Estimating the Intensive Margin Effect in Panel Data Settings*

Javier Viviens †

February 14, 2026

Abstract

Many policies operate through two different channels: the extensive margin (e.g., the decision to participate) and the intensive margin (e.g., the intensity of the response among participants). This paper develops a novel identification strategy to estimate the intensive margin effect in panel data settings. I adapt the Horowitz-Manski-Lee bounds to the Changes-in-Changes framework to partially identify both the average and quantile intensive margin treatment effects. Additionally, I explore how to leverage multiple sources of sample selection to relax the monotonicity assumption in the original Horowitz-Manski-Lee bounds, which may be of independent interest. Alongside the identification strategy, I present estimators and inference results. I illustrate the relevance of the proposed methodology by analyzing a job training program in Colombia.

JEL-Code: C14, C21, C23.

Keywords: Intensive Margin, Difference-in-differences, Sample selection, Principal stratification, Partial identification.

*A previous version of this paper has been circulated with the title "Difference-in-Differences and Changes-in-Changes with Sample Selection". An R package for implementing the proposed methodology is under development. I am thankful to Fabrizia Mealli, Andrea Ichino, and Alessandro Tarozzi for insightful discussions and guidance. I am also grateful to Francesco Drago, Erin Gabriel, Ellen Greaves, Bo Honoré, Guido Imbens, Miguel Nunes, Jörn-Steffen Pischke, and seminar participants at the European University Institute, JRC Ispra, EuroCIM 2025, ACIC 2025, iCEBDA 2025, and SAEe 2025 for valuable comments and suggestions. All errors are mine.

†Department of Economics, European University Institute. Email: javier.viviens@eui.eu.

1 Introduction

A central concern in economics is understanding the mechanisms through which policies operate. Aggregate policy effects often mask distinct channels through which individuals are affected. A prominent instance arises in policies that impact the extensive margin (e.g., the decision to participate) and the intensive margin (e.g., the intensity of the response among participants). A canonical example of such policies is job training programs, which affect employment (extensive margin) and also wages for those employed (intensive margin) (Ham & Lalonde, 1996; Heckman et al., 1999; Angrist, 2001).[1] Distinguishing between these margins is essential for both welfare analysis and policy design.

The main contribution of this paper is to develop an identification strategy for estimating the intensive margin effect in panel data settings. The existing literature has proposed identification and estimation strategies for intensive margin effects in settings where the treatment can be assumed uncounfounded (e.g., experimental settings).[2] This assumption is often implausible in panel data settings, where treatment is confounded with unobservable characteristics (Arkhangelsky & Imbens, 2024; Liu et al., 2024; Ghanem et al., 2025). This paper fills this gap by adapting intensive-margin estimands and identification strategies to panel data settings in which treatment is not randomly assigned.

I formalize the intensive margin effect using principal stratification analysis (Frangakis & Rubin, 2002). This framework characterizes each unit by its potential participation decision under both treatment arms, thereby partitioning the sample into four distinct subpopulations

[1] Additional examples include, but are not limited to, education policies that affect enrolment and performance for those enrolled (Krueger, 1999; Cornwell et al., 2005); carbon taxes that affect firms' entry and exit and emissions for active firms (Greenstone et al., 2012); weather shocks that affect the decision to migrate and income for those who do not migrate (Gröger & Zylberberg, 2016). In the medical literature, this phenomenon is known as truncation due to death (Zhang & Rubin, 2003; Rubin, 2006; Ding et al., 2011).

[2] This literature was pioneered by Zhang and Rubin (2003), Imai (2008), Lee (2009), and Zhang et al. (2009), who introduced a principled approach to partially identify the intensive margin effect. Subsequent research has extended this literature in multiple directions, including sharpening bounds with covariates (Grilli & Mealli, 2008; Ding et al., 2011; Long & Hudgens, 2013; Samii et al., 2025; Semenova, 2025), instruments (Mattei et al., 2014), post-censoring outcomes (Yang & Small, 2016), or a structural model for selection (Honoré & Hu, 2020), incorporating longitudinal data (Grossi et al., 2023; Comment et al., 2025), considering noncompliance (Chen & Flores, 2015; Blanco et al., 2020), targeting different estimands (Huber & Mellace, 2015; Bartalotti et al., 2023), and exploring non-binary treatment regimes (Lee & Liu, 2025). However, all these papers rely on the assumption that treatment is unconfounded.

or principal strata. For example, in the job training case, those who are always employed, those employed only if trained, those employed only if not trained, and those who are always unemployed. The extensive margin effect refers to the treatment effect on participation (employment in the job training example). The intensive margin effect is defined as the treatment effect on the response for the subset of units that would participate under both treatment arms, and hence for whom there is no effect on participation, i.e., no extensive margin effect. In the job training case, these are the workers who are employed regardless of whether they receive the training or not. For these workers, once participation is defined in terms of employment, any remaining treatment effect operates through the intensive margin, such as changes in wages.

The identification strategy developed in this paper focuses on the Always-Observed principal stratum, i.e., the units with observed outcomes regardless of their treatment status.[3] The effect on these units can be interpreted as the policy's intensive margin effect. To achieve identification, I extend the Howorowitz-Manski-Lee bounds (Horowitz & Manski, 1995; Lee, 2009) to the Changes-in-Changes (CiC) model (Athey & Imbens, 2006). The identification strategy relies on the assumption that all the unobserved confounders affecting the outcome for the Always-Observed units can be captured in a single latent index (for example, ability in the case of wages). The distribution of this index may differ for the Always-Observed units in the treatment and control groups, but it is assumed to be constant over time within each group (that is, the distribution of ability for workers who are always employed can be different in the treatment and control groups as long as this distribution does not vary over time). Combined with information on the principal strata proportions, this assumption delivers partial identification of two causal estimands: (i) the Average Treatment Effect on the Treated for the Always-Observed stratum (ATT_{AO}) and (ii) the Quantile Treatment Effect on the Treated for the same stratum (QTT_{AO}).

Another contribution of this paper is the estimation of the principal strata proportions.

[3]The methodology developed in this paper can also be applied in settings where outcomes are observed for all units, but we are interested in isolating the intensive margin effect. Consider the 'bad control' example in Angrist and Pischke (2009), where we are interested in the effect of education on earnings. Part of this effect operates through occupational choice. In this example, we may be interested in the effect of education on earnings net of occupational changes. In this case, we would estimate the effect on the subpopulation of units that always have a white-collar occupation, regardless of their education level.

I also employ the CiC approach to model participation. Combined with a monotonicity assumption, as in the original Horowitz-Manski-Lee bounds, this method delivers point identification of the principal strata proportions. The required monotonicity assumption implies that treatment can affect participation only in 'one direction' for all the units, ruling out the existence of one of the strata (e.g., units observed only if treated).[4] In the job training context, this assumption implies that all workers employed without training would also be employed if trained. I relax this assumption when the data include multiple sources of participation. For instance, wages may be unobserved for both unemployed workers and workers who decide to migrate. I assume each source is monotonic while allowing the direction of monotonicity to differ across sources. For example, the training may reduce the probability of unemployment but increase the likelihood of migration for all units. This extension enables all four principal strata to exist while still achieving point identification of their proportions. This novel extension of the original Horowitz-Manski-Lee bounds could also be of interest in settings where treatment is unconfounded.

I illustrate the proposed methodology by analyzing a job training program in Colombia. Using data from Attanasio et al. (2011), I estimate the intensive margin effect of the training on salaried earnings. I contrast these findings with those obtained using a 'naive' approach, in which the CiC estimator is applied to units with observed wages. The 'naive' approach suggests that the training increased wages by 13% for trained units. In contrast, using the methodology presented in this paper, I fail to reject the null that the training's intensive margin effect was zero. When examining the QTT_{AO}, I find that the estimated bounds vary across the outcome distribution, suggesting heterogeneous intensive margin effects with potential distributional implications.

To date, the literature on panel data has focused on correcting selection bias and targeting the ATT (Hausman & Wise, 1979; Heckman, 1979; Fitzgerald et al., 1998; Bhattacharya, 2008; Semykina & Wooldridge, 2010; Carlson & Semykina, 2024; Ghanem et al., 2024). While these approaches recover causal effects under sample selection, their estimand can-

[4]This assumption is analogous to the monotonicity assumption in IV settings (Angrist et al., 1996) that rules out the existence of defiers.

not be interpreted as the intensive margin effect.[5] More recently, there has been growing attention to estimating the intensive margin effect in panel data settings. In parallel with this work, Rathnayake et al. (2024) and Shin (2024) extend Horowitz-Manski-Lee bounds to a Difference-in-Differences (DiD) setting and target one of the estimands proposed in this paper (the ATT on the Always-Observed principal stratum). The methodology proposed by this paper can also be applied to the DiD design, as illustrated in Appendix A.1. A key distinction is that both Rathnayake et al. (2024) and Shin (2024) identify the principal strata proportions under a parallel trends assumption in participation, whereas I identify them using the Changes-in-Changes model. An advantage of my approach is that it guarantees estimated probabilities that lie in the unit interval. Moreover, this paper is the first to establish partial identification of distributional effects, the QTT_{AO}. This estimand is particularly relevant when policymakers are interested not only in the average impact of a policy but also in its distributional effects. By shifting the focus from means to distributions, the CiC framework enables the exploration of heterogeneity of treatment effects along the outcome distribution. Finally, I consider several practical extensions—including covariates, binary outcomes, and repeated cross sections—which further enhance the applicability of the proposed framework.

The remainder of this paper is organized as follows. Section 2 introduces the notation and set-up. Section 3 introduces the identification strategy to estimate causal effects for the Always-Observed units. Section 4 derives the estimation and inference results. Section 5 applies the methodology to data from a job training program from Attanasio et al. (2011). Section 6 concludes. The Appendix sections describe the extensions of the main results (A), prove them (B) and provide supplementary material (C).

2 Notation and set up

For simplicity, I consider the two-groups two-periods case. Consider a random sample of N units observed during two periods, $t \in \{1, 2\}$. Let w_{it} denote the treatment of unit i at

[5] This estimand identifies the effect for all treated units and thus integrates over all the principal strata, mixing intensive and extensive margin effects.

time t. No one is treated in the pre-treatment period, $w_{i1} = 0 \; \forall i$. Some units are treated in the post-treatment period, $w_{i2} \in \{0, 1\}$. Let G denote a treatment group indicator, with $G_i = 1$ if unit i is treated in the second period and zero otherwise, $G_i \equiv w_{i2}$. Let Y_{it} be the outcome of unit i at time t. I adopt the potential outcome framework and assume the Stable Unit Treatment Value Assumption (SUTVA) (Rubin, 1980, 1990). This assumption implies that there is no interference between units and there are no hidden versions of treatment, allowing me to relate the observed Y_{it} to the potential outcomes of unit i, $Y_{it}(w_{i1}, w_{i2})$:

$$Y_{it} = Y_{it}(0, 1)G_i + Y_{it}(0, 0)(1 - G_i).$$

The outcome for some sampled units may not be observed. Let S_{it} be a binary selection indicator equal to 1 if the outcome of unit i is observed at time t and zero otherwise. A crucial aspect of this setup is the recognition that the selection indicator, S, is also a post-treatment variable. Consequently, I must also relate the observed selection outcome, S_{it}, to the potential selection outcomes:

$$S_{it} = S_{it}(0, 1)G_i + S_{it}(0, 0)(1 - G_i).$$

Throughout the paper, I will maintain the assumption of no treatment anticipation for both the outcome and the selection, a standard assumption in the literature (Arkhangelsky & Imbens, 2024).

Assumption 1 *No treatment anticipation*

$$\begin{aligned} S_{i1}(0, w_{i2}) = S_{i1}(0, w'_{i2}) = S_{i1} \quad & \forall w_{i2}, w'_{i2} \forall i, \\ Y_{i1}(0, w_{i2}) = Y_{i1}(0, w'_{i2}) = Y_{i1} \quad & \forall w_{i2}, w'_{i2} \forall i. \end{aligned}$$

Assumption 1 implies that the treatment does not affect the outcome and the selection in the pre-treatment period. Since no unit is treated in the first period, it is sufficient to index

the second-period potential outcomes with w_{i2}:

$$S_{i2} = S_{i2}(1)G_i + S_{i2}(0)(1 - G_i),$$
$$Y_{i2} = Y_{i2}(1)G_i + Y_{i2}(0)(1 - G_i).$$

Since selection is a post-treatment variable, the population can be partitioned into subpopulations defined by joint values of the potential selection indicator in the post-treatment period, i.e., $(S_{i2}(1), S_{i2}(0))$. This approach is known as principal stratification (Frangakis & Rubin, 2002). Every unit belongs to one of the four principal strata. Table 1 presents the principal strata, where V_i denotes the principal stratum of unit i. Because potential outcomes are not affected by the treatment, the principal strata exist prior to the treatment assignment. While the observed selection status is a post-treatment outcome, the principal strata act as pre-treatment covariates.[6]

Table 1: Principal Strata.

$S_{i2}(0)$	$S_{i2}(1)$	Description	V_i
1	1	Always Observed	AO
0	0	Never Observed	NO
1	0	Observed only in Control	OC
0	1	Observed only in Treatment	OT

Notes: Principal Strata defined by the joint value of the potential selection indicators, $S(0)$ and $S(1)$. For example, consider the case where the treatment is a job training program aiming to increase workers' wages. The AO stratum comprises workers who are always employed, both with and without training. The NO stratum consists of workers who are always unemployed, regardless of whether they are treated or not. The OC stratum refers to workers who would find a job without the training but would be unemployed if they were treated. Conversely, the OT stratum includes workers who are unemployed in the absence of treatment but find a job if they receive it.

To facilitate the exposition of the proposed methodology, I impose the following assumption on missingness:

Assumption 2 *Missingness as an absorbing state*

$$S_{i1} = 0 \implies S_{i2} = 0.$$

[6]In the example discussed in footnote 3, occupational choice is a post-treatment variable and deemed a bad control (Rosenbaum, 1984; Angrist & Pischke, 2009). Nevertheless, principal strata can be defined in terms of the potential occupational choices under each treatment arm. The treatment determines which of these potential choices is realized, but does not determine to which principal stratum the unit belongs.

Assumption 2 implies that once a unit's outcome is unobserved, it remains unobserved in all subsequent periods. Thus, all the units with $S_{i1} = 0$ will belong to the Never-Observed stratum at $t = 2$. In contrast, units with $S_{i1} = 1$ can transition to any stratum in the second period. Beyond this, Assumptions 1 and 2 do not impose additional constraints on the proportion of the strata in the population, which may vary across groups, nor on the distribution of potential outcomes $Y(0)$ and $Y(1)$ across strata.

Assumption 2 is plausible in settings with censoring due to death or similar events, such as clinical trials where patients cannot be observed after death, or educational contexts where students who drop out or graduate do not subsequently re-enroll. In other settings, however, researchers may be reluctant to impose this assumption, as outcomes that are unobserved in the pre-treatment period may be observed at a later date. This assumption imposes testable restrictions on the data, since S_{i1} and S_{i2} are observed for all units. I impose Assumption 2 to simplify the exposition of the methodology, but it is not required for the identification of the causal effects considered in this paper. Appendix A.2 presents the identification results of this paper without this assumption.

It is also important to note that this assumption does not correspond to restricting the sample to those observed in the first period. Allowing for missingness in the pre-treatment period is crucial for imputing the missing potential selection outcomes, which is central to the methodology developed in this paper. Restricting the sample to units with observed outcomes at baseline would discard valuable information contained in baseline selection. At the same time, as discussed in Appendix A.2, under Assumption 2, the identification strategy in the main text targets causal effects only for units with observed outcomes at baseline. Nonetheless, this strategy can be extended to target causal effects regardless of whether the units were observed in the pre-treatment period, as shown in Appendix A.2.

Finally, I formalize the random sampling assumption in Assumption 3.

Assumption 3 *Random Sampling.* $\{Y_{i1}, Y_{i2}, S_{i1}, S_{i2}, G_i\}_{i=1}^{N}$ *are independent and identically distributed with joint distribution* $P_{Y,G,S}$*, and* $supp(Y_{it}) \subset \mathbb{R} \cup \{*\}$*,* $supp(S_{it}) = \{0, 1\}$*, and* $supp(G_i) = \{0, 1\}$*.*

Assumption 3 implies that the units are sampled from a larger population. Even when the outcome Y_{it} is not observed for some units, the researcher knows that those units were included in the sample. I use the convention that $Y_{it} = *$ for units with $S_{it} = 0$. For example, wages are only observed for employed workers, but researchers also observe which workers are unemployed; quality of life is only observed for survivors, but researchers know which patients died; and academic performance is only observed for enrolled students, but researchers know which students dropped out. Accordingly, the expectations throughout the paper are meant to capture the sampling-based uncertainty (Abadie et al., 2020).

3 Identification of intensive margin effects

This section presents a novel identification strategy to estimate the intensive margin effect in longitudinal settings. Unlike previous literature on intensive margin effects,[7] I do not assume treatment unconfoundedness. Instead, I allow the treatment to be confounded with unobservable unit-level characteristics. However, I impose restrictions on these unobservables' distribution over time. In what follows, I propose two causal estimands and discuss their identification.

Estimand 1 *Average Treatment Effect on the Treated Always-Observed units (ATT_{AO})*

$$ATT_{AO} = \mathbb{E}[Y_{i2}(1) - Y_{i2}(0) \mid G_i = 1, V_i = AO]$$

Estimand 1 is the ATT for the subpopulation of Always-Observed units. This estimand captures the average effect for treated units whose outcome is observed under both treatment arms, and therefore can be interpreted as the policy's intensive margin effect. This estimand is especially relevant in settings where the researcher is interested in the causal effect net of any compositional changes.

In many applications where treatment effects are likely heterogeneous, policymakers may be interested not only in the average treatment effect but also in its distributional impact.

[7] See footnote 2.

For instance, consider the case of a job training program. A positive ATT_{AO} indicates that, on average, the training increases wages for the workers employed both with and without the program. However, this average positive effect might be driven primarily by workers at the lower end of the wage distribution or, conversely, by those at the top. These contrasting scenarios carry significant policy implications, as the distributional effects of the training differ substantially. To fill this gap, I propose a new estimand that targets the treatment effect at various quantiles of the outcome distribution for the Always-Observed treated units.

Estimand 2 *Quantile Treatment Effect on the Treated Always-Observed units (QTT_{AO})*

$$\begin{aligned} QTT_{AO}(q) &= Q_{Y_2(1)|G=1,V=AO}(q) - Q_{Y_2(0)|G=1,V=AO}(q), \\ Q_Y(q) &= \inf\{y : F_Y(y) \geq q\}, \\ q &\in (0,1) \end{aligned}$$

where $F_{Y_t(w)|G=1,V=AO}$ denotes the cumulative distribution of the Potential Outcome $Y_t(w)$ at period t for units that belong to the the treatment group, $G = 1$, and to the Always-Observed stratum, $V = AO \equiv S_2(1) = S_2(0) = 1$.

Estimand 2 contrasts the quantile q of the treated potential outcome distribution for AO units in the treatment group with the same quantile of the control potential outcome distribution for these units. For example, $\text{QTT}_{\text{AO}}(0.5)$ captures the treatment effect at the median of the outcome distribution. This estimand captures a distributional effect: it does not require any rank-invariance assumption across potential outcomes and should not be confused with the median treatment effect (Imbens & Rubin, 2015).

Both Estimands 1 and 2 are principal causal effects, in that they define causal effects for a subpopulation characterized by a principal stratum. Because principal stratum membership depends on joint potential outcomes, which are never simultaneously observed, this subpopulation is not identifiable without additional assumptions. Analogously to the IV setting, where individual compliance types cannot be directly observed, it may not be possible to know which units belong to the Always-Observed stratum. This feature may be viewed as a limitation of the proposed methodology. Nevertheless, because these estimands admit

a clear interpretation as intensive margin effects, they are policy-relevant and practically meaningful for evaluating interventions and informing scale-up decisions. Finally, as in every principal stratification analysis, it is advisable to characterize the units belonging to the Always-Observed principal stratum (Mealli & Mattei, 2012).

An alternative to targeting principal causal effects is given by bias-correction methods, that attempt to recover the ATT in the presence of sample selection. This estimand can be interpreted as "the treatment effect if all the treated units were observed" and therefore cannot be interpreted as the intensive-margin effect without further assumptions. Moreover, bias-correction methods have several limitations. First, they assume that outcomes are always well-defined, even if they are not directly observed (Zhang et al., 2009). In the job training example, bias-correction methods assume the existence of a 'wage offer', allowing us to consider what the wage of an unemployed worker is. Second, they believe that the selection status is manipulable (Mealli & Mattei, 2012), meaning that employment status could, in principle, be directly manipulated by the researcher.[8] Third, bias-correction methods often rely on exclusion restriction assumptions, requiring variables that affect employment but not wages (Lee, 2009) and place restrictions on how potential outcomes are distributed across principal strata.

Next, I outline the assumptions required for partial identification of these estimands. First, I discuss the partial identification of both the ATT_{AO} and the $\text{QTT}_{\text{AO}}(q)$ in a Changes-in-Changes setting. I then examine the identification of the principal strata proportions, which constitutes a key component of the proposed methodology, under alternative sets of assumptions.

[8]When selection status is manipulable, one can define potential outcomes that can never be observed in the data and are 'a priori counterfactuals', such that the potential wage without training for a worker who is only employed if trained. Another example of this type of potential outcome is given by the IV setting, where we can hypothesize about the potential outcome under treatment for a never-taker. See Mealli and Mattei (2012) for a discussion.

3.1 Identification of causal effects

The existing literature estimating the effects for the Always-Observed units has relied on the assumption that treatment is unconfounded. However, this assumption is often unrealistic in causal models for panel data, where treatment assignment is typically non-random and likely correlated with unobserved factors (Liu et al., 2024; Ghanem et al., 2025). To address this limitation, I reformulate the Horowitz-Manski-Lee bounds (Horowitz & Manski, 1995; Lee, 2009) by replacing the unconfoundedness assumption with the Changes-in-Changes (CiC) model as in Athey and Imbens (2006), allowing the treatment assignment to be confounded with unobservable characteristics that affect the potential outcomes.

The choice of the CiC outcome modeling is meant to address several limitations in the canonical DiD research design. The canonical DiD may be sensitive to functional form (Roth & Sant'Anna, 2023) and implicitly relies on additivity and linearity assumptions on the potential outcomes. To address these issues, Athey and Imbens (2006) generalize the DiD identification strategy to a nonlinear and scale-free model, Changes-in-Changes. In the CiC model, individual and time 'fixed effects' can influence the outcome in more flexible ways beyond the linear and additive fashion typically imposed in DiD. The CiC setup also offers additional advantages. First, shifting the focus from means to distributions enables the identification of distributional treatment effects, such as the QTT_{AO}. Second, it avoids extrapolation outside the outcome's support, ensuring that the counterfactual outcomes $Y(0)$ for treated units lie within the support of Y. On the other hand, because the method reconstructs the full distribution of $Y(0)$ for treated units, it requires assumptions that are not needed in the DiD framework, where identification concerns only the first moment of this distribution. As a result, the canonical DiD is not nested as a particular case of the CiC research design (see Roth and Sant'Anna (2023, Remark 4) for an example). Nevertheless, the methodology developed in this paper can be adapted to the canonical DiD setting, as shown in Appendix A.1.

Assumption 4 *Outcome model for Always-Observed units*

$$(Y_{it}(0) \mid V_i = AO) \quad = \quad m(\mathcal{U}_{it}, t),$$

where $m(u,t)$ *is strictly increasing in* u $\forall t$*, and* $\mathcal{U}_{it}$ *is an unobservable scalar for unit* i *at time* t *with constant distribution over time within groups,*

$$\mathcal{U}_{i1} \mid G_i, V_i = AO \sim \mathcal{U}_{i2} \mid G_i, V_i = AO.$$

Assumption 4 is the principal counterpart to assumptions 3.1-3.4 in Athey and Imbens (2006). It models the untreated potential outcome as a function of an unobservable individual characteristic, $\mathcal{U}_{it}$. This model embeds several restrictions on the unobservable, its distribution, and the outcome function $m(\cdot)$. These assumptions do not restrict the data in any way and, thus, are not testable.

This model assumes that all the individual characteristics affecting the outcome can be captured in a single index, $\mathcal{U}_{it}$. The function $m(\cdot)$ being strictly increasing implies that units with higher values of the unobservable also have higher outcomes. This assumption is intuitive when the unobservable is interpreted as an individual characteristic such as ability or productivity. It holds by construction in additively separable models, such as the two-way fixed effects model. However, it also accommodates a broad range of non-additive nonlinear outcome functions.

The distribution of the unobservable may differ between the treatment and control groups. However, it remains constant over time within each group. This assumption is crucial, as it allows interpreting changes in the outcome in the control group as changes in the function $m(\cdot)$. Once the trend in $m(\cdot)$ is estimated, the missing potential outcome for the treatment group can be identified.

Example 1 *(Two-way fixed effects model). Assume a separable additive model. Let* $m(\mathcal{U}_{it}, t) = \mathcal{U}_{it} + \lambda_t$*, and assume* $\mathcal{U}_{it} = \alpha_i + \varepsilon_{it}$*, where* ε_{it} *is random noise. Then, the untreated potential outcome for AO units can be expressed as:*

$$Y_{it}(0) = \alpha_i + \lambda_t + \varepsilon_{it}.$$

This example illustrates that the standard two-way fixed effect model can be seen as a special case of the CiC outcome model.

Proposition 1 *Let $Y_{it}(1)$ and $Y_{it}(0)$ be continuous with compact support. Furthermore, let the support of $Y_{it}(0)$ for the treatment group be contained in the support of $Y_{it}(0)$ for the control group. Then, if Assumptions 1, 2, 3 and 4 hold, then $\Lambda^{LB}(q)$ and $\Lambda^{UB}(q)$ are lower and upper bounds for the Quantile Treatment Effect on the Treated Always-Observed units ($QTT_{AO}(q)$), where:*

$$\Lambda^{LB}(q) = Q_{Y_2|G=1,S_2=1}(q\pi_1) - Q_{Y_2|G=0,S_2=1}\left(F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q\pi_1 + 1 - \pi_1)\right) + 1 - \pi_0\right),$$

$$\Lambda^{UB}(q) = Q_{Y_2|G=1,S_2=1}(q\pi_1 + 1 - \pi_1) - Q_{Y_2|G=0,S_2=1}\left(F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S=1}(q\pi_1)\right) - (1 - \pi_0)\right),$$

$$\pi_1 = Pr(S_{i2}(0) = 1 \mid G_i = 1, S_{i2}(1) = 1),$$

$$\pi_0 = Pr(S_{i2}(1) = 1 \mid G_i = 0, S_{i2}(0) = 1),$$

$$F_Y(y) := Pr(Y \leq y)$$

$$Q_Y(q) := \inf\{y : F_Y(y) \geq q\}$$

provided that

$$F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q\pi_1 + 1 - \pi_1)\right) \leq \pi_0$$

$$F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=0}(q\pi_1)\right) \geq 1 - \pi_0.$$

If $F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q\pi_1 + 1 - \pi_1)\right) > \pi_0$, then

$$\Lambda^{LB}(q) = Q_{Y_2|G=1,S_2=1}(q\pi_1) - Q_{Y_2|G=0,S_2=1}(1).$$

If $F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=0}(q\pi_1)\right) < 1 - \pi_0$, then

$$\Lambda^{UB}(q) = Q_{Y_2|G=1,S_2=1}(q\pi_1 + 1 - \pi_1) - Q_{Y_2|G=0,S_2=1}(0)$$

Proof: See Appendix B.1.

