Sharp bounds on the MTE with sample selection

Possebom, Vitor

Yale University

30 October 2018

Sharp Bounds on the MTE with Sample Selection*

Vitor Possebom†
Yale University

First Draft: October 2018
This Draft: January 2019

Abstract

I propose a Generalized Roy Model with sample selection that can be used to analyze treatment effects in a variety of empirical problems. First, I decompose, under a monotonicity assumption on the sample selection indicator, the MTR function for the observable outcome when treated as a weighted average of (i) the MTR on the outcome of interest for the always-observed sub-population and (ii) the MTE on the observable outcome for the observed-only-when-treated sub-population, and show that such decomposition can provide point-wise sharp bounds on the MTE of interest for the always-observed sub-population. Moreover, I impose an extra mean dominance assumption and tighten the previous bounds. I, then, show how to point-identify those bounds when the support of the propensity score is continuous. After that, I show how to (partially) identify the MTE of interest when the support of the propensity score is discrete. At the end, I estimate bounds on the MTE of the Job Corps Training Program on hourly wages for the always-employed sub-population and find that it is decreasing with the likelihood of attending the program for the Non-Hispanic group. For example, I find that the ATT is between \$.38 and \$1.17 while the ATU is between \$.73 and \$3.14.

Keywords: Marginal Treatment Effect, Sample Selection, Partial Identification, Principal Stratification, Program Evaluation, Training Programs.

JEL Codes: C31, C35, C36, J38

*I would like to thank Joseph Altonji, Nathan Barker, Ivan Canay, John Finlay, Yuichi Kitamura, Marianne Köhli, Helena Laneuville, Jaewon Lee, Yusuke Narita, Pedro Sant'anna, Masayuki Sawada, Edward Vytlacil, Siuyuat Wong and seminar participants at Yale University for helpful suggestions.
†*Email:* vitoraugusto.possebom@yale.edu

1 Introduction

I propose a Generalized Roy Model (Heckman & Vytlacil (1999)) with sample selection in which there is one outcome of interest that is observed only if the individual self-selects into the sample. So, in addition to the fundamental problem of causal analysis in which I only observe one of the potential outcomes due to endogenous self-selection into treatment, I also face a problem of endogenous sample selection. Such framework is useful to analyze many empirical problems: the effect of a job training program on wages (Heckman et al. (1999), Lee (2009), Chen & Flores (2015)), the college wage premium (Altonji (1993), Card (1999), Carneiro et al. (2011)), scarring effects (Heckman & Borjas (1980), Farber (1993), Jacobson et al. (1993)), the effect of an educational intervention on short- and long-term outcomes (Krueger & Whitmore (2001), Angrist et al. (2006), Angrist et al. (2009), Chetty et al. (2011), Dobbie & Jr. (2015)), the effect of a medical treatment on health quality (CASS (1984), Sexton & Hebel (1984), U.S. Department of Health and Human Services (2004)), the effect of procedural laws on litigation outcomes (Helland & Yoon (2017)), and any randomized control trial that faces an attrition problem (DeMel et al. (2013), Angelucci et al. (2015)). For example, in the case of a job training program, I am interested in its effect on workers' hourly wages (outcome of interest), but I only observe their hourly labor earnings (observable outcome). Note that, in such context, I face two endogeneity problems: self-selection into the training program and self-selection into employment.

Under a monotonicity assumption on the sample selection indicator, I decompose the Marginal Treatment Response (MTR) function for the potential observable outcome when treated as a weighted average of (i) the MTR on the outcome of interest for the sub-population who is always observed and (ii) the Marginal Treatment Effect (MTE) on the observable outcome for the sub-population who is observed only when treated. Under a bounded (in one direction) support condition, such decomposition is useful because it allows me to propose point-wise sharp bounds on the MTE on the outcome of interest for the always-observed sub-population (MTE^{OO}) as a function of the MTR functions on the observable outcome, the maximum and (or) minimum of the support of the potential outcome, and the proportions of always-observed individuals and observed-only-when-treated individuals. I also show that it is impossible to construct bounds without extra assumptions when the support of the potential outcome is the entire real line. After that, I impose an extra mean dominance assumption that compares the always-observed population against the observed-only-when-treated population, tightening the previous bounds. Moreover, under this new assumption, I show that those tighter bounds are also sharp and derive an informative lower bound even when the support of the potential outcome is the entire real line.

I, then, proceed to show that those bounds are well-identified. When the support of the propensity score is an interval, the relevant objects are point-identified by applying the local instrumental variable approach (LIV, see Heckman & Vytlacil (1999)) to the expectations of the observable outcome and of the selection indicator conditional on the propensity score and the treatment status. However, in many empirical applications, the support of the propensity score is a finite set. In such context, I can identify bounds on the MTE^{OO} of interest by adapting the nonparametric bounds proposed by Mogstad et al. (2018) or the flexible parametric approach suggested by Brinch et al. (2017) to encompass a sample selection problem. When using the nonparametric approach, the bounds on the MTE^{OO} of interest are simply an outer set that contains the true MTE^{OO}, i.e., they are not point-wise sharp anymore.

Partial identification of the MTE^{OO} of interest is useful for two reasons. First, bounds on the MTE^{OO} can be used to shed light on the heterogeneity of treatment effects, allowing the researcher to understand who benefits and who loses with a specific treatment. Such knowledge can be used to optimally design policies that provide incentives to agents to take a treatment. Second, bounds on the MTE^{OO} can be used to construct bounds in any treatment effect parameter that is written as a weighted integral of the MTE^{OO}. For example, by taking a weighted average of the point-wise sharp bounds on the MTE^{OO}, one can bound the average treatment effect (ATE), the average treatment effect on the treated (ATT), any local average treatment effect (LATE, Imbens & Angrist (1994)) and any policy-relevant treatment effect (PRTE, Heckman & Vytlacil (2001*b*)) for the always-observed sub-population. Although such bounds may not be sharp for any specific parameter, they are a general and easy-to-apply solution to many empirical problems. Therefore, if the applied researcher is interested in a parameter that already has specific bounds for it (e.g., intention-to-treat treatment effect (ITT^{OO}) by Lee (2009) and local average treatment effect ($LATE^{OO}$) by Chen & Flores (2015) for the always-observed subpopulation), he or she should use a specialized tool. However, if the applied researcher is interested in parameters without specialized bounds (e.g., ATE, ATT and the Average Treatment Effect on the Untreated (ATU) in the case of imperfect compliance), he or she may take a weighted integral of point-wise sharp bounds on the MTE^{OO} of interest. In other words, facing a trade-off between empirical flexibility and sharpness, the partial identification tool proposed in this paper focus on empirical flexibility while still ensuring some notion of sharpness.

At the end, I illustrate the usefulness of the proposed bounds on the MTE^{OO} of interest by analyzing the effect of the Job Corps Training Program (JCTP) on hourly wages for the Non-Hispanic always-employed sub-population. My framework is ideal to analyze such important experiment because it simultaneously addresses the imperfect compliance issue (self-selection into treatment) by focusing on the MTE, and the endogenous employment decision (sample selection) by using a partial identification strategy. Although my MTE^{OO} bounds are uninformative when using only the monotonicity assumption, they are tight and positive under a mean dominance assumption, illustrating the identification power of extra assumptions in a context of partial identification. Most interestingly, I find that the bounds of the MTE^{OO} on hourly wages are decreasing on the likelihood of attending the program, implying that the agents who benefit the most from the JCTP are the least likely to attend it. As a consequence of this result, my estimates suggest that ATU is greater than the ATT for the always-employed subpopulation. Moreover, my bounds on the $LATE^{OO}$ are in line with the estimates of Chen & Flores (2015) and the effect of the JCTP on employment is positive for every agent according to the test proposed by Machado et al. (2018). Finally, as a by product of my estimation strategy, I also find that the MTE on employment and hourly labor earnings are decreasing on the likelihood of attending the JCTP.

I make contributions to three literatures: identification of treatment effects using an instrument, identification of treatment effects with sample selection, and the effect of job training programs.

The literature about treatment effects with an instrument is enormous and I only briefly discuss it. Imbens & Angrist (1994) show that we can identify the LATE. Heckman & Vytlacil (1999), Heckman & Vytlacil (2005) and Heckman et al. (2006) define the MTE and explain how to compute any treatment effect as a weighted average of the MTE. However, if the support of the propensity score is not the unit interval, then it is not possible to recover some important treatment effects, such as the ATE, the ATT and the ATU. A parametric solution to this problem is given by Brinch et al. (2017), who identify a flexible polynomial function for the MTE whose degree is defined by the cardinality of the propensity score support.

A nonparametric solution to the impossibility of identifying the ATE and the ATT is bounding them. Mogstad et al. (2018) use the information contained on IV-like estimands to construct non-parametrically worst- and best- case bounds on policy-relevant treatment effects. Other authors focus on imposing weak monotonicity assumptions or a structural model. In the first group, Manski (1990), Manski (1997) and Manski & Pepper (2000) propose bounds for the ATE and ATT. Chen et al. (2017) propose an average monotonicity condition combined with a mean dominance condition across subpopulation groups and sharpen the bounds previously proposed. Huber et al. (2017) add a mean independence condition within subpopulation groups and bound not only the ATE and ATT when there is noncompliance, but also the Average Treatment Effect on the Untreated (ATU) and the ATE for always-takers and never-takers (ATE-AT and ATE-NT).

Complementing the weak monotonicity approach, the structural approach has focused mainly on the binary outcome case due to the need to impose bounded outcome variables. Heckman & Vytlacil (2001*a*), Bhattacharya et al. (2008), Chesher (2010), Chiburis (2010), Shaikh & Vytlacil (2011) and Bhattacharya et al. (2012) made important contributions to this literature, bounding the ATE and the ATT. While Bhattacharya et al. (2008), Shaikh & Vytlacil (2011) and Bhattacharya et al. (2012) consider a thresholding crossing model on the treatment and the outcome variable, Chiburis (2010) assumes a thresholding crossing model only on the outcome variable.

I contribute to this literature about identifying treatment effects using an instrument by extending the non-parametric approach by Mogstad et al. (2018) and the flexible parametric approach by Brinch et al. (2017) to encompass a sample-selection problem. By doing so, I can partially identify the MTE function on the outcome of interest instead of on the observable outcome.

The literature about identification of treatment effects with sample selection is vast and I only briefly discuss it. The control function approach is a possible solution to it and is analyzed by Heckman (1979), Ahn & Powell (1993) and Newey et al. (1999), encompassing parametric, semiparametric and nonparametric tools. Using auxiliary data is another possible solution and is studied by Chen et al. (2008). A nonparametric solution that requires weaker conditions is bounding. In a seminal paper, Lee (2009) imposes a weak monotonicity assumption on the relationship between sample selection and treatment assignment to sharply bound the ITT for the subpopulation of always-observed individuals (ITT^{OO}). Using techniques developed by Frangakis & Rubin (2002), Blundell et al. (2007) and Imai (2008) and a weak monotonicity assumption, Blanco et al. (2013*a*) bound the Intention-to-Treat Quantile Treatment Effect for the always-observed individuals ($Q - ITT^{OO}$). Moreover, by imposing weak dominance assumptions across subpopulation groups, they can sharpen the ITT^{OO} bounds proposed by Lee (2009). Huber & Mellace (2015) additionally impose a bounded support for the outcome variable and propose bounds on the ITT for two other subpopulations: observed-only-when-treated individuals (ITT^{NO}), and observed-only-when-untreated individuals ITT^{ON}. Complementary to those studies, Lechner & Mell (2010) derive bounds for the ITT and the Q-ITT for the treated-and-observed subpopulation, Mealli & Pacini (2013) derive bounds for the ITT when the exclusion restriction is violated and there are two outcome variables, and Behaghel et al. (2015) combines techniques developed by Heckman (1979) and Lee (2009) to propose bounds for the ATE in a survey framework in which the interviewer tries to contact the surveyed individual multiple times.

In the intersection of both literatures, a few authors address the problem of sample selection and endogenous treatment simultaneously. Huber (2014) point-identifies the ATE and the Quantile Treatment Effect (QTE) for the observed sub-population and for the entire population using a nested propensity score based on an instrument for sample selection. Fricke et al. (2015) point-identify the LATE by using a random treatment assignment and a continuous exogenous variable to instrument for treatment status and sample selection. Lee & Salanie (2016), who also include sample selection in a Generalized Roy Model, use two continuous instruments to provide control functions for the selection into treatment and sample selection problems, allowing them to point-identify the MTE.

Although the three previous contributions are important, finding a credible instrument for sample selection is hard, especially in Labor Economics. For this reason, it is important to develop tools that do not rely on the existence of an instrument for sample selection. Frolich & Huber (2014) point-identify the LATE under a predetermined sample-selection assumption, ruling out an contemporaneous relationship between the potential outcomes and the sample selection problem. Chen & Flores (2015) derive bounds for Average Treatment Effect for the always-observed compliers (LATE-OO) by combining one instrument with a double exclusion restriction with monotonicity assumptions on the sample selection and the selection into treatment problems. Moreover, Blanco et al. (2017) and Steinmayr (2014) extend the work by Chen & Flores (2015) by, respectively, considering a censored outcome variable and analyzing mixture variables combining four strata.

I contribute to the literature about identification of treatment effects with sample selection by partially identifying the MTE on the always-observed subsample allowing for a contemporaneous relationship between the potential outcomes and the sample selection problem, and using only one (discrete) instrument combined with a monotonicity assumption. Doing so is theoretically important, because it can unify, in one framework, the bounds for different treatment effects with sample selection, and empirically relevant, because it allow us to partially identify any treatment effect on the outcome of interest in many empirical problems. For example, when analyzing the effect of a job training program on wages, it is important to compare the ATT with the ATU in order to understand whether the workers who benefit the most from such policy are actually the ones who receive training.

The literature about job training programs is immense and I only briefly discuss it. Heckman et al. (1999) wrote an influential survey paper about it, summarizing its main results and challenges. In particular, after a randomized experiment funded by the U.S. Department of Labor in 1995, many papers were written about the effects of the Job Corps Training Program, such as Schochet et al. (2001) and Schochet et al. (2008). They find that the ITT and the LATE are positive for educational attainment (GED and vocational certificates), negative for criminal activity and, positive for employment and earnings beginning in the third year after random assignment. With respect to the heterogeneity of treatment effects, their most interesting result states that there were no employment or earnings effects for Hispanic youths, a result that is further investigated by Flores-Lagunes et al. (2010). Complementing those estimates, Chen et al. (2017) partially identifies the ATE and the ATT on earnings, employment and welfare benefits, finding that they are positive for the first two variables and negative for the last one. When analyzing heterogeneous treatment effects, their lower bounds suggest that the treatment is more effective for a treated youth than for a randomly chosen youth, while their upper bounds support the opposite conclusion.

Finally, the papers that are closer to mine were written by Lee (2009), Blanco et al. (2013*a*) and Chen & Flores (2015), who analyze the effect of the Job Corps Training Program on wages by focusing, respectively, on the ITT, the Q-ITT and the LATE parameters for the always-observed sub-population. Lee (2009) rules out a zero effect after accounting for the loss in labor market experience generated by the extra education acquired by Job Corps participants. Blanco et al. (2013*a*) complement this analysis by finding that the statistically significant $Q-ITT^{OO}$ for non-Hispanic youths is between 2.7% and 14% and relatively stable across different quantiles, while the $Q-ITT^{OO}$ bounds for Hispanic youths are very wide and include the zero. Chen & Flores (2015) find that the $LATE^{OO}$ on hourly wages four years after randomization is between 5.7% and 13.9% for the entire population and between 7.7% and 17.5% for the non-Hispanic population under monotonicity and mean dominance assumptions. Overall, all authors find positive results for the effects of the Job Corps Training

Program.

I contribute to literature about the Job Corps Training Program by analyzing the MTE for the Non-Hispanic group, allowing me to understand heterogeneous treatment effects over the likelihood of attending the program. To summarize those results, I also compute estimates of the ATE^{OO}, the $LATE^{OO}$, the ATT^{OO} and the ATU^{OO}. Moreover, I formally test whether this training program has a monotone effect on employment by implementing the test proposed by Machado et al. (2018) for the the non-Hispanic and Hispanic sub-populations. My empirical results suggests that the agents who are more likely to benefit from the JCTP are the least likely to attend the program.

This paper proceeds as follows: section 2 details the Generalized Roy Model with sample selection; section 3 explains how to derive bounds for the MTE^{OO} of interest; sections 4 and 5 discuss identification of the MTE^{OO} bounds when the support of the propensity score is continuous or discrete; and section 6 analyzes the effect of the Job Corps Training Program on hourly wages. Finally, section 7 concludes.

2 Framework

I begin with the classical potential outcome framework by Rubin (1974) and modify it to include a sample selection problem. Let Z be an instrumental variable whose support is given by $\mathcal{Z}$, X be a vector of covariates whose support is given by $\mathcal{X}$, $W := (X, Z)$ be a vector that combines the covariates and the instrument whose support is given by $\mathcal{W} := \mathcal{X} \times \mathcal{Z}$, D be a treatment status indicator, Y_0^* be the potential outcome of interest when the person is not treated, and Y_1^* be the potential outcome of interest when the person is treated. The outcome variable of interest (e.g., wages) is $Y^* := D \cdot Y_1^* + (1 - D) \cdot Y_0^*$. Moreover, let S_1 and S_0 be potential sample selection indicators when treated and when not treated, and define $S := D \cdot S_1 + (1 - D) \cdot S_0$ as the sample selection indicator (e.g., employment status). Define $Y := S \cdot Y^*$ as the observable outcome (e.g., labor earnings). I also define $Y_1 := S_1 \cdot Y_1^*$ and $Y_0 := S_0 \cdot Y_0^*$ as the potential observable outcomes. Observe that, following Lee (2009) and Chen & Flores (2015), my notation implicit imposes two exclusion restrictions: Z has no direct impact on the potential outcome of interest nor on the sample selection indicator. The second exclusion restriction requires attention in empirical applications. On the one hand, it may be a strong assumption in randomized control trials if sample selection is due to attrition and initial assignment has an effect on the subject's willingness to contact the researchers. On the other hand, it may be a reasonable assumption in many labor market applications, such as the evaluation of a job training program. For example, in my empirical section, it is reasonable that the initial random assignment to the Job Corps Training Program (JCTP) has no impact on future employment status.

I model sample selection and selection into treatment using the Generalized Roy Model (Heckman & Vytlacil 1999). Let U and V be random variables, and $P : \mathcal{W} \to \mathbb{R}$ and $Q : \{0,1\} \times \mathcal{X} \to \mathbb{R}$ be unknown functions. I assume that:

$$D := \mathbf{1}\left\{P\left(W\right) \geq U\right\} \tag{1}$$

and

$$S := \mathbf{1}\left\{Q\left(D, X\right) \geq V\right\}. \tag{2}$$

As Vytlacil (2002) shows, equations (1) and (2) are equivalent to assuming monotonicity conditions on the selection into treatment problem (Imbens & Angrist (1994)) and on the sample selection problem (Lee (2009)). I stress that both monotonicity assumptions are testable using the tools developed by Machado et al. (2018). Note also that, given equation (2), $S_0 = \mathbf{1}\left\{Q\left(0, X\right) \geq V\right\}$ and $S_1 = \mathbf{1}\left\{Q\left(1, X\right) \geq V\right\}$.

The random variables U and V are jointly continuously distributed conditional on X with density $f_{U,V|X} : \mathbb{R}^2 \times \mathcal{X} \to \mathbb{R}$ and cumulative distribution function $F_{U,V|X} : \mathbb{R}^2 \times \mathcal{X} \to \mathbb{R}$. As is well known in the literature, equations (1) and (2) can be rewritten as

$$D = \mathbf{1}\left\{F_{U|X}\left(P\left(W\right)|X\right) \geq F_{U|X}\left(U\,|X\right)\right\} = \mathbf{1}\left\{\tilde{P}\left(W\right) \geq \tilde{U}\right\}$$
$$S = \mathbf{1}\left\{F_{V|X}\left(Q\left(D, X\right)|X\right) \geq F_{V|X}\left(V\,|X\right)\right\} = \mathbf{1}\left\{\tilde{Q}\left(D, X\right) \geq \tilde{V}\right\}$$

where $\tilde{P}(W) \coloneqq F_{U|X}(P(W)|X)$, $\tilde{U} \coloneqq F_{U|X}(U|X)$, $\tilde{Q}(D,X) \coloneqq F_{V|X}(Q(D,X)|X)$, and $\tilde{V} \coloneqq F_{V|X}(V|X)$. Consequently, the marginal distributions of $\tilde{U}$ and $\tilde{V}$ conditional on X follow the standard uniform distribution. Since this is merely a normalization, I drop the tilde and mantain throughout the paper the normalization that the marginal distributions of U and V conditional on X follow the standard uniform distribution and that $(P(w), Q(d,x)) \in [0,1]^2$ for any $(x,z,d) \in \mathcal{W} \times \{0,1\}$. I also assume that:

Assumption 1 *The instrument Z is independent of all latent variables given the covariates X, i.e., $Z \perp\!\!\!\perp (U, V, Y_0^*, Y_1^*)|X$.*

Assumption 2 *The distribution of $P(W)$ given X is nondegenerate.*

Assumption 3 *The first and second population moments of the counterfactual variables are finite, i.e., $\mathbb{E}[|Y_d^*|] < +\infty$, $\mathbb{E}\left[(Y_d^*)^2\right] < +\infty$, and $\mathbb{E}[|S_d|] < +\infty$ for any $d \in \{0,1\}$.*

Assumption 4 *Both treatment groups exist for any value of X, i.e., $0 < \mathbb{P}[D = 1|X] < 1$.*

Assumption 5 *The covariates X are invariant to counterfactual manipulations, i.e., $X_0 = X_1 = X$, where X_0 and X_1 are the counterfactual values of X that would be observed when the person is, respectively, not treated or treated.*

Assumption 6 *The potential outcomes Y_0^* and Y_1^* have the same support, i.e., $\mathcal{Y}^* \coloneqq \mathcal{Y}_0^* = \mathcal{Y}_1^*$, where $\mathcal{Y}_0^* \subseteq \mathbb{R}$ is the support of Y_0^* and $\mathcal{Y}_1^* \subseteq \mathbb{R}$ is the support of Y_1^*.*

Assumption 7 *Define $\underline{y}^* \coloneqq \inf\{y \in \mathcal{Y}^*\} \in \mathbb{R} \cup \{-\infty\}$ and $\overline{y}^* \coloneqq \sup\{y \in \mathcal{Y}^*\} \in \mathbb{R} \cup \{\infty\}$. I assume that $\underline{y}^*$ and $\overline{y}^*$ are known, and that*

1. *$\underline{y}^* > -\infty$, $\overline{y}^* = \infty$ and $\mathcal{Y}^*$ is an interval, or*
2. *$\underline{y}^* = -\infty$, $\overline{y}^* < \infty$ and $\mathcal{Y}^*$ is an interval, or*
3. *$\underline{y}^* > -\infty$, $\overline{y}^* < \infty$ and*
   (a) *$\mathcal{Y}^*$ is an interval or*

(b) $\underline{y}^* \in \mathcal{Y}^*$ and $\overline{y}^* \in \mathcal{Y}^*$.

I stress that assumption 7 is fairly general. Case 1 covers continuous random variables whose support is convex and bounded below (e.g.: wages), while Case 3.a covers continuous variables with bounded convex support (e.g.: test scores). Case 3.b encompasses not only binary variables, but also any discrete variable whose support is finite (e.g.: years of education). It also includes mixed random variables whose support is not an interval but achieves its maximum and minimum. I also highlight that proposition 13 shows that assumption 7 is partially necessary to the existence of bounds on the MTE^{OO} of interest in the sense that, if $\underline{y}^* = -\infty$ and $\overline{y}^* = +\infty$, then it is impossible to bound the marginal treatment effect on the outcome of interest for the always-observed sub-population without any extra assumption.

Assumption 8 *Treatment has a positive effect on the sample selection indicator for all individuals, i.e.,* $Q(1,x) > Q(0,x) > 0$ *for any* $x \in \mathcal{X}$.

Assumption 8 goes beyond the monotonicity condition implicitly imposed by equation (2) by assuming that the direction of the effect of treatment on the sample selection indicator is known and positive, i.e., $Q(1,x) \geq Q(0,x)$ for any $x \in \mathcal{X}$. In this sense, it is a standard assumption in the literature.[1] Most importantly, it is also a testable assumption using the tools developed by Machado et al. (2018), because, under monotone sample selection (equation (2)), identification of the sign of the ATE on the selection indicator provides a test for Assumption 8. However, Assumption 8 is slightly stronger than what is usually imposed in the literature, because it additionally imposes $Q(0,x) > 0$ and $Q(1,x) > Q(0,x)$ for any $x \in \mathcal{X}$. I stress that the first inequality implies that there is a sub-population who is always observed, allowing me to properly define my target parameter — the marginal treatment effect on the outcome of interest for the always-observed population (MTE^{OO}). I also highlight that the second inequality implies that there is a sub-population who is observed only when treated, making the problem theoretically interesting by eliminating trivial cases of point-identification of the MTE^{OO} of interest as discussed in proposition 10. Finally, I emphasize that all my results can

[1] Lee (2009) and Chen & Flores (2015) write it in an equivalent way as $S_1 \geq S_0$, while Manski (1997) and Manski & Pepper (2000) call it the "monotone treatment response" assumption.

be stated and derived with some obvious changes if I impose $Q(0, x) > Q(1, x) > 0$ for any $x \in \mathcal{X}$ instead of Assumption 8, as it is done in Appendix C. I also discuss, in Appendix D, an agnostic approach to monotonicity in the sample selection problem (equation (2)) and show, in Appendix E, that bounds derived with non-monotone sample selection are uninformative (i.e., equal to $\left(\underline{y}^* - \overline{y}^*, \overline{y}^* - \underline{y}^*\right)$) under mild regularity conditions.

In my empirical application, Assumption 8 imposes that the JCTP has a positive effect on employment for all individuals, which is plausible given the objectives and services provided by this training program. As discussed by Chen & Flores (2015), the two potential threats against it — the "lock-in" effect (van Ours (2004)) and an increase in the reservation wage of treated individuals — are likely to become less relevant in the long-run, justifying my focus on the hourly wage after 208 weeks from randomization. Most importantly, this assumption is formally tested by the method developed by Machado et al. (2018) and I reject, at the 1%-significance level, the null hypothesis that Assumption 8 is invalid for the Non-Hispanic group.

Finally, in partial identification contexts, extra assumptions may have a lot of identification power. In the specific case of identifying treatment effects with sample selection, it is common to use mean or stochastic dominance assumptions to tighten the bounds on the parameter of interest (Imai (2008), Blanco et al. (2013*a*), Huber & Mellace (2015) and Huber et al. (2017)) and justify them based on the intuitive argument that some population sub-groups have more favorable underlying characteristics than others. In particular, I discuss the identifying power of the following mean dominance assumption[2]:

Assumption 9 *The potential outcome when treated for the always-observed sub-population is greater than or equal to the same parameter for the observed-only-when-treated sub-population:*

$$\mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 1, S_1 = 1\right.\right] \geq \mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 0, S_1 = 1\right.\right]$$

for any $x \in \mathcal{X}$ *and* $u \in [0, 1]$.

[2]In appendix F, I derive bounds on the MTE of interest when the above inequality holds in the other direction.

Unfortunately, such assumption is empirically untestable, implying that its use must be justified for each application based on qualitative or theoretical arguments. In particular, in my empirical application, it imposes that the marginal treatment response function of wages when treated for the always-employed population is greater than the same object for the employed-only-when-treated population. Similarly to the case discussed by Chen & Flores (2015, section 2.3), Assumption (9) implies a positive correlation between employment and wages, which is supported by standard models of labor supply.

3 Bounds on the MTE$^{\text{OO}}$ on the outcome of interest

The target parameter, the MTE on the outcome of interest for the sub-population who is always observed (MTE^{OO}), is given by

$$\begin{aligned}\Delta_{Y^*}^{OO}(x,u) &:= \mathbb{E}\left[Y_1^* - Y_0^* \,|X=x, U=u, S_0=1, S_1=1\right] \\ &= \mathbb{E}\left[Y_1^* \,|X=x, U=u, S_0=1, S_1=1\right] - \mathbb{E}\left[Y_0^* \,|X=x, U=u, S_0=1, S_1=1\right]\end{aligned} \tag{3}$$

for any $u \in [0,1]$ and any $x \in \mathcal{X}$, and is a natural parameter of interest. In labor market applications where sample selection is due to observing wages only when agents are employed, it is the effect on wages for the subpopulation who is always employed. In medical applications where sample selection is due to the death of a patient, it is the effect on health quality for the subpopulation who survives regardless of the treatment status. In the education literature where sample selection is due to students quiting school, it is the effect on test scores for the subpopulation who do not drop out of school regardless of the treatment status. In all those cases, the target parameter captures the intensive margin of the treatment effect.[3]

Other possibly interesting parameters are the MTE on the outcome of interest for the sub-population who is never observed ($\mathbb{E}\left[Y_1^* - Y_0^* \,|X=x, U=u, S_0=0, S_1=0\right]$, MTE^{NN}),

[3] If the researcher is interested in the extensive margin of the treatment effect, captured by the MTE on the observable outcome ($\mathbb{E}\left[Y_1 - Y_0 \,|X=x, U=u\right]$) and by the MTE on the selection indicator ($\mathbb{E}\left[S_1 - S_0 \,|X=x, U=u\right]$), he or she can apply the identification strategies described by Heckman et al. (2006), Brinch et al. (2017) and Mogstad et al. (2018).

the MTR function under no treatment for the outcome of interest for the sub-population who is observed only when treated ($\mathbb{E}\left[Y_0^* \left| X=x, U=u, S_0=0, S_1=1\right.\right]$, MTR_0^{NO}) and MTR function under treatment for the outcome of interest for the sub-population who is observed only when treated ($\mathbb{E}\left[Y_1^* \left| X=x, U=u, S_0=0, S_1=1\right.\right]$, MTR_1^{NO}). While the last parameter can be partially identified (Appendix B), the first two parameters are impossible to point-identify or bound in a informative way because the outcome of interest (Y_0^* or Y_1^*) is never observed for the conditioning sub-populations. Note also that the sub-population who is observed only when not treated ($S_0 = 1$ and $S_1 = 0$) do not exist by Assumption 8. I also stress that the conditioning subpopulations in all the above-mentioned parameters are determined by post-treatment outcomes and, as a consequence, are connected to the statistical literature known as principal stratification (Frangakis & Rubin (2002)).

I, now, focus on the target parameter $\Delta_{Y^*}^{OO}(x,u)$ given by equation (3). While subsection 3.1 derives bounds on the MTE^{OO} of interest (equation (3)) using only a monotonicity assumption (assumptions 1-8), subsection 3.2 tighten those bounds by additionally imposing the Mean Dominance Assumption 9. Finally, subsection 3.3 discusses the empirical relevance of such bounds.

3.1 Partial Identification with only a Monotonicity Assumption

Here, my goal is to derive bounds on $\Delta_{Y^*}^{OO}(x,u)$ under assumptions 1-8. Note that the second right-hand term in equation (3) can be written as[4]

$$\mathbb{E}\left[Y_0^* \left| X=x, U=u, S_0=1, S_1=1\right.\right] = \frac{m_0^Y(x,u)}{m_0^S(x,u)}, \tag{4}$$

where I define $m_0^Y(x,u) \coloneqq \mathbb{E}\left[Y_0 \left| X=x, U=u\right.\right]$ and $m_0^S(x,u) \coloneqq \mathbb{E}\left[S_0 \left| X=x, U=u\right.\right]$ as the MTR functions associated to the counterfactual variables Y_0 and S_0 respectively. In this section, I assume that all terms in the right-hand side of equation (4) are point-identified, postponing the discussion about their identification to sections 4 and 5.

[4] Appendix A.1 contains a proof of this claim.

The first right-hand term in equation (3) can be written as[5]

$$\mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 1, S_1 = 1\right.\right] = \frac{m_1^Y(x,u) - \Delta_Y^{NO}(x,u) \cdot \Delta_S(x,u)}{m_0^S(x,u)}, \tag{5}$$

where $m_1^Y(x,u) \coloneqq \mathbb{E}\left[Y_1 \left| X = x, U = p\right.\right]$ is the MTR function associated to the counterfactual variable Y_1, $\Delta_Y^{NO}(x,u) \coloneqq \mathbb{E}\left[Y_1 - Y_0 \left| X = x, U = u, S_0 = 0, S_1 = 1\right.\right]$ is the MTE on the observable outcome Y for the sub-population who is observed only when treated, $\Delta_S(x,u) \coloneqq \mathbb{E}\left[S_1 - S_0 \left| X = x, U = u\right.\right] = m_1^S(x,u) - m_0^S(x,u)$ is the MTE on the selection indicator, and $m_1^S(x,u) \coloneqq \mathbb{E}\left[S_1 \left| X = x, U = u\right.\right]$ is the MTR function associated to the counterfactual variable S_1. In this section, I also assume that $m_1^Y(x,u)$ and $\Delta_S(x,u)$ are point-identified, postponing the discussion about their identification to sections 4 and 5.

Although point-identification of $\mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 1, S_1 = 1\right.\right]$ is not possible, I can find identifiable bounds for it.[6]

Proposition 10 *Suppose that $m_0^Y(x,u)$, $m_1^Y(x,u)$, $m_0^S(x,u)$ and $\Delta_S(x,u)$ are point-identified.*

Under Assumptions 1-6, 7.1 and 8, $\mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 1, S_1 = 1\right.\right]$ *must satisfy*

$$\underline{y}^* \leq \mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 1, S_1 = 1\right.\right] \leq \frac{m_1^Y(x,u) - \underline{y}^* \cdot \Delta_S(x,u)}{m_0^S(x,u)}. \tag{6}$$

Under Assumptions 1-6, 7.2 and 8, $\mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 1, S_1 = 1\right.\right]$ *must satisfy*

$$\frac{m_1^Y(x,u) - \overline{y}^* \cdot \Delta_S(x,u)}{m_0^S(x,u)} \leq \mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 1, S_1 = 1\right.\right] \leq \overline{y}^*. \tag{7}$$

Under Assumptions 1-6, 7.3 (sub-case (a) or (b)) and 8, $\mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 1, S_1 = 1\right.\right]$ *must satisfy*

$$\begin{aligned}\frac{m_1^Y(x,u) - \overline{y}^* \cdot \Delta_S(x,u)}{m_0^S(x,u)} &\leq \mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 1, S_1 = 1\right.\right] \\ &\leq \frac{m_1^Y(x,u) - \underline{y}^* \cdot \Delta_S(x,u)}{m_0^S(x,u)}.\end{aligned} \tag{8}$$

[5] Appendix A.2 contains a proof of this claim.
[6] Appendix A.3 contains a proof of this proposition.

