] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 0, R_{i1} = 1, R_{i2} = 1]\} \\
&\quad\quad \times \Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i = 0, R_{i1} = 1) \\
&= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 0, R_{i1} = 1, R_{i2} = 1] \\
&\quad + \{\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1, R_{i2} = 0] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1, R_{i2} = 1]\} \\
&\quad\quad \times \Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i = 0, R_{i1} = 1)
\end{aligned}
$$

By **parallel trends of observed outcome**, we have

$$
\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1] = \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 0, R_{i1} = 1]
$$

Thus,

$$
\begin{aligned}
&\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1, R_{i2} = 1] \\
&\quad + \{\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1, R_{i2} = 0] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1, R_{i2} = 1]\} \\
&\quad\quad \times \Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1) \\
&= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 0, R_{i1} = 1, R_{i2} = 1] \\
&\quad + \{\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1, R_{i2} = 0] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1, R_{i2} = 1]\} \\
&\quad\quad \times \Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i = 0, R_{i1} = 1)
\end{aligned}
$$

which leads to

$$
\begin{aligned}
&\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1, R_{i2} = 0] \\
&= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1, R_{i2} = 1]
\end{aligned}
$$

$$
\begin{aligned}
&+ \big\{\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1, R_{i2} = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 0, R_{i1} = 1, R_{i2} = 1]\big\} \\
&\quad \times \big\{ \Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i = 0, R_{i1} = 1) - \Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1)\big\}^{-1}
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 0, R_{i1} = 1, R_{i2} = 0] \\
&= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 0, R_{i1} = 1, R_{i2} = 1] \\
&\quad + \big\{\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1, R_{i2} = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 0, R_{i1} = 1, R_{i2} = 1]\big\} \\
&\quad\quad \times \big\{ \Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i = 0, R_{i1} = 1) - \Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1)\big\}^{-1}
\end{aligned}
$$

By **bias homogeneity**,

$$
\begin{aligned}
&\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, R_{i1} = 1, R_{i2} = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, R_{i1} = 1, R_{i2} = 0] \\
&= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1, R_{i2} = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1, R_{i2} = 0] \\
&= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 0, R_{i1} = 1, R_{i2} = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 0, R_{i1} = 1, R_{i2} = 0] \\
&= -\big\{\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1, R_{i2} = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 0, R_{i1} = 1, R_{i2} = 1]\big\} \\
&\quad\quad \times \big\{ \Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i = 0, R_{i1} = 1) - \Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1)\big\}^{-1}
\end{aligned}
$$

Plugging this into the following equation,

$$
\begin{aligned}
&\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, R_{i2} = 0] \\
&= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, R_{i1} = 1, R_{i2} = 0] \\
&= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, R_{i1} = 1, R_{i2} = 1] \\
&\quad + \big\{\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1, R_{i2} = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i = 0, R_{i1} = 1, R_{i2} = 1]\big\} \\
&\quad\quad \times \big\{ \Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i = 0, R_{i1} = 1) - \Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i = 1, R_{i1} = 1)\big\}^{-1}
\end{aligned}
$$

Similarly, we can show the equation with $d = 0$. □

B.2 Proofs of Section 3.2

B.2.1 Proof of Proposition 4

Proof. By Assumption 6, we have

$$\mathbb{E}[R_{i2}(0) \mid D_i = 1] = \mathbb{E}[R_{i2}(0) - R_{i1}(0) \mid D_i = 0] + \mathbb{E}[R_{i1}(0) \mid D_i = 1].$$

Since R_{it} is binary, we have

$$\Pr(R_{i2}(0) = 1 \mid D_i = 1) = \Pr(R_{i2} = 1 \mid D_i = 0) - \Pr(R_{i1} = 1 \mid D_i = 0) + \Pr(R_{i1} = 1 \mid D_i = 1).$$