Remark 1 *The two conditions,*

$$F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q\pi_1 + 1 - \pi_1)\right) \leq \pi_0$$

and

$$F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=0}(q\pi_1)\right) \geq 1-\pi_0,$$

ensure that extrapolation is not required outside the support of $Y_2 \mid G=0$*. When these conditions do not hold, the lower bound for* $Q_{Y_2(0)|G=1,V=AO}(q)$ *is given by the infimum of the support of* $Y_2 \mid G=0$*, and the upper bound by its supremum. This implies that, for extreme quantiles, the bounds on* $QTT_{AO}(q)$ *may be uninformative. Moreover, the smaller the proportion of Always-Observed units,* π_1 *and* π_0*, the larger the region where these bounds may be uninformative. Note that when* $\pi_0 = 1$*, both conditions are always satisfied.*

Remark 2 *Once the distributions of* $Y_2(1)_{|G=1,V=AO}$ *and* $Y_2(0)_{|G=1,V=AO}$ *are partially identified, the bounds for the* ATT_{AO} *follow immediately:*

$$\int_0^1 \Lambda^{LB}(q)dq \leq ATT_{AO} \leq \int_0^1 \Lambda^{UB}(q)dq.$$

Remark 3 *Proposition 1 builds on continuity and support assumptions. These results can be extended to cases where the support of* $Y_{it} \mid G_i = 1$ *does not fully overlap with that of* $Y_{it} \mid G_i = 0$*; see Athey and Imbens (2006, Corollary 3.1) for a discussion. Appendix A.5 explores the identification for discrete outcomes. However, the identification is unsuitable when* Y *is mixed or has non-compact support.*

Remark 4 *When the OT and OC strata are empty, it follows that* $\pi_1 = \pi_0 = 1$*,* $\Lambda^{LB} = \Lambda^{UB}$*, and the* $QTT_{AO}(q)$ *is point identified. The intuition behind this result is that all the units observed in the post-treatment period belong to the AO stratum. In this scenario,* S *becomes an indicator of membership in the AO stratum.*

Proposition 1 describes a trimming procedure to bound the $\text{QTT}_{\text{AO}}(q)$. It builds on Horowitz and Manski (1995), who derive bounds for the components of a mixture distribution when the mixture proportions are known—here, the proportions of Always-Observed units in each group. I combine this with results from Athey and Imbens (2006), who derive the counterfactual distribution of $Y_2(0)$ for treated units. Once the distributions of $Y_2(1)$ and $Y_2(0)$ for the treated AO units are bounded, the identification of the $\text{QTT}_{\text{AO}}(q)$ and the ATT_{AO} follows.

The intuition behind Proposition 1 is as follows. Suppose that the proportion of AO units among treated units with observed outcomes, π_1, is known. For example, if $\pi_1 = 0.9$, the observed distribution for treated units trimmed at the 90% percentile constitutes a lower bound for the real latent distribution for these 90% AO treated units. Similarly, the top 90% of the observed distribution constitutes an upper bound. Analogously, the distribution of outcomes for control AO units can be partially identified using the corresponding proportion, π_0. Once the outcome distributions for AO units are partially identified, they can be combined with the results from Athey and Imbens (2006) to partially identify the $\text{QTT}_{\text{AO}}(q)$. Under the CiC model for the AO units, any change in the control group's outcome distribution must come from changes in the outcome mapping $m(\cdot)$, as the distribution of $\mathcal{U}_{it}$ is assumed to be stable over time within groups. This lets me use the control group to recover how $m(\cdot)$ shifts over time. I can then apply the same shift to the pre-treatment distribution of $\mathcal{U}_{it}$ in the treatment group, inferred from their pre-treatment distribution of the outcome, to construct the counterfactual post-treatment distribution. Figure 1 provides graphical intuition behind Proposition 1.

3.2 Identification of principal strata proportions

A pivotal factor of Proposition 1 is the proportion of AO units in each group, π_1 and π_0. Next, I impose additional assumptions on the selection mechanism that enable the identification of these proportions.

Assumption 5 *Monotonicity*

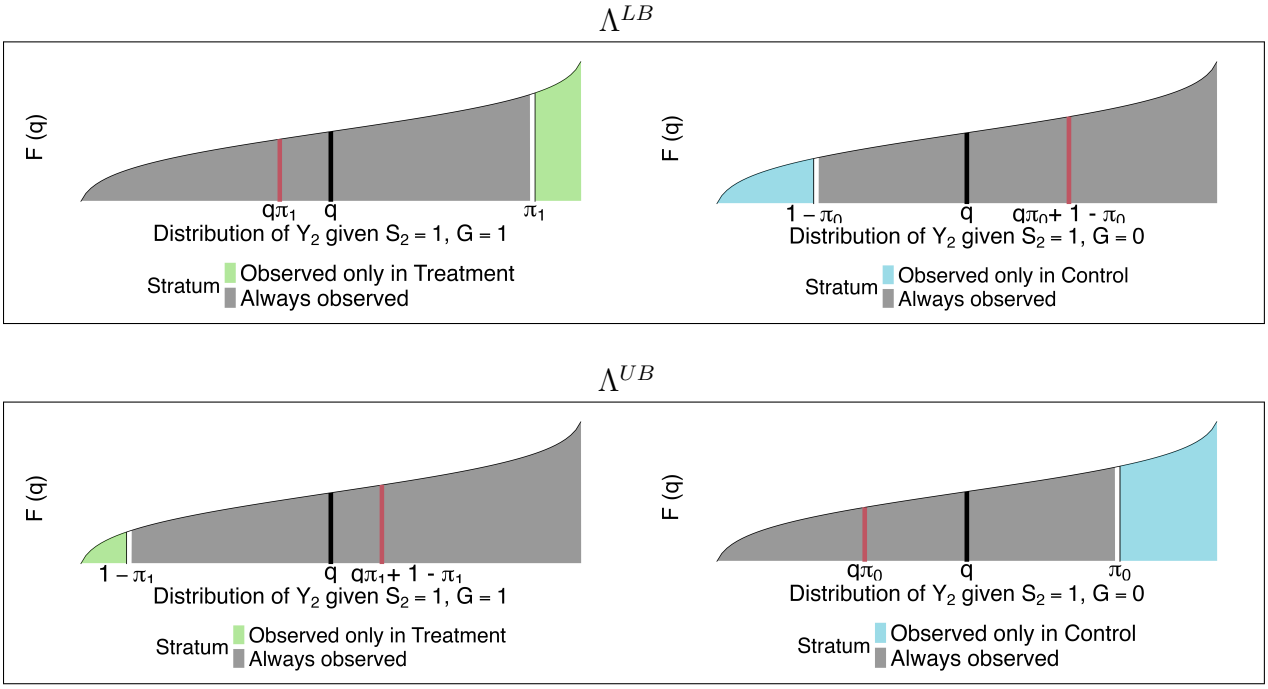
$$S_{i2}(1) \geq S_{i2}(0) \quad \forall i \quad \textit{Positive Monotonicity}$$

$$\textit{or}$$

$$S_{i2}(1) \leq S_{i2}(0) \quad \forall i \quad \textit{Negative Monotonicity}$$

Assumption 5 implies that treatment can affect selection only in 'one direction' for all the units and rules out one of the four principal strata. Positive monotonicity excludes the OC stratum. This implies that the control units observed in both periods consist solely of

Figure 1: Graphical intuition on Proposition 1.



Notes: This figure plots a hypothetical distribution of $Y_2 \mid S_2 = 1$. The distributions on the left correspond to the Treated group ($G = 1$) and the one on the right to the Control group ($G = 0$). The shaded areas correspond to the proportion of strata in the given group: black corresponds to the AO units (π_1 on the left and π_0 on the right), green corresponds to the OT stratum and blue to the OC. The vertical black lines denote where a given quantile q lies in the observed distribution. The vertical red lines denotes where the same quantile q lies in the trimmed distributions, used in Proposition 1. This figure also illustrates Remark 4: as π_1 and π_0 go to 1, the black shadowed areas expand, and the trimmed (red) quantiles converge to those of the entire distribution (black).

Always-Observed units: $S_{i2}(1) \geq S_{i2}(0) \implies Pr(S_{i2}(1) = 0 \mid S_{i2}(0) = 1) = 0 \implies \pi_0 = 1$. Conversely, negative monotonicity rules out the OT stratum, assuming that all the treated units observed in both periods belong to the AO stratum. While restrictive, this assumption is standard in the literature and fundamental to ensure point identification of principal strata proportions. It is analogous to the monotonicity assumption that rules out the existence of defiers in IV settings. In section 3.3, I propose a novel methodology to relax this assumption when different sources of selection are available in the data.

The preference for positive or negative monotonicity depends on the context. For example, consider a researcher estimating the effect of job training on wages. It is reasonable to assume that a worker employed without training would also be employed with it. This assumption makes positive monotonicity the preferred assumption. Conversely, take the example of a researcher studying the effect of increased tuition fees on university students'

performance. In this case, sample selection arises as some students drop out. Under the assumption that higher education costs increase dropouts, it is reasonable to conclude that if a student remains enrolled despite the fee increase, they would also stay enrolled without it. In this case, negative monotonicity is preferred.

When positive (negative) monotonicity is assumed, π_0 (π_1) is equal to one, and only the treatment (control) group needs to be trimmed. I adopt the CiC (Athey & Imbens, 2006) methodology to estimate the proportion of AO units in this group. This identification strategy assumes that an unknown function of individual unobservables determines the potential selection outcomes. This approach has some key advantages over the canonical DiD framework when applied to selection: it allows the identification of the missing potential selection outcomes for both groups. This is crucial when trimming the control group under negative monotonicity. Furthermore, it ensures that all the estimated probabilities lie in the unit interval.

Assumption 6 *Selection Model*

Under Positive Monotonicity:

$$S_{it}(0) = h^0(U_{it}, t).$$

Under Negative Monotonicity:

$$S_{it}(1) = h^1(U_{it}, t),$$

where U_{it} is an unobservable scalar for unit i at time t and $h^w(u,t)$ is non decreasing function in u $\forall$ $t \in \{1,2\}$, $w = \{0,1\}$.

The unobservable U is continuously distributed and has the same compact support in both groups. Its distribution is constant over time within groups

$$U_{i1} \mid G_i \sim U_{i2} \mid G_i.$$

Additionally, given the realized selection outcome, the distribution of U is independent of the group in a given time period.

$$U_{it} \perp\!\!\!\perp G_i \mid S_{it}$$

Assumption 6 models the selection mechanism as a function of an unobservable individual

characteristic, U_{it}. Depending on the monotonicity direction in Assumption 5, a model for $S_{it}(0)$ or $S_{it}(1)$ needs to be assumed. Under Positive (Negative) Monotonicity, the control (treatment) group is composed solely of AO units, and the treatment (control) group is a mixture of AO and OT (OC) units, and therefore $\mathbb{E}[S_{i2}(0) \mid G_i = 1, S_2 = 1]$ ($\mathbb{E}[S_{i2}(1) \mid G_i = 0, S_2 = 1]$) needs to be imputed. This selection model shares significant similarities with the outcome model in Assumption 4. The main components of both models are the unobservable scalars, U in the case of selection, and $\mathcal{U}$ in the case of the outcome. While they may be independent, the potential correlation between the unobserved factors influencing selection and those determining the outcome is central to the non-ignorable sample selection problem.

There are three noteworthy differences between the outcome model in Assumption 4 and the selection model in Assumption 6, albeit both share the general intuition behind a CiC model. First, depending on the direction of the monotonicity assumption, imputing the missing potential selection outcomes may be necessary for the treated or control units. This requires specifying functions for both potential selection outcomes, $S(1)$ in case negative monotonicity holds, and $S(0)$ in case positive monotonicity holds. In contrast, only the missing potential outcome $Y(0)$ for treated units needs to be imputed. Thus, modeling $Y_{it}(1)$ is not required. Second, the unknown function $m(\cdot)$ is assumed to be strictly increasing in the unobservable. Since S is binary, the equivalent assumption cannot be made for $h^0(\cdot)$ and $h^1(\cdot)$. Nevertheless, the non-decreasing assumption ensures that observed units will not have lower values of the unobservable than the units that leave the sample. Third, Assumption 6 states that $U_{it} \perp\!\!\!\perp G_i \mid S_{it}$. While the corresponding assumption, $\mathcal{U}_{it} \perp\!\!\!\perp G_i \mid Y_i, V_i = AO$, is not explicitly made, it is embedded in Assumption 4. Since the function $m(u, t)$ is strictly increasing in u, it follows that $\mathcal{U}_{it} = m^{-1}(Y_{it}, t)$. Thus, when conditioning on Y and t, the distribution of $\mathcal{U}$ becomes degenerate and, therefore, identical for both groups. As a result, the conditional independence assumption is trivially satisfied.

Lemma 1 *If Assumptions 1, 2, 3, 5, and 6 hold, then the missing potential selection outcomes are given by:*

Under Positive Monotonicity,

$$\mathbb{E}[S_{i2}(0) \mid G_i = 1] = \mathbb{E}[S_{i1} \mid G_i = 1]\frac{\mathbb{E}[S_{i2} \mid G_i = 0]}{\mathbb{E}[S_{i1} \mid G_i = 0]}.$$

Under Negative Monotonicity,

$$\mathbb{E}[S_{i2}(1) \mid G_i = 0] = \mathbb{E}[S_{i1} \mid G_i = 0]\frac{\mathbb{E}[S_{i2} \mid G_i = 1]}{\mathbb{E}[S_{i1} \mid G_i = 1]}.$$

Remark 5 *Both* $\mathbb{E}[S_{i2}(0) \mid G_i = 1]$ *and* $\mathbb{E}[S_{i2}(1) \mid G_i = 0]$ *will always lie in the unit interval. They will always be well-defined, except for the case when* $\mathbb{E}[S_{i1} \mid G = g] = 0$*, which would mean that no unit is observed in the pre-treatment period in group* g*.*

Lemma 1 is equivalent to Theorem 4.2 in Athey and Imbens (2006), and the proof can be found in Athey and Imbens (2006, p. 458). Under the CiC model for selection, the missing selection outcome in the post-treatment period for group g is given by their baseline selection in the pre-treatment period scaled by the proportional change for the other group. Table 2 provides a numerical example for illustration purposes.

Table 2: Numerical example of Lemma 1

Group (G)	$\mathbb{E}[S_1]$	$\mathbb{E}[S_2(G)]$	$\mathbb{E}[S_2(1-G)]$
Treatment ($G = 1$)	0.9	0.55	0.45
Control ($G = 0$)	0.7	0.35	0.43

Notes: This table illustrates Lemma 1 with a numerical example. Columns 2 and 3 display the observed mean in the selection indicator in each of the groups. Column 2 corresponds to the pre-treatment period and column 3 to the post-treatment period. The fourth column presents the missing potential selection outcome for each of the groups, imputed using Lemma 1.

Remark 6 *The treatment effect on selection,* $\mathbb{E}[S_{i2}(1) - S_{i2}(0) \mid G_i = g]$ *will have the same sign for both groups* $g \in \{0, 1\}$*. This is consistent with the monotonicity assumption. Furthermore, these two effects will be the same if the expected selection in the pre-treatment period is the same across groups, i.e.,* $\mathbb{E}[S_{i1} \mid G_i = 1] = \mathbb{E}[S_{i1} \mid G_i = 0]$*. This aligns with the selection model in Assumption 6 as*

$$\mathbb{E}[S_{i1} \mid G_i = 1] = \mathbb{E}[S_{i1} \mid G_i = 0] \iff U_{it} \perp\!\!\!\perp G_i \implies \mathbb{E}[S_{i2}(1) - S_{i2}(0)] \perp\!\!\!\perp G_i$$

Proof: See Appendix B.2

Proposition 2 *Under Assumptions 1, 2, 3, 5, and 6, the proportion of Always-Observed units in the treatment group (π_1) and control group (π_0) are identified as follows:*

If Positive Monotonicity holds:

$$\pi_0 = 1$$
$$\pi_1 = \frac{\mathbb{E}[S_{i2}(0) \mid G_i = 1]}{\mathbb{E}[S_{i2} \mid G_i = 1]} = \frac{\mathbb{E}[S_{i1} \mid G_i = 1]}{\mathbb{E}[S_{i2} \mid G_i = 1]} \frac{\mathbb{E}[S_{i2} \mid G_i = 0]}{\mathbb{E}[S_{i1} \mid G_i = 0]} \in [0, 1]$$

If Negative Monotonicity holds:

$$\pi_0 = \frac{\mathbb{E}[S_{i2}(1) \mid G_i = 0]}{\mathbb{E}[S_{i2} \mid G_i = 0]} = \frac{\mathbb{E}[S_{i1} \mid G_i = 0]}{\mathbb{E}[S_{i2} \mid G_i = 0]} \frac{\mathbb{E}[S_{i2} \mid G_i = 1]}{\mathbb{E}[S_{i1} \mid G_i = 1]} \in [0, 1]$$
$$\pi_1 = 1$$

Proof: See Appendix B.3.

3.3 Relaxing the monotonicity assumption

Assumption 5 can be restrictive, as it requires the treatment to affect selection in only one direction for all the units, even when selection may respond to treatment through multiple channels. Similarly, Assumption 6 implies that all unobservables affecting selection can be summarized in a single scalar, ruling out different unobservables affecting selection in distinct ways. When multiple sources of selection are observed, the monotonicity assumption can be relaxed by allowing each source to exhibit monotonicity independently, potentially with different signs. For instance, consider a scholarship program designed to improve academic performance. Such a scholarship may affect selection through two distinct mechanisms: reducing dropout rates (positive monotonicity) while increasing graduation rates (negative monotonicity).

Suppose that there are J different sources of sample selection, where s_{it}^j is a binary variable equal to 0 if unit i's outcome is not observed because of source j and 1 otherwise. In the previous example, $J = 2$, with s_{it}^1 equal to 0 if the student i drops out at period t, and $s_{it}^2 = 0$ if the student graduates. By definition, these sources are mutually exclusive,

meaning that when a unit leaves the sample, it is due to a specific source j. For instance, students may leave the sample either because they graduated or because they dropped out before completing the degree, but both events cannot happen simultaneously. This framework allows the expression of the generic selection indicator as the product of the different sources of selection: $S_{it} = \prod_{j=1}^{J} s_{it}^{j}$, where the mutual exclusivity of the sources implies that $\sum_{j=1}^{J} s_{it}^{j} \in \{J-1, J\}\ \forall i, t$. This notation extends to potential selection indicators as well: $S_{it}(w) = \prod_{j=1}^{J} s_{it}^{j}(w)$.

The four strata defined in Table 1 exist for each source j. I can define analogously a variable V_i^j indicating the principal stratum of unit i defined by the joint values of the potential selection indicator of the source j, $(s_{i2}^{j}(1), s_{i2}^{j}(0))$.

Assumption 7 *Source-specific monotonicity*
For any source $j \in \{1, ..., J\}$*:*

$$s_{i2}^{j}(1) \geq s_{i2}^{j}(0) \quad \forall i \quad \textit{Positive Monotonicity}$$

$$\textit{or}$$

$$s_{i2}^{j}(1) \leq s_{i2}^{j}(0) \quad \forall i \quad \textit{Negative Monotonicity}$$

Assumption 7 is analogous to Assumption 5 defined for each specific source of selection. It rules out the existence of one stratum for each source j ($V_i^j = OC$ if positive monotonicity is assumed and $V_i^j = OT$ if negative monotonicity is assumed). Similarly, I can formulate an analogous, source-specific Assumption 6:

Assumption 8 *Source-specific Selection Model*
For any source $j \in \{1, ..., J\}$*:*
Under Positive Monotonicity,

$$s_{it}^{j}(0) = h_j^0(u_{it}^{j}, t).$$

Under Negative Monotonicity,

$$s_{it}^{j}(1) = h_j^1(u_{it}^{j}, t),$$

where u_{it}^{j} *is an unobservable scalar for unit* i *at time* t *and* $h_j^w(u, t)$ *is non decreasing function*

in $u \ \forall \ t \in \{1, 2\}$, $w = \{0, 1\}$.

The unobservable u^j *is continuously distributed and has the same compact support in both groups. Its distribution is constant over time within groups*

$$u_{i1}^j \mid G_i \sim u_{i2}^j \mid G_i.$$

Additionally, given the realized selection outcome, the distribution of u^j *is independent of the group in a given time period.*

$$u_{it}^j \perp\!\!\!\perp G_i \mid s_{it}^j$$

Assumption 6 states that unobservables affecting selection can be summarized by a single scalar that affects the eventual selection in a monotonic way. Assumption 8 relaxes this assumption by allowing for multiple unobservable scalars that may affect selection in different directions. Going back to the college example, one can think of two distinct unobservables affecting selection: ability and motivation. Higher ability increases the likelihood of graduation, whereas lower motivation increases the likelihood of dropping out. Because students with high ability and high motivation may leave the sample through graduation, while those with low ability and low motivation may leave through dropout, it is difficult to summarize these distinct selection forces using a single index.

Let $\mathcal{X}_j$ denote the set of units that belong to stratum X according to source j. For instance, $\mathcal{AO}_j = \{i : V_i^j = AO\} \equiv \{i : s_{i2}^j(1) = 1, s_{i2}^j(0) = 1\}$.

Assumption 9 *No intersection of OC and OT strata.*

$$\mathcal{OT}_j \cap \mathcal{OC}_k = \varnothing \quad \forall j, k \in \{1, ..., J\}$$

Assumption 9 rules out the scenario where a unit belongs to the OT stratum defined by source j and to the OC stratum defined by a different source. Consequently, units that belong to the OC or OT strata according to source j can only belong to the AO stratum according to all the other sources. Formally, $s_{i2}^j(1) \neq s_{i2}^j(0) \implies s_{i2}^k(1) = s_{i2}^k(0) = 1 \ \forall k \neq j$.

For instance, consider a student who drops out in the control group but remains enrolled if treated ($s_{i2}^1(0) = 0, s_{i2}^1(1) = 1$). Since sources are mutually exclusive and $s_{i2}^1(0) = 0$, it

must be true that this unit does not graduate in the control group, $s^2_{i2}(0) = 1$. Assumption 9 rules out the possibility that this student, who drops out in the control group, graduates in the treatment group, i.e., assumes that $s^2_{i2}(1) = 1$. Hence, given that the student belongs to the OT stratum defined by the joint values of the dropout source, $(s^1_{i2}(0), s^1_{i2}(1)) = (0, 1)$, they must belong to the AO stratum defined by the joint values of the graduation source, $(s^2_{i2}(1), s^2_{i2}(0)) = (1, 1)$.

Remark 7 *If sources are mutually exclusive, and Assumption 5 holds, then Assumptions 7 and 9 also hold, with all the sources of selection having the same sign of monotonicity. On the other hand, when Assumptions 7 and 9 hold, then:*

- *If all sources exhibit the same sign of monotonicity, then Assumption 5 also holds.*
- *If some sources exhibit positive monotonicity while others exhibit negative monotonicity, then Assumption 5 is violated.*

While I acknowledge that Assumptions 7 and 9 are restrictive, they relax Assumption 5 as noted in Remark 7. This relaxation, which allows the existence of the four principal strata, comes at no cost, as the proportion of AO units in each group remains point-identified. This section can therefore be viewed as a generalization of Section 3.2: when Assumptions 5 and 6 hold, then Assumptions 7, 8, and 9 also hold; when Assumptions 5 and 6 fail, then Assumptions 7, 8, and 9 may still hold. A significant limitation, though, is that different sources must exist and be identifiable within the data, which is not always guaranteed. Intuitively, if selection is driven by multiple channels operating in different directions, the assumptions in Section 3.2 are violated. If these channels manifest themselves in distinct, observable sources of selection, the assumptions in this section may still be satisfied. If not, these assumptions are violated as well.

Lemma 2 *If assumptions 7 and 9 hold, the four principal strata defined by the joint values of* $(S_{i2}(1), S_{i2}(0))$ *that partition the population are given by:*

$$\mathcal{AO} = \bigcap_j \mathcal{AO}_j \quad ; \quad \mathcal{NO} = \bigcup_j \mathcal{NO}_j \quad ; \quad \mathcal{OT} = \bigcup_j \mathcal{OT}_j \quad ; \quad \mathcal{OC} = \bigcup_j \mathcal{OC}_j$$

Proof: See Appendix B.4.

Lemma 2 states that to belong to the AO stratum, a unit must belong to the AO stratum defined by the potential selection indicators of all sources. Conversely, if a unit belongs to the NO, OT, or OC stratum for any source, it will also belong to the corresponding principal stratum defined by the overall selection indicator S. Table C1 provides an example with two different sources of selection.

Proposition 3 *Let J^+ denote the set of sources for which positive monotonicity holds and J^- the set of sources for which negative monotonicity holds. Under Assumptions 1, 2, 3, 7, 8, and 9, the proportion of Always-Observed units in the control group (π_0) and the treatment group (π_1) are identified as follows:*

$$\pi_0 = \frac{1}{\mathbb{E}[S_{i2} \mid G_i = 0]} \left(1 - \sum_{j \in J^+} \left(1 - \mathbb{E}[s_{i2}^j \mid G_i = 0]\right) - \sum_{j \in J^-} \left(1 - \mathbb{E}[s_{i1}^j \mid G_i = 0] \frac{\mathbb{E}[s_{i2}^j \mid G_i = 1]}{\mathbb{E}[s_{i1}^j \mid G_i = 1]} \right) \right)$$

$$\pi_1 = \frac{1}{\mathbb{E}[S_{i2} \mid G_i = 1]} \left(1 - \sum_{j \in J^+} \left(1 - \mathbb{E}[s_{i1}^j \mid G_i = 1] \frac{\mathbb{E}[s_{i2}^j \mid G_i = 0]}{\mathbb{E}[s_{i1}^j \mid G_i = 0]} \right) - \sum_{j \in J^-} \left(1 - \mathbb{E}[s_{i2}^j \mid G_i = 1]\right) \right)$$

Proof: See Appendix B.5.

