MTE with Misspecification*

Julián Martínez-Iriarte †
Department of Economics
UC Santa Cruz

Pietro Emilio Spini ‡
Department of Economics
University of Bristol

This draft: June 2023
First draft: June 2021

Abstract

This paper studies the implication of a fraction of the population not responding to the instrument when selecting into treatment. We show that, in general, the presence of non-responders biases the Marginal Treatment Effect (MTE) curve and many of its functionals. We obtain partial identification results for the MTE, the LATE and the MPRTE. Moreover, in a linear-in-covariates model we obtain a testable implication for the presence of non-responders. Simulation results show that our proposed estimators, using the range of the propensity score, work well.

Keywords: Marginal Treatment Effects, Local Instrumental Variable, Misspecification, Partial Identification, Inference on Extrema

*We thank Sukjin Han, Vitor Possebom, Yixiao Sun, Kaspar Wuthrich, and seminar participants at UC San Diego for helpful comments. All remaining errors are ours.
†Email: jmart425@ucsc.edu
‡Email: pietro.spini@bristol.ac.uk

1 Introduction

Marginal treatment effects (MTEs) have unified the identification theory of several policy parameters. While the MTE framework is essentially non-parametric,[1] it is required that the recipient's participation into treatment follows a (generalized) Roy model. Such a feature can be interpreted as a form of additive separability: a comparison of costs on one side and benefits on the other side determines selection into treatment. Identification of the MTE is achieved via the local instrumental variable (LIV) approach (Heckman and Vytlacil (2001, 2005)). An excellent survey is provided by Mogstad and Torgovitsky (2018). An early effort to analyze MTE under misspecification can be found in the appendix of the seminal paper by Heckman and Vytlacil (2001). They consider a case where the additive separability in the selection equation does not hold. The most serious consequence is that the LIV approach does not identify the MTE curve.

In this paper we analyze a different type of misspecification. We model a situation in which, under additive separability, a proportion of the population does not take into account the instrumental variable when deciding whether to take up treatment or not. We refer to them as *non-responders*. To analyze the resulting bias, we define a pseudo-MTE curve which results from applying the LIV approach over a population where a fraction does not respond the instrumental variable. Under no misspecification, the pseudo-MTE curve would coincide with the MTE curve. The resulting bias can be interpreted as a location-scale change of the MTE curve, parameterized by the proportion of non-responders and their propensity score.

We have two main results. First we show that it is possible to bound the proportion of non-responders. This, in turn, allows to provide bounds on the MTE curve, and in parameters derived from it. In particular, we focus on the local average treatment effect (LATE) and the marginal policy relevant treatment effect (MPRTE). Second, we show that in a setting with linear-in-covariates potential outcomes, there is a testable implication for the presence of non-responders.

The bounds depend on the range of the observed propensity score. We refer to this quantity as a re-scaling correction. Our estimator follows the intersection bounds method proposed by Chernozhukov, Lee, and Rosen (2013). We perform a Monte Carlo simulations to asses the performance of the method.

Recently, Acerenza, Ban, and Kedagni (2021), Possebom (2021), and Acerenza (2023) focus on the effect of measurement error in treatment status on the MTE curve. We complement such results by noting that a simple change to our setup can cover the case of misclassification. In a setting where treatment status is misclassified, the observed outcome is generated with the true treatment status. In our setting of misclassification, the observed outcome can be regarded as a mixture of responders and non-responders. The proportion of non-responders is analogous to the proportion of misreporters. Indeed, our results also hold if instead of having a fraction of non-responders, we have a fraction of misreporters.

The rest of the paper is organized as follows: Section 2 introduces the model; Section 3

[1] Linearity is sometimes assumed to facilitate estimation. See, *e.g.*, Appendix B in Heckman, Urzua, and Vytlacil (2006). In Section 4 of this paper, we show that linearity can help detect non-responders.

contains the main partial identification results; Section 4 shows how to detect the presence of non-responders in linear models; and Section 7 concludes.

2 Misspecification and MTE

In this section we introduce our model for misspecification in the MTE framework (Bjorklund and Moffitt (1987), Heckman and Vytlacil (2001, 2005)). We analyze the consequences of misspecification from the identification point of view.

2.1 The Model

We start with a general non-separable potential outcome model

$$\begin{aligned} Y(0) &= h_0(X, U_0), \\ Y(1) &= h_1(X, U_1), \\ Y &= D^*Y(1) + (1 - D^*)Y(0), \end{aligned}$$

where D^* is the observed treatment status, X are observable covariates with support denoted by $\mathcal{X}$, and $\{Y(0), Y(1)\}$ and Y are potential and observed outcomes, respectively. The functions h_0 and h_1 are unknown.

We model misspecification as a situation where there are two types of individuals: responders and non-responders. Responders select into treatment taking into account the incentives in Z. Their selection equation is given by $D = \mathbb{1}\{\mu(X, Z) \geq V\}$. On the other hand, non-responders do not react to incentives in Z at all. Their selection equation is given by $\tilde{D} = \mathbb{1}\{\tilde{\mu}(X) \geq \tilde{V}\}$. Notice how Z is not featured in $\tilde{\mu}(\cdot)$. For the non-responders, Z fails the relevance condition of the standard MTE model. Both μ and $\tilde{\mu}$ are unknown.

Let R be the latent status of an individual: $R = 1$ for a responder and $R = 0$ for a non-responder. The observed treatment status D^* is given by:

$$D^* = R \cdot D + (1 - R) \cdot \tilde{D}. \tag{1}$$

We allow for the proportion of non-responders to vary with X. To this end, we define $\delta_X = \Pr(R = 0|X) = \Pr(D^* = \tilde{D}|X)$. Thus, for every subpopulation with characteristics $X = x$ there is a proportion $\delta_x \in [0, 1)$ of non-responders. We consider values where $\sup_{x \in \mathcal{X}} \delta_x < 1$ to avoid a situation where no-one responds to the instrumental variable.

Remark 1
We observe Y according to $Y = D^*Y(1) + (1 - D^*)Y(0)$, which is given by the actual choice D^*. If, instead, we have $Y = DY(1) + (1 - D)Y(0)$, then we can interpret D^* as a misclassified treatment status. In this case, all individuals decide according to $D = \mathbb{1}\{\mu(X, Z) \geq V\}$, but a fraction of them

reports according to $\tilde{D} = \mathbb{1}\left\{\tilde{\mu}(X) \geq \tilde{V}\right\}$ *See Acerenza, Ban, and Kedagni (2021) and Possebom (2021) for recent studies on MTE under misclassification.*

The econometrician observes a cross section of (Y_i, D_i^*, X_i, Z_i). The latent status R is not observed. When $\delta_X = 0$ almost surely, then $D^* = D$ and we are in the familiar MTE framework of Heckman and Vytlacil (2001, 2005). Otherwise, if $\delta_X \neq 0$ almost surely, for an observation of D_i^*, we do not know whether we are observing the treatment status of a non-responder or of a responder. That is, it is unknown if we are observing D_i or $\tilde{D}_i$.

Assumption 1
Type Conditional Independence. *$R \perp Z \| X$.*

Assumption 1 states that once we control for X, the latent status of a individuals does not vary with the instrumental variable Z. In this paper, we largely regard the non-respondent type as a pre-existing type, possibly (and likely) correlated to the covariates X. Because generally Z may be (conditionally) randomized by the policy-maker, we believe Assumption 1 is realistic in most scenarios where an econometrician would want to elicit a policy parameter like the MTE.

Assumption 2
Relevance and Exogeneity

1. *$\mu(X, Z)$ is a nondegenerate random variable conditional on X.*
2. *$(U_0, U_1, V, \tilde{V})$ are independent of Z conditional on X.*

Note that, for the subpopulation of non-responders, the instrument is valid but totally irrelevant. The larger the value of δ_x, the "weaker" the instrument Z, since most participants with $X = x$ are non-responders. With the exception of the requirement that $\tilde{V} \perp Z \| X$, these are the same conditions of Heckman and Vytlacil (2001, 2005). Our additional requirement covers the subpopulation of non-responders: neither the "cost" of treatment $\tilde{V}$ nor the "benefit" $\tilde{\mu}(X)$ depend on Z when conditioned on X.

Example 1
To fix ideas, we can think of a two part cost of providing the incentive. A fixed cost associated to targeting a particular subpopulation with covariates $X = x$ and the cost of the incentive itself. If Z is a voucher, there could be administrative costs associated to making it available to subpopulation $X = x$. For non-responders who do not redeem the voucher, the cost of the incentive is zero. Such a scenario would satisfy Assumption 2.

The misclassification structure of Equation (1) allows to define three different propensity scores. An observed/identified one which is based on the observables (D^*, X, Z), and two latent/unobserved propensity scores: one for the reponders and one for the non-responders. Formally, they are given by

$$P^*(X,Z) := \Pr(D^* = 1|X,Z) \qquad \textbf{(Observed)}$$

$$P(X,Z) := \Pr(D = 1|R = 1, X, Z) \qquad \textbf{(Latent: responders)}$$

$$\tilde{P}(X) := \Pr(\tilde{D} = 1|R = 0, X) \qquad \textbf{(Latent: non-responders)}$$

Lemma 1
Under Assumptions 1 and 2.2 we can relate the different propensity scores by

$$P^*(X,Z) = (1-\delta_X)\cdot P(X,Z) + \delta_X \cdot \tilde{P}(X). \tag{2}$$

For a fixed $X = x$, the result in Lemma 1 shows that the observed propensity (still random through Z) is an affine transformation of the propensity score for the responders. If, additionally, we take two different values of Z, for example z and z', we can remove the contribution of $\tilde{P}(X)$, which is invariant with respect to z and obtain[2]

$$P^*(x,z) - P^*(x,z') = (1-\delta_x)\cdot\left[P(x,z) - P(x,z')\right] \tag{3}$$

Equation (3) says that the changes on the observed propensity score induced by varying Z are proportional to the changes on the true propensity score induced by varying Z. Thus, if we knew δ_x, we could recover the change in the propensity score for the responders. When Z is continuous, we can take a limiting version of this argument, *e.g.*, as $z' \to z$, to obtain

$$\frac{\partial P^*(x,z)}{\partial z} = (1-\delta_x)\cdot\frac{\partial P(x,z)}{\partial z}. \tag{4}$$

Both the discrete (equation (3)), and the continuous (equation (4)) change in the propensity score play a role in the relationship between the MTE curve (defined below) and certain parameters of interest. In both cases, there is an attenuation effect.