By Assumption 7, we have

$$\mathbb{E}[R_{i2}(1) \mid D_i = 0] = \mathbb{E}[R_{i2}(1) - R_{i2}(0) \mid D_i = 1] + \mathbb{E}[R_{i2}(0) \mid D_i = 0]$$

which gives us

$$\Pr(R_{i2}(1) = 1 \mid D_i = 0) = \Pr(R_{i2} = 1 \mid D_i = 1) - \Pr(R_{i2}(0) = 1 \mid D_i = 1) + \Pr(R_{i2} = 1 \mid D_i = 0).$$

By plugging in the previous result, we have

$$\Pr(R_{i2}(1) = 1 \mid D_i = 0) = \Pr(R_{i2} = 1 \mid D_i = 1) - \Pr(R_{i1} = 1 \mid D_i = 1) + \Pr(R_{i1} = 1 \mid D_i = 0).$$

□

B.2.2 Additional Result: Bounds for Principal Strata Proportion

Without monotonicity, we can bound the proportion of principal strata, $\pi_{11}(0)$ and $\pi_{11}(1)$, using Assumption 6 and 7.

$$\begin{aligned}
&\max\{0, \Pr(R_{i2}(0) = 1 \mid D_i = 1) - \Pr(R_{i2} = 0 \mid D_i = 1)\} \\
&\quad \leq \pi_{11}(1) \leq \min\{\Pr(R_{i2} = 1 \mid D_i = 1), \Pr(R_{i2}(0) = 1 \mid D_i = 1)\} \\
&\max\{0, \Pr(R_{i2}(1) = 1 \mid D_i = 0) - \Pr(R_{i2} = 0 \mid D_i = 0)\}
\end{aligned}$$

$$\leq \pi_{11}(0) \leq \min\left\{\Pr(R_{i2}=1 \mid D_i=0), \Pr(R_{i2}(1)=1 \mid D_i=0)\right\}$$

where

$$\Pr(R_{i2}(0)=1 \mid D_i=1) = \Pr(R_{i2}=1 \mid D_i=0) - \Pr(R_{i1}=1 \mid D_i=0) + \Pr(R_{i1}=1 \mid D_i=1)$$
$$\Pr(R_{i2}(1)=1 \mid D_i=0) = \Pr(R_{i2}=1 \mid D_i=1) - \Pr(R_{i1}=1 \mid D_i=1) + \Pr(R_{i1}=1 \mid D_i=0).$$

Proof. Part 1: Bounding $\pi_{11}(1)$. We have the following observed quantities and constraint

$$\begin{cases} \pi_{r_1,r_0}(1) \in [0,1] \quad \forall r_0, r_1 = 0,1 \\ \pi_{11}(1) + \pi_{01}(1) + \pi_{10}(1) + \pi_{00}(1) = 1 \\ \pi_{10}(1) = \Pr(R_{i2}=1 \mid D_i=1) - \pi_{11}(1) \\ \pi_{01}(1) = \Pr(R_{i2}(0)=1 \mid D_i=1) - \pi_{11}(1) \end{cases}$$

where $\Pr(R_{i2}(0)=1 \mid D_i=1)$ is identified with Assumption 6. Then, combining last three conditions we have

$$\begin{aligned} \pi_{00}(1) &= 1 - \Pr(R_{i2}=1 \mid D_i=1) - \Pr(R_{i2}(0)=1 \mid D_i=1) + \pi_{11}(1) \\ &= \Pr(R_{i2}=0 \mid D_i=1) - \Pr(R_{i2}(0)=1 \mid D_i=1) + \pi_{11}(1) \end{aligned}$$

