Sensitivity of Bounds on ATEs under Survey Non-response

Roman Nedela* Lukáš Lafférs†

February 16, 2019

Abstract

This paper reformulates the problem of bounding average treatment effects under survey non-response studied in Behaghel et al. (2015) as an optimization problem. We provide a proof that this formulation is equivalent to the original problem and further extend it into a sensitivity analysis of the identifying assumptions. We control the departure from the assumption of treatment exogeneity via an interpretable parameter and thus allow to quantify the importance of the crucial identification assumption.

JEL: C4, C6.

Keywords: Bounds; Survey non-response; Average treatment effects; Sensitivity analysis;

*Matej Bel University, Dept. of Mathematics. E-mail: `roman.nedela3@umb.sk`

†Matej Bel University, Dept. of Mathematics. E-mail: `lukas.laffers@gmail.com`, Web: `http://www.lukaslaffers.com`

1 Introduction

In 2007-08, the French unemployment benefit provider (Unédic) mandated private companies to provide intensive counselling to job seekers. Behaghel et al. (2015) looked at the problem of evaluating this job training program using data from questionaries with a notable survey nonresponse.

The program was implemented as a randomized control trial with job training as a treatment variable. About 10 months after the end of the program, individuals (job training participants and a control group) were repeatedly attempted to reach before obtaining a response (employment status) in a phone survey. The number of maximum calls to reach individuals was fixed and the number of actual calls until response was recorded for each individual. Since the treatment affected individual's survey response behaviour, respondents sample was non-random. The quantity of interest was the average treatment effect (henceforth ATE) of the job training participation on the employment status.

The model of Behaghel et al. (2015) lies on a crossroad between Heckit (Heckman, 1977) and Lee (2009), as it uses information about the number of calls as a substitute for an instrument to gain narrower bounds on the ATE. Individuals are assumed to respond according to an unobserved variable V, interpreted as a reluctance to respond. This quantity is unaffected by the treatment, and is related to incentives to respond and to the treatment assignment via a function p. A higher effort to reach a person is assumed to have a positive impact on the probability of obtaining the response, so p is a monotone function in V. The reluctance to respond V is normalized to a uniform distribution, so that the function p can be interpreted as a response rate for a given the number of calls and the treatment status.

Our paper presents a method for conducting sensitivity analysis of the results in Behaghel et al. (2015) to the identifying assumptions and therefore extends its applicability. This relaxation is of particular practical interest as the assumption of a strict treatment exogeneity is often not satisfied. Examples include sample selection problem, treatment non-compliance, sample attrition, measurement error, among many others (Berk et al. (1988)). Departure from the assumption of treatment exogeneity is controlled via a relaxation parameter.

We formulate the problem of finding bounds on the ATE as an optimization problem, where we conduct a search in the space of all possible probability distributions that are compatible with the observed data and that satisfy the assumptions of the model. This paper is similar to Lafférs and Nedela Jr (2017), where a sensitivity analysis of the model studied in Lee (2009) was conducted. There is one substantial complication that makes the approach of Lafférs and Nedela Jr (2017) not directly applicable in this paper. Formulating the model of Behaghel et al. (2015) in the optimization framework requires discretizing an unobserved continuous variable and this renders a straightforward implementation infeasible.

This paper contributes to the literature of applications of mathematical programming in the identification analysis (e.g. Balke and Pearl (1997), Honoré and Tamer (2006), Manski (2007), Molinari (2008), Freyberger and Horowitz (2015), Lafférs (2018) and Torgovitsky (2016) among many others).

2 Problem formulation

The following subsection provides a mathematical formulation of the problem proposed in Behaghel et al. (2015).

2.1 Mathematical formulation

Let us define a random vector $U = (Y^1, Y^0, R^1, R^0, V, N, Z)$, with a support $\mathcal{U}$ and let $\pi : \mathcal{U} \mapsto \langle 0, 1\rangle$ denotes a joint probability distribution of U. Let $\Pi(\mathcal{U})$ stands for a set of all possible probability distributions with a support $\mathcal{U}$.

The components of $\mathcal{U}$ are the following:

- Treatment (assignment to the job programme) $Z \in \mathcal{Z} = \{0, 1\}$.
- Potential outcomes (employment status 10 months after the training) $Y^z \in \mathcal{Y} = \{0, 1\},\ \forall z \in \mathcal{Z}$.[1]
- Potential responses (indicator variable of survey response) $R^z \in \mathcal{R} = \{0, 1\},\ \forall z \in \mathcal{Z}$.
- Reluctance to respond $V \in \mathcal{V} = \langle 0, 1\rangle$.
- Number of call attempts $N \in \{1, 2, \ldots, w_{\max}, w_\infty\} = \mathcal{N} \subset \mathbb{N}$. A fixed constant $w_{\max}$ stands for the maximum number of call attempts and $w_\infty \equiv w_{\max} + 1$ denotes individuals not reached until $w_{\max}$ call attempts.

Note that the random vector U is not completely observed. Let $O = (Y, R, N, Z)$ denote the vector of observed random variables with a support $\mathcal{O}$ and with probability distribution $\pi_O : \mathcal{O} \mapsto \langle 0, 1\rangle$. The components of $\mathcal{O}$ are:

- Observed outcome $Y \in \mathcal{Y}$.
- Observed survey response $R \in \mathcal{R}$.
- Number of call attempts $N \in \mathcal{N}$.
- Treatment $Z \in \mathcal{Z}$.