Proposition 3 identifies the proportion of always observed in both the treatment and control groups. By Assumption 7, some sources exhibit positive monotonicity, while others exhibit negative monotonicity. Following the same logic as in Proposition 2, the former can identify the AO units in the treatment group, and the latter can identify the AO units in the control group. Appendix B.6 shows how Proposition 2 is a specific case of Proposition 3 with only one source of selection, $J = 1$.

4 Estimation and inference

This section proposes estimators for the bounds of the ATT_{AO} and $\text{QTT}_{\text{AO}}(q)$. I show that these estimators are consistent and asymptotically normal and discuss how to construct

confidence intervals for the estimated bounds. The estimators are constructed using the sample analogs of the objects in Propositions 1 and 3. For simplicity and without loss of generality, this section considers the case with two different sources of sample selection. For the first one, $j = 1$, positive monotonicity holds. For the second one, $j = 2$, negative monotonicity holds. All the summations in this section are over the entire sample.

4.1 Estimation

Let N denote the sample size, N_0 the number of units in the control group, and N_1 the number of units in the treatment group. The estimators of the proportions of Always-Observed units defined in Proposition 3 are given by:

$$\hat{\pi}_0 = \frac{1}{\frac{\sum_i S_{i2}(1-G_i)}{N_0}} \left(1 - \left(1 - \frac{\sum_i s^1_{i2}(1-G_i)}{N_0}\right) - \left(1 - \frac{\sum_i s^2_{i1}(1-G_i)}{N_0}\frac{\sum_i s^2_{i2}G_i}{\sum_i s^2_{i1}G_i}\right)\right) \tag{1}$$

$$\hat{\pi}_1 = \frac{1}{\frac{\sum_i S_{i2}G_i}{N_1}} \left(1 - \left(1 - \frac{\sum_i s^1_{i1}G_i}{N_1}\frac{\sum_i s^1_{i2}(1-G_i)}{\sum_i s^1_{i1}(1-G_i)}\right) - \left(1 - \frac{\sum_i s^2_{i2}G_i}{N_1}\right)\right) \tag{2}$$

where I have substituted the expectations in Proposition 3 with their sample analogs.

Given the estimated proportions, $\hat{\pi}_0$ and $\hat{\pi}_1$, the estimators for the bounds of the $\text{QTT}_{\text{AO}}(q)$ defined in Proposition 1 are given by:

$$\widehat{\Lambda^{LB}}(q) = \widehat{Q}_{Y_2|G=1,S_2=1}(q\hat{\pi}_1) - \widehat{Q}_{Y_2|G=0,S_2=1}\left(\hat{q}^*_{LB}\right), \tag{3}$$

$$\hat{q}^*_{LB} = \min\left\{\widehat{F}_{Y_1|G=0,S_2=1}\left(\widehat{Q}_{Y_1|G=1,S_2=1}(q\hat{\pi}_1 + 1 - \hat{\pi}_1)\right) + 1 - \hat{\pi}_0, 1\right\}, \tag{4}$$

$$\widehat{\Lambda^{UB}}(q) = \widehat{Q}_{Y_2|G=1,S_2=1}(q\hat{\pi}_1 + 1 - \hat{\pi}_1) - \widehat{Q}_{Y_2|G=0,S_2=1}\left(\hat{q}^*_{UB}\right), \tag{5}$$

$$\hat{q}^*_{UB} = \max\left\{\widehat{F}_{Y_1|G=0,S_2=1}\left(\widehat{Q}_{Y_1|G=1,S_2=1}(q\hat{\pi}_1)\right) - (1 - \hat{\pi}_0), 0\right\}, \tag{6}$$

where $\widehat{F}_Y$ denotes the empirical distribution of Y. For instance,

$$\widehat{F}_{Y_1|G=0,S_2=1}(y) = \frac{\sum_i S_{i2}(1-G_i)\mathbb{1}(Y_{i1} \leq y)}{\sum_i S_{i2}(1-G_i)}$$

$$\widehat{Q}_{Y_2|G=0,S_2=1}(q) = \inf\left\{y : \widehat{F}_{Y_2|G=0,S_2=1}(y) \geq q\right\}.$$

The bounds for the ATT_{AO} can be estimated under the CiC framework using the esti-

mated bounds for the QTT_{AO}, as stated in Remark 2:

$$\widehat{ATT}_{AO} \in \left[\int_0^1 \widehat{\Lambda^{LB}}(q)dq \quad , \quad \int_0^1 \widehat{\Lambda^{UB}}(q)dq \right] \tag{7}$$

4.2 Asymptotic Normality

The next propositions establish the consistency and asymptotic normality of the proposed estimators. These results are indispensable for conducting valid inference on the estimated bounds.

Proposition 4 *Asymptotic normality of the bounds of the* QTT_{AO}

$$\sqrt{n}(\widehat{\Lambda^{LB}}(q) - \Lambda^{LB}(q)) \xrightarrow{d} \mathcal{N}\left(0, \varsigma^2_{LB}\right) \tag{8}$$

$$\sqrt{n}(\widehat{\Lambda^{UB}}(q) - \Lambda^{UB}(q)) \xrightarrow{d} \mathcal{N}\left(0, \varsigma^2_{UB}\right) \tag{9}$$

Provided that

$$q \in (0,1),$$

$$F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q\pi_1 + 1 - \pi_1)\right) + 1 - \pi_0 < 1,$$

and

$$F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=0}(q\pi_1)\right) - (1 - \pi_0) > 0.$$

Proof: See Appendix B.7

Under Positive Monotonicity, $\pi_0 = 1$ and the bounds are asymptotically normal for any $q \in (0,1)$. When $\pi_0 \neq 1$, there might be some extreme quantiles in which the estimators in equations (3-6) are the sample maximum/minimum of $Y_2 \mid G = 0$. In these cases, the estimator for $Q_{Y_2(0)|G_i=1,S_2=1}(q)$ follows a extreme value distribution.

4.3 Confidence Intervals

Provided that the estimators of the bounds are asymptotically normal, Imbens and Manski (2004) provide an expression to construct confidence intervals around a partially identified parameter:

$$CI_\alpha = \left[\widehat{\Lambda^{LB}} - Z_\alpha \frac{\hat{\sigma}_{LB}}{\sqrt{n}} \quad , \quad \widehat{\Lambda^{UB}} + Z_\alpha \frac{\hat{\sigma}_{UB}}{\sqrt{n}}\right] \tag{10}$$

where Z_α is such that:

$$\Phi\left(Z_\alpha + \sqrt{n}\frac{\widehat{\Lambda^{UB}} - \widehat{\Lambda^{LB}}}{\max\{\hat{\sigma}_{UB}, \hat{\sigma}_{LB}\}}\right) - \Phi\left(-Z_\alpha\right) = \alpha,$$

where $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution. The interval described in equation (10) contains the set $[\Lambda^{LB}, \Lambda^{UB}]$ at least $\alpha\%$ of the times. Intuitively, when the bounds are far away from each other and/or the standard errors of the bounds are small, the confidence interval is constructed by adding or subtracting the one-side critical value associated with a confidence level α (e.g., 1.645 for a 95% confidence interval) times the standard error of the bound. This critical value increases when the bounds are close to each other or have large standard errors. In the extreme case where $\widehat{\Lambda^{LB}} = \widehat{\Lambda^{UB}}$ and the ATT_{AO} is point identified, $Z_{0.95} \approx 1.96$ and equation (10) becomes the standard confidence interval constructed around a point-identified parameter.

5 Empirical application

This section illustrates the methodology using a job training program, *Jóvenes en Acción,* implemented in Colombia between 2002 and 2005. The program provided three months of in-classroom training and three months of on-the-job training to disadvantaged youth to improve their labor market outcomes. The program has been previously studied by Attanasio et al. (2011), who found a positive effect on wages and the probability of having paid employment for women. For men, they find a positive effect on job quality but no significant impact on earnings. In a follow-up study, Attanasio et al. (2017) examined the program's long-term

effects, finding similar and persistent impacts on earnings and job quality for both men and women. More recently, Possebom and Riva (2024) revisited these results using principal stratification analysis. They partially identify the effect of the training on the probability of being employed in the formal sector for the women who are employed, regardless of their treatment status. For this stratum, they cannot reject the null that the program had no impact on formal employment.

In this paper, I use data from Attanasio et al. (2011), which consists of survey responses from 3,955 individuals. The participants were interviewed twice: initially in January 2005, before the training, and then between August and October 2006, after the training. Using the baseline data, I construct a panel dataset. Further details on data collection and program implementation are available in Attanasio et al. (2011).

I analyze the effect of the training on salaried wages. This outcome is observed only among those with paid employment, as unemployed and self-employed workers do not receive salaried wages. Of the 3,955 individuals in the sample, 1453 have paid employment in the pre-treatment period. Among these, 20% were no longer employed in salaried positions during the post-treatment period, and 19% could not be reached in the post-treatment survey. As a result, I observe salaried wages in both periods for 888 individuals. Table C3 presents these numbers separately for each treatment group.

I can define two different sources of sample selection in my sample: transitioning to unemployment/self-employment and survey non-response. For those with missing salaried earnings, I observe the specific source of selection in the data. I assume negative monotonicity for unemployment/self-employment and positive monotonicity for survey non-response.[9] Under these assumptions, I estimate the proportion of units that have paid employment in

[9]The negative monotonicity assumption implies that all workers who did not transition into unemployment or self-employment after receiving training would also have retained their salaried positions in the absence of treatment. In the context of disadvantaged youth in Colombia, many salaried jobs are informal. As a result, training may reduce the probability of holding a salaried position by increasing treated workers' reservation wages, enhancing their entrepreneurial skills and likelihood of self-employment, or raising aspirations that lead to longer job-search durations in the short run following program completion. By contrast, the positive monotonicity assumption implies that all control units that were reachable in the post-treatment period would also have responded to the survey had they been treated. This assumption is plausible if program participation increases respondents' commitment to survey participation and if attrition is correlated with migration, which the treatment may reduce.

both treatment arms among my final sample of 888 workers. That is, I estimate the proportions of Always-Observed units, π_1 and π_0 as defined in Proposition 5, using the estimators described in equations (1) and (2). I estimate $\hat{\pi}_1 = 0.93$ and $\hat{\pi}_0 = 0.96$. In other words, of the 501 units observed in both periods in the treatment group and the 387 units in the control group, 93% of the treated units and 96% of the control group units would have also been observed in the opposite group.

Using these estimated proportions, I partially identify the ATT_{AO}. Table 3 reports the estimates of the bounds with and without the inclusion of covariates.[10] For comparison, columns 3 and 4 present results from a complete-case analysis, which discards observations with missing outcomes and applies the Changes-in-Changes estimator to the subsample of units with observed outcomes.

Table 3: Estimates of the bounds for the ATT_{AO}

Outcome	Log of salaried earnings			
Estimand	ATT_{AO}		Complete Case	
Estimate	[-0.11 , 0.429]	[-0.095 , 0.319]	0.129	0.123
95% CI	(-0.169 , 0.527)	(-0.158 , 0.416)	(0.037 , 0.22)	(0.031 , 0.215)
Covariates	No	Yes	No	Yes
N	888	888	888	888

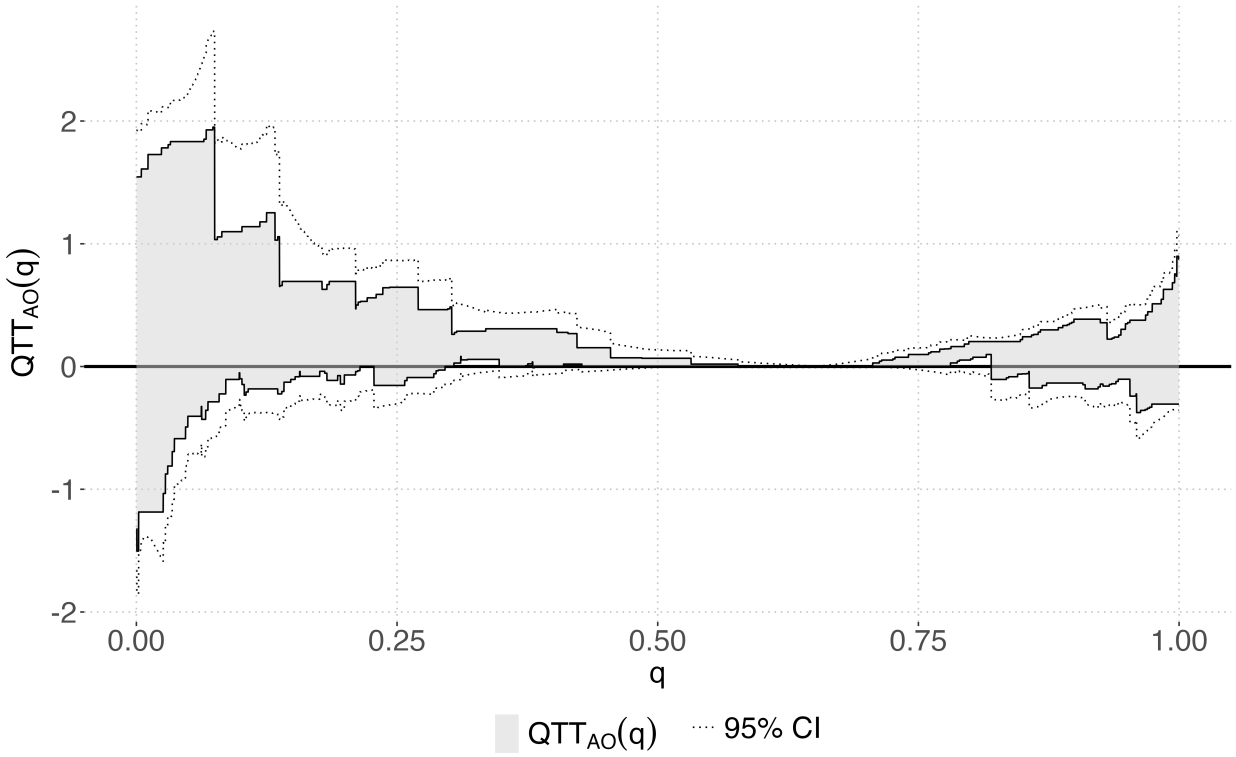
Notes: Columns 1 and 2 present the estimates of the bounds for the ATT_{AO} using the estimated proportions $\hat{\pi}_1 = 0.93$ and $\hat{\pi}_0 = 0.96$. Column 1 reports the bounds estimated without covariates with the estimators presented in Section 4, while Column 2 includes education as a covariate as explained in Appendix A.3. Columns 3 and 4 show estimates using the complete case analysis without and with covariates respectively. The 95% confidence intervals are computed using equation (10).

Table 3 illustrates the importance of the methodology developed in this paper. Disregarding the problem of sample selection and estimating the ATT using units with observed outcomes suggests that the training had a positive and significant effect on salaried earnings, increasing them by 12% for treated units. However, this increase may be driven by differences in the distribution of outcomes across strata. Instead, I estimate the intensive-margin effect of the training by targeting treated units with paid employment regardless of their treatment status. The bounds for this effect include 0. Accordingly, I cannot reject the null hypothesis that the program had no impact on wages at any significance level.

[10] The covariate adjustment is described in Appendix A.3.

In this context, it is relevant to examine the distributional effects. Suppose the treatment effects are heterogeneous along the outcome distribution. In that case, the bounds for the average effect may include zero even if the effect is positive and significant in some parts of the distribution. For instance, consider two different training programs. The first only benefits workers at the top of the salaried earnings distribution. The other only benefits workers at the bottom. These two policies may generate the same bounds for the ATT_{AO}. However, their policy implications change drastically: while the first program increases wage dispersion, the second helps reduce earnings inequality. Figure 2 plots the $\text{QTT}_{\text{AO}}(q)$ for different quantiles q.

Figure 2: $\text{QTT}_{\text{AO}}(q)$. Outcome: log of salaried earnings



Notes: Quantile Treatment effect on the Treated Always Observed units ($\text{QTT}_{\text{AO}}(q)$) for different values of q. The number of units used to compute these bounds is 888.

Figure 2 shows that the $\text{QTT}_{\text{AO}}(q)$ bounds vary across different quantiles, reflecting heterogeneity in the treatment effects. The 95% confidence interval includes zero for all the quantiles. However, their widths and ranges differ substantially across the outcome distribution. The bounds are vast and not very informative at the lower quantiles. However, as q approaches the median, the bounds tighten and the lower bound approaches or even

exceeds zero. A precise zero effect is estimated between the median and the third quartile. Finally, a moderate but positive effect is estimated around the eighth decile, before the bounds widen again towards the upper tail of the distribution.

6 Conclusion and extensions

This paper develops a novel identification strategy for estimating the intensive margin effect in panel data settings. I propose two new estimands, the ATT_{AO} and the $\text{QTT}_{\text{AO}}(q)$, which can be interpreted as the average and the quantile intensive-margin effects, respectively. I explore partial identification of these causal quantities within a CiC framework, which allows the treatment to be confounded with unobservable unit-level characteristics. Appendix A extends the methodology presented in the main text in different directions. Appendix A.1 replaces the Changes-in-Changes outcome model with the canonical Difference-in-Differences framework. Appendix A.2 revisits the methodology without Assumption 2, which treats missingness as an absorbing state. Appendix A.3 studies how to incorporate covariates into the identification and estimation strategies. Appendix A.4 extends the identification to repeated cross-sectional data. Appendix A.5 explores the identification when the outcome is discrete. Yet several extensions remain for future research.

The intensive margin effects are instrumental for policy evaluation. Beyond estimating the intensive margin effect, I encourage practitioners to examine the extensive margin effect by characterizing all the principal strata, including those in the Observed-if-Control and Observed-if-Treated strata. Furthermore, the two estimands proposed in this paper focus on the treatment group. Since the treatment is not unconfounded, its effect will likely differ across the treatment and control groups. However, the CiC framework can be extended to impute the missing potential outcome for control units, thereby enabling the identification of the average treatment effect for the control group. As a result, the proposed methodology can accommodate the identification of two additional estimands, the ATE_{AO} and the $\text{QTE}_{\text{AO}}(q)$, enhancing the external validity of the conclusions drawn from the analysis.

This paper focuses on the two-group, two-period case. However, many applications in-

volve multiple pre- and post-treatment periods. Incorporating this information is essential for both assessing the identification assumptions and gaining a deeper understanding of the policy's impacts. Moreover, many policies are implemented in a staggered fashion. While the proposed methodology can be smoothly extended to the multiple groups, multiple periods setting through the estimation and aggregation of pairwise comparisons (Athey & Imbens, 2006; Sun & Abraham, 2021), a more ambitious and valuable direction for research would be to integrate it with doubly robust estimators that combine outcome modeling with propensity score weighting (Sant'Anna & Zhao, 2020; Arkhangelsky & Imbens, 2022). Another promising direction for future research is the extension to time-varying estimands, as proposed by Lin et al. (2008) and Comment et al. (2025), which may be particularly policy-relevant in dynamic settings.

One limitation of the proposed methodology is that the derived bounds may be uninformative. In general, lower proportions of Always-Observed units lead to wider bounds, and even when this proportion is close to one, certain outcome distributions may still generate wide identification regions. Adopting a Bayesian perspective may help infer where the target estimand is most likely to lie within the bounds. Another approach is to tighten the bounds by incorporating covariates, as discussed in Appendix A.3. Under treatment unconfoundedness, covariate-adjusted bounds are weakly tighter than unadjusted bounds (Grilli & Mealli, 2008; Lee, 2009; Long & Hudgens, 2013; Mealli & Pacini, 2013). However, in the setting studied here, where treatment is confounded and covariate distributions can differ substantially across groups, there is no guarantee that covariate adjustment will tighten the bounds. Determining whether, how, and which covariates should be included in panel data settings remains an important topic for future research.

Finally, this paper relaxes the monotonicity assumptions while maintaining the point identification of principal strata proportions. This extension may also be of interest in settings under unconfoundedness. Nevertheless, it relies on multiple sources of sample selection. Furthermore, it maintains the monotonicity assumption for each available source. If the researcher does not observe multiple sources of selection, believes that these multiple sources do not capture all the unobservables that may affect selection in different directions, and/or

does not want to rely on monotonicity of any form, Rathnayake et al. (2024) and Shin (2024) provide identification of the ATT_{AO} without monotonicity for the DiD case. However, they replace the monotonicity assumption with the assumption that the average treatment effect on selection is the same in the treatment and control groups[11], as in settings where treatment is unconfounded. An interesting approach to address the monotonicity assumption is the automated method proposed by Duarte et al. (2024), which can be used to assess the plausibility of the assumption and derive sharp bounds that relax or even remove it.

References

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2020). Sampling-Based versus Design-Based Uncertainty in Regression Analysis. *Econometrica*, *88*(1), 265–296. https://doi.org/10.3982/ECTA12675

Angrist, J. D. (2001). Estimation of Limited Dependent Variable Models With Dummy Endogenous Regressors: Simple Strategies for Empirical Practice. *Journal of Business & Economic Statistics*, *19*(1), 2–28. https://doi.org/10.1198/07350010152472571

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, *91*(434), 444–455. https://doi.org/10.1080/01621459.1996.10476902

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press. https://doi.org/10.2307/j.ctvcm4j72

Arkhangelsky, D., & Imbens, G. (2024). Causal models for longitudinal and panel data: A survey. *The Econometrics Journal*, *27*(3), C1–C61. https://doi.org/10.1093/ectj/utae014

Arkhangelsky, D., & Imbens, G. W. (2022). Doubly robust identification for causal panel data models. *The Econometrics Journal*, *25*(3), 649–674. https://doi.org/10.1093/ectj/utac019

Athey, S., & Imbens, G. W. (2006). Identification and Inference in Nonlinear Difference-in-Differences Models. *Econometrica*, *74*(2), 431–497. https://doi.org/10.1111/j.1468-0262.2006.00668.x

Attanasio, O., Guarín, A., Medina, C., & Meghir, C. (2017). Vocational Training for Disadvantaged Youth in Colombia: A Long-Term Follow-Up. *American Economic Journal: Applied Economics*, *9*(2), 131–143. https://doi.org/10.1257/app.20150554

Attanasio, O., Kugler, A., & Meghir, C. (2011). Subsidizing Vocational Training for Disadvantaged Youth in Colombia: Evidence from a Randomized Trial. *American Economic Journal: Applied Economics*, *3*(3), 188–220. https://doi.org/10.1257/app.3.3.188

Bartalotti, O., Kédagni, D., & Possebom, V. (2023). Identifying marginal treatment effects in the presence of sample selection. *Journal of Econometrics*, *234*(2), 565–584. https://doi.org/10.1016/j.jeconom.2021.11.011

Bhattacharya, D. (2008). Inference in panel data models under attrition caused by unobservables. *Journal of Econometrics*, *144*(2), 430–446. https://doi.org/10.1016/j.jeconom.2008.03.002

[11] In the case of Rathnayake et al. (2024), conditional on the pre-treatment selection outcome.

Blanco, G., Chen, X., Flores, C. A., & Flores-Lagunes, A. (2020). Bounds on Average and Quantile Treatment Effects on Duration Outcomes Under Censoring, Selection, and Noncompliance. *Journal of Business & Economic Statistics*, *38*(4), 901–920. https://doi.org/10.1080/07350015.2019.1609975

Carlson, A. L., & Semykina, A. (2024, September). Addressing Attrition in Nonlinear Dynamic Panel Data Models with an Application to Health. https://doi.org/10.2139/ssrn.4965274

Chen, X., & Flores, C. A. (2015). Bounds on Treatment Effects in the Presence of Sample Selection and Noncompliance: The Wage Effects of Job Corps. *Journal of Business & Economic Statistics*.