Remark 2
Weak Instrument *Note that a possible indication of the presence of a weak instrument Z is given by the average derivative of the observed propensity score. In our framework, for a given value of δ_x, this equals an attenuated version of the average derivative of the true propensity score, because of equation (4):*

$$E\left[\frac{\partial P^*(x,Z)}{\partial z}\right] = (1-\delta_x)E\left[\frac{\partial P(x,Z)}{\partial z}\right].$$

Thus, even if the instrument is strong for the subpopulation of responders, it might appear as weak in the sample. See Assumption 2 in Possebom (2021) for a related assumption.

[2]We write $P^*(x,z)$ for $\Pr(D^* = 1|X = x, Z = z)$, and $P(x,z)$ for $\Pr(D = 1|R = 1, X = x, Z = z)$.

2.2 The MTE for Responders

For the subpopulation of responders, the standard MTE framework holds. This motivates us to define an MTE curve for this subpopulation. In doing so, we are implicitly assuming that this is our object of interest. The reason for this is that many times we can also control the instrumental variable Z. Thus, to asses the effects of manipulations of the incentives in Z we look at the MTE curve for responders.

Let $\mathcal{P}_x$ and $\mathcal{P}_x^*$ denote the support of $P(x, Z) := \Pr(D = 1|X = x, Z)$ and $P^*(x, Z) := \Pr(D^* = 1|X = x, Z)$ respectively. For the subpopulation of responders, we rewrite the selection equation as $D = \mathbb{1}\{P(X, Z) \geq U_D\}$ where $U_D \sim U_{(0,1)}$.[3] Thus, we define the MTE curve for responders as

$$\text{MTE}(u, x) := \mathbb{E}[Y(1) - Y(0)|R = 1, U_D = u, X = x].$$

By the LIV approach we have the following equivalence result:[4]

$$\text{MTE}(u, x) = \frac{\partial \mathbb{E}[Y|R = 1, P(X, Z) = u, X = x]}{\partial u} \text{ for } u \in \mathcal{P}_x. \tag{5}$$

Since we do not observe D, we cannot estimate $P(X, Z)$, and hence this is *not* an identification result in our setting. In a similar fashion, we *define* the following pseudo-MTE curve:

$$\text{MTE}^*(u, x; \delta_x) := \frac{\partial \mathbb{E}[Y|P^*(X, Z) = u, X = x]}{\partial u} \text{ for } u \in \mathcal{P}_x^*. \tag{6}$$

We emphasize that the pseudo-MTE curve is indexed by δ_x because it depends implicitly on the proportion of the nonresponders. From the data, we can only compute $\text{MTE}^*(u, x; \delta_x)$ over $\mathcal{P}_x^*$, not $\text{MTE}(u, x)$. The pseudo-MTE curve is the curve that would be mistakenly taken to be the MTE curve. Indeed, in the absence of non-responders, $\text{MTE}^*(u, x; 0) = \text{MTE}(u, x)$. If non-responders are present in the $X = x$ subpopulation, that is if $\delta_x > 0$, the observed $\text{MTE}^*(u, x; \delta_x)$ does not identify $\text{MTE}(u, x)$. In another words, the LIV approach is biased. We can now fully characterize the bias induced by δ_x on the MTE curve.

Lemma 2
Under Assumptions 1 and 2, we can write

$$MTE(v, x) = (1 - \delta_x) MTE^*\left((1 - \delta_x)v + \delta_x \tilde{P}(x), x; \delta_x\right) \text{ for } v \in \mathcal{P}_x, \tag{7}$$

and

$$MTE^*(u, x; \delta_x) = \frac{1}{1 - \delta_x} MTE\left(\frac{u - \delta_x \tilde{P}(x)}{1 - \delta_x}, x\right) \text{ for } u \in \mathcal{P}_x^*. \tag{8}$$

[3]This follows from $D = \mathbb{1}\{F_{V|R,X,Z}(\mu(X, Z)|1, X, Z) \geq F_{V|R,X,Z}(V|1, X, Z)\}$. Noting that by assumptions 1 and 2.(2), we have $D = \mathbb{1}\{P(X, Z) \geq F_{V|R,X}(V|1, X)\}$. Finally, we take $U_D := F_{V|R,X}(V|1, X)$.

[4]See the section B in the appendix for a derivation. Also see Heckman and Vytlacil (2001) for sufficient conditions on the standard case (with no non-responders).

Lemma 2 shows that the bias is in the form of both location and scale. Equation (8), which is equivalent to Equation (7),[5] shows that MTE* is obtained by changing the location from u to $u - \delta_x \tilde{P}(x)$, and rescaling by $(1-\delta_x)^{-1}$. Thus, as in a location-scale family of densities, we can regard MTE* as a family of curves, defined over $\mathcal{P}_x^*$, which is indexed by δ_x and $\tilde{P}(x)$.

The usefulness of Lemma 2 stems from the fact that it is possible to recover parameters that are based on the MTE curve *without* having to recover the MTE curve in the first place. The is covered in the next corollary. We focus on two parameters: the local average treatment effect (LATE) and the marginal policy relevant treatment effect (MPRTE). We focus on these two parameters because they are identified under virtually no support restriccions on the propensity score.[6] The parameters are given by:

$$\text{LATE}(x, P(x,z), P(x,z')) = \frac{1}{P(x,z) - P(x,z')} \int_{P(x,z')}^{P(x,z)} \text{MTE}(u,x)du.$$

and[7]

$$\text{MPRTE}(x) = \int_{\mathcal{Z}} \text{MTE}(P(x,z),x) \frac{\partial P(x,z)}{\partial z} \left(E\left[\frac{\partial [P(x,Z)]}{\partial z} \right] \right)^{-1} f_{Z|X}(z|x)dz.$$

The versions under misspecification are given by

$$\text{LATE}^*(x, P^*(x,z), P^*(x,z')) = \frac{1}{P^*(x,z) - P^*(x,z')} \int_{P^*(x,z')}^{P^*(x,z)} \text{MTE}^*(u,x;\delta_x)du \tag{9}$$

and

$$\text{MPRTE}^*(x) = \int_{\mathcal{Z}} \text{MTE}^*(P^*(x,z),x;\delta_x) \frac{\partial P^*(x,z)}{\partial z} \left(E\left[\frac{\partial [P^*(x,Z)]}{\partial z} \right] \right)^{-1} f_{Z|X}(z|x)dz.$$

The next corollary relates both sets of parameters. In particular, it shows that there is a local attenuation and that the signs are preserved.

Corollary 1

Under the assumptions of Lemma 2, the LATE, for $P(x,z') < P(x,z)$ with $z, z' \in \mathcal{Z}$ satisfies

$$LATE(x, P(x,z), P(x,z')) = (1-\delta_x) LATE^*(x, P^*(x,z), P^*(x,z')).$$

The marginal policy relevant treatment effect (MPRTE) is an average of the $MTE(u,x)$ along the margin of indifference: when $U_D = P(X,Z)$. It satisfies

$$MPRTE(x) = (1-\delta_x) MPRTE^*(x).$$

[5] Note the changes in the domain of integration between (7) and (8).
[6] For example, the CATE requires full support on the propensity score.
[7] This representation of the MPRTE is based on Martinez-Iriarte and Sun (2022).

3 Partial identification

The result of Corollary 1 implies that if we can bound the proportion of non-responders δ_x, then we can bound both the LATE and the MPRTE. The next assumption will yield a partial identified interval for δ_x.

Assumption 3
Interval Support. *The support of $P(x, Z)$ is an interval $\mathcal{P}_x := [\underline{p_x}, \overline{p_x}]$ for every x in a subset $\mathcal{X}_B \subseteq \mathcal{X}$.*

The combination of Assumption 3 and equation (2) yields that the support of the observed propensity score, denoted $\mathcal{P}_x^* := [\underline{p_x^*}, \overline{p_x^*}]$, is given by

$$\underline{p_x^*} = (1 - \delta_x)\underline{p_x} + \delta_x \tilde{P}(x)$$
$$\overline{p_x^*} = (1 - \delta_x)\overline{p_x} + \delta_x \tilde{P}(x)$$

Taking the difference, we obtain $\overline{p_x^*} - \underline{p_x^*} = (1 - \delta_x)(\overline{p_x} - \underline{p_x})$. Since $\overline{p_x} - \underline{p_x}$, the unobserved range of $P(x, Z)$, has to be less than 1, then, $\overline{p_x^*} - \underline{p_x^*} \leq 1 - \delta_x$. This implies that the proportion of non-responders is bounded by

$$0 \leq \delta_x \leq 1 - (\overline{p_x^*} - \underline{p_x^*}).$$

Naturally, in the above, we cannot provide a lower bound. That is, we always have to allow for the possibility that there are no non-responders. Conversely,

$$(\overline{p_x^*} - \underline{p_x^*}) \leq 1 - \delta_x \leq 1$$

We refer to $\overline{p_x^*} - \underline{p_x^*}$ as the *re-scaling correction*. In view of the results of Corollary 1, we have the following bounds on the magnitudes of LATE and the MPRTE:

$$(\overline{p_x^*} - \underline{p_x^*}) \left|\text{LATE}^*(x, P^*(x,z), P^*(x,z'))\right| \leq \left|\text{LATE}(x, P(x,z), P(x,z'))\right| \leq \left|\text{LATE}^*(x, P^*(x,z), P^*(x,z'))\right|, \tag{10}$$

and

$$(\overline{p_x^*} - \underline{p_x^*}) \left|\text{MPRTE}^*(x)\right| \leq \left|\text{MPRTE}(x)\right| \leq \left|\text{MPRTE}^*(x)\right|.$$

In particular, in both cases the sign is preserved. While we bound the magnitudes, we note that the sign of the true parameters coincide with the sign of the parameters under misspecification.