We will also define a function $p : \mathcal{N} \times \mathcal{Z} \mapsto \langle 0, 1\rangle$, that stands for the threshold crossing function. If the reluctance to respond V is smaller than $p(w, z)$ for a given value of a number of call attempts $N = w$ and of a treatment $Z = z$, then we observe the response, so that $R^z = 1$.

Here we list the complete set of the model assumptions:

(C1) Compatibility with observed data: $Y = Y^1 Z + Y^0(1 - Z)$ and $R = R^1 Z + R^0(1 - Z)$.
(C2) Latent variable threshold-crossing response model: $\forall z \in \mathcal{Z} : R^z = \mathbb{1}\Big(V < p(w_{\max}, z)\Big)$.[2]
(C3) Treatment exogeneity: $Z \perp \Big(Y^0, Y^1, V\Big)$.
(C4) Monotonicity of p in w: $\forall z \in \mathcal{Z}, \forall w_1 > w_2 \in \mathcal{N} : p(w_1, z) \geq p(w_2, z)$.
(C5) Uniformity of unobserved reluctance: $V \sim \mathrm{Unif}(0, 1)$.[3]

---

[1] We use binary variables as employment status, although any bounded finite support $\mathcal{Y} \subset \mathbb{R}$ could be used as support for outcome variables Y^1, Y^0 in the model specification.

[2] We use formulation of this condition as in Vytlacil (2002), where the condition holds in both potential outcomes of the response variable.

[3] It is important to note that the assumption (C5) serves as a normalization, it allows us to interpret the $p(w, z)$ as cumulative distribution of observing w call attempts given $Z = z$. For the purposes of identification of ATE_c it is only the *ordering* of $p(w, z)$ that matters.

(C6) Definition of N: $\forall z \in \mathcal{Z}, \forall w \in \mathcal{N} : \Big(V < p(w,z), Z = z \iff N \le w, Z = z \Big)$.[4]

This is a set of restrictions on π and let us call Φ the set of all probability distributions for which there exists a threshold crossing function p, so that they together satisfy (C1)-(C6):

$$\Phi = \{\pi \in \Pi(\mathcal{U}) : \exists p : (\pi, p) \text{ satisfy (C1)-(C6)}\}.$$

Under the model assumptions (C1)-(C6), the function p is identified from the observed distribution π_O. If some of these assumptions were relaxed or dropped, this need not be the case anymore. We address this in the following section.

The aim is to find the sharp bounds on average treatment effect $ATE_c \equiv \mathbb{E}_\pi \Big(Y^1 - Y^0 | V < \overline{p} \Big)$, where $\overline{p} = \min\{p(w_{\max}, 0), p(w_{\max}, 1)\}$. The identification is *local* as it is often the case under the sample selection problem.

By conducting a mechanical search in the space of all joint probability distributions in Φ (that is: those that are compatible with the model assumptions) we find the distribution π that maximizes/minimizes ATE_c. All variables are discrete but V. The following subsection provides a way of discretization of V, so that the identified set for the parameter ATE_c is unaffected. This is important because then the question of identification of ATE_c reduces to a *finite* dimensional optimization problem, more specifically, to a linear program.

2.2 Discretization

Here we define a model Φ^* equivalent to model Φ in the sense of the observed probability distribution π_O and the parameter of interest, ATE_c.

Let us define a random vector $U^* = (Y^1, Y^0, R^1, R^0, V^*, N, Z)$, with joint distribution $\pi^* : \mathcal{U}^* \mapsto \langle 0, 1 \rangle$. The support of $\mathcal{U}^*$ is the same as the support of $\mathcal{U}$, except for V. We define the discretized reluctance to respond: $V^* \in \mathcal{V}^* \subseteq \langle 0, 1 \rangle$, whose finite support is $\mathcal{V}^* = \Big\{ v_1^*, \dots, v_{2 \cdot w_{\max}}^*, v_{2 \cdot w_{\max}+1}^* \Big\}$, which is the ordered set $\{p(w,z); \forall z \in \mathcal{Z}, \forall w \in \mathcal{N} \setminus \{w_\infty\}\} \cup \{1\}$. In the rest of the paper we shall assume that values $p(w,z)$ are different $\forall z \in \mathcal{Z}, \forall w \in \mathcal{N} \setminus \{w_\infty\}$.[5] Now let $v \in \langle v_{k-1}^*, v_k^*)$ maps to v_k^*, $\forall k \in \{1, \dots, 2w_{\max}\}$, $v \in \langle v_{2w_{\max}}^*, v_{2w_{\max}+1}^* \rangle$ maps to $v_{2w_{\max}+1}^* \equiv 1$ and $v_0^* \equiv 0$.

The assumptions that define the model with discrete V^* are the following:

(D1) Compatibility with observed data: $Y = Y^1 Z + Y^0 (1 - Z)$ and $R = R^1 Z + R^0 (1 - Z)$.

(D2) Latent variable threshold-crossing response model: $\forall z \in \mathcal{Z} : R^z = \mathbb{1}\Big(V^* \le p(w_{\max}, z) \Big)$.

(D3) Treatment exogeneity: $Z \perp Y^0, Y^1, V^*$.

(D4) Monotonicity in w: $\forall z \in \mathcal{Z}, \forall w_1 > w_2 \in \mathcal{N} : p(w_1, z) \ge p(w_2, z)$.