Comment, L., Mealli, F., Haneuse, S., & Zigler, C. M. (2025). Survivor Average Causal Effects for Continuous Time: A Principal Stratification Approach to Causal Inference With Semicompeting Risks. *Biometrical Journal*, *67*(2), e70041. https://doi.org/10.1002/bimj.70041

Cornwell, C. M., Lee, K. H., & Mustard, D. B. (2005). Student Responses to Merit Scholarship Retention Rules. *The Journal of Human Resources*, *40*(4), 895–917. https://www.jstor.org/stable/4129544

Ding, P., Geng, Z., Yan, W., & Zhou, X.-H. (2011). Identifiability and Estimation of Causal Effects by Principal Stratification With Outcomes Truncated by Death. *Journal of the American Statistical Association*, *106*(496), 1578–1591. https://www.jstor.org/stable/23239560

Duarte, G., Finkelstein, N., Knox, D., Mummolo, J., & Shpitser, I. (2024). An Automated Approach to Causal Inference in Discrete Settings. *Journal of the American Statistical Association*, *119*(547), 1778–1793. https://doi.org/10.1080/01621459.2023.2216909

Fitzgerald, J., Gottschalk, P., & Moffitt, R. (1998). An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics. *The Journal of Human Resources*, *33*(2), 251–299. https://doi.org/10.2307/146433

Frangakis, C. E., & Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, *58*(1), 21–29. https://doi.org/10.1111/j.0006-341x.2002.00021.x

Ghanem, D., Hirshleifer, S., Kédagni, D., & Ortiz-Becerra, K. (2024). Correcting attrition bias using changes-in-changes. *Journal of Econometrics*, *241*(2), 105737. https://doi.org/10.1016/j.jeconom.2024.105737

Ghanem, D., Sant'Anna, P. H. C., & Wüthrich, K. (2025). Selection and parallel trends [arXiv preprint arXiv:2203.09001]. https://doi.org/10.48550/arXiv.2203.09001

Greenstone, M., List, J. A., & Syverson, C. (2012). The Effects of Environmental Regulation on the Competitiveness of U.S. Manufacturing [NBER Working Paper]. https://doi.org/10.3386/w18392

Grilli, L., & Mealli, F. (2008). Nonparametric Bounds on the Causal Effect of University Studies on Job Opportunities Using Principal Stratification. *Journal of Educational and Behavioral Statistics*, *33*(1), 111–130. Retrieved May 22, 2024, from https://www.jstor.org/stable/20172106

Gröger, A., & Zylberberg, Y. (2016). Internal Labor Migration as a Shock Coping Strategy: Evidence from a Typhoon. *American Economic Journal: Applied Economics*, *8*(2), 123–153. https://doi.org/10.1257/app.20140362

Grossi, G., Mariani, M., Mattei, A., & Mealli, F. (2023, December). Bayesian principal stratification with longitudinal data and truncation by death. https://doi.org/10.48550/arXiv.2401.00196

Ham, J. C., & Lalonde, R. J. (1996). The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training. *Econometrica*, *64*(1), 175–205. https://doi.org/10.2307/2171928

Hausman, J. A., & Wise, D. A. (1979). Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment. *Econometrica*, *47*(2), 455–473. https://doi.org/10.2307/1914193

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, *47*(1), 153–161. https://doi.org/10.2307/1912352

Heckman, J. J., Lalonde, R. J., & Smith, J. A. (1999). The Economics and Econometrics of Active Labor Market Programs. In *Handbook of Labor Economics* (pp. 1865–2097, Vol. 3). Elsevier. https://linkinghub.elsevier.com/retrieve/pii/S1573446399030126

Honoré, B. E., & Hu, L. (2020). Selection Without Exclusion. *Econometrica*, *88*(3), 1007–1029. https://doi.org/10.3982/ECTA16481

Horowitz, J. L., & Manski, C. F. (1995). Identification and Robustness with Contaminated and Corrupted Data. *Econometrica*, *63*(2), 281–302. https://doi.org/10.2307/2951627

Horowitz, J. L., & Manski, C. F. (2000). Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data. *Journal of the American Statistical Association*, *95*(449), 77–84. https://doi.org/10.2307/2669526

Horowitz, J. L., & Manski, C. F. (2006). Identification and estimation of statistical functionals using incomplete data. *Journal of Econometrics*, *132*(2), 445–459. https://doi.org/10.1016/j.jeconom.2005.02.007

Huber, M., & Mellace, G. (2015). Sharp Bounds on Causal Effects under Sample Selection. *Oxford Bulletin of Economics and Statistics*, *77*(1), 129–151. https://doi.org/10.1111/obes.12056

Imai, K. (2008). Sharp bounds on the causal effects in randomized experiments with "truncation-by-death". *Statistics & Probability Letters*, *78*(2), 144–149. https://doi.org/10.1016/j.spl.2007.05.015

Imbens, G. W., & Manski, C. F. (2004). Confidence Intervals for Partially Identified Parameters. *Econometrica*, *72*(6), 1845–1857. https://doi.org/10.1111/j.1468-0262.2004.00555.x

Imbens, G. W., & Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press. https://doi.org/10.1017/CBO9781139025751

Kang, B., & Lee, S. (2024). Robust Asymptotic and Bootstrap Inference for Nonsmooth GMM [SSRN Working Paper]. https://doi.org/10.2139/ssrn.4836402

Krueger, A. B. (1999). Experimental Estimates of Education Production Functions*. *The Quarterly Journal of Economics*, *114*(2), 497–532. https://doi.org/10.1162/003355399556052

Lee, D. S. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *The Review of Economic Studies*, *76*(3), 1071–1102. https://doi.org/10.1111/j.1467-937X.2009.00536.x

Lee, Y.-Y., & Liu, C.-A. (2025). Lee Bounds with a Continuous Treatment in Sample Selection [arXiv preprint arXiv:2411.04312]. https://doi.org/10.48550/arXiv.2411.04312

Lin, J. Y., Have, T. R. T., & Elliott, M. R. (2008). Longitudinal Nested Compliance Class Model in the Presence of Time-Varying Noncompliance. *Journal of the American Statistical Association*, *103*(482), 462–473. https://www.jstor.org/stable/27640071

Liu, L., Wang, Y., & Xu, Y. (2024). A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data. *American Journal of Political Science*, *68*(1), 160–176. https://doi.org/10.1111/ajps.12723

Long, D. M., & Hudgens, M. G. (2013). Sharpening Bounds on Principal Effects with Covariates. *Biometrics*, *69*(4), 812–819. https://doi.org/10.1111/biom.12103

Manski, C. F. (1989). Anatomy of the Selection Problem. *The Journal of Human Resources*, *24*(3), 343–360. https://doi.org/10.2307/145818

Manski, C. F. (1990). Nonparametric Bounds on Treatment Effects. *The American Economic Review*, *80*(2), 319–323.

Mattei, A., Mealli, F., & Pacini, B. (2014). Identification of Causal Effects in the Presence of Nonignorable Missing Outcome Values. *Biometrics*, *70*(2), 278–288.

Mealli, F., & Mattei, A. (2012). A refreshing account of principal stratification. *The International Journal of Biostatistics*, *8*(1). https://doi.org/10.1515/1557-4679.1380

Mealli, F., & Pacini, B. (2013). Using Secondary Outcomes to Sharpen Inference in Randomized Experiments With Noncompliance. *Journal of the American Statistical Association*, *108*(503), 1120–1131.

Newey, W. K., & McFadden, D. (1994). Chapter 36 Large sample estimation and hypothesis testing. In *Handbook of Econometrics* (pp. 2111–2245, Vol. 4). Elsevier.

Possebom, V., & Riva, F. (2024). Probability of Causation with Sample Selection: A Reanalysis of the Impacts of Jóvenes en Acción on Formality. *Journal of Business & Economic Statistics*, *0*(0), 1–10. https://doi.org/10.1080/07350015.2024.2388639

Rathnayake, G., Negi, A., Bartalotti, O., & Zhao, X. (2024). Difference-in-Differences with Sample Selection [arXiv preprint arXiv:2411.09221]. https://doi.org/10.48550/arXiv.2411.09221

Rosenbaum, P. R. (1984). The Consequences of Adjustment for a Concomitant Variable that Has Been Affected by the Treatment. *Royal Statistical Society. Journal. Series A: General*, *147*(5), 656–666. https://doi.org/10.2307/2981697

Roth, J., & Sant'Anna, P. H. C. (2023). When Is Parallel Trends Sensitive to Functional Form? *Econometrica*, *91*(2), 737–747. https://doi.org/10.3982/ECTA19402

Rubin, D. B. (1980). Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment. *Journal of the American Statistical Association*, *75*(371), 591–593. https://doi.org/10.2307/2287653

Rubin, D. B. (1990). [On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9.] Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies. *Statistical Science*, *5*(4), 472–480. https://doi.org/10.1214/ss/1177012032

Rubin, D. B. (2006). Causal Inference through Potential Outcomes and Principal Stratification: Application to Studies with "Censoring" Due to Death. *Statistical Science*, *21*(3), 299–309. https://www.jstor.org/stable/27645760

Samii, C., Wang, Y., & Zhou, J. A. (2025). Generalizing Trimming Bounds for Endogenously Missing Outcome Data Using Random Forests. *Political Analysis*, 1–15. https://doi.org/10.1017/pan.2025.10001

Sant'Anna, P. H. C., & Zhao, J. (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics*, *219*(1), 101–122. https://doi.org/10.1016/j.jeconom.2020.06.003

Semenova, V. (2025). Generalized Lee bounds. *Journal of Econometrics*, *251*, 106055. https://doi.org/10.1016/j.jeconom.2025.106055

Semykina, A., & Wooldridge, J. M. (2010). Estimating panel data models in the presence of endogeneity and selection. *Journal of Econometrics*, *157*(2), 375–380. https://doi.org/10.1016/j.jeconom.2010.03.039

Shin, S. (2024). Difference-in-differences Design with Outcomes Missing Not at Random [arXiv preprint arXiv:2411.18772]. https://doi.org/10.48550/arXiv.2411.18772

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, *225*(2), 175–199. https://doi.org/10.1016/j.jeconom.2020.09.006

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

Zhang, J. L., & Rubin, D. B. (2003). Estimation of Causal Effects via Principal Stratification When Some Outcomes Are Truncated by "Death". *Journal of Educational and Behavioral Statistics*, *28*(4), 353–368. https://www.jstor.org.eui.idm.oclc.org/stable/3701340

Zhang, J. L., Rubin, D. B., & Mealli, F. (2009). Likelihood-Based Analysis of Causal Effects of Job-Training Programs Using Principal Stratification. *Journal of the American Statistical Association*, *104*(485), 166–176.

A Extensions and additional results

A.1 Identification in the DiD setup

Assumption 10 *Principal Parallel Trends (PPT)*

$$\mathbb{E}[Y_{i2}(0) - Y_{i1} \mid G_i = 1, V_i = AO] = \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid G_i = 0, V_i = AO]$$

Assumption 10 states that, in the absence of treatment, the difference in the outcome between groups remains constant over time for the AO units. It is a conditional version of the canonical parallel trends assumption. Like other conditional parallel trends assumptions, it is neither a sufficient nor a necessary condition for the unconditional parallel trend assumption. Under Assumptions 1, 2 and 10, the missing potential outcome for treated Always-Observed units is given by:

$$\mathbb{E}[Y_{i2}(0) \mid G_i = 1, V_i = AO] = \mathbb{E}[Y_{i1} \mid G_i = 1, V_i = AO] + \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid G_i = 0, V_i = AO] \tag{A1}$$

Then, I could plug in the term in equation (A1) into the Estimand 1 to obtain:

$$\begin{aligned} \text{ATT}_{\text{AO}} &= \mathbb{E}[Y_{i2}(1) - Y_{i1} \mid G_i = 1, V_i = AO] - \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid G_i = 0, V_i = AO] \\ &= \mathbb{E}[Y_{i2} - Y_{i1} \mid G_i = 1, V_i = AO] - \mathbb{E}[Y_{i2} - Y_{i1} \mid G_i = 0, V_i = AO]. \end{aligned} \tag{A2}$$

Proposition 5 *Let $Y_{it}(1)$ and $Y_{it}(0)$ be continuous. If Assumptions 1, 2, 3 and 10 hold, then Δ^{LB} and Δ^{UB} are lower and upper bounds for the Average Treatment Effect on the Treated Always-Observed units (ATT_{AO}), where*

$$\begin{aligned} \Delta^{LB} &= \mathbb{E}[\ddot{Y}_i \mid G_i = 1, S_{i2} = 1, \ddot{Y}_i \leq \ddot{y}^1_{\pi_1}] - \mathbb{E}[\ddot{Y}_i \mid G_i = 0, S_{i2} = 1, \ddot{Y}_i \geq \ddot{y}^0_{1-\pi_0}] \\ \Delta^{UB} &= \mathbb{E}[\ddot{Y}_i \mid G_i = 1, S_{i2} = 1, \ddot{Y}_i \geq \ddot{y}^1_{1-\pi_1}] - \mathbb{E}[\ddot{Y}_i \mid G_i = 0, S_{i2} = 1, \ddot{Y}_i \leq \ddot{y}^0_{\pi_0}] \\ \ddot{Y}_i &= Y_{i2} - Y_{i1} \\ \ddot{y}^g_q &= \inf\{\ddot{y} : F(\ddot{y}) \geq q\}, \text{ with } F \text{ the c.d.f. of } \ddot{Y} \text{ conditional on } S_2 = 1 \text{ and } G = g \\ \pi_1 &= Pr(S_{i2}(0) = 1 \mid G_i = 1, S_{i2}(1) = 1) \\ \pi_0 &= Pr(S_{i2}(1) = 1 \mid G_i = 0, S_{i2}(0) = 1). \end{aligned}$$

Proof: See Appendix B.8.

Proposition 5 describes a trimming procedure to bound the ATT_{AO}, as proposed by Lee (2009). If the proportion of Always-Observed units in each group is known, one can use them to trim the observed distribution of $\ddot{Y} = Y_2 - Y_1$ and partially identify $\mathbb{E}[\ddot{Y}_i \mid G_i = g, V_i = AO]$. For instance, suppose 90% of the units in the treatment group belong to the AO stratum. In this case, the expected value of $\ddot{Y}$ using only the bottom 90% of the distribution of $\ddot{Y}$ provides a lower bound of the expected value of $\ddot{Y}$ for these units. Similarly, using the top 90% of the distribution provides an upper bound for the expected value of $\ddot{Y}$ for the AO units in the treatment group.

Figure A1 provides graphical intuition for Proposition 5. The distributions on the left correspond to the treatment group and are trimmed using π_1. The ones on the right correspond to the control group and are trimmed using π_0. The lower bound is computed using the bottom π_1 units from the treatment group and the top π_0 units from the control group. Likewise, the upper bound is obtained using the top π_1 units from the treatment group and the bottom π_0 units from the control group.

A.1.1 Estimation and inference in the DiD setup

The estimators for the bounds of the ATT_{AO} defined in Proposition 5 are given by:

$$\hat{\ddot{y}}^1_q = \min\left\{\ddot{y} : \frac{\sum_i S_{i2} G_i \mathbb{1}(\ddot{Y}_i \leq \ddot{y})}{\sum_i S_{i2} G_i} \geq q\right\} \tag{A3}$$

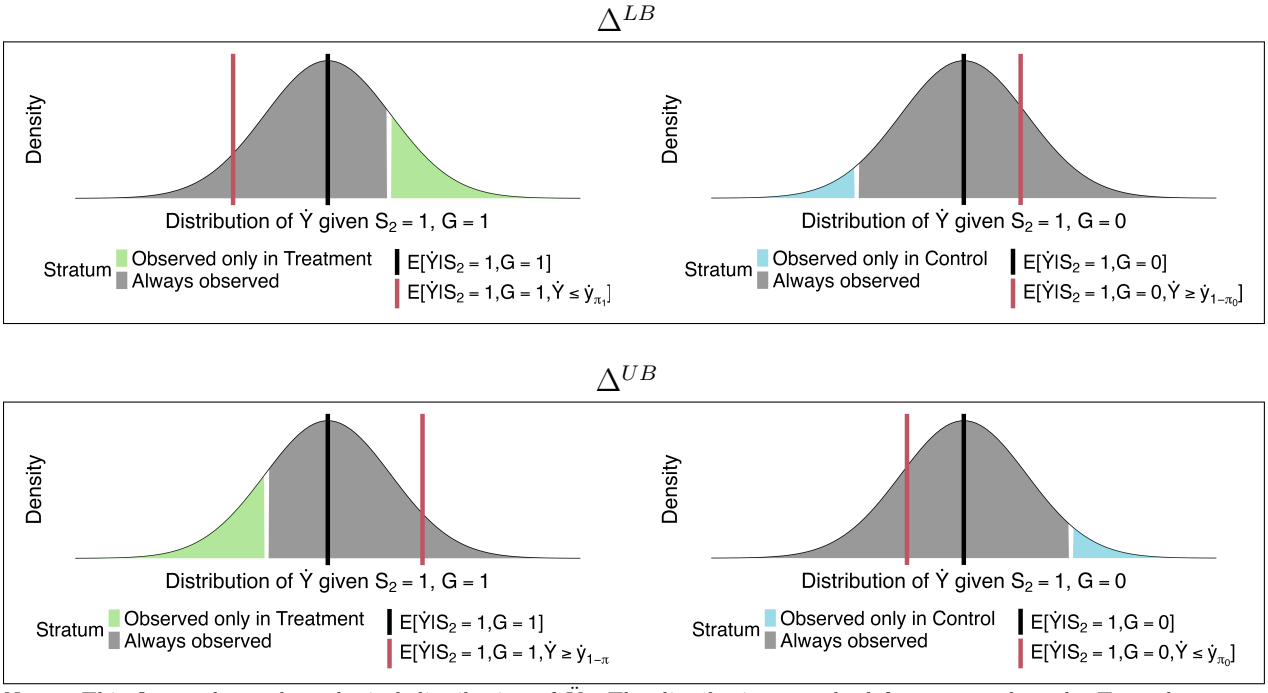
$$\hat{\ddot{y}}^0_q = \min\left\{\ddot{y} : \frac{\sum_i S_{i2}(1-G_i) \mathbb{1}(\ddot{Y}_i \leq \ddot{y})}{\sum_i S_{i2}(1-G_i)} \geq q\right\}. \tag{A4}$$

$$\widehat{\Delta^{LB}} = \frac{\sum_i S_{i2} G_i \mathbb{1}(\ddot{Y}_i \leq \hat{\ddot{y}}^1_{\hat{\pi}_1}) \ddot{Y}_i}{\sum_i S_{i2} G_i \mathbb{1}(\ddot{Y}_i \leq \hat{\ddot{y}}^1_{\hat{\pi}_1})} - \frac{\sum_i S_{i2}(1-G_i) \mathbb{1}(\ddot{Y}_i \geq \hat{\ddot{y}}^0_{1-\hat{\pi}_0}) \ddot{Y}_i}{\sum_i S_{i2}(1-G_i) \mathbb{1}(\ddot{Y}_i \geq \hat{\ddot{y}}^0_{1-\hat{\pi}_0})} \tag{A5}$$

$$\widehat{\Delta^{UB}} = \frac{\sum_i S_{i2} G_i \mathbb{1}(\ddot{Y}_i \geq \hat{\ddot{y}}^1_{1-\hat{\pi}_1}) \ddot{Y}_i}{\sum_i S_{i2} G_i \mathbb{1}(\ddot{Y}_i \geq \hat{\ddot{y}}^1_{1-\hat{\pi}_1})} - \frac{\sum_i S_{i2}(1-G_i) \mathbb{1}(\ddot{Y}_i \leq \hat{\ddot{y}}^0_{\hat{\pi}_0}) \ddot{Y}_i}{\sum_i S_{i2}(1-G_i) \mathbb{1}(\ddot{Y}_i \leq \hat{\ddot{y}}^0_{\hat{\pi}_0})} \tag{A6}$$

Equations (A3) and (A4) estimate the trimming threshold for a given trimming proportion. Equations (A5) and (A6) estimate the expectations with their sample analogs, taking the trimming threshold and proportions as given.

Figure A1: Graphical intuition on Proposition 5.



Notes: This figure plots a hypothetical distribution of $\ddot{Y}$. The distributions on the left correspond to the Treated group ($G = 1$) and the one on the right to the Control group ($G = 0$). The shaded areas correspond to the proportion of strata in the given group: black corresponds to the AO units (π_1 on the left and π_0 on the right), green corresponds to the OT stratum and blue to the OC. The vertical black lines are the expectation of the plotted distribution. The vertical red lines are the trimmed distributions used in Proposition 5. This figure also illustrates Remark 4: as π_1 and π_0 go to 1, the black shadowed areas expand, and the trimmed (red) expectations converge to those of the entire distribution (black).

Proposition 6 *Asymptotic normality of the bounds of the* ATT_{AO}

$$\sqrt{n}(\widehat{\Delta^{LB}} - \Delta^{LB}) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2_{LB}\right) \tag{A7}$$

$$\sqrt{n}(\widehat{\Delta^{UB}} - \Delta^{UB}) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2_{UB}\right), \tag{A8}$$

Proof: See Appendix B.9

A.1.2 Empirical Application

Table A1: Estimates of the bounds for the ATT_{AO}

Outcome	Log of salaried earnings	
Estimand	ATT_{AO}	Complete Case
Estimate	[-0.013 , 0.322]	0.158
95% CI	(-0.082 , 0.39)	(0.075 , 0.241)
Covariates	No	No
N	888	888

Notes: Columns 1 presents the estimates of the bounds for the ATT_{AO} using the estimated proportions $\hat{\pi}_1 = 0.93$ and $\hat{\pi}_0 = 0.96$, with the estimators presented in Section A.1.1, while Column 2 shows estimates using the complete case analysis. The 95% confidence intervals are computed using equation (10).

A.2 Identification without Assumption 2, Missingness as an absorbing state.

Assumption 2 states that once a unit's outcome is unobserved, it remains unobserved in all subsequent periods. There are settings in which this assumption is not plausible. For instance, a job-training program in which a worker is unemployed during the pre-treatment period but later finds a job. In this case, there would be some i in the data such that $S_{i1} = 0$ but $S_{i2} = 1$, violating Assumption 2. Next, I discuss the identification for settings in which missingness is not an absorbing state.

Consider the Estimand 1, ATT_{AO}. Using the Law of Total Probability, I can write it as:

$$\begin{aligned}\text{ATT}_{\text{AO}} =&\mathbb{E}[Y_{i2}(1) - Y_{i2}(0) \mid G_i = 1, V_i = AO, S_{i1} = 1]Pr(S_{i1} = 1 \mid G_i = 1, V_i = AO)+ \\ &\mathbb{E}[Y_{i2}(1) - Y_{i2}(0) \mid G_i = 1, V_i = AO, S_{i1} = 0]Pr(S_{i1} = 0 \mid G_i = 1, V_i = AO)\end{aligned} \tag{A9}$$

Under Assumption 2,

$$V_i = AO \implies S_{i2} = 1 \implies S_{i1} \neq 0$$

and therefore, $Pr(S_{i1} = 0 \mid G_i = 1, V_i = AO) = 0$. Hence, under Assumption 2, the following equality is true:

$$\text{ATT}_{\text{AO}} = \mathbb{E}[Y_{i2}(1) - Y_{i2}(0) \mid G_i = 1, V_i = AO, S_{i1} = 1]. \tag{A10}$$

Sections 3 and 4 discuss identification, estimation, and inference for the RHS of equation (A10). The conditioning on $S_1 = 1$ is not being made explicit as under Assumption 2 it is implied by the conditioning on $V_i = AO$.

When missingness is not an absorbing state, some units may have observed outcomes only in the post-treatment period, i.e., $S_{i1} = 0$ and $S_{i2} = 1$. Such units may belong to the Always-Observed stratum. In that case, $Pr(S_{i1} = 0 \mid V_i = AO)$ is no longer 0 and the equality in equation (A10) is no longer true.[12] Importantly, without Assumption 2, the identification strategy developed in Sections 3 and 4 still identifies a causal effect, since S_1 is not affected by the treatment.[13] However, the resulting estimand is no longer the ATT_{AO} but rather the ATT_{AO} for units observed in the first period, as in the RHS of equation (A10).

In some contexts, researchers may want to estimate the causal effect for all the treated units in the AO stratum, regardless of whether they were observed in the pre-treatment period. Consider the case of the ATT_{AO}:

$$\text{ATT}_{\text{AO}} = \mathbb{E}[Y_{i2}(1) - Y_{i2}(0) \mid G_i = 1, V_i = AO] \tag{A11}$$

$$= \mathbb{E}[Y_{i2}(1) \mid G_i = 1, V_i = AO] - \mathbb{E}[Y_{i2}(0) \mid G_i = 1, V_i = AO] \tag{A12}$$

The first term in equation (A12) can be partially identified using the identification results from Section 3 using all the units observed at $t = 2$ instead of conditioning on those with $S_{i1} = 1$. However, because now $\mathbb{E}[S_{i2}]$ may be larger than $\mathbb{E}[S_{i1}]$, Proposition 2 needs to be more general:

Proposition 7 *Under Assumptions 1, 3, 5, and 6, the proportion of Always-Observed units in the treatment group (π_1) and control group (π_0) are identified as follows:*

[12] Unless one is willing to assume that pre-treatment selection is ignorable, that is, $Y_{i2}(w_{i2}) \mid G_i = 1, V_i = AO \perp\!\!\!\perp S_{i1} \quad \forall \quad w_{i2}$.

[13] Albeit the conditioning on $S_{i1} = 1$ must be made explicit in all the assumptions, estimands, and identification results.

If Positive Monotonicity holds:

$$\pi_0 = 1$$

$$\pi_1 = \frac{\mathbb{E}[S_{i2}(0) \mid G_i = 1]}{\mathbb{E}[S_{i2} \mid G_i = 1]} = \frac{\mathbb{E}[S_{i1} \mid G_i = 1]}{\mathbb{E}[S_{i2} \mid G_i = 1]} \frac{\mathbb{E}[S_{i2} \mid G_i = 0]}{\mathbb{E}[S_{i1} \mid G_i = 0]} \quad \text{if} \quad \mathbb{E}[S_{i2} \mid G_i = 0] \leq \mathbb{E}[S_{i1} \mid G_i = 0]$$

$$\pi_1 = \frac{\mathbb{E}[S_{i2}(0) \mid G_i = 1]}{\mathbb{E}[S_{i2} \mid G_i = 1]} = \frac{1 - (1 - \mathbb{E}[S_{i1} \mid G_i = 1]) \frac{1-\mathbb{E}[S_{i2}|G_i=0]}{1-\mathbb{E}[S_{i1}|G_i=0]}}{\mathbb{E}[S_{i2} \mid G_i = 1]} \quad \text{if} \quad \mathbb{E}[S_{i2} \mid G_i = 0] > \mathbb{E}[S_{i1} \mid G_i = 0]$$

If Negative Monotonicity holds:

$$\pi_0 = \frac{\mathbb{E}[S_{i2}(1) \mid G_i = 0]}{\mathbb{E}[S_{i2} \mid G_i = 0]} = \frac{\mathbb{E}[S_{i1} \mid G_i = 0]}{\mathbb{E}[S_{i2} \mid G_i = 0]} \frac{\mathbb{E}[S_{i2} \mid G_i = 1]}{\mathbb{E}[S_{i1} \mid G_i = 1]} \quad \text{if} \quad \mathbb{E}[S_{i2} \mid G_i = 1] \leq \mathbb{E}[S_{i1} \mid G_i = 1]$$

$$\pi_0 = \frac{\mathbb{E}[S_{i2}(1) \mid G_i = 0]}{\mathbb{E}[S_{i2} \mid G_i = 0]} = \frac{1 - (1 - \mathbb{E}[S_{i1} \mid G_i = 0]) \frac{1-\mathbb{E}[S_{i2}|G_i=1]}{1-\mathbb{E}[S_{i1}|G_i=1]}}{\mathbb{E}[S_{i2} \mid G_i = 0]} \quad \text{if} \quad \mathbb{E}[S_{i2} \mid G_i = 1] > \mathbb{E}[S_{i1} \mid G_i = 1]$$

$$\pi_1 = 1$$

Proof: The proof is analogous to the proof of Proposition 2 (Appendix B.3), complementing Lemma 1 with the full results in Theorem 4.2 in Athey and Imbens (2006) when $\mathbb{E}[S_{i2} \mid G_i = g] \geq \mathbb{E}[S_{i1} \mid G_i = g]$.

Notice that these proportions are not the ones used to partially identify $\mathbb{E}[Y_{i2}(1) - Y_{i2}(0) \mid G_i = 1, V_i = AO, S_{i1} = 1]$. In that case, as indicated in footnote 13, we would need to condition on $S_{i1} = 1$ in Proposition 2, that is,

Remark 8 *Under Assumptions 1, 3, 5, and 6, the proportion of Always-Observed units that were also observed in the pre-treatment period in the treatment group* ($\tilde{\pi}_1$) *and control group* ($\tilde{\pi}_0$) *are identified as follows:*

If Positive Monotonicity holds:

$$\tilde{\pi}_0 = 1$$

$$\tilde{\pi}_1 = \frac{\mathbb{E}[S_{i2}(0) \mid G_i = 1, S_{i1} = 1]}{\mathbb{E}[S_{i2} \mid G_i = 1, S_{i1} = 1]} = \frac{\mathbb{E}[S_{i2} \mid G_i = 0, S_{i1} = 1]}{\mathbb{E}[S_{i2} \mid G_i = 1, S_{i1} = 1]}$$

If Negative Monotonicity holds:

$$\tilde{\pi}_0 = \frac{\mathbb{E}[S_{i2}(1) \mid G_i = 0, S_{i1} = 1]}{\mathbb{E}[S_{i2} \mid G_i = 0, S_{i1} = 1]} = \frac{\mathbb{E}[S_{i2} \mid G_i = 1, S_{i1} = 1]}{\mathbb{E}[S_{i2} \mid G_i = 0, S_{i1} = 1]}$$

$$\tilde{\pi}_1 = 1$$

Remark 9 *Under Assumption 2,* $\mathbb{E}[S_{i2} \mid G_i = g] \leq \mathbb{E}[S_{i1} \mid G_i = g]$ *and* $\pi_g = \tilde{\pi}_g \ \forall g$.