There is a important remark to be made about the bounds of proposition 10. Note that, even when the support is bounded in only one direction (assumptions 7.1 and 7.2), it is possible to derive lower and upper bounds on $\mathbb{E}\left[Y_1^* \,|X = x, U = u, S_0 = 1, S_1 = 1\right]$.

At this point, it is also important to understand the determinants of the width of those bounds. First, if there is no sample selection problem at all ($\mathbb{P}\left[S_0 = 1, S_1 = 1 \,|X = x, U = u\right] = 1$), then $m_0^S\left(x, u\right) = 1$, $\Delta_S\left(x, u\right) = 0$, implying tighter bounds in equations (6) and (7) and point-identification in equation (8). Second and most importantly, if there is no problem of differential sample selection with respect to treatment status ($\mathbb{P}\left[S_0 = 0, S_1 = 1 \,|X = x, U = u\right] = 0$), then $\Delta_S\left(x, u\right) = 0$, once more implying tighter bounds in equations (6) and (7) and point-identification in equation (8). Both cases are theoretically uninteresting and ruled out by Assumption 8.

Finally, combining equations (3) and (4) and proposition 10, I can partially identify the target parameter $\Delta_{Y^*}^{OO}\left(x, u\right)$:

Corollary 11 *Suppose that $m_0^Y\left(x, u\right)$, $m_1^Y\left(x, u\right)$, $m_0^S\left(x, u\right)$ and $\Delta_S\left(x, u\right)$ are point-identified.*

Under Assumptions 1-6, 7.1 and 8, the bounds on $\Delta_{Y^*}^{OO}\left(x, u\right)$ are given by

$$\Delta_{Y^*}^{OO}\left(x, u\right) \geq \underline{y}^* - \frac{m_0^Y\left(x, u\right)}{m_0^S\left(x, u\right)} =: \underline{\Delta_{Y^*}^{OO}}\left(x, u\right) \tag{9}$$

and that

$$\Delta_{Y^*}^{OO}\left(x, u\right) \leq \frac{m_1^Y\left(x, u\right) - \underline{y}^* \cdot \Delta_S\left(x, u\right)}{m_0^S\left(x, u\right)} - \frac{m_0^Y\left(x, u\right)}{m_0^S\left(x, u\right)} =: \overline{\Delta_{Y^*}^{OO}}\left(x, u\right). \tag{10}$$

Under Assumptions 1-6, 7.2 and 8, the bounds on $\Delta_{Y^*}^{OO}\left(x, u\right)$ are given by

$$\Delta_{Y^*}^{OO}\left(x, u\right) \geq \frac{m_1^Y\left(x, u\right) - \overline{y}^* \cdot \Delta_S\left(x, u\right)}{m_0^S\left(x, u\right)} - \frac{m_0^Y\left(x, u\right)}{m_0^S\left(x, u\right)} =: \underline{\Delta_{Y^*}^{OO}}\left(x, u\right) \tag{11}$$

and that

$$\Delta_{Y^*}^{OO}\left(x, u\right) \leq \overline{y}^* - \frac{m_0^Y\left(x, u\right)}{m_0^S\left(x, u\right)} =: \overline{\Delta_{Y^*}^{OO}}\left(x, u\right). \tag{12}$$

Under Assumptions 1-6, 7.3 (sub-case (a) or (b)) and 8, the bounds on $\Delta_{Y^*}^{OO}\left(x, u\right)$ are *given by*

$$\Delta_{Y^*}^{OO}(x,u) \geq \max\left\{\frac{m_1^Y(x,u) - \overline{y}^* \cdot \Delta_S(x,u)}{m_0^S(x,u)}, \underline{y}^*\right\} - \frac{m_0^Y(x,u)}{m_0^S(x,u)} =: \underline{\Delta_{Y^*}^{OO}}(x,u) \tag{13}$$

and that

$$\Delta_{Y^*}^{OO}(x,u) \leq \min\left\{\frac{m_1^Y(x,u) - \underline{y}^* \cdot \Delta_S(x,u)}{m_0^S(x,u)}, \overline{y}^*\right\} - \frac{m_0^Y(x,u)}{m_0^S(x,u)} =: \overline{\Delta_{Y^*}^{OO}}(x,u). \tag{14}$$

Most importantly, I can show that[7]:

Proposition 12 *Suppose that the functions m_0^Y, m_1^Y, m_0^S and Δ_S are point-identified at every pair $(x,u) \in \mathcal{X} \times [0,1]$. Under Assumptions 1-6, 7 (sub-cases 1, 2, 3(a) or 3(b)) and 8, the bounds $\underline{\Delta_{Y^*}^{OO}}$ and $\overline{\Delta_{Y^*}^{OO}}$, given by corollary 11, are point-wise sharp, i.e., for any $\overline{u} \in [0,1]$, $\overline{x} \in \mathcal{X}$ and $\delta(\overline{x},\overline{u}) \in \left(\underline{\Delta_{Y^*}^{OO}}(\overline{x},\overline{u}), \overline{\Delta_{Y^*}^{OO}}(\overline{x},\overline{u})\right)$, there exist random variables $\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)$ such that*

$$\Delta_{\tilde{Y}^*}^{OO}(\overline{x},\overline{u}) := \mathbb{E}\left[\tilde{Y}_1^* - \tilde{Y}_0^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right] = \delta(\overline{x},\overline{u}), \tag{15}$$

$$\mathbb{P}\left[\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{V}\right) \in \mathcal{Y}^* \times \mathcal{Y}^* \times [0,1] \middle| X = \overline{x}, \tilde{U} = u\right] = 1 \text{ for any } u \in [0,1], \tag{16}$$

and

$$F_{\tilde{Y},\tilde{D},\tilde{S},Z,X}(y,d,s,z,\overline{x}) = F_{Y,D,S,Z,X}(y,d,s,z,\overline{x}) \tag{17}$$

for any $(y,d,s,z) \in \mathbb{R}^4$, where $\tilde{D} := \mathbf{1}\left\{P(X,Z) \geq \tilde{U}\right\}$, $\tilde{S}_0 = \mathbf{1}\left\{Q(0,X) \geq \tilde{V}\right\}$, $\tilde{S}_1 = \mathbf{1}\left\{Q(1,X) \geq \tilde{V}\right\}$, $\tilde{Y}_0 = \tilde{S}_0 \cdot \tilde{Y}_0^*$, $\tilde{Y}_1 = \tilde{S}_1 \cdot \tilde{Y}_1^*$ and $\tilde{Y} = \tilde{D} \cdot \tilde{Y}_1 + \left(1 - \tilde{D}\right) \cdot \tilde{Y}_0$.

Intuitively, proposition 12 says that, for any $\delta(\overline{x},\overline{u}) \in \left(\underline{\Delta_{Y^*}^{OO}}(\overline{x},\overline{u}), \overline{\Delta_{Y^*}^{OO}}(\overline{x},\overline{u})\right)$, it is possible to create candidate random variables $\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)$ that generate the candidate marginal treatment effect $\delta(\overline{x},\overline{u})$ (equation (15)), satisfy the bounded support condition — a restriction imposed by my model (Assumption 7) and summarized in equation (16) — and

[7] Appendix A.4 contains the proof of this proposition. Note that, if the functions m_0^Y, m_1^Y, m_0^S and Δ_S are point-identified only in a subset of the unit interval, then point-wise sharpness holds only in that subset.

generate the same distribution of the observable variables — a restriction imposed by the data and summarized in equation (17). In other words, the data and the model in section 2 do not generate enough restrictions to refute that the true target parameter $\Delta^{OO}_{Y^*}(\overline{x}, \overline{u})$ is equal to the candidate target parameter $\delta(\overline{x}, \overline{u})$.

Moreover, the bounded support condition (Assumption 7) is partially necessary to the existence of bounds on the target parameter $\Delta^{OO}_{Y^*}(\overline{x}, \overline{u})$. When the support is unbounded in both directions (i.e., $\underline{y}^* = -\infty$ and $\overline{y}^* = +\infty$), then it is impossible to derive bounds on the target parameter $\Delta^{OO}_{Y^*}(\overline{x}, \overline{u})$ without any extra assumption. Proposition 13 formalizes this last statement.[8]

Proposition 13 *Suppose that the functions m^Y_0, m^Y_1, m^S_0 and Δ_S are point-identified at every pair $(x, u) \in \mathcal{X} \times [0, 1]$. Impose assumptions 1-6 and 8. If $\mathcal{Y}^* = \mathbb{R}$, then, for any $\overline{u} \in [0, 1]$, $\overline{x} \in \mathcal{X}$ and $\delta(\overline{x}, \overline{u}) \in \mathbb{R}$, there exist random variables $\left(\tilde{Y}^*_0, \tilde{Y}^*_1, \tilde{U}, \tilde{V}\right)$ such that*

$$\Delta^{OO}_{\tilde{Y}^*}(\overline{x}, \overline{u}) := \mathbb{E}\left[\tilde{Y}^*_1 - \tilde{Y}^*_0 \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right] = \delta(\overline{x}, \overline{u}), \tag{18}$$

$$\mathbb{P}\left[\left(\tilde{Y}^*_0, \tilde{Y}^*_1, \tilde{V}\right) \in \mathcal{Y}^* \times \mathcal{Y}^* \times [0, 1] \middle| X = \overline{x}, \tilde{U} = u\right] = 1 \text{ for any } u \in [0, 1], \tag{19}$$

and

$$F_{\tilde{Y}, \tilde{D}, \tilde{S}, Z, X}(y, d, s, z, \overline{x}) = F_{Y, D, S, Z, X}(y, d, s, z, \overline{x}) \tag{20}$$

for any $(y, d, s, z) \in \mathbb{R}^4$, where $\tilde{D} := \mathbf{1}\left\{P(X, Z) \geq \tilde{U}\right\}$, $\tilde{S}_0 = \mathbf{1}\left\{Q(0, X) \geq \tilde{V}\right\}$, $\tilde{S}_1 = \mathbf{1}\left\{Q(1, X) \geq \tilde{V}\right\}$, $\tilde{Y}_0 = \tilde{S}_0 \cdot \tilde{Y}^*_0$, $\tilde{Y}_1 = \tilde{S}_1 \cdot \tilde{Y}^*_1$ and $\tilde{Y} = \tilde{D} \cdot \tilde{Y}_1 + \left(1 - \tilde{D}\right) \cdot \tilde{Y}_0$.

In other words, when the support of the potential outcome is the entire real line, the data and the model in section 2 do not generate enough restrictions to refute that the true target parameter $\Delta^{OO}_{Y^*}(\overline{x}, \overline{u})$ is equal to an arbitrarily large effect in magnitude. This impossibility result is interesting in light of the previous literature about partial identification of treatment effects with sample selection. In the case of the ITT^{OO} (Lee (2009)) and the $LATE^{OO}$ (Chen & Flores (2015)), it is possible to construct informative bounds even when the support of the

[8] Appendix A.5 contains the proof of this proposition.

potential outcome is the entire real line. However, when focusing on a specific point of the MTE^{OO} function, it is impossible to construct informative bounds when $\mathcal{Y}^* = \mathbb{R}$ due to the local nature of the target parameter.

There is one important remark about the results I have just derived. Note that propositions 12 and 13 do not impose any smoothness condition on the joint distribution of $(Y_0^*, Y_1^*, U, V, Z, X)$. In particular, the conditional cumulative distribution functions $F_{V|X,U}$, $F_{Y_0^*|X,U,V}$ and $F_{Y_1^*|X,U,V}$ are allowed to be discontinuous functions of U at the point $\overline{u}$. Appendix G states and proves a sharpness result similar to proposition 12 and an impossibility result similar to proposition 13 when $F_{V|X,U}$, $F_{Y_0^*|X,U,V}$ and $F_{Y_1^*|X,U,V}$ must be continuous functions of U.

3.2 Partial Identification with an Extra Mean Dominance Assumption

Here, I use the Mean Dominance Assumption 9 to tighten the bounds on the target parameter $\Delta_{Y^*}^{OO}$ (equation (3)) given by corollary 11. Note that Assumption 9 implies that $\Delta_{Y}^{NO}(x,u) \leq \dfrac{m_1^Y(x,u)}{m_1^S(x,u)} \leq \mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 1, S_1 = 1\right.\right]$ by equations (A.4) and (A.5). As a consequence, by following the same steps of the proof of corollary 11, I can derive:

Corollary 14 *Fix $u \in [0,1]$ and $x \in \mathcal{X}$ arbitrarily. Suppose that the $m_0^Y(x,u)$, $m_1^Y(x,u)$, $m_0^S(x,u)$ and $\Delta_S(x,u)$ are point-identified.*

Under Assumptions 1-6, 7.1, 8 and 9, $\Delta_{Y^*}^{OO}(x,u)$ must satisfy

$$\Delta_{Y^*}^{OO}(x,u) \geq \frac{m_1^Y(x,u)}{m_1^S(x,u)} - \frac{m_0^Y(x,u)}{m_0^S(x,u)} =: \underline{\Delta_{Y^*}^{OO}}(x,u) \tag{21}$$

and that

$$\Delta_{Y^*}^{OO}(x,u) \leq \frac{m_1^Y(x,u) - \underline{y}^* \cdot \Delta_S(x,u)}{m_0^S(x,u)} - \frac{m_0^Y(x,u)}{m_0^S(x,u)} =: \overline{\Delta_{Y^*}^{OO}}(x,u). \tag{22}$$

Under Assumptions 1-6, 7.2, 8 and 9, $\Delta_{Y^*}^{OO}(x,u)$ must satisfy

$$\Delta_{Y^*}^{OO}(x,u) \geq \frac{m_1^Y(x,u)}{m_1^S(x,u)} - \frac{m_0^Y(x,u)}{m_0^S(x,u)} =: \underline{\Delta_{Y^*}^{OO}}(x,u) \tag{23}$$

and that

$$\Delta_{Y^*}^{OO}(x,u) \leq \overline{y}^* - \frac{m_0^Y(x,u)}{m_0^S(x,u)} =: \overline{\Delta_{Y^*}^{OO}}(x,u). \tag{24}$$

Under Assumptions 1-6, 7.3 (sub-case (a) or (b)), 8 and 9, $\Delta_{Y^*}^{OO}(x,u)$ *must satisfy*

$$\Delta_{Y^*}^{OO}(x,u) \geq \frac{m_1^Y(x,u)}{m_1^S(x,u)} - \frac{m_0^Y(x,u)}{m_0^S(x,u)} =: \underline{\Delta_{Y^*}^{OO}}(x,u) \tag{25}$$

and that

$$\Delta_{Y^*}^{OO}(x,u) \leq \min\left\{\frac{m_1^Y(x,u) - \underline{y}^* \cdot \Delta_S(x,u)}{m_0^S(x,u)}, \overline{y}^*\right\} - \frac{m_0^Y(x,u)}{m_0^S(x,u)} =: \overline{\Delta_{Y^*}^{OO}}(x,u). \tag{26}$$

When $\mathcal{Y}^* = \mathbb{R}$ *and Assumptions 1-6, 8 and 9 hold,* $\Delta_{Y^*}^{OO}(x,u)$ *must satisfy*

$$\Delta_{Y^*}^{OO}(x,u) \geq \frac{m_1^Y(x,u)}{m_1^S(x,u)} - \frac{m_0^Y(x,u)}{m_0^S(x,u)} =: \underline{\Delta_{Y^*}^{OO}}(x,u) \tag{27}$$

and that

$$\Delta_{Y^*}^{OO}(x,u) \leq \infty =: \overline{\Delta_{Y^*}^{OO}}(x,u). \tag{28}$$

Notice that, under Mean Dominance Assumption 9, I can increase the lower bounds proposed in corollary 11 under Assumption 7 and provide an informative lower bound even when the support of the outcome of interest is the entire real line, a result in stark contrast with proposition 13. These improvements clearly show the identifying power of the Mean Dominance Assumption 9.

As in subsection 3.1, I assume that $m_0^Y(x,u)$, $m_1^Y(x,u)$, $m_0^S(x,u)$, $m_1^S(x,u)$, and $\Delta_S(x,u)$ are point-identified, postponing the discussion about their identification to sections 4 and 5.

Now, using the above corollary, I can combine the sharpness and the impossibility results of subsection 3.1 in one single proposition[9]:

Proposition 15 *Suppose that the functions* m_0^Y, m_1^Y, m_0^S, m_1^S *and* Δ_S *are point-identified at every pair* $(x,u) \in \mathcal{X} \times [0,1]$. *Under Assumptions 1-6, 8 and 9, the bounds* $\underline{\Delta_{Y^*}^{OO}}$ *and* $\overline{\Delta_{Y^*}^{OO}}$,

[9] Appendix A.6 contains a proof of this proposition.

given by corollary 14, are point-wise sharp, i.e., for any $\overline{u} \in [0,1]$, $\overline{x} \in \mathcal{X}$ *and* $\delta(\overline{x},\overline{u}) \in \left(\underline{\Delta_{Y^*}^{OO}}(\overline{x},\overline{u}), \overline{\Delta_{Y^*}^{OO}}(\overline{x},\overline{u})\right)$, *there exist random variables* $\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)$ *such that*

$$\Delta_{\tilde{Y}^*}^{OO}(\overline{x},\overline{u}) \coloneqq \mathbb{E}\left[\tilde{Y}_1^* - \tilde{Y}_0^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right] = \delta(\overline{x},\overline{u}), \tag{29}$$

$$\mathbb{P}\left[\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{V}\right) \in \mathcal{Y}^* \times \mathcal{Y}^* \times [0,1] \middle| X = \overline{x}, \tilde{U} = u\right] = 1 \text{ for any } u \in [0,1], \tag{30}$$

$$\mathbb{E}\left[\tilde{Y}_1^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right] \geq \mathbb{E}\left[\tilde{Y}_1^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 0, \tilde{S}_1 = 1\right], \tag{31}$$

and

$$F_{\tilde{Y},\tilde{D},\tilde{S},Z,X}(y,d,s,z,\overline{x}) = F_{Y,D,S,Z,X}(y,d,s,z,\overline{x}) \tag{32}$$

for any $(y,d,s,z) \in \mathbb{R}^4$, *where* $\tilde{D} \coloneqq \mathbf{1}\left\{P(X,Z) \geq \tilde{U}\right\}$, $\tilde{S}_0 = \mathbf{1}\left\{Q(0,X) \geq \tilde{V}\right\}$, $\tilde{S}_1 = \mathbf{1}\left\{Q(1,X) \geq \tilde{V}\right\}$, $\tilde{Y}_0 = \tilde{S}_0 \cdot \tilde{Y}_0^*$, $\tilde{Y}_1 = \tilde{S}_1 \cdot \tilde{Y}_1^*$ *and* $\tilde{Y} = \tilde{D} \cdot \tilde{Y}_1 + \left(1 - \tilde{D}\right) \cdot \tilde{Y}_0$.

Note that, in addition to all the restriction imposed by proposition 12, the candidate random variables $\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)$ must also satisfy an extra model restriction (equation (31)) associated to the Mean Dominance Assumption 9.

3.3 Empirical Relevance of Bounds on the MTE$^{\text{OO}}$ of Interest

Now, it is important to discuss the empirical relevance of partially identifying the MTE^{OO} of interest. First, bounds on the MTE^{OO} can illuminate the heterogeneity of the treatment effect, allowing the researcher to understand who benefits and who loses with a specific treatment. This is important because common parameters (e.g.: ATE^{OO}, ATT^{OO}, ATU^{OO}, $LATE^{OO}$) can be positive even when most people lose with a policy if the few winners have very large gains. Moreover, knowing, even partially, the MTE^{OO} function can be useful to optimally design policies that provides incentives to agents to take some treatment. Second, I can use the MTE^{OO} bounds to partially identify any treatment effect that is described as a weighted integral of $\Delta_{Y^*}^{OO}(x,u)$ because

$$\int_0^1 \left(\underline{\Delta_{Y^*}^{OO}}(x,u)\right) \cdot \omega(x,u) \, \mathrm{d}u \leq \int_0^1 \Delta_{Y^*}^{OO}(x,u) \cdot \omega(x,u) \, \mathrm{d}u$$

$$\leq \int_0^1 \left(\overline{\Delta_{Y^*}^{OO}}(x,u)\right) \cdot \omega(x,u)\, \mathrm{d}u, \tag{33}$$

where $\omega(x,\cdot)$ is a known or identifiable weighting function. Even though such bounds may not be sharp for any specific parameter, they are a general and off-the-shelf solution to many empirical problems. As a consequence of this trade-off, I recommend the applied researcher to use a specialized tool if he or she is interested in a parameter that already has specific bounds for it (e.g., ITT by Lee (2009) and LATE by Chen & Flores (2015)). However, I suggest the applied research to easily compute a weighted integral of point-wise sharp bounds on the MTE of interest if he or she is interested in parameters without specialized bounds (e.g., ATE, ATT and ATU in the case with imperfect compliance). In other words, facing a trade-off between empirical flexibility and sharpness, the partial identification tool proposed in this paper focus on empirical flexibility while still ensuring point-wise sharpness of the bounds on the MTE of interest.

Tables 1 and 2 show some of the treatment effect parameters that can be partially identified using inequality (33). More examples are given by Heckman et al. (2006, Tables 1A and 1B) and Mogstad et al. (2018, Table 1).

Table 1: Treatment Effects as Weighted Integrals of the Marginal Treatment Effect

$ATE^{OO} = \mathbb{E}\left[Y_1^* - Y_0^* \mid S_0 = 1, S_1 = 1\right] = \int_0^1 \Delta_{Y^*}^{OO}(u)\, \mathrm{d}u$
$ATT^{OO} = \mathbb{E}\left[Y_1^* - Y_0^* \mid D = 1, S_0 = 1, S_1 = 1\right] = \int_0^1 \Delta_{Y^*}^{OO}(u) \cdot \omega_{ATT}(u)\, \mathrm{d}u$
$ATU^{OO} = \mathbb{E}\left[Y_1^* - Y_0^* \mid D = 0, S_0 = 1, S_1 = 1\right] = \int_0^1 \Delta_{Y^*}^{OO}(u) \cdot \omega_{ATU}(u)\, \mathrm{d}u$
$LATE^{OO}(\underline{u}, \overline{u}) = \mathbb{E}\left[Y_1^* - Y_0^* \mid U \in [\underline{u}, \overline{u}], S_0 = 1, S_1 = 1\right] = \int_0^1 \Delta_{Y^*}^{OO}(u) \cdot \omega_{LATE}(u)\, \mathrm{d}u$

Source: Heckman et al. (2006) and Mogstad et al. (2018). Note: Conditioning on X is kept implicit in this table for brevity.

Table 2: Weights

$\omega_{ATT}(x,u) = \dfrac{\int_u^1 f_{P(W)\vert X}(p\,\vert x)\,\mathrm{d}p}{\mathbb{E}[P(W)\,\vert X=x]}$
$\omega_{ATU}(x,u) == \dfrac{\int_0^u f_{P(W)\vert X}(p\,\vert x)\,\mathrm{d}p}{1-\mathbb{E}[P(W)\,\vert X=x]}$
$\omega_{LATE}(x,u) = \dfrac{\mathbf{1}\{u \in [\underline{u},\overline{u}]\}}{\overline{u}-\underline{u}}$

Source: Heckman et al. (2006) and Mogstad et al. (2018).

4 Partial identification when the support of the propensity score is an interval

Here, I fix $x \in \mathcal{X}$ and impose that the support of the propensity score, defined by $\mathcal{P}_x \coloneqq \{P(x,z) : z \in \mathcal{Z}\}$, is an interval[10]. Then, under assumptions 1-5, the MTR functions associated to any variable $A \in \{Y,S\}$ are point-identified by[11]:

$$m_0^A(x,p) = \mathbb{E}[A\,\vert X=x, P(W)=p, D=0] - \frac{\partial \mathbb{E}[A\,\vert X=x, P(W)=p, D=0]}{\partial p} \cdot (1-p)\,, \tag{34}$$

and

$$m_1^A(x,p) = \mathbb{E}[A\,\vert X=x, P(W)=p, D=1] + \frac{\partial \mathbb{E}[A\,\vert X=x, P(W)=p, D=1]}{\partial p} \cdot p \tag{35}$$

for any $p \in \mathcal{P}_x$.

Finally, the point-wise sharp bounds on $\Delta_{Y^*}^{OO}(x,p)$ are point-identified by combining equations (34) and (35), the fact that $\Delta_S(x,p) = m_1^S(x,p) - m_0^S(x,p)$, and Corollaries 11 or 14.

[10] $\mathcal{P}_x$ as an interval may be achieved by a continuous instrument Z or by the existence of independent covariates (Carneiro et al. 2011).

[11] Appendix A.7 contains a proof of this claim based on the Local Instrumental Variable (LIV) approach described by Heckman & Vytlacil (2005).

5 Partial identification when the support of the propensity score is discrete

When the support of the propensity score is not an interval, I cannot point-identify $m_0^Y(x,u)$, $m_1^Y(x,u)$, $m_0^S(x,u)$, $m_1^S(x,u)$, and $\Delta_S(x,u)$ without extra assumptions, implying that I cannot identify the bounds on $\Delta_{Y^*}^{OO}(x,u)$ given by Corollaries 11 or 14. There are two solutions for such lack of identification: I can non-parametrically bound those four objects (Mogstad et al. (2018)) or I can impose flexible parametric assumptions (Brinch et al. (2017)) to point-identify them. While the first approach is discussed in subsection 5.1, the second one is detailed in subsection 5.2.

5.1 Non-parametric outer set around the MTE$^{\mathbf{OO}}$ of interest

For any $u \in [0,1]$ and $x \in \mathcal{X}$, I can bound $m_0^Y(x,u)$, $m_1^Y(x,u)$, $m_0^S(x,u)$, $m_1^S(x,u)$, and $\Delta_S(x,u)$ using the machinery proposed by Mogstad et al. (2018). To do so, fix $A \in \{Y,S\}$ and $d \in \{0,1\}$ and define the pair of functions $m^A := \left(m_0^A, m_1^A\right)$ and the set of admissible MTR functions $\mathcal{M}^A \ni m^A$. Furthermore, fix $(x,u) \in \mathcal{X} \times [0,1]$ and define the functions $\Gamma_1^*\colon \mathcal{M}^Y \to \mathbb{R}$, $\Gamma_2^*\colon \mathcal{M}^Y \to \mathbb{R}$, $\Gamma_3^*\colon \mathcal{M}^S \to \mathbb{R}$, $\Gamma_4^*\colon \mathcal{M}^S \to \mathbb{R}$ and $\Gamma_5^*\colon \mathcal{M}^S \to \mathbb{R}$ as:

$$
\begin{aligned}
\Gamma_1^*\left(\tilde{m}^Y\right) &= \tilde{m}_1^Y(x,u) + 0 \cdot \tilde{m}_0^Y(x,u) \\
\Gamma_2^*\left(\tilde{m}^Y\right) &= 0 \cdot \tilde{m}_1^Y(x,u) + \tilde{m}_0^Y(x,u) \\
\Gamma_3^*\left(\tilde{m}^S\right) &= 0 \cdot \tilde{m}_1^S(x,u) + \tilde{m}_0^S(x,u) \\
\Gamma_4^*\left(\tilde{m}^S\right) &= \tilde{m}_1^S(x,u) + 0 \cdot \tilde{m}_0^S(x,u) \\
\Gamma_5^*\left(\tilde{m}^S\right) &= \tilde{m}_1^S(x,u) - \tilde{m}_0^S(x,u),
\end{aligned}
$$

and observe that $\Gamma_1^*\left(m^Y\right) = m_1^Y(x,u)$, $\Gamma_2^*\left(m^Y\right) = m_0^Y(x,u)$, $\Gamma_3^*\left(m^S\right) = m_0^S(x,u)$, $\Gamma_4^*\left(m^S\right) = m_1^S(x,u)$, and $\Gamma_5^*\left(m^S\right) = \Delta_S(x,u)$. Moreover, define, for each $A \in \{Y,S\}$, $\mathcal{G}_A$ to be a collection of known or identified measurable functions $g_A\colon \{0,1\} \times \mathcal{X} \times \mathcal{Z} \to \mathbb{R}$ whose second moment is finite. For each IV-like specification $g_A \in \mathcal{G}_A$, define also $\beta_{g_A} := \mathbb{E}\left[g_A(D,Z)\,A\,|X=x\right]$.

According to proposition 1 by Mogstad et al. (2018), the function $\Gamma_{g_A}\colon \mathcal{M}^A \to \mathbb{R}$, defined as

$$\begin{aligned}\Gamma_{g_A}\left(\tilde{m}^A\right) = \mathbb{E}\left[\int_0^1 \tilde{m}_0^A\left(X,u\right)\cdot g_A\left(0,Z\right)\cdot \mathbf{1}\left\{p\left(W\right)<u\right\}\,\mathrm{d}u\middle| X=x\right]\\ +\mathbb{E}\left[\int_0^1 \tilde{m}_1^A\left(X,u\right)\cdot g_A\left(1,Z\right)\cdot \mathbf{1}\left\{p\left(W\right)\geq u\right\}\,\mathrm{d}u\middle| X=x\right],\end{aligned}$$

satisfies $\Gamma_{g_A}\left(m^A\right) = \beta_{g_A}$. As a result, m^A must lie in the set $\mathcal{M}_{\mathcal{G}_A}$ of admissible functions that satisfy the restrictions imposed by the data through the IV-like specifications, where:

$$\mathcal{M}_{\mathcal{G}_A} \coloneqq \left\{\tilde{m}^A \in \mathcal{M}^A\colon \Gamma_{g_A}\left(\tilde{m}^A\right) = \beta_{g_A} \text{ for all } g_A \in \mathcal{G}_A\right\}.$$

Assuming that $\mathcal{M}^A$ is convex and $M_{\mathcal{G}_A} \neq \emptyset$ for every $A \in \{Y, S\}$, proposition 2 by Mogstad et al. (2018) ensures that:

$$\begin{aligned}
\inf_{\tilde{m}^Y\in\mathcal{M}_{\mathcal{G}_Y}} \Gamma_1^*\left(\tilde{m}^Y\right) &\eqqcolon \underline{m_1^Y\left(x,u\right)} \leq & m_1^Y\left(x,u\right) &\leq \overline{m_1^Y\left(x,u\right)} \coloneqq & \sup_{\tilde{m}^Y\in\mathcal{M}_{\mathcal{G}_Y}} \Gamma_3^*\left(\tilde{m}^Y\right)\\
\inf_{\tilde{m}^Y\in\mathcal{M}_{\mathcal{G}_Y}} \Gamma_2^*\left(\tilde{m}^Y\right) &\eqqcolon \underline{m_0^Y\left(x,u\right)} \leq & m_0^Y\left(x,u\right) &\leq \overline{m_0^Y\left(x,u\right)} \coloneqq & \sup_{\tilde{m}^Y\in\mathcal{M}_{\mathcal{G}_Y}} \Gamma_2^*\left(\tilde{m}^Y\right)\\
\inf_{\tilde{m}^S\in\mathcal{M}_{\mathcal{G}_S}} \Gamma_3^*\left(\tilde{m}^S\right) &\eqqcolon \underline{m_0^S\left(x,u\right)} \leq & m_0^S\left(x,u\right) &\leq \overline{m_0^S\left(x,u\right)} \coloneqq & \sup_{\tilde{m}^S\in\mathcal{M}_{\mathcal{G}_S}} \Gamma_3^*\left(\tilde{m}^S\right)\\
\inf_{\tilde{m}^S\in\mathcal{M}_{\mathcal{G}_S}} \Gamma_4^*\left(\tilde{m}^S\right) &\eqqcolon \underline{m_1^S\left(x,u\right)} \leq & m_1^S\left(x,u\right) &\leq \overline{m_1^S\left(x,u\right)} \coloneqq & \sup_{\tilde{m}^S\in\mathcal{M}_{\mathcal{G}_S}} \Gamma_4^*\left(\tilde{m}^S\right)\\
\inf_{\tilde{m}^S\in\mathcal{M}_{\mathcal{G}_S}} \Gamma_5^*\left(\tilde{m}^S\right) &\eqqcolon \underline{\Delta_S\left(x,u\right)} \leq & \Delta_S\left(x,u\right) &\leq \overline{\Delta_S\left(x,u\right)} \coloneqq & \sup_{\tilde{m}^S\in\mathcal{M}_{\mathcal{G}_S}} \Gamma_5^*\left(\tilde{m}^S\right)
\end{aligned} \tag{36}$$

As a consequence, I can combine Corollary 11 and inequalities (36) to provide a non-parametrically identified outer set around $\Delta_{Y^*}^{OO}\left(x,u\right)$:

Corollary 16 *Fix $u \in [0,1]$ and $x \in \mathcal{X}$ arbitrarily.*

Under assumptions 1-6, 7.1 and 8, the bounds of an outer set around $\Delta_{Y^*}^{OO}\left(x,u\right)$ are given by

$$\Delta_{Y^*}^{OO}\left(x,u\right) \geq \underline{y}^* - \frac{\overline{m_0^Y\left(x,u\right)}}{\underline{m_0^S\left(x,u\right)}}, \tag{37}$$

and that

$$\Delta_{Y^*}^{OO}\left(x,u\right) \leq \frac{\overline{m_1^Y\left(x,u\right)}}{\underline{m_0^S\left(x,u\right)}} - \frac{\underline{y}^*\cdot\underline{\underline{\Delta_S\left(x,u\right)}}}{\overline{m_0^S\left(x,u\right)}} - \frac{\underline{\underline{m_0^Y\left(x,u\right)}}}{m_0^S\left(x,u\right)}. \tag{38}$$

Under assumptions 1-6, 7.2 and 8, the bounds of an outer set around $\Delta_{Y^*}^{OO}(x,u)$ are given by

$$\Delta_{Y^*}^{OO}(x,u) \geq \frac{\underline{m_1^Y(x,u)}}{\overline{m_0^S(x,u)}} - \frac{\overline{y}^* \cdot \overline{\Delta_S(x,u)}}{\underline{m_0^S(x,u)}} - \frac{\overline{m_0^Y(x,u)}}{\underline{m_0^S(x,u)}}, \tag{39}$$

and that

$$\Delta_{Y^*}^{OO}(x,u) \leq \overline{y}^* - \frac{\underline{m_0^Y(x,u)}}{\overline{m_0^S(x,u)}}. \tag{40}$$

Under assumptions 1-6, 7.3 (sub-case (a) or (b)) and 8, the bounds of an outer set around $\Delta_{Y^*}^{OO}(x,u)$ are given by

$$\Delta_{Y^*}^{OO}(x,u) \geq \max\left\{\max\left\{\frac{\underline{m_1^Y(x,u)}}{\overline{m_0^S(x,u)}} - \frac{\overline{y}^* \cdot \overline{\Delta_S(x,u)}}{\underline{m_0^S(x,u)}}, \underline{y}^*\right\} - \frac{\overline{m_0^Y(x,u)}}{\underline{m_0^S(x,u)}}, \underline{y}^* - \overline{y}^*\right\}, \tag{41}$$

and that

$$\Delta_{Y^*}^{OO}(x,u) \leq \min\left\{\min\left\{\frac{\overline{m_1^Y(x,u)}}{\underline{m_0^S(x,u)}} - \frac{\underline{y}^* \cdot \underline{\Delta_S(x,u)}}{\overline{m_0^S(x,u)}}, \overline{y}^*\right\} - \frac{\underline{m_0^Y(x,u)}}{\overline{m_0^S(x,u)}}, \overline{y}^* - \underline{y}^*\right\}. \tag{42}$$

Note that I can obviously combine Corollary 14 and inequalities (36) to derive the bounds of an outer set around $\Delta_{Y^*}^{OO}(x,u)$ under the Mean Dominance Assumption 9. Moreover, I stress that the cost of non-parametric partial identification of $m_0^Y(x,u)$, $m_1^Y(x,u)$, $m_0^S(x,u)$, $m_1^S(x,u)$, and $\Delta_S(x,u)$ is losing the point-wise sharpness of the bounds around the target parameter $\Delta_{Y^*}^{OO}$. For that reason, Corollary 16 is stated in terms of bounds of an outer set around $\Delta_{Y^*}^{OO}(x,u)$, that contains the true target parameter $\Delta_{Y^*}^{OO}(x,u)$ by construction.