Thus, we have

$$\begin{cases} 0 \leq \pi_{11}(1) \leq 1 \\ -\Pr(R_{i2}=0 \mid D_i=1) \leq \pi_{11}(1) \leq \Pr(R_{i2}=1 \mid D_i=1) \\ -\Pr(R_{i2}(0)=0 \mid D_i=1) \leq \pi_{11}(1) \leq \Pr(R_{i2}(0)=1 \mid D_i=1) \\ \Pr(R_{i2}(0)=1 \mid D_i=1) - \Pr(R_{i2}=0 \mid D_i=1) \leq \pi_{11}(1) \leq \Pr(R_{i2}=1 \mid D_i=1) + \Pr(R_{i2}(0)=1 \mid D_i=1) \end{cases}$$

which gives us the bounds for $\pi_{11}(1)$

$$\begin{aligned} &\max\left\{0, \Pr(R_{i2}(0)=1 \mid D_i=1) - \Pr(R_{i2}=0 \mid D_i=1)\right\} \\ &\leq \pi_{11}(1) \leq \min\left\{\Pr(R_{i2}=1 \mid D_i=1), \Pr(R_{i2}(0)=1 \mid D_i=1)\right\} \end{aligned}$$

Part 2: Bounding $\pi_{11}(0)$. By the assumption $\mathbb{E}[R_i(1)-R_i(0) \mid D_i=1] = \mathbb{E}[R_i(1)-R_i(0) \mid D_i=0]$ we have $\Pr(R_i(1)=1 \mid D_i=0) = \Pr(R_i=1 \mid D_i=1) + \Pr(R_i=1 \mid D_i=0) - \Pr(R_i(0)=1 \mid D_i=1)$. Plugging in $\Pr(R_{i2}(0)=1 \mid D_i=1) = \Pr(R_i=1 \mid D_i=0) - \Pr(R_{i1}=1 \mid D_i=0) + \Pr(R_{i1}=1 \mid D_i=1)$, we have

$$\Pr(R_i(1)=1 \mid D_i=0) = \Pr(R_i=1 \mid D_i=1) - \Pr(R_{i1}=1 \mid D_i=1) + \Pr(R_{i1}=1 \mid D_i=0)$$

Accordingly, we have

$$\begin{cases} \pi_{r_1,r_0}(0) \in [0,1] \quad \forall r_0, r_1 = 0,1 \\ \pi_{11}(0) + \pi_{01}(0) + \pi_{10}(0) + \pi_{00}(0) = 1 \\ \pi_{10}(0) = \Pr(R_{i2}=1 \mid D_i=0) - \pi_{11}(0) \\ \pi_{01}(0) = \Pr(R_{i2}(1)=1 \mid D_i=0) - \pi_{11}(0) \end{cases}$$

Analogous to the previous case, we have

$$\begin{aligned} &\max\{0, \Pr(R_{i2}(1)=1 \mid D_i=0) - \Pr(R_{i2}=0 \mid D_i=0)\} \\ &\quad \leq \pi_{11}(0) \leq \min\{\Pr(R_{i2}=1 \mid D_i=0), \Pr(R_{i2}(1)=1 \mid D_i=0)\} \end{aligned}$$

□

C Extension to Multiple Baseline Response Indicators

ID	$W_i^{(1)}$	$W_i^{(2)}$	Y_{i1}	D_i	Y_{i2}
1	0	0	0	0	1
2	2	1	NA	0	NA
3	0	1	3	1	NA
4	3	NA	NA	0	NA
5	1	1	2	1	4
6	2	2	3	1	NA
7	2	1	0	1	1
8	1	NA	3	1	NA
9	NA	1	2	1	4

ID	$\widetilde{R}_i^{(1)}$	$\widetilde{R}_i^{(2)}$	R_{i1}	D_i	R_{i2}
1	1	1	1	0	1
2	1	1	0	0	0
3	1	1	1	1	0
4	1	0	0	0	0
5	1	1	1	1	1
6	1	1	1	1	0
7	1	1	1	1	1
8	1	0	1	1	0
9	0	1	1	1	1