(D5) Distribution of discretized reluctance: $\forall v^* \in \mathcal{V}^* : F_{V^*}^*(v^*) = v^*$.

(D6) Definition of N: $\forall z \in \mathcal{Z}, \forall w \in \mathcal{N} : \Big(V^* \le p(w,z), Z = z \iff N \le w, Z = z \Big)$.

[4] We note that in Behaghel et al. (2015), the variable N is defined as "the number of attempts made before a person is actually reached". This definition together with assumptions (C2), (C4) imply (C6). The (C6) assumption is therefore introduced here in order to *define* N in terms of the other variables in the model.

[5] This assumption implies that $P(N = w_1 | Z = z_1)$ and $P(N = w_0 | Z = z_0)$ are different $\forall w_0 \neq w_1 \in \mathcal{N} \setminus \{w_\infty\}$ and $\forall z_0, z_1 \in \mathcal{Z}$. We make this assumption for the sake of exposition. Our method for sensitivity analysis is valid even if such assumption would not hold, but the notation and proofs would be much more complicated.

Similarly, we define

$$\Phi^* = \{\pi^* \in \Pi(\mathcal{U}^*) : \exists p : (\pi^*, p) \text{ satisfy (D1)-(D6)}\}.$$

We have to show that this discretized model Φ^* identifies the same bounds on ATE_c as the original model Φ. After this is established, we can use the linear programming approach to find sharp bounds on the conditional average treatment effect: $ATE_c^* = \mathbb{E}_{\pi^*}\Big(Y^1 - Y^0 | V^* \leq \overline{p}\Big)$.

The following proposition formalizes the equivalence of the original model and the discretized problem.

Proposition 1.

$$\forall \theta \in \mathbb{R} : \bigg(\exists \pi \in \Phi, \theta = \mathbb{E}_{\pi}\Big(Y^1 - Y^0 | V < \overline{p}\Big)\bigg) \iff \bigg(\exists \pi^* \in \Phi^*, \theta = \mathbb{E}_{\pi^*}\Big(Y^1 - Y^0 | V^* \leq \overline{p}\Big)\bigg).$$

This proposition allows to state the identification of bounds on ATE_c as a problem of finding global extrema (through all $\pi^* \in \Phi^*$) of ATE_c given the observed probability π_O and restrictions imposed on Φ. The proof is given in Appendix A.

2.3 Linear program formulation

This subsection presents how finding the bounds on ATE_c may be formulated as a linear program. More precisely, we express the ATE_c as a linear function of the joint probability distribution $\pi*$ and we translate the identifying assumptions into linear restrictions on π^*.

Let n and m denote the size of support of V^* and $(Y^1, Y^0, R^1, R^0, N, Z)$ respectively and let $m.n$ be the size of the discrete support $\mathcal{U}^*$ in the probability space $(\Omega, \mathcal{F}, \pi)$. Given that all the variables have finite support, the support $\mathcal{U}^*$ is also finite. It is therefore sufficient to have a discrete set for $\Omega \equiv \sqcup_{i=1}^{m \cdot n} \omega_i$ in order to represent the probabilistic behaviour of U^*. Let the elements of $\mathcal{U}^*$ be denoted as $u_i^* \equiv U(\omega_i) = (Y^1(\omega_i), Y^0(\omega_i), R^1(\omega_i), R^0(\omega_i), V^*(\omega_i), N(\omega_i), Z(\omega_i))$, $\forall \omega_i \in \Omega$, with corresponding probabilities $\pi_i^* \equiv \pi^*(U^* = u_i^*)$, $\forall i \in \{1, \ldots, m \cdot n\}$. The index i represents the i-th combination of values that the vector $(Y^1, Y^0, R^1, R^0, V^*, N, Z)$. We denote $y^1(i) \equiv Y^1(\omega_i)$ the value of the first component of the vector u_i, where the elements u_i's are lexicographically ordered. The notation for other vector components follows similarly. Let $\overline{z} \equiv \operatorname{argmin}_{z \in \mathcal{Z}} p(w_{\max}, z)$.

Define c_i, $\forall i \in \{1, \ldots, m \cdot n\}$:

$$c_i = \begin{cases} (y^1(i) - y^0(i))/\pi_O(R = 1 | Z = \overline{z}), & \text{if } v^*(i) \leq \overline{p}, \\ 0, & \text{otherwise.} \end{cases}$$

When assumptions (D1), (D2) and (D3) are satisfied, the following equality holds:

$$\pi_O(R = 1 | Z = \overline{z}) = \pi^*(R^{\overline{z}} = 1 | Z = \overline{z}) = \pi^*(V^* \leq \overline{p} | Z = \overline{z}) = \pi^*(V^* \leq \overline{p}).$$

Bounds on ATE_c can be written as $\min_{\pi^*}\Big(\sum_{i=1}^{m \cdot n} c_i \cdot \pi_i^*\Big)$ and $\max_{\pi^*}\Big(\sum_{i=1}^{m \cdot n} c_i \cdot \pi_i^*\Big)$, $\forall \pi^* \in \Phi^*$, with constraints:

(D1) Observed distribution

$$\pi^*\Big(Y^1Z + Y^0(1-Z) = \tilde{y}, R^1Z + R^0(1-Z) = \tilde{r}, N = \tilde{w}, Z = \tilde{z}\Big) =$$

$$= \sum_{i=1}^{m\cdot n} \pi_i^* \cdot \mathbb{1}\Big(y^1(i)z(i) + y^0(i)(1-z(i)) = \tilde{y}, r^1(i)z(i) + r^0(i)(1-z(i)) = \tilde{r}, w(i) = \tilde{w}, z(i) = \tilde{z}\Big) =$$

$$= \pi_O(O = \tilde{o}),$$

$$\forall \tilde{o} = (\tilde{y}, \tilde{r}, \tilde{w}, \tilde{z}) \in \mathcal{O}.$$