5.2 Parametric identification of the MTE$^{\mathbf{OO}}$ bounds

The fully non-parametric approach explained in subsection 5.1 may provide an uninformative outer set (e.g., equal to $\overline{y}^* - \underline{y}^*$ or $\underline{y}^* - \overline{y}^*$ when the support of the potential outcome is bounded). In such cases, parametric assumptions on the marginal treatment response function may buy a lot of identifying power. Although restrictive in principle, parametric assumptions may be flexible enough to provide credible bounds on $\Delta_{Y^*}^{OO}(x,u)$.

I fix $x \in \mathcal{X}$ and assume that the support of the propensity score $P(x,Z)$ is discrete and given by $\mathcal{P}_x = \{p_{x,1}, \ldots, p_{x,N}\}$ for some $N \in \mathbb{N}$. I could directly apply the identification strategy proposed by Brinch et al. (2017) by assuming that the MTR functions associated to Y and S are polynomial functions of U. However, this assumption is problematic for binary variables, such as the selection indicator S. For this reason, I make a small modification to the procedure created by Brinch et al. (2017): for $d \in \{0,1\}$ and $A \in \{Y, S\}$, the MTR function is given by

$$m_d^A(x,u) = M^A\left(u, \boldsymbol{\theta}_{x,d}^A\right) \tag{43}$$

for any $u \in [0,1]$, where $\Theta_x^A \subset \mathbb{R}^{2L}$ is a set of feasible parameters, $L \in \{1, \ldots, N\}$ is the number of parameters for each treatment group d, $\left(\boldsymbol{\theta}_{x,0}^A, \boldsymbol{\theta}_{x,1}^A\right) \in \Theta_x^A$ is a vector of pseudo-true unknown parameters, and $M^A \colon [0,1] \times \mathbb{R}^{2L} \to \mathbb{R}$ is a known function. For example, in the case of a binary variable, a reasonable choice of M^A is the Bernstein Polynomial $\left(M^A\left(u, \boldsymbol{\theta}_{x,d}^A\right) = \sum_{l=0}^{L-1} \theta_{x,d,l}^A \cdot \binom{L-1}{l} \cdot u^l \cdot (1-u)^{L-1-l}\right)$ with feasible set $\Theta_x^A = [0,1]^{2L}$. In the case of the selection indicator, the feasible set would be further restricted by Assumption 8 to $\Theta_x^A = \left\{\left(\tilde{\boldsymbol{\theta}}_{x,0}^A, \tilde{\boldsymbol{\theta}}_{x,1}^A\right) \in [0,1]^{2L} : \tilde{\boldsymbol{\theta}}_{x,1}^A \geq \tilde{\boldsymbol{\theta}}_{x,0}^A\right\}$. I stress that the only difference between the Bernstein polynomial model and the simple polynomial model proposed by Brinch et al. (2017) is that it is easier to impose feasibility restrictions on the former model.

Back to the parametric model given by equation (43), I define the parameters $\left(\boldsymbol{\theta}_{x,0}^A, \boldsymbol{\theta}_{x,1}^A\right)$ as pseudo-true parameters in the sense that the parametric model in equation (43) is an approximation to the true data generating process via the moments $\mathbb{E}\left[A \,|X = x, P(W) = p_n, D = d\right]$ for any $d \in \{0,1\}$ and $n \in \{1, \ldots, N\}$. Formally, I define

$$\begin{aligned}\left(\boldsymbol{\theta}_{x,0}^A, \boldsymbol{\theta}_{x,1}^A\right) \coloneqq \underset{\left(\tilde{\boldsymbol{\theta}}_{x,0}^A, \tilde{\boldsymbol{\theta}}_{x,1}^A\right) \in \Theta_x^A}{\operatorname{argmin}} \sum_{n=1}^{N} \Bigg\{ & \left(\mathbb{E}\left[A \,|X = x, P(W) = p_n, D = 0\right] - \frac{\int_{p_n}^1 M^A\left(u, \tilde{\boldsymbol{\theta}}_{x,0}^A\right) \mathrm{d}u}{1 - p_n}\right)^2 \\ & + \left(\mathbb{E}\left[A \,|X = x, P(W) = p_n, D = 1\right] - \frac{\int_0^{p_n} M^A\left(u, \tilde{\boldsymbol{\theta}}_{x,1}^A\right) \mathrm{d}u}{p_n}\right)^2 \Bigg\}.\end{aligned} \tag{44}$$

Note that, to estimate parameters $\left(\boldsymbol{\theta}_{x,0}^A, \boldsymbol{\theta}_{x,1}^A\right)$, I can simply use the sample analogue of equation (44), i.e., I only have to estimate a constrained OLS regression whose restrictions are given by the set Θ_x^A. If the model restrictions imposed through the set of feasible parameters Θ_x^A are valid and $L = N$, then my parametric model collapses to the model proposed by Brinch et al. (2017) and I find that[12], for any $p_n \in \mathcal{P}_x$,

$$\mathbb{E}\left[A\,|X = x, P\left(W\right) = p_n, D = 0\right] = \frac{\int_{p_n}^{1} M^A\left(u, \boldsymbol{\theta}_{x,0}^A\right)\, \mathrm{d}u}{1 - p_n} \tag{45}$$

$$\mathbb{E}\left[A\,|X = x, P\left(W\right) = p_n, D = 1\right] = \frac{\int_{0}^{p_n} M^A\left(u, \boldsymbol{\theta}_{x,1}^A\right)\, \mathrm{d}u}{p_n}. \tag{46}$$

I can, then, combine Corollary 11 and equations (43) and (44) to bound $\Delta_{Y^*}^{OO}\left(x, u\right)$:

Corollary 17 *Fix $u \in [0,1]$ and $x \in \mathcal{X}$ arbitrarily.*

Under assumptions 1-6, 7.1 and 8, the bounds on $\Delta_{Y^*}^{OO}\left(x, u\right)$ are given by

$$\Delta_{Y^*}^{OO}\left(x, u\right) \geq \underline{y}^* - \frac{M^Y\left(u, \boldsymbol{\theta}_{x,0}^Y\right)}{M^S\left(u, \boldsymbol{\theta}_{x,0}^S\right)}, \tag{47}$$

and that

$$\Delta_{Y^*}^{OO}\left(x, u\right) \leq \frac{M^Y\left(u, \boldsymbol{\theta}_{x,1}^Y\right) - \underline{y}^* \cdot \left[M^S\left(u, \boldsymbol{\theta}_{x,1}^S\right) - M^S\left(u, \boldsymbol{\theta}_{x,0}^S\right)\right]}{M^S\left(u, \boldsymbol{\theta}_{x,0}^S\right)} - \frac{M^Y\left(u, \boldsymbol{\theta}_{x,0}^Y\right)}{M^S\left(u, \boldsymbol{\theta}_{x,0}^S\right)}. \tag{48}$$

Under assumptions 1-6, 7.2 and 8, the bounds on $\Delta_{Y^*}^{OO}\left(x, u\right)$ are given by

$$\Delta_{Y^*}^{OO}\left(x, u\right) \geq \frac{M^Y\left(u, \boldsymbol{\theta}_{x,1}^Y\right) - \overline{y}^* \cdot \left[M^S\left(u, \boldsymbol{\theta}_{x,1}^S\right) - M^S\left(u, \boldsymbol{\theta}_{x,0}^S\right)\right]}{M^S\left(u, \boldsymbol{\theta}_{x,0}^S\right)} - \frac{M^Y\left(u, \boldsymbol{\theta}_{x,0}^Y\right)}{M^S\left(u, \boldsymbol{\theta}_{x,0}^S\right)}, \tag{49}$$

and that

$$\Delta_{Y^*}^{OO}\left(x, u\right) \leq \overline{y}^* - \frac{M^Y\left(u, \boldsymbol{\theta}_{x,0}^Y\right)}{M^S\left(u, \boldsymbol{\theta}_{x,0}^S\right)}. \tag{50}$$

Under assumptions 1-6, 7.3 (sub-case (a) or (b)) and 8, the bounds on $\Delta_{Y^*}^{OO}\left(x, u\right)$ are

[12] Appendix A.8 contains a proof of this claim.

given by

$$\Delta_{Y^*}^{OO}(x,u) \geq \max\left\{\frac{M^Y\left(u,\boldsymbol{\theta}_{x,1}^Y\right) - \overline{y}^* \cdot \left[M^S\left(u,\boldsymbol{\theta}_{x,1}^S\right) - M^S\left(u,\boldsymbol{\theta}_{x,0}^S\right)\right]}{M^S\left(u,\boldsymbol{\theta}_{x,0}^S\right)}, \underline{y}^*\right\} - \frac{M^Y\left(u,\boldsymbol{\theta}_{x,0}^Y\right)}{M^S\left(u,\boldsymbol{\theta}_{x,0}^S\right)}, \tag{51}$$

and that

$$\Delta_{Y^*}^{OO}(x,u) \leq \min\left\{\frac{M^Y\left(u,\boldsymbol{\theta}_{x,1}^Y\right) - \underline{y}^* \cdot \left[M^S\left(u,\boldsymbol{\theta}_{x,1}^S\right) - M^S\left(u,\boldsymbol{\theta}_{x,0}^S\right)\right]}{M^S\left(u,\boldsymbol{\theta}_{x,0}^S\right)}, \overline{y}^*\right\} - \frac{M^Y\left(u,\boldsymbol{\theta}_{x,0}^Y\right)}{M^S\left(u,\boldsymbol{\theta}_{x,0}^S\right)}. \tag{52}$$

Note that I can obviously combine Corollary 14 and equations (43) and (44) to bound $\Delta_{Y^*}^{OO}(x,u)$ under the Mean Dominance Assumption 9.

6 Empirical Application: Job Corps Training Program

Active Labor Market Programs are a common way to possibly fight unemployment and increase wages by providing public employment services, labor market training and subsidized employment. Given their economic importance, they were extensively studied in the literature: Heckman et al. (1999), Heckman & Smith (1999), Abadie et al. (2002) and van Ours (2004). In particular, the Job Corps Training Program (JCTP) received great academic attention (Schochet et al. (2001), Schochet et al. (2008), Flores-Lagunes et al. (2010), Flores et al. (2012), Blanco et al. (2013*a*), Blanco et al. (2013*b*), Chen & Flores (2015), Chen et al. (2017), Blanco et al. (2017)) due to its randomized evaluation funded by the U.S. Department of Labor in the mid 1990's.

For its social and academic importance, I focus on analyzing the Marginal Treatment Effect of the JCTP on wages for the always-employed sub-population (MTE^{OO}). This program provides free education and vocational training to individuals who are legal residents of the U.S., are between the ages of 16 and 24 and come from a low-income household (Schochet et al. (2001) and Lee (2009)). Besides receiving education and vocational training, the trainees reside in the Job Corps center, that offers meals and a small cash allowance.

In the mid 1990's, the U.S. Department of Labor hired Mathematica Policy Resarch, Inc., to evaluate the JCTP through a randomized experiment. According to Chen & Flores (2015), eligible people who applied to JCTP for the first time between November 1994 and December 1995 (80,833 applicants) were randomly assigned into a treatment group and a control group. People in the control group (5,977) were embargoed from the program for 3 years, while those in the treatment group (74,856) were allowed to enroll in JC. However, in this randomized control trial, there was non-compliance (selection into treatment) because some individuals in the treated group decided not to participate in the program and some individuals in the control group were able to attend the JCTP even though they were officially embargoed.

To evaluate the JCTP, I start by describing the dataset, providing summary statistics and, most importantly, formally testing the assumptions that the potential treatment status is monotone on the instrument (equation (1)) and that the potential employment (sample selection status) is positively monotone on the treatment (Assumption 8) using the test elaborated by Machado et al. (2018). I, then, estimate and discuss the marginal treatment responses and effects on employment and labor earnings using the parametric tool developed by Brinch et al. (2017). Finally, I estimate and discuss the bounds on the MTE^{OO} on wages without and with the mean dominance assumption (Assumption 9), given, respectively, by Corollaries 11 and 14.

6.1 Descriptive Statistics and the Monotonicity Assumptions

The National Job Corps Study (NJCS) sample contains 15,386 individuals — all 5,977 control group individuals and 9,409 randomly selected treatment group individuals. All of them were interviewed at random assignment and at 12, 30 and 48 months after random assignment. Following Lee (2009), I only keep individuals with non-missing values for weekly earnings and weekly hours worked for every week after randomization (9,145). Following Chen & Flores (2015), I also add to the dataset a dummy variable that is equal to one if the individual was ever enrolled in the JCTP during the 208 weeks after random assignment. As a consequence, I drop 51 observations with missing values for the enrollment variable. I stress that this variable is my treatment dummy (D), while random treatment assignment is my instrument (Z).

The dataset contains information about demographic covariates (sex, age, race, marriage, number of children, years of schooling, criminal behavior, personal income) and pre- and post-treatment labor market outcomes (employment and earnings). Following Chen & Flores (2015), hourly wages at week 208 are created by dividing weekly earnings by weekly hours worked at that week, implying that a missing wage is equivalent to zero weekly hours worked. I consider the person to be unemployed ($S = 0$) when the wage is missing and to be employed ($S = 1$) when the wage is non-missing. Differently from Lee (2009) and Chen & Flores (2015), who use log hourly wages as their main outcome variable, my outcome of interest (Y^*) is the level of the hourly wage because Assumption 7.1 requires that the support $\mathcal{Y}^*$ has a finite lower bound. As a consequence, the observable outcome Y is defined as hourly labor earnings. Finally, I use the NJCS design weights in my empirical analysis because some subpopulations were randomized with different, but known, probabilities (Schochet et al. (2001)).

Considering the results found by Flores-Lagunes et al. (2010), who focus on explaining the negative but insignificant effects on employment and labor earnings for the Hispanic subpopulation, I separately analyze two subsamples from the NJCS sample: the Non-Hispanics subsample and the Hispanics subsample. Table 3 shows descriptive statistics for both subsamples. Note that, as expected, the pre-treatment covariates are, on average, very similar between the groups defined by the random treatment assignment. Consequently, both subsamples maintains the balance of baseline variables. When comparing Non-Hispanics and Hispanics, I find numerically small differences with respect to the variables *female*, *never married*, *has children*, *ever arrested*, *has a job at baseline*, and *had a job*, suggesting that it is important to separately analyze those two groups.

Table 4 shows preliminary effects for the Non-Hispanic and the Hispanic subsamples. The first row shows that a large number of individuals did not comply to their treatment assignment. As is expected for any voluntary treatment, a large share of individuals (around 30% for both subsamples) decided not to take the treatment even though they were assigned to the treatment group. There are also some individuals (5% among Non-Hispanics and 3%

Table 3: Summary Statistics of Selected Baseline Variables

	Non-Hispanic Sample			Hispanic Sample		
	Z = 1	Z = 0	Diff.	Z = 1	Z = 0	Diff.
Female	.443	.454	-.011	.502	.473	.030
			(.011)			(.025)
Age at baseline	18.436	18.342	.095*	18.438	18.398	.040
			(.049)			(.109)
White	.318	.318	.000	—	—	—
			(.011)			
Black	.595	.592	.002	—	—	—
			(.011)			
Never married	.926	.924	.002	.875	.874	.001
			(.006)			(.017)
Has children	.186	.190	-.004	.201	.206	-.004
			(.009)			(.020)
Years of Schooling	10.137	10.115	.022	10.022	10.057	-.034
			(.036)			(.084)
Ever arrested	.255	.257	-.002	.216	.211	.005
			(.010)			(.021)
Personal Inc.: <3000	.787	.788	-.001	.789	.794	-.005
			(.010)			(.022)
Has a job at baseline	.204	.188	.016*	.170	.211	-.041**
			(.009)			(.020)
A year before baseline:						
Had a job	.642	.627	.015	.601	.630	-.029
			(.011)			(.025)
Months employed	3.652	3.513	.140	3.344	3.616	-.272
			(.098)			(.214)
Earnings	2899.41	2795.62	103.79	2956.38	2885.47	70.91
			(103.81)			(477.08)
Observations	4554	2977	Total: 7531	942	621	Total: 1563

Note: Z indicates random treatment assignment. Standard errors are in parenthesis. ***, ** and * denote that difference is statistically significant at the 1%, at 5% and 10% level, respectively. Estimation uses design weights.

among Hispanics) who attended the JCTP even though their were embargoed. Moreover, the instrument (treatment assignment) is clearly strong for both subsamples. When analyzing the treatment effects and similarly to the previous literature (e.g.: Schochet et al. (2008), Flores-Lagunes et al. (2010) and Chen & Flores (2015)), we find that the JCTP has a positive and significant effect on Non-Hispanics and a negative but insignificant effect on Hispanics.

Table 4: Preliminary Effects

	Non-Hispanic Sample			Hispanic Sample		
	Z = 1	Z = 0	Diff.	Z = 1	Z = 0	Diff.
Ever enrolled in JCTP	.737	.047	.689***	.747	.028	.719***
			(.008)			(.016)
ITT estimates						
Hours per week	28.06	25.54	2.52***	26.63	27.30	-.670
			(.60)			(1.28)
Earnings per week	230.24	194.72	35.52***	218.34	228.63	-1.29
			(5.49)			(12.68)
Employed	.613	.564	.049***	.605	.607	-.002
			(.011)			(.025)
LATE estimates						
Hours per week			3.66***			-.930
			(.880)			(1.78)
Earnings per week			51.52***			-14.31
			(8.00)			(17.64)
Employed			.071***			-.003
			(.016)			(.034)

Note: Z indicates random treatment assignment. Outcome variables are measured at week 208 after randomization. Standard errors are in parenthesis. ***, ** and * denote that difference is statistically significant at the 1%, at 5% and 10% level, respectively. Estimation uses design weights.

This last result, particularly with respect to the employment status, is paramountly important for my analysis. Similarly to Lee (2009) and Chen & Flores (2015), I assume that the effect of the treatment on employment (i.e., sample selection) is monotone and positive. However, a negative effect of JCTP on employment is evidence against this assumption as discussed by Flores-Lagunes et al. (2010) and Chen & Flores (2015). For this reason, it important to formally test it. To do so, I implement the procedure developed by Machado et al. (2018), that simultaneously tests instrument exogeneity (Assumption 1), monotonicity of treatment take-up on treatment assignment (equation (1)) and monotonicity of employment on the treatment (equation (2)). Their procedure also uses this last test as a gate-keeper to test that the effect of the treatment on employment is positive (Assumption (8)).

In a more detailed way, the test proposed by Machado et al. (2018) has three steps. In the first step, the null hypothesis is that the instrument is not exogenous, or treatment take-up is not monotone on treatment assignment, or employment is not monotone on treatment take-up. As a consequence, the alternative hypothesis is that Assumption 1 and equations (1) and (2) hold. In the second step, that is implemented only if the first step rejects its null hypothesis, the second null hypothesis is that the effect of the treatment on employment is non-positive. Consequently, its alternative hypothesis is that Assumptions 1 and 8 and equations (1) and (2) hold. Finally, in the third step, that is implemented only if the second step does not reject its null hypothesis, the third null hypothesis is that the effect of the treatment on employment is non-negative. Consequently, its alternative hypothesis is that, while Assumption 1 and equations (1) and (2) are valid, Assumption 8 holds in the opposite direction (see Assumption C.1).

Table 5 shows the results of the test described above. For the Non-Hispanics subsample, steps 1 and 2 reject their null hypotheses at the 1%-significance level, implying that Assumptions 1 and 8 and equations (1) and (2) are valid. Consequently, I can use Corollary 11 to bound the MTE^{OO} of the JCTP on wages for the Non-Hispanics subsample. For the Hispanics subsample, step 1 rejects its null hypothesis at the 1%-significance level, while neither step 2 nor step 3 reject their null hypotheses at the 10%-significance level. As a consequence, Assumption 1 and equations (1) and (2) are valid, but it seems that there is no effect of the treatment on employment, i.e., $S_1 = S_0$ for all individuals. With no differential sample selection for the Hispanic population, partial identification of the MTE of interest is trivial as discussed immediately after proposition 10. For this reason, I focus my empirical analysis on the Non-Hispanic subsample.

Table 5: Testing the Identification Assumptions

	Non-Hispanics Subsample				Hispanics Subsample			
	Estimated	Critical Value			Estimated	Critical Value		
	Test Statistic	10%	5%	1%	Test Statistic	10%	5%	1%
Step 1	.282	.034	.039	.043	.308	.044	.047	.050
Step 2	.070	.033	.036	.039	-.003	.032	.036	.038
Step 3	-.070	.033	.036	.039	.003	.032	.036	.038

Note: The alternative hypothesis of step 1 is that Assumption 1 and equations (1) and (2) are valid. The alternative hypothesis of step 2 is that Assumptions 1 and 8 and equations (1) and (2) are valid. The alternative hypothesis of step 3 is that Assumptions 1 and C.1 and equations (1) and (2) are valid. Critical values were computed using 10,000 bootstrap repetitions and are related to the 10%, 5% and 1% significance levels. Estimation uses design weights.

6.2 MTR and MTE on Employment and Labor Earnings: Non-Hispanics Sub-population

As a preliminary step to estimate the bounds on the MTE^{OO} of the JCTP on hourly wages for the Non-Hispanic subsample, I need to estimate the MTR functions on employment and hourly labor earnings, i.e., I need to estimate the functions m_0^S, m_1^S, m_0^Y, and m_1^Y. To do so, I use the procedure described in Subsection 5.2, that adapts the method developed by Brinch et al. (2017) to a constrained framework. Specifically, I model the MTR functions of Y and S using Bernstein polynomials with four parameters, i.e., $M^A\left(u, \boldsymbol{\theta}_d^A\right) = \theta_{d,0}^A \cdot (1-u) + \theta_{d,1}^A \cdot u$ for any $A \in \{Y, S\}$ and $d \in \{0,1\}$ with feasible sets $\Theta^Y = \mathbb{R}_+^4$ and $\Theta^S = \left\{\left(\boldsymbol{\theta}_0^S, \boldsymbol{\theta}_1^S\right) \in [0,1]^4 : \boldsymbol{\theta}_1^S \geq \boldsymbol{\theta}_0^S\right\}$. To estimate $\left(\boldsymbol{\theta}_0^A, \boldsymbol{\theta}_1^A\right)$. I run the following constrained OLS model:[13]

$$A = a_0^A \cdot (1-D) + b_0^A \cdot (1-D) \cdot P(Z) + a_1^A \cdot D + b_1^A \cdot D \cdot P(Z) + e, \tag{53}$$

where e is the error term, $\theta_{0,0}^A = a_0^A - b_0^A$, $\theta_{0,1}^A = a_0^A + b_0^A$, $\theta_{1,0}^A = a_1^A$, $\theta_{1,1}^A = a_1^A + 2 \cdot b_1^A$ and the constraints on $\left(a_0^A, b_0^A, a_1^A, b_1^A\right)$ are given by Θ^A.

Tabel 6 reports the point-estimates and confidence intervals of the parametric models for the MTR functions on employment and hourly labor earnings. Note that the feasibility constraint $\theta_{1,0}^S \geq \theta_{0,0}^S$ is binding even though Assumption 8 is valid according to the test

[13] Appendix A.9 connects the above OLS model to the minimization problem (44) when the instrument is binary and there are no covariates.

proposed by Machado et al. (2018). Moreover, for the upper bound of the 90%-confidence interval, the feasibility constraint $\theta_{1,0}^{S} \leq 1$ is also binding.

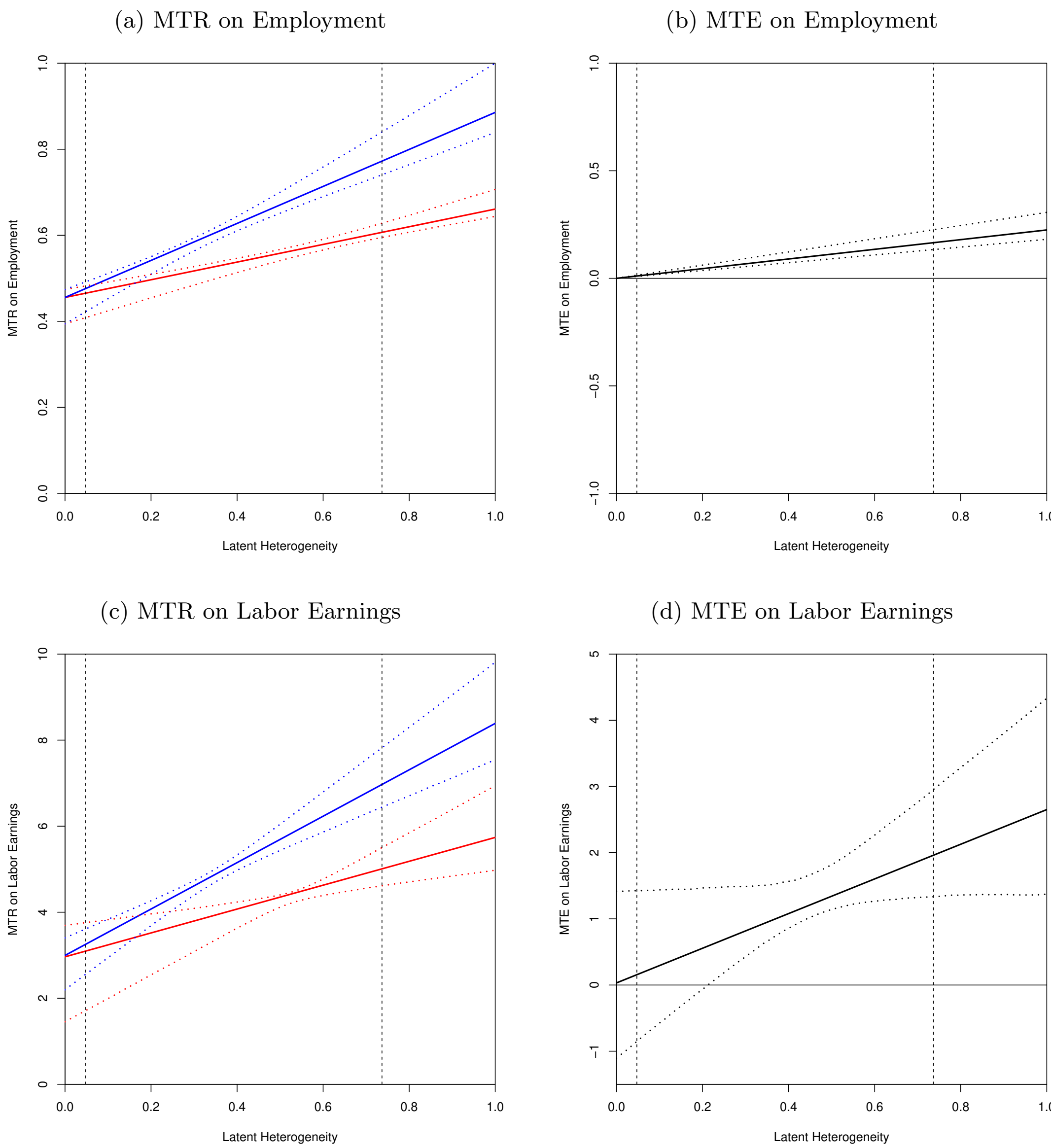
Table 6: Parametric MTR Functions: Non-Hispanic Subsample

Outcome Variable	Parameters for any for any $A \in \{Y, S\}$			
	$\theta_{0,0}^{A}$	$\theta_{0,1}^{A}$	$\theta_{1,0}^{A}$	$\theta_{1,1}^{A}$
Employment (S)	0.46	0.66	0.46	0.89
	$[0.39, 0.47]$	$[0.64, 0.71]$	$[0.39, 0.47]$	$[0.84, 1.00]$
Labor Earnings (Y)	2.96	5.74	3.00	8.39
	$[1.45, 3.69]$	$[4.98, 6.94]$	$[2.20, 3.41]$	$[7.54, 9.81]$

Note: The MTR on Employment is given by $M^{S}\left(u, \boldsymbol{\theta}_{d}^{S}\right) = \theta_{d,0}^{S} \cdot (1-u) + \theta_{d,1}^{S} \cdot u$ with feasibility set given by $\Theta^{S} = \left\{\left(\boldsymbol{\theta}_{0}^{S}, \boldsymbol{\theta}_{1}^{S}\right) \in [0,1]^{4} : \boldsymbol{\theta}_{1}^{S} \geq \boldsymbol{\theta}_{0}^{S}\right\}$. The MTR on Labor Earnings is given by $M^{Y}\left(u, \boldsymbol{\theta}_{d}^{Y}\right) = \theta_{d,0}^{Y} \cdot (1-u) + \theta_{d,1}^{Y} \cdot u$ with feasibility set given by $\Theta^{Y} = \mathbb{R}_{+}^{4}$. In brackets, I report 90%-confidence interval based on 5,000 bootstrap repetitions. Estimation uses design weights.

It is easier to understand and interpret those estimate using Figure 1. The solid lines are the point-estimates of the MTR and MTE functions based on the parameters reported in Table 6. The dotted lines are point-wise 90%-confidence intervals around the estimated functions based on 5,000 bootstrap repetitions. Blue colored lines are associated with treated potential outcomes, while red colored lines are associated with untreated outcomes. In Subfigure 1a, I find that, although the employment probability for the agents who are most likely to attend the JCTP is similar between treated and untreated individuals, the employment probability for the agents who are less likely to attend the JCTP is much higher for treated individuals than for untreated ones. As a consequence, the MTE on employment for the Non-Hispanic subsample (Subfigure 1b) is a increasing function of the latent heterogeneity. Similarly, in Subfigure 1c, I find that, although expected hourly labor earnings for the agents who are most likely to attend the JCTP is similar between treated and untreated individuals, expected hourly labor earnings for the agents who are less likely to attend the JCTP is much higher for treated individuals than for untreated ones. As a consequence, the MTE on hourly labor earnings for the Non-Hispanic subsample (Subfigure 1d) is a increasing function of the latent heterogeneity.

Figure 1: Parametric MTR and MTE Functions: Non-Hispanic subsample



Notes: The solid lines are the point-estimates of the MTR and MTE functions based on the parameters reported in Table 6. The dotted lines are point-wise 90%-confidence intervals around the estimated functions based on 5,000 bootstrap repetitions. Blue colored lines are associated with treated potential outcomes, while red colored lines are associated with untreated outcomes. The vertical dashed lines represent the sample values of the propensity score $P[D=1|Z]$. Estimation uses design weights.

6.3 Bounds on the MTE$^{\mathbf{OO}}$ on Wages: Non-Hispanic Sub-population

To partially identify the MTE^{OO} of the JCTP on wages for the Non-Hispanic subsample, I can combine the functions estimated in Subsection 6.2 with Corollaries 11 and 14. I stress that, while the first corollary imposes only assumptions that are valid by the experimental design (Assumption 1), technical (Assumptions 3-7) or testable (Assumptions 2 and 8, and equation 1), Corollary 14 additionally uses the Mean Dominance Assumption 9. This last assumption imposes that the marginal treatment response function of wages when treated for the always-employed population is greater than the same object for the employed-only-when-treated population, implying a positive correlation between employment and wages, which is supported by standard models of labor supply according to Chen & Flores (2015).

Another important issue when estimating bounds on a parameter of interest is that there are two ways to construct confidence intervals. The conservative method finds the ζ-confidence intervals around the upper and lower MTE^{OO} bounds and, then, uses their upper most and lower most bounds to construct a confidence interval that contains the identified region with probability ζ. Since the parameter of interest has to be inside the identified region, such confidence interval contain the parameter of interest with probability at least ζ. An alternative method is proposed by Imbens & Manski (2004), who directly construct a ζ-confidence interval that contains the parameter of interest. Since they take into account that the parameter of interest has to be inside the identified region by construction, their confidence interval is tighter than the conservative method.

Figure 2 shows the parametric bounds of the MTE^{OO} on wages using Corollary 11 (Subfigure 2a) and using Corollary 14 (Subfigure 2b). The solid lines are the point-estimates of the parametric bounds of the MTE on wages, while the dotted lines are point-wise conservative 90%-confidence intervals around the identified region based on 5,000 bootstrap repetitions and the dashed lines (that are almost on top of the solid lines) are point-wise 90%-confidence intervals of the parameter of interest (Imbens & Manski (2004)) based also on 5,000 bootstraps repetitions.

As a way to understand the magnitude of the effects, I compare the estimated MTE^{OO}

bounds against the average observed hourly wage of the Non-Hispanics assigned to the control group, \$7.72. Note that the bounds that do not use the mean dominance assumption (Subfigure 2a) are implausibly negative. Even for the agents who are the least likely to attend the JCTP, the lower bound of the MTE^{OO} on wages (-\$6.51) imply that the JCTP would drive their hourly wages almost to zero. Such implausibly negative lower bound is based on the worst-case scenario of a perfectly negative correlation between employment and wages, which is not supported by standard labor models as pointed out by Chen & Flores (2015).