Note that the previous bounds do not rely on bounding the MTE curve using the pseudo-MTE curve. However, in some cases, the MTE curve itself is of independent interest. For example, see

Possebom (2021). We use Lemma 2. Indeed, using (7) we obtain

$$\begin{aligned}|\text{MTE}(v,x)| &= (1-\delta_x)\left|\text{MTE}^*\left((1-\delta_x)v+\delta_x\tilde{P}(x),x;\delta_x\right)\right|\\ &\leq \left|\text{MTE}^*\left((1-\delta_x)v+\delta_x\tilde{P}(x),x;\delta_x\right)\right|\\ &\leq \max_{u\in\mathcal{P}_x^*}\left|\text{MTE}^*(u,x;\delta_x)\right|.\end{aligned}$$

The first inequality follows from the fact that $(1-\delta_x)\leq 1$. The max in the second inequality is well defined by the continuity of the pseudo-MTE curve. To derive a lower bound, we use re-scaling correction which satisfies $\overline{p_x^*}-\underline{p_x^*}\leq 1-\delta_x$. Thus,

$$\begin{aligned}|\text{MTE}(v,x)| &= (1-\delta_x)\left|\text{MTE}^*\left((1-\delta_x)v+\delta_x\tilde{P}(x),x;\delta_x\right)\right|\\ &\geq (\overline{p_x^*}-\underline{p_x^*})\left|\text{MTE}^*\left((1-\delta_x)v+\delta_x\tilde{P}(x),x;\delta_x\right)\right|\\ &\geq (\overline{p_x^*}-\underline{p_x^*})\min_{u\in\mathcal{P}_x^*}\left|\text{MTE}^*(u,x;\delta_x)\right|.\end{aligned}$$

Putting all together, we obtain that, for $v\in\mathcal{P}_x$, the following bounds hold

$$(\overline{p_x^*}-\underline{p_x^*})\min_{u\in\mathcal{P}_x^*}\left|\text{MTE}^*(u,x;\delta_x)\right|\leq|\text{MTE}(v,x)|\leq\max_{u\in\mathcal{P}_x^*}\left|\text{MTE}^*(u,x;\delta_x)\right|.$$

In contrast to the case of LATE and MPRTE, the signs are not preserved in the case of the MTE curve.

4 Detecting non-responders

In this section we show that the presence of non-responders has a testable implication. We follow Appendix B in Heckman, Urzua, and Vytlacil (2006). The model is

$$\begin{aligned}Y(0) &= \alpha+\beta_0X+U_0\\ Y(1) &= \alpha+\varphi+\beta_1X+U_1.\end{aligned}$$

The selection equation for responders and non-responders are $D=\mathbb{1}\{\gamma Z\geq V\}$ and $\tilde{D}=\mathbb{1}\{0\geq\tilde{V}\}$ respectively. As before, let $R=1$ for a responder and $R=0$ for a non-responder. Therefore, observed treatment is $D^*=R\cdot D+(1-R)\cdot\tilde{D}$.

To derive the LIV representation, we use equation (B2) in Heckman, Urzua, and Vytlacil (2006) to arrive at

$$\mathbb{E}[Y|R=1,P(X,Z)=u,X=x]=\alpha+\beta_0x+(\beta_1-\beta_0)xu+K(u)$$

where $K(u) = \varphi u + \mathbb{E}[U_1 - U_0 | R = 1, D = 1, P(Z) = u]u$. Therefore, by equation (5), we obtain

$$\begin{aligned}\text{MTE}(u, x) &= \frac{\partial \mathbb{E}\left[Y | R = 1, P(X, Z) = u, X = x\right]}{\partial u} \text{ for } u \in \mathcal{P}_x \\ &= (\beta_1 - \beta_0)x + K'(u).\end{aligned}$$

Of course, this is unfeasible since we do not observe the latent status R. Using the result in Lemma 2 we can write an expression for the pseudo-true MTE curve:

$$\begin{aligned}\text{MTE}^*(u, x; \delta_x) &= \frac{1}{1 - \delta_x}\text{MTE}\left(\frac{u - \delta_x \tilde{P}(x)}{1 - \delta_x}, x\right) \text{ for } u \in \mathcal{P}_x^* \\ &= \frac{1}{1 - \delta_x}(\beta_1 - \beta_0)x + \frac{1}{1 - \delta_x}K'\left(\frac{u - \delta_x \tilde{P}(x)}{1 - \delta_x}\right).\end{aligned}$$

For a fixed x, we have the following result for the map $u \mapsto \text{MTE}(u, x)$:

$$\frac{\partial \text{MTE}(u, x)}{\partial u} = K''(u).$$

which does not depend on x. However,

$$\frac{\partial \text{MTE}^*(u, x; \delta_x)}{\partial u} = \frac{1}{(1 - \delta_x)^2}K''\left(\frac{u - \delta_x \tilde{P}(x)}{1 - \delta_x}\right)$$

does depend on x. We can use this to test for the presence of non-responders. Note that $\frac{\partial \text{MTE}^*(u,x;\delta_x)}{\partial u}$ is identified from the data as it is the second derivative of

$$\frac{\partial \text{MTE}^*(u, x; \delta_x)}{\partial u} = \frac{\partial^2 \mathbb{E}\left[Y | P^*(X, Z) = u, X = x\right]}{\partial u^2} \text{ for } u \in \mathcal{P}_x^*.$$

We state this formally in the lemma below.

Lemma 3

Let x_1 and x_2 be two elements of $\mathcal{X}$ such that $\mathcal{P}_{x_1}^* \cap \mathcal{P}_{x_2}^* \neq \emptyset$. If, for some $\tilde{u} \in \mathcal{P}_{x_1}^* \cap \mathcal{P}_{x_2}^*$,

$$\left.\frac{\partial^2 \mathbb{E}\left[Y | P^*(X, Z) = u, X = x_1\right]}{\partial u^2}\right|_{u=\tilde{u}} \neq \left.\frac{\partial^2 \mathbb{E}\left[Y | P^*(X, Z) = u, X = x_2\right]}{\partial u^2}\right|_{u=\tilde{u}},$$

then either $\delta_{x_1} \neq 0$ or $\delta_{x_2} \neq 0$.

The key assumption is that we can find at least two overlapping (conditional) propensity scores $\mathcal{P}_{x_1}^*$ and $\mathcal{P}_{x_2}^*$. A simple implementation of the lemma is the following. Suppose X is univariate, and run a (series) regression

$$\begin{aligned}Y_i = \alpha_0 &+ \alpha_1 \hat{P}(X_i, Z_i) + \alpha_2 \hat{P}(X_i, Z_i)^2 + \alpha_3 X_i + \alpha_4 X_i^2 \\ &+ \alpha_5 \hat{P}(X_i, Z_i)X_i + \alpha_6 \hat{P}(X_i, Z_i)^2 X_i + \alpha_7 \hat{P}(X_i, Z_i)X_i^2 + \alpha_8 \hat{P}(X_i, Z_i)^2 X_i^2 + \varepsilon_i\end{aligned}$$

Here $\hat{P}(X,Z)$ is a preliminary estimator of the propensity score $P^*(X,Z)$. Then compute the second derivative with respect to $\hat{P}(X_i,Z_i)$, and evaluate it at both $(\tilde{u},x_1)$ and $(\tilde{u},x_2)$. This yields

$$\left.\frac{\partial^2\hat{\mathbb{E}}\left[Y|\hat{P}(X,Z)=\tilde{u},X=x_1\right]}{\partial u^2}\right|_{u=\tilde{u}}=2\hat{\alpha}_2\tilde{u}+2\hat{\alpha}_6\tilde{u}x_1+2\hat{\alpha}_8\tilde{u}x_1^2$$

and

$$\left.\frac{\partial^2\hat{\mathbb{E}}\left[Y|\hat{P}(X,Z)=\tilde{u},X=x_2\right]}{\partial u^2}\right|_{u=\tilde{u}}=2\hat{\alpha}_2\tilde{u}+2\hat{\alpha}_6\tilde{u}x_2+2\hat{\alpha}_8\tilde{u}x_2^2$$

which amounts to testing whether $\alpha_6=\alpha_8=0$.

5 Estimation and Inference

In this section we provide estimation results for the MTE, the LATE, and the re-scaling correction.

Assumption 4
The observed data $\{Y_i,X_i,Z_i,D_i^*\}_{i=1}^n$ *is i.i.d., where* $Y_i\in\mathbb{R}$, $X_i\in\mathbb{R}^d$, $Z_i\in\mathbb{R}$, *and* $D_i^*\in\{0,1\}$.

5.1 MTE and LATE

The parameter to estimate is given by (9). We fix two values z and z' in the support of Z, and x in the support of X.

First we start with the estimation of the propensity score. One option is via via maximum likelihood using a probit or logit model:

$$\hat{p}(x,z)=G(x'\hat{\alpha}_x+\hat{\alpha}_z z),$$

and

$$\hat{p}(x,z')=G(x'\hat{\alpha}_x+\hat{\alpha}_z z'),$$

where G is either the standard normal CDF or the logistic CDF. Another option is a non-parametric estimator.

The pseudo-MTE curve is given in (6). To estimate it we ran a non-parametric regression of Y_i on $\hat{p}(X_i,Z_i)$ and X_i, denoted by $\hat{\mathbb{E}}\left[Y|\hat{p}(X,Z)=u,X=x\right]$. The estimated pseudo-MTE curve is then

$$\widehat{\text{MTE}}^*(u,x)=\frac{\partial\hat{\mathbb{E}}\left[Y|\hat{p}(X,Z)=u,X=x\right]}{\partial u}.$$

Finally, the estimated misspecified LATE is then

$$\widehat{\text{LATE}}^*(x, \hat{p}(x,z), \hat{p}(x,z')) = \frac{1}{\hat{p}(x,z) - \hat{p}(x,z')} \sum_{j=1}^{J-1} \widehat{\text{MTE}}^*(u_j, x)(u_{j+1} - u_j)du$$

where $\{u_1, \ldots, u_J\}$ is a grid with $u_1 := \hat{p}(x,z')$ and $u_J := \hat{p}(x,z)$.

5.2 Re-scaling correction

In this section we focus on estimation and inference for $(\overline{p_x^*}, \underline{p_x^*})$ in order to construct the re-scaling correction $\overline{p_x^*} - \underline{p_x^*}$. We adapt and implement the methods in Chernozhukov, Lee, and Rosen (2013). Importantly, we extend the analysis to the case where the infimum and supremum bounding function discussed in Chernozhukov, Lee, and Rosen (2013) depend on a conditioning set. To this end, suppose the propensity score model $\hat{p}(x,z)$ is estimated through either a parametric model, or a nonparametric series or kernel regression. Then one can construct an estimator for $(\overline{p_x^*}, \underline{p_x^*})$ based, respectively, on extensions of Theorems 4, 5 and 6 in Chernozhukov, Lee, and Rosen (2013). Their procedure requires corrections of the nonparametric estimate of $\hat{P}(x,z)$ by an additional terms $k(\gamma_n^u, x) \cdot s(x,z)$ and $k(\gamma_n^l, x) \cdot s(x,z)$, composed by a point-wise measure of variability $s(x,z)$ and a z-uniform, x-point-wise critical value $k(\gamma_n, x)$ for an appropriate γ_n quantiles dependent on sample size. Thus, instead of the naive proposal where $\widehat{\overline{p_x^*}} = \sup_{z \in Z(x)} \hat{P}(x,z)$ and $\widehat{\underline{p_x^*}} = \inf_{z \in \hat{Z}_n(x)} \hat{P}(x,z)$ we consider $\widehat{\overline{p_x^*}} = \sup_{z \in \hat{Z}_n(x)} \hat{P}(x,z) + k(\gamma_n^u, x) \cdot s(x,z)$ and $\widehat{\underline{p_x^*}} = \inf_{z \in Z(x)} \hat{P}(x,z) + k(\gamma_n^l, x) \cdot s(x,z)$. The exact construction of the correction term depends on the method used to estimate the function $\hat{P}(\cdot)$ but we give an illustration here using a series estimator based on splines.