(D2) Threshold crossing

$$\pi^*\Big(\mathbb{1}(V^* \le p(w_{\max}, \tilde{z})) = R^{\tilde{z}}\Big) = \sum_{i=1}^{m\cdot n} \pi_i^* \cdot \mathbb{1}\Big(\mathbb{1}(v^*(i) \le p(w_{\max}, \tilde{z})) = r^{\tilde{z}}(i)\Big) = 1,$$

$$\forall \tilde{z} \in \mathcal{Z}.$$

(D3) Treatment exogeneity

$$\pi^*(Y^1 = \tilde{y^1}, Y^0 = \tilde{y^0}, V^* = \tilde{v}^*, Z = 1)/\pi^*(Z = 1) -$$

$$-\pi^*(Y^1 = \tilde{y^1}, Y^0 = \tilde{y^0}, V^* = \tilde{v}^*, Z = 0)/\pi^*(Z = 0) =$$

$$\sum_{i=1}^{m\cdot n} \pi_i^* \cdot \Big(\mathbb{1}\Big(y^1(i) = \tilde{y^1}, y^0(i) = \tilde{y^0}, v^*(i) = \tilde{v}^*, z(i) = 1\Big)/\pi_O\Big(Z = 1\Big) -$$

$$-\mathbb{1}\Big(y^1(i) = \tilde{y^1}, y^0(i) = \tilde{y^0}, v^*(i) = \tilde{v}^*, z(i) = 0\Big)/\pi_O\Big(Z = 0\Big)\Big) = 0,$$

$$\forall \tilde{y^1}, \tilde{y^0} \in \mathcal{Y}, \forall \tilde{v}^* \in \mathcal{V}^*.$$

(D4) Monotonicity of p in z

$$p(w_1, \tilde{z}) \ge p(w_2, \tilde{z}), \forall \tilde{z} \in \mathcal{Z}, \forall w_1 > w_2 \in \mathcal{N}.$$

(D5) Distribution of V^*

$$F^*_{V^*}(v^*) = \tilde{v}^*, \forall \tilde{v}^* \in \mathcal{V}^*.$$

(D6) Definition of N

$$\pi^*\Big(\mathbb{1}\Big(V^* \le p(w, \tilde{z}), Z = \tilde{z}\Big) = \mathbb{1}\Big(N \le w, Z = \tilde{z}\Big)\Big) =$$

$$= \sum_{i=1}^{m\cdot n} \pi_i^* \cdot \mathbb{1}\Big(\mathbb{1}\Big(v^*(i) \le p(w, \tilde{z}), z(i) = \tilde{z}\Big) = \mathbb{1}\Big(w(i) \le w, z(i) = \tilde{z}\Big)\Big) = 1,$$

$$\forall \tilde{z} \in \mathcal{Z}, \forall w \in \mathcal{N}.$$

We note that values of $p(w, z)$ are observed $\forall z \in \mathcal{Z}, \forall w \in \mathcal{N}$, when assumptions (D3)-(D6) hold, as they represent the proportion of (treated or non-treated) respondents until the w-th call. Also, the distribution of condition (C5) or (D5) was chosen for interpretation purposes only and could be chosen arbitrarily (see Behaghel et al. (2015)), it has no use in linear program implementation except for labelling the support of V^*. Condition (D4) is embeded in condition (D6) as the marginal probabilities $\pi^*\Big(V^* \le p(w, \tilde{z}), Z = \tilde{z}\Big) = \pi^*\Big(N \le w, Z = \tilde{z}\Big)$ are monotone in w given $\tilde{z}$ for all $w \in \mathcal{N}$.

Hence the problem of bounding ATE_c may stated as two linear optimalization problems $\min_{\pi^*} / \max_{\pi^*}\{c^\intercal \pi^* : \pi^* \text{ satisfies (D1)-(D3), (D6)}, \pi^* \ge 0\}$.

3 Sensitivity Analysis

This section formulates a sensitivity analysis of the bounds on ATE on identifying assumptions in terms of an optimization problem and this constitutes the main contribution of this paper. Instead of deriving analytical results under relaxed assumptions and proving their validity and sharpness, we can translate the relaxed assumption as a restriction in the optimization framework and get the bounds in a straightforward manner. The usefulness of linear programming formulation for sensitivity analysis of ATE bounds is not new and was previously used in Lafférs (2018) and Lafférs (2019). This paper is, to the best of our knowledge, the first one that explores the robustness of the Behaghel et al. (2015) bounds to treatment exogeneity.

in many applications, the assumption of treatment exogeneity might be violated because of treatment noncompliance, endogenous sample selection and/or mismeasurement of outcomes and thus leading to misleading results. The purpose of the presented method is to extend the usefulness of the method of Behaghel et al. (2015) for situations when the treatment assignment is not completely random.

In the following subsections we provide mathematical formulation of the problem with relaxed treatment exogeneity assumption (C3) and formulate the problem of finding bounds on ATE_c as an optimization problem.