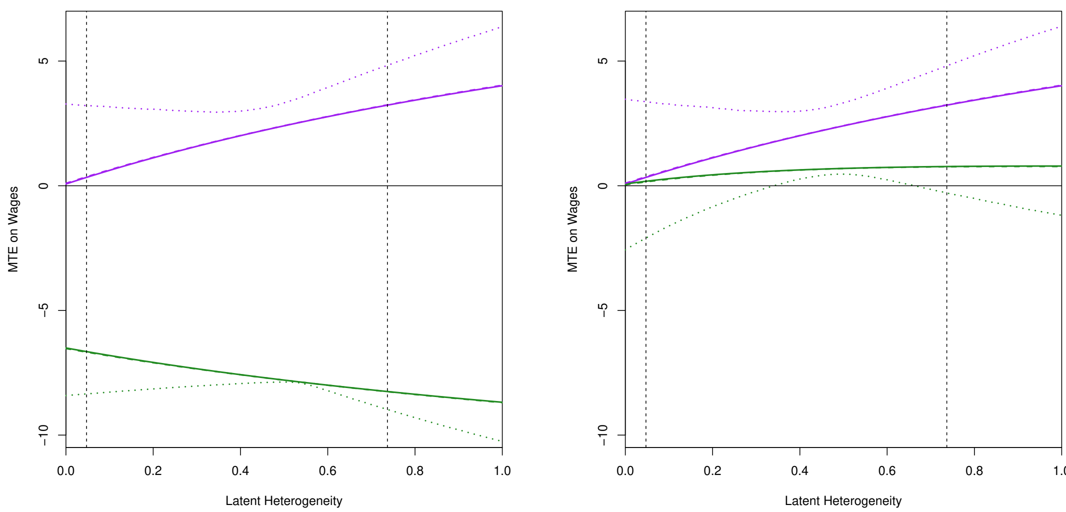
By imposing the Mean Dominance Assumption 9, I rule out this extreme case and can increase the lower bound from equation (9) to equation (21), narrowing the bounds of the MTE^{OO} on wages (Subfigure 2b). Under this extra assumption, the MTE^{OO} on wages is significant at the 10%-confidence level for latent heterogeneity values between 0.34 and 0.66 even when I use the conservative confidence interval given by the dotted lines. When considering the confidence interval based on Imbens & Manski (2004), I find that the JCTP has a significantly positive effect for all agents. Most interestingly, the MTE^{OO} on wages seems to be larger for the agents who are least likely to attend the JCTP.

To better understand the magnitude of those effects and compare my results with the previous literature, I summarize the bounds on the MTE^{OO} function using four key parameters — ATE^{OO}, ATT^{OO}, ATU^{OO} and $LATE^{OO}$ — that are described in Tables 1 and 2 as integrals of the MTE^{OO} function. Table 7 reports those bounds in brackets, the 90%-conservative confidence intervals of the identified region in parenthesis and the 90%-confidence intervals of the parameter of interest (Imbens & Manski (2004)) in braces. As expected, the bounds without the mean dominance assumption are wide and uninformative, while, when imposing Assumption 9, the LATE is significant at 10% according to the conservative confidence interval, and all treatment effects are significant at 10% according to the confidence intervals proposed by Imbens & Manski (2004).

I stress that my $LATE^{OO}$ estimates represent an effect between 7.51% and 24.74% of the average observed hourly wage of the Non-Hispanics assigned to the control group, which are comparable to the bounds of the $LATE^{OO}$ parameter derived by Chen et al. (2017) —

Figure 2: Parametric Bounds of the MTE^{OO} on Wages: Non-Hispanic subsample

(a) Without Mean Dominance Assumption (b) With Mean Dominance Assumption



Notes: The solid lines are the point-estimates of the parametric bounds of the MTE^{OO} on wages. The dotted lines are point-wise conservative 90%-confidence intervals around the identified region based on 5,000 bootstrap repetitions. The dashed lines (that are almost on top of the solid lines) are point-wise 90%-confidence intervals of the parameter of interest (Imbens & Manski (2004)) based on 5,000 bootstraps repetitions. The vertical dashed lines represent the sample values of the propensity score $P[D = 1| Z]$. Estimation uses design weights.

Table 7: Bounds of the ATE^{OO}, ATT^{OO}, ATU^{OO} and $LATE^{OO}$ on Wages: Non-Hispanic subsample

Mean Dominance Asssumption 9	Treatment Effect			
	ATE^{OO}	ATT^{OO}	ATU^{OO}	$LATE^{OO}$
	$[-7.73, 2.28]$	$[-7.11, 1.17]$	$[-8.20, 3.14]$	$[-7.52, 1.91]$
NO	$(-7.88, 3.14)$	$(-8.19, 3.09)$	$(-8.55, 4.32)$	$(-7.96, 2.97)$
	$\{-7.73, 2.29\}$	$\{-7.13, 1.19\}$	$\{-8.21, 3.15\}$	$\{-7.53, 1.92\}$
	$[0.61, 2.28]$	$[0.38, 1.17]$	$[0.73, 3.14]$	$[0.58, 1.91]$
YES	$(0.38, 3.13)$	$(-1.19, 3.17)$	$(-0.01, 4.29)$	$(0.11, 2.96)$
	$\{0.61, 2.29\}$	$\{0.37, 1.19\}$	$\{0.72, 3.15\}$	$\{0.57, 2.96\}$

Note: In brackets, I report the bounds on the parameter of interest that are integrals of the bounds on the MTE^{OO} function. In parenthesis, I report conservative 90%-confidence intervals around the identified region based on 5,000 bootstrap repetitions, while, in braces, I report 90%-confidence intervals of the parameter of interest (Imbens & Manski (2004)) based on 5,000 bootstraps repetitions Estimation uses design weights.

approximately between 7.7% and 17.5% under a similar set of assumptions. The finding that their bounds are tighter than mine for the $LATE^{OO}$ is not surprising because their method leverages all the available information to specifically identify the $LATE^{OO}$ while my tool bounds the MTE^{OO} function and, then, flexibly bounds the other treatment effects for the always-employed population.

As a consequence of such flexibility, I can partially identify other treatment effects that may be policy-relevant. For example, the ATE^{OO} is bounded between 7.90% and 29.30% of the average observed hourly wage of the Non-Hispanics assigned to the control group. Most interestingly, the ATT^{OO} and the ATU^{OO} are, respectively, bounded between 4.92% and 15.16%, and 9.46% and 40.67%, suggesting that the agents who do not to attend the JCTP are the ones who benefit the most from it. To conclude, I stress that, even though the upper bound of the treatment effects on wages may be unrealistically large, the magnitude of the lower bounds are similar to the results found by Chen et al. (2017) and are reasonable when compared to an ITT effect of 16.70% on earnings per week and of 9.87% on hours per week.

7 Conclusion

My main empirical findings suggest that the marginal treatment effect on hourly labor earnings, employment and hourly wages increases with the latent heterogeneity variable for the Non-Hispanic group. More specifically, while MTEs for the agents who are the most likely to attend the JCTP are very small, the MTEs for the agents who are least likely to attend the JCTP are considerably large. Economically, this result implies that the agents who are more likely to benefit from the JCTP are not attending it due to some unobserved constraint. As a consequence, providing stronger incentives for attendance (e.g.: a larger monetary allowance) is possibly efficient.

References

Abadie, A., Angrist, J. & Imbens, G. (2002), 'Instrumental Variables Estimation of the Effect of Subsidezed Training on the Quantiles of Trainee Earnings Treatment', *Econometrica*

70(1), pp. 91–117.

Ahn, H. & Powell, J. L. (1993), 'Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism', *Journal of Econometrics* **58**, pp. 3–29.

Altonji, J. (1993), 'The Demand for and Return to Education When Education Outcomes are Uncertain', *Journal of Labor Economics* **11**(1), pp. 48–83.

Angelucci, M., Karlan, D. & Zinman, J. (2015), 'Microcredit Impacts: Evidence from a Randomized Microcredit Program Placement Experiment by Comportamos Banco', *American Economic Journal: Applied Economics* **7**(1), pp. 151–182.

Angrist, J., Bettinger, E. & Kremer, M. (2006), 'Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia', *The American Economic Review* **96**(3), 847–862.
URL: *http://www.jstor.org/stable/30034075*

Angrist, J., Lang, D. & Oreopoulos, P. (2009), 'Incentives and Services for College Achievement: Evidence from a Randomized Trial', *American Economic Journal: Applied Economics* **1**(1), pp. 1–28.

Behaghel, L., Crepon, B., Gurgand, M. & Barbanchon, T. L. (2015), 'Please Call Again: Correcting Nonresponse Bias in Treatment Effect Models', *The Review of Economics and Statistics* **97**(5), pp. 1070–1080.

Bhattacharya, J., Shaikh, A. M. & Vytlacil, E. (2008), 'Treatment Effect Bounds under Monotonicity Assumptions: An Application to Swan-Ganz Catheterization', *The American Economic Review: Papers and Proceedings* **98**(2), pp. 351–356.

Bhattacharya, J., Shaikh, A. M. & Vytlacil, E. (2012), 'Treatment Effect Bounds: An Application to Swan-Ganz Catheterization', *Journal of Econometrics* **168**(2), pp. 223–243.

Blanco, G., Chen, X., Flores, C. A. & Flores-Lagunes, A. (2017), Bounding Average and Quantile Effects of Training on Employment and Unemployment Durations under Selection, Censoring, and Noncompliance. Available at: http://conference.iza.org/conference_files/EVAL_2017/blanco_g7367.pdf.

Blanco, G., Flores, C. A. & Flores-Lagunes, A. (2013*a*), 'Bounds on Average and Quantile Treatment Effects of Job Corps Training on Wages', *Journal of Human Resources* **48**(3), pp. 659–701.

Blanco, G., Flores, C. A. & Flores-Lagunes, A. (2013*b*), 'The Effects of Job Corps Training Program on Wages for Adolescents and Young Adults', *American Economic Review: Papers and Proceedings* **103**(3), pp. 418–422.

Blundell, R., Gosling, A., Ichimura, H. & Meghir, C. (2007), 'Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds', *Econometrica* **75**(2), pp. 323–363.

Brinch, C. N., Mogstad, M. & Wiswall, M. (2017), 'Beyond LATE with a Discrete Instrument', *Journal of Political Economy* **125**(4), pp. 985–1039.

Card, D. (1999), The Causal Effect of Education on Earnings, *in* O. Ashenfelter & D. Card, eds, 'Handbook of Labor Economics', Vol. 3A, Elsevier, pp. pp. 1801–1863.

Carneiro, P., Heckman, J. J. & Vytlacil, E. (2011), 'Estimating Marginal Returns to Education', *The American Economic Review* **101**(6), pp. 2754–2781.

CASS (1984), 'Myocardial Infarction and Mortality in the Coronary Artery Surgery Study (CASS) Randomized Trial', *The New England Journal of Medicine* **310**(12), pp. 750–758.

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 and Economic Statistics* **33**(4), pp. 523–540.

Chen, X., Flores, C. A. & Flores-Lagunes, A. (2017), 'Going beyond LATE: Bounding Average Treatment Effects of Job Corps Training', *Journal of Human Resources* .

Chen, X., Hong, H. & Tarozzi, A. (2008), 'Semiparametric Efficiency in GMM Models with Auxiliary Data', *The Annals of Statistics* **36**(2), 808–843.

Chesher, A. (2010), 'Instrumental Variable Models for Discrete Outcomes', *Econometrica* **78**(2), pp. 575–601.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W. & Yagan, D. (2011), 'How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star', *The Quarterly Journal of Economics* **126**(4), 1593–1660.
URL: *http://dx.doi.org/10.1093/qje/qjr041*

Chiburis, R. C. (2010), 'Semiparametric Bounds on Treatment Effects', *Journal of Econometrics* **159**(2), 267–275.

DeMel, S., McKenzie, D. & Woodruff, C. (2013), 'The Demand for, and Consequences of, Formalization among Informal First in Sri Lanka', *American Economic Journal: Applied Economics* **5**(2), 122–150.

Dobbie, W. & Jr., R. G. F. (2015), 'The Medium-Term Impacts of High-Achieving Charter Schools', *Journal of Political Economy* **123**(5), pp. 985–1037.

Farber, H. S. (1993), 'The Incidence and Costs of Job Loss: 1982-91', *Brookings Papers: Microeconomics* pp. pp. 73–132.

Flores, C., Flores-Lagunes, A., Gonzales, A. & Neumann, T. (2012), 'Estimating the Effects of Lenght of Exposure to Instruction in a Training Program: The Case of Job Corps', *The Review of Economics and Statistics* **94**(1), pp. 153–171.

Flores-Lagunes, A., Gonzales, A. & Neumann, T. (2010), 'Learning but no Earning? The Impact of Job Corps Training on Hispanic Youth', *Economic Inquiry* **48**(3), pp. 651–667.

Frangakis, C. & Rubin, D. (2002), 'Principal Stratification in Causal Inference', *Biometrics* **58**(1), pp. 21–29.

Fricke, H., Frolich, M., Huber, M. & Lechner, M. (2015), Endogeneity and Non-Response Bias in Treatment Evaluation: Nonparametric Identification of Causal Effects by Instruments. Available at: https://www.iza.org/publications/dp/9428/endogeneity-and-non-response-bias-in-treatment-evaluation-nonparametric-identification-of-

Frolich, M. & Huber, M. (2014), 'Treatment Evaluation with Multiple Outcome Periods Under Endogeneity and Attrition', *Journal of the American Statistical Association* **109**(508), 1697–1711.

Heckman, J. J. (1979), 'Sample Selection Bias as a Specification Error', *Econometrica* **47**(1), pp. 153–161.

Heckman, J. J. & Borjas, G. J. (1980), 'Does unemployment cause future unemployment? definitions, questions and answers from a continuous time model of heterogeneity and state dependence', *Economica* **47**(187), 247–283.
URL: *http://www.jstor.org/stable/2553150*

Heckman, J. J., Urzua, S. & Vytlacil, E. (2006), 'Understanding Instrumental Variables in Models with Essential Heterogeneity', *The Review of Economics and Statistics* **88**(3), pp. 389–432.

Heckman, J. J. & Vytlacil, E. (2001*a*), Instrumental Variables, Selection Models, and Tight Bounds on the Average Treatment Effect, *in* M. Lechner & J. Pfeifer, eds, 'Econometric Evaluation of Labour Market Policies', Vol. 13 of *ZEW Economic Studies*, Physica, Heidelberg.

Heckman, J. J. & Vytlacil, E. (2005), 'Structural Equations, Treatment Effects and Econometric Policy Evaluation', *Econometrica* **73**(3), pp. 669–738.

Heckman, J., LaLonde, R. & Smith, J. (1999), The Economics and Econometrics of Active Labor Market Programs, *in* O. Ashenfelter & D. Card, eds, 'Handbook of Labor Economics', Vol. 3A, Elsevier, pp. pp. 1865–2097.

Heckman, J. & Smith, J. (1999), 'The Pre-Programme Earnings Dip and the Determinants of Participation in a Social Programme: Implications for Simple Programme Evaluation Strategies', *Economic Journal* **109**(2), pp. 313–348.

Heckman, J. & Vytlacil, E. (1999), 'Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects', *Proceedings of the National Academy of Sciences of the United States of America* **96**, 4730–4734.

Heckman, J. & Vytlacil, E. (2001*b*), 'Policy-Relevant Treatment Effects', *American Economic Review: Papers and Proceedings* **91**(2), pp. 107–111.

Helland, E. & Yoon, J. (2017), 'Estimating the Effects of the English Rule on Litigation Outcomes', *The Review of Economics and Statistics* **99**(4), pp. 678–682.

Horowitz, J. & Manski, C. (2000), 'Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data', *Journal of the American Statistical Association* **95**, pp. 77–84.

Huber, M. (2014), 'Treatment Evaluation in the Presence of Sample Selection', *Econometric Reviews* **33**(8), pp. 869–905.

Huber, M., Laffers, L. & Mellace, G. (2017), 'Sharp IV Bounds on Average Treatment Effects on the Treated and Other Populations under Endogeneity and Noncompliance', *Journal of Applied Econometrics* **32**, pp. 56–79.

Huber, M. & Mellace, G. (2015), 'Sharp Bounds on Causal Effects under Sample Selection', *Oxford Bulletin of Economics and Statistics* **77**(1), pp. 129–151.

Imai, K. (2008), 'Sharp Bounds on the Causal Effects in Randomized Experiments with Truncation- by- Death', *Statistics and Probability Letters* **78**(2), pp. 144–149.

Imbens, G. & Manski, C. (2004), 'Confidence Intervals for Partially Identified Parameters', *Econometrica* **72**(6), pp. 1845–1857.

Imbens, G. W. & Angrist, J. D. (1994), 'Identification and Estimation of Local Average Treatment Effects', *Econometrica* **62**(2), pp. 467–475.

Jacobson, L. S., LaLonde, R. J. & Sullivan, D. G. (1993), 'Earnings Losses of Displaced Workers', *The American Economic Review* **83**(4), pp. 685–709.

Krueger, A. B. & Whitmore, D. M. (2001), 'The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR', *The Economic Journal* **111**(468), pp. 1–28.

Lechner, M. & Mell, B. (2010), Partial Idendification of Wage Effects of Training Programs. Available at: https://ideas.repec.org/p/bro/econwp/2010-8.html.

Lee, D. S. (2009), 'Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects', *The Review of Economic Studies* **76**, pp. 1071–1102.

Lee, S. & Salanie, B. (2016), Identifying Effects of Multivalued Treatments.

Machado, C., Shaikh, A. M. & Vytlacil, E. (2018), Instrumental Variables and the Sign of the Average Treatment Effect. Forthcoming on the Journal of Econometrics. Available at: https://drive.google.com/file/d/1yIdbyzVaCj5f9cOHN8K_nhSKGDKDhCzY/view.

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

Manski, C. F. (1997), 'Monotone Treatment Response', *Econometrica* **65**(6), pp. 1311–1334.

Manski, C. F. & Pepper, J. V. (2000), 'Monotone Instrumental Variables: With an Application to the Returns to Schooling', *Econometrica* **68**(4), pp. 997–1010.

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

Mogstad, M., Santos, A. & Torgovitsky, A. (2018), 'Using Instrumental Variables for Inference about Policy Relevant Treatment Effects', *Econometrica* **86**(5), pp. 1589–1619. NBER Working Paper 23568.

Newey, W. K., Powell, J. L. & Vella, F. (1999), 'Nonparametric Estimation of Triangular Simultaneous Equations Models', *Econometrica* **67**(3), pp. 565–603.

Rubin, D. (1974), 'Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies', *Journal of Educational Psychology* **66**(5), 688–701.

Schochet, P. Z., Bughardt, J. & Glazerman, S. (2001), National Job Corps on Participants' Employment and Related Outcomes, Technical Report MPR Reference No. 8140-530, Mathematic Policy Research.

Schochet, P. Z., Bughardt, J. & McConnell, S. (2008), 'Does Job Corps Work? Impact Findings from the National Job Corps Study', *American Economic Review* **98**(5), pp. 1864–1886.

Sexton, M. & Hebel, R. (1984), 'A Clinical Trial of Change in Maternal Smoking and its Effects on Birth Weight', *Journal of the American Medical Association* **251**(7), pp. 911–915.

Shaikh, A. & Vytlacil, E. (2011), 'Partial Identification in Triangular Systems of Equations with Binary Dependent Variables', *Econometrica* **79**(3), pp. 949–955.

Steinmayr, A. (2014), When a Random Sample is not Random: Bounds on the Effect of Migration on Children Left Behind. Working Paper.

U.S. Department of Health and Human Services (2004), *The Health Consequences of Smoking: A Report of the Surgeon General*, U.S. Department of Health and Human Services, Public Health Service, Office on Smoking and Health. Available at: https://www.cdc.gov/tobacco/data_statistics/sgr/2004/index.htm.

van Ours, J. C. (2004), 'The Locking-in Effect of Subsidized Jobs', *Journal of Comparative Economics* **32**(1), pp. 37–55.

Vytlacil, E. (2002), 'Independence, Monotonicity and Latent Index Models: An Equivalence Result', *Econometrica* **70**(1), pp. 331–341.

Supporting Information
(Online Appendix)

A Proofs of the main results

A.1 Proof of Equation (4)

Note that

$$\begin{aligned}
\mathbb{E}\left[Y_0^* \,|X=x,U=u,S_0=1,S_1=1\right] &= \mathbb{E}\left[Y_0^* \,|X=x,U=u,S_0=1\right] \\
&\quad \text{by Assumption 8} \\
&= \frac{\mathbb{E}\left[S_0 \cdot Y_0^* \,|X=x,U=u\right]}{\mathbb{P}\left[S_0=1 \,|X=x,U=u\right]} \\
&\quad \text{by the definition of conditional expectation} \\
&= \frac{\mathbb{E}\left[Y_0 \,|X=x,U=u\right]}{\mathbb{E}\left[S_0 \,|X=x,U=u\right]} \\
&= \frac{m_0^Y\left(x,u\right)}{m_0^S\left(x,u\right)}. \qquad \blacksquare
\end{aligned}$$

A.2 Proof of Equation (5)

First, observe that

$$\begin{aligned}
m_0^S\left(x,u\right) &:= \mathbb{E}\left[S_0 \,|X=x,U=u\right] \\
&= \mathbb{P}\left[Q\left(0,X\right) \geq V \,|X=x,U=u\right] \qquad \text{(A.1)} \\
&\quad \text{by equation (2)}, \\
m_1^S\left(x,u\right) &:= \mathbb{E}\left[S_1 \,|X=x,U=u\right] \\
&= \mathbb{P}\left[Q\left(1,X\right) \geq V \,|X=x,U=u\right] \qquad \text{(A.2)} \\
&\quad \text{by equation (2)}, \\
\Delta_S\left(x,u\right) &:= \mathbb{E}\left[S_1 - S_0 \,|X=x,U=u\right] \\
&= m_1^S\left(x,u\right) - m_0^S\left(x,u\right)
\end{aligned}$$

$$= \mathbb{P}\left[Q(1,X) \geq V > Q(0,X) \,|X = x, U = u\right]$$

by equations (A.1) and (A.2) and Assumption (8)

$$= \mathbb{P}\left[S_0 = 0, S_1 = 1 \,|X = x, U = u\right] \tag{A.3}$$

by equation (2), and

$$\Delta_Y^{NO}(x,u) \coloneqq \mathbb{E}\left[Y_1 - Y_0 \,|X = x, U = u, S_0 = 0, S_1 = 1\right]$$

$$= \mathbb{E}\left[S_1 \cdot Y_1^* - S_0 \cdot Y_0^* \,|X = x, U = u, S_0 = 0, S_1 = 1\right]$$

$$= \mathbb{E}\left[Y_1^* \,|X = x, U = u, S_0 = 0, S_1 = 1\right]. \tag{A.4}$$

Note also that:

$$m_1^Y(x,u) \coloneqq \mathbb{E}\left[Y_1 \,|X = x, U = u\right]$$

$$= \mathbb{E}\left[S_1 \cdot Y_1^* \,|X = x, U = u\right]$$

$$= \mathbb{E}\left[Y_1^* \,|X = x, U = u, S_0 = 1, S_1 = 1\right] \cdot \mathbb{P}\left[S_0 = 1 \,|X = x, U = u\right]$$

$$+ \mathbb{E}\left[Y_1^* \,|X = x, U = u, S_0 = 0, S_1 = 1\right] \cdot \mathbb{P}\left[S_0 = 0, S_1 = 1 \,|X = x, U = u\right]$$

by Assumption 8 and the Law of Iterated Expectations

$$= \mathbb{E}\left[Y_1^* \,|X = x, U = u, S_0 = 1, S_1 = 1\right] \cdot m_0^S(x,u) + \Delta_Y^{NO}(x,u) \cdot \Delta_S(x,u) \tag{A.5}$$

by equations (A.1), (A.3) and (A.4),

implying equation (5) after some rearrangement. ■

A.3 Proof of Proposition 10

Note that

$$\underline{y}^* \leq \mathbb{E}\left[Y_1^* \,|X = x, U = u, S_0 = 1, S_1 = 1\right] \leq \overline{y}^* \tag{A.6}$$

by the definition of $\underline{y}^*$ and $\overline{y}^*$. Observe also that

$$\underline{y}^* \leq \Delta_Y^{NO}(x,u) \leq \overline{y}^*$$

by equation (A.4) and the definition of $\underline{y}^*$ and $\overline{y}^*$, implying, by equation (5), that

$$\mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 1, S_1 = 1\right.\right] \leq \frac{m_1^Y(x,u) - \underline{y}^* \cdot \Delta_S(x,u)}{m_0^S(x,u)} \tag{A.7}$$

under assumption 7.1,

$$\frac{m_1^Y(x,u) - \overline{y}^* \cdot \Delta_S(x,u)}{m_0^S(x,u)} \leq \mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 1, S_1 = 1\right.\right] \tag{A.8}$$

under assumption 7.2, and

$$\begin{aligned} \frac{m_1^Y(x,u) - \overline{y}^* \cdot \Delta_S(x,u)}{m_0^S(x,u)} &\leq \mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 1, S_1 = 1\right.\right] \\ &\leq \frac{m_1^Y(x,u) - \underline{y}^* \cdot \Delta_S(x,u)}{m_0^S(x,u)}. \end{aligned} \tag{A.9}$$

under assumption 7.3 (sub-case (a) or (b)). Combining equations (A.6)-(A.9), it is easy to show that proposition 10 holds. ■

A.4 Proof of Proposition 12

First, I prove proposition 12 under assumption 7.3 (sub-cases (a) and (b)). At the end of this section, I prove proposition 12 under assumptions 7.1 and 7.2.

A.4.1 Proof under Assumption 7.3 (sub-cases (a) and (b))

Fix $\overline{u} \in [0,1]$, $\overline{x} \in \mathcal{X}$ and $\delta(\overline{x},\overline{u}) \in \left(\underline{\Delta_{Y^*}^{OO}}(\overline{x},\overline{u}), \overline{\Delta_{Y^*}^{OO}}(\overline{x},\overline{u})\right)$ arbitrarily. For brevity, define $\alpha(\overline{x},\overline{u}) \coloneqq \delta(\overline{x},\overline{u}) + \dfrac{m_0^Y(\overline{x},\overline{u})}{m_0^S(\overline{x},\overline{u})}$ and $\gamma(\overline{x},\overline{u}) \coloneqq \dfrac{m_1^Y(\overline{x},\overline{u}) - \alpha(\overline{x},\overline{u}) \cdot m_0^S(\overline{x},\overline{u})}{\Delta_S(\overline{x},\overline{u})}$.

Note that

$$
\begin{aligned}
\delta\left(\overline{x},\overline{u}\right) &\in \left(\underline{\Delta_{Y^*}^{OO}}\left(\overline{x},\overline{u}\right),\overline{\Delta_{Y^*}^{OO}}\left(\overline{x},\overline{u}\right)\right)\\
\Leftrightarrow\quad \alpha\left(\overline{x},\overline{u}\right) &\in \left(\max\left\{\frac{m_1^Y\left(x,u\right)-\overline{y}^*\cdot\Delta_S\left(x,u\right)}{m_0^S\left(x,u\right)},\underline{y}^*\right\},\right.\\
&\qquad\qquad \left.\min\left\{\frac{m_1^Y\left(x,u\right)-\underline{y}^*\cdot\Delta_S\left(x,u\right)}{m_0^S\left(x,u\right)},\overline{y}^*\right\}\right)\\
&\subseteq\left(\underline{y}^*,\overline{y}^*\right),
\end{aligned}
\tag{A.10}
$$

and that

$$
\begin{aligned}
\alpha\left(\overline{x},\overline{u}\right) &\in \left(\frac{m_1^Y\left(x,u\right)-\overline{y}^*\cdot\Delta_S\left(x,u\right)}{m_0^S\left(x,u\right)},\frac{m_1^Y\left(x,u\right)-\underline{y}^*\cdot\Delta_S\left(x,u\right)}{m_0^S\left(x,u\right)}\right)\\
\Leftrightarrow\quad \gamma\left(\overline{x},\overline{u}\right) &\in \left(\underline{y}^*,\overline{y}^*\right).
\end{aligned}
\tag{A.11}
$$

The strategy of this proof consists of defining random variables $\left(\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}\right)$ through their joint cumulative distribution function $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}$ and, then, checking that equations (15), (16) and (17) are satisfied. I fix $(y_0,y_1,u,v,z,x)\in\mathbb{R}^6$ and define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}$ in twelve steps:

Step 1. For $x\notin\mathcal{X}$, $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}\left(y_0,y_1,u,v,z,x\right)=F_{Y_0^*,Y_1^*,U,V,Z,X}\left(y_0,y_1,u,v,z,x\right)$.

Step 2. From now on, consider $x\in\mathcal{X}$. Since

$$
F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}\left(y_0,y_1,u,v,z,x\right)=F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0,y_1,u,v,z\,|x\right)\cdot F_X\left(x\right),
$$

it suffices to define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0,y_1,u,v,z\,|x\right)$. Moreover, I impose

$$
Z\perp\!\!\!\perp\left(\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}\right)|X
$$

by writing

$$F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0,y_1,u,v,z\,|x\right)=F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}|X}\left(y_0,y_1,u,v\,|x\right)\cdot F_{Z|X}\left(z\,|x\right),$$

implying that it is sufficient to define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}|X}\left(y_0,y_1,u,v\,|x\right)$.

Step 3. For $u\notin[0,1]$, I define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}|X}\left(y_0,y_1,u,v\,|x\right)=F_{Y_0^*,Y_1^*,U,V|X}\left(y_0,y_1,u,v\,|x\right)$.

Step 4. From now on, consider $u\in[0,1]$. Since

$$F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}|X}\left(y_0,y_1,u,v\,|x\right)=F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}|X,\tilde{U}}\left(y_0,y_1,v\,|x,u\right)\cdot F_{\tilde{U}|X}\left(u\,|x\right),$$

it suffices to define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}|X,\tilde{U}}\left(y_0,y_1,v\,|x,u\right)$ and $F_{\tilde{U}|X}\left(u\,|x\right)$.

Step 5. I define $F_{\tilde{U}|X}\left(u\,|x\right)=F_{U|X}\left(u\,|x\right)=u$.

Step 6. For any $u\neq\overline{u}$, I define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}|X,\tilde{U}}\left(y_0,y_1,v\,|x,u\right)=F_{Y_0^*,Y_1^*,V|X,U}\left(y_0,y_1,v\,|x,u\right)$.

Step 7. For any $v\notin[0,1]$, I define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}|X,\tilde{U}}\left(y_0,y_1,v\,|x,\overline{u}\right)=F_{Y_0^*,Y_1^*,V|X,U}\left(y_0,y_1,v\,|x,\overline{u}\right)$.

Step 8. From now on, consider $v\in[0,1]$. Since

$$F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}|X,\tilde{U}}\left(y_0,y_1,v\,|x,\overline{u}\right)=F_{\tilde{Y}_0^*,\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}\left(y_0,y_1\,|x,\overline{u},v\right)\cdot F_{\tilde{V}|X,\tilde{U}}\left(v\,|x,\overline{u}\right),$$

it is sufficient to define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}\left(y_0,y_1\,|x,\overline{u},v\right)$ and $F_{\tilde{V}|X,\tilde{U}}\left(v\,|x,\overline{u}\right)$.

Step 9. I define

$$F_{\tilde{V}|X,\tilde{U}}\left(v\,|x,\overline{u}\right)=\begin{cases} m_0^S\left(x,\overline{u}\right)\cdot\dfrac{v}{Q\left(0,x\right)} & \text{if } v\leq Q\left(0,x\right)\\[2ex] m_0^S\left(x,\overline{u}\right)+\Delta_S\left(x,\overline{u}\right)\cdot\dfrac{v-Q\left(0,x\right)}{Q\left(1,x\right)-Q\left(0,x\right)} & \text{if } Q\left(0,x\right)<v\leq Q\left(1,x\right)\\[2ex] m_1^S\left(x,\overline{u}\right)+\left(1-m_1^S\left(x,\overline{u}\right)\right)\dfrac{v-Q\left(1,x\right)}{1-Q\left(1,x\right)} & \text{if } Q\left(1,x\right)<v\end{cases}.$$

Step 10. I write $F_{\tilde{Y}_0^*, \tilde{Y}_1^* | X, \tilde{U}, \tilde{V}}\left(y_0, y_1 \left| x, \overline{u}, v\right.\right) = F_{\tilde{Y}_0^* | X, \tilde{U}, \tilde{V}}\left(y_0 \left| x, \overline{u}, v\right.\right) \cdot F_{\tilde{Y}_1^* | X, \tilde{U}, \tilde{V}}\left(y_1 \left| x, \overline{u}, v\right.\right)$, implying that I can separately define $F_{\tilde{Y}_0^* | X, \tilde{U}, \tilde{V}}\left(y_0 \left| x, \overline{u}, v\right.\right)$ and $F_{\tilde{Y}_1^* | X, \tilde{U}, \tilde{V}}\left(y_1 \left| x, \overline{u}, v\right.\right)$.

Step 11. When $\mathcal{Y}^*$ is a bounded interval (sub-case (a) in assumption 7.3), I define

$$F_{\tilde{Y}_0^* | X, \tilde{U}, \tilde{V}}\left(y_0 \left| x, \overline{u}, v\right.\right) = \begin{cases} \mathbf{1}\left\{y_0 \geq \dfrac{m_0^Y\left(\overline{x}, \overline{u}\right)}{m_0^S\left(\overline{x}, \overline{u}\right)}\right\} & \text{if } v \leq Q\left(0, x\right) \\ ----------- & -------- \\ \mathbf{1}\left\{y_0 \geq \dfrac{\underline{y}^* + \overline{y}^*}{2}\right\} & \text{if } Q\left(0, x\right) < v \end{cases}.$$

When $\overline{y}^* = \max\left\{y \in \mathcal{Y}^*\right\}$ and $\underline{y}^* = \min\left\{y \in \mathcal{Y}^*\right\}$ (case (b) in assumption 7.3), I define

$$F_{\tilde{Y}_0^* | X, \tilde{U}, \tilde{V}}\left(y_0 \left| x, \overline{u}, v\right.\right) = \begin{cases} 0 & \text{if } y_0 < \underline{y}^* \text{ and } v \leq Q\left(0, x\right) \\ 1 - \dfrac{\dfrac{m_0^Y\left(\overline{x}, \overline{u}\right)}{m_0^S\left(\overline{x}, \overline{u}\right)} - \underline{y}^*}{\overline{y}^* - \underline{y}^*} & \text{if } \underline{y}^* \leq y_0 < \overline{y}^* \text{ and } v \leq Q\left(0, x\right) \\ 1 & \text{if } \overline{y}^* \leq y_0 \text{ and } v \leq Q\left(0, x\right) \\ ----------- & --------------- \\ \mathbf{1}\left\{y_0 \geq \overline{y}^*\right\} & \text{if } Q\left(0, x\right) < v \end{cases}.$$

which are valid cumulative distribution functions because $\dfrac{m_0^Y\left(\overline{x}, \overline{u}\right)}{m_0^S\left(\overline{x}, \overline{u}\right)} \in \left[\underline{y}^*, \overline{y}^*\right]$.