ID	$W_i^{(1)}$	$W_i^{(2)}$	Y_{i1}	D_i	Y_{i2}
1	0	0	0	0	1
2	2	1	NA	0	NA
3	0	1	3	1	NA
4	3	NA	NA	0	NA
5	1	1	2	1	4
6	2	2	3	1	NA
7	2	1	0	1	1
8	1	NA	3	1	NA
9	NA	1	2	1	4

ID	$\widetilde{R}_i^{(1)}$	$\widetilde{R}_i^{(2)}$	R_{i1}	D_i	R_{i2}
1	1	1	1	0	1
2	1	1	0	0	0
3	1	1	1	1	0
4	1	0	0	0	0
5	1	1	1	1	1
6	1	1	1	1	0
7	1	1	1	1	1
8	1	0	1	1	0
9	0	1	1	1	1

Table 6: Toy Example of Observed Data with Auxiliary Variables.

Assumption 9 (*Baseline response indicators as IV for time trend*). *Suppose we have baseline response indicators* $\widetilde{R}_i^{(1)}$ *and* $\widetilde{R}_i^{(2)}$. *We assume the following set of assumptions, for each treatment group whose pre-treatment outcome is observed:*

1. ***Parallel difference in trends*** *The difference in time trends of the outcome between respondents and nonrespondents is parallel for two auxiliary variables.*

$$\begin{aligned}&\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, \widetilde{R}_i^{(1)} = 0, R_{i1} = 1] \\ &= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, \widetilde{R}_i^{(2)} = 1, R_{i1} = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, \widetilde{R}_i^{(2)} = 0, R_{i1} = 1]\end{aligned}$$

   for $d = 0, 1$.

2. ***Bias homogeneity****: The bias due to missingness in the time trend of the outcome is*

homogeneous across the subgroups defined by the baseline response indicators, and is same as the marginalized version.

$$\begin{aligned}
&\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, R_{i1} = 1, R_{i2} = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, R_{i1} = 1, R_{i2} = 0] \\
&= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, \widetilde{R}_i^{(1)} = r, R_{i1} = 1, R_{i2} = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, \widetilde{R}_i^{(1)} = r, R_{i1} = 1, R_{i2} = 0] \\
&= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, \widetilde{R}_i^{(2)} = r, R_{i1} = 1, R_{i2} = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, \widetilde{R}_i^{(2)} = r, R_{i1} = 1, R_{i2} = 0]
\end{aligned}$$

for $d = 0, 1$ *and* $r = 0, 1$.

3. ***Relevance to missingness****: The baseline response indicators are relevant to the missingness of post-treatment outcome.*

$$\begin{aligned}
&\Pr(R_{i2} = 0 \mid D_i = d, \widetilde{R}_i^{(1)} = 0, R_{i1} = 1) \neq \Pr(R_{i2} = 0 \mid D_i = d, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1) \\
&\Pr(R_{i2} = 0 \mid D_i = d, \widetilde{R}_i^{(2)} = 0, R_{i1} = 1) \neq \Pr(R_{i2} = 0 \mid D_i = d, \widetilde{R}_i^{(2)} = 1, R_{i1} = 1)
\end{aligned}$$

for $d = 0, 1$.

Lemma 2 *Under Assumption 9, we have*

$$\begin{aligned}
&\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1, R_{i2} = 0] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1, R_{i2} = 1] \\
&= \big[\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1, R_{i2} = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, \widetilde{R}_i^{(1)} = 0, R_{i1} = 1, R_{i2} = 1] \\
&\quad - \{\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, \widetilde{R}_i^{(2)} = 1, R_{i1} = 1, R_{i2} = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, \widetilde{R}_i^{(2)} = 0, R_{i1} = 1, R_{i2} = 1]\}\big] \\
&\quad \times \big[\Pr(R_{i2} = 0 \mid D_i = d, \widetilde{R}_i^{(1)} = 0, R_{i1} = 1) - \Pr(R_{i2} = 0 \mid D_i = d, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1) \\
&\quad - \{\Pr(R_{i2} = 0 \mid D_i = d, \widetilde{R}_i^{(2)} = 1, R_{i1} = 1) - \Pr(R_{i2} = 0 \mid D_i = d, \widetilde{R}_i^{(2)} = 0, R_{i1} = 1)\}\big]^{-1}
\end{aligned}$$

for $d = 0, 1$.