We compute the pair $(\overline{p_x^*}, \underline{p_x^*})$ using the a generalized version of the Chernozhukov, Lee, and Rosen (2013) algorithm, allowing for further dependence on covariates. Algorithm 1 is adapted for parametric and series estimators while Algorithm 2 is adapted for kernel based estimators.

Algorithm 1
Point-wise p-unbiased, estimators of $(\widehat{\overline{p_x^*}}, \widehat{\underline{p_x^*}})$ based on parametric or series first step estimators

Step 1 – Set $\gamma_n^u = 1 - 0.1/\log(n)$ and $\gamma_n^l = 1 - 0.1/\log(n)$ and K_n as the number of series terms. Simulate M draws, denoted $(Z_1, \cdots, Z_M)$ from $\mathcal{N}(0, I_{K_n})$.

Step 2 – Obtain an estimate for $\hat{\Omega}_n$, the variance-covariance matrix of $\sqrt{n}(\hat{\beta}_n - \beta_n)$, where $\hat{\beta}_n$ collects the series estimator coefficients.

Step 3 – For $(x,z) \in \mathcal{Z} \times \mathcal{X}$ we compute the standardized bases expansion of the propensity score: $\hat{g}(x,z) = B(x,z)'\hat{\Omega}_n^{-\frac{1}{2}}$, where $B(x,z) = \text{vec}(B(x) \otimes B(z))$ is a vector of basis functions (in our case, *splines). Compute* $s_n(x,z) = \frac{\|\hat{g}(x,z)\|}{\sqrt{n}}$

Step 4 – Obtain the quantiles $k^u_{\gamma_n}(x) := q_{\gamma^u_n}\left(\sup_{z\in Z(x)} \frac{\hat{g}(x,z)'Z_r}{\|\hat{g}(x,z)\|}\right)$ *and* $k^l_{\gamma_n}(x) := q_{\gamma^l_n}\left(\sup_{z\in Z(x)} \frac{\hat{g}(x,z)'Z_r}{\|\hat{g}(x,z)\|}\right)$ *for* $r = 1, \cdots, R$*. and obtain the preliminary optimizing collection of sets*

$$\hat{Z}^l_n(x) = \left\{(x,z) \in \mathcal{Z} \times \mathcal{X} : \hat{P}_n(x,z) \leq \min_{z\in Z}\left(\hat{P}_n(x,z) + k^l_{\gamma_n}(x)\cdot s_n(x,z)\right) + 2\cdot k^l_{\gamma_n}(x)\cdot s_n(x,z)\right\}$$
$$\hat{Z}^u_n(x) = \left\{(x,z) \in \mathcal{Z} \times \mathcal{X} : \hat{P}_n(x,z) \geq \max_{z\in Z}\left(\hat{P}_n(x,z) + k^u_{\gamma_n}(x)\cdot s_n(x,z)\right) + 2\cdot k^u_{\gamma_n}(x)\cdot s_n(x,z)\right\}$$

Step 5 – Finally compute:

$$k^u_{p,\hat{Z}_n}(x) = q_p\left(\sup_{z\in\hat{Z}^u_n(x)} \frac{\hat{g}(x,z)'Z_r}{\|\hat{g}(x,z)\|} \textit{ for } r = 1,\cdots R\right);$$
$$k^l_{p,\hat{Z}_n}(x) = q_p\left(\inf_{z\in\hat{Z}^l_n(x)} \frac{\hat{g}(x,z)'Z_r}{\|\hat{g}(x,z)\|} \textit{ for } r = 1,\cdots R\right)$$

And obtain estimators:

$$\widehat{\overline{p^*_x}} = \sup_{z\in Z}\left(\hat{P}_n(x,z) + k^u_{p,\hat{Z}_n}(x)\cdot\frac{\|\hat{g}(x,z)\|}{\sqrt{n}}\right)$$
$$\widehat{\underline{p^*_x}} = \inf_{z\in Z}\left(\hat{P}_n(x,z) + k^l_{p,\hat{Z}_n}(x)\cdot\frac{\|\hat{g}(x,z)\|}{\sqrt{n}}\right)$$

Algorithm 2

Point-wise Point-wise p-unbiased, estimators of $(\widehat{\overline{p^*_x}}, \widehat{\underline{p^*_x}})$ *based on kernels*

Step 1 – Set $\gamma^u_n = 1 - 0.1/\log(n)$ *and* $\gamma^l_n = 1 - 0.1/\log(n)$ *and H as the bandwidth matrix (for example* $H = h\cdot I_{(d_x+d_z)}$*). Simulate* $M\times n$ *draws, with* $i = 1,\cdots,n$ *and* $r = 1,\cdots,M$ *denoted* η_{ri} *from* $\mathcal{N}(0,1)$.

Step 2 – Denote $W_i = (X_i, Z_i)$*. For each* (x,z) *on an grid in* $\mathcal{Z}\times\mathcal{X}$ *and for each* $r = 1,\cdots,R$ *obtain:*

$$\hat{U}_i := D - \hat{P}(X_i, Z_i)$$
$$\hat{g}(x,z,\hat{U}_i,X_i,Z_i) := \frac{\hat{U}_i}{(h_n)^{d_x+d_z}\hat{f}_n(x,z)} K_H\left(\begin{pmatrix}x\\z\end{pmatrix} - \begin{pmatrix}X_i\\Z_i\end{pmatrix}\right)$$
$$s_n(x,z) := \frac{\mathbb{E}_n[\hat{g}(x,z,U_i,X_i,Z_i)^2]}{nh_n^{d_x+d_z}}$$
$$\mathbb{G}_n(\hat{g},r) := \frac{1}{\sqrt{n}}\sum_{i=1}^n \eta_{ir}\times\hat{g}(x,z,\hat{U}_i,X_i,Z_i)$$

Step 3 – Obtain the quantiles $k^u_{\gamma_n}(x) := q_{\gamma^u_n}\left(\inf_{z\in Z(x)} \frac{\mathbb{G}_n(\hat{g},r)}{\sqrt{\mathbb{E}_n[\hat{g}(x,z,\hat{U}_i,X_i,Z_i)]}}\right)$ *and*

$k^l_{\gamma_n}(x) := q_{\gamma^l_n}\left(\sup_{z\in Z(x)} \frac{\mathbb{G}_n(\hat{g},r)}{\sqrt{\mathbb{E}_n[\hat{g}(x,z,\hat{U}_i,X_i,Z_i)]}}\right)$ *for* $r = 1, \cdots, R$ *and obtain the preliminary optimizing collection of sets*

$$\hat{Z}^l_n(x) = \left\{(x,z) \in \mathcal{Z} \times \mathcal{X} : \hat{P}_n(x,z) \leq \min_{z\in Z}\left(\hat{P}_n(x,z) + k^l_{\gamma_n}(x)\cdot s_n(x,z)\right) + 2\cdot k^l_{\gamma_n}(x)\cdot s_n(x,z)\right\}$$
$$\hat{Z}^u_n(x) = \left\{(x,z) \in \mathcal{Z} \times \mathcal{X} : \hat{P}_n(x,z) \geq \max_{z\in Z}\left(\hat{P}_n(x,z) + k^u_{\gamma_n}(x)\cdot s_n(x,z)\right) + 2\cdot k^u_{\gamma_n}(x)\cdot s_n(x,z)\right\}$$

Step 4 – Finally compute:

$$k^u_{p,\hat{Z}_n}(x) = q_p\left(\sup_{z\in\hat{Z}^u_n(x)} \frac{\mathbb{G}_n(\hat{g},r)}{\sqrt{\mathbb{E}_n[\hat{g}(x,z,\hat{U}_i,X_i,Z_i)]}} \text{ for } r = 1,\cdots,R\right);$$
$$k^l_{p,\hat{Z}_n}(x) = q_p\left(\inf_{z\in\hat{Z}^l_n(x)} \frac{\mathbb{G}_n(\hat{g},r)}{\sqrt{\mathbb{E}_n[\hat{g}(x,z,\hat{U}_i,X_i,Z_i)]}} \text{ for } r = 1,\cdots,R\right)$$

for $r = 1, \cdots R$. *And obtain estimators:*

$$\widehat{\overline{p^*_x}}(p) = \sup_{z\in Z}\left(\hat{P}_n(x,z) + k^u_{p,\hat{Z}_n}(x)\cdot s_n(x,z)\right)$$
$$\widehat{\underline{p^*_x}}(p) = \inf_{z\in Z}\left(\hat{P}_n(x,z) + k^l_{p,\hat{Z}_n}(x)\cdot s_n(x,z)\right)$$

Finally, we give a formal result for consistency of our estimator $(\widehat{\overline{p^*_x}}, \widehat{\underline{p^*_x}})$ under both parametric or series and kernel based estimators. Like the algorithms above, these are point-wise conditional versions of Theorems 4, 5 and 6 in Chernozhukov, Lee, and Rosen (2013). While the point-wise consistency of these conditional results does not appear in the form we present here in Chernozhukov, Lee, and Rosen (2013), the derivation of these theorem, given their results, is straightforward.

Theorem 1 (Conditional Bounds when $\hat{P}(\cdot)$ is a parametric estimator)
Let p *be the quantile index, so that, for example,* $p = 0.5$ *results in median unbiased estimators, while* $p = 0.025$ *and* 0.975 *result in the upper and lower point-wise conditional bound. Then the estimators defined by Algorithm 1, for any fixed* $x \in \mathcal{X}_B$, *satisfies the following properties:*
i) $P_n[\underline{p^*_x} \leq \widehat{\underline{p^*_x}}(p)] \geq p - o(1)$
ii) $|\underline{p^*_x} - \widehat{\underline{p^*_x}}(p)| = O_{P_n}\left(\frac{1}{\sqrt{n}}\right)$.
iii) $P_n(\underline{p^*_x} + \mu_n\frac{1}{\sqrt{n}}) \to 1$ *for any* $\mu_n \xrightarrow{P_n} \infty$.
iv) In addition, if condition V in Chernozhukov, Lee, and Rosen (2013) holds then $P_n[\underline{p^*_x} \leq \widehat{\underline{p^*_x}}(p)] = p - o(1)$

Theorem 2 (Conditional Bounds when $\hat{P}(\cdot)$ is a series based nonparametric estimator)
Let p *be the quantile index, so that, for example,* $p = 0.5$ *results in median unbiased estimators, while* $p = 0.025$ *and* 0.975 *result in the upper and lower point-wise conditional bound. Then the estimators*

defined by Algorithm 1, for any fixed $x \in \mathcal{X}_B$*, satisfies the following properties:*

i) $P_n[\underline{p_x^*} \leq \widehat{\underline{p_x^*}}(p)] \geq p - o(1)$

ii) $|\underline{p_x^*} - \widehat{\underline{p_x^*}}(p)| = O_{P_n}\left(\sqrt{\log n}\frac{\zeta_n}{\sqrt{n}}\right)$.

iii) $P_n(\underline{p_x^*} + \mu_n\sqrt{\log n}\frac{\zeta_n}{\sqrt{n}}) \to 1$ *for any* $\mu_n \xrightarrow{P_n} \infty$.