The second term in equation (A12) can be decomposed as follows using the Law of Total Expectation:

$$\begin{aligned}\mathbb{E}[Y_2(0) \mid G_i = 1, V_i = AO] =& \mathbb{E}[Y_2(0) \mid G_i = 1, V_i = AO, S_{i1} = 1] Pr(S_{i1} = 1 \mid G_i = 1, V_i = AO) + \\ & \mathbb{E}[Y_2(0) \mid G_i = 1, V_i = AO, S_{i1} = 0] Pr(S_{i1} = 0 \mid G_i = 1, V_i = AO),\end{aligned} \tag{A13}$$

where the partial identification of $\mathbb{E}[Y_2(0) \mid G_i = 1, V_i = AO, S_{i1} = 1]$ is discussed in Sections 3 and 4, $Pr(S_{i1} = 1 \mid G_i = 1, V_i = AO)$ is identified from Baye's rule[14] by

$$Pr(S_{i1} = 1 \mid G_i = 1, V_i = AO) = \frac{\tilde{\pi}_1 \mathbb{E}[S_{i1} \mid G_i = 1, S_{i2} = 1]}{\pi_1},$$

and $E[Y_2(0) \mid G_i = 1, V_i = AO, S_{i1} = 0]$ can be bounded using the infimum and supremum of the distribution of Y (Manski, 1990), provided once again that Y has compact support.

All the discussion in this section directly applies to the Estimand 2 ($\text{QTT}_{\text{AO}}(q)$) as well.

A.3 Covariates

Up to this point, this paper abstracts from the role of covariates. However, they have substantial potential to strengthen the identification strategy (Grilli & Mealli, 2008; Lee, 2009; Mealli & Pacini, 2013; Long & Hudgens, 2013; Semenova, 2025). This section incorporates covariates into the analysis presented in the main text.

[14]

$$\tilde{\pi}_1 = Pr(S_2(0) = 1 \mid G_i = 1, S_2(1) = 1, S_1 = 1) = Pr(V_i = AO \mid G_i = 1, S_{i2} = 1, S_{i1} = 1)$$
$$= \frac{Pr(S_{i1} = 1 \mid V_i = AO, G_i = 1, S_2 = 1) Pr(V_i = AO \mid G_i = 1, S_{i2} = 1)}{Pr(S_{i1} = 1 \mid G_i = 1, S_{i2} = 1)} = \frac{Pr(S_{i1} = 1 \mid V_i = AO, G_i = 1, S_2 = 1)\pi_1}{Pr(S_{i1} = 1 \mid G_i = 1, S_{i2} = 1)}$$

Introducing covariates can improve inference in two different ways. First, conditioning on covariates can make identifying assumptions more credible by allowing them to hold locally rather than globally. For example, the monotonicity assumption can be restrictive, as it requires treatment to affect selection in the same direction for all units. Conditioning monotonicity on covariates relaxes this restriction by allowing treatment to affect selection in one direction for some covariate cells and in the opposite direction for others.

Second, covariates can be used to tighten the bounds, making them more informative. As proposed by Grilli and Mealli (2008), researchers can partition the covariate space, X, and estimate the Conditional ATT_{AO} in each cell. Then, the unconditional ATT_{AO} could be recovered by aggregating the conditional ATT_{AO}, marginalizing over the distribution of X. Lee (2009) and Mealli and Pacini (2013) show that the resulting bounds will never be wider than the unconditional ones. Long and Hudgens (2013) characterize the conditions under which they will be strictly narrower. Whether these results extend to settings without treatment unconfoundedness remains an open question for future research.

To be explicit, I next introduce the conditional analogs of estimands, assumptions, and identification results for the identification of the ATT_{AO} in the DiD framework. Identical reasoning applies to the CiC framework and the $\text{QTT}_{\text{AO}}(q)$ estimand. Let $X \subseteq \mathbb{R}^d$ be the covariate space, and $X_i \in X$ be the vector of covariates of unit i. Let $\{x^{(k)}\}_{k=1}^{K}$ be a partition of the covariate space, so that $X_i \in x \implies X_i \notin \{x^{(k)}\}_{k=1}^{K} \backslash \{x\}$.

Estimand 3 *Conditional Average Treatment Effect on the Treated ($CATT_{AO}(x)$)*

$$\mathbb{E}[Y_{i2}(1) - Y_{i2}(0) \mid G_i = 1, V_i = AO, X_i = x]$$

Assumption 11 *Conditional Principal Parallel Trends (CPPT)*

$$\mathbb{E}[Y_{i2}(0) - Y_{i1} \mid G_i = 1, V_i = AO, X_i = x] = \mathbb{E}[Y_{i2}(0) - Y_{i1} \mid G_i = 0, V_i = AO, X_i = x]$$

Assumption 12 *Conditional Monotonicity*

$$S_{i2}(1) \geq S_{i2}(0) \quad \forall i : X_i = x \quad \textit{Positive Monotonicity}$$

$$\textit{or}$$

$$S_{i2}(1) \leq S_{i2}(0) \quad \forall i : X_i = x \quad \textit{Negative Monotonicity}$$

Assumption 13 *Conditional Selection Model*

Under Positive Monotonicity:

$$(S_{it}(0) \mid X_i = x) = h_x^0(U_{it}^x, t).$$

Under Negative Monotonicity:

$$(S_{it}(1) \mid X_i = x) = h_x^1(U_{it}^x, t),$$

where U_{it}^x is an unobservable scalar for unit i at time t and $h_x^w(u,t)$ is non decreasing function in u $\forall$ $t \in \{1,2\}$, $w = \{0,1\}$.

The unobservable U^x is continuously distributed and has the same compact support in both groups. Its distribution is constant over time within groups

$$(U_{i1}^x \mid G_i, X_i = x) \sim (U_{i2}^x \mid G_i, X_i = x).$$

Additionally, given the realized selection outcome, the distribution of U is independent of the group in a given time period.

$$U_{it}^x \perp\!\!\!\perp G_i \mid S_{it}, X_i = x$$

Remark 10 *Let $Y_{it}(1)$ and $Y_{it}(0)$ be continuous. If Assumptions 1, 2, 3, 11, 12 and 13 hold, then $\Delta^{LB}(x)$ and $\Delta^{UB}(x)$ are lower and upper bounds for the Conditional Average Treatment Effect on the Treated Always-Observed units ($ATT_{AO}(x)$), where*

$$\Delta^{LB}(x) = \mathbb{E}[\ddot{Y}_i \mid G_i = 1, S_{i2} = 1, \ddot{Y}_i \leq \ddot{y}^1_{\pi_1}, X_i = x] - \mathbb{E}[\ddot{Y}_i \mid G_i = 0, S_{i2} = 1, \ddot{Y}_i \geq \ddot{y}^0_{1-\pi_0}, X_i = x]$$

$$\Delta^{UB}(x) = \mathbb{E}[\ddot{Y}_i \mid G_i = 1, S_{i2} = 1, \ddot{Y}_i \geq \ddot{y}^1_{1-\pi_1}, X_i = x] - \mathbb{E}[\ddot{Y}_i \mid G_i = 0, S_{i2} = 1, \ddot{Y}_i \leq \ddot{y}^0_{\pi_0}, X_i = x]$$

$$\ddot{Y}_i = Y_{i2} - Y_{i1}$$

$$\ddot{y}^g_q = \inf\{\ddot{y} : F(\ddot{y}) \geq q\}, \textit{ with } F \textit{ the c.d.f. of } \ddot{Y} \textit{ conditional on } S_2 = 1, X_i = x \textit{ and } G = g$$

If Conditional Positive Monotonicity holds:

$$\pi_0 = 1$$
$$\pi_1 = \frac{\mathbb{E}[S_{i1} \mid G_i = 1, X_i = x]}{\mathbb{E}[S_{i2} \mid G_i = 1, X_i = x]} \frac{\mathbb{E}[S_{i2} \mid G_i = 0, X_i = x]}{\mathbb{E}[S_{i1} \mid G_i = 0, X_i = x]}$$

If Conditional Negative Monotonicity holds:

$$\pi_0 = \frac{\mathbb{E}[S_{i1} \mid G_i = 0, X_i = x]}{\mathbb{E}[S_{i2} \mid G_i = 0, X_i = x]} \frac{\mathbb{E}[S_{i2} \mid G_i = 1, X_i = x]}{\mathbb{E}[S_{i1} \mid G_i = 1, X_i = x]}$$
$$\pi_1 = 1$$

The proof follows verbatim by conditioning throughout and using Propositions 2 and 5.

The adjusted $\overline{\text{ATT}_{\text{AO}}}$ is equal to the weighted average of $C\text{ATT}_{\text{AO}}(x)$, marginalizing over the distribution of X for the treated Always-Observed units:

$$\overline{\text{ATT}_{\text{AO}}} \in \left[\overline{\Delta^{LB}}, \overline{\Delta^{UB}}\right],$$

where

$$\overline{\Delta^{LB}} = \sum_{x \in X} \Delta^{LB}(x) p(x),$$
$$\overline{\Delta^{UB}} = \sum_{x \in X} \Delta^{UB}(x) p(x),$$
$$p(x) = Pr(X_i = x \mid G_i = 1, V_i = AO)$$

Conjecture 1 *The adjusted $\overline{ATT_{AO}}$ will always lie inside the unadjusted ATT_{AO}*

$$\left[\overline{\Delta^{LB}}, \overline{\Delta^{UB}}\right] \subseteq \left[\Delta^{LB}, \Delta^{UB}\right]$$

This statement is proved in Long and Hudgens (2013, Proposition 1) under the assumption of treatment unconfoundedness.

When it comes to estimation, each of the $\text{CATT}_{\text{AO}}(x)$ can be estimated as described in Section 4 in each of the cells that partition the covariate space. These estimates can be

aggregated as expressed in (A14), where each of the $\text{CATT}_{\text{AO}}(x)$ is weighted by the number of treated units and the estimated proportion of AO units in this cell.

$$\begin{aligned}
\widehat{\Delta^{LB}} &= \sum_{x \in X} \omega_x \widehat{\Delta^{LB}}(x) \\
\widehat{\Delta^{UB}} &= \sum_{x \in X} \omega_x \widehat{\Delta^{UB}}(x) \\
\omega_x &= \frac{\widehat{\pi}_1^x}{\Pi_1} \frac{\sum_i \mathbb{1}(G_i = 1, S_{i2} = 1, X_i = x)}{N_1} \\
\Pi_1 &= \sum_{x \in X} \widehat{\pi}_1^x
\end{aligned} \tag{A14}$$

The discretized, cell-based procedure just described can help make the identification strategy more credible and informative. This procedure goes in the direction of Semenova (2025) and could be generalized and improved by adapting the Generalized Lee Bounds in Semenova (2025) to the panel data setting considered in this paper.

A.4 Repeated Cross-Sections

Throughout the paper, I have conditioned on both S_{i1} and S_{i2} and, for the DiD case, used $\ddot{Y}_i = Y_{i2} - Y_{i1}$. When panel data are unavailable, and only repeated cross-sections are observed, these quantities cannot be constructed. This section extends the results to the repeated cross-section setting.

Lemma A1 *Let $Y_{it}(1)$ and $Y_{it}(0)$ be continuous with compact support. Furthermore, let the support of $Y_{it}(0)$ for the treatment group be contained in the support of $Y_{it}(0)$ for the control group. Then, if Assumptions 1, 2, 3 and 4 hold, then*

$$QTT_{AO}(q) = Q_{Y_2|G=1,V=AO}(q) - Q_{Y_2|G=0,V=AO}\left(F_{Y_1|G=0,V=AO}\left(Q_{Y_1|G=1,V=AO}(q)\right)\right)$$

Proof: See equation (B17).

Lemma A2 *Let $Y_{it}(1)$ and $Y_{it}(0)$ be continuous. If Assumptions 1, 2, 3 and 10 hold, then*

$$\begin{aligned} ATT_{AO} =& \mathbb{E}[Y_{i2} \mid G_i = 1, V_i = AO] - \mathbb{E}[Y_{i1} \mid G_i = 1, V_i = AO] \\ &+ \mathbb{E}[Y_{i2} \mid G_i = 0, V_i = AO] - \mathbb{E}[Y_{i1} \mid G_i = 0, V_i = AO] \end{aligned}$$

Proof: See equations (A1) and (A2).

Lemmas A1 and A2 identify the two causal estimands as functions of the expectations or c.d.f.s of observed outcomes, conditional on a latent stratum membership. Proposition 2 could be directly applied to the repeated cross-section case to identify the proportion of AO units in each group at $t = 2$, that is, $Pr(V_i = AO \mid G = g, S_2 = 1)$. These proportions can be used to partially identify the distribution of $Y_2 \mid G = g, S_2 = 1$ and its expectation.

In the panel setting, these proportions can also be used to partially identify the distribution of $Y_{i1} \mid G = g, V_i = AO$ and its expectation, exploiting the observed distribution of $Y_{i1} \mid G = g, V_i = AO, S_{i2} = 1$. However, in the repeated cross-sections setting, this is not feasible, and identification must instead rely on the distribution of $Y_{i1} \mid G = g, S_1 = 1$.

Proposition 8 *Let $Y_{it}(1)$ and $Y_{it}(0)$ be continuous with compact support. Furthermore, let the support of $Y_{it}(0)$ for the treatment group be contained in the support of $Y_{it}(0)$ for the control group. Then, if Assumptions 1, 2, 3 and 4 hold, then $\Lambda^{LB}(q)$ and $\Lambda^{UB}(q)$ are lower and upper bounds for the Quantile Treatment Effect on the Treated Always-Observed units*

($QTT_{AO}(q)$), where:

$$\Lambda^{LB}(q) = Q_{Y_2|G=1,S_2=1}(q\pi_1) - Q_{Y_2|G=0,S_2=1}\left(F_{Y_1|G=0,S_1=1}\left(Q_{Y_1|G=1,S_1=1}(q\varpi_1 + 1 - \varpi_1)\right)\frac{\pi_0}{\varpi_0} + 1 - \pi_0\right),$$

$$\Lambda^{UB}(q) = Q_{Y_2|G=1,S_2=1}(q\pi_1 + 1 - \pi_1) - Q_{Y_2|G=0,S_2=1}\left(\left(F_{Y_1|G=0,S_1=1}\left(Q_{Y_1|G=1,S_1=1}(q\varpi_1)\right) - (1 - \varpi_0)\right)\frac{\pi_0}{\varpi_0}\right),$$

$$\pi_1 := Pr(V_i = AO \mid G_i = 1, S_{i2} = 1)$$

$$\pi_0 := Pr(V_i = AO \mid G_i = 0, S_{i2} = 1)$$

$$\varpi_1 := Pr(V_i = AO \mid G_i = 1, S_{i1} = 1)$$

$$\varpi_0 := Pr(V_i = AO \mid G_i = 0, S_{i1} = 1)$$

$$F_Y(y) := Pr(Y \leq y)$$

$$Q_Y(q) := \inf\{y : F_Y(y) \geq q\}$$

provided that

$$F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q\varpi_1 + 1 - \varpi_1)\right) \leq \varpi_0$$

$$F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=0}(q\varpi_1)\right) \geq 1 - \varpi_0.$$

If $F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q\varpi_1 + 1 - \varpi_1)\right) > \varpi_0$, *then*

$$\Lambda^{LB}(q) = Q_{Y_2|G=1,S_2=1}(q\pi_1) - Q_{Y_2|G=0,S_2=1}(1).$$

If $F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=0}(q\varpi_1)\right) < 1 - \varpi_0$, *then*

$$\Lambda^{UB}(q) = Q_{Y_2|G=1,S_2=1}(q\pi_1 + 1 - \pi_1) - Q_{Y_2|G=0,S_2=1}(0)$$

Proposition 9 *Let $Y_{it}(1)$ and $Y_{it}(0)$ be continuous. If Assumptions 1, 2, 3 and 10 hold, then Δ^{LB} and Δ^{UB} are lower and upper bounds for the Average Treatment Effect on the*

Treated Always-Observed units (ATT_{AO}), where

$$\begin{aligned}
\Delta^{LB} &= \mathbb{E}[Y_{i2} \mid G_i = 1, Y_{i2} \leq y^1_{2,\pi_1}] - \mathbb{E}[Y_{i1} \mid G_i = 1, Y_{i1} \geq y^1_{1,1-\varpi_1}] \\
&\quad + \mathbb{E}[Y_{i2} \mid G_i = 0, Y_{i2} \leq y^0_{2,\pi_0}] - \mathbb{E}[Y_{i1} \mid G_i = 0, Y_{i1} \geq y^1_{1,1-\varpi_0}] \\
\Delta^{UB} &= \mathbb{E}[Y_{i2} \mid G_i = 1, Y_{i2} \geq y^1_{2,1-\pi_1}] - \mathbb{E}[Y_{i1} \mid G_i = 1, Y_{i1} \leq y^1_{1,\varpi_1}] \\
&\quad + \mathbb{E}[Y_{i2} \mid G_i = 0, Y_{i2} \geq y^0_{2,1-\pi_0}] - \mathbb{E}[Y_{i1} \mid G_i = 0, Y_{i1} \leq y^1_{1,\varpi_0}]
\end{aligned}$$

$\ddot{y}^g_{t,q} := \inf\{\ddot{y} : F_{Y_t|G=g,S_t=1}(\ddot{y}) \geq q\}$, *with* $F_{Y_t|G=g,S_t=1}$ *the c.d.f. of* Y_t *conditional on* $S_t = 1$ *and* $G = g$

$\pi_1 := Pr(V_i = AO \mid G_i = 1, S_{i2} = 1)$

$\pi_0 := Pr(V_i = AO \mid G_i = 0, S_{i2} = 1)$

$\varpi_1 := Pr(V_i = AO \mid G_i = 1, S_{i1} = 1)$

$\varpi_0 := Pr(V_i = AO \mid G_i = 0, S_{i1} = 1)$.

The proof of Propositions 8 and 9 are similar to the proofs of Propositions 1 and 5 respectively and are omitted.

The remaining challenge is the identification of ϖ_0 and ϖ_1, used to partially identify the distribution of $Y_1 \mid G = g, V_i = AO$ from the distribution of $Y_1 \mid G_i = g, S_1 = 1$.

Proposition 10 *Under Assumptions 1, 2, and 3, the proportions of AO units in the pre-treatment period are given by*

$$\varpi_1 = \pi_1 \frac{\mathbb{E}[S_{i2} \mid G_i = 1]}{\mathbb{E}[S_{i1} \mid G_i = 1]}$$
$$\varpi_0 = \pi_0 \frac{\mathbb{E}[S_{i2} \mid G_i = 0]}{\mathbb{E}[S_{i1} \mid G_i = 0]}$$

Proof: See Appendix B.10.

Finally, notice that without panel data, Assumption 2 cannot be tested since the pair (S_{i1}, S_{i2}) is never jointly observed for any unit. Nevertheless, if $\mathbb{E}[S_{i1} \mid G_i = g] \geq \mathbb{E}[S_{i2} \mid G_i = g]$, this could be interpreted as evidence against this assumption. When Assumption 2 does not hold, π_1 and π_0 can still be identified as described in Proposition 7. On the other hand, ϖ_1 and ϖ_0 can only be partially identified as follows:

Proposition 11 *Under Assumptions 1 and 3, the proportions of AO units in the pre-treatment period are bounded by*

$$\max\left\{0, \frac{\pi_1 \mathbb{E}[S_{i2} \mid G_i = 1] - (1 - \mathbb{E}[S_{i1} \mid G_i = 1])}{\mathbb{E}[S_{i1} \mid G_i = 1]}\right\} \leq \varpi_1 \leq \min\left\{1, \pi_1 \frac{\mathbb{E}[S_{i2} \mid G_i = 1]}{\mathbb{E}[S_{i1} \mid G_i = 1]}\right\}$$
$$max\left\{0, \frac{\pi_0 \mathbb{E}[S_{i2} \mid G_i = 0] - (1 - \mathbb{E}[S_{i1} \mid G_i = 0])}{\mathbb{E}[S_{i1} \mid G_i = 0]}\right\} \leq \varpi_0 \leq \min\left\{1, \pi_0 \frac{\mathbb{E}[S_{i2} \mid G_i = 0]}{\mathbb{E}[S_{i1} \mid G_i = 0]}\right\}$$

Proof: See Appendix B.11.

Intuitively, once the proportion of AO units in the observed post-treatment period is identified (π_g), then the total proportion of AO units in the group g is identified as well. Because V_i is defined using the post-treatment potential selection outcomes, under Assumption 3, the proportion of units with $V_i = AO$ is the same in the pre and post-treatment periods. Under Assumption 2, all the units such that $V_i = AO$ would also be observed in the pre-treatment period and hence $Pr(V_i = AO \mid G_i = g, S_1 = 1)$ is identified. When Assumption 2 does not hold, it could be that some of the missing units in the pre-treatment period have $V_i = AO$. In the worst-case scenario, all the missing units in the pre-treatment period belong to the AO stratum. In the best-case scenario, all the units with $V_i = AO$ are observed at $t = 1$. These extreme cases yield bounds on $Pr(V_i = AO \mid G_i = g, S_{i1} = 1)$ as described in Proposition 11 and Appendix B.11. These bounds can then be used in Propositions 8 and 9.

Example 2 *Consider the case where 90% of the units are observed in the pre-treatment period in group $G = g$, 80% of the units are observed in the post-treatment period in the same group, out of which 90% are AO. That is, $\mathbb{E}[S_{i1} \mid G_i = g] = 0.9$, $\mathbb{E}[S_{i2} \mid G_i = g] = 0.8$ and $Pr(V_i = AO \mid G_i = g, S_{i2} = 1) = 0.9$.*

90% of the observed units in the post-treatment period are AO, meaning that $0.9 \times 0.8 = 0.72$ of the total units in group g belong to this stratum. The remaining $0.1 \times 0.8 = 0.08$ belongs to the OT stratum, and the 0.2 not observed belongs to the OC and NO strata.

This means that 72% of the units belong to the AO stratum. If $V_i = AO \implies S_{i1} = 1$, it follows that $0.72/0.9 = 0.8$ units observed in the pre-treatment period would belong to the AO stratum in the post-treatment period. On the other hand, if all the missing units in the pre-treatment period (10%) belong to the AO stratum, then it follows that there are still 62%

of units who have $S_{i1} = 1$ *and* $V_i = AO$*, which is* $0.62/0.9 \approx 0.68$ *of the observed treated group at* $t = 1$*.*

A.5 Discrete outcomes

So far, I have assumed that the outcome Y is continuously distributed over a compact support. When the outcome is discrete, it is still possible to define the quantile function $Q_Y(q) = \{\inf y : F(y) \geq q\}$. However, because discrete data typically exhibit ties, this function is no longer strictly increasing in q; instead, it becomes a step function. As a result, the identification strategy developed for continuous outcomes is no longer directly applicable. In the DiD case, the trimming procedure developed in Proposition 5 relies on the identification of the trimming values, $\ddot{y}_q$. However, with discrete outcomes (or continuous outcomes with ties), it is possible that $F(\ddot{y}_q) \neq q$. In the CiC case, the strict monotonicity of the function $m(\cdot)$ in Assumption 4 has too strong consequences. Both points are illustrated in Example 3.

Example 3 *Assume that* $Y_i \sim Bernoulli(0.7)$ *and that* $\pi = 0.9$*. That is, I have a binary outcome, and I know that 90% of the observed units belong to the AO stratum. In this case,* $y_{0.9} = Q(0.9) = 1$*, and the trimming procedure described in Proposition 5 would not trim any unit, with* $\mathbb{E}[Y_i \mid Y_i \leq y_{0.9}] = 0.7$*. If, instead, one were to apply a strict inequality, the procedure would trim 30% of the units, with* $\mathbb{E}[Y_i \mid y_{0.9} < 1] = 0$*. It is clear that neither approach delivers a satisfactory trimming rule in this setting.*

Consider now the CiC outcome model in Assumption 4. According to this model, $\mathcal{U}_i > \mathcal{U}_j \iff Y_i > Y_j$*. With a binary outcome, this implies that the unobservable* $\mathcal{U}_i$ *can take only two values:* $\mathcal{U}_0$ *for all the units with* $Y_i = 0$ *and* $\mathcal{U}_1$ *for all the units with* $Y_i = 1$*. Furthermore, since this distribution is constant over time, it implies that the distribution of* Y *is also constant, which is restrictive.*

A.5.1 Identification in the CiC setup with Discrete outcomes

For the CiC case, I extend the methodoloy developed in this paper to the Discrete Changes-in-Changes (DCIC) design Athey and Imbens (2006, Section 4).

Assumption 14 *Outcome model*

$$(Y_{it}(0) \mid V_i = AO) \quad = \quad m(\mathcal{U}_{it}, t),$$

where $m(u,t)$ *is non-decreasing in* u $\forall t$*, and* $\mathcal{U}_{it}$ *is an unobservable scalar for unit* i *at time* t *with continuous and constant distribution over time within groups,*

$$\mathcal{U}_{i1} \mid G_i, V_i = AO \sim \mathcal{U}_{i2} \mid G_i, V_i = AO.$$

In Assumption 4, the function $m(\mathcal{U}_{it}, t)$ was assumed to be strictly increasing in $\mathcal{U}_{it}$. In Assumption 14, this function is assumed to be non-decreasing. This assumption is weaker and eliminates the strong implications of the CiC model with a discrete outcome highlighted earlier. The downside is that the distribution of the missing potential outcome for the treated group can only be partially identified (Athey & Imbens, 2006). Furthermore, the unobservable $\mathcal{U}_{it}$ is assumed to be continuous.