Proof.

$$\begin{aligned}
&\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1] \\
&= \sum_{r=0,1} \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1, R_{i2} = r] \Pr(R_{i2} = r \mid D_i = 1, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1) \\
&= \sum_{r=0,1} \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1, R_{i2} = r] \Pr(R_{i2} = r \mid D_i = 1, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1)
\end{aligned}$$

$$
\begin{aligned}
&+ \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1, R_{i2} = 1] \Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1) \\
&- \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1, R_{i2} = 1] \Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1) \\
&= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = d, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1, R_{i2} = 1] \\
&\quad + \{\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1, R_{i2} = 0] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1, R_{i2} = 1]\} \\
&\quad \times \Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1)
\end{aligned}
$$

With similar steps,

$$
\begin{aligned}
&\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(1)} = 0, R_{i1} = 1] \\
&= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(1)} = 0, R_{i1} = 1, R_{i2} = 1] \\
&\quad + \{\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(1)} = 0, R_{i1} = 1, R_{i2} = 0] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(1)} = 0, R_{i1} = 1, R_{i2} = 1]\} \\
&\quad \times \Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i^{(1)} = 0, R_{i1} = 1) \\
&= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(1)} = 0, R_{i1} = 1, R_{i2} = 1] \\
&\quad + \{\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1, R_{i2} = 0] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1, R_{i2} = 1]\} \\
&\quad \times \Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i^{(1)} = 0, R_{i1} = 1)
\end{aligned}
$$

where the last equality holds by **bias homogeneity** in Assumption 9. Because of the **parallel difference in trends**, we have

$$
\begin{aligned}
&\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1, R_{i2} = 1] \\
&\quad + \{\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1, R_{i2} = 0] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1, R_{i2} = 1]\} \\
&\quad \times \Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1) \\
&- \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(1)} = 0, R_{i1} = 1, R_{i2} = 1] \\
&\quad - \{\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1, R_{i2} = 0] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1, R_{i2} = 1]\} \\
&\quad \times \Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i^{(1)} = 0, R_{i1} = 1) \\
&= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(2)} = 1, R_{i1} = 1, R_{i2} = 1] \\
&\quad + \{\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(2)} = 1, R_{i1} = 1, R_{i2} = 0] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(2)} = 1, R_{i1} = 1, R_{i2} = 1]\} \\
&\quad \times \Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i^{(2)} = 1, R_{i1} = 1)
\end{aligned}
$$

$$
\begin{aligned}
&- \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(2)} = 0, R_{i1} = 1, R_{i2} = 1] \\
&\quad - \{\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(2)} = 1, R_{i1} = 1, R_{i2} = 0] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(2)} = 1, R_{i1} = 1, R_{i2} = 1]\} \\
&\quad\quad \times \Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i^{(2)} = 0, R_{i1} = 1)
\end{aligned}
$$

By rearranging the terms,

$$
\begin{aligned}
&\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1, R_{i2} = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(1)} = 0, R_{i1} = 1, R_{i2} = 1] \\
&\quad + \{\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1, R_{i2} = 0] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1, R_{i2} = 1]\} \\
&\quad\quad \times \{\Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1) - \Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i^{(1)} = 0, R_{i1} = 1)\} \\
&= \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(2)} = 1, R_{i1} = 1, R_{i2} = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(2)} = 0, R_{i1} = 1, R_{i2} = 1] \\
&\quad + \{\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(2)} = 1, R_{i1} = 1, R_{i2} = 0] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(2)} = 1, R_{i1} = 1, R_{i2} = 1]\} \\
&\quad\quad \times \{\Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i^{(2)} = 1, R_{i1} = 1) - \Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i^{(2)} = 0, R_{i1} = 1)\}
\end{aligned}
$$