3.1 Mathematical formulation

We reformulate the problem using relaxation parameter α_E to allow treatment exogeneity to be relaxed in a meaningful and controlable manner while still preserving linear optimisation problem formulation. Thus we are able to model the effect of treatment exogeneity violation on ATE_c. Given the value of $\alpha_E \in \langle 0, 1\rangle$ we can relax treatment exogeneity in the following way:

(rC3) Relaxed treatment exogeneity:

$$\left|\pi(Y^1 = \tilde{y}^1, Y^0 = \tilde{y}^0, V = \tilde{v}|Z = 1) - \pi(Y^1 = \tilde{y}^1, Y^0 = \tilde{y}^0, V = \tilde{v}|Z = 0)\right| \leq \alpha_{\tilde{y}^1\tilde{y}^0}(\tilde{v}),$$

$$\forall \tilde{y}^1 \times \tilde{y}^0 \times \tilde{v} \in \mathcal{Y} \times \mathcal{Y} \times \mathcal{V}, \text{ where } \sum_{\tilde{y}^1,\tilde{y}^0 \in \mathcal{Y}} \int_0^1 \alpha_{\tilde{y}^1\tilde{y}^0}(\tilde{v}) \mathrm{d}\tilde{v} = \alpha_E, \alpha_{\tilde{y}^1\tilde{y}^0}(\tilde{v}) \geq 0.$$

Nonnegative bounding functions $\alpha_{\tilde{y}^1\tilde{y}^0}(\tilde{v})$ are defined on support of variable V.

In the discretized version of the problem we use $n^2_{\mathcal{Y}} \cdot n$ parameters $\alpha_{\tilde{y}^1\tilde{y}^0k}$ corresponding to each of the sets $\tilde{y}^1 \times \tilde{y}^0 \times \langle v^*_{k-1}, v^*_k\rangle$, where $n_{\mathcal{Y}}$ is size of set $\mathcal{Y}$.

(rD3) Relaxed treatment exogeneity:

$$\left|\pi(Y^1 = \tilde{y}^1, Y^0 = \tilde{y}^0, V^* = \tilde{v}^*_k|Z = 1) - \pi(Y^1 = \tilde{y}^1, Y^0 = \tilde{y}^0, V^* = \tilde{v}^*_k|Z = 0)\right| \leq \alpha^*_{\tilde{y}^1\tilde{y}^0k},$$

$$\forall \tilde{y}^1 \times \tilde{y}^0 \times \tilde{v}^*_k \in \mathcal{Y} \times \mathcal{Y} \times \mathcal{V}^*, \forall k = 1, \ldots, n, \text{ where } \sum_{\tilde{y}^1,\tilde{y}^0 \in \mathcal{Y}} \sum_{k=1}^{n} \alpha^*_{\tilde{y}^1\tilde{y}^0k} = \alpha_E, \alpha^*_{\tilde{y}^1\tilde{y}^0k} \geq 0.$$

Set of inequalities (rC3) is equivalent to setting the total variation distance of conditional probability distributions $\pi^*(.|Z=0)$ and $\pi^*(.|Z=1)$ over $\mathcal{Y}\times\mathcal{Y}\times\mathcal{V}^*$ to be less or equal than $\frac{\alpha_E}{2}$. The total variation distance between two probability measures π_1 and π_2 is defined as $TV(\pi_1,\pi_1)\equiv\sup_{A\in\mathcal{B}}|\pi_1(A)-\pi_2(A)|$, where $\mathcal{B}$ stands for all Borel sets. In case that π_1 and π_2 are discrete distributions, then $TV(\pi_1,\pi_2)=\frac{1}{2}\sum_{x\in\Omega}|\pi_1(x)-\pi_2(x)|$.

Similarly as in the original (nonrelaxed) version of the problem, if the discretized relaxed model identifies identical parameter bounds, we can find them by the search through all possible probability distributions satisfying discretized model assumptions using finite dimensional optimalization. We define the set of probability distributions compatible with the relaxed continuous model as

$$\Phi_r=\{\pi\in\Pi(\mathcal{U}):\exists p:(\pi,p)\text{ satisfies (C1),(C2),(rC3),(C4)-(C6)}\},$$

and with the relaxed discretized model as

$$\Phi_r^*=\{\pi^*\in\Pi(\mathcal{U}^*):\exists p:(\pi,p)\text{ satisfies (D1),(D2),(rD3),(D4)-(D6)}\}.$$

The following proposition shows that this discretization does not change the bounds for ATE_c.

Proposition 2.

$$\forall\theta\in\mathbb{R}:\Big(\exists\pi\in\Phi_r,\theta=\mathbb{E}_\pi\Big(Y^1-Y^0|V<\overline{p}\Big)\Big)\iff\Big(\exists\pi^*\in\Phi_r^*,\theta=\mathbb{E}_{\pi_\zeta^*}\Big(Y^1-Y^0|V^*\le\overline{p}\Big)\Big).$$

The proof is provided in Appendix A.

3.2 Optimization formulation

We begin the optimization formulation with changing the objective function of the linear program by adding $n_{\mathcal{Y}}^2\cdot n+1$ zero coefficients d_i and removing the denominator with the share of always-responding individuals.