Moreover, if condition V in Chernozhukov, Lee, and Rosen (2013) holds then:

vi)$P_n[\underline{p_x^*} \leq \widehat{\underline{p_x^*}}(p)] = p - o(1)$

v) $|\underline{p_x^*} - \widehat{\underline{p_x^*}}(p)| = O_{P_n}\left(\frac{\zeta_n}{\sqrt{n}}\right)$.

vi) $P_n(\underline{p_x^*} + \mu_n\frac{\zeta_n}{\sqrt{n}}) \to 1$ *for any* $\mu_n \xrightarrow{P_n} \infty$.

Theorem 3 (Conditional Bounds when $\hat{P}(\cdot)$ is a kernel based nonparametric estimator)
Let p *be the quantile index, so that, for example,* $p = 0.5$ *results in median unbiased estimators, while* $p = 0.025$ *and* 0.975 *result in the upper and lower point-wise conditional bound. Then the estimators defined by Algorithm 2, for any fixed* $x \in \mathcal{X}_B$*, satisfies the following properties:*

i) $P_n[\underline{p_x^*} \leq \widehat{\underline{p_x^*}}(p)] \geq p - o(1)$

ii) $|\underline{p_x^*} - \widehat{\underline{p_x^*}}(p)| = O_{P_n}\left(\sqrt{\log n}\frac{1}{\sqrt{nh_n^{d_x+d_z}}}\right)$.

iii) $P_n(\underline{p_x^*} + \mu_n\sqrt{\log n}\frac{1}{\sqrt{nh_n^{d_x+d_z}}}) \to 1$ *for any* $\mu_n \xrightarrow{P_n} \infty$.

Moreover, if condition V in Chernozhukov, Lee, and Rosen (2013) holds then:

vi)$P_n[\underline{p_x^*} \leq \widehat{\underline{p_x^*}}(p)] = p - o(1)$

v) $|\underline{p_x^*} - \widehat{\underline{p_x^*}}(p)| = O_{P_n}\left(\frac{1}{\sqrt{nh_n^{d_x+d_z}}}\right)$.

vi) $P_n(\underline{p_x^*} + \mu_n\frac{1}{\sqrt{nh_n^{d_x+d_z}}}) \to 1$ *for any* $\mu_n \xrightarrow{P_n} \infty$.

6 A synthetic example

We look at one example for a DGP which helps higlight the procedure. Consider the model from the previous section

$$\begin{aligned}
Y(0) &= \alpha + \beta_0 X + U_0 \\
Y(1) &= \alpha + \varphi + \beta_1 X + U_1 \\
D &= \mathbb{1}\{\gamma Z \geq V\} \\
\tilde{D} &= \mathbb{1}\{0 \geq \tilde{V}\} \\
D^* &= R \cdot D + (1 - R) \cdot \tilde{D}
\end{aligned}$$

with the following parametrization

$$\gamma = 2,\ \beta_0 = 2,\ \beta_1 = 3,\ \phi = 2.5,\ \rho_0 = -0.5,\ \rho_1 = 0.5$$

for the selection and potential outcome equations in the previous section, and the non-responders probability specified below. Observables and unobservables are distributed according to:

$$\begin{aligned}
X &\sim \text{Uniform}(0,1) \\
Z &\sim \text{Uniform}(-1,1) \\
X &\perp Z \\
\begin{pmatrix} U_0 \\ U_1 \\ V \end{pmatrix} &\sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & \rho_0 \\ 0 & 1 & \rho_1 \\ \rho_0 & \rho_1 & 1 \end{pmatrix} \right).
\end{aligned}$$

Here, ρ_0 and ρ_1 control the degree of endogeneity. Recall that R is the latent status of an individual: $R = 1$ for a responder and $R = 0$ for a non-responder. We simulate the proportion of non-responders as

$$\delta_x := \Pr(R = 0 | X = x) = \max\{0, x - 0.5\}$$

so that below $x = 0.5$ there are no non-responders. Past the $x = 0.5$ point, the probability that someone is a non-responder increases linearly with x. At $x = 1$ half are non-responders. In this scenario $\tilde{P}(X) := \Pr(\tilde{D} = 1 | R = 0, X) = \Phi(0) = 0.5$ so the conditional and unconditional propensity score coincide. On the other hand, the true propensity score for responders, $P(X, Z) := \Pr(D = 1 | R = 1, X, Z)$, is supported between $\Phi(-2)$ and $\Phi(2)$ and it is fixed as a function of x. This means that $\underline{p_x}, \overline{p_x} = [\Phi(-2), \Phi(2)]$. As prescribed by Proposition 1, the bounds on the observed propensity score $[\underline{p_x}^*, \overline{p_x}^*]$, as a function of x are captured by:

$$\begin{aligned}
\underline{p_x}^* &= \max(0, 0.5 - x) \cdot \Phi(-2) + 0.5 \cdot (1 - \max(0, 0.5 - x)) \\
\overline{p_x}^* &= \max(0, 0.5 - x) \cdot \Phi(2) + 0.5 \cdot (1 - \max(0, 0.5 - x))
\end{aligned}$$

Below, we report the implied estimates for the observed upper and lower bounds of the propensity score, based on a series regression (as in Algorithm 1) and based on a kernel regression (as in Algorithm 2).

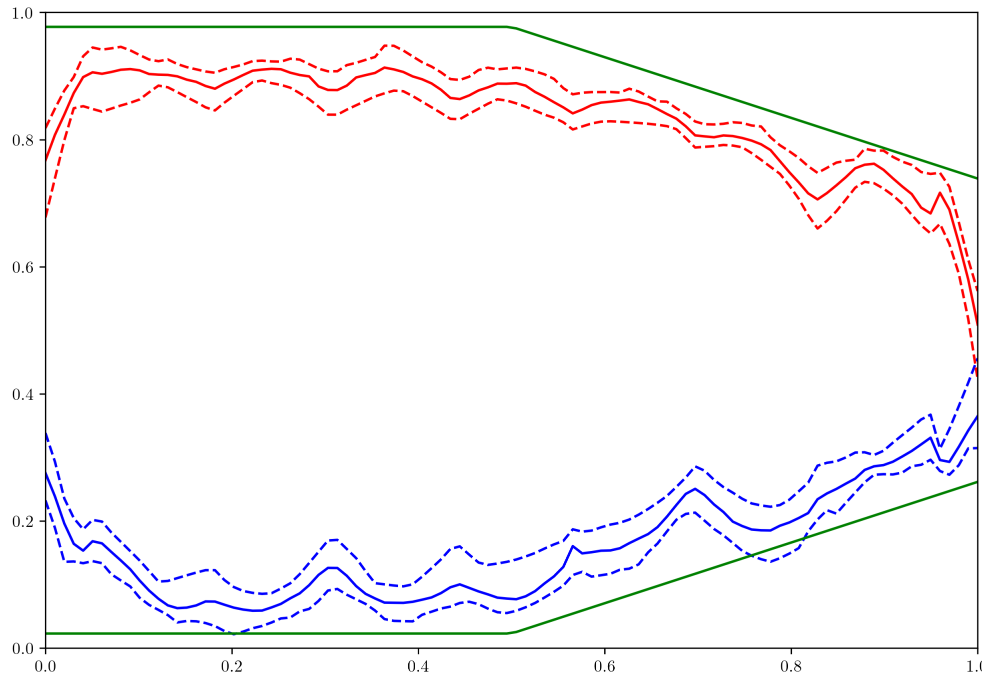

Figure 1: Bounds estimated through a series estimator based on Algorithm 1. Solid line is the median unbiased estimator for upper (red) and lower (blue) bounds respectively. Dashed lines are the 95% pointwise confidence intervals. Sample size $n = 1000$, replications $M = 1000$.

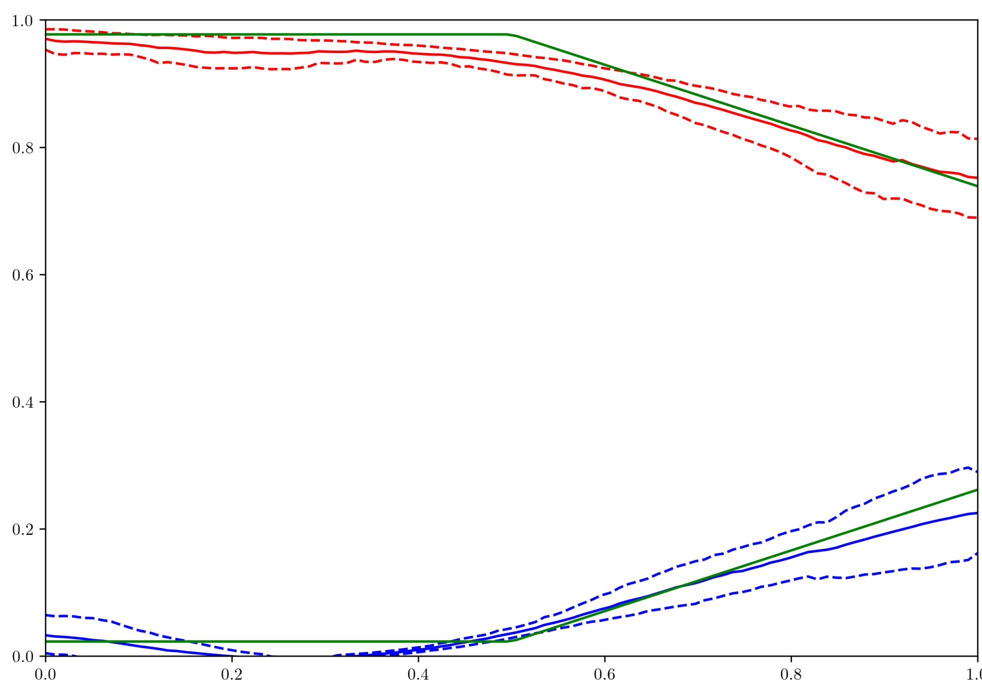

Figure 2: Bounds estimated through a kernel estimator based on Algorithm 2. Solid line is the median unbiased estimator for upper (red) and lower (blue) bounds respectively. Dashed lines are the 95% pointwise confidence intervals. Sample size $n = 1000$, replications $M = 1000$.