Again, by **bias homogeneity**, we have

$$
\begin{aligned}
&\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1, R_{i2} = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(1)} = 0, R_{i1} = 1, R_{i2} = 1] \\
&- \{\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(2)} = 1, R_{i1} = 1, R_{i2} = 1] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(2)} = 0, R_{i1} = 1, R_{i2} = 1]\} \\
&= \{\mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(2)} = 1, R_{i1} = 1, R_{i2} = 0] - \mathbb{E}[Y_{i2} - Y_{i1} \mid D_i = 1, \widetilde{R}_i^{(2)} = 1, R_{i1} = 1, R_{i2} = 1]\} \\
&\quad\quad \times \big[\{\Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i^{(2)} = 1, R_{i1} = 1) - \Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i^{(2)} = 0, R_{i1} = 1)\} \\
&\quad\quad\quad - \{\Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i^{(1)} = 1, R_{i1} = 1) - \Pr(R_{i2} = 0 \mid D_i = 1, \widetilde{R}_i^{(1)} = 0, R_{i1} = 1)\}\big]
\end{aligned}
$$

We can rearrange the terms to get the desired result. Similarly, we can show the equation with $d = 0$. □

D Identification of ATT with Principal Ignorability

Assumption 10 (*Principal ignorability conditioning on the treatment*).

$$(Y_{i2} - Y_{i1}) \perp\!\!\!\perp S_i \mid X_i, D_i$$

where X_i is the pre-treatment covariates.

Assumption 10 (Principal ignorability conditioning on the treatment) is a variant of principal ignorability assumption (Ding and Lu, 2016), that has been adjusted to parallel trends setup. It implies that

$$\mathbb{E}[Y_{i2}(1) - Y_{i1} \mid S_i = s, X_i, D_i = 1] = \mathbb{E}[Y_{i2}(1) - Y_{i1} \mid S_i = s', X_i, D_i = 1]$$
$$\mathbb{E}[Y_{i2}(0) - Y_{i1} \mid S_i = s, X_i, D_i = 0] = \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid S_i = s', X_i, D_i = 0]$$

To further illustrate this, consider the following set of assumptions that are stronger yet intuitive — if these two assumptions hold, then Assumption 10 (Principal ignorability conditioning on the treatment) also holds:

$$Y_{i2}(1) - Y_{i2}(0) \perp\!\!\!\perp S_i \mid X_i, D_i \text{ and } Y_{i2}(0) - Y_{i1} \perp\!\!\!\perp S_i \mid X_i, D_i.$$

This implies (1) homogeneous effect across principal strata conditioning on covariates and treatment and (2) conditional parallel trends across principal strata within each treatment group. Another set of assumptions that are more minimal are:

$$Y_{i2}(1) \perp\!\!\!\perp S_i \mid X_i, D_i \text{ and } Y_{i2}(0) \perp\!\!\!\perp S_i \mid X_i, D_i.$$

This is equivalent to principal ignorability assumption from the original paper (which assumes a randomized treatment), yet conditioning on the treatment as well.

Remark 4 (*Principal ignorability under PT as MAR*). *How does Assumption 10 (Principal ignorability conditioning on the treatment) related to missing at random assumption? Specif-*

ically, we can compare Assumption 10 with the following missing at random assumption:

$$(Y_{i2} - Y_{i1}) \perp\!\!\!\perp R_i \mid X_i, D_i.$$

Note that $S_i \equiv (R_i(1), R_i(0))$. *Thus Assumption 10 (Principal ignorability conditioning on the treatment) which assumes a joint independence implies the marginal independence above. As a result, one may think of this assumtion as a specific type of outcome missing at random.*

Here, we introduce two parallel trend assumptions conditioning on pre-treatment covaraites.