Step 12. When $\mathcal{Y}^*$ is a bounded interval (case (a) in assumption 7.3), I define

$$F_{\tilde{Y}_1^* | X, \tilde{U}, \tilde{V}}\left(y_1 \left| x, \overline{u}, v\right.\right) = \begin{cases} \mathbf{1}\left\{y_1 \geq \alpha\left(\overline{x}, \overline{u}\right)\right\} & \text{if } v \leq Q\left(0, x\right) \\ --------- & ------------ \\ \mathbf{1}\left\{y_1 \geq \gamma\left(\overline{x}, \overline{u}\right)\right\} & \text{if } Q\left(0, x\right) < v \leq Q\left(1, x\right) \\ --------- & ------------ \\ \mathbf{1}\left\{y_1 \geq \dfrac{\underline{y}^* + \overline{y}^*}{2}\right\} & \text{if } Q\left(1, x\right) < v \end{cases}.$$

When $\overline{y}^* = \max\{y \in \mathcal{Y}^*\}$ and $\underline{y}^* = \min\{y \in \mathcal{Y}^*\}$ (case (b) in assumption 7.3), I define

$$F_{\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}\left(y_1 \left| x, \overline{u}, v\right.\right) = \begin{cases} 0 & \text{if } y_1 < \underline{y}^* \text{ and } v \leq Q(0,x) \\ 1 - \dfrac{\alpha(\overline{x},\overline{u}) - \underline{y}^*}{\overline{y}^* - \underline{y}^*} & \text{if } \underline{y}^* \leq y_1 < \overline{y}^* \text{ and } v \leq Q(0,x) \\ 1 & \text{if } \overline{y}^* \leq y_1 \text{ and } v \leq Q(0,x) \\ --------- & --------------------- \\ 0 & \text{if } y_1 < \underline{y}^* \text{ and } Q(0,x) < v \leq Q(1,x) \\ 1 - \dfrac{\gamma(\overline{x},\overline{u}) - \underline{y}^*}{\overline{y}^* - \underline{y}^*} & \text{if } \underline{y}^* \leq y_1 < \overline{y}^* \text{ and } Q(0,x) < v \leq Q(1,x) \\ 1 & \text{if } \overline{y}^* \leq y_1 \text{ and } Q(0,x) < v \leq Q(1,x) \\ --------- & --------------------- \\ \mathbf{1}\{y_1 \geq \overline{y}^*\} & \text{if } Q(1,x) < v \end{cases}.$$

which are valid cumulative distribution functions because of equations (A.10) and (A.11).

Having defined the joint cumulative distribution function $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}$, note that equations (A.10) and (A.11), $\dfrac{m_0^Y(\overline{x},\overline{u})}{m_0^S(\overline{x},\overline{u})} \in \left[\underline{y}^*, \overline{y}^*\right]$ and steps 7-12 ensure that equation (16) holds.

Now, I show, in three steps, that equation (15) holds.

Step 13. Observe that

$$\begin{aligned} &\mathbb{E}\left[\tilde{Y}_1^* \left| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right.\right] \\ &\quad = \mathbb{E}\left[\tilde{Y}_1^* \left| X = \overline{x}, \tilde{U} = \overline{u}, Q(0,\overline{x}) \geq \tilde{V}\right.\right] \\ &\qquad \text{by the definition of } \tilde{S}_0 \text{ and } \tilde{S}_1 \\ &\quad = \frac{\mathbb{E}\left[\mathbf{1}\left\{Q(0,\overline{x}) \geq \tilde{V}\right\} \cdot \tilde{Y}_1^* \left| X = \overline{x}, \tilde{U} = \overline{u}\right.\right]}{\mathbb{P}\left[Q(0,\overline{x}) \geq \tilde{V} \left| X = \overline{x}, \tilde{U} = \overline{u}\right.\right]} \end{aligned}$$

by the definition of conditional expectation

$$= \frac{\mathbb{E}\left[\mathbf{1}\left\{Q\left(0,\overline{x}\right) \geq \tilde{V}\right\} \cdot \mathbb{E}\left[\tilde{Y}_1^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{V}\right] \middle| X = \overline{x}, \tilde{U} = \overline{u}\right]}{\mathbb{P}\left[Q\left(0,\overline{x}\right) \geq \tilde{V} \middle| X = \overline{x}, \tilde{U} = \overline{u}\right]}$$

by the Law of Iterated Expectations

$$= \frac{\int_0^{Q(0,\overline{x})} \mathbb{E}\left[\tilde{Y}_1^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{V} = v\right] \mathrm{d}F_{\tilde{V}|X,\tilde{U}}\left(v \,|x, \overline{u}\right)}{\mathbb{P}\left[Q\left(0,\overline{x}\right) \geq \tilde{V} \middle| X = \overline{x}, \tilde{U} = \overline{u}\right]}$$

by the definition of expectation and by step 7

$$= \frac{\int_0^{Q(0,\overline{x})} \alpha\left(\overline{x}, \overline{u}\right) \mathrm{d}F_{\tilde{V}|X,\tilde{U}}\left(v \,|x, \overline{u}\right)}{\mathbb{P}\left[Q\left(0,\overline{x}\right) \geq \tilde{V} \middle| X = \overline{x}, \tilde{U} = \overline{u}\right]}$$

by step 12

$$= \alpha\left(\overline{x}, \overline{u}\right) \tag{A.12}$$

by linearity of the Lebesgue Integral

Step 14. Similarly to the last step, notice that

$$\begin{aligned}
&\mathbb{E}\left[\tilde{Y}_0^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right] \\
&= \mathbb{E}\left[\tilde{Y}_0^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, Q\left(0,\overline{x}\right) \geq \tilde{V}\right] \\
&= \frac{\mathbb{E}\left[\mathbf{1}\left\{Q\left(0,\overline{x}\right) \geq \tilde{V}\right\} \cdot \tilde{Y}_0^* \middle| X = \overline{x}, \tilde{U} = \overline{u}\right]}{\mathbb{P}\left[Q\left(0,\overline{x}\right) \geq \tilde{V} \middle| X = \overline{x}, \tilde{U} = \overline{u}\right]} \\
&= \frac{\mathbb{E}\left[\mathbf{1}\left\{Q\left(0,\overline{x}\right) \geq \tilde{V}\right\} \cdot \mathbb{E}\left[\tilde{Y}_0^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{V}\right] \middle| X = \overline{x}, \tilde{U} = \overline{u}\right]}{\mathbb{P}\left[Q\left(0,\overline{x}\right) \geq \tilde{V} \middle| X = \overline{x}, \tilde{U} = \overline{u}\right]} \\
&= \frac{\int_0^{Q(0,\overline{x})} \mathbb{E}\left[\tilde{Y}_0^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{V} = v\right] \mathrm{d}F_{\tilde{V}|X,\tilde{U}}\left(v \,|x, \overline{u}\right)}{\mathbb{P}\left[Q\left(0,\overline{x}\right) \geq \tilde{V} \middle| X = \overline{x}, \tilde{U} = \overline{u}\right]}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\displaystyle\int_0^{Q(0,\overline{x})} \frac{m_0^Y(\overline{x},\overline{u})}{m_0^S(\overline{x},\overline{u})} \mathrm{d}F_{\tilde{V}|X,\tilde{U}}(v\,|x,\overline{u})}{\mathbb{P}\left[Q(0,\overline{x}) \geq \tilde{V} \,\middle|\, X=\overline{x}, \tilde{U}=\overline{u}\right]} \text{ by step 11} \\
&= \frac{m_0^Y(\overline{x},\overline{u})}{m_0^S(\overline{x},\overline{u})}. \qquad \text{(A.13)}
\end{aligned}
$$

Step 15. Note that

$$
\begin{aligned}
\Delta_{\tilde{Y}^*}^{OO}(\overline{x},\overline{u}) &\coloneqq \mathbb{E}\left[\tilde{Y}_1^* - \tilde{Y}_0^* \,\middle|\, X=\overline{x}, \tilde{U}=\overline{u}, \tilde{S}_0=1, \tilde{S}_1=1\right] \\
&= \mathbb{E}\left[\tilde{Y}_1^* \,\middle|\, X=\overline{x}, \tilde{U}=\overline{u}, \tilde{S}_0=1, \tilde{S}_1=1\right] \\
&\quad - \mathbb{E}\left[\tilde{Y}_0^* \,\middle|\, X=\overline{x}, \tilde{U}=\overline{u}, \tilde{S}_0=1, \tilde{S}_1=1\right] \\
&= \alpha(\overline{x},\overline{u}) - \frac{m_0^Y(\overline{x},\overline{u})}{m_0^S(\overline{x},\overline{u})} \\
&\qquad \text{by equations (A.12) and (A.13)} \\
&= \delta(\overline{x},\overline{u}) \\
&\qquad \text{by the definition of } \alpha(\overline{x},\overline{u}),
\end{aligned}
$$

ensuring that equation (15) holds.

Finally, I show, in two steps, that equation (17) holds.

Step 16. Fix $(y,d,s,z) \in \mathbb{R}^4$ arbitrarily and observe that equation (17) can be simplified to:

$$
\begin{aligned}
&F_{\tilde{Y},\tilde{D},\tilde{S},Z,X}(y,d,s,z,\overline{x}) = F_{Y,D,S,Z,X}(y,d,s,z,\overline{x}) \\
\Leftrightarrow &F_{\tilde{Y},\tilde{D},\tilde{S},Z|X}(y,d,s,z\,|\overline{x}) \cdot F_X(\overline{x}) = F_{Y,D,S,Z|X}(y,d,s,z\,|\overline{x}) \cdot F_X(\overline{x}) \\
\Leftrightarrow &F_{\tilde{Y},\tilde{D},\tilde{S},Z|X}(y,d,s,z\,|\overline{x}) = F_{Y,D,S,Z|X}(y,d,s,z\,|\overline{x}) \qquad \text{(A.14)}
\end{aligned}
$$

Step 17. Notice that

$$
F_{\tilde{Y},\tilde{D},\tilde{S},Z|X}(y,d,s,z\,|\overline{x})
$$

$$
\begin{aligned}
&= \mathbb{E}\left[\mathbf{1}\left\{\left(\tilde{Y}, \tilde{D}, \tilde{S}, Z\right) \leq (y, d, s, z)\right\}\middle| X = \overline{x}\right] \\
&= \int \mathbf{1}\left\{\left(\tilde{Y}, \tilde{D}, \tilde{S}, Z\right) \leq (y, d, s, z)\right\} \mathrm{d}F_{\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}, Z|X}\left(y_0, y_1, u, v, z \left| \overline{x}\right.\right) \\
&\quad \text{because } \left(\tilde{Y}, \tilde{D}, \tilde{S}, Z\right) \text{ are functions of } \left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}, Z\right) \\
&= \int \left[\mathbf{1}\left\{\left(\tilde{Y}, \tilde{D}, \tilde{S}, Z\right) \leq (y, d, s, z)\right\} \cdot \mathbf{1}\left\{u \neq \overline{u}\right\}\right] \mathrm{d}F_{\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}, Z|X}\left(y_0, y_1, u, v, z \left| \overline{x}\right.\right) \\
&\quad + \int \left[\mathbf{1}\left\{\left(\tilde{Y}, \tilde{D}, \tilde{S}, Z\right) \leq (y, d, s, z)\right\} \cdot \mathbf{1}\left\{u = \overline{u}\right\}\right] \mathrm{d}F_{\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}, Z|X}\left(y_0, y_1, u, v, z \left| \overline{x}\right.\right) \\
&\quad \text{by linearity of the Lebesgue Integral} \\
&= \int \left[\mathbf{1}\left\{\left(\tilde{Y}, \tilde{D}, \tilde{S}, Z\right) \leq (y, d, s, z)\right\} \cdot \mathbf{1}\left\{u \neq \overline{u}\right\}\right] \mathrm{d}F_{\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}, Z|X}\left(y_0, y_1, u, v, z \left| \overline{x}\right.\right) \\
&\quad \text{because } \mathbb{P}\left[\tilde{U} = \overline{u}\middle| X = \overline{x}\right] = 0 \text{ by step 5} \\
&= \int \left[\mathbf{1}\left\{(Y, D, S, Z) \leq (y, d, s, z)\right\} \cdot \mathbf{1}\left\{u \neq \overline{u}\right\}\right] \mathrm{d}F_{Y_0^*, Y_1^*, U, V, Z|X}\left(y_0, y_1, u, v, z \left| \overline{x}\right.\right) \\
&\quad \text{by steps 2-6} \\
&= \int \left[\mathbf{1}\left\{(Y, D, S, Z) \leq (y, d, s, z)\right\} \cdot \mathbf{1}\left\{u \neq \overline{u}\right\}\right] \mathrm{d}F_{Y_0^*, Y_1^*, U, V, Z|X}\left(y_0, y_1, u, v, z \left| \overline{x}\right.\right) \\
&\quad + \int \left[\mathbf{1}\left\{(Y, D, S, Z) \leq (y, d, s, z)\right\} \cdot \mathbf{1}\left\{u = \overline{u}\right\}\right] \mathrm{d}F_{Y_0^*, Y_1^*, U, V, Z|X}\left(y_0, y_1, u, v, z \left| \overline{x}\right.\right) \\
&\quad \text{because } \mathbb{P}\left[U = \overline{u} \left| X = \overline{x}\right.\right] = 0 \\
&= \int \mathbf{1}\left\{(Y, D, S, Z) \leq (y, d, s, z)\right\} \mathrm{d}F_{Y_0^*, Y_1^*, U, V, Z|X}\left(y_0, y_1, u, v, z \left| \overline{x}\right.\right) \\
&\quad \text{by linearity of the Lebesgue Integral} \\
&= \mathbb{E}\left[\mathbf{1}\left\{(Y, D, S, Z) \leq (y, d, s, z)\right\}\middle| X = \overline{x}\right] \\
&= F_{Y,D,S,Z|X}\left(y, d, s, z \left| \overline{x}\right.\right),
\end{aligned}
$$

implying equation (17) according to equation (A.14).

I can, then, conclude that proposition 12 is true. ■

As a remark, the above constructive proof defines random variables $\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)$ that matches other important moments of the true data generating process besides the ones imposed by proposition 12.

Remark 1. Note that

$$\mathbb{P}\left[\tilde{S}_0 = 1, \tilde{S}_1 = 1 \middle| X = \overline{x}, \tilde{U} = \overline{u}\right] = \mathbb{P}\left[Q\left(0, \overline{x}\right) \geq \tilde{V} \middle| X = \overline{x}, \tilde{U} = \overline{u}\right]$$

by the definition of $\tilde{S}_0$ and $\tilde{S}_1$

$$= m_0^S\left(\overline{x}, \overline{u}\right) \tag{A.15}$$

by step 9,

and, similarly, that

$$\mathbb{P}\left[\tilde{S}_0 = 0, \tilde{S}_1 = 1 \middle| X = \overline{x}, \tilde{U} = \overline{u}\right] = \mathbb{P}\left[Q\left(1, \overline{x}\right) \geq \tilde{V} > Q\left(0, \overline{x}\right) \middle| X = \overline{x}, \tilde{U} = \overline{u}\right]$$
$$= \Delta_S\left(\overline{x}, \overline{u}\right). \tag{A.16}$$

Remark 2. Analogously to equation (A.12), I find that

$$\mathbb{E}\left[\tilde{Y}_1^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 0, \tilde{S}_1 = 1\right] = \gamma\left(\overline{x}, \overline{u}\right). \tag{A.17}$$

Remark 3. Combining equations (A.5), (A.12) and (A.15)-(A.17), I have that

$$\mathbb{E}\left[\tilde{Y}_1 \middle| X = x, \tilde{U} = \overline{u}\right] = m_1^Y\left(\overline{x}, \overline{u}\right).$$

A.4.2 Proof under Assumptions 7.1 and 7.2

I, now, prove Proposition 12 under Assumptions 7.1 and 7.2. In particular, I focus on the case $\underline{y}^* > -\infty$ and $\overline{y}^* = +\infty$ (Assumption 7.1) because it is more common in empirical applications. The case $\underline{y}^* = -\infty$ and $\overline{y}^* < +\infty$ (Assumption 7.2) is symmetric.

The proof under Assumption 7.1 is equal to the proof under Assumption 7.3(a). The only difference is that

$$
\begin{aligned}
\delta\left(\overline{x},\overline{u}\right) &\in \left(\underline{\Delta_{Y^*}^{OO}}\left(\overline{x},\overline{u}\right),\overline{\Delta_{Y^*}^{OO}}\left(\overline{x},\overline{u}\right)\right)\\
\Leftrightarrow\quad \alpha\left(\overline{x},\overline{u}\right) &\in \left(\underline{y}^*,\frac{m_1^Y\left(x,u\right)-\underline{y}^*\cdot\Delta_S\left(x,u\right)}{m_0^S\left(x,u\right)}\right)\\
&\subseteq\left(\underline{y}^*,+\infty\right),
\end{aligned}
\tag{A.18}
$$

and that

$$
\begin{aligned}
\alpha\left(\overline{x},\overline{u}\right) &\in \left(\underline{y}^*,\frac{m_1^Y\left(x,u\right)-\underline{y}^*\cdot\Delta_S\left(x,u\right)}{m_0^S\left(x,u\right)}\right)\\
\Leftrightarrow\quad \gamma\left(\overline{x},\overline{u}\right) &\in \left(\underline{y}^*,+\infty\right).
\end{aligned}
\tag{A.19}
$$

A.5 Proof of Proposition 13

This proof is essentially the same proof of Proposition 12 under Assumption 7.3.(a) (appendix A.4.1). Fix $\overline{u}\in[0,1]$, $\overline{x}\in\mathcal{X}$ and $\delta\left(\overline{x},\overline{u}\right)\in\mathbb{R}$ arbitrarily. For brevity, define $\alpha\left(\overline{x},\overline{u}\right):=\delta\left(\overline{x},\overline{u}\right)+\dfrac{m_0^Y\left(\overline{x},\overline{u}\right)}{m_0^S\left(\overline{x},\overline{u}\right)}$ and $\gamma\left(\overline{x},\overline{u}\right):=\dfrac{m_1^Y\left(\overline{x},\overline{u}\right)-\alpha\left(\overline{x},\overline{u}\right)\cdot m_0^S\left(\overline{x},\overline{u}\right)}{\Delta_S\left(\overline{x},\overline{u}\right)}$. Note that $\alpha\left(\overline{x},\overline{u}\right)\in\mathbb{R}=\mathcal{Y}^*$ and $\gamma\left(\overline{x},\overline{u}\right)\in\mathbb{R}=\mathcal{Y}^*$.

I define the random variables $\left(\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}\right)$ using the joint cumulative distribution function $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}$ described by steps 1-12 in Appendix A.4.1 for the case of convex support $\mathcal{Y}^*$. Note that equation (19) is trivially true when $\mathcal{Y}^*=\mathbb{R}$. Moreover, equations (18) and (20) are valid by the argument described in steps 13-17 in Appendix A.4.1.

I can, then, conclude that proposition 13 is true. ■

A.6 Proof of Proposition 15

This proof is essentially the same proof of Propositions 12 and 13 (Appendices A.4 and A.5). Fix $\overline{u}\in[0,1]$, $\overline{x}\in\mathcal{X}$ and $\delta\left(\overline{x},\overline{u}\right)\in\left(\underline{\Delta_{Y^*}^{OO}}\left(\overline{x},\overline{u}\right),\overline{\Delta_{Y^*}^{OO}}\left(\overline{x},\overline{u}\right)\right)$ arbitrarily. For brevity, define $\alpha\left(\overline{x},\overline{u}\right):=\delta\left(\overline{x},\overline{u}\right)+\dfrac{m_0^Y\left(\overline{x},\overline{u}\right)}{m_0^S\left(\overline{x},\overline{u}\right)}$ and $\gamma\left(\overline{x},\overline{u}\right):=\dfrac{m_1^Y\left(\overline{x},\overline{u}\right)-\alpha\left(\overline{x},\overline{u}\right)\cdot m_0^S\left(\overline{x},\overline{u}\right)}{\Delta_S\left(\overline{x},\overline{u}\right)}$. The only difference from the previous proofs is that, now,

$$\begin{aligned}
\mathbb{E}\left[\tilde{Y}_1^* \middle| X=\overline{x}, \tilde{U}=\overline{u}, \tilde{S}_0=1, \tilde{S}_1=1\right] &= \alpha\left(\overline{x}, \overline{u}\right) \\
&\qquad \text{by equation (A.12)} \\
&\geq \frac{m_1^Y\left(\overline{x}, \overline{u}\right)}{m_1^S\left(\overline{x}, \overline{u}\right)} \qquad\qquad \text{(A.20)} \\
&\qquad \text{because } \delta\left(\overline{x}, \overline{u}\right) \geq \underline{\Delta_{Y^*}^{OO}}\left(\overline{x}, \overline{u}\right)
\end{aligned}$$

and that

$$\begin{aligned}
\mathbb{E}\left[\tilde{Y}_1^* \middle| X=\overline{x}, \tilde{U}=\overline{u}, \tilde{S}_0=0, \tilde{S}_1=1\right] &= \gamma\left(\overline{x}, \overline{u}\right) \\
&\qquad \text{by equation (A.17)} \\
&= \frac{m_1^Y\left(\overline{x}, \overline{u}\right) - \alpha\left(\overline{x}, \overline{u}\right) \cdot m_0^S\left(\overline{x}, \overline{u}\right)}{\Delta_S\left(\overline{x}, \overline{u}\right)} \\
&\leq \frac{m_1^Y\left(\overline{x}, \overline{u}\right) - \dfrac{m_1^Y\left(\overline{x}, \overline{u}\right)}{m_1^S\left(\overline{x}, \overline{u}\right)} \cdot m_0^S\left(\overline{x}, \overline{u}\right)}{\Delta_S\left(\overline{x}, \overline{u}\right)} \\
&\qquad \text{by equation (A.20)} \\
&= \frac{m_1^Y\left(\overline{x}, \overline{u}\right)}{m_1^S\left(\overline{x}, \overline{u}\right)},
\end{aligned}$$

implying that the model restriction (31) holds.

A.7 Proof of Equations (34) and (35)

I first prove that equation (34) holds. For any $A \in \{Y, S\}$, observe that

$$\begin{aligned}
\mathbb{E}\left[A \,|X=x, P\left(W\right)=p, D=0\right] &= \mathbb{E}\left[A_0 \,|X=x, P\left(W\right)=p, D=0\right] \\
&= \mathbb{E}\left[A_0 \,|X=x, P\left(W\right)=p, P\left(W\right)<U\right] \\
&\qquad \text{by equation (1)} \\
&= \mathbb{E}\left[A_0 \,|X=x, P\left(W\right)=p, p<U\right] \\
&= \mathbb{E}\left[A_0 \,|X=x, p<U\right]
\end{aligned}$$

$$\text{by assumption (1)}$$

$$= \frac{\mathbb{E}\left[\mathbf{1}\left\{p < U\right\} \cdot A_0 \left| X = x\right.\right]}{\mathbb{P}\left[p < U \left| X = x\right.\right]}$$

$$\text{by the definition of conditional expectation}$$

$$= \frac{\mathbb{E}\left[\mathbf{1}\left\{p < U\right\} \cdot A_0 \left| X = x\right.\right]}{1 - p}$$

$$\text{by the normalization } U \left| X \right. \sim \text{Uniform}\left[0, 1\right]$$

$$= \frac{\mathbb{E}\left[\mathbf{1}\left\{p < U\right\} \cdot \mathbb{E}\left[A_0 \left| X = x, U = u\right.\right] \left| X = x\right.\right]}{1 - p}$$

$$\text{by the Law of Iterated Expectations}$$

$$= \frac{\int_p^1 m_0^A\left(x, u\right) \mathrm{d}u}{1 - p}$$

$$\text{by the normalization } U \left| X \right. \sim \text{Uniform}\left[0, 1\right],$$

implying that

$$\begin{aligned}
\frac{\partial \mathbb{E}\left[A \left| X = x, P\left(W\right) = p, D = 0\right.\right]}{\partial p} &= \frac{-m_0^A\left(x, p\right)}{1 - p} + \frac{\mathbb{E}\left[\mathbf{1}\left\{p < U\right\} \cdot A_0 \left| X = x\right.\right]}{\left(1 - p\right)^2} \\
&= \frac{-m_0^A\left(x, p\right)}{1 - p} + \frac{\mathbb{E}\left[\mathbf{1}\left\{p < U\right\} \cdot A_0 \left| X = x\right.\right]}{\left(1 - p\right) \cdot \mathbb{P}\left[p < U \left| X = x\right.\right]} \\
&\qquad \text{by the normalization } U \left| X \right. \sim \text{Uniform}\left[0, 1\right] \\
&= \frac{-m_0^A\left(x, p\right)}{1 - p} + \frac{\mathbb{E}\left[A \left| X = x, P\left(W\right) = p, D = 0\right.\right]}{1 - p}
\end{aligned}$$

Rearranging the last expression, I can derive equation (34):

$$\begin{aligned}
m_0^A\left(x, p\right) = {} & \mathbb{E}\left[A \left| X = x, P\left(W\right) = p, D = 0\right.\right] \\
& - \frac{\partial \mathbb{E}\left[A \left| X = x, P\left(W\right) = p, D = 0\right.\right]}{\partial p} \cdot \left(1 - p\right).
\end{aligned}$$

Equation (35) is derived in an analogous way using $\mathbb{E}\left[A \left| X = x, P\left(W\right) = p, D = 1\right.\right]$ and its derivative with respect to the propensity score. ■

A.8 Proof of Equations (45) and (46)

We first prove that equation (45) holds. For any $A \in \{Y, S\}$, observe that

$$\mathbb{E}\left[A \,|X = x, P(W) = p_n, D = 0\right] = \frac{\int_{p_n}^{1} m_0^A(x, u)\, \mathrm{d}u}{1 - p_n}$$

according to Appendix A.7

$$= \frac{\int_{p_n}^{1} M^A\left(u, \boldsymbol{\theta}_{x,0}^A\right) \mathrm{d}u}{1 - p_n}$$

by equation (43).

Equation (46) is derived in an analogous way using $\mathbb{E}\left[A \,|X = x, P(W) = p_n, D = 1\right]$. ■

A.9 Connecting OLS Model (53) to the Minimization Problem (44)

Note that, for any $z \in \{0, 1\}$,

$$\begin{aligned}
\frac{\int_{P(z)}^{1} M^A\left(u, \tilde{\boldsymbol{\theta}}_0^A\right) \mathrm{d}u}{1 - P(z)} &= \frac{\int_{P(z)}^{1} \left(\theta_{0,0}^A \cdot (1-u) + \theta_{0,1}^A \cdot u\right) \mathrm{d}u}{1 - P(z)} \\
&= \frac{\theta_{0,0}^A + \theta_{0,1}^A}{2} + \frac{-\theta_{0,0}^A + \theta_{0,1}^A}{2} \cdot P(z) \\
&= a_0^A + b_0^A \cdot P(z),
\end{aligned} \tag{A.21}$$

where $a_0^A := \dfrac{\theta_{0,0}^A + \theta_{0,1}^A}{2}$ and $b_0^A := \dfrac{-\theta_{0,0}^A + \theta_{0,1}^A}{2}$, and

$$\begin{aligned}
\frac{\int_{0}^{P(z)} M^A\left(u, \tilde{\boldsymbol{\theta}}_1^A\right) \mathrm{d}u}{P(z)} &= \frac{\int_{0}^{P(z)} \left(\theta_{1,0}^A \cdot (1-u) + \theta_{1,1}^A \cdot u\right) \mathrm{d}u}{P(z)} \\
&= \theta_{1,0}^A + \frac{-\theta_{1,0}^A + \theta_{1,1}^A}{2} \cdot P(z) \\
&= a_1^A + b_1^A \cdot P(z),
\end{aligned} \tag{A.22}$$

where $a_1^A := \theta_{1,0}^A$ and $b_1^A := \dfrac{-\theta_{1,0}^A + \theta_{1,1}^A}{2}$.

When I combine equations (44), (A.21) and (A.22), I find the OLS model given by equation

(53). Moreover, by solving the linear system given by $a_0^A = \frac{\theta_{0,0}^A + \theta_{0,1}^A}{2}$, $b_0^A = \frac{-\theta_{0,0}^A + \theta_{0,1}^A}{2}$, $a_1^A = \theta_{1,0}^A$ and $b_1^A = \frac{-\theta_{1,0}^A + \theta_{1,1}^A}{2}$, I find that $\theta_{0,0}^A = a_0^A - b_0^A$, $\theta_{0,1}^A = a_0^A + b_0^A$, $\theta_{1,0}^A = a_1^A$, $\theta_{1,1}^A = a_1^A + 2 \cdot b_1^A$.

B Bounds on the MTE for the Observed-only-when-treated Sub-population

Here, I use the same notation of section 3 and I am interested in the following target parameter: $m_1^{NO}(x,u) \coloneqq \mathbb{E}[Y_1^* | X = x, U = u, S_0 = 0, S_1 = 1]$, which is equal to Δ_Y^{NO} according to equation (A.4). Following the same steps of the proof of Proposition 10, I can show that:

Corollary B.1 *Suppose that the $m_0^Y(x,u)$, $m_1^Y(x,u)$, $m_0^S(x,u)$ and $\Delta_S(x,u)$ are point-identified.*

Under assumptions 1-6, 7.1 and 8, the bounds on $m_1^{NO}(x,u)$ are given by

$$\underline{m_1^{NO}}(x,u) \coloneqq \underline{y}^* \leq m_1^{NO}(x,u) \leq \frac{m_1^Y(x,u) - \underline{y}^* \cdot m_0^S(x,u)}{\Delta_S(x,u)} \eqqcolon \overline{m_1^{NO}}(x,u). \tag{B.1}$$

Under assumptions 1-6, 7.2 and 8, the bounds on $m_1^{NO}(x,u)$ are given by

$$\underline{m_1^{NO}}(x,u) \coloneqq \frac{m_1^Y(x,u) - \overline{y}^* \cdot m_0^S(x,u)}{\Delta_S(x,u)} \leq m_1^{NO}(x,u) \leq \overline{y}^* \eqqcolon \overline{m_1^{NO}}(x,u). \tag{B.2}$$

Under assumptions 1-6, 7.3 (sub-case (a) or (b)) and 8, the bounds on $m_1^{NO}(x,u)$ are given by

$$\underline{m_1^{NO}}(x,u) \coloneqq \frac{m_1^Y(x,u) - \overline{y}^* \cdot m_0^S(x,u)}{\Delta_S(x,u)} \leq m_1^{NO}(x,u) \leq \frac{m_1^Y(x,u) - \underline{y}^* \cdot m_0^S(x,u)}{\Delta_S(x,u)} \eqqcolon \overline{m_1^{NO}}(x,u). \tag{B.3}$$

Following the same proof of proposition 12 (see Remark 2 at the end of Appendix A.4.1), I can also show that:

Proposition B.2 *Suppose that the functions m_0^Y, m_1^Y, m_0^S and Δ_S are point-identified at every pair $(x,u) \in \mathcal{X} \times [0,1]$. Under assumptions 1-6, 7 (sub-cases 1, 2, 3(a) or 3(b)) and 8, the bounds $\underline{m_1^{NO}}$ and $\overline{m_1^{NO}}$, given by Proposition B.1, are point-wise sharp, i.e., for any $\overline{u} \in [0,1]$, $\overline{x} \in \mathcal{X}$ and $\gamma(\overline{x},\overline{u}) \in \left(\underline{m_1^{NO}}(\overline{x},\overline{u}), \overline{m_1^{NO}}(\overline{x},\overline{u})\right)$, there exist random variables*

$\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)$ *such that*

$$\tilde{m}_1^{NO}\left(\overline{x}, \overline{u}\right) \coloneqq \mathbb{E}\left[\tilde{Y}_1^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 0, \tilde{S}_1 = 1\right] = \gamma\left(\overline{x}, \overline{u}\right), \tag{B.4}$$

$$\mathbb{P}\left[\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{V}\right) \in \mathcal{Y}^* \times \mathcal{Y}^* \times [0,1] \middle| X = \overline{x}, \tilde{U} = u\right] = 1 \text{ for any } u \in [0,1], \tag{B.5}$$

and

$$F_{\tilde{Y},\tilde{D},\tilde{S},Z,X}\left(y, d, s, z, \overline{x}\right) = F_{Y,D,S,Z,X}\left(y, d, s, z, \overline{x}\right) \tag{B.6}$$

for any $(y, d, s, z) \in \mathbb{R}^4$*, where* $\tilde{D} \coloneqq \mathbf{1}\left\{P\left(X, Z\right) \geq \tilde{U}\right\}$*,* $\tilde{S}_0 = \mathbf{1}\left\{Q\left(0, X\right) \geq \tilde{V}\right\}$*,* $\tilde{S}_1 = \mathbf{1}\left\{Q\left(1, X\right) \geq \tilde{V}\right\}$*,* $\tilde{Y}_0 = \tilde{S}_0 \cdot \tilde{Y}_0^*$*,* $\tilde{Y}_1 = \tilde{S}_1 \cdot \tilde{Y}_1^*$ *and* $\tilde{Y} = \tilde{D} \cdot \tilde{Y}_1 + \left(1 - \tilde{D}\right) \cdot \tilde{Y}_0$*.*

Finally, following the same proof of proposition 13, I can also show that:

Proposition B.3 *Suppose that the functions* m_0^Y*,* m_1^Y*,* m_0^S *and* Δ_S *are point-identified at every pair* $(x, u) \in \mathcal{X} \times [0,1]$*. Impose assumptions 1-6 and 8. If* $\mathcal{Y}^* = \mathbb{R}$*, then, for any* $\overline{u} \in [0,1]$*,* $\overline{x} \in \mathcal{X}$ *and* $\gamma\left(\overline{x}, \overline{u}\right) \in \mathbb{R}$*, there exist random variables* $\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)$ *such that*

$$\tilde{m}_1^{NO}\left(\overline{x}, \overline{u}\right) \coloneqq \mathbb{E}\left[\tilde{Y}_1^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 0, \tilde{S}_1 = 1\right] = \gamma\left(\overline{x}, \overline{u}\right), \tag{B.7}$$

$$\mathbb{P}\left[\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{V}\right) \in \mathcal{Y}^* \times \mathcal{Y}^* \times [0,1] \middle| X = \overline{x}, \tilde{U} = u\right] = 1 \text{ for any } u \in [0,1], \tag{B.8}$$

and

$$F_{\tilde{Y},\tilde{D},\tilde{S},Z,X}\left(y, d, s, z, \overline{x}\right) = F_{Y,D,S,Z,X}\left(y, d, s, z, \overline{x}\right) \tag{B.9}$$

for any $(y, d, s, z) \in \mathbb{R}^4$*, where* $\tilde{D} \coloneqq \mathbf{1}\left\{P\left(X, Z\right) \geq \tilde{U}\right\}$*,* $\tilde{S}_0 = \mathbf{1}\left\{Q\left(0, X\right) \geq \tilde{V}\right\}$*,* $\tilde{S}_1 = \mathbf{1}\left\{Q\left(1, X\right) \geq \tilde{V}\right\}$*,* $\tilde{Y}_0 = \tilde{S}_0 \cdot \tilde{Y}_0^*$*,* $\tilde{Y}_1 = \tilde{S}_1 \cdot \tilde{Y}_1^*$ *and* $\tilde{Y} = \tilde{D} \cdot \tilde{Y}_1 + \left(1 - \tilde{D}\right) \cdot \tilde{Y}_0$*.*

C Negative Treatment Effect on the Selection Indicator

Even when sample selection is monotone (equation (2)), Assumption 8 may be invalid in some empirical applications. In particular, it might be the case that the following assumption holds:

Assumption C.1 *Treatment has a negative effect on the sample selection indicator for all individuals, i.e.,* $Q(0,x) > Q(1,x) > 0$ *for any* $x \in \mathcal{X}$.