Now observe that if we estimate the bounds of the propensity score by simply taking the empirical minimum and maximum, thus disregarding the correction proposed by Chernozhukov, Lee, and Rosen (2013), the upper and lower bounds of the propensity score will not be result in the correct inference. The interpretation is discussed in detail in Chernozhukov, Lee, and Rosen (2013). For this specific DGP, the bias may not be too dramatic, as the curvature of the green line is very limited. We illustrate this in the figure below:

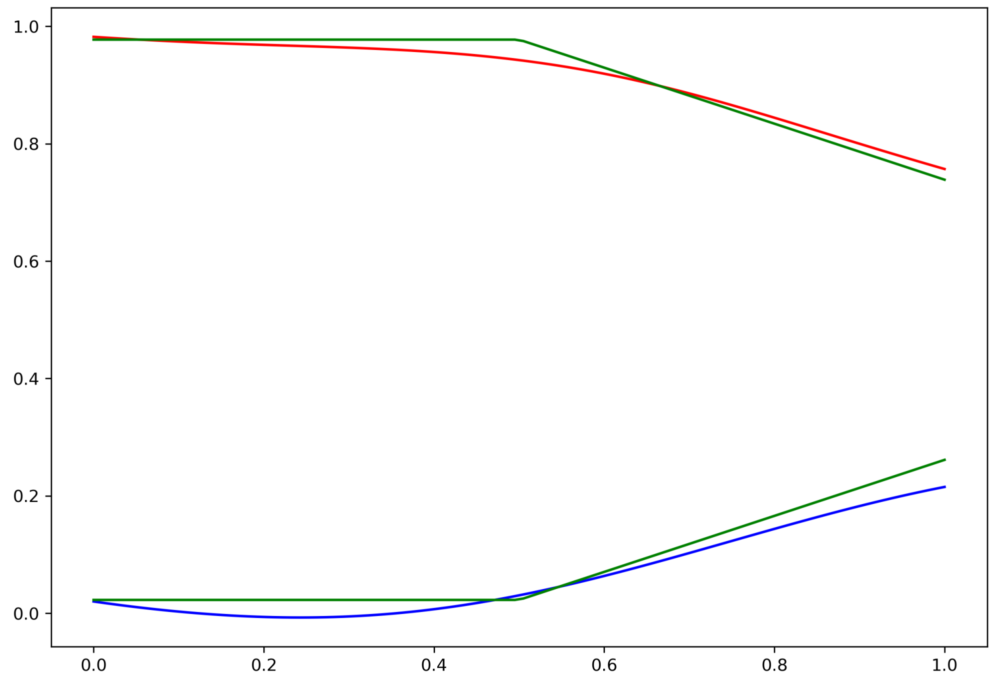

Figure 3: Naive estimator that takes the sup and inf of the kernel estimator.

To conduct inference on the re-scaling correction, that is $\overline{p_x^*} - \underline{p_x^*}$, one could construct the intersection of the upper and lower p-quantile bounds for $\tilde{p} = (1+p)/2$ based on the Bonferroni correction for multiple hypothesis. This inference procedure, albeit conservative, bypasses the need to construct a covariance function between $\overline{p_x^*}$ and $\underline{p_x^*}$ which may be very complicated to derive analytically.[8] The figure below presents an estimate of the range, together with the confidence interval obtained through the Bonferroni correction, in the case of the kernel based propensity score estimator.

[8]This is the approach suggested by Chernozhukov, Lee, and Rosen (2013)

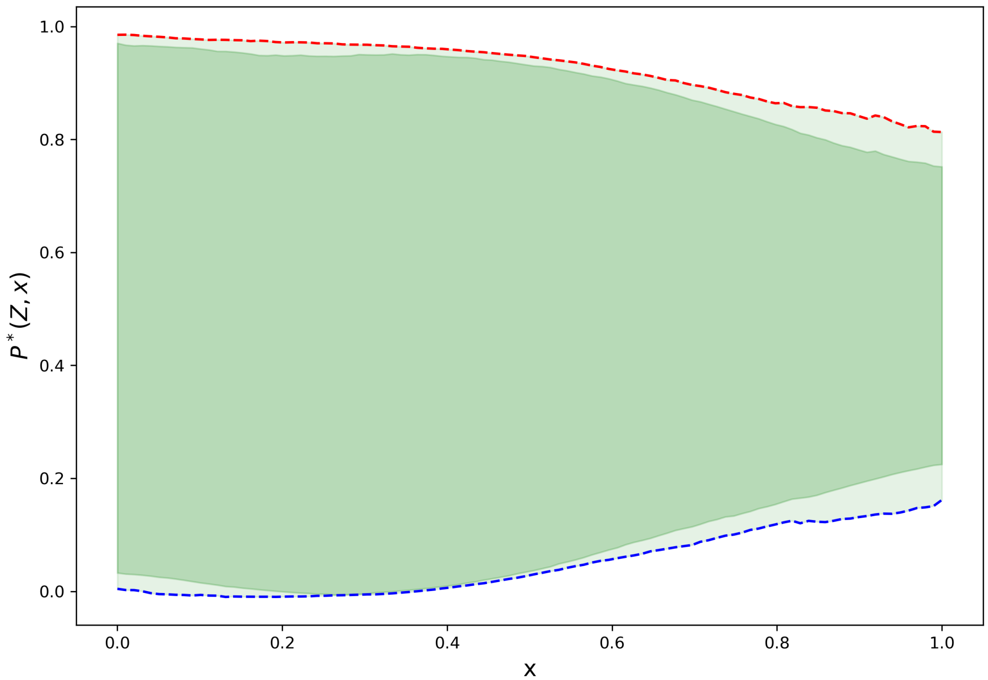

Figure 4: 90% confidence interval for the range, obtained through the 95% confidence interval for the upper and lower bound. The red and blue dashed lines are the one sided 95% confidence intervals.

We may estimate the pseudo-MTE obtained by just ignoring the issue with the estimated propensity score. That is, we proceed by estimating the observed propensity score and then estimate the conditional mean of Y given $X = x$ and the propensity score value $P(Z, x)$, and take its derivative. That is the estimated MTE curve.

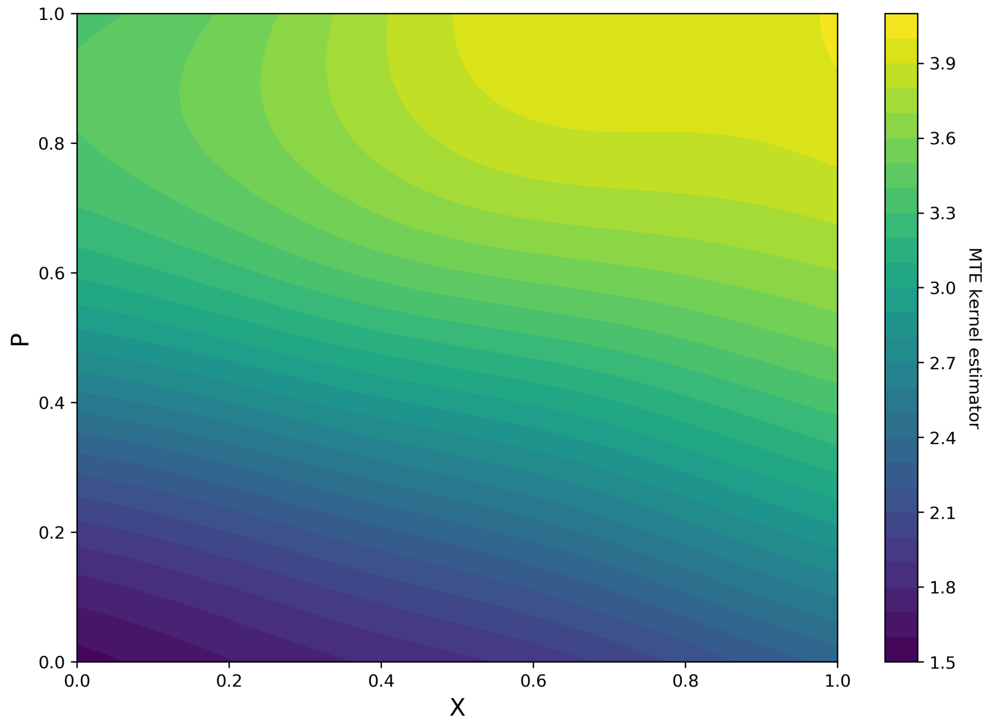

Figure 5: MTE kernel estimate with Kernel estimated propensity score.

7 Conclusion

This paper analyzes and MTE model where a fraction of the population is unresponsive to the instrumental variable. We show that the MTE curve, and parameters associated with it can be point identified. In a linear model, we present a testable implication for the presence of non-responders. We also present estimation results.

References

Acerenza, S. (2023): "Partial Identification of Marginal Treatment Effects with Discrete Instruments and Misreported Treatment," Working Paper.

Acerenza, S., K. Ban, and D. Kedagni (2021): "Marginal Treatment Effects with Misclassified Treatment," Working Paper.

Bjorklund, A., and R. Moffitt (1987): "The Estimation of Wage Gains and Welfare Gains in Self-Selection," *The Review of Economics and Statistics*, 69(1), 42–49.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013): "Intersection bounds: Estimation and inference," *Econometrica*, 81(2), 667–737.

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

Heckman, J. J., and E. Vytlacil (2001): "Local Instrumental Variables," in *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, ed. by C. Hsiao, K. Morimune, and J. Powell, pp. 1–46. Cambrigde University Press.

——— (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 73(3), 669–738.

Martinez-Iriarte, J., and Y. Sun (2022): "Identification and Estimation of Unconditional Policy Effects of an Endogenous Binary Treatment," Working Paper.

Mogstad, M., and A. Torgovitsky (2018): "Identification and Extrapolation of Causal Effects with Instrumental Variables," *Annual Review of Economics*, 10, 577–613.

Possebom, V. (2021): "Crime and Mismeasured Punishment: Marginal Treatment Effect with Misclassification," Working Paper.