Define d_i, $\forall i\in\{1,\dots,m\cdot n+n_{\mathcal{Y}}^2\cdot n+1\}$:

$$d_i=\begin{cases}y^1(i)-y^0(i), & \text{if } v^*(i)\le\overline{p}\text{ and } i\le m\cdot n\\ 0, & \text{otherwise.}\end{cases}$$

After relaxing independence between V^* and Z the quantity $\eta^*\equiv\pi^*(V^*\le\overline{p})$ is not observed anymore. This variable, however, can be bounded and used as an additional unknown variable in the linear optimization formulation.
Expressing η^* yields

$$\eta^*=\pi^*(V^*\le\overline{p})=\pi^*(V^*\le\overline{p}|Z=\overline{z})\cdot\pi^*(Z=\overline{z})+\pi^*(V^*\le\overline{p}|Z=1-\overline{z})\cdot\pi^*(Z=1-\overline{z})=$$
$$=\pi_O(R=1|Z=\overline{z})\cdot\pi_O(Z=\overline{z})+\pi^*(V^*\le\overline{p}|Z=1-\overline{z})\cdot\pi_O(Z=1-\overline{z}),$$

where $\pi^*(V^*\le\overline{p}|Z=1-\overline{z})$ is unobserved and bounded by

$$\pi_O(N\le\underline{w}|Z=1-\overline{z})\le\pi^*(V^*\le\overline{p}|Z=1-\overline{z})\le\pi_O(N\le\overline{w}|Z=1-\overline{z})$$

and values $\underline{w}$ and $\overline{w}$ are defined as

$$\underline{w}\equiv\underset{w\in\mathcal{N}}{\operatorname{argmin}}\{p(w,1-\overline{z})\in\mathcal{V}^*:p(w,1-\overline{z})\ge\overline{p}\},$$

$$\overline{w} \equiv \underset{w \in \mathcal{N}}{\text{argmax}} \{p(w, 1 - \overline{z}) \in \mathcal{V}^* : p(w, 1 - \overline{z}) \leq \overline{p}\}.$$

From now on, suppose that the function p is fixed. Using reparametrization $\pi_i^{**} \equiv \pi_i^*/\eta^*$, $\alpha_{\tilde{y}^1\tilde{y}^0k}^{**} \equiv \alpha_{\tilde{y}^1\tilde{y}^0k}^*/\eta^*$ and $\eta^{**} \equiv 1/\eta^*$ we conduct a joint search through $\pi^{**} \equiv (\pi_1^{**}, ..., \pi_{m \cdot n}^{**})$, $\alpha^{**} \equiv (\alpha_{\underline{y},\underline{y},1}^{**}, \alpha_{\underline{y},\underline{y},2}^{**}, ..., \alpha_{\overline{y},\overline{y},n}^{**})$ and η^{**} in a linear program with constraints on $(\pi^{**}, \alpha^{**}, \eta^{**})$, where $\underline{y} \equiv \min(\mathcal{Y})$, $\overline{y} \equiv \max(\mathcal{Y})$.

(rD1) Observed distribution

$$\sum_{i=1}^{m \cdot n} \pi_i^{**} \cdot \mathbb{1}\Big(y^1(i)z(i) + y^0(i)(1 - z(i)) = \tilde{y}, r^1(i)z(i) + r^0(i)(1 - z(i)) = \tilde{r}, w(i) = \tilde{w}, z(i) = \tilde{z}\Big) -$$

$$-\eta^{**} \cdot \pi_O(O = \tilde{o}) = 0,$$

$$\forall \tilde{o} = (\tilde{y}, \tilde{r}, \tilde{w}, \tilde{z}) \in \mathcal{O}.$$

(rD2) Treshold crossing

$$\sum_{i=1}^{m \cdot n} \pi_i^{**} \cdot \mathbb{1}\Big(\mathbb{1}(v^*(i) \leq p(w_{\max}, \tilde{z})) = r^{\tilde{z}}(i)\Big) - \eta^{**} = 0,$$

$$\forall \tilde{z} \in \mathcal{Z}.$$

(rD3 a) Relaxed treatment exogeneity

$$\sum_{i=1}^{m \cdot n} \pi_i^{**} \cdot \Big| \mathbb{1}\Big(y^1(i) = \tilde{y}^1, y^0(i) = \tilde{y}^0, v^*(i) = \tilde{v}_k^*, d(i) = 1\Big) / \pi_O\Big(Z = 1\Big) -$$

$$-\mathbb{1}\Big(y^1(i) = \tilde{y}^1, y^0(i) = \tilde{y}^0, v^*(i) = \tilde{v}_k^*, d(i) = 0\Big) / \pi_O\Big(Z = 0\Big)\Big| - \alpha_{\tilde{y}^1\tilde{y}^0k}^{**} \leq 0,$$

$$\forall \tilde{y}^1 \times \tilde{y}^0 \times \tilde{v}_k^* \in \mathcal{Y} \times \mathcal{Y} \times \mathcal{V}^*, \forall k = 1, \ldots, n.$$

(rD3 b) Total variation

$$\sum_{\tilde{y}^1, \tilde{y} \in \mathcal{Y}} \sum_{k=1}^{n} \alpha_{\tilde{y}^1\tilde{y}^0k}^{**} - \eta^{**}\alpha_E = 0.$$

(rD4) Monotonicity of p in z

$$p(w_1, \tilde{z}) \geq p(w_2, \tilde{z}), \forall \tilde{z} \in \mathcal{Z}, \forall w_1 > w_2 \in \mathcal{N}.$$