I stress that this assumption is testable according to Machado et al. (2018).

With obvious modifications to the proofs of Corollary 11 and Propositions 12 and 13 (see the proofs of Propositions D.3 and D.4), I can show that the target parameter in section 3 can be bounded, that its bounds are sharp and that it is impossible to derive bounds for the target parameter with only assumptions 1-6 and C.1. First, I state a result that is analogous to Corollary 11.

Corollary C.2 *Fix* $u \in [0,1]$ *and* $x \in \mathcal{X}$ *arbitrarily. Suppose that the* $m_0^Y(x,u)$, $m_1^Y(x,u)$, $m_0^S(x,u)$ *and* $\Delta_S(x,u)$ *are point-identified.*

Under Assumptions 1-6, 7.1 and C.1, the bounds on $\Delta_{Y^*}^{OO}(x,u)$ *are given by*

$$\Delta_{Y^*}^{OO}(x,u) \geq \frac{m_1^Y(x,u)}{m_1^S(x,u)} - \frac{m_0^Y(x,u) - \underline{y}^* \cdot (-\Delta_S(x,u))}{m_1^S(x,u)} =: \underline{\Lambda_{Y^*}^{OO}}(x,u) \tag{C.1}$$

and that

$$\Delta_{Y^*}^{OO}(x,u) \leq \frac{m_1^Y(x,u)}{m_1^S(x,u)} - \underline{y}^* =: \overline{\Lambda_{Y^*}^{OO}}(x,u). \tag{C.2}$$

Under Assumptions 1-6, 7.2 and C.1, the bounds on $\Delta_{Y^*}^{OO}(x,u)$ *are given by*

$$\Delta_{Y^*}^{OO}(x,u) \geq \frac{m_1^Y(x,u)}{m_1^S(x,u)} - \overline{y}^* =: \underline{\Lambda_{Y^*}^{OO}}(x,u) \tag{C.3}$$

and that

$$\Delta_{Y^*}^{OO}(x,u) \leq \frac{m_1^Y(x,u)}{m_1^S(x,u)} - \frac{m_0^Y(x,u) - \overline{y}^* \cdot (-\Delta_S(x,u))}{m_1^S(x,u)} =: \overline{\Lambda_{Y^*}^{OO}}(x,u). \tag{C.4}$$

Under Assumptions 1-6, 7.3 (sub-case (a) or (b)) and C.1, the bounds on $\Delta_{Y^*}^{OO}(x,u)$ are given by

$$\Delta_{Y^*}^{OO}(x,u) \geq \frac{m_1^Y(x,u)}{m_1^S(x,u)} - \min\left\{\frac{m_0^Y(x,u) - \underline{y}^* \cdot (-\Delta_S(x,u))}{m_1^S(x,u)}, \overline{y}^*\right\} =: \underline{\Lambda_{Y^*}^{OO}}(x,u) \tag{C.5}$$

and that

$$\Delta_{Y^*}^{OO}(x,u) \leq \frac{m_1^Y(x,u)}{m_1^S(x,u)} - \max\left\{\frac{m_0^Y(x,u) - \overline{y}^* \cdot (-\Delta_S(x,u))}{m_1^S(x,u)}, \underline{y}^*\right\} =: \overline{\Lambda_{Y^*}^{OO}}(x,u). \tag{C.6}$$

Second, I state a result that is analogous to Proposition 12.

Proposition C.3 *Suppose that the functions m_0^Y, m_1^Y, m_0^S and Δ_S are point-identified at every pair $(x,u) \in \mathcal{X} \times [0,1]$. Under Assumptions 1-6, 7 (sub-cases 1, 2, 3(a) or 3(b)) and C.1, the bounds $\underline{\Lambda_{Y^*}^{OO}}$ and $\overline{\Lambda_{Y^*}^{OO}}$, given by Proposition C.2, are point-wise sharp, i.e., for any $\overline{u} \in [0,1]$, $\overline{x} \in \mathcal{X}$ and $\delta(\overline{x},\overline{u}) \in \left(\underline{\Lambda_{Y^*}^{OO}}(\overline{x},\overline{u}), \overline{\Lambda_{Y^*}^{OO}}(\overline{x},\overline{u})\right)$, there exist random variables $\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)$ such that*

$$\Delta_{\tilde{Y}^*}^{OO}(\overline{x},\overline{u}) \coloneqq \mathbb{E}\left[\tilde{Y}_1^* - \tilde{Y}_0^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right] = \delta(\overline{x},\overline{u}), \tag{C.7}$$

$$\mathbb{P}\left[\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{V}\right) \in \mathcal{Y}^* \times \mathcal{Y}^* \times [0,1] \middle| X = \overline{x}, \tilde{U} = u\right] = 1 \text{ for any } u \in [0,1], \tag{C.8}$$

and

$$F_{\tilde{Y},\tilde{D},\tilde{S},Z,X}(y,d,s,z,\overline{x}) = F_{Y,D,S,Z,X}(y,d,s,z,\overline{x}) \tag{C.9}$$

for any $(y,d,s,z) \in \mathbb{R}^4$, where $\tilde{D} \coloneqq \mathbf{1}\left\{P(X,Z) \geq \tilde{U}\right\}$, $\tilde{S}_0 = \mathbf{1}\left\{Q(0,X) \geq \tilde{V}\right\}$, $\tilde{S}_1 = \mathbf{1}\left\{Q(1,X) \geq \tilde{V}\right\}$, $\tilde{Y}_0 = \tilde{S}_0 \cdot \tilde{Y}_0^*$, $\tilde{Y}_1 = \tilde{S}_1 \cdot \tilde{Y}_1^*$ and $\tilde{Y} = \tilde{D} \cdot \tilde{Y}_1 + \left(1 - \tilde{D}\right) \cdot \tilde{Y}_0$.

Finally, I state a result that is analogous to Proposition 13.

Proposition C.4 *Suppose that the functions m_0^Y, m_1^Y, m_0^S and Δ_S are point-identified at every pair $(x,u) \in \mathcal{X} \times [0,1]$. Impose Assumptions 1-6 and C.1. If $\mathcal{Y}^* = \mathbb{R}$, then, for any*

$\overline{u} \in [0,1]$, $\overline{x} \in \mathcal{X}$ and $\delta(\overline{x}, \overline{u}) \in \mathbb{R}$, there exist random variables $\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)$ such that

$$\Delta_{\tilde{Y}^*}^{OO}(\overline{x}, \overline{u}) \coloneqq \mathbb{E}\left[\tilde{Y}_1^* - \tilde{Y}_0^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right] = \delta(\overline{x}, \overline{u}), \tag{C.10}$$

$$\mathbb{P}\left[\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{V}\right) \in \mathcal{Y}^* \times \mathcal{Y}^* \times [0,1] \middle| X = \overline{x}, \tilde{U} = u\right] = 1 \text{ for any } u \in [0,1], \tag{C.11}$$

and

$$F_{\tilde{Y}, \tilde{D}, \tilde{S}, Z, X}(y, d, s, z, \overline{x}) = F_{Y, D, S, Z, X}(y, d, s, z, \overline{x}) \tag{C.12}$$

for any $(y, d, s, z) \in \mathbb{R}^4$, where $\tilde{D} \coloneqq \mathbf{1}\left\{P(X, Z) \geq \tilde{U}\right\}$, $\tilde{S}_0 = \mathbf{1}\left\{Q(0, X) \geq \tilde{V}\right\}$, $\tilde{S}_1 = \mathbf{1}\left\{Q(1, X) \geq \tilde{V}\right\}$, $\tilde{Y}_0 = \tilde{S}_0 \cdot \tilde{Y}_0^*$, $\tilde{Y}_1 = \tilde{S}_1 \cdot \tilde{Y}_1^*$ and $\tilde{Y} = \tilde{D} \cdot \tilde{Y}_1 + \left(1 - \tilde{D}\right) \cdot \tilde{Y}_0$.

D Monotone Sample Selection

Depending on the results of the test proposed by Machado et al. (2018), a researcher may want to be agnostic about the direction of the monotone selection problem and impose only equation (2), while ruling out uninteresting cases. In such situation, it is reasonable to assume:

Assumption D.1 *Treatment has a monotone effect on the sample selection indicator for all individuals, i.e.,* ***either*** *(i)* $Q(1,x) > Q(0,x) > 0$ *for any* $x \in \mathcal{X}$ ***or*** *(ii)* $Q(0,x) > Q(1,x) > 0$ *for any* $x \in \mathcal{X}$.

I stress that assumption D.1 only strengthens equation (2) by ruling out the theoretically uninteresting cases mentioned after Assumption (8).

By combining Corollaries 11 and C.2, I find that:

Corollary D.2 *Fix* $u \in [0,1]$ *and* $x \in \mathcal{X}$ *arbitrarily. Suppose that the* $m_0^Y(x,u)$, $m_1^Y(x,u)$, $m_0^S(x,u)$ *and* $\Delta_S(x,u)$ *are point-identified. Under Assumptions 1-6, 7 and D.1, the bounds on* $\Delta_{Y^*}^{OO}(x,u)$ *are given by*

$$\begin{aligned}\underline{\Upsilon_{Y^*}^{OO}}(x,u) &:= \min\left\{\underline{\Delta_{Y^*}^{OO}}(x,u), \underline{\Lambda_{Y^*}^{OO}}(x,u)\right\} \\ &\leq \Delta_{Y^*}^{OO}(x,u) \\ &\leq \max\left\{\overline{\Delta_{Y^*}^{OO}}(x,u), \overline{\Lambda_{Y^*}^{OO}}(x,u)\right\} =: \overline{\Upsilon_{Y^*}^{OO}}(x,u)\end{aligned} \tag{D.1}$$

Most importantly, those bounds are also point-wise sharp:[14]

Proposition D.3 *Suppose that the functions* m_0^Y, m_1^Y, m_0^S *and* Δ_S *are point-identified at every pair* $(x,u) \in \mathcal{X} \times [0,1]$. *Under Assumptions 1-6, 7 (sub-cases 1, 2, 3(a) or 3(b)) and D.1, the bounds* $\underline{\Upsilon_{Y^*}^{OO}}$ *and* $\overline{\Upsilon_{Y^*}^{OO}}$, *given by Corollary D.2, are point-wise sharp, i.e., for any* $\overline{u} \in [0,1]$, $\overline{x} \in \mathcal{X}$ *and* $\delta(\overline{x},\overline{u}) \in \left(\underline{\Upsilon_{Y^*}^{OO}}(\overline{x},\overline{u}), \overline{\Upsilon_{Y^*}^{OO}}(\overline{x},\overline{u})\right)$, *there exist random variables*

[14]The proof of propositions D.3 and D.4 are located at the end of Appendix D.

$\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)$ *such that*

$$\Delta_{\tilde{Y}^*}^{OO}\left(\overline{x}, \overline{u}\right) := \mathbb{E}\left[\tilde{Y}_1^* - \tilde{Y}_0^* \left| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right.\right] = \delta\left(\overline{x}, \overline{u}\right), \tag{D.2}$$

$$\mathbb{P}\left[\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{V}\right) \in \mathcal{Y}^* \times \mathcal{Y}^* \times [0,1] \right| X = \overline{x}, \tilde{U} = u\Big] = 1 \text{ for any } u \in [0,1], \tag{D.3}$$

and

$$F_{\tilde{Y}, \tilde{D}, \tilde{S}, Z, X}\left(y, d, s, z, \overline{x}\right) = F_{Y, D, S, Z, X}\left(y, d, s, z, \overline{x}\right) \tag{D.4}$$

for any $(y, d, s, z) \in \mathbb{R}^4$, *where* $\tilde{D} := \mathbf{1}\left\{P(X, Z) \geq \tilde{U}\right\}$, $\tilde{S}_0 = \mathbf{1}\left\{Q(0, X) \geq \tilde{V}\right\}$, $\tilde{S}_1 = \mathbf{1}\left\{Q(1, X) \geq \tilde{V}\right\}$, $\tilde{Y}_0 = \tilde{S}_0 \cdot \tilde{Y}_0^*$, $\tilde{Y}_1 = \tilde{S}_1 \cdot \tilde{Y}_1^*$ *and* $\tilde{Y} = \tilde{D} \cdot \tilde{Y}_1 + \left(1 - \tilde{D}\right) \cdot \tilde{Y}_0$.

Finally, I state an impossibility result that is analogous to Proposition 13.

Proposition D.4 *Suppose that the functions* m_0^Y, m_1^Y, m_0^S *and* Δ_S *are point-identified at every pair* $(x, u) \in \mathcal{X} \times [0,1]$. *Impose assumptions 1-6 and D.1. If* $\mathcal{Y}^* = \mathbb{R}$, *then, for any* $\overline{u} \in [0,1]$, $\overline{x} \in \mathcal{X}$ *and* $\delta\left(\overline{x}, \overline{u}\right) \in \mathbb{R}$, *there exist random variables* $\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)$ *such that*

$$\Delta_{\tilde{Y}^*}^{OO}\left(\overline{x}, \overline{u}\right) := \mathbb{E}\left[\tilde{Y}_1^* - \tilde{Y}_0^* \left| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right.\right] = \delta\left(\overline{x}, \overline{u}\right), \tag{D.5}$$

$$\mathbb{P}\left[\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{V}\right) \in \mathcal{Y}^* \times \mathcal{Y}^* \times [0,1] \right| X = \overline{x}, \tilde{U} = u\Big] = 1 \text{ for any } u \in [0,1], \tag{D.6}$$

and

$$F_{\tilde{Y}, \tilde{D}, \tilde{S}, Z, X}\left(y, d, s, z, \overline{x}\right) = F_{Y, D, S, Z, X}\left(y, d, s, z, \overline{x}\right) \tag{D.7}$$

for any $(y, d, s, z) \in \mathbb{R}^4$, *where* $\tilde{D} := \mathbf{1}\left\{P(X, Z) \geq \tilde{U}\right\}$, $\tilde{S}_0 = \mathbf{1}\left\{Q(0, X) \geq \tilde{V}\right\}$, $\tilde{S}_1 = \mathbf{1}\left\{Q(1, X) \geq \tilde{V}\right\}$, $\tilde{Y}_0 = \tilde{S}_0 \cdot \tilde{Y}_0^*$, $\tilde{Y}_1 = \tilde{S}_1 \cdot \tilde{Y}_1^*$ *and* $\tilde{Y} = \tilde{D} \cdot \tilde{Y}_1 + \left(1 - \tilde{D}\right) \cdot \tilde{Y}_0$.

Proof of Proposition D.3. I only prove proposition D.3 under assumption 7.3 (subcases (a) and (b)).The proofs of proposition D.3 under assumptions 7.1 and 7.2 are trivial modifications of the proof presented below.

Fix $\overline{u} \in [0,1]$, $\overline{x} \in \mathcal{X}$ and $\delta\left(\overline{x}, \overline{u}\right) \in \left(\underline{\Upsilon_{Y^*}^{OO}}\left(\overline{x}, \overline{u}\right), \overline{\Upsilon_{Y^*}^{OO}}\left(\overline{x}, \overline{u}\right)\right)$ arbitrarily. For brevity, define

$$\alpha\left(\overline{x},\overline{u}\right) \coloneqq \mathbf{1}\left\{Q\left(1,x\right) > Q\left(0,x\right)\right\} \cdot \left(\delta\left(\overline{x},\overline{u}\right) + \frac{m_0^Y\left(\overline{x},\overline{u}\right)}{m_0^S\left(\overline{x},\overline{u}\right)}\right) + \mathbf{1}\left\{Q\left(1,x\right) < Q\left(0,x\right)\right\} \cdot \left(-\delta\left(\overline{x},\overline{u}\right) + \frac{m_1^Y\left(\overline{x},\overline{u}\right)}{m_1^S\left(\overline{x},\overline{u}\right)}\right),$$

$$\gamma\left(\overline{x},\overline{u}\right) \coloneqq \mathbf{1}\left\{Q\left(1,x\right) > Q\left(0,x\right)\right\} \cdot \left(\frac{m_1^Y\left(\overline{x},\overline{u}\right) - \alpha\left(\overline{x},\overline{u}\right) \cdot m_0^S\left(\overline{x},\overline{u}\right)}{\Delta_S\left(\overline{x},\overline{u}\right)}\right) + \mathbf{1}\left\{Q\left(1,x\right) < Q\left(0,x\right)\right\} \cdot \left(\frac{m_0^Y\left(\overline{x},\overline{u}\right) - \alpha\left(\overline{x},\overline{u}\right) \cdot m_1^S\left(\overline{x},\overline{u}\right)}{-\Delta_S\left(\overline{x},\overline{u}\right)}\right),$$

$$\underline{Q}\left(x\right) = \min\left\{Q\left(0,x\right),Q\left(1,x\right)\right\},$$

$$\overline{Q}\left(x\right) = \max\left\{Q\left(0,x\right),Q\left(1,x\right)\right\},$$

$$\underline{m}^S\left(x,\overline{u}\right) = \min\left\{m_0^S\left(x,\overline{u}\right), m_1^S\left(x,\overline{u}\right)\right\} \text{ for any } x \in \mathcal{X},$$

and

$$\overline{m}^S\left(x,\overline{u}\right) = \max\left\{m_0^S\left(x,\overline{u}\right), m_1^S\left(x,\overline{u}\right)\right\} \text{ for any } x \in \mathcal{X}.$$

Note that

$$\alpha\left(\overline{x},\overline{u}\right) \in \left(\underline{y}^*,\overline{y}^*\right), \tag{D.8}$$

and that

$$\gamma\left(\overline{x},\overline{u}\right) \in \left(\underline{y}^*,\overline{y}^*\right). \tag{D.9}$$

The strategy of this proof consists of defining random variables $\left(\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}\right)$ through their joint cumulative distribution function $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}$ and, then, checking that equations (D.2), (D.3) and (D.4) are satisfied. I fix $(y_0,y_1,u,v,z,x) \in \mathbb{R}^6$ and define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}$ in twelve steps:

Step 1. For $x \notin \mathcal{X}$, $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}\left(y_0,y_1,u,v,z,x\right) = F_{Y_0^*,Y_1^*,U,V,Z,X}\left(y_0,y_1,u,v,z,x\right)$.

Step 2. From now on, consider $x \in \mathcal{X}$. Since

$$F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}\left(y_0,y_1,u,v,z,x\right) = F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0,y_1,u,v,z\,|x\right)\cdot F_X\left(x\right),$$

it suffices to define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0,y_1,u,v,z\,|x\right)$. Moreover, I impose

$$Z \perp\!\!\!\perp \left(\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}\right)|X$$

by writing

$$F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0,y_1,u,v,z\,|x\right) = F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}|X}\left(y_0,y_1,u,v\,|x\right)\cdot F_{Z|X}\left(z\,|x\right),$$

implying that it is sufficient to define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}|X}\left(y_0,y_1,u,v\,|x\right)$.

Step 3. For $u \notin [0,1]$, I define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}|X}\left(y_0,y_1,u,v\,|x\right) = F_{Y_0^*,Y_1^*,U,V|X}\left(y_0,y_1,u,v\,|x\right)$.

Step 4. From now on, consider $u \in [0,1]$. Since

$$F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}|X}\left(y_0,y_1,u,v\,|x\right) = F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}|X,\tilde{U}}\left(y_0,y_1,v\,|x,u\right)\cdot F_{\tilde{U}|X}\left(u\,|x\right),$$

it suffices to define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}|X,\tilde{U}}\left(y_0,y_1,v\,|x,u\right)$ and $F_{\tilde{U}|X}\left(u\,|x\right)$.

Step 5. I define $F_{\tilde{U}|X}\left(u\,|x\right) = F_{U|X}\left(u\,|x\right) = u$.

Step 6. For any $u \neq \overline{u}$, I define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}|X,\tilde{U}}\left(y_0,y_1,v\,|x,u\right) = F_{Y_0^*,Y_1^*,V|X,U}\left(y_0,y_1,v\,|x,u\right)$.

Step 7. For any $v \notin [0,1]$, I define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}|X,\tilde{U}}\left(y_0,y_1,v\,|x,\overline{u}\right) = F_{Y_0^*,Y_1^*,V|X,U}\left(y_0,y_1,v\,|x,\overline{u}\right)$.

Step 8. From now on, assume that $v \in [0,1]$. Since

$$F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}|X,\tilde{U}}\left(y_0,y_1,v\,|x,\overline{u}\right) = F_{\tilde{Y}_0^*,\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}\left(y_0,y_1\,|x,\overline{u},v\right)\cdot F_{\tilde{V}|X,\tilde{U}}\left(v\,|x,\overline{u}\right),$$

it is sufficient to define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}\left(y_0,y_1\,|x,\overline{u},v\right)$ and $F_{\tilde{V}|X,\tilde{U}}\left(v\,|x,\overline{u}\right)$.

Step 9. I define

$$F_{\tilde{V}|X,\tilde{U}}\left(v\,|x,\overline{u}\right)=\begin{cases}\underline{m}^{S}\left(x,\overline{u}\right)\cdot\dfrac{v}{\underline{Q}\left(x\right)} & \text{if } v\leq\underline{Q}\left(x\right)\\ \\ \underline{m}^{S}\left(x,\overline{u}\right)+\left(\overline{m}^{S}\left(x,\overline{u}\right)-\underline{m}^{S}\left(x,\overline{u}\right)\right)\cdot\dfrac{v-\underline{Q}\left(x\right)}{\overline{Q}\left(x\right)-\underline{Q}\left(x\right)} & \text{if } \underline{Q}\left(x\right)<v\leq\overline{Q}\left(x\right)\\ \\ \overline{m}^{S}\left(x,\overline{u}\right)+\left(1-\overline{m}^{S}\left(x,\overline{u}\right)\right)\dfrac{v-\overline{Q}\left(x\right)}{1-\overline{Q}\left(x\right)} & \text{if } \overline{Q}\left(x\right)<v\end{cases}.$$

Step 10. I write $F_{\tilde{Y}_0^*,\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}\left(y_0,y_1\,|x,\overline{u},v\right)=F_{\tilde{Y}_0^*|X,\tilde{U},\tilde{V}}\left(y_0\,|x,\overline{u},v\right)\cdot F_{\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}\left(y_1\,|x,\overline{u},v\right)$, implying that I can separately define $F_{\tilde{Y}_0^*|X,\tilde{U},\tilde{V}}\left(y_0\,|x,\overline{u},v\right)$ and $F_{\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}\left(y_1\,|x,\overline{u},v\right)$.

Step 11. When $Q\left(1,x\right)>Q\left(0,x\right)$ and $\mathcal{Y}^*$ is a bounded interval (sub-case (a) in assumption 7.3), I define

$$F_{\tilde{Y}_0^*|X,\tilde{U},\tilde{V}}\left(y_0\,|x,\overline{u},v\right)=\begin{cases}\mathbf{1}\left\{y_0\geq\dfrac{m_0^Y\left(\overline{x},\overline{u}\right)}{m_0^S\left(\overline{x},\overline{u}\right)}\right\} & \text{if } v\leq\underline{Q}\left(x\right)\\ -------- & --------\\ \mathbf{1}\left\{y_0\geq\dfrac{\underline{y}^*+\overline{y}^*}{2}\right\} & \text{if } \underline{Q}\left(x\right)<v\end{cases}.$$

When $Q\left(1,x\right)>Q\left(0,x\right)$ and $\overline{y}^*=\max\left\{y\in\mathcal{Y}^*\right\}$ and $\underline{y}^*=\min\left\{y\in\mathcal{Y}^*\right\}$ (case (b) in assumption 7.3), I define

$$F_{\tilde{Y}_0^*|X,\tilde{U},\tilde{V}}\left(y_0\,|x,\overline{u},v\right)=\begin{cases}0 & \text{if } y_0<\underline{y}^*\text{ and } v\leq\underline{Q}\left(x\right)\\ \\ 1-\dfrac{\dfrac{m_0^Y\left(\overline{x},\overline{u}\right)}{m_0^S\left(\overline{x},\overline{u}\right)}-\underline{y}^*}{\overline{y}^*-\underline{y}^*} & \text{if } \underline{y}^*\leq y_0<\overline{y}^*\text{ and } v\leq\underline{Q}\left(x\right)\\ \\ 1 & \text{if } \overline{y}^*\leq y_0\text{ and } v\leq\underline{Q}\left(x\right)\\ -------- & --------\\ \mathbf{1}\left\{y_0\geq\overline{y}^*\right\} & \text{if } \underline{Q}\left(x\right)<v\end{cases}.$$

which are valid cumulative distribution functions because $\dfrac{m_0^Y(\overline{x},\overline{u})}{m_0^S(\overline{x},\overline{u})} \in \left[\underline{y}^*, \overline{y}^*\right]$.

When $Q(1,x) < Q(0,x)$ and $\mathcal{Y}^*$ is a bounded interval (case (a) in assumption 7.3), I define

$$F_{\tilde{Y}_0^*|X,\tilde{U},\tilde{V}}(y_0|x,\overline{u},v) = \begin{cases} \mathbf{1}\{y_0 \geq \alpha(\overline{x},\overline{u})\} & \text{if } v \leq \underline{Q}(x) \\ ----------- & ------------- \\ \mathbf{1}\{y_0 \geq \gamma(\overline{x},\overline{u})\} & \text{if } \underline{Q}(x) < v \leq \overline{Q}(x) \\ ----------- & ------------- \\ \mathbf{1}\left\{y_0 \geq \dfrac{\underline{y}^* + \overline{y}^*}{2}\right\} & \text{if } \overline{Q}(x) < v \end{cases}.$$

When $Q(1,x) < Q(0,x)$ and $\overline{y}^* = \max\{y \in \mathcal{Y}^*\}$ and $\underline{y}^* = \min\{y \in \mathcal{Y}^*\}$ (case (b) in assumption 7.3), I define

$$F_{\tilde{Y}_0^*|X,\tilde{U},\tilde{V}}(y_0|x,\overline{u},v) = \begin{cases} 0 & \text{if } y_0 < \underline{y}^* \text{ and } v \leq \underline{Q}(x) \\ 1 - \dfrac{\alpha(\overline{x},\overline{u}) - \underline{y}^*}{\overline{y}^* - \underline{y}^*} & \text{if } \underline{y}^* \leq y_0 < \overline{y}^* \text{ and } v \leq \underline{Q}(x) \\ 1 & \text{if } \overline{y}^* \leq y_0 \text{ and } v \leq \underline{Q}(x) \\ ----------- & --------------------- \\ 0 & \text{if } y_0 < \underline{y}^* \text{ and } \underline{Q}(x) < v \leq \overline{Q}(x) \\ 1 - \dfrac{\gamma(\overline{x},\overline{u}) - \underline{y}^*}{\overline{y}^* - \underline{y}^*} & \text{if } \underline{y}^* \leq y_0 < \overline{y}^* \text{ and } \underline{Q}(x) < v \leq \overline{Q}(x) \\ 1 & \text{if } \overline{y}^* \leq y_0 \text{ and } \underline{Q}(x) < v \leq \overline{Q}(x) \\ ----------- & --------------------- \\ \mathbf{1}\{y_0 \geq \overline{y}^*\} & \text{if } \overline{Q}(x) < v \end{cases}.$$

which are valid cumulative distribution functions because of equations (D.8) and (D.9).

Step 12. When $Q(1,x) > Q(0,x)$ and $\mathcal{Y}^*$ is a bounded interval (case (a) in assumption 7.3), I define

$$F_{\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}(y_1|x,\overline{u},v) = \begin{cases} \mathbf{1}\{y_1 \geq \alpha(\overline{x},\overline{u})\} & \text{if } v \leq \underline{Q}(x) \\ ------ & ------- \\ \mathbf{1}\{y_1 \geq \gamma(\overline{x},\overline{u})\} & \text{if } \underline{Q}(x) < v \leq \overline{Q}(x) \\ ------ & ------- \\ \mathbf{1}\left\{y_1 \geq \dfrac{\underline{y}^* + \overline{y}^*}{2}\right\} & \text{if } \overline{Q}(x) < v \end{cases}.$$

When $Q(1,x) > Q(0,x)$ and $\overline{y}^* = \max\{y \in \mathcal{Y}^*\}$ and $\underline{y}^* = \min\{y \in \mathcal{Y}^*\}$ (case (b) in assumption 7.3), I define

$$F_{\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}(y_1|x,\overline{u},v) = \begin{cases} 0 & \text{if } y_1 < \underline{y}^* \text{ and } v \leq \underline{Q}(x) \\ 1 - \dfrac{\alpha(\overline{x},\overline{u}) - \underline{y}^*}{\overline{y}^* - \underline{y}^*} & \text{if } \underline{y}^* \leq y_1 < \overline{y}^* \text{ and } v \leq \underline{Q}(x) \\ 1 & \text{if } \overline{y}^* \leq y_1 \text{ and } v \leq \underline{Q}(x) \\ ------ & ----------- \\ 0 & \text{if } y_1 < \underline{y}^* \text{ and } \underline{Q}(x) < v \leq \overline{Q}(x) \\ 1 - \dfrac{\gamma(\overline{x},\overline{u}) - \underline{y}^*}{\overline{y}^* - \underline{y}^*} & \text{if } \underline{y}^* \leq y_1 < \overline{y}^* \text{ and } \underline{Q}(x) < v \leq \overline{Q}(x) \\ 1 & \text{if } \overline{y}^* \leq y_1 \text{ and } \underline{Q}(x) < v \leq \overline{Q}(x) \\ ------ & ----------- \\ \mathbf{1}\{y_1 \geq \overline{y}^*\} & \text{if } \overline{Q}(x) < v \end{cases}.$$

which are valid cumulative distribution functions because of equations (A.10) and (A.11).

When $Q(1,x) < Q(0,x)$ and $\mathcal{Y}^*$ is a bounded interval (sub-case (a) in assumption 7.3), I define

$$F_{\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}\left(y_1\left|x,\overline{u},v\right.\right)=\begin{cases}\mathbf{1}\left\{y_1\geq\dfrac{m_1^Y\left(\overline{x},\overline{u}\right)}{m_1^S\left(\overline{x},\overline{u}\right)}\right\} & \text{if } v\leq\underline{Q}\left(x\right)\\ ------------ & --------\\ \mathbf{1}\left\{y_1\geq\dfrac{\underline{y}^*+\overline{y}^*}{2}\right\} & \text{if } \underline{Q}\left(x\right)<v\end{cases}.$$

When $Q\left(1,x\right)<Q\left(0,x\right)$ and $\overline{y}^*=\max\left\{y\in\mathcal{Y}^*\right\}$ and $\underline{y}^*=\min\left\{y\in\mathcal{Y}^*\right\}$ (case (b) in assumption 7.3), I define

$$F_{\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}\left(y_1\left|x,\overline{u},v\right.\right)=\begin{cases}0 & \text{if } y_1<\underline{y}^* \text{ and } v\leq\underline{Q}\left(x\right)\\ 1-\dfrac{\dfrac{m_1^Y\left(\overline{x},\overline{u}\right)}{m_1^S\left(\overline{x},\overline{u}\right)}-\underline{y}^*}{\overline{y}^*-\underline{y}^*} & \text{if } \underline{y}^*\leq y_1<\overline{y}^* \text{ and } v\leq\underline{Q}\left(x\right)\\ 1 & \text{if } \overline{y}^*\leq y_1 \text{ and } v\leq\underline{Q}\left(x\right)\\ ------------ & ----------------\\ \mathbf{1}\left\{y_1\geq\overline{y}^*\right\} & \text{if } \underline{Q}\left(x\right)<v\end{cases}.$$

which are valid cumulative distribution functions because $\dfrac{m_1^Y\left(\overline{x},\overline{u}\right)}{m_1^S\left(\overline{x},\overline{u}\right)}\in\left[\underline{y}^*,\overline{y}^*\right]$.

Having defined the joint cumulative distribution function $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}$, note that equations (D.8) and (D.9), the facts $\dfrac{m_0^Y\left(\overline{x},\overline{u}\right)}{m_0^S\left(\overline{x},\overline{u}\right)}\in\left[\underline{y}^*,\overline{y}^*\right]$ and $\dfrac{m_1^Y\left(\overline{x},\overline{u}\right)}{m_1^S\left(\overline{x},\overline{u}\right)}\in\left[\underline{y}^*,\overline{y}^*\right]$, and steps 7-12 ensure that equation (D.3) holds.

Now, I show, in three steps, that equation (D.2) holds.

Step 13. Observe that

$$\begin{aligned}&\mathbb{E}\left[\tilde{Y}_1^*\left|X=\overline{x},\tilde{U}=\overline{u},\tilde{S}_0=1,\tilde{S}_1=1\right.\right]\\&\quad=\mathbf{1}\left\{Q\left(1,x\right)>Q\left(0,x\right)\right\}\cdot\alpha\left(\overline{x},\overline{u}\right)+\mathbf{1}\left\{Q\left(1,x\right)<Q\left(0,x\right)\right\}\cdot\frac{m_1^Y\left(\overline{x},\overline{u}\right)}{m_1^S\left(\overline{x},\overline{u}\right)}.\quad\text{(D.10)}\end{aligned}$$

Step 14. Notice that

$$\mathbb{E}\left[\tilde{Y}_0^* \middle| X=\overline{x}, \tilde{U}=\overline{u}, \tilde{S}_0=1, \tilde{S}_1=1\right]$$
$$= \mathbf{1}\left\{Q(1,x)>Q(0,x)\right\} \cdot \frac{m_0^Y(\overline{x},\overline{u})}{m_0^S(\overline{x},\overline{u})} + \mathbf{1}\left\{Q(1,x)<Q(0,x)\right\} \cdot \alpha(\overline{x},\overline{u}). \quad \text{(D.11)}$$

Step 15. Note that Steps 13 and 14 imply that

$$\Delta_{\tilde{Y}^*}^{OO}(\overline{x},\overline{u}) \coloneqq \mathbb{E}\left[\tilde{Y}_1^* - \tilde{Y}_0^* \middle| X=\overline{x}, \tilde{U}=\overline{u}, \tilde{S}_0=1, \tilde{S}_1=1\right] = \delta(\overline{x},\overline{u}),$$

ensuring that equation (D.2) holds.