Appendices

A Proofs

Proof. Proof of Lemma 1

Starting with the model in (1) we can write

$$\begin{aligned}\Pr(D^* = 1|X,Z) = \Pr(R = 1|X,Z) \cdot \Pr(D = 1|R = 1, X, Z) \\ + \Pr(R = 0|X,Z) \cdot \Pr(\tilde{D} = 1|R = 0, X, Z).\end{aligned}$$

Assumption 1 simplifies the mixing probabilities to $\Pr(R = 1|X,Z) = \Pr(R = 1|X) = 1 - \delta_X$ and $\Pr(R = 0|X,Z) = \Pr(R = 0|X) = \delta_X$. We obtain

$$\Pr(D^* = 1|X,Z) = (1 - \delta_X) \cdot \Pr(D = 1|R = 1, X, Z) + \delta_X \cdot \Pr(\tilde{D} = 1|R = 0, X, Z).$$

To see that $Pr(\tilde{D} = 1|R = 0, X, Z) = Pr(\tilde{D} = 1|R = 0, X)$, we note that By Assumptions 1 and 2.2, $Z \perp X, \tilde{V} \| X$ so that:

$$\begin{aligned}\Pr(\tilde{D} = 1|R = 0, X, Z) &= \Pr(\tilde{\mu}(X) \geq \tilde{V}|R = 0, X, Z) \\ &= \Pr(\tilde{\mu}(X) \geq \tilde{V}|R = 0, X) \\ &= \Pr(\tilde{D} = 1|R = 0, X).\end{aligned}$$

Therefore

$$\begin{aligned}\Pr(D^* = 1|X,Z) &= (1-\delta_X)\cdot \Pr(D=1|R=1,X,Z) + \delta_X \cdot \Pr(\tilde{D}=1|R=0,X)\\ &= (1-\delta_X)\cdot P(X,Z) + \delta_X \cdot \tilde{P}(X).\end{aligned}$$

□

Proof. Proof of Lemma 2

Using (2), for $u \in \mathcal{P}_x^*$, we can write

$$\begin{aligned}\mathbb{E}\left[Y|P^*(X,Z)=u, X=x\right] &= \mathbb{E}\left[Y|(1-\delta_x)\cdot P(X,Z) + \delta_x \cdot \tilde{P}(X) = u, X=x\right]\\ &= \mathbb{E}\left[Y\middle|P(X,Z) = \frac{u-\delta_x\tilde{P}(x)}{1-\delta_x}, X=x\right]\end{aligned}$$

Differentiating with respect to u, we obtain

$$\text{MTE}^*(u,x;\delta_x) = \frac{1}{1-\delta_x}\text{MTE}\left(\frac{u-\delta_x\tilde{P}(x)}{1-\delta_x}, x\right) \text{ for } u \in \mathcal{P}_x^*.$$

since $\frac{u-\delta_x\tilde{P}(x)}{1-\delta} \in \mathcal{P}_x$ by (2). Alternatively, $u \in \mathcal{P}_x^*$ if and only if $v \in \mathcal{P}_x$, so that

$$\text{MTE}(v,x) = (1-\delta_x)\text{MTE}^*\left((1-\delta_x)v + \delta_x\tilde{P}(x), x; \delta_x\right) \text{ for } v \in \mathcal{P}_x.$$

□

Proof. Proof of Corollary 1 Consider the LATE, for $P(x,z') < P(x,z)$ with $z, z' \in \mathcal{Z}$, which can be obtained from MTE curve as

$$\text{LATE}(x, P(x,z), P(x,z')) = \frac{1}{P(x,z)-P(x,z')}\int_{P(x,z')}^{P(x,z)} \text{MTE}(u,x)du.$$

Under misspecification, for the same $z, z' \in \mathcal{Z}$, we have

$$\begin{aligned}\text{LATE}^*(x, P^*(x,z), P^*(x,z')) &= \frac{1}{P^*(x,z)-P^*(x,z')}\int_{P^*(x,z')}^{P^*(x,z)} \text{MTE}^*(u,x;\delta_x)du\\ &= \frac{(1-\delta_x)^{-1}}{P(x,z)-P(x,z')}\int_{(1-\delta_x)P(x,z')+\delta_x\tilde{P}(x)}^{(1-\delta_x)P(x,z)+\delta_x\tilde{P}(x)} \frac{1}{1-\delta_x}\\ &\times \text{MTE}\left(\frac{u-\delta_x\tilde{P}(x)}{1-\delta_x}, x\right) du.\end{aligned}$$

Note that to go from MTE* to MTE we used Lemma 2. We did not use Corollary 3. Defining the change of variables $\tilde{u} = \frac{u-\delta_x \tilde{P}(x)}{1-\delta_x}$, we get $(1-\delta_x)d\tilde{u} = du$. We then write

$$\begin{aligned}
\text{LATE}^*(x, P^*(x,z), P^*(x,z')) &= \frac{(1-\delta_x)^{-1}}{P(x,z)-P(x,z')} \int_{(1-\delta_x)P(x,z')+\delta_x\tilde{P}(x)}^{(1-\delta_x)P(x,z)+\delta_x\tilde{P}(x)} \frac{1}{1-\delta_x} \\
&\times \text{MTE}\left(\frac{u-\delta_x\tilde{P}(x)}{1-\delta_x}, x\right) du \\
&= \frac{(1-\delta_x)^{-1}}{P(x,z)-P(x,z')} \int_{P(x,z')}^{P(x,z)} \text{MTE}(u,x)du \\
&= \frac{1}{1-\delta_x} \text{LATE}(x, P(x,z), P(x,z')).
\end{aligned}$$

The marginal policy relevant treatment effect (MPRTE) is given by

$$\text{MPRTE}(x) = \int_{\mathcal{Z}} \text{MTE}(P(x,z),x) \frac{\partial P(x,z)}{\partial z} \left(E\left[\frac{\partial [P(x,Z)]}{\partial z}\right]\right)^{-1} f_{Z|X}(z|x)dz$$

Then, using Equations (4) and (7) we get

$$\begin{aligned}
\text{MPRTE}^*(x) &= \int_{\mathcal{Z}} \text{MTE}^*(P^*(x,z), x; \delta_x) \frac{\partial P^*(x,z)}{\partial z} \left(E\left[\frac{\partial [P^*(x,Z)]}{\partial z}\right]\right)^{-1} f_{Z|X}(z|x)dz \\
&= \int_{\mathcal{Z}} \frac{1}{1-\delta_x} \text{MTE}(P(x,z),x) \frac{\partial P(x,z)}{\partial z} \left(E\left[\frac{\partial [P(X,Z)]}{\partial z}\right]\right)^{-1} f_{Z|X}(z|x)dz \\
&= \frac{1}{1-\delta_x} \text{MPRTE}(x).
\end{aligned}$$

□

B Validity of the LIV approach for responders.

The model is the same as described in section 2.1. We want to show that the MTE curve for responders,

$$\text{MTE}(u,x) := \mathbb{E}\left[Y(1) - Y(0) | R = 1, U_D = u, X = x\right]$$

can be identified by

$$\text{MTE}(u,x) = \frac{\partial \mathbb{E}\left[Y | R = 1, P(X,Z) = u, X = x\right]}{\partial u} \text{ for } u \in \mathcal{P}_x.$$

The selection equation for responders can be written as

$$D = \mathbb{1}\{P(X,Z) \geq U_D\},$$

where $P(X,Z) := \Pr(D = 1 | R = 1, X, Z)$ is the propensity score for responders and $U_D := F_{V|R,X,Z}(V|1,X,Z) = F_{V|R,X}(V|1,X)$ because by assumptions 1 and 2.(2) Z is independent of V

and R conditional on X. Moreover, conditional on $R = 1$, U_D is independent of X and uniformly distributed on $(0, 1)$.[9]

To show the validity of the LIV representation result in this setting, consider $\mathbb{E}[Y|R = 1, P(X, Z) = u, X = x]$. We start with $Y = DY(1) + (1 - D)Y(0)$. Thus, consider first

$$\begin{aligned}\mathbb{E}[DY(1)|R = 1, P(X, Z) = u, X = x] = \mathbb{E}[Y(1)|R = 1, P(X, Z) = u, X = x, D = 1] \\ \times \Pr(D = 1|R = 1, P(X, Z) = u, X = x)\end{aligned}$$

This last probability is

$$\begin{aligned}\Pr(D = 1|R = 1, P(X, Z) = u, X = x) &= \Pr(P(X, Z) \geq U_D|R = 1, P(X, Z) = u, X = x) \\ &= \Pr(u \geq U_D|R = 1, P(X, Z) = u, X = x) \\ &= \Pr(u \geq U_D|R = 1) \\ &= u.\end{aligned}$$

And the conditional expectation of $Y(1)$ is

$$\begin{aligned}\mathbb{E}[Y(1)|R = 1, P(X, Z) = u, X = x, D = 1] &= \mathbb{E}[Y(1)|R = 1, P(X, Z) = u, X = x, P(X, Z) \geq U_D] \\ &= \mathbb{E}[Y(1)|R = 1, X = x, u \geq U_D] \\ &= \int_{-\infty}^{\infty} y f_{Y(1)|R=1,X=x,U_D \leq u}(y) dy\end{aligned}$$

This conditional density is derivative of conditional cdf $F_{Y(1)|R=1,X=x,U_D \leq u}(y)$, which can be expressed as

$$\begin{aligned}F_{Y(1)|R=1,X=x,U_D \leq u}(y) &= \Pr(Y(1) \leq y|R = 1, X = x, U_D \leq u) \\ &= \frac{\Pr(Y(1) \leq y, U_D \leq u|R = 1, X = x)}{\Pr(U_D \leq u|R = 1, X = x)} \\ &= \frac{1}{u} \int_{-\infty}^{y} \int_{0}^{u} f_{Y(1),U_D|R=1,X=x}(\tilde{y}, \tilde{u}) d\tilde{y} d\tilde{u} \\ &= \frac{1}{u} \int_{-\infty}^{y} \int_{0}^{u} f_{Y(1)|U_D=\tilde{u},R=1,X=x}(\tilde{y}) d\tilde{y} d\tilde{u}\end{aligned}$$

[9] To see this, note that $\Pr(U_D \leq u|R = 1, X = x) = \Pr(F_{V|R,X}(V|1, x) \leq u|R = 1, X = x) = \Pr(V \leq F^{-1}_{V|R,X}(u|1, x)|R = 1, X = x) = F_{V|R,X}(F^{-1}_{V|R,X}(u|1, x)|1, x) = u$ for any value of x. Integrating against $F_{X|R}(x|1)$ shows that, conditional on $R = 1$, U_D is uniformly distributed on $(0, 1)$.