(rD5) Distribution V^*

$$F_{V^*}^*(v^*) = \tilde{v}^*, \forall \tilde{v}^* \in \mathcal{V}^*.$$

(rD6) Definition of N

$$\sum_{i=1}^{m \cdot n} \pi_i^{**} \cdot \mathbb{1}\Big(\mathbb{1}\Big(v^*(i) \leq p(w, \tilde{z}), z(i) = \tilde{z}\Big) = \mathbb{1}\Big(w(i) \leq w, z(i) = \tilde{z}\Big)\Big) - \eta^{**} = 0,$$

$$\forall \tilde{z} \in \mathcal{Z}, \forall w \in \mathcal{N}.$$

(rD7) Always responding

$$\sum_{i=1}^{m\cdot n} \pi_i^{**} \cdot \mathbb{1}\Big(v^*(i) \leq \overline{p}\Big) = 1.$$

Now the set of all possible values of π^{**} is

$$\Phi_r^{**}\Big\{\pi^{**} \in \Pi^{**}(\mathcal{U}^*) : \exists p : \exists \alpha^{**} \geq 0 : \exists \eta^{**} \geq 0 : (\pi^{**}, \alpha^{**}, \theta^{**})$$

$$\text{satisfy (rD1),(rD2),(rD3a),(rD3b),(rD4)-(rD7)}\Big\},$$

where $\Pi^{**}(\mathcal{U}^*) \equiv \{\pi^{**} = \pi^*/\pi^*(V^* \leq \overline{p}) : \pi^* \in \Pi(\mathcal{U}^*)\}$. We obtain the following bounds on ATE_c: $\min\limits_{\pi^{**}\in\Phi_r^{**}} / \max\limits_{\pi^{**}\in\Phi_r^{**}} \left(\sum\limits_{i=1}^{m\cdot n} d_i \pi_i^{**}\right)$. For a given function p, these optimization problems are still within the linear programming framework.

Since the treatment exogeneity is now relaxed, we loose observability of $\pi^*(V^* \leq p(w,z))$ which is not necessarily equal $\pi^*(V^* \leq p(w,z)|Z = z) = \pi_O(N \leq w|Z = z)$. The implementation now requires consideration of all possible shapes of the function p. By shape of function p we mean the order of function values $p(w,z)$, $\forall z \in \mathcal{Z}$, $\forall w \in \mathcal{N}$.

Each order of function values formulates different condition (rD6) as it sets different cell probabilities of π^{**} to zero while monotonicity of p in z (rD4) is satified. The values of $p(w_\infty, 0)$ and $p(w_\infty, 1)$ are set to zero. The number of all possible function shapes satisfying condition (rD4) depends on size of support of $\mathcal{N}$ and is equal to $\binom{n-1}{w_{\max}}$. For each of these shapes we run a different linear program and then we find bounds as minimum/maximum over all possible functions p.

We use an artificial example with known ATE_c value in the next section in order to illustrate the use of optimization with relaxed assumptions.

4 Illustration

In this section we provide an example of ATE_c bounds estimates with confidence intervals on generated data from known unobserved probability distribution. We consider an artificial dataset with 9999 observations, where unobserved proportions meet conditions (D1)–(D6) and the function p is known and fixed. Number of maximum call attempts $w_{max} = 3$, observed and potential outcomes are binary and treatment exogeneity in the generated data is satisfied (so that $\alpha_E = 0$). The shape of p function of generated dataset is depicted in figure 1. The real size of parameter ATE_c in generated dataset is -0.2504.

Except for the case $\alpha_E = 0$ we have to run the linear program for every possible shape of p. In the case of 3 maximum call attempts, for this setup $\binom{7-1}{3} = 20$ linear program runs are needed for each bound evaluation. The 99% confidence intervals were computed by subsampling from observed dataset, see (Politis et al., 1999). The ATE_c parameter bounds along with 99% confidence interval are plotted for different relaxion parameter α_E values in figure 2. [6]

[6] MATLAB code is available from the authors upon request.

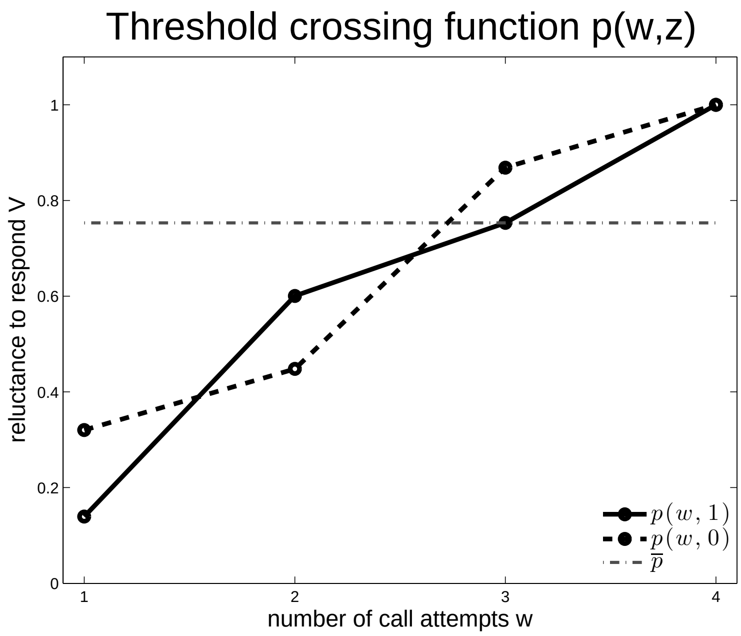

Figure 1: Threshold crossing function $p(w,z)$ in the illustrative example.