Finally, to show that equation (D.4) holds, it suffices to follow steps 16 and 17 in Appendix A.4.1.

I can, then, conclude that proposition D.3 is true. ■

Proof of Proposition D.4. This proof is essentially the same proof of proposition D.3 under assumption 7.3.(a). Fix $\overline{u} \in [0,1]$, $\overline{x} \in \mathcal{X}$ and $\delta(\overline{x},\overline{u}) \in \mathbb{R}$ arbitrarily. For brevity, define

$$\alpha(\overline{x},\overline{u}) \coloneqq \mathbf{1}\left\{Q(1,x)>Q(0,x)\right\} \cdot \left(\delta(\overline{x},\overline{u}) + \frac{m_0^Y(\overline{x},\overline{u})}{m_0^S(\overline{x},\overline{u})}\right)$$
$$+ \mathbf{1}\left\{Q(1,x)<Q(0,x)\right\} \cdot \left(-\delta(\overline{x},\overline{u}) + \frac{m_1^Y(\overline{x},\overline{u})}{m_1^S(\overline{x},\overline{u})}\right),$$

and

$$\gamma(\overline{x},\overline{u}) \coloneqq \mathbf{1}\left\{Q(1,x)>Q(0,x)\right\} \cdot \left(\frac{m_1^Y(\overline{x},\overline{u}) - \alpha(\overline{x},\overline{u}) \cdot m_0^S(\overline{x},\overline{u})}{\Delta_S(\overline{x},\overline{u})}\right)$$
$$+ \mathbf{1}\left\{Q(1,x)<Q(0,x)\right\} \cdot \left(\frac{m_0^Y(\overline{x},\overline{u}) - \alpha(\overline{x},\overline{u}) \cdot m_1^S(\overline{x},\overline{u})}{-\Delta_S(\overline{x},\overline{u})}\right).$$

Note that $\alpha(\overline{x},\overline{u}) \in \mathbb{R} = \mathcal{Y}^*$ and $\gamma(\overline{x},\overline{u}) \in \mathbb{R} = \mathcal{Y}^*$.

I define the random variables $\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)$ using the joint cumulative distribution function $F_{\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}, Z, X}$ described by steps 1-12 in the last proof for the case of convex support

$\mathcal{Y}^*$. Note that equation (D.6) is trivially true when $\mathcal{Y}^* = \mathbb{R}$. Moreover, equations (D.5) and (D.7) are valid by the argument described in the last proof

I can, then, conclude that proposition D.4 is true. ■

E Uninformative Bounds with Non-monotone Sample Selection

In the main text and in appendices C and D, I impose some monotonicity condition on the sample selection problem through equation (2). However, in some empirical applications, this assumption may be invalid. For example, in the short run, a job training program may move some individuals from unemployment to employment by increasing their human capital or from employment to unemployment by decreasing their labor market experience. Since this is a frequent feature in empirical economics, it is important to understand what can be discovered about the marginal treatment effect when sample selection is not monotone. To do so, I drop equation (2) and impose equation (1), Assumptions 1-6, a small generalization of Assumption 7

Assumption E.1 *I assume that $\underline{y}^*$ and $\overline{y}^*$ are known, and that*

1. $\underline{y}^* = -\infty$, $\overline{y}^* = \infty$ *and* $\mathcal{Y}^* = \mathbb{R}$, *or*
2. $\underline{y}^* > -\infty$, $\overline{y}^* = \infty$ *and* $\mathcal{Y}^*$ *is an interval, or*
3. $\underline{y}^* = -\infty$, $\overline{y}^* < \infty$ *and* $\mathcal{Y}^*$ *is an interval, or*
4. $\underline{y}^* > -\infty$, $\overline{y}^* < \infty$ *and*
   (a) $\mathcal{Y}^*$ *is an interval or*
   (b) $\underline{y}^* \in \mathcal{Y}^*$ *and* $\overline{y}^* \in \mathcal{Y}^*$.

and mild regularity conditions to ensure that all objects are well-defined

Assumption E.2 *For any $x \in \mathcal{X}$ and $u \in [0, 1]$,*

$$\mathbb{P}\left[S_0 = 1, S_1 = 1\right] > 0, \tag{E.1}$$

$$\mathbb{P}\left[S_0 = 1, S_1 = 0\right] > 0, \tag{E.2}$$

$$\mathbb{P}\left[S_0 = 0, S_1 = 1\right] > 0, \tag{E.3}$$

$$\overline{y}^* \cdot m_d^S(x,u) - m_d^Y(x,u) > 0 \text{ for any } d \in \{0,1\}, \tag{E.4}$$

and

$$m_d^Y(x,u) - \underline{y}^* \cdot m_d^S(x,u) > 0 \text{ for any } d \in \{0,1\}. \tag{E.5}$$

I stress that conditions (E.4) and (E.5) are implied by a non-degenerate conditional distribution for each potential outcome of interest. Most importantly, the above assumptions are sufficient to construct bounds for the ITT^{OO} (Horowitz & Manski (2000)) and for the $LATE^{OO}$ (Chen & Flores 2015, section 2.4) that are shorter than the entire support of the treatment effect.

I, now, show that, differently from the ITT^{OO} and the $LATE^{OO}$, the bounds on the MTE^{OO} on the outcome of interest (equation (3)) without equation (2) are uninformative, i.e., the bounds without monotone sample selection are equal to $\left(\underline{y}^* - \overline{y}^*, \overline{y}^* - \underline{y}^*\right)$. Formally, I have that:

Proposition E.3 *Suppose that the functions m_0^Y, m_1^Y, m_0^S and Δ_S are point-identified at every pair $(x,u) \in \mathcal{X} \times [0,1]$. Impose equation (1) and assumptions 1-6 and E.1-E.2. Then, for any $\overline{u} \in [0,1]$, $\overline{x} \in \mathcal{X}$ and $\delta(\overline{x},\overline{u}) \in \left(\underline{y}^* - \overline{y}^*, \overline{y}^* - \underline{y}^*\right)$, there exist random variables $\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{S}_0, \tilde{S}_1\right)$ such that*

$$\Delta_{\tilde{Y}^*}^{OO}(\overline{x},\overline{u}) \coloneqq \mathbb{E}\left[\tilde{Y}_1^* - \tilde{Y}_0^* \,\middle|\, X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right] = \delta(\overline{x},\overline{u}), \tag{E.6}$$

$$\mathbb{P}\left[\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{S}_0, \tilde{S}_1\right) \in \mathcal{Y}^* \times \mathcal{Y}^* \times \{0,1\} \times \{0,1\} \,\middle|\, X = \overline{x}, \tilde{U} = u\right] = 1 \text{ for any } u \in [0,1], \tag{E.7}$$

and

$$F_{\tilde{Y},\tilde{D},\tilde{S},Z,X}(y,d,s,z,\overline{x}) = F_{Y,D,S,Z,X}(y,d,s,z,\overline{x}) \tag{E.8}$$

for any $(y,d,s,z) \in \mathbb{R}^4$, where $\tilde{D} \coloneqq \mathbf{1}\left\{P(X,Z) \geq \tilde{U}\right\}$, $\tilde{S} = \tilde{D} \cdot \tilde{S}_1 + \left(1 - \tilde{D}\right) \cdot \tilde{S}_0$, $\tilde{Y}_0 = \tilde{S}_0 \cdot \tilde{Y}_0^*$, $\tilde{Y}_1 = \tilde{S}_1 \cdot \tilde{Y}_1^*$ and $\tilde{Y} = \tilde{D} \cdot \tilde{Y}_1 + \left(1 - \tilde{D}\right) \cdot \tilde{Y}_0$.

Proof of Proposition E.3. I only prove proposition E.3 under assumption E.1.4 (sub-cases (a) or (b)) because this is the more demanding case and because the other cases are trivial extensions of this one.

Fix $\overline{u} \in [0,1]$, $\overline{x} \in \mathcal{X}$ and $\delta(\overline{x},\overline{u}) \in \left(\underline{y}^* - \overline{y}^*, \overline{y}^* - \underline{y}^*\right)$ arbitrarily. For brevity, define $(\alpha_0(\overline{x},\overline{u}), \alpha_1(\overline{x},\overline{u})) \in \left(\underline{y}^*, \overline{y}^*\right)^2$ such that $\delta(\overline{x},\overline{u}) = \alpha_1(\overline{x},\overline{u}) - \alpha_0(\overline{x},\overline{u})$, $\pi(\overline{x},\overline{u}) := \frac{1}{2} \cdot \min_{d \in \{0,1\}} \left\{ \min \left\{ m_d^S(\overline{x},\overline{u}), \frac{\overline{y}^* \cdot m_d^S(x,u) - m_d^Y(x,u)}{\overline{y}^* - \alpha_d(\overline{x},\overline{u})}, \frac{m_d^Y(x,u) - \underline{y}^* \cdot m_d^S(x,u)}{\alpha_d(\overline{x},\overline{u}) - \underline{y}^*} \right\} \right\}$, $\gamma_0(\overline{x},\overline{u}) := \frac{m_0^Y(\overline{x},\overline{u}) - \alpha_0(\overline{x},\overline{u}) \cdot \pi(\overline{x},\overline{u})}{m_0^S(\overline{x},\overline{u}) - \pi(\overline{x},\overline{u})}$ and $\gamma_1(\overline{x},\overline{u}) := \frac{m_1^Y(\overline{x},\overline{u}) - \alpha_1(\overline{x},\overline{u}) \cdot \pi(\overline{x},\overline{u})}{m_1^S(\overline{x},\overline{u}) - \pi(\overline{x},\overline{u})}$. Note that, by construction, $\pi(\overline{x},\overline{u}) > 0$ and $(\gamma_0(\overline{x},\overline{u}), \gamma_1(\overline{x},\overline{u})) \in \left(\underline{y}^*, \overline{y}^*\right)^2$.

The strategy of this proof consists of defining random variables $\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{S}_0, \tilde{S}_1\right)$ through their joint cumulative distribution function $F_{\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{S}_0, \tilde{S}_1, Z, X}$ and, then, checking that equations (E.6), (E.7) and (E.8) are satisfied. I fix $(y_0, y_1, u, s_0, s_1, z, x) \in \mathbb{R}^7$ and define $F_{\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{S}_0, \tilde{S}_1, Z, X}$ in twelve steps:

Step 1. For $x \notin \mathcal{X}$, $F_{\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{S}_0, \tilde{S}_1, Z, X}(y_0, y_1, u, s_0, s_1, z, x) = F_{Y_0^*, Y_1^*, U, S_0, S_1, Z, X}(y_0, y_1, u, s_0, s_1, z, x)$.

Step 2. From now on, consider $x \in \mathcal{X}$. Since

$$F_{\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{S}_0, \tilde{S}_1, Z, X}(y_0, y_1, u, s_0, s_1, z, x) = F_{\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{S}_0, \tilde{S}_1, Z|X}(y_0, y_1, u, s_0, s_1, z \,|x) \cdot F_X(x),$$

it suffices to define $F_{\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{S}_0, \tilde{S}_1, Z, X}(y_0, y_1, u, s_0, s_1, z, x)$. Moreover, I impose

$$Z \perp\!\!\!\perp \left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{S}_0, \tilde{S}_1\right) |X$$

by writing

$$F_{\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{S}_0, \tilde{S}_1, Z, X}(y_0, y_1, u, s_0, s_1, z, x) = F_{\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{S}_0, \tilde{S}_1|X}(y_0, y_1, u, s_0, s_1 \,|x) \cdot F_{Z|X}(z\,|x),$$

implying that it is sufficient to define $F_{\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{S}_0, \tilde{S}_1|X}(y_0, y_1, u, s_0, s_1 \,|x)$.

Step 3. For $u \notin [0,1]$, I define $F_{\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{S}_0, \tilde{S}_1|X}(y_0, y_1, u, s_0, s_1 \,|x) = F_{Y_0^*, Y_1^*, U, S_0, S_1|X}(y_0, y_1, u, s_0, s_1 \,|x)$.

Step 4. From now on, consider $u \in [0,1]$. Since

$$F_{\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{S}_0, \tilde{S}_1|X}(y_0, y_1, u, s_0, s_1 \,|x) = F_{\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{S}_0, \tilde{S}_1|X, \tilde{U}}(y_0, y_1, s_0, s_1 \,|x, u) \cdot F_{\tilde{U}|X}(u\,|x),$$

it suffices to define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{S}_0,\tilde{S}_1|X,\tilde{U}}\left(y_0,y_1,s_0,s_1\,|x,u\right)$ and $F_{\tilde{U}|X}\left(u\,|x\right)$.

Step 5. I define $F_{\tilde{U}|X}\left(u\,|x\right)=F_{U|X}\left(u\,|x\right)=u$.

Step 6. For any $u\neq\overline{u}$, I define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{S}_0,\tilde{S}_1|X,\tilde{U}}\left(y_0,y_1,s_0,s_1\,|x,u\right)=F_{Y_0^*,Y_1^*,S_0,S_1|X,U}\left(y_0,y_1,s_0,s_1\,|x,u\right)$.

Step 7. For any $(s_0,s_1)\notin\{0,1\}^2$, I define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{S}_0,\tilde{S}_1|X,\tilde{U}}\left(y_0,y_1,s_0,s_1\,|x,\overline{u}\right)=F_{Y_0^*,Y_1^*,S_0,S_1|X,U}\left(y_0,y_1,s_0,s_1\,|x,\overline{u}\right)$

Step 8. From now on, consider $(s_0,s_1)\in\{0,1\}^2$. Since

$$F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{S}_0,\tilde{S}_1|X,\tilde{U}}\left(y_0,y_1,s_0,s_1\,|x,\overline{u}\right)=F_{\tilde{Y}_0^*,\tilde{Y}_1^*|X,\tilde{U},\tilde{S}_0,\tilde{S}_1}\left(y_0,y_1\,|x,\overline{u},s_0,s_1\right)\cdot F_{\tilde{S}_0,\tilde{S}_1|X,\tilde{U}}\left(s_0,s_1\,|x,\overline{u}\right),$$

it is sufficient to define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*|X,\tilde{U},\tilde{S}_0,\tilde{S}_1}\left(y_0,y_1\,|x,\overline{u},s_0,s_1\right)$ and $F_{\tilde{S}_0,\tilde{S}_1|X,\tilde{U}}\left(s_0,s_1\,|x,\overline{u}\right)$.

Step 9. I define $F_{\tilde{S}_0,\tilde{S}_1|X,\tilde{U}}\left(s_0,s_1\,|x,\overline{u}\right)$ by writing

$$\mathbb{P}\left[\tilde{S}_0=1,\tilde{S}_1=1\middle|X=x,\tilde{U}=\overline{u}\right]=\pi\left(\overline{x},\overline{u}\right)>0,$$

$$\mathbb{P}\left[\tilde{S}_0=1,\tilde{S}_1=0\middle|X=x,\tilde{U}=\overline{u}\right]=m_0^S\left(\overline{x},\overline{u}\right)-\pi\left(\overline{x},\overline{u}\right)>0,$$

$$\mathbb{P}\left[\tilde{S}_0=0,\tilde{S}_1=1\middle|X=x,\tilde{U}=\overline{u}\right]=m_1^S\left(\overline{x},\overline{u}\right)-\pi\left(\overline{x},\overline{u}\right)>0,\ \text{and}$$

$$\mathbb{P}\left[\tilde{S}_0=0,\tilde{S}_1=0\middle|X=x,\tilde{U}=\overline{u}\right]=1-m_1^S\left(\overline{x},\overline{u}\right)-m_0^S\left(\overline{x},\overline{u}\right)+\pi\left(\overline{x},\overline{u}\right).$$

Step 10. I write $F_{\tilde{Y}_0^*,\tilde{Y}_1^*|X,\tilde{U},\tilde{S}_0,\tilde{S}_1}\left(y_0,y_1\,|x,\overline{u},s_0,s_1\right)=F_{\tilde{Y}_0^*|X,\tilde{U},\tilde{S}_0,\tilde{S}_1}\left(y_0\,|x,\overline{u},s_0,s_1\right)\cdot F_{\tilde{Y}_1^*|X,\tilde{U},\tilde{S}_0,\tilde{S}_1}\left(y_1\,|x,\overline{u},s_0,s_1\right)$, implying that I can separately define $F_{\tilde{Y}_0^*|X,\tilde{U},\tilde{S}_0,\tilde{S}_1}\left(y_0\,|x,\overline{u},s_0,s_1\right)$ and $F_{\tilde{Y}_1^*|X,\tilde{U},\tilde{S}_0,\tilde{S}_1}\left(y_1\,|x,\overline{u},s_0,s_1\right)$.

Step 11. When $\mathcal{Y}^*$ is a bounded interval (sub-case (a) in assumption 7.3), I define

$$F_{\tilde{Y}_0^*|X,\tilde{U},\tilde{S}_0,\tilde{S}_1}\left(y_0\,|x,\overline{u},s_0,s_1\right)=\begin{cases}\mathbf{1}\left\{y_0\geq\alpha_0\left(\overline{x},\overline{u}\right)\right\} & \text{if }\ (s_0,s_1)=(1,1)\\ -----------&-------------\\ \mathbf{1}\left\{y_0\geq\gamma_0\left(\overline{x},\overline{u}\right)\right\} & \text{if }\ (s_0,s_1)=(1,0)\\ -----------&-------------\\ \mathbf{1}\left\{y_0\geq\dfrac{\underline{y}^*+\overline{y}^*}{2}\right\} & \text{if }\ (s_0,s_1)\in\left\{(0,0),(0,1)\right\}\end{cases}.$$

When $\overline{y}^* = \max\{y \in \mathcal{Y}^*\}$ and $\underline{y}^* = \min\{y \in \mathcal{Y}^*\}$ (case (b) in assumption 7.3), I define

$$F_{\tilde{Y}_0^*|X,\tilde{U},\tilde{V}}\left(y_0\,|x,\overline{u},v\right) = \begin{cases} 0 & \text{if } y_0 < \underline{y}^* \text{ and } (s_0,s_1) = (1,1) \\ 1 - \dfrac{\alpha_0\left(\overline{x},\overline{u}\right) - \underline{y}^*}{\overline{y}^* - \underline{y}^*} & \text{if } \underline{y}^* \leq y_0 < \overline{y}^* \text{ and } (s_0,s_1) = (1,1) \\ 1 & \text{if } \overline{y}^* \leq y_0 \text{ and } (s_0,s_1) = (1,1) \\ ------------ & ------------------- \\ 0 & \text{if } y_0 < \underline{y}^* \text{ and } (s_0,s_1) = (1,0) \\ 1 - \dfrac{\gamma_0\left(\overline{x},\overline{u}\right) - \underline{y}^*}{\overline{y}^* - \underline{y}^*} & \text{if } \underline{y}^* \leq y_0 < \overline{y}^* \text{ and } (s_0,s_1) = (1,0) \\ 1 & \text{if } \overline{y}^* \leq y_0 \text{ and } (s_0,s_1) = (1,0) \\ ------------ & ------------------- \\ \mathbf{1}\left\{y_0 \geq \overline{y}^*\right\} & (s_0,s_1) \in \left\{(0,0),(0,1)\right\} \end{cases}.$$

which are valid cumulative distribution functions because $\alpha_0\left(\overline{x},\overline{u}\right) \in \left(\underline{y}^*, \overline{y}^*\right)$ and $\gamma_0\left(\overline{x},\overline{u}\right) \in \left(\underline{y}^*, \overline{y}^*\right)$.

Step 12. When $\mathcal{Y}^*$ is a bounded interval (sub-case (a) in assumption 7.3), I define

$$F_{\tilde{Y}_1^*|X,\tilde{U},\tilde{S}_0,\tilde{S}_1}\left(y_1\,|x,\overline{u},s_0,s_1\right) = \begin{cases} \mathbf{1}\left\{y_1 \geq \alpha_1\left(\overline{x},\overline{u}\right)\right\} & \text{if } (s_0,s_1) = (1,1) \\ ------------ & -------------- \\ \mathbf{1}\left\{y_1 \geq \gamma_1\left(\overline{x},\overline{u}\right)\right\} & \text{if } (s_0,s_1) = (0,1) \\ ------------ & -------------- \\ \mathbf{1}\left\{y_1 \geq \dfrac{\underline{y}^* + \overline{y}^*}{2}\right\} & \text{if } (s_0,s_1) \in \left\{(0,0),(1,0)\right\} \end{cases}.$$

When $\overline{y}^* = \max\{y \in \mathcal{Y}^*\}$ and $\underline{y}^* = \min\{y \in \mathcal{Y}^*\}$ (case (b) in assumption 7.3), I define

$$F_{\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}(y_1 | x, \overline{u}, v) = \begin{cases} 0 & \text{if } y_1 < \underline{y}^* \text{ and } (s_0, s_1) = (1,1) \\ 1 - \dfrac{\alpha_1(\overline{x}, \overline{u}) - \underline{y}^*}{\overline{y}^* - \underline{y}^*} & \text{if } \underline{y}^* \leq y_1 < \overline{y}^* \text{ and } (s_0, s_1) = (1,1) \\ 1 & \text{if } \overline{y}^* \leq y_1 \text{ and } (s_0, s_1) = (1,1) \\ ---------- & --------------- \\ 0 & \text{if } y_1 < \underline{y}^* \text{ and } (s_0, s_1) = (0,1) \\ 1 - \dfrac{\gamma_1(\overline{x}, \overline{u}) - \underline{y}^*}{\overline{y}^* - \underline{y}^*} & \text{if } \underline{y}^* \leq y_1 < \overline{y}^* \text{ and } (s_0, s_1) = (0,1) \\ 1 & \text{if } \overline{y}^* \leq y_1 \text{ and } (s_0, s_1) = (0,1) \\ ---------- & --------------- \\ \mathbf{1}\{y_1 \geq \overline{y}^*\} & (s_0, s_1) \in \{(0,0), (1,0)\} \end{cases}.$$

which are valid cumulative distribution functions because $\alpha_1(\overline{x}, \overline{u}) \in (\underline{y}^*, \overline{y}^*)$ and $\gamma_1(\overline{x}, \overline{u}) \in (\underline{y}^*, \overline{y}^*)$.

Having defined the joint cumulative distribution function $F_{\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{S}_0, \tilde{S}_1, Z, X}$, note that steps 7-12 ensure that equation (E.7) holds.

Now, observe equation (E.6) holds because steps 11 and 12 ensure that $\alpha_1(\overline{x}, \overline{u}) = \mathbb{E}\left[\tilde{Y}_1^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right]$ and $\alpha_0(\overline{x}, \overline{u}) = \mathbb{E}\left[\tilde{Y}_0^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right]$.

Finally, equation (E.8) holds according to the same argument described at the end of appendix A.4.1.

I can, then, conclude that proposition E.3 is true. ■

F MTE bounds under a Mean Dominance Assumption

Here, I modify the Mean Dominance Assumption (9) by changing the direction of the inequality, i.e., I assume that:

Assumption F.1 *The potential outcome when treated for the always-observed sub-population is less than or equal to the same parameter for the observed-only-when-treated sub-population:*

$$\mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 1, S_1 = 1\right.\right] \leq \mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 0, S_1 = 1\right.\right]$$

for any $x \in \mathcal{X}$ and $u \in [0, 1]$.

Note that assumption F.1 implies that $\Delta_Y^{NO}(x, u) \geq \dfrac{m_1^Y(x, u)}{m_1^S(x, u)} \geq \mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 1, S_1 = 1\right.\right]$. As a consequence, by following the same steps of the proof of Corollary 14, I can derive:

Corollary F.2 *Fix $u \in [0, 1]$ and $x \in \mathcal{X}$ arbitrarily. Suppose that the $m_0^Y(x, u)$, $m_1^Y(x, u)$, $m_0^S(x, u)$ and $\Delta_S(x, u)$ are point-identified.*

Under assumptions 1-6, 7.1, 8 and F.1, $\Delta_{Y^*}^{OO}(x, u)$ must satisfy

$$\Delta_{Y^*}^{OO}(x, u) \geq \underline{y}^* - \frac{m_0^Y(x, u)}{m_0^S(x, u)} =: \underline{\Delta_{Y^*}^{OO}}(x, u) \tag{F.1}$$

and that

$$\Delta_{Y^*}^{OO}(x, u) \leq \frac{m_1^Y(x, u)}{m_1^S(x, u)} - \frac{m_0^Y(x, u)}{m_0^S(x, u)} =: \overline{\Delta_{Y^*}^{OO}}(x, u). \tag{F.2}$$

Under assumptions 1-6, 7.2, 8 and F.1, $\Delta_{Y^*}^{OO}(x, u)$ must satisfy

$$\Delta_{Y^*}^{OO}(x, u) \geq \frac{m_1^Y(x, u) - \overline{y}^* \cdot \Delta_S(x, u)}{m_0^S(x, u)} - \frac{m_0^Y(x, u)}{m_0^S(x, u)} =: \underline{\Delta_{Y^*}^{OO}}(x, u) \tag{F.3}$$

and that

$$\Delta_{Y^*}^{OO}(x, u) \leq \frac{m_1^Y(x, u)}{m_1^S(x, u)} - \frac{m_0^Y(x, u)}{m_0^S(x, u)} =: \overline{\Delta_{Y^*}^{OO}}(x, u). \tag{F.4}$$

Under assumptions 1-6, 7.3 (sub-case (a) or (b)), 8 and F.1, $\Delta_{Y^*}^{OO}(x,u)$ *must satisfy*

$$\Delta_{Y^*}^{OO}(x,u) \geq \max\left\{\frac{m_1^Y(x,u) - \overline{y}^* \cdot \Delta_S(x,u)}{m_0^S(x,u)}, \underline{y}^*\right\} - \frac{m_0^Y(x,u)}{m_0^S(x,u)} =: \underline{\Delta_{Y^*}^{OO}}(x,u) \tag{F.5}$$

and that

$$\Delta_{Y^*}^{OO}(x,u) \leq \frac{m_1^Y(x,u)}{m_1^S(x,u)} - \frac{m_0^Y(x,u)}{m_0^S(x,u)} =: \overline{\Delta_{Y^*}^{OO}}(x,u). \tag{F.6}$$

When $\mathcal{Y}^* = \mathbb{R}$ *and assumptions 1-6, 8 and F.1 hold,* $\Delta_{Y^*}^{OO}(x,u)$ *must satisfy*

$$\Delta_{Y^*}^{OO}(x,u) \geq -\infty =: \underline{\Delta_{Y^*}^{OO}}(x,u) \tag{F.7}$$

and that

$$\Delta_{Y^*}^{OO}(x,u) \leq \frac{m_1^Y(x,u)}{m_1^S(x,u)} - \frac{m_0^Y(x,u)}{m_0^S(x,u)} =: \overline{\Delta_{Y^*}^{OO}}(x,u). \tag{F.8}$$

I highlight that the bounds in corollary F.2 can be identified using the strategies that were described in sections 4 and 5. Most importantly, I can derive a result similar to proposition 15:

Proposition F.3 *Suppose that the functions* m_0^Y, m_1^Y, m_0^S, m_1^S *and* Δ_S *are point-identified at every pair* $(x,u) \in \mathcal{X} \times [0,1]$. *Under assumptions 1-6, 8 and F.1, the bounds* $\underline{\Delta_{Y^*}^{OO}}$ *and* $\overline{\Delta_{Y^*}^{OO}}$*, given by corollary F.2, are point-wise sharp, i.e., for any* $\overline{u} \in [0,1]$, $\overline{x} \in \mathcal{X}$ *and* $\delta(\overline{x},\overline{u}) \in \left(\underline{\Delta_{Y^*}^{OO}}(\overline{x},\overline{u}), \overline{\Delta_{Y^*}^{OO}}(\overline{x},\overline{u})\right)$*, there exist random variables* $\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)$ *such that*

$$\Delta_{\tilde{Y}^*}^{OO}(\overline{x},\overline{u}) := \mathbb{E}\left[\tilde{Y}_1^* - \tilde{Y}_0^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right] = \delta(\overline{x},\overline{u}), \tag{F.9}$$

$$\mathbb{P}\left[\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{V}\right) \in \mathcal{Y}^* \times \mathcal{Y}^* \times [0,1] \middle| X = \overline{x}, \tilde{U} = u\right] = 1 \text{ for any } u \in [0,1], \tag{F.10}$$

$$\mathbb{E}\left[\tilde{Y}_1^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right] \leq \mathbb{E}\left[\tilde{Y}_1^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 0, \tilde{S}_1 = 1\right], \tag{F.11}$$

and

$$F_{\tilde{Y},\tilde{D},\tilde{S},Z,X}(y,d,s,z,\overline{x}) = F_{Y,D,S,Z,X}(y,d,s,z,\overline{x}) \tag{F.12}$$

for any $(y,d,s,z) \in \mathbb{R}^4$*, where* $\tilde{D} := \mathbf{1}\left\{P(X,Z) \geq \tilde{U}\right\}$, $\tilde{S}_0 = \mathbf{1}\left\{Q(0,X) \geq \tilde{V}\right\}$, $\tilde{S}_1 =$

$\mathbf{1}\left\{Q\left(1,X\right)\geq\tilde{V}\right\}$, $\tilde{Y}_0=\tilde{S}_0\cdot\tilde{Y}_0^*$, $\tilde{Y}_1=\tilde{S}_1\cdot\tilde{Y}_1^*$ *and* $\tilde{Y}=\tilde{D}\cdot\tilde{Y}_1+\left(1-\tilde{D}\right)\cdot\tilde{Y}_0$.

I stress that the proof of proposition F.3 is symmetric to the proof of proposition 15 (Appendix A.6).

G Sharpness and Impossibility Results with Smoothness Restrictions

In the main text, I imposed no smoothness condition on the joint distribution of $(Y_0^*, Y_1^*, U, V, Z, X)$. Here, I impose the following smoothness condition:

Assumption G.1 *The conditional cumulative distribution functions $F_{V|X,U}$ are $F_{Y_0^*,Y_1^*|X,U,V}$ are continuous functions of the value of U.*

As a consequence of this new assumption, Propositions 12 and 13 have to be modified to accommodate infinitesimal violations of the data restriction and to ensure that the extra model restrictions imposed by assumption G.1 are also satisfied.

Proposition G.2 *Suppose that the functions m_0^Y, m_1^Y, m_0^S and Δ_S are point-identified at every pair $(x, u) \in \mathcal{X} \times [0, 1]$. Under Assumptions 1-6, 7 (sub-cases 1, 2, 3(a) or 3(b)), 8 and G.1, the bounds $\underline{\Delta_{Y^*}^{OO}}$ and $\overline{\Delta_{Y^*}^{OO}}$, given by Corollary 11 are infinitesimally point-wise sharp, i.e., for any $\epsilon \in \mathbb{R}_{++}$, $\overline{u} \in [0, 1]$, $\overline{x} \in \mathcal{X}$ and $\delta(\overline{x}, \overline{u}) \in \left(\underline{\Delta_{Y^*}^{OO}}(\overline{x}, \overline{u}), \overline{\Delta_{Y^*}^{OO}}(\overline{x}, \overline{u})\right)$, there exist random variables $\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)$ such that*

$$\Delta_{\tilde{Y}^*}^{OO}(\overline{x}, \overline{u}) \coloneqq \mathbb{E}\left[\tilde{Y}_1^* - \tilde{Y}_0^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right] = \delta(\overline{x}, \overline{u}), \tag{G.1}$$

$$\mathbb{P}\left[\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{V}\right) \in \mathcal{Y}^* \times \mathcal{Y}^* \times [0, 1] \middle| X = \overline{x}, \tilde{U} = u\right] = 1 \text{ for any } u \in [0, 1], \tag{G.2}$$

$$F_{\tilde{V}|X,\tilde{U}} \text{ is a continuous function of the value of } \tilde{U}, \tag{G.3}$$

$$F_{\tilde{Y}_0^*,\tilde{Y}_1^*|X,\tilde{U},\tilde{V}} \text{ is a continuous function of the value of } \tilde{U}, \tag{G.4}$$

and

$$\left|F_{\tilde{Y},\tilde{D},\tilde{S},Z,X}(y, d, s, z, \overline{x}) - F_{Y,D,S,Z,X}(y, d, s, z, \overline{x})\right| \leq \epsilon \tag{G.5}$$

for any $(y, d, s, z) \in \mathbb{R}^4$, where $\tilde{D} \coloneqq \mathbf{1}\left\{P(X, Z) \geq \tilde{U}\right\}$, $\tilde{S}_0 = \mathbf{1}\left\{Q(0, X) \geq \tilde{V}\right\}$, $\tilde{S}_1 = \mathbf{1}\left\{Q(1, X) \geq \tilde{V}\right\}$, $\tilde{Y}_0 = \tilde{S}_0 \cdot \tilde{Y}_0^*$, $\tilde{Y}_1 = \tilde{S}_1 \cdot \tilde{Y}_1^*$ and $\tilde{Y} = \tilde{D} \cdot \tilde{Y}_1 + \left(1 - \tilde{D}\right) \cdot \tilde{Y}_0$.

Proposition G.3 *Suppose that the functions m_0^Y, m_1^Y, m_0^S and Δ_S are point-identified at every pair $(x, u) \in \mathcal{X} \times [0, 1]$. Impose Assumptions 1-6, 8 and G.1. If $\mathcal{Y}^* = \mathbb{R}$, then, for any*

$\epsilon \in \mathbb{R}_{++}$, $\overline{u} \in [0,1]$, $\overline{x} \in \mathcal{X}$ and $\delta(\overline{x},\overline{u}) \in \mathbb{R}$, there exist random variables $\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)$ such that

$$\Delta_{\tilde{Y}^*}^{OO}(\overline{x},\overline{u}) \coloneqq \mathbb{E}\left[\tilde{Y}_1^* - \tilde{Y}_0^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right] = \delta(\overline{x},\overline{u}), \tag{G.6}$$

$$\mathbb{P}\left[\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{V}\right) \in \mathcal{Y}^* \times \mathcal{Y}^* \times [0,1] \middle| X = \overline{x}, \tilde{U} = u\right] = 1 \text{ for any } u \in [0,1], \tag{G.7}$$

$$F_{\tilde{V}|X,\tilde{U}} \text{ is a continuous function of the value of } \tilde{U}, \tag{G.8}$$

$$F_{\tilde{Y}_0^*,\tilde{Y}_1^*|X,\tilde{U},\tilde{V}} \text{ is a continuous function of the value of } \tilde{U}, \tag{G.9}$$

and

$$\left|F_{\tilde{Y},\tilde{D},\tilde{S},Z,X}(y,d,s,z,\overline{x}) - F_{Y,D,S,Z,X}(y,d,s,z,\overline{x})\right| \leq \epsilon \tag{G.10}$$

for any $(y,d,s,z) \in \mathbb{R}^4$, where $\tilde{D} \coloneqq \mathbf{1}\left\{P(X,Z) \geq \tilde{U}\right\}$, $\tilde{S}_0 = \mathbf{1}\left\{Q(0,X) \geq \tilde{V}\right\}$, $\tilde{S}_1 = \mathbf{1}\left\{Q(1,X) \geq \tilde{V}\right\}$, $\tilde{Y}_0 = \tilde{S}_0 \cdot \tilde{Y}_0^*$, $\tilde{Y}_1 = \tilde{S}_1 \cdot \tilde{Y}_1^*$ and $\tilde{Y} = \tilde{D} \cdot \tilde{Y}_1 + \left(1 - \tilde{D}\right) \cdot \tilde{Y}_0$.