This means that

$$f_{Y(1)|R=1,X=x,U_D\leq u}(y) = \frac{1}{u}\int_0^u f_{Y(1)|U_D=\tilde{u},R=1,X=x}(y)d\tilde{u}$$

Therefore, we can write

$$\begin{aligned}\mathbb{E}[DY(1)|R=1,P(X,Z)=u,X=x] &= \mathbb{E}[Y(1)|R=1,P(X,Z)=u,X=x,D=1]\\ &\quad\times \Pr(D=1|R=1,P(X,Z)=u,X=x)\\ &= \frac{1}{u}\int_{-\infty}^{\infty} y\int_0^u f_{Y(1)|U_D=\tilde{u},R=1,X=x}(y)d\tilde{u}dy \times u\end{aligned}$$

So that

$$\begin{aligned}\frac{\partial\mathbb{E}[DY(1)|R=1,P(X,Z)=u,X=x]}{\partial u} &= \int_{-\infty}^{\infty} y f_{Y(1)|U_D=u,R=1,X=x}(y)dy\\ &= \mathbb{E}[Y(1)|R=1,U_D=u,X=x].\end{aligned}$$

Similar calculations for $\mathbb{E}[(1-D)Y(0)|R=1,P(X,Z)=u,X=x]$ yield

$$\frac{\partial\mathbb{E}[(1-D)Y(0)|R=1,P(X,Z)=u,X=x]}{\partial u} = \mathbb{E}[Y(0)|R=1,U_D=u,X=x].$$

Thus,

$$\text{MTE}(u,x) = \frac{\partial\mathbb{E}\left[Y|R=1,P(X,Z)=u,X=x\right]}{\partial u} \text{ for } u\in\mathcal{P}_x.$$

C Identification under known support of the true propensity score

In this section we investigate some consequences of *a priori* knowing the support of $\mathcal{P}_x$.

The next result shows that there is some form of automatic debiasing when integrating the MTE^* curve. In the special case of full support of the true propensity score, there is no bias in the conditional average treatment effect. First we make an assumption regarding $\mathcal{P}_x$.

Theorem 4
Let Assumptions 1, 2, 3. Then

$$\int_{\inf\mathcal{P}_x^*}^{\sup\mathcal{P}_x^*} MTE^*(u,x;\delta_x)du = \int_{\inf\mathcal{P}_x}^{\sup\mathcal{P}_x} MTE(v,x)dv.$$

Proof. Denote $\mathcal{P}_x = [\underline{p_x},\overline{p_x}]$, then $\mathcal{P}_x^* := [\underline{p_x^*},\overline{p_x^*}]$ where $\underline{p_x^*} := \inf\mathcal{P}_x^* = (1-\delta_x)\underline{p_x} + \delta_x\tilde{P}(x)$ and $\overline{p_x^*} := \sup\mathcal{P}_x^* = (1-\delta_x)\overline{p_x} + \delta_x\tilde{P}(x)$. Consider the integrating the pseudo-MTE curve over the support of the observed propensity score :

$$\int_{\underline{p_x^*}}^{\overline{p_x^*}} \text{MTE}^*(u, x; \delta_x)du = \int_{(1-\delta_x)\underline{p_x}+\delta_x \tilde{P}(x)}^{(1-\delta_x)\overline{p_x}+\delta_x \tilde{P}(x)} \text{MTE}^*(u, x; \delta_x)du.$$

$$= \int_{(1-\delta_x)\underline{p_x}+\delta_x \tilde{P}(x)}^{(1-\delta_x)\overline{p_x}+\delta_x \tilde{P}(x)} \frac{1}{1-\delta_x}\text{MTE}\left(\frac{u-\delta_x \tilde{P}(x)}{1-\delta_x}, x\right) du$$

$$= \int_{\underline{p_x}}^{\overline{p_x}} \text{MTE}(v, x)dv$$

where we have done the change of variables $v = \frac{u-\delta_x \tilde{P}(x)}{1-\delta_x}$. □

In general, the subpopulation whose propensity score is supported on $\mathcal{P}_x$ is determined by the instrument. Under the stronger condition that the propensity score has full support, Theorem 4 identifies the CATE(x). In this case, Assumption 3 says that the incentive in the instrument Z is strong enough to induce any individual in the $X = x$ subpopulation into or out of treatment.

Corollary 2

Let Assumptions 1, 2 hold. If Assumption 3 holds for $x \in \mathcal{X}_B$ with $\mathcal{P}_x = [0, 1]$, then $CATE(x)$ is identified for any such $\mathcal{X}_B$.

Remark 3

The result of Theorem 4 (and Corollary 2) states that by integrating the observed (and biased) marginal treatment effect curve over the support of the observed (and biased) propensity score leads to the correct averaging for $CATE(x)$ as a special case, provided that the propensity score for responders is supported on an interval. Thus, under the type of misspecification described in Equation (1), $CATE(x)$ is robust to $\delta_x \neq 0$.

Remark 4

This result also hold in a setting of misclassification and was our original motivation. That is, in a setting where instead of $Y = D^*Y(1) + (1 - D^*)Y(0)$, we have $Y = DY(1) + (1 - D)Y(0)$ and we interpret D^* as a misclassified treatment status.

Unfortunately, the automatic “de-biasing" in Theorem 4 does not hold for the other policy parameters that can be obtained via the MTE curve. On the other hand, we show that the if the bounds on the support are known, they can be used to identify δ_x which allows an explicit “de-biasing" procedure. Proposition 1 says that we can actually identify both δ_x and $\tilde{P}(x)$. It then follows from Lemma 2 that we can recover the MTE(u, x) curve, which was identified up to these two parameters.

Proposition 1

Let Assumptions 1, 2 hold. Let assumption 3 hold for $x \in \mathcal{X}_B$ with $\mathcal{P}_x = [\underline{p_x}, \overline{p_x}]$. If $\overline{p_x} - \underline{p_x}$ is known,

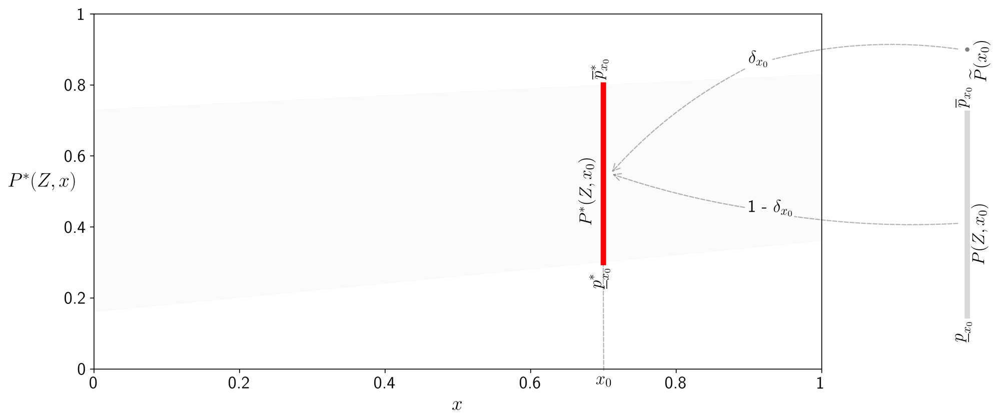

Figure 6: Identifying δ_x: The figure shows the link between the non-responders propensity score, the proportion of non-responders and the observed propensity score. Because the non-responders propensity score does not vary with the instrument Z and $supp(P(x,Z)) = [\underline{p_x}, \overline{p_x}]$ the δ_x can be recovered from observing the discrepancy from the observed support $[\underline{p_x}^*, \overline{p_x}^*]$ and $[\underline{p_x}, \overline{p_x}]$, like described in Proposition 1. The picture shows one of those points, x_0.

then, for any $x \in \mathcal{X}_B$, δ_x, $\tilde{P}(x)$, and $P(x,Z)$ are identified by

$$\delta_x = 1 - \frac{(\overline{p_x^*} - \underline{p_x^*})}{(\overline{p_x} - \underline{p_x})},$$

$$\tilde{P}(x) = \frac{\underline{p_x^*} - \frac{(\overline{p_x^*} - \underline{p_x^*})}{(\overline{p_x} - \underline{p_x})}}{1 - \frac{(\overline{p_x^*} - \underline{p_x^*})}{(\overline{p_x} - \underline{p_x})}},$$

$$P(x,Z) = \frac{P^*(x,Z) - \delta_x \cdot \tilde{P}(x)}{1 - \delta_x}.$$

The intuition for this result is simple. Because the original propensity score $P(x,Z)$, for any fixed x, is supported on the unit interval, the observed support $\mathcal{P}_x^* = [\underline{p_x^*}, \overline{p_x^*}]$ will contain enough information to identify δ_x. This is summarized Figure 6. Having identified δ_x, then we use equation (8) to identify the MTE curve. Note that only the length $\overline{p_x} - \underline{p_x}$ is required to be known.

Corollary 3

Let Assumptions 1, 2 hold. Let assumption 3 hold for $x \in \mathcal{X}_B$ with $\mathcal{P}_x = [\underline{p_x}, \overline{p_x}]$. If $\underline{p_x} - \overline{p_x}$ is known, then the MTE curve is identified by:

$$MTE(v,x) = \frac{(\overline{p_x^*} - \underline{p_x^*})}{(\overline{p_x} - \underline{p_x})} \cdot MTE^*\left(\frac{(\overline{p_x^*} - \underline{p_x^*})}{(\overline{p_x} - \underline{p_x})} v + \left(\underline{p_x^*} - \frac{(\overline{p_x^*} - \underline{p_x^*})}{(\overline{p_x} - \underline{p_x})}\right), x; 1 - \frac{(\overline{p_x^*} - \underline{p_x^*})}{(\overline{p_x} - \underline{p_x})}\right)$$

for $v \in \mathcal{P}_x$, where $\underline{p_x^*} = \inf \mathcal{P}_x^*$ and $\overline{p_x^*} = \sup \mathcal{P}_x^*$.

Proof. Note that the MTE^* function is identified. The result then follows immediately by substituting, in Equation (7), the expressions that identify δ_x and $\tilde{P}(x)$ in Proposition 1 . □

This corollary provides the correct "de-biasing" to be performed on the observed MTE curve to match the true MTE curve. However, it is possible to recover parameters that are based on the MTE curve *without* having to recover the MTE curve in the first place. We provide two examples based on Corollary 1 and Proposition 1. We have

$$(\overline{p_x^*} - \underline{p_x^*})\text{LATE}^*(x, P^*(x,z), P^*(x,z')) = \text{LATE}(x, P(x,z), P(x,z')),$$

and

$$(\overline{p_x^*} - \underline{p_x^*})\text{MPRTE}^*(x) = \text{MPRTE}(x).$$

In the previous examples, proceeding as if there were no misspecification, yields biased parameters. Thus, the automatic "de-biasing" in CATE is the exception rather than the rule.