Proposition 12 *Let* $Y_{it}(1)$ *and* $Y_{it}(0)$ *be discrete with compact support. Furthermore, let the support of* $Y_{it}(0)$ *for the treatment group be contained in the support of* $Y_{it}(0)$ *for the control group. Then, if Assumptions 1, 2, 3 and 14 hold, then* $\Lambda^{LB}(q)$ *and* $\Lambda^{UB}(q)$ *are lower and upper bounds for the Quantile Treatment Effect on the Treated Always-Observed units*

($QTT_{AO}(q)$), where:

$$\Lambda^{LB}(q) = Q_{Y_2|G=1,S_2=1}(q\pi_1) - Q_{Y_2|G=0,S_2=1}\left(F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q\pi_1 + 1 - \pi_1)\right) + 1 - \pi_0\right),$$
$$\Lambda^{UB}(q) = Q_{Y_2|G=1,S_2=1}(q\pi_1 + 1 - \pi_1) - \widetilde{Q}_{Y_2|G=0,S_2=1}\left(F_{Y_1|G=0,S_2=1}\left(\widetilde{Q}_{Y_1|G=1,S=1}(q\pi_1)\right) - (1 - \pi_0)\right),$$
$$\pi_1 = Pr(S_{i2}(0) = 1 \mid G_i = 1, S_{i2}(1) = 1),$$
$$\pi_0 = Pr(S_{i2}(1) = 1 \mid G_i = 0, S_{i2}(0) = 1),$$
$$F_Y(y) := Pr(Y \leq y),$$
$$Q_Y(q) := \inf\{y : F_Y(y) \geq q\},$$
$$\widetilde{Q}_Y(q) := \sup\{y \cup \{-\infty\} : F_Y(y) \leq q\},$$

provided that

$$F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q\pi_1 + 1 - \pi_1)\right) \leq \pi_0$$
$$F_{Y_1|G=0,S_2=1}\left(\widetilde{Q}_{Y_1|G=1,S_2=0}(q\pi_1)\right) \geq 1 - \pi_0.$$

If $F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q\pi_1 + 1 - \pi_1)\right) > \pi_0$, *then*

$$\Lambda^{LB}(q) = Q_{Y_2|G=1,S_2=1}(q\pi_1) - Q_{Y_2|G=0,S_2=1}(1).$$

If $F_{Y_1|G=0,S_2=1}\left(\widetilde{Q}_{Y_1|G=1,S_2=0}(q\pi_1)\right) < 1 - \pi_0$, *then*

$$\Lambda^{UB}(q) = Q_{Y_2|G=1,S_2=1}(q\pi_1 + 1 - \pi_1) - \widetilde{Q}_{Y_2|G=0,S_2=1}(0).$$

Where I use the convention that $F_Y(-\infty) = 0$.

Proof: Similar as proof in Appendix B.1, replacing Lemma B3 by Theorem 4.1 in Athey and Imbens (2006).

Proposition 12 partially identifies the $\text{QTT}_{\text{AO}}(q)$ under Assumption 14. The key difference relative to Proposition 1 concerns the lower bound, Δ^{LB}, where the left-continuous quantile function, $Q_Y(q)$, is replaced by the right-continuous quantile function, $\widetilde{Q}_Y(q)$. This modification follows from the results in Section 4 in Athey and Imbens (2006). When the outcome is discrete, or more generally, there are mass points that generate ties in the data, the dis-

tribution $F_{Y_2(0)|G=1}(y)$ can only be partially identified. The upper bound is given using the left-continuous quantile function, Q, as it 'jumps' at the first y where $F_Y(y)$ reaches or exceeds q. On the other hand, the right-continuous quantile function, $\widetilde{Q}$, 'jumps' at the last value of y such that $F_Y(y)$ is not larger than q. When $F_Y(y)$ is strictly increasing, then $Q_Y(q) = \widetilde{Q}_Y(q)$. When $F_Y(y)$ is flat, $Q_Y(q) > \widetilde{Q}_Y(q)$. Consequently, Q is used to compute the upper bound for AO units, Δ^{LB}, while $\widetilde{Q}$ is used for the lower bound, Δ^{LB}.

A.5.2 Identification in the DiD setup with Discrete outcomes

The solution to the problem illustrated in Example 3 is to reweight units so that the weighted average of the outcome successfully captures the upper or lower bound of the latent distribution for the AO units. That is, define a new p.m.f. that is equivalent to the original p.m.f. after trimming $\pi \times 100$ percent of the units. To that end, define:

$$\underline{\mathcal{P}}_Y^{\pi}(y) = \begin{cases} \frac{Pr(Y=y)}{\pi} & \text{if} \quad y < Q_Y(\pi) \\ \frac{\pi - Pr(y<Q_Y(\pi))}{\pi} & \text{if} \quad y = Q_Y(\pi) \\ 0 & \text{if} \quad y > Q_Y(\pi) \end{cases} \tag{A15}$$

$$\overline{\mathcal{P}}_Y^{\pi}(y) = \begin{cases} 0 & \text{if} \quad y < Q_Y(1-\pi) \\ \frac{\pi - Pr(Y>Q_Y(1-\pi))}{\pi} & \text{if} \quad y = Q_Y(1-\pi) \\ \frac{Pr(Y=y)}{\pi} & \text{if} \quad y > Q_Y(1-\pi) \end{cases} \tag{A16}$$

Equations (A15) and (A16) describe a trimming function for discrete outcomes given a trimming proportion π. In equation (A15), the trimming threshold is given by $Q_Y(\pi)$. All the outcomes strictly above this threshold receive zero weight, and all the units strictly below receive a weight equal to their probability. Because Y is discrete, there could be a mass at $Y = Q_y(\pi)$, implying that $Pr(Y < Q_Y(\pi)) < \pi < Pr(Y \leq Q_Y(\pi))$. If that is the case, the units with $Y = Q_Y(\pi)$ are reweighted, using $\pi - Pr(Y < Q_Y(\pi))$ as weight instead of $Pr(Y = Q(\pi)) = Pr(Y \leq Q_Y(\pi)) - Pr(Y > Q_Y(\pi))$. Finally, the weights are normalized by π, ensuring that $\underline{\mathcal{P}}_Y^{\pi}(y)$ is a proper probability mass function. Exactly the same intuition is behind equation (A16).

Proposition 13 *Let $Y_{it}(1)$ and $Y_{it}(0)$ be discrete. If Assumptions 1, 2, 3 and 10 hold, then Δ^{LB} and Δ^{UB} are lower and upper bounds for the Average Treatment Effect on the Treated Always-Observed units (ATT_{AO}), where*

$$\Delta^{LB} = \sum_{i:G_i=1,S_{i2}=1} \ddot{Y}_i \underline{\mathcal{P}}^{\pi_1}_{\ddot{Y}_i|G_i=1,S_{i2}=1}(y) - \sum_{i:G_i=0,S_{i2}=1} \ddot{Y}_i \overline{\mathcal{P}}^{\pi_0}_{\ddot{Y}_i|G_i=0,S_{i2}=1}(y)$$

$$\Delta^{UB} = \sum_{i:G_i=1,S_{i2}=1} \ddot{Y}_i \overline{\mathcal{P}}^{\pi_1}_{\ddot{Y}_i|G_i=1,S_{i2}=1}(y) - \sum_{i:G_i=0,S_{i2}=1} \ddot{Y}_i \underline{\mathcal{P}}^{\pi_0}_{\ddot{Y}_i|G_i=0,S_{i2}=1}(y)$$

$$\pi_1 = Pr(S_{i2}(0) = 1 \mid G_i = 1, S_{i2}(1) = 1)$$

$$\pi_0 = Pr(S_{i2}(1) = 1 \mid G_i = 0, S_{i2}(0) = 1).$$

B Proofs

B.1 Proof of Proposition 1

The structure of the proof is as follows. First, I draw on results from Horowitz and Manski (1995) to partially identify the outcome distribution for the Always-Observed units in all the periods and groups. Next, I use results from Athey and Imbens (2006) to express the Causal Estimand 2, QTT_{AO}, as a function of distributions of observed outcomes. Finally, I combine these results to partially identify the $\text{QTT}_{\text{AO}}(q)$.

Proof.

Lemma B1 *Let Y be a continuous random variable and a mixture of two random variables, Y_0 and Y_1, with known mixing probability, p. The c.d.f. of the observed Y is $F(y) = (1-p)F_0(y) + pF_1(y)$, where F_0 and F_1 are the c.d.f.s of Y_0 and Y_1 respectively. Then, the c.d.f. of Y_1 is bounded as follows:*

$$F_1(y) \in \left[\max\left\{0, \frac{F(y)-(1-p)}{p}\right\}, \min\left\{\frac{F(y)}{p}, 1\right\}\right]$$

These bounds are sharp.

Proof: See Horowitz and Manski (1995, Proposition 2).

Consider the observed outcome for the treatment group in the post-treatment period. Denote its c.d.f. by $F_{Y_2|G=1,S_2=1}(y)$. This distribution is a mixture of the outcome distribution for AO and OT units with mixing probability π_1. Accordingly, I can write

$$F_{Y_2|G=1,S_2=1}(y) = \pi_1 F_{Y_2|G=1,V=AO}(y) + (1-\pi_1) F_{Y_2|G=1,V=OT}(y). \tag{B1}$$

From Lemma B1, it follows that

$$F_{Y_2|G=1,V=AO}(y) \in \left[\max\left\{0, \frac{F_{Y_2|G=1,S_2=1}(y) - (1-\pi_1)}{\pi_1}\right\}, \min\left\{\frac{F_{Y_2|G=1,S_2=1}(y)}{\pi_1}, 1\right\}\right]. \tag{B2}$$

Consider now the observed outcome for the treatment group in the pre-treatment period, restricted to units also observed in the post-treatment period. Denote its c.d.f. by $F_{Y_1|G=1,S_2=1}(y)$. Because I am conditioning on being observed in the post-treatment period, this distribution is also a mixture of the outcome distribution for AO and OT units with mixing probability π_1. Therefore, I can write

$$F_{Y_1|G=1,S_2=1}(y) = \pi_1 F_{Y_1|G=1,V=AO}(y) + (1-\pi_1) F_{Y_1|G=1,V=OT}(y). \tag{B3}$$

From Lemma B1, it follows that

$$F_{Y_1|G=1,V=AO}(y) \in \left[\max\left\{0, \frac{F_{Y_1|G=1,S_2=1}(y) - (1-\pi_1)}{\pi_1}\right\}, \min\left\{\frac{F_{Y_1|G=1,S_2=}(y)}{\pi_1}, 1\right\}\right]. \tag{B4}$$

Symmetric arguments can be used to partially identify the outcome distribution for AO units in the control group. Table B1 summarizes the resulting bounds for the outcome distribution of AO units across all periods and groups.

Table B1: Bounds for the outcome distribution for Always-Observed units.

Distribution	Lower Bound	Upper Bound
$F_{Y_1\|G=1,V=AO}(y)$	$\max\left\{0, \frac{F_{Y_1\|G=1,S_2=1}(y)-(1-\pi_1)}{\pi_1}\right\}$	$\min\left\{\frac{F_{Y_1\|G=1,S_2=}(y)}{\pi_1}, 1\right\}$
$F_{Y_2\|G=1,V=AO}(y)$	$\max\left\{0, \frac{F_{Y_2\|G=1,S_2=1}(y)-(1-\pi_1)}{\pi_1}\right\}$	$\min\left\{\frac{F_{Y_2\|G=1,S_2=1}(y)}{\pi_1}, 1\right\}$
$F_{Y_1\|G=0,V=AO}(y)$	$\max\left\{0, \frac{F_{Y_1\|G=0,S_2=1}(y)-(1-\pi_0)}{\pi_0}\right\}$	$\min\left\{\frac{F_{Y_1\|G=0,S_2=}(y)}{\pi_0}, 1\right\}$
$F_{Y_2\|G=0,V=AO}(y)$	$\max\left\{0, \frac{F_{Y_2\|G=0,S_2=1}(y)-(1-\pi_0)}{\pi_0}\right\}$	$\min\left\{\frac{F_{Y_2\|G=0,S_2=1}(y)}{\pi_0}, 1\right\}$

Notes: Sharp bounds on the outcome distribution for the Always-Observed units in all periods ans groups. $F_{Y_t|G=g,V=AO}(y)$ denotes the c.d.f. of the outcome at period t in group g for the AO principal stratum. $F_{Y_t|G=g,S_2=1}(y)$ denotes the c.d.f. of the outcome at period t in group g for units observed at $t=2$. Results follow from Lemma B1.

Lemma B2 *Let Y be a continuous random variable and a mixture of two random variables, Y_0 and Y_1, with known mixing probability, p. The c.d.f. of the observed Y is $F(y) = (1-p)F_0(y) + pF_1(y)$, where F_0 and F_1 are the c.d.f.s of Y_0 and Y_1 respectively. The q quantile of the distribution F_1 is bounded as follows:*

$$Q_1(q) \in [Q(qp), Q(qp+1-p)],$$

where $Q(\cdot)$ is the quantile function,

$$Q(q) = \inf\{y : F(y) \geq q\}.$$

These bounds are sharp.

Proof: See Horowitz and Manski (1995, Proposition 3).

As shown in equations (B1) and (B3), I can write the observed outcome in the treatment group as a mixture of the distributions for AO and OT units. Therefore, by Lemma B2, the quantile functions for the AO units are bounded as follows:

$$Q_{Y_2|G=1,V=AO}(q) \in \left[Q_{Y_2|G=1,S_2=1}(q\pi_1) \quad , \quad Q_{Y_2|G=1,S_2=1}(q\pi_1 + 1 - \pi_1)\right], \tag{B5}$$

$$Q_{Y_1|G=1,V=AO}(q) \in \left[Q_{Y_1|G=1,S_2=1}(q\pi_1) \quad , \quad Q_{Y_1|G=1,S_2=1}(q\pi_1 + 1 - \pi_1)\right]. \tag{B6}$$

Note that, because both q and π_1 lie in the unit interval, it follows that $0 \leq q\pi_1 \leq 1$ and $0 \leq q\pi_1 + 1 - \pi_1 \leq 1$.

Symmetric arguments can be used to partially identify the quantiles of the distribution for AO units in the control group. Table B2 summarizes the resulting bounds for the q quantile of the outcome distribution for AO units across all periods and groups.

Table B2: Bounds for the quantile function for Always-Observed units.

Quantile function	Lower Bound	Upper Bound
$Q_{Y_1\|G=1,V=AO}(q)$	$Q_{Y_1\|G=1,S_2=1}(q\pi_1)$	$Q_{Y_1\|G=1,S_2=1}(q\pi_1 + 1 - \pi_1)$
$Q_{Y_2\|G=1,V=AO}(q)$	$Q_{Y_2\|G=1,S_2=1}(q\pi_1)$	$Q_{Y_2\|G=1,S_2=1}(q\pi_1 + 1 - \pi_1)$
$Q_{Y_1\|G=0,V=AO}(q)$	$Q_{Y_1\|G=0,S_2=1}(q\pi_0)$	$Q_{Y_1\|G=0,S_2=1}(q\pi_0 + 1 - \pi_0)$
$Q_{Y_2\|G=0,V=AO}(q)$	$Q_{Y_2\|G=0,S_2=1}(q\pi_0)$	$Q_{Y_2\|G=0,S_2=1}(q\pi_0 + 1 - \pi_0)$

Notes: Sharp bounds on the quantiles of the outcome distribution for the Always-Observed units in all periods ans groups. $Q_{Y_t|G=g,S_2=1}(p) := \inf\{y : F_{Y_t|G=g,S_2=1} \geq p\}$, and $F_{Y_t|G=g,S_2=1}$ denotes the c.d.f. of the outcome at period t in group g for units observed in the post-treatment outcome. Results follow from Lemma B2.

Now, recall the expression for the $\text{QTT}_{\text{AO}}(q)$,

$$\text{QTT}_{\text{AO}}(q) = Q_{Y_2(1)|G=1,V=AO}(q) - Q_{Y_2(0)|G=1,V=AO}(q).$$

I can substitute the first term in the estimand by the realized quantile for AO units, $Q_{Y_2|G=1,V=AO}(q)$, which is partially identified as expressed in the second row of Table B2. The second term, $Q_{Y_2(0)|G=1,V=AO}(q)$, involves the quantile of the distribution of the missing potential outcome, and needs to be imputed.

Lemma B3 *Under Assumption 4,*

$$F_{Y_2(0)|G=1,V=AO}(y) = F_{Y_1|G=1,V=AO}\left(Q_{Y_1|G=0,V=AO}\left(F_{Y_2|G=0,V=AO}(y)\right)\right).$$

Proof: See Athey and Imbens (2006, Theorem 3.1).

Note that Athey and Imbens (2006) use different notation: the distribution $F_{Y_t|G=g}$ corresponds to $F_{Y,gt}$ in their notation, and they index time as $t \in \{0,1\}$ rather than $t \in \{1,2\}$. Furthermore, they write $Y(0)$ and $Y(1)$ as Y^N and Y^I respectively. For example, $F_{Y_1|G=1}$ is written as $F_{Y,10}$ in their notation. Finally, they write $Q(\cdot)$ as $F^{-1}(\cdot)$.

Notice that Assumption 4 is the principal counterpart to assumptions 3.1-3.4 in Athey and Imbens (2006), under which Theorem 3.1 is derived. However, as argued in Athey and Imbens (2006, p. 443), by conditioning all the assumptions on a covariate X, Theorem 3.1 can be used to identify the distribution of $Y_2(0) \mid G = 1, X$, where X in this case is the latent principal stratum, V.

Because Y is continuous and has compact support, it follows that $F(Q(q)) = q$ and $Q(F(y)) = y$ for any y in the support of Y. Then, from Lemma B3,

$$F_{Y_2(0)|G=1,V=AO}(y) = q = F_{Y_1|G=1,V=AO}\left(Q_{Y_1|G=0,V=AO}\left(F_{Y_2|G=0,V=AO}(y)\right)\right) \tag{B7}$$

$$Q_{Y_1|G=1,V=AO}(q) = Q_{Y_1|G=1,V=AO}\left(F_{Y_1|G=1,V=AO}\left(Q_{Y_1|G=0,V=AO}\left(F_{Y_2|G=0,V=AO}(y)\right)\right)\right) \tag{B8}$$

$$Q_{Y_1|G=1,V=AO}(q) = Q_{Y_1|G=0,V=AO}\left(F_{Y_2|G=0,V=AO}(y)\right) \tag{B9}$$

$$F_{Y_1|G=0,V=AO}\left(Q_{Y_1|G=1,V=AO}(q)\right) = F_{Y_1|G=0,V=AO}\left(Q_{Y_1|G=0,V=AO}\left(F_{Y_2|G=0,V=AO}(y)\right)\right) \tag{B10}$$

$$F_{Y_1|G=0,V=AO}\left(Q_{Y_1|G=1,V=AO}(q)\right) = F_{Y_2|G=0,V=AO}(y) \tag{B11}$$

$$Q_{Y_2|G=0,V=AO}\left(F_{Y_1|G=0,V=AO}\left(Q_{Y_1|G=1,V=AO}(q)\right)\right) = Q_{Y_2|G=0,V=AO}\left(F_{Y_2|G=0,V=AO}(y)\right) \tag{B12}$$

$$Q_{Y_2|G=0,V=AO}\left(F_{Y_1|G=0,V=AO}\left(Q_{Y_1|G=1,V=AO}(q)\right)\right) = y = Q_{Y_2(0)|G=1,V=AO}(q), \tag{B13}$$

where the last equality comes from

$$F_{Y_2(0)|G=1,V=AO}(y) = q \tag{B14}$$

$$Q_{Y_2(0)|G=1,V=AO}\left(F_{Y_2(0)|G=1,V=AO}(y)\right) = Q_{Y_2(0)|G=1,V=AO}(q) \tag{B15}$$

$$y = Q_{Y_2(0)|G=1,V=AO}(q). \tag{B16}$$

Hence, I can replace the potential outcomes in the Causal Estimand 2 by observed outcomes as follows:

$$\text{QTT}_{\text{AO}}(q) = Q_{Y_2|G=1,V=AO}(q) - Q_{Y_2|G=0,V=AO}\left(F_{Y_1|G=0,V=AO}\left(Q_{Y_1|G=1,V=AO}(q)\right)\right), \tag{B17}$$

which is equivalent to equation (18) in Athey and Imbens (2006) conditioning on $V = AO$. All these quantities can be partially identified as noted above in Tables B1 and B2. Therefore, the partial identification of $\text{QTT}_{\text{AO}}(q)$ follows. Specifically, the upper (lower) bound of the $\text{QTT}_{\text{AO}}(q)$ is given by the upper (lower) bound of $Q_{Y_2|G=1,V=AO}(q)$ and the lower (upper) bound of $Q_{Y_2|G=0,V=AO}\left(F_{Y_1|G=0,V=AO}\left(Q_{Y_1|G=1,V=AO}(q)\right)\right)$. The bounds of the former can be found in Table B2. I derive the bounds of the latter next.

First, I derive the lower bound of $Q_{Y_2|G=0,V=AO}\left(F_{Y_1|G=0,V=AO}\left(Q_{Y_1|G=1,V=AO}(q)\right)\right)$. From Lemma B2, as shown in Table B2, the lower bound of $Q_{Y_2|G=0,V=AO}(q)$ is given by $Q_{Y_2|G=0,S_2=1}(q\pi_0)$. So it follows that

$$\begin{aligned} &Q_{Y_2|G=0,V=AO}\left(F_{Y_1|G=0,V=AO}\left(Q_{Y_1|G=1,V=AO}(q)\right)\right) \geq \\ &Q_{Y_2|G=0,S_2=1}\left(F_{Y_1|G=0,V=AO}\left(Q_{Y_1|G=1,V=AO}(q)\right)\pi_0\right). \end{aligned} \tag{B18}$$

From Lemma B1, as shown in Table B1, the lower bound of $F_{Y_1|G=0,V=AO}(y)$ is given by

$\max\left\{0, \frac{F_{Y_1|G=0,S_2=1}(y)-(1-\pi_0)}{\pi_0}\right\}$. Then, it follows that:

$$Q_{Y_2|G=0,V=AO}\left(F_{Y_1|G=0,V=AO}\left(Q_{Y_1|G=1,V=AO}(q)\right)\right) \geq$$
$$Q_{Y_2|G=0,S_2=1}\left(\max\left\{0, \frac{F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,V=AO}(q)\right)-(1-\pi_0)}{\pi_0}\right\}\pi_0\right). \tag{B19}$$

From Lemma B2, as shown in Table B2, the lower bound of $Q_{Y_1|G=1,V=AO}(q)$ is given by $Q_{Y_1|G=1,S_2=1}(q\pi_1)$. So it follows that

$$Q_{Y_2|G=0,V=AO}\left(F_{Y_1|G=0,V=AO}\left(Q_{Y_1|G=1,V=AO}(q)\right)\right) \geq$$
$$Q_{Y_2|G=0,S_2=1}\left(\max\left\{0, \frac{F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q\pi_1)\right)-(1-\pi_0)}{\pi_0}\right\}\pi_0\right). \tag{B20}$$

Because I am assuming that Y has compact support and the support of $Y_t \mid G=1$ is contained in the support of $Y_t \mid G=0$, then $F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q)\right)$ is identified for any q. Because both q and π_g lie in the unit interval, their product also lies in the unit interval. All in all, the lower bound for $Q_{Y_2|G=0,V=AO}\left(F_{Y_1|G=0,V=AO}\left(Q_{Y_1|G=1,V=AO}(q)\right)\right)$ is given by:

$$\begin{cases} Q_{Y_2|G=0,S_2=1}\left(F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q\pi_1)\right)-(1-\pi_0)\right) & \text{if} \quad F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=0}(q\pi_1)\right) \geq 1 \\ Q_{Y_2|G=0,S_2=1}(0) & \text{otherwise} \end{cases} \tag{B21}$$

I can follow the same steps to derive the upper bound of $Q_{Y_2|G=0,V=AO}\left(F_{Y_1|G=0,V=AO}\left(Q_{Y_1|G=1,V=AO}(q)\right)\right)$:

$$Q_{Y_2|G=0,V=AO}\left(F_{Y_1|G=0,V=AO}\left(Q_{Y_1|G=1,V=AO}(q)\right)\right) \leq$$
$$Q_{Y_2|G=0,S_2=1}\left(F_{Y_1|G=0,V=AO}\left(Q_{Y_1|G=1,V=AO}(q)\right)\pi_0+1-\pi_0\right) \tag{B22}$$

$$Q_{Y_2|G=0,V=AO}\left(F_{Y_1|G=0,V=AO}\left(Q_{Y_1|G=1,V=AO}(q)\right)\right) \leq$$
$$Q_{Y_2|G=0,S_2=1}\left(\min\left\{\frac{F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,V=AO}(q)\right)}{\pi_0}, 1\right\}\pi_0+1-\pi_0\right) \tag{B23}$$

$$Q_{Y_2|G=0,V=AO}\left(F_{Y_1|G=0,V=AO}\left(Q_{Y_1|G=1,V=AO}(q)\right)\right) \leq$$
$$Q_{Y_2|G=0,S_2=1}\left(\min\left\{\frac{F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q\pi_1+1-\pi_1)\right)}{\pi_0},1\right\}\pi_0+1-\pi_0\right). \tag{B24}$$

So the upper bound of $Q_{Y_2|G=0,V=AO}\left(F_{Y_1|G=0,V=AO}\left(Q_{Y_1|G=1,V=AO}(q)\right)\right)$ is given by:

$$\begin{cases} Q_{Y_2|G=0,S_2=1}\left(F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q\pi_1+1-\pi_1)\right)+1-\pi_0\right) & \text{if } F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q\pi_1+1-\pi_1)\right)+1-\pi_0 \leq 1 \\ Q_{Y_2|G=0,S_2=1}(1) & \text{otherwise} \end{cases} \tag{B25}$$

To conclude, equation (B17) presents the $\text{QTT}_{\text{AO}}(q)$ as a function of the distributions of observed outcomes for the AO units. These distributions are partially identified. The upper bound of the $\text{QTT}_{\text{AO}}(q)$ is given by the upper bound of the first term (second row in Table B2) and the lower bound of the second term (equation (B21)). The lower bound of the $\text{QTT}_{\text{AO}}(q)$ is given by the lower bound of the first term (second row in Table B2) and the upper bound of the second term (equation (B25)). That is,

$$\Lambda^{LB}(q) = Q_{Y_2|G=1,S_2=1}(q\pi_1) - Q_{Y_2|G=0,S_2=1}\left(F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q\pi_1+1-\pi_1)\right)+1-\pi_0\right) \tag{B26}$$

if

$$F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q\pi_1+1-\pi_1)\right) \leq \pi_0$$

and

$$\Lambda^{LB}(q) = Q_{Y_2|G=1,S_2=1}(q\pi_1) - Q_{Y_2|G=0,S_2=1}(1) \tag{B27}$$

otherwise, and

$$\Lambda^{UB}(q) = Q_{Y_2|G=1,S_2=1}(q\pi_1+1-\pi_1) - Q_{Y_2|G=0,S_2=1}\left(F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q\pi_1)\right)-(1-\pi_0)\right) \tag{B28}$$

if

$$F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=0}(q\pi_1)\right) \geq 1-\pi_0$$

and

$$\Lambda^{UB}(q) = Q_{Y_2|G=1,S_2=1}(q\pi_1+1-\pi_1) - Q_{Y_2|G=0,S_2=1}(0) \tag{B29}$$

otherwise. ■

B.2 Proof of Remark 6

The treatment effect on selection, $\mathbb{E}[S_{i2}(1) - S_{i2}(0) \mid G_i = g]$ will have the same sign for both groups $g \in \{0, 1\}$.