Assumption 11 (*Principal strata parallel trend with covariates*).

$$\mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 0, S_i = s, X_i = x] = \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid D_i = 1, S_i = s, X_i = x]$$

for $s \in \{(0,0), (0,1), (1,0), (1,1)\}$.

Assumption 12 (*Parallel trend of missingness with covariates*).

$$\mathbb{E}[R_{i2}(0) - R_{i1}(0) \mid D_i = 1, X_i = x] = \mathbb{E}[R_{i2}(0) - R_{i1}(0) \mid D_i = 0, X_i = x]$$

Proposition 5 (*Principal score*). *Let* $e_{r_1,r_0}(x) = \Pr(R_{i2}(1) = r_1, R_{i2}(0) = r_0 \mid D_i = 1, X_i = x)$ *for* $r_1, r_0 = 0, 1$. *Then, under Assumption 5 (Monotonicity) and Assumption 12 (Parallel trend of missingness with covariates), we have*

$$\begin{aligned}
e_{11}(x) &= \Pr(R_{i1} = 1 \mid D_i = 1, X_i = x) \\
&\quad + \Pr(R_{i2} = 1 \mid D_i = 0, X_i = x) - \Pr(R_{i1} = 1 \mid D_i = 0, X_i = x) \\
e_{00}(x) &= \Pr(R_{i2} = 0 \mid D_i = 1, X_i = x) \\
e_{10}(x) &= \Pr(R_{i2} = 1 \mid D_i = 1, X_i = x) - e_{11}(x)
\end{aligned}$$

Theorem 3 (*Identification with principal ignorability*). *Under Assumption 5 (Monotonicity), Assumption 10 (Principal ignorability conditioning on the treatment), Assumption 11 (Principal strata parallel trend with covariates), and Assumption 12 (Parallel trend of miss-*

ingness with covariates), we have

$$\begin{aligned}
&\mathbb{E}[Y_{i2}(1) - Y_{i2}(0) \mid D_i = 1, R_{i2}(1) = r_1, R_{i2}(0) = r_0] \\
&\quad = \mathbb{E}[w_{r_1,r_0}(X_i)Y_{i2} \mid D_i = 1, R_{i2} = 1] - \mathbb{E}[w_{r_1,r_0}(X_i)Y_{i1} \mid D_i = 1] \\
&\qquad - \mathbb{E}[w_{r_1,r_0}(X_i)Y_{i2} \mid D_i = 0, R_{i2} = 1] + \mathbb{E}[w_{r_1,r_0}(X_i)Y_{i1} \mid D_i = 0]
\end{aligned}$$

where

$$w_{r_1,r_0}(x) = \frac{e_{r_1,r_0}(x)}{\mathbb{E}[e_{r_1,r_0}(X_i)]}$$

for $(r_1, r_0) \in \{(1,1),(0,0),(1,0)\}$. *Accordingly, we have*

$$\begin{aligned}
ATT = \sum_{(r_1,r_0)} &\mathbb{E}[e_{r_1,r_0}(X_i)Y_{i2} \mid D_i = 1, R_{i2} = 1] - \mathbb{E}[e_{r_1,r_0}(X_i)Y_{i1} \mid D_i = 1] \\
&- \mathbb{E}[e_{r_1,r_0}(X_i)Y_{i2} \mid D_i = 0, R_{i2} = 1] + \mathbb{E}[e_{r_1,r_0}(X_i)Y_{i1} \mid D_i = 0]
\end{aligned}$$

Limitations of Principal Ignorability A list of notable limitations of principal ignorability under this setup is:

- It is in essence a missing at random assumption.
- In practice, it is likely that we also have missingness in pre-treatment covariates.
- With highdimensional pre-treatment covariates, one may potentially need to assume a parametric model for the estimation.