The proofs of propositions G.2 and G.3 are below. They are small modification of the previous proofs.

Proof of Proposition G.2. I only prove proposition G.2 under assumption 7.3 (subcases (a) and (b)).The proofs of proposition G.2 under assumptions 7.1 and 7.2 are trivial modifications of the proof presented below.

Fix any $\overline{u} \in [0,1]$, any $\overline{x} \in \mathcal{X}$, any $\delta(\overline{x},\overline{u}) \in \left(\underline{\Delta_{Y^*}^{OO}}(\overline{x},\overline{u}), \overline{\Delta_{Y^*}^{OO}}(\overline{x},\overline{u})\right)$ and any $\epsilon \in \mathbb{R}_{++}$ such that $\min\left\{\overline{u} - \dfrac{\epsilon}{2 \cdot F_X(\overline{x})}, 1 - \left(\overline{u} - \dfrac{\epsilon}{2 \cdot F_X(\overline{x})}\right)\right\} > 0$. For brevity, define $\alpha(\overline{x},\overline{u}) \coloneqq \delta(\overline{x},\overline{u}) + \dfrac{m_0^Y(\overline{x},\overline{u})}{m_0^S(\overline{x},\overline{u})}$, $\gamma(\overline{x},\overline{u}) \coloneqq \dfrac{m_1^Y(\overline{x},\overline{u}) - \alpha(\overline{x},\overline{u}) \cdot m_0^S(\overline{x},\overline{u})}{\Delta_S(\overline{x},\overline{u})}$ and $\overline{\epsilon} \coloneqq \dfrac{\epsilon}{2 \cdot F_X(\overline{x})}$.

Note that

$$
\begin{aligned}
\delta\left(\overline{x},\overline{u}\right) &\in \left(\underline{\Delta_{Y^*}^{OO}}\left(\overline{x},\overline{u}\right), \overline{\Delta_{Y^*}^{OO}}\left(\overline{x},\overline{u}\right)\right) \\
\Leftrightarrow \quad \alpha\left(\overline{x},\overline{u}\right) &\in \left(\max\left\{\frac{m_1^Y\left(x,u\right) - \overline{y}^* \cdot \Delta_S\left(x,u\right)}{m_0^S\left(x,u\right)}, \underline{y}^*\right\},\right. \\
&\qquad \left.\min\left\{\frac{m_1^Y\left(x,u\right) - \underline{y}^* \cdot \Delta_S\left(x,u\right)}{m_0^S\left(x,u\right)}, \overline{y}^*\right\}\right) \\
&\subseteq \left(\underline{y}^*, \overline{y}^*\right),
\end{aligned} \tag{G.11}
$$

and that

$$
\begin{aligned}
\alpha\left(\overline{x},\overline{u}\right) &\in \left(\frac{m_1^Y\left(x,u\right) - \overline{y}^* \cdot \Delta_S\left(x,u\right)}{m_0^S\left(x,u\right)}, \frac{m_1^Y\left(x,u\right) - \underline{y}^* \cdot \Delta_S\left(x,u\right)}{m_0^S\left(x,u\right)}\right) \\
\Leftrightarrow \quad \gamma\left(\overline{x},\overline{u}\right) &\in \left(\underline{y}^*, \overline{y}^*\right).
\end{aligned} \tag{G.12}
$$

The strategy of this proof consists of defining random variables $\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)$ through their joint cumulative distribution function $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}$ and, then, checking that conditions (G.1)-(G.5) are satisfied. I fix $(y_0, y_1, u, v, z, x) \in \mathbb{R}^6$ and define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}$ in fourteen steps:

Step 1. For $x \notin \mathcal{X}$, $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}\left(y_0, y_1, u, v, z, x\right) = F_{Y_0^*,Y_1^*,U,V,Z,X}\left(y_0, y_1, u, v, z, x\right)$.

Step 2. From now on, consider $x \in \mathcal{X}$. Since

$$
F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}\left(y_0, y_1, u, v, z, x\right) = F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0, y_1, u, v, z\left|x\right.\right) \cdot F_X\left(x\right),
$$

it suffices to define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0, y_1, u, v, z\left|x\right.\right)$. Moreover, I impose

$$
Z \perp\!\!\!\perp \left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right) |X
$$

by writing

$$F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0,y_1,u,v,z\,|x\right)=F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}|X}\left(y_0,y_1,u,v\,|x\right)\cdot F_{Z|X}\left(z\,|x\right),$$

implying that it is sufficient to define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}|X}\left(y_0,y_1,u,v\,|x\right)$.

Step 3. For $u\notin[0,1]$, I define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}|X}\left(y_0,y_1,u,v\,|x\right)=F_{Y_0^*,Y_1^*,U,V|X}\left(y_0,y_1,u,v\,|x\right)$.

Step 4. From now on, consider $u\in[0,1]$. Since

$$F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}|X}\left(y_0,y_1,u,v\,|x\right)=F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}\left|X,\tilde{U}\right.}\left(y_0,y_1,v\,|x,u\right)\cdot F_{\tilde{U}|X}\left(u\,|x\right),$$

it suffices to define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}\left|X,\tilde{U}\right.}\left(y_0,y_1,v\,|x,u\right)$ and $F_{\tilde{U}|X}\left(u\,|x\right)$.

Step 5. I define $F_{\tilde{U}|X}\left(u\,|x\right)=F_{U|X}\left(u\,|x\right)=u$.

Step 6. For any $u\notin(\overline{u}-\overline{\epsilon},\overline{u}+\overline{\epsilon})$, I define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}\left|X,\tilde{U}\right.}\left(y_0,y_1,v\,|x,u\right)=F_{Y_0^*,Y_1^*,V|X,U}\left(y_0,y_1,v\,|x,u\right)$.

Step 7. For any $v\notin[0,1]$, I define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}\left|X,\tilde{U}\right.}\left(y_0,y_1,v\,|x,\overline{u}\right)=F_{Y_0^*,Y_1^*,V|X,U}\left(y_0,y_1,v\,|x,\overline{u}\right)$.

Step 8. From now on, consider $v\in[0,1]$. Since

$$F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}\left|X,\tilde{U}\right.}\left(y_0,y_1,v\,|x,\overline{u}\right)=F_{\tilde{Y}_0^*,\tilde{Y}_1^*\left|X,\tilde{U},\tilde{V}\right.}\left(y_0,y_1\,|x,\overline{u},v\right)\cdot F_{\tilde{V}\left|X,\tilde{U}\right.}\left(v\,|x,\overline{u}\right),$$

it is sufficient to define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*\left|X,\tilde{U},\tilde{V}\right.}\left(y_0,y_1\,|x,\overline{u},v\right)$ and $F_{\tilde{V}\left|X,\tilde{U}\right.}\left(v\,|x,\overline{u}\right)$.

Step 9. I define

$$F_{\tilde{V}\left|X,\tilde{U}\right.}\left(v\,|x,\overline{u}\right)=\begin{cases}m_0^S\left(x,\overline{u}\right)\cdot\dfrac{v}{Q\left(0,x\right)} & \text{if } v\leq Q\left(0,x\right)\\[2ex] m_0^S\left(x,\overline{u}\right)+\Delta_S\left(x,\overline{u}\right)\cdot\dfrac{v-Q\left(0,x\right)}{Q\left(1,x\right)-Q\left(0,x\right)} & \text{if } Q\left(0,x\right)<v\leq Q\left(1,x\right)\\[2ex] m_1^S\left(x,\overline{u}\right)+\left(1-m_1^S\left(x,\overline{u}\right)\right)\dfrac{v-Q\left(1,x\right)}{1-Q\left(1,x\right)} & \text{if } Q\left(1,x\right)<v\end{cases}.$$

Step 10. For any $u \in (\overline{u} - \overline{\epsilon}, \overline{u})$, I define

$$F_{\tilde{V}|X,\tilde{U}}\left(v\,|x,u\right) = F_{\tilde{V}|X,\tilde{U}}\left(v\,|x,\overline{u}-\overline{\epsilon}\right)\cdot\left(\frac{\overline{u}-u}{\overline{\epsilon}}\right) + F_{\tilde{V}|X,\tilde{U}}\left(v\,|x,\overline{u}\right)\cdot\left(\frac{u-\overline{u}+\overline{\epsilon}}{\overline{\epsilon}}\right),$$

which are valid cumulative distribution functions because a convex combination of cumulative distribution functions is a cumulative distribution function.

For any $u \in (\overline{u}, \overline{u} + \overline{\epsilon})$, I define

$$F_{\tilde{V}|X,\tilde{U}}\left(v\,|x,u\right) = F_{\tilde{V}|X,\tilde{U}}\left(v\,|x,\overline{u}\right)\cdot\left(\frac{\overline{u}+\overline{\epsilon}-u}{\overline{\epsilon}}\right) + F_{\tilde{V}|X,\tilde{U}}\left(v\,|x,\overline{u}+\overline{\epsilon}\right)\cdot\left(\frac{u-\overline{u}}{\overline{\epsilon}}\right),$$

which are valid cumulative distribution functions because a convex combination of cumulative distribution functions is a cumulative distribution function.

Note that $F_{\tilde{V}|X,\tilde{U}}$ is a continuous function of the value of $\tilde{U}$, i.e., it satisfies restriction (G.3).

Step 11. I write $F_{\tilde{Y}_0^*,\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}\left(y_0, y_1\,|x,\overline{u},v\right) = F_{\tilde{Y}_0^*|X,\tilde{U},\tilde{V}}\left(y_0\,|x,\overline{u},v\right)\cdot F_{\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}\left(y_1\,|x,\overline{u},v\right)$, implying that I can separately define $F_{\tilde{Y}_0^*|X,\tilde{U},\tilde{V}}\left(y_0\,|x,\overline{u},v\right)$ and $F_{\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}\left(y_1\,|x,\overline{u},v\right)$.

Step 12. When $\mathcal{Y}^*$ is a bounded interval (sub-case (a) in assumption 7.3), I define

$$F_{\tilde{Y}_0^*|X,\tilde{U},\tilde{V}}\left(y_0\,|x,\overline{u},v\right) = \begin{cases} \mathbf{1}\left\{y_0 \geq \dfrac{m_0^Y\left(\overline{x},\overline{u}\right)}{m_0^S\left(\overline{x},\overline{u}\right)}\right\} & \text{if } v \leq Q\left(0,x\right) \\ ------------ & -------- \\ \mathbf{1}\left\{y_0 \geq \dfrac{\underline{y}^* + \overline{y}^*}{2}\right\} & \text{if } Q\left(0,x\right) < v \end{cases}.$$

When $\overline{y}^* = \max\{y \in \mathcal{Y}^*\}$ and $\underline{y}^* = \min\{y \in \mathcal{Y}^*\}$ (case (b) in assumption 7.3), I define

$$F_{\tilde{Y}_0^*|X,\tilde{U},\tilde{V}}(y_0|x,\overline{u},v) = \begin{cases} 0 & \text{if } y_0 < \underline{y}^* \text{ and } v \leq Q(0,x) \\ 1 - \dfrac{\dfrac{m_0^Y(\overline{x},\overline{u})}{m_0^S(\overline{x},\overline{u})} - \underline{y}^*}{\overline{y}^* - \underline{y}^*} & \text{if } \underline{y}^* \leq y_0 < \overline{y}^* \text{ and } v \leq Q(0,x) \\ 1 & \text{if } \overline{y}^* \leq y_0 \text{ and } v \leq Q(0,x) \\ ----------- & --------------- \\ \mathbf{1}\{y_0 \geq \overline{y}^*\} & \text{if } Q(0,x) < v \end{cases}.$$

which are valid cumulative distribution functions because $\dfrac{m_0^Y(\overline{x},\overline{u})}{m_0^S(\overline{x},\overline{u})} \in \left[\underline{y}^*, \overline{y}^*\right]$.

Step 13. When $\mathcal{Y}^*$ is a bounded interval (case (a) in assumption 7.3), I define

$$F_{\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}(y_1|x,\overline{u},v) = \begin{cases} \mathbf{1}\{y_1 \geq \alpha(\overline{x},\overline{u})\} & \text{if } v \leq Q(0,x) \\ --------- & ------------ \\ \mathbf{1}\{y_1 \geq \gamma(\overline{x},\overline{u})\} & \text{if } Q(0,x) < v \leq Q(1,x) \\ --------- & ------------ \\ \mathbf{1}\left\{y_1 \geq \dfrac{\underline{y}^* + \overline{y}^*}{2}\right\} & \text{if } Q(1,x) < v \end{cases}.$$

When $\overline{y}^* = \max\{y \in \mathcal{Y}^*\}$ and $\underline{y}^* = \min\{y \in \mathcal{Y}^*\}$ (case (b) in assumption 7.3), I define

$$F_{\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}(y_1|x,\overline{u},v) = \begin{cases} 0 & \text{if } y_1 < \underline{y}^* \text{ and } v \leq Q(0,x) \\ 1 - \dfrac{\alpha(\overline{x},\overline{u}) - \underline{y}^*}{\overline{y}^* - \underline{y}^*} & \text{if } \underline{y}^* \leq y_1 < \overline{y}^* \text{ and } v \leq Q(0,x) \\ 1 & \text{if } \overline{y}^* \leq y_1 \text{ and } v \leq Q(0,x) \\ ------ & ------------ \\ 0 & \text{if } y_1 < \underline{y}^* \text{ and } Q(0,x) < v \leq Q(1,x) \\ 1 - \dfrac{\gamma(\overline{x},\overline{u}) - \underline{y}^*}{\overline{y}^* - \underline{y}^*} & \text{if } \underline{y}^* \leq y_1 < \overline{y}^* \text{ and } Q(0,x) < v \leq Q(1,x) \\ 1 & \text{if } \overline{y}^* \leq y_1 \text{ and } Q(0,x) < v \leq Q(1,x) \\ ------ & ------------ \\ \mathbf{1}\{y_1 \geq \overline{y}^*\} & \text{if } Q(1,x) < v \end{cases}.$$

which are valid cumulative distribution functions because of equations (G.11) and (G.12).

Step 14. For any $u \in (\overline{u} - \overline{\epsilon}, \overline{u})$, I define

$$\begin{aligned} F_{\tilde{Y}_0^*,\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}(y_0,y_1|x,u,v) = {} & F_{\tilde{Y}_0^*,\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}(y_0,y_1|x,\overline{u}-\overline{\epsilon},v) \cdot \left(\frac{\overline{u}-u}{\overline{\epsilon}}\right) \\ & + F_{\tilde{Y}_0^*,\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}(y_0,y_1|x,\overline{u},v) \cdot \left(\frac{u-\overline{u}+\overline{\epsilon}}{\overline{\epsilon}}\right), \end{aligned}$$

which are valid cumulative distribution functions because a convex combination of cumulative distribution functions is a cumulative distribution function.

For any $u \in (\overline{u}, \overline{u} + \overline{\epsilon})$, I define

$$\begin{aligned} F_{\tilde{Y}_0^*,\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}(y_0,y_1|x,u,v) = {} & F_{\tilde{Y}_0^*,\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}(y_0,y_1|x,\overline{u},v) \cdot \left(\frac{\overline{u}+\overline{\epsilon}-u}{\overline{\epsilon}}\right) \\ & + F_{\tilde{Y}_0^*,\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}(y_0,y_1|x,\overline{u}+\overline{\epsilon},v)\left(\frac{u-\overline{u}}{\overline{\epsilon}}\right), \end{aligned}$$

which are valid cumulative distribution functions because a convex combination of cumulative distribution functions is a cumulative distribution function.

Note that $F_{\tilde{Y}_0^*,\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}$ is a continuous function of the value of $\tilde{U}$, i.e., it satisfies restriction (G.4).

Having defined the joint cumulative distribution function $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}$, note that equations (G.11) and (G.12), $\dfrac{m_0^Y(\overline{x},\overline{u})}{m_0^S(\overline{x},\overline{u})} \in \left[\underline{y}^*, \overline{y}^*\right]$ and steps 7-14 ensure that equation (G.2) holds.

Now, I show, in three steps, that equation (G.1) holds.

Step 15. Observe that

$$
\begin{aligned}
&\mathbb{E}\left[\tilde{Y}_1^* \,\middle|\, X=\overline{x}, \tilde{U}=\overline{u}, \tilde{S}_0=1, \tilde{S}_1=1\right] \\
&\quad= \mathbb{E}\left[\tilde{Y}_1^* \,\middle|\, X=\overline{x}, \tilde{U}=\overline{u}, Q(0,\overline{x}) \geq \tilde{V}\right] \\
&\quad= \frac{\mathbb{E}\left[\mathbf{1}\left\{Q(0,\overline{x}) \geq \tilde{V}\right\} \cdot \tilde{Y}_1^* \,\middle|\, X=\overline{x}, \tilde{U}=\overline{u}\right]}{\mathbb{P}\left[Q(0,\overline{x}) \geq \tilde{V} \,\middle|\, X=\overline{x}, \tilde{U}=\overline{u}\right]} \\
&\quad= \frac{\mathbb{E}\left[\mathbf{1}\left\{Q(0,\overline{x}) \geq \tilde{V}\right\} \cdot \mathbb{E}\left[\tilde{Y}_1^* \,\middle|\, X=\overline{x}, \tilde{U}=\overline{u}, \tilde{V}\right] \,\middle|\, X=\overline{x}, \tilde{U}=\overline{u}\right]}{\mathbb{P}\left[Q(0,\overline{x}) \geq \tilde{V} \,\middle|\, X=\overline{x}, \tilde{U}=\overline{u}\right]} \\
&\quad= \frac{\int_0^{Q(0,\overline{x})} \mathbb{E}\left[\tilde{Y}_1^* \,\middle|\, X=\overline{x}, \tilde{U}=\overline{u}, \tilde{V}=v\right] \mathrm{d}F_{\tilde{V}|X,\tilde{U}}(v\,|x,\overline{u})}{\mathbb{P}\left[Q(0,\overline{x}) \geq \tilde{V} \,\middle|\, X=\overline{x}, \tilde{U}=\overline{u}\right]} \\
&\quad= \frac{\int_0^{Q(0,\overline{x})} \alpha(\overline{x},\overline{u})\, \mathrm{d}F_{\tilde{V}|X,\tilde{U}}(v\,|x,\overline{u})}{\mathbb{P}\left[Q(0,\overline{x}) \geq \tilde{V} \,\middle|\, X=\overline{x}, \tilde{U}=\overline{u}\right]} \\
&\qquad \text{by step 13} \\
&\quad= \alpha(\overline{x},\overline{u}). \qquad \text{(G.13)}
\end{aligned}
$$

Step 16. Notice that

$$\begin{aligned}
&\mathbb{E}\left[\tilde{Y}_0^* \,\middle|\, X=\overline{x}, \tilde{U}=\overline{u}, \tilde{S}_0=1, \tilde{S}_1=1\right] \\
&\quad = \mathbb{E}\left[\tilde{Y}_0^* \,\middle|\, X=\overline{x}, \tilde{U}=\overline{u}, Q\left(0,\overline{x}\right) \geq \tilde{V}\right] \\
&\quad = \frac{\mathbb{E}\left[\mathbf{1}\left\{Q\left(0,\overline{x}\right) \geq \tilde{V}\right\} \cdot \tilde{Y}_0^* \,\middle|\, X=\overline{x}, \tilde{U}=\overline{u}\right]}{\mathbb{P}\left[Q\left(0,\overline{x}\right) \geq \tilde{V} \,\middle|\, X=\overline{x}, \tilde{U}=\overline{u}\right]} \\
&\quad = \frac{\mathbb{E}\left[\mathbf{1}\left\{Q\left(0,\overline{x}\right) \geq \tilde{V}\right\} \cdot \mathbb{E}\left[\tilde{Y}_0^* \,\middle|\, X=\overline{x}, \tilde{U}=\overline{u}, \tilde{V}\right] \,\middle|\, X=\overline{x}, \tilde{U}=\overline{u}\right]}{\mathbb{P}\left[Q\left(0,\overline{x}\right) \geq \tilde{V} \,\middle|\, X=\overline{x}, \tilde{U}=\overline{u}\right]} \\
&\quad = \frac{\int\limits_0^{Q(0,\overline{x})} \mathbb{E}\left[\tilde{Y}_0^* \,\middle|\, X=\overline{x}, \tilde{U}=\overline{u}, \tilde{V}=v\right] \, \mathrm{d}F_{\tilde{V}|X,\tilde{U}}\left(v \,|x, \overline{u}\right)}{\mathbb{P}\left[Q\left(0,\overline{x}\right) \geq \tilde{V} \,\middle|\, X=\overline{x}, \tilde{U}=\overline{u}\right]} \\
&\quad = \frac{\int\limits_0^{Q(0,\overline{x})} \frac{m_0^Y\left(\overline{x},\overline{u}\right)}{m_0^S\left(\overline{x},\overline{u}\right)} \, \mathrm{d}F_{\tilde{V}|X,\tilde{U}}\left(v \,|x, \overline{u}\right)}{\mathbb{P}\left[Q\left(0,\overline{x}\right) \geq \tilde{V} \,\middle|\, X=\overline{x}, \tilde{U}=\overline{u}\right]} \\
&\qquad \text{by step 12} \\
&\quad = \frac{m_0^Y\left(\overline{x},\overline{u}\right)}{m_0^S\left(\overline{x},\overline{u}\right)}. \qquad \text{(G.14)}
\end{aligned}$$

Step 17. Note that

$$\begin{aligned}
\Delta_{\tilde{Y}^*}^{OO}\left(\overline{x},\overline{u}\right) &\coloneqq \mathbb{E}\left[\tilde{Y}_1^* - \tilde{Y}_0^* \,\middle|\, X=\overline{x}, \tilde{U}=\overline{u}, \tilde{S}_0=1, \tilde{S}_1=1\right] \\
&= \mathbb{E}\left[\tilde{Y}_1^* \,\middle|\, X=\overline{x}, \tilde{U}=\overline{u}, \tilde{S}_0=1, \tilde{S}_1=1\right] \\
&\quad - \mathbb{E}\left[\tilde{Y}_0^* \,\middle|\, X=\overline{x}, \tilde{U}=\overline{u}, \tilde{S}_0=1, \tilde{S}_1=1\right] \\
&= \alpha\left(\overline{x},\overline{u}\right) - \frac{m_0^Y\left(\overline{x},\overline{u}\right)}{m_0^S\left(\overline{x},\overline{u}\right)} \\
&\qquad \text{by equations (G.13) and (G.14)} \\
&= \delta\left(\overline{x},\overline{u}\right) \\
&\qquad \text{by the definition of } \alpha\left(\overline{x},\overline{u}\right),
\end{aligned}$$

ensuring that equation (G.1) holds.

Finally, I show, in four steps, that equation (G.5) holds.

Step 18. Fix $(y, d, s, z) \in \mathbb{R}^4$ arbitrarily and observe that expression (G.5) can be simplified to:

$$\begin{aligned}
&\left| F_{\tilde{Y},\tilde{D},\tilde{S},Z,X}\left(y, d, s, z, \overline{x}\right) - F_{Y,D,S,Z,X}\left(y, d, s, z, \overline{x}\right) \right| \leq \epsilon \\
\Leftrightarrow &\left| F_{\tilde{Y},\tilde{D},\tilde{S},Z|X}\left(y, d, s, z \,|\overline{x}\right) \cdot F_X\left(\overline{x}\right) - F_{Y,D,S,Z|X}\left(y, d, s, z \,|\overline{x}\right) \cdot F_X\left(\overline{x}\right) \right| \leq \epsilon \\
\Leftrightarrow &\left| F_{\tilde{Y},\tilde{D},\tilde{S},Z|X}\left(y, d, s, z \,|\overline{x}\right) - F_{Y,D,S,Z|X}\left(y, d, s, z \,|\overline{x}\right) \right| \leq \frac{\epsilon}{F_X\left(\overline{x}\right)} \\
\Leftrightarrow &\left| F_{\tilde{Y},\tilde{D},\tilde{S},Z|X}\left(y, d, s, z \,|\overline{x}\right) - F_{Y,D,S,Z|X}\left(y, d, s, z \,|\overline{x}\right) \right| \leq 2 \cdot \overline{\epsilon} \qquad \text{(G.15)}
\end{aligned}$$

by the definition of $\overline{\epsilon}$.

Step 19. Notice that

$$\begin{aligned}
&F_{\tilde{Y},\tilde{D},\tilde{S},Z|X}\left(y, d, s, z \,|\overline{x}\right) - F_{Y,D,S,Z|X}\left(y, d, s, z \,|\overline{x}\right) \\
&= \mathbb{E}\left[\mathbf{1}\left\{\left(\tilde{Y}, \tilde{D}, \tilde{S}, Z\right) \leq (y, d, s, z)\right\}\middle| X = \overline{x}\right] - \mathbb{E}\left[\mathbf{1}\left\{(Y, D, S, Z) \leq (y, d, s, z)\right\}\middle| X = \overline{x}\right] \\
&= \int \mathbf{1}\left\{\left(\tilde{Y}, \tilde{D}, \tilde{S}, Z\right) \leq (y, d, s, z)\right\} \mathrm{d}F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0, y_1, u, v, z \,|\overline{x}\right) \\
&\quad - \int \mathbf{1}\left\{(Y, D, S, Z) \leq (y, d, s, z)\right\} \mathrm{d}F_{Y_0^*,Y_1^*,U,V,Z|X}\left(y_0, y_1, u, v, z \,|\overline{x}\right) \\
&= \int \left[\mathbf{1}\left\{\left(\tilde{Y}, \tilde{D}, \tilde{S}, Z\right) \leq (y, d, s, z)\right\} \cdot \mathbf{1}\left\{u \notin \left(\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}\right)\right\}\right] \mathrm{d}F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0, y_1, u, v, z \,|\overline{x}\right) \\
&\quad + \int \left[\mathbf{1}\left\{\left(\tilde{Y}, \tilde{D}, \tilde{S}, Z\right) \leq (y, d, s, z)\right\} \cdot \mathbf{1}\left\{u \in \left(\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}\right)\right\}\right] \mathrm{d}F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0, y_1, u, v, z \,|\overline{x}\right) \\
&\quad - \int \left[\mathbf{1}\left\{(Y, D, S, Z) \leq (y, d, s, z)\right\} \cdot \mathbf{1}\left\{u \notin \left(\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}\right)\right\}\right] \mathrm{d}F_{Y_0^*,Y_1^*,U,V,Z|X}\left(y_0, y_1, u, v, z \,|\overline{x}\right) \\
&\quad - \int \left[\mathbf{1}\left\{(Y, D, S, Z) \leq (y, d, s, z)\right\} \cdot \mathbf{1}\left\{u \in \left(\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}\right)\right\}\right] \mathrm{d}F_{Y_0^*,Y_1^*,U,V,Z|X}\left(y_0, y_1, u, v, z \,|\overline{x}\right) \\
&\quad \text{by linearity of the Lebesgue Integral} \\
&= \int \left[\mathbf{1}\left\{(Y, D, S, Z) \leq (y, d, s, z)\right\} \cdot \mathbf{1}\left\{u \notin \left(\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}\right)\right\}\right] \mathrm{d}F_{Y_0^*,Y_1^*,U,V,Z|X}\left(y_0, y_1, u, v, z \,|\overline{x}\right) \\
&\quad + \int \left[\mathbf{1}\left\{\left(\tilde{Y}, \tilde{D}, \tilde{S}, Z\right) \leq (y, d, s, z)\right\} \cdot \mathbf{1}\left\{u \in \left(\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}\right)\right\}\right] \mathrm{d}F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0, y_1, u, v, z \,|\overline{x}\right) \\
&\quad - \int \left[\mathbf{1}\left\{(Y, D, S, Z) \leq (y, d, s, z)\right\} \cdot \mathbf{1}\left\{u \notin \left(\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}\right)\right\}\right] \mathrm{d}F_{Y_0^*,Y_1^*,U,V,Z|X}\left(y_0, y_1, u, v, z \,|\overline{x}\right) \\
&\quad - \int \left[\mathbf{1}\left\{(Y, D, S, Z) \leq (y, d, s, z)\right\} \cdot \mathbf{1}\left\{u \in \left(\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}\right)\right\}\right] \mathrm{d}F_{Y_0^*,Y_1^*,U,V,Z|X}\left(y_0, y_1, u, v, z \,|\overline{x}\right) \\
&\quad \text{by steps 2-6}
\end{aligned}$$

$$
\begin{aligned}
&= \int \left[\mathbf{1}\left\{ \left(\tilde{Y}, \tilde{D}, \tilde{S}, Z\right) \leq (y,d,s,z) \right\} \cdot \mathbf{1}\left\{ u \in (\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}) \right\} \right] \mathrm{d}F_{\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}, Z|X}\left(y_0, y_1, u, v, z \,|\overline{x}\right) \\
&\qquad - \int \left[\mathbf{1}\left\{ (Y,D,S,Z) \leq (y,d,s,z) \right\} \cdot \mathbf{1}\left\{ u \in (\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}) \right\} \right] \mathrm{d}F_{Y_0^*, Y_1^*, U, V, Z|X}\left(y_0, y_1, u, v, z \,|\overline{x}\right) \\
&\leq \int \mathbf{1}\left\{ u \in (\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}) \right\} \mathrm{d}F_{\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}, Z|X}\left(y_0, y_1, u, v, z \,|\overline{x}\right) \\
&= \int \mathbf{1}\left\{ u \in (\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}) \right\} \mathrm{d}F_{\tilde{U}|X}\left(u \,|\overline{x}\right) \\
&= 2 \cdot \overline{\epsilon}
\end{aligned}
$$

by step 5.

Step 20. Following the same procedure of step 19, I have that:

$$
\begin{aligned}
&F_{\tilde{Y}, \tilde{D}, \tilde{S}, Z|X}\left(y,d,s,z \,|\overline{x}\right) - F_{Y,D,S,Z|X}\left(y,d,s,z \,|\overline{x}\right) \\
&\quad= \int \left[\mathbf{1}\left\{ \left(\tilde{Y}, \tilde{D}, \tilde{S}, Z\right) \leq (y,d,s,z) \right\} \cdot \mathbf{1}\left\{ u \in (\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}) \right\} \right] \mathrm{d}F_{\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}, Z|X}\left(y_0, y_1, u, v, z \,|\overline{x}\right) \\
&\qquad - \int \left[\mathbf{1}\left\{ (Y,D,S,Z) \leq (y,d,s,z) \right\} \cdot \mathbf{1}\left\{ u \in (\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}) \right\} \right] \mathrm{d}F_{Y_0^*, Y_1^*, U, V, Z|X}\left(y_0, y_1, u, v, z \,|\overline{x}\right) \\
&\quad\geq - \int \mathbf{1}\left\{ u \in (\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}) \right\} \mathrm{d}F_{Y_0^*, Y_1^*, U, V, Z|X}\left(y_0, y_1, u, v, z \,|\overline{x}\right) \\
&\quad= - \int \mathbf{1}\left\{ u \in (\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}) \right\} \mathrm{d}F_{U|X}\left(u \,|\overline{x}\right) \\
&\quad= -2 \cdot \overline{\epsilon}
\end{aligned}
$$

Step 21. Combining steps 19 and 20, I find that

$$
\left| F_{\tilde{Y}, \tilde{D}, \tilde{S}, Z|X}\left(y,d,s,z \,|\overline{x}\right) - F_{Y,D,S,Z|X}\left(y,d,s,z \,|\overline{x}\right) \right| \leq 2 \cdot \overline{\epsilon},
$$

implying equation (G.5) according to equation (G.15).

I can, then, conclude that proposition G.2 is true. ■

Proof of Proposition G.3. This proof is essentially the same proof of Proposition G.2 under assumption 7.3.(a). Fix any $\overline{u} \in [0,1]$, any $\overline{x} \in \mathcal{X}$, any $\delta(\overline{x}, \overline{u}) \in \left(\underline{\Delta_{Y^*}^{OO}}(\overline{x}, \overline{u}), \overline{\Delta_{Y^*}^{OO}}(\overline{x}, \overline{u}) \right)$ and any $\epsilon \in \mathbb{R}_{++}$ such that $\min\left\{ \overline{u} - \dfrac{\epsilon}{2 \cdot F_X(\overline{x})}, 1 - \left(\overline{u} - \dfrac{\epsilon}{2 \cdot F_X(\overline{x})} \right) \right\} > 0$. For brevity, define $\alpha\left(\overline{x},\overline{u}\right) := \delta\left(\overline{x},\overline{u}\right) + \dfrac{m_0^Y\left(\overline{x},\overline{u}\right)}{m_0^S\left(\overline{x},\overline{u}\right)}$, $\gamma\left(\overline{x},\overline{u}\right) := \dfrac{m_1^Y\left(\overline{x},\overline{u}\right) - \alpha\left(\overline{x},\overline{u}\right)\cdot m_0^S\left(\overline{x},\overline{u}\right)}{\Delta_S\left(\overline{x},\overline{u}\right)}$ and $\overline{\epsilon} := \dfrac{\epsilon}{2\cdot F_X\left(\overline{x}\right)}$. Note that $\alpha\left(\overline{x},\overline{u}\right) \in \mathbb{R} = \mathcal{Y}^*$ and $\gamma\left(\overline{x},\overline{u}\right) \in \mathbb{R} = \mathcal{Y}^*$.

I define the random variables $\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)$ using the joint cumulative distribution function $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}$ described by steps 1-14 in the proof of proposition G.2 for the case of convex support $\mathcal{Y}^*$. Note that equation (G.7) is trivially true when $\mathcal{Y}^* = \mathbb{R}$. Moreover, equations (G.6) and (G.10) are valid by the argument described in steps 15-21 in the previous proof.

I can, then, conclude that proposition G.3 is true. ■