Proof. Assume without loss of generality that the treatment effect on selection for the treated units is positive:

$$\mathbb{E}[S_{i2}(1) - S_{i2}(0) \mid G_i = 1] > 0 \tag{B30}$$

$$\mathbb{E}[S_{i2} \mid G_i = 1] - \mathbb{E}[S_{i1} \mid G_i = 1]\frac{\mathbb{E}[S_{i2} \mid G_i = 0]}{\mathbb{E}[S_{i1} \mid G_i = 0]} > 0 \tag{B31}$$

$$\mathbb{E}[S_{i2} \mid G_i = 1] > \mathbb{E}[S_{i1} \mid G_i = 1]\frac{\mathbb{E}[S_{i2} \mid G_i = 0]}{\mathbb{E}[S_{i1} \mid G_i = 0]} \tag{B32}$$

$$\mathbb{E}[S_{i1} \mid G_i = 0]\frac{\mathbb{E}[S_{i2} \mid G_i = 1]}{\mathbb{E}[S_{i1} \mid G_i = 1]} > \mathbb{E}[S_{i2} \mid G_i = 0] \tag{B33}$$

$$\mathbb{E}[S_{i1} \mid G_i = 0]\frac{\mathbb{E}[S_{i2} \mid G_i = 1]}{\mathbb{E}[S_{i1} \mid G_i = 1]} - \mathbb{E}[S_{i2} \mid G_i = 0] > 0 \tag{B34}$$

$$\mathbb{E}[S_{i2}(1) - S_{i2}(0) \mid G_i = 0] > 0 \tag{B35}$$

Then, the treatment effect on selection for the control units is also positive. ■

$\mathbb{E}[S_{i1} \mid G_i = 1] = \mathbb{E}[S_{i1} \mid G_i = 0] \implies \mathbb{E}[S_{i2}(1) - S_{i2}(0) \mid G_i = 0] = \mathbb{E}[S_{i2}(1) - S_{i2}(0) \mid G_i = 1]$.

Proof.

$$\mathbb{E}[S_{i2}(1) - S_{i2}(0) \mid G_i = 1] = \mathbb{E}[S_{i2} \mid G_i = 1] - \mathbb{E}[S_{i1} \mid G_i = 1]\frac{\mathbb{E}[S_{i2} \mid G_i = 0]}{\mathbb{E}[S_{i1} \mid G_i = 0]} \tag{B36}$$

$$= \mathbb{E}[S_{i2} \mid G_i = 1]\frac{\mathbb{E}[S_{i1} \mid G_i = 1]}{\mathbb{E}[S_{i1} \mid G_i = 1]} - \mathbb{E}[S_{i1} \mid G_i = 1]\frac{\mathbb{E}[S_{i2} \mid G_i = 0]}{\mathbb{E}[S_{i1} \mid G_i = 0]} \tag{B37}$$

$$= \mathbb{E}[S_{i2} \mid G_i = 1]\frac{\mathbb{E}[S_{i1} \mid G_i = 0]}{\mathbb{E}[S_{i1} \mid G_i = 1]} - \mathbb{E}[S_{i1} \mid G_i = 0]\frac{\mathbb{E}[S_{i2} \mid G_i = 0]}{\mathbb{E}[S_{i1} \mid G_i = 0]} \tag{B38}$$

$$= \mathbb{E}[S_{i1} \mid G_i = 0]\frac{\mathbb{E}[S_{i2} \mid G_i = 1]}{\mathbb{E}[S_{i1} \mid G_i = 1]} - \mathbb{E}[S_{i2} \mid G_i = 0] \tag{B39}$$

$$= \mathbb{E}[S_{i2}(1) - S_{i2}(0) \mid G_i = 0] \tag{B40}$$

■

It can also be shown that there are only two cases when the treatment effect on selection is the same in both groups: when the baseline selection is the same across groups, as I just showed, and when the treatment effect is zero.

B.3 Proof of Proposition 2

Below, I prove Proposition 2 under positive monotonicity. The prove under negative monotonicity follows by symmetry.

Proof.

$$\pi_0 = Pr(S_{i2}(1) = 1 \mid G_i = 0, S_{i2}(0) = 1) = 1 \tag{B41}$$

Equation (B41) follows immediately from the positive monotonicity assumption.

$$\pi_1 = Pr(S_{i2}(0) = 1 \mid G_i = 1, S_{i2}(1) = 1) \tag{B42}$$

$$= \frac{Pr(S_{i2}(1) = 1 \mid G_i = 1, S_{i2}(0) = 1) Pr(S_{i2}(0) = 1 \mid G_i = 1)}{Pr(S_{i2}(1) = 1 \mid G_i = 1)} \tag{B43}$$

$$= \frac{Pr(S_{i2}(0) = 1 \mid G_i = 1)}{Pr(S_{i2}(1) = 1 \mid G_i = 1)} \tag{B44}$$

$$= \frac{\mathbb{E}[S_{i2}(0) \mid G_i = 1]}{\mathbb{E}[S_{i2} \mid G_i = 1]} \tag{B45}$$

$$= \frac{\mathbb{E}[S_{i1} \mid G_i = 1]}{\mathbb{E}[S_{i2} \mid G_i = 1]} \frac{\mathbb{E}[S_{i2} \mid G_i = 0]}{\mathbb{E}[S_{i1} \mid G_i = 0]} \tag{B46}$$

Where (B43) follows from Bayes' theorem, (B44) follows from the positive monotonicity assumption and (B46) follows from Lemma 1. ■

B.4 Proof of Lemma 2

Proof.

$$\mathcal{AO} = \{i : S_{i2}(0) = 1, S_{i2}(1) = 1\} = \{i : \prod_j s^j_{i2}(0) = 1, \prod_j s^j_{i2}(1) = 1\} = \{i : s^j_{i2}(0) = s^j_{i2}(1) = 1 \quad \forall j\}$$

$$= \{i : V^j_i = AO \quad \forall j\} = \{i : i \in \mathcal{AO}_j \quad \forall j\} = \bigcap_j \mathcal{AO}_j \tag{B47}$$

$$\mathcal{NO} = \{i : S_{i2}(0) = 0, S_{i2}(1) = 0\} = \{i : \prod_j s^j_{i2}(0) = 0, \prod_j s^j_{i2}(1) = 0\}$$

$$= \{i : s^k_{i2}(0) = 0 \text{ for one } k \in J, s^\kappa_{i2}(1) \text{ for one } \kappa \in J\} \tag{B48}$$

$$= \{i : s^k_{i2}(0) = 0, s^k_{i2}(1) = 0 \text{ for one } k \in J\} = \{i : i \in NO_k \text{ for one } k \in J\} = \bigcup_j \mathcal{NO}_j \tag{B49}$$

$$\mathcal{OT} = \{i : S_{i2}(0) = 0, S_{i2}(1) = 1\} = \{i : \prod_j s^j_{i2}(0) = 0, \prod_j s^j_{i2}(1) = 1\}$$

$$\{i : s^k_{i2}(0) = 0 \text{ for one } k \in J, s^j_{i2}(0) = 1 \quad \forall j \neq k, \quad s^j_{i2}(1) = 1 \quad \forall j\} \tag{B50}$$

$$= \{i : i \in \mathcal{OT}_k \text{ for one } k \in J, i \in \mathcal{AO}_j \quad \forall j \neq k\} = \bigcup_j \mathcal{OT}_j \tag{B51}$$

$$\mathcal{OC} = \{i : S_{i2}(0) = 1, S_{i2}(1) = 0\} = \{i : \prod_j s^j_{i2}(0) = 1, \prod_j s^j_{i2}(1) = 0\}$$

$$\{i : s^j_{i2}(0) = 1 \quad \forall j, \quad s^k_{i2}(1) = 0 \text{ for one } k \in J, s^j_{i2}(1) = 1 \quad \forall j \neq k \quad \forall j\} \tag{B52}$$

$$= \{i : i \in \mathcal{OC}_k \text{ for one } k \in J, i \in \mathcal{AO}_j \quad \forall j \neq k\} = \bigcup_j \mathcal{OC}_j \tag{B53}$$

where (B48), (B50) and (B52) follow from the sources being mutually exclusive and (B49) from the no intersection of OC and OT strata. As shown in the text, when a unit belongs to the NO stratum according to one source, then it belongs to the AO stratum according to all the remaining sources. It suffices to belong to the NO stratum according to one source to belong to the NO stratum overall. Therefore, the NO stratum is given by the union of NO strata according to all the sources. The same reasoning can be used for OC and OT strata.

These sets constitute a partition of the sample. Consider any source j. All the units must belong to one of the four strata defined by the source j, i.e., $i \in \{\mathcal{AO}_j, \mathcal{NO}_j, \mathcal{OT}_j, \mathcal{OC}_j\}$. If a unit belongs to the NO, OT or OC stratum according to this source, then mutual

exclusivity and no intersection of OC and OT strata imply that this unit belongs to the AO stratum for all the other sources. This implies that no unit belongs to more than one set in $\{\mathcal{AO}, \mathcal{NO}, \mathcal{OT}, \mathcal{OC}\}$ simultaneously. Furthermore, the union of these sets equal to the whole sample by construction, so they constitute a partition of the sample. ■

B.5 Proof of Proposition 3

For illustration purposes, I derive first the expression for π_1 with $J = 2$. For $j = 1$ positive monotonicity holds while for $j = 2$ negative monotonicity holds. Following the same logic, I prove then the general case for both π_1 and π_0.

Proof. Consider the case with $J = 2$, where positive monotonicity holds for $j = 1$ ($s_{i2}^1(1) \geq$

$s^1_{i2}(0)$) and negative monotonicity holds for $j = 2$ ($s^2_{i2}(1) \leq s^2_{i2}(0)$).

$$\pi_1 = Pr(S_{i2}(0) = 1 \mid G_i = 1, S_{i2}(1) = 1) \tag{B54}$$

$$= \frac{Pr(S_{i2}(0) = 1, S_{i2}(1) = 1 \mid G_i = 1)}{Pr(S_{i2}(1) = 1 \mid G_i = 1)} \tag{B55}$$

$$= \frac{1 - Pr(S_{i2}(0) = 0, S_{i2}(1) = 1 \mid G_i = 1) - Pr(S_{i2}(0) = 1, S_{i2}(1) = 0 \mid G_i = 1) - Pr(S_{i2}(0) = 0, S_{i2}(1) =}{\mathbb{E}[S_{i2} \mid G_i = 1]} \tag{B56}$$

$$= \frac{1 - Pr(s^1_{i2}(0) = 0, s^1_{i2}(1) = 1 \mid G_i = 1) - Pr(s^2_{i2}(0) = 1, s^2_{i2}(1) = 0 \mid G_i = 1) - Pr(S_{i2}(0) = 0, S_{i2}(1) =}{\mathbb{E}[S_{i2} \mid G_i = 1]} \tag{B57}$$

$$\begin{aligned} = &\frac{1 - Pr(s^1_{i2}(0) = 0, s^1_{i2}(1) = 1 \mid G_i = 1) - Pr(s^2_{i2}(0) = 1, s^2_{i2}(1) = 0 \mid G_i = 1)}{\mathbb{E}[S_{i2} \mid G_i = 1]} \\ &- \frac{Pr(s^1_{i2}(0) = 0, s^1_{i2}(1) = 0 \mid G_i = 1) + Pr(s^2_{i2}(0) = 0, s^2_{i2}(1) = 0 \mid G_i = 1)}{\mathbb{E}[S_{i2} \mid G_i = 1]} \\ &+ \underbrace{\frac{Pr(s^1_{i2}(0) = 0, s^1_{i2}(1) = 0, s^2_{i2}(0) = 0, s^2_{i2}(1) = 0 \mid G_i = 1)}{\mathbb{E}[S_{i2} \mid G_i = 1]}}_{0 \text{ (by mutual exclusivity)}} \end{aligned} \tag{B58}$$

$$= \frac{1 - Pr(s^1_{i2}(0) = 0 \mid G_i = 1) - Pr(s^2_{i2}(1) = 0 \mid G_i = 1)}{\mathbb{E}[S_{i2} \mid G_i = 1]} \tag{B59}$$

$$= \frac{1 - (1 - \mathbb{E}[s^1_{i2}(0) \mid G_i = 1]) - (1 - \mathbb{E}[s^2_{i2} \mid G_i = 1])}{\mathbb{E}[S_{i2} \mid G_i = 1]} \tag{B60}$$

$$= \frac{1}{\mathbb{E}[S_{i2} \mid G_i = 1]} \left(1 - \left(1 - \mathbb{E}[s^1_{i1} \mid G_i = 1] \frac{\mathbb{E}[s^1_{i2} \mid G_i = 0]}{\mathbb{E}[s^1_{i1} \mid G_i = 0]}\right) - \left(1 - \mathbb{E}[s^2_{i2} \mid G_i = 1]\right)\right) \tag{B61}$$

where (B57) follows from the monotonicity assumptions and (B61) from the selection model in Assumption 6 applied to each source of attrition. ■

Proof. Consider the generic case with J different sources. Let J^+ denote the set of sources for which positive monotonicity holds and J^- the set of sources for which negative monotonicity holds.

Mutual exclusivity of the attrition sources implies:

$$Pr(V_i = NO) = \Pr\left(i \in \bigcup_j \mathcal{NO}_j\right) = \sum_j Pr\left(s^j_{i2}(0) = 0, s^j_{i2}(1) = 0\right) \tag{B62}$$

as

$$\mathcal{NO}_j \bigcap \mathcal{NO}_k = \varnothing \quad \forall j,k \in J.$$

Similarly,

$$Pr(V_i = OT) = \Pr\left(i \in \bigcup_j \mathcal{OT}_j\right) = \sum_j Pr\left(s^j_{i2}(0) = 0, s^j_{i2}(1) = 1\right) \tag{B63}$$

$$Pr(V_i = OC) = \Pr\left(i \in \bigcup_j \mathcal{OC}_j\right) = \sum_j Pr\left(s^j_{i2}(0) = 1, s^j_{i2}(1) = 0\right) \tag{B64}$$

Then,

$$\pi_1 = Pr(S_{i2}(0) = 1 \mid G_i = 1, S_{i2}(1) = 1) \tag{B65}$$

$$= \frac{Pr(S_{i2}(0) = 1, S_{i2}(1) = 1 \mid G_i = 1)}{Pr(S_{i2}(1) = 1 \mid G_i = 1)} \tag{B66}$$

$$= \frac{1 - Pr(V_i = OT \mid G_i = 1) - Pr(V_i = OC \mid G_i = 1) - Pr(V_i = NO \mid G_i = 1)}{\mathbb{E}[S_{i2} \mid G_i = 1]} \tag{B67}$$

$$= \frac{1 - \sum_j Pr\left(s^j_{i2}(0) = 0, s^j_{i2}(1) = 1 \mid G_i = 1\right) - \sum_j Pr\left(s^j_{i2}(0) = 1, s^j_{i2}(1) = 0 \mid G_i = 1\right)}{\mathbb{E}[S_{i2} \mid G_i = 1]}$$
$$- \frac{\sum_j Pr\left(s^j_{i2}(0) = 0, s^j_{i2}(1) = 0 \mid G_i = 1\right)}{\mathbb{E}[S_{i2} \mid G_i = 1]} \tag{B68}$$

$$= \frac{1 - \sum_{j \in J^+} Pr\left(s^j_{i2}(0) = 0, s^j_{i2}(1) = 1 \mid G_i = 1\right) - \sum_{j \in J^-} Pr\left(s^j_{i2}(0) = 1, s^j_{i2}(1) = 0 \mid G_i = 1\right)}{\mathbb{E}[S_{i2} \mid G_i = 1]}$$
$$- \frac{\sum_j Pr\left(s^j_{i2}(0) = 0, s^j_{i2}(1) = 0 \mid G_i = 1\right)}{\mathbb{E}[S_{i2} \mid G_i = 1]} \tag{B69}$$

$$= \frac{1 - \sum_{j \in J^+} Pr\left(s^j_{i2}(0) = 0 \mid G_i = 1\right) - \sum_{j \in J^-} Pr\left(s^j_{i2}(1) = 0 \mid G_i = 1\right)}{\mathbb{E}[S_{i2} \mid G_i = 1]} \tag{B70}$$

$$= \frac{1 - \sum_{j \in J^+} \left(1 - \mathbb{E}[s^j_{i2}(0) \mid G_i = 1]\right) - \sum_{j \in J^-} \left(1 - \mathbb{E}[s^j_{i2} \mid G_i = 1]\right)}{\mathbb{E}[S_{i2} \mid G_i = 1]} \tag{B71}$$

$$= \frac{1}{\mathbb{E}[S_{i2} \mid G_i = 1]} \left(1 - \sum_{j \in J^+} \left(1 - \mathbb{E}[s^j_{i1} \mid G_i = 1] \frac{\mathbb{E}[s^j_{i2} \mid G_i = 0]}{\mathbb{E}[s^j_{i1} \mid G_i = 0]}\right) - \sum_{j \in J^-} \left(1 - \mathbb{E}[s^j_{i2} \mid G_i = 1]\right)\right), \tag{B72}$$

where equation (B68) follows from equations (B62-B64), equation (B69) follows from the monotonicity assumptions and equation (B72) follows from the selection model in Assump-

tion 6 applied to each source of attrition. Similarly,

$$\pi_0 = Pr(S_{i2}(1) = 1 \mid G_i = 0, S_{i2}(0) = 1) \tag{B73}$$

$$= \frac{Pr(S_{i2}(0) = 1, S_{i2}(1) = 1 \mid G_i = 0)}{Pr(S_{i2}(0) = 1 \mid G_i = 0)} \tag{B74}$$

$$= \frac{1 - Pr(V_i = OT \mid G_i = 0) - Pr(V_i = OC \mid G_i = 0) - Pr(V_i = NO \mid G_i = 0)}{\mathbb{E}[S_{i2} \mid G_i = 0]} \tag{B75}$$

$$= \frac{1 - \sum_j Pr\left(s^j_{i2}(0) = 0, s^j_{i2}(1) = 1 \mid G_i = 0\right) - \sum_j Pr\left(s^j_{i2}(0) = 1, s^j_{i2}(1) = 0 \mid G_i = 0\right)}{\mathbb{E}[S_{i2} \mid G_i = 0]}$$
$$- \frac{\sum_j Pr\left(s^j_{i2}(0) = 0, s^j_{i2}(1) = 0 \mid G_i = 0\right)}{\mathbb{E}[S_{i2} \mid G_i = 0]} \tag{B76}$$

$$= \frac{1 - \sum_{j \in J^+} Pr\left(s^j_{i2}(0) = 0, s^j_{i2}(1) = 1 \mid G_i = 0\right) - \sum_{j \in J^-} Pr\left(s^j_{i2}(0) = 1, s^j_{i2}(1) = 0 \mid G_i = 0\right)}{\mathbb{E}[S_{i2} \mid G_i = 0]}$$
$$- \frac{\sum_j Pr\left(s^j_{i2}(0) = 0, s^j_{i2}(1) = 0 \mid G_i = 0\right)}{\mathbb{E}[S_{i2} \mid G_i = 0]} \tag{B77}$$

$$= \frac{1 - \sum_{j \in J^+} Pr\left(s^j_{i2}(0) = 0 \mid G_i = 0\right) - \sum_{j \in J^-} Pr\left(s^j_{i2}(1) = 0 \mid G_i = 0\right)}{\mathbb{E}[S_{i2} \mid G_i = 0]} \tag{B78}$$

$$= \frac{1 - \sum_{j \in J^+} \left(1 - \mathbb{E}[s^j_{i2} \mid G_i = 0]\right) - \sum_{j \in J^-} \left(1 - \mathbb{E}[s^j_{i2}(1) \mid G_i = 0]\right)}{\mathbb{E}[S_{i2} \mid G_i = 0]} \tag{B79}$$

$$= \frac{1}{\mathbb{E}[S_{i2} \mid G_i = 0]} \left(1 - \sum_{j \in J^+} \left(1 - \mathbb{E}[s^j_{i2} \mid G_i = 0]\right) - \sum_{j \in J^-} \left(1 - \mathbb{E}[s^j_{i2} \mid G_i = 0] \frac{\mathbb{E}[s^j_{i2} \mid G_i = 1]}{\mathbb{E}[s^j_{i1} \mid G_i = 1]}\right)\right). \tag{B80}$$

■

B.6 Proof of Proposition 2 as a specific case of Proposition 3

Proof. Consider the case where $J = 1$. Without loss of generality, assume the unique source of attrition observable in the data, S, is such that $S_{i2}(1) \geq S_{i2}(0) \quad \forall i$, i.e., $J^- = \varnothing$. From Proposition 3:

$$\pi_0 = \frac{1}{\mathbb{E}[S_{i2} \mid G_i = 0]} \left(1 - (1 - \mathbb{E}[S_{i2} \mid G_i = 0])\right) = \frac{\mathbb{E}[S_{i2} \mid G_i = 0]}{\mathbb{E}[S_{i2} \mid G_i = 0]} = 1 \tag{B81}$$

$$\pi_1 = \frac{1}{\mathbb{E}[S_{i2} \mid G_i = 1]} \left(1 - \left(1 - \mathbb{E}[S_{i1} \mid G_i = 1] \frac{\mathbb{E}[S_{i2} \mid G_i = 0]}{\mathbb{E}[S_{i1} \mid G_i = 0]}\right)\right) = \frac{\mathbb{E}[S_{i1} \mid G_i = 1]}{\mathbb{E}[S_{i2} \mid G_i = 0]} \frac{\mathbb{E}[S_{i2} \mid G_i = 0]}{\mathbb{E}[S_{i1} \mid G_i = 0]} \tag{B82}$$

■

B.7 Proof of Proposition 4

This proof uses the following results from Athey and Imbens (2006):

Lemma B4 *The CiC quantile estimator has an asymptotically linear representation.*

Let $\tau(q) = F^{-1}_{Y_2(1)|G=1}(q) - F^{-1}_{Y_2(0)|G=1}(q)$. *Let* $\hat{\tau}(q)$ *be CiC quantile estimator,* $\hat{\tau}(q) = \widehat{F}^{-1}_{Y_2(1)|G=1}(q) - \widehat{F}^{-1}_{Y_2|G=0}\left(\widehat{F}_{Y_1|G=0}\left(\widehat{F}^{-1}_{Y_1|G=1}(q)\right)\right)$. *Then,*

$$\hat{\tau}(q) = \sum_{g,t} \hat{\tau}_{g,t}(q) + o_p(n^{-1/2})$$

Proof: See Athey and Imbens (2006, Theorem 5.3).

Corollary 14 *The CiC quantile estimator is asymptotically normal.*

$$\sqrt{n}(\hat{\tau}(q) - \tau(q)) \xrightarrow{d} \mathcal{N}(0, V_q^{12} + V_q^{02} + V_q^{01} + V_q^{11})$$

Proof: See Athey and Imbens (2006, Theorem 5.3).

In general, the asymptotic variance of the CiC quantile estimator is difficult to interpret. However, it is important to note that it is a function of the quantile q, and that the variance is also linear in g and t.

Lemma B5 *The estimated proportions are asymptotically normal.*

$$\sqrt{n}(\hat{\pi}_1 - \pi_1) \xrightarrow{d} \mathcal{N}(0, V_{\pi_1})$$
$$\sqrt{n}(\hat{\pi}_0 - \pi_0) \xrightarrow{d} \mathcal{N}(0, V_{\pi_0})$$

Proof: See Athey and Imbens (2006, Theorem 5.4).

Proof. I will derive the asymptotic distribution for $\widehat{\Lambda^{LB}}(q)$. The same steps could be followed to derive the asymptotic distribution of $\widehat{\Lambda^{UB}}(q)$.

First, denote with

$$\tilde{q} := F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q\pi_1 + 1 - \pi_1)\right) + 1 - \pi_0$$

and

$$q^*_{LB} := \min\{\tilde{q}, 1\}$$

The structure of the proof is as follows. First, I show that $\hat{q}^*_{LB}$ converges in probability to q^*_{LB}. Then, I use the Delta method to prove asymptotic normality for the interior case when $\tilde{q} < 1$, which is guaranteed when $q \in (0,1)$ and $F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q\pi_1 + 1 - \pi_1)\right) + 1 - \pi_0 < 1$.

- $\hat{q}^*_{LB} \xrightarrow{p} q^*_{LB}$

From Lemma B5, it follows that $\hat{\pi}_0 \xrightarrow{p} \pi_0$ and $\hat{\pi}_1 \xrightarrow{p} \pi_1$. From the Glivenko–Cantelli theorem, I know that the empirical distribution function converges almost surely to the true distribution function, $\widehat{F}_Y(y) \xrightarrow{a.s.} F_Y(y)$. Because Y is continuous with compact support, it follows that $\widehat{Q}_Y(q) \xrightarrow{p} Q_Y(q)$. Then, by the continuous mapping theorem,

$$\widehat{Q}_{Y_1|G=1,S_2=1}(q\hat{\pi}_1 + 1 - \hat{\pi}_1) \xrightarrow{p} Q_{Y_1|G=1,S_2=1}(q\pi_1 + 1 - \pi_1) \quad \text{(B83)}$$

$$\widehat{F}_{Y_1|G=0,S_2=1}\left(\widehat{Q}_{Y_1|G=1,S_2=1}(q\hat{\pi}_1 + 1 - \hat{\pi}_1)\right) + 1 - \hat{\pi}_0 \xrightarrow{p}$$

$$F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q\pi_1 + 1 - \pi_1)\right) + 1 - \pi_0 = \tilde{q} \quad \text{(B84)}$$

$$\hat{q}^*_{LB} = \min\left\{\widehat{F}_{Y_1|G=0,S_2=1}\left(\widehat{Q}_{Y_1|G=1,S_2=1}(q\hat{\pi}_1 + 1 - \hat{\pi}_1)\right) + 1 - \hat{\pi}_0, 1\right\} \xrightarrow{p} \min\{\tilde{q}, 1\} = q^*_{LB} \quad \text{(B85)}$$

- Interior case, $\tilde{q} < 1$

In an abuse of notation, denote:

$$\widehat{\Lambda^{LB}}(q) = \widehat{Q}_{Y_2|G=1,S_2=1}(q\hat{\pi}_1) - \widehat{Q}_{Y_2|G=0,S_2=1}\left(\widehat{F}_{Y_1|G=0,S_2=1}\left(\widehat{Q}_{Y_1|G=1,S_2=1}(q\hat{\pi}_1 + 1 - \hat{\pi}_1)\right) + 1 - \hat{\pi}_0\right) = \widehat{\Lambda}_q(\hat{\pi}_1, \hat{\pi}_0)$$

$$\Lambda^{LB}(q) = Q_{Y_2|G=1,S_2=1}(q\pi_1) - Q_{Y_2|G=0,S_2=1}\left(F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q\pi_1 + 1 - \pi_1)\right) + 1 - \pi_0\right) = \Lambda_q(\pi_1, \pi_0)$$

Using a first-order Taylor approximation, I can write:

$$\widehat{\Lambda}_q\left(\hat{\pi}_1,\hat{\pi}_0\right) \approx \widehat{\Lambda}_q\left(\pi_1,\pi_0\right) + \frac{\partial\widehat{\Lambda}_q\left(\pi_1,\pi_0\right)}{\partial\pi_1}\left(\hat{\pi}_1-\pi_1\right) + \frac{\partial\widehat{\Lambda}_q\left(\pi_1,\pi_0\right)}{\partial\pi_0}\left(\hat{\pi}_0-\pi_0\right) \tag{B86}$$

$$\widehat{\Lambda}_q\left(\hat{\pi}_1,\hat{\pi}_0\right) - \Lambda_q\left(\pi_1,\pi_0\right) \approx \widehat{\Lambda}_q\left(\pi_1,\pi_0\right) - \Lambda_q\left(\pi_1,\pi_0\right) + \frac{\partial\widehat{\Lambda}_q\left(\pi_1,\pi_0\right)}{\partial\pi_1}\left(\hat{\pi}_1-\pi_1\right) + \frac{\partial\widehat{\Lambda}_q\left(\pi_1,\pi_0\right)}{\partial\pi_0}\left(\hat{\pi}_0-\pi_0\right) \tag{B87}$$

From Corollary 14, it follows that

$$\sqrt{n}\left(\widehat{\Lambda}_q(\pi_1,\pi_0) - \Lambda_q(\pi_1,\pi_0)\right) \xrightarrow{d} \mathcal{N}\left(0, V^{12}_{q\pi_1} + V^{02}_{q+1-\pi_0} + V^{01}_{q} + V^{11}_{q\pi_1+1-\pi_1}\right). \tag{B88}$$

From Corollary 14, and provided that $\widehat{\Lambda}_q(\pi_1,\pi_0)$ uniformly converges to $\Lambda_q(\pi_1,\pi_0)$, I can use the mean value theorem to write:

$$\frac{\partial\widehat{\Lambda}_q\left(\pi_1,\pi_0\right)}{\partial\pi_1} = \lim_{h\to 0}\frac{\widehat{\Lambda}_q\left(\pi_1+h,\pi_0\right)-\widehat{\Lambda}_q\left(\pi_1,\pi_0\right)}{h} \xrightarrow{p} \lim_{h\to 0}\frac{\Lambda_q\left(\pi_1+h,\pi_0\right)-\Lambda_q\left(\pi_1,\pi_0\right)}{h} = \frac{\partial\Lambda_q\left(\pi_1,\pi_0\right)}{\partial\pi_1} \tag{B89}$$

From equation (B89), Lemma B5 and Slutsky theorem, it follows that:

$$\frac{\partial\widehat{\Lambda}_q\left(\pi_1,\pi_0\right)}{\partial\pi_1}\left(\hat{\pi}_1-\pi_1\right) \xrightarrow{d} \mathcal{N}\left(0, \left(\frac{\partial\Lambda_q\left(\pi_1,\pi_0\right)}{\partial\pi_1}\right)^2 V_{\pi_1}\right) \tag{B90}$$

Similarly,

$$\frac{\partial\widehat{\Lambda}_q\left(\pi_1,\pi_0\right)}{\partial\pi_0}\left(\hat{\pi}_0-\pi_0\right) \xrightarrow{d} \mathcal{N}\left(0, \left(\frac{\partial\Lambda_q\left(\pi_1,\pi_0\right)}{\partial\pi_0}\right)^2 V_{\pi_0}\right) \tag{B91}$$

Combining equations (B88) and (B90) and (B91), it follows that

$$\sqrt{n}\left(\widehat{\Lambda}_q\left(\hat{\pi}_1,\hat{\pi}_0\right) - \Lambda_q\left(\pi_1,\pi_0\right)\right) \xrightarrow{d} \mathcal{N}\left(0,\varsigma^2_{LB}\right), \tag{B92}$$

where

$$\varsigma^2_{LB} = \left(V^{12}_{q\pi_1} + V^{02}_{q+1-\pi_0} + V^{01}_{q} + V^{11}_{q\pi_1+1-\pi_1}\right) + \left(\frac{\partial\Lambda_q\left(\pi_1,\pi_0\right)}{\partial\pi_1}\right)^2 V_{\pi_1} + \left(\frac{\partial\Lambda_q\left(\pi_1,\pi_0\right)}{\partial\pi_0}\right)^2 V_{\pi_0}. \tag{B93}$$

The second and third terms in ς^2_{LB} come from estimating the proportion of AO units in each of the groups. If these proportions were known, these two terms would disappear. Similarly, the less precisely these proportions are estimated, i.e., the larger V_{π_1} and V_{π_0} are, the more

they contribute to the variance of the bounds.

■

B.8 Proof of Proposition 5

This proof uses the following result from Horowitz and Manski (1995):

Lemma B6 *Let Y be a continuous random variable and a mixture of two random variables, Y_0 and Y_1, with known mixing probability, p. The c.d.f. of the observed Y is $F(y) = (1-p)F_0(y) + pF_1(y)$, where F_0 and F_1 are the c.d.f.s of Y_0 and Y_1 respectively. The c.d.f. of Y after trimming the $1-p$ upper tail of Y is given by $G(y) = \min\left\{\frac{F(y)}{p}, 1\right\}$. Then, $\int_{-\infty}^{\infty} ydG(y) \leq \int_{-\infty}^{\infty} ydF_1(y)$ and $\int_{-\infty}^{\infty} ydG(y)$ is a sharp lower bound for $\int_{-\infty}^{\infty} ydF_1(y)$. The c.d.f. of Y after trimming the $1-p$ bottom tail of Y is given by $G'(y) = \max\left\{0, \frac{F(y)-(1-p)}{p}\right\}$. Then, $\int_{-\infty}^{\infty} ydG'(y) \geq \int_{-\infty}^{\infty} ydF_1(y)$ and $\int_{-\infty}^{\infty} ydG'(y)$ is a sharp upper bound for $\int_{-\infty}^{\infty} ydF_1(y)$. Proof: See Horowitz and Manski (1995, Proposition 4).*

Proof. From equation (A2), it follows that the lower bound for the ATT_{AO} is given by the lower bound of $\mathbb{E}[\ddot{Y}_i \mid G_i = 1, V_i = AO]$ and the upper bound of $\mathbb{E}[\ddot{Y}_i \mid G_i = 0, V_i = AO]$. To prove Proposition 5, I prove next that $\mathbb{E}[\ddot{Y}_i \mid G_i = 1, S_{i2} = 1, \ddot{Y}_i \leq \ddot{y}^1_{\pi_1}]$ is a lower bound for $\mathbb{E}[\ddot{Y}_i \mid G_i = 1, V_i = AO]$. Similar arguments can be employed for the upper bound of $\mathbb{E}[\ddot{Y}_i \mid G_i = 0, V_i = AO]$ and the upper bound of the ATT_{AO}.

Let $F(\ddot{y})$ be the c.d.f. of $\ddot{Y}$ given $S_2 = 1$, $G = 1$. It can be expressed as the convex combination of the c.d.f. of $\ddot{Y}$ for the AO and OT strata:

$$F(\ddot{y}) = \pi_1 F^{AO}(\ddot{y}) + (1-\pi_1)F^{OT}(\ddot{y}), \tag{B94}$$

where $F^v(\ddot{y})$ is the c.d.f. of $\ddot{Y}$ given $G = 1$ and $V = v$. $\pi_1 = Pr(S_{i2}(0) = 1 \mid G_i = 1, S_{i2}(1) = 1)$, as defined in Proposition 5, is taken as given. Then, it follows that

$$\mathbb{E}[\ddot{Y}_i \mid G_i = 1, S_{i2} = 1, \ddot{Y}_i \leq \ddot{y}^1_{\pi_1}] = \frac{1}{\pi_1}\int_{-\infty}^{\ddot{y}^1_{\pi_1}} ydF(\ddot{y}) \leq \int_{-\infty}^{\infty} ydF^{AO}(\ddot{y}) = \mathbb{E}[\ddot{Y} \mid G = 1, V = AO], \tag{B95}$$

where $\ddot{y}^1_{\pi_1}$ is the π_1th quantile of the distribution of $\ddot{Y}$ given $S_2 = 1$ and $G = 1$. The inequality is given by Lemma B6. From Lemma B6 it also follows that this bound is sharp, provided that π_1 is uniquely identified in the data. ■

B.9 Proof of Proposition 6

This proof is similar to the proofs of Propositions 2 and 3 in Lee (2009) and uses the following results from Newey and McFadden (1994):

Lemma B7 *Suppose that* $z_i, (i = 1, 2...)$, *are i.i.d.,* $\mathbb{E}[g(z, \theta)] = 0$ *if and only if* $\theta = \theta^*$; $\theta^* \in \Theta$, *which is compact;* $g(z, \theta)$ *is continuous at each* $\theta \in \Theta$ *with probability one;* $\mathbb{E}[\sup_{\theta \in \Theta} \|g(z, \theta)\|] < \infty$. *Then,* $\hat{\theta} \xrightarrow{p} \theta^*$.
Proof: See Newey and McFadden (1994, Theorem 2.6).

Lemma B8 *Suppose that* $\hat{g}_n(\hat{\theta})' \hat{g}_n(\hat{\theta}) \leq \inf_{\theta \in \Theta} \hat{g}_n(\theta)' \hat{g}_n(\theta) + o_p(n^{-1})$, $\hat{\theta} \xrightarrow{p} \theta^*$, *where there is* $g^*(\theta^*)$ *is such that (i)* $g^*(\theta^*) = 0$, *(ii)* $g^*(\theta^*)$ *is differentiable at* θ^* *with derivative* G *such that* $G'G$ *is nonsingular, (iii)* θ^* *is an interior point of* Θ, *(iv)* $\sqrt{n}\hat{g}_n(\theta^*) \xrightarrow{\mathrm{d}} N(0, \Sigma)$; (v) *for any* $\delta_n \longrightarrow 0, \sup_{\|\theta - \theta^*\| \leqslant \delta_n} \sqrt{n} \|\hat{g}_n(\theta) - \hat{g}_n(\theta^*) - g^*(\theta)\| / [1 + \sqrt{n}\|\theta - \theta^*\|] \xrightarrow{\mathrm{p}} 0$. *Then* $\sqrt{n}\left(\hat{\theta} - \theta^*\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, (G'G)^{-1} G'\Sigma G (G'G)^{-1}\right)$.
Proof: See Newey and McFadden (1994, Theorem 7.2).

Proof. To prove consistency, it suffices to prove that $\frac{\sum_i S_{i2} G_i \mathbb{1}(\ddot{Y}_i \leq \hat{\ddot{y}}^1_{\hat{\pi}_1}) \ddot{Y}_i}{\sum_i S_{i2} G_i \mathbb{1}(\ddot{Y}_i \leq \hat{\ddot{y}}^1_{\hat{\pi}_1})}$ is a consistent estimator of $\mathbb{E}[\ddot{Y}_i \mid G_i = 1, S_{i2} = 1, \ddot{Y}_i \leq \ddot{y}^1_{\pi_1}]$. Similar steps can be used to prove that $\frac{\sum_i S_{i2}(1-G_i) \mathbb{1}(\ddot{Y}_i \geq \hat{\ddot{y}}^0_{1-\hat{\pi}_0}) \ddot{Y}_i}{\sum_i S_{i2}(1-G_i) \mathbb{1}(\ddot{Y}_i \geq \hat{\ddot{y}}^0_{1-\hat{\pi}_0})}$ is a consistent estimator of $\mathbb{E}[\ddot{Y}_i \mid G_i = 0, S_{i2} = 1, \ddot{Y}_i \leq \ddot{y}^0_{\pi_0}]$. Then, by the Continuous Mapping Theorem for p-convergence, it follows that $\widehat{\Delta^{LB}}$ is a consistent estimator for Δ^{LB}. Symmetric arguments can be used to prove that $\widehat{\Delta^{UB}}$ is consistent for Δ^{UB}.

Consider the case with $J = 2$. For $j = 1$ positive monotonicity holds while for $j = 2$

negative monotonicity holds. Define the moment function

$$g(z_i,\theta)=\begin{pmatrix}(\ddot{Y}_i-\mu)S_{i2}G_i\mathbb{1}(\ddot{Y}_i\leq\ddot{y}^1_{\pi_1})\\(\mathbb{1}(\ddot{Y}_i\leq\ddot{y}^1_{\pi_1}-\pi^1_1)S_{i2}G_i\\(s^1_{i1}-\alpha^1_{1,1})G_i\\(s^1_{i1}-\alpha^1_{1,0})(1-G_i)\\(s^1_{i2}-\alpha^1_{2,0})(1-G_i)\\(s^2_{i2}-\alpha^2_{2,1})G_i\\\left(S_{i2}-\frac{\alpha^1_{1,1}\frac{\alpha^1_{2,0}}{\alpha^1_{1,0}}+\alpha^2_{2,1}}{\pi_1}\right)G_i\end{pmatrix},$$

where $\theta'=(\mu,\ddot{y}^1_{\pi_1},\alpha^1_{1,1},\alpha^1_{1,0},\alpha^1_{2,0},\alpha^2_{2,1},\pi_1)$, $\theta'^*=(\mu^*,\ddot{y}^{1*}_{\pi_1},\alpha^{1*}_{1,1},\alpha^{1*}_{1,0},\alpha^{1*}_{2,0},\alpha^{2*}_{2,1},\pi^*_1)$, $\mu^*\equiv\mathbb{E}[\ddot{Y}_i\mid G_i=1,S_{i2}=1,\ddot{Y}_i\leq\ddot{y}^1_{\pi_1}]$, $\alpha^{j*}_{t,g}\equiv Pr(s^j_t=1\mid G=g)$. The estimators provided in equations (2), (A3) and (A5) are the solution to

$$\min_{\theta\in\Theta}\left(\sum_i g(z_i,\theta)\right)'\left(\sum_i g(z_i,\theta)\right).$$

This just-identified system yields only one solution as long as any of the following expectations are equal to zero: $\mathbb{E}[s^1_1\mid G=1],\mathbb{E}[s^1_1\mid G=0],\mathbb{E}[s^1_2\mid G=0],\mathbb{E}[s^2_2\mid G=1]$. Θ is compact provided that it consists of the bounds of the support for the trimmed mean and quantiles, and the unit interval for the probabilities. Given that the supports are bounded and $g(z,\theta)$ is continuous at each $\theta\in\Theta$, from Lemma B7 it follows that $\frac{\sum_i S_{i2}G_i\mathbb{1}(\ddot{Y}_i\leq\hat{\ddot{y}}^1_{\hat{\pi}_1})\ddot{Y}_i}{\sum_i S_{i2}G_i\mathbb{1}(\ddot{Y}_i\leq\hat{\ddot{y}}^1_{\hat{\pi}_1})}$ is a consistent estimator of $\mathbb{E}[\ddot{Y}_i\mid G_i=1,S_{i2}=1,\ddot{Y}_i\leq\ddot{y}^1_{\pi_1}]$.

Once consistency is proved, asymptotic normality follows from Lemma B8 as shown in Lee (2009, Proposition 3) and Kang and Lee (2024).

■

B.10 Proof of Proposition 10

Proof. Let π_g be the proportion of AO units in group g in the post-treatment period, that is, $\pi_g = Pr(V_i = AO \mid G_i = g, S_{i2} = 1)$. Using the Law of Total Probability, I can write:

$$Pr(V_i = AO \mid G_i = g) = Pr(V_i = AO, S_{i2} = 1 \mid G_i = g) + Pr(V_i = AO, S_{i2} = 0 \mid G_i = g). \tag{B96}$$

Because $V_i = AO \implies S_{i2} = 1$, it follows that $Pr(V_i = AO, S_{i2} = 0 \mid G_i = g) = 0$. Therefore, equation (B96) becomes:

$$\begin{aligned} Pr(V_i = AO \mid G_i = g) &= Pr(V_i = AO, S_{i2} = 1 \mid G_i = g) && \text{(B97)} \\ &= Pr(V_i = AO \mid G_i = g, S_{i2} = 1) Pr(S_{i2} = 1 \mid G_i = g) && \text{(B98)} \\ &= \pi_g \mathbb{E}[S_{i2} \mid G_i = g] && \text{(B99)} \end{aligned}$$

By assumption 3, this probability is the same in the pre-treatment period and in the post-treatment period. Recall that V_i is defined using the potential selection outcomes in the post-treatment period. Assumption 3 ensures that units sampled in the pre-treatment period have the same distribution of post-treatment selection outcomes as units sampled in the post-treatment period.

Using the same logic for the pre-treatment period, it follows that:

$$Pr(V_i = AO \mid G_i = g) = Pr(V_i = AO, S_{i1} = 1 \mid G_i = g) + Pr(V_i = AO, S_{i1} = 0 \mid G_i = g). \tag{B100}$$

By Assumption 2, it follows that $V_i = AO \implies S_{i1} = 1$, which implies that $Pr(V_i = AO, S_{i1} = 0 \mid G_i = g) = 0$. That is, all the units sampled in the pre-treatment period that would belong to the AO stratum in the post-treatment period have to have $S_{i1} = 1$. From this, it follows that

$$\begin{aligned} Pr(V_i = AO \mid G_i = g) &= Pr(V_i = AO, S_{i1} = 1 \mid G_i = g) && \text{(B101)} \\ &= Pr(V_i = AO \mid G_i = g, S_{i1} = 1) Pr(S_{i1} = 1 \mid G_i = g) && \text{(B102)} \\ &= \varpi_g \mathbb{E}[S_{i1} \mid G_i = g] && \text{(B103)} \end{aligned}$$

Finally, combining equations (B99) and , it follows that

$$\pi_g \mathbb{E}[S_{i2} \mid G_i = g] = \varpi_g \mathbb{E}[S_{i1} \mid G_i = g] \tag{B104}$$

$$\varpi_g = \pi_g \frac{\mathbb{E}[S_{i2} \mid G_i = g]}{\mathbb{E}[S_{i1} \mid G_i = g]} \tag{B105}$$

■

B.11 Proof of Proposition 11

If Assumption 2 does not hold, then ϖ_g can only be partially identified.

Proof.

$$\varpi_g = Pr(V_i = AO \mid G_i = g, S_{i1} = 1) = \frac{Pr(V_i = AO, S_{i1} = 1 \mid G_i = g)}{\mathbb{E}[S_{i1} = 1 \mid G_i = g]}$$

From equation (B99),

$$Pr(V_i = AO \mid G_i = g) = \pi_g \mathbb{E}[S_{i2} \mid Gi = g].$$

By the Law of Total Probability,

$$Pr(V_i = AO \mid G_i = g) = Pr(V_i = AO, S_{i1} = 1 \mid G_i = g) + Pr(V_i = AO, S_{i1} = 0 \mid G_i = g), \tag{B106}$$

$$\pi_g \mathbb{E}[S_{i2} \mid Gi = g] = Pr(V_i = AO, S_{i1} = 1 \mid G_i = g) + Pr(V_i = AO, S_{i1} = 0 \mid G_i = g). \tag{B107}$$

So that,

$$Pr(V_i = AO, S_{i1} = 1 \mid G_i = g) = \pi_g \mathbb{E}[S_{i2} \mid Gi = g] - Pr(V_i = AO, S_{i1} = 0 \mid G_i = g) \tag{B108}$$

Because

$$0 \leq Pr(V_i = AO, S_{i1} = 0 \mid G_i = g) \leq Pr(S_{i1} = 0 \mid G_i = g) = 1 - \mathbb{E}[S_{i1} \mid G_i = g], \tag{B109}$$

It follows that

$$\pi_g \mathbb{E}[S_{i2} \mid Gi = g] - (1 - \mathbb{E}[S_{i1} \mid G_i = g]) \leq Pr(V_i = AO, S_{i1} = 1 \mid G_i = g) \tag{B110}$$

and

$$Pr(V_i = AO, S_{i1} = 1 \mid G_i = g) \leq \pi_g \mathbb{E}[S_{i2} \mid Gi = g] \tag{B111}$$

All in all, it follows that

$$\varpi_g = \frac{Pr(V_i = AO, S_{i1} = 1 \mid G_i = g)}{\mathbb{E}[S_{i1} = 1 \mid G_i = g]} \leq \min \left\{ 1, \pi_g \frac{\mathbb{E}[S_{i2} \mid G_i = g]}{\mathbb{E}[S_{i1} \mid G_i = g]} \right\}. \tag{B112}$$

and

$$\varpi_g = \frac{Pr(V_i = AO, S_{i1} = 1 \mid G_i = g)}{\mathbb{E}[S_{i1} = 1 \mid G_i = g]} \geq \max \left\{ 0, \frac{\pi_g \mathbb{E}[S_{i2} \mid G_i = g] - (1 - \mathbb{E}[S_{i1} \mid G_i = g])}{\mathbb{E}[S_{i1} \mid G_i = g]} \right\} \tag{B113}$$

■

The upper bound corresponds to the case where all the units in the AO stratum are observed in the pre-treatment period, and is given by equation (B105). This is always the case under Assumption 2. Without this assumption, however, the proportion of units observed in the pre-treatment period may be smaller than the proportion of AO. In that case, the upper bound of ϖ_g is equal to 1, implying that all the observed units belong to the AO stratum. The lower bound corresponds to the case in which all the units not observed in the pre-treatment period belong to the *AO* stratum in the post-treatment period. Consequently, if the proportion of units with not observed outcomes in the pre-treatment period is larger than the proportion of the AO stratum, then the lower bound is just 0.

C Supplementary material

Table C1 illustrates Lemma 3 under Assumptions 7 and 9 when multiple sources of attrition are available in the data. Table C2 shows how the subpopulations based on the observed selection outcomes identify the proportions of the 4 strata.

Table C1: Principal Strata in the post-treatment period with multiple sources of attrition

	Source 1			Source 2			Overall Selection			
Row	$\boldsymbol{s^1(0)}$	$\boldsymbol{s^1(1)}$	$\boldsymbol{V^1}$	$\boldsymbol{s^2(0)}$	$\boldsymbol{s^2(1)}$	$\boldsymbol{V^2}$	$\boldsymbol{S(0)}$	$\boldsymbol{S(1)}$	$\boldsymbol{V}$	**Ruled out**
1	0	0	NO_1	0	0	NO_2	0	0	NO	Mutual Exclusivity
2	0	0	NO_1	0	1	OT_2	0	0	NO	Mutual Exclusivity, Monotonicity
3	0	0	NO_1	1	0	OC_2	0	0	NO	Mutual Exclusivity
4	0	0	NO_1	1	1	AO_2	0	0	NO	No
5	1	1	AO_1	0	0	NO_2	0	0	NO	No
6	1	1	AO_1	0	1	OT_2	0	1	OT	Negative Monotonicity Source 2
7	1	1	AO_1	1	0	OC_2	1	0	OC	No
8	1	1	AO_1	1	1	AO_2	1	1	AO	No
9	0	1	OT_1	0	0	NO_2	0	0	NO	Mutual Exclusivity
10	0	1	OT_1	0	1	OT_2	0	1	OT	Mutual Exclusivity, Monotonicity
11	0	1	OT_1	1	0	OC_2	0	0	NO	No Overlap
12	0	1	OT_1	1	1	AO_2	0	1	OT	No
13	1	0	OC_1	0	0	NO_2	0	0	NO	Mutual Exclusivity, Monotonicity
14	1	0	OC_1	0	1	OT_2	0	0	NO	No Overlap, Monotonicity x 2
15	1	0	OC_1	1	0	OC_2	1	0	OC	Mutual Exclusivity, Monotonicity
16	1	0	OC_1	1	1	AO_2	1	0	OC	Positive Monotonicity Source 1

Notes: For Source 1 positive monotonicity is assumed, i.e., $s^1(1) \geq s^1(0)$. For Source 2 negative monotonicity is assumed, i.e., $s^2(1) \leq s^2(0)$. All the rows that involve the OC_1 stratum according to Source 1 or the OT_2 according to Source 2 are ruled out by these monotonicity assumptions. All the rows such that $s^1(0) = s^2(0) = 0$ or $s^1(1) = s^2(1) = 0$ or both are ruled out by the fact that sources are mutually exclusive. Finally, rows 11 and 14, which constitute the cases in which a unit belongs to the OT stratum according to one source and to the OC stratum according to the other, are ruled out by Assumption 9. All in all, we are left with 5 rows.

Table C2: Map between subpopulations defined based on observed outcomes and Principal Strata.

Subpopulation (s^1, s^2, G)	Strata	Rows in Table C1
$(1,1,1)$	AO, OT	8, 12
$(1,0,1)$	NO, OC	5, 7
$(0,1,1)$	NO	4
$(1,1,0)$	OC, AO	7, 8
$(1,0,0)$	NO	5
$(0,1,0)$	NO, OT	4, 12

Notes: The first column represents the partition of the sample into 6 different subpopulations based on the observed outcomes. The second column represents the composition in terms of the principal strata of these subpopulations. The third column is the correspondence between the principal strata in the second column con the rows in Table C1. For example, consider the first row. This subpopulation, $(s^1, s^2, G) = (1,1,1)$ are the treated units that are observed. In terms of potential outcomes, these are the units such that $s^1(1) = s^2(1) = 1 \implies S(1) = 1$. So that these units are AO or OT. If we look at table C1, once the rows that are ruled out by assumption are excluded, the units in this subpopulation can only belong to rows 8 and 12.

Table C3: Post-treatment selection outcomes for units observed in the pre-treatment period

	Pre-treatment	Post-treatment			
	$S_1 = 1$	$S_1 = 1,\ S_2 = 0$		$S_1 = 1,\ S_2 = 1$	
		Unemployment /Self-employment	Non-response	Paid employment	Total
Control	661	124	150	387	661
Treatment	792	166	125	501	792
Total	1453	290	275	888	1453

Notes: The first column indicates the number of units with salaried earnings observed in the pre-treatment period. Columns 2-4 indicate the post-treatment selection outcome for these units. Specifically, Column 2 shows the number of units observed in the first period who did not have paid employment in the second period. Column 3 reports the number of units that had paid employment in the first period but did not answer the survey in the second period. Column 4 shows the number of units for whom salaried earnings are observed both in the pre-treatment and the post-treatment periods.