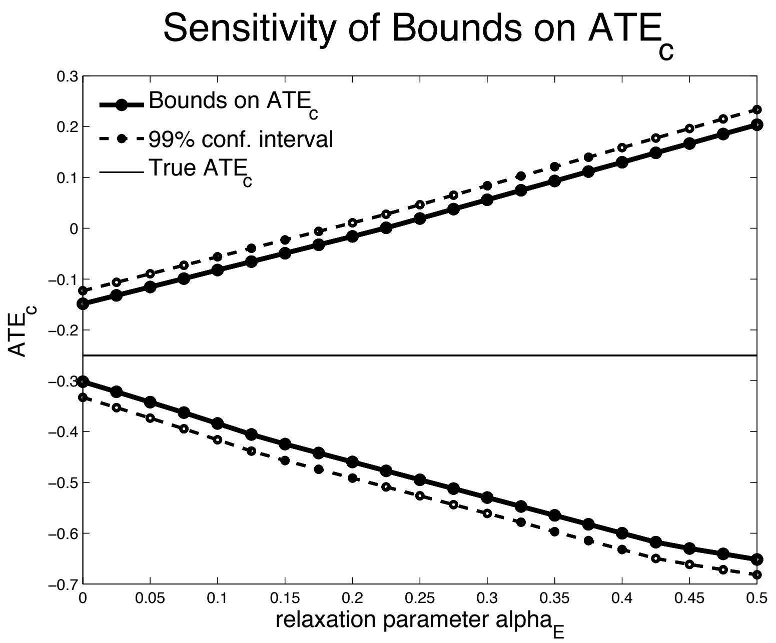

Figure 2: Sensitivity of the ATE_c on the departure from treatment exogeneity assumption controlled via relaxation parameter α_E. 99% confidence intervals based on subsampling (Politis et al., 1999) with subsample size $m = \lfloor n^{0.6} \rfloor$. Solid line in the middle depicts the true value of ATE_c = -0.2504.

5 Conclusion

This paper considered the problem of identification of average treatment effect in case of survey non-response, when the number of attempts to obtain response from individuals is recorded. We formulated the identification problem as a finite dimensional optimization problem and such formulation allowed us to further extend the results of Behaghel et al. (2015) by considering sensitivity analysis to the main identification assumption - treatment assignment exogeneity. We control the departure from this assumption via an interpretable parameter. We suggest such sensitivity analysis to be reported whenever Behaghel et al. (2015) bounds are estimated and the validity of the exogeneity assumption is questioned.

6 Acknowledgements

This research was supported by VEGA grant 1/0843/17 and Slovak Research and Development Agency, grant number APVV-17-0329.

A Appendix - Proofs

Further notation for proofs

Let m be the size of support of marginal distribution of π^* without V^* denoted as $\pi^{*,-V^*}$, which is the same as size of support of marginal distribution of π for V denoted as π^{-V}. Given that all the variables but V are discrete, the support of π^{-V} is finite. We use an index j for all the possible combinations that the vector $(Y^1, Y^0, R^1, R^0, N, Z)$ can attain. Symbols $\pi_j^{*,-V^*}$, π_j^{-V} will stand for a probability of this j-th combination and $\pi_j(v)$ (or $\pi_j^*(v^*)$) will denote the joint probability of the j-th combination and simultaneously $V = v$ (or $V^* = v^*$).

Let $n = 2w_{\max} + 1$ be the size of support of marginal distribution of π^* for V^* denoted as π^{*,V^*}. Furthermore, let F_X, F_X^* denote cumulative distribution functions for a random vector X.

Proof of Proposition 1.

Part "($\implies$)":

First we will show that given $\pi \in \Phi$ we can find π^* that satisfies the model restrictions Φ^* and that ATE_c and ATE_c^* are equal.

For $\forall j \in \{1, \ldots, m\}, \forall k \in \{1, \ldots, n\}$, and for an ordered sequence $\left\{v_k^*\right\}_{k=1}^n$ with $v_k^* \in \mathcal{V}^*$ and $v_0^* = 0$, we define

$$\pi_{j,k}^* \equiv \pi_j^*(v_k^*) = \int_{v_{k-1}^*}^{v_k^*} \pi_j(v) \mathrm{d}v. \tag{A.1}$$

From equation (A.1) we observe that the marginal distributions π_j^{-V} and $\pi_j^{*,-V^*}$ are identical: $\pi_j^{*,-V^*} = \sum_{k=1}^n \pi_{j,k}^* = \sum_{k=1}^n \int_{v_{k-1}^*}^{v_k^*} \pi_j(v) \mathrm{d}v = \pi_j^{-V}$.

Since the marginal distributions $\pi_j^{*,-V^*}$, π_j^{-V} are equal, the observed distribution of (Y, R, N, Z) is the same and conditions (C1) and (D1) are identical.

(C2) $\implies$ (D2)
It is sufficient to prove that $\forall z \in \mathcal{Z} : V < p(w_{\max}, z) \implies V^* \leq p(w_{\max}, z)$. This can be seen from definition of V^*. There exists $v_K^* \in \mathcal{V}^*$ such that $v_K^* = p(w_{\max}, z)$. If V takes values in interval $\langle 0, v_K^*)$, then V maps to ordered values in $\{v_1^*, \ldots, v_K^*\}$ and thus $V^* \leq v_K^*$.

(C3) $\implies$ (D3)
Condition (C4) can be translated as $\forall z \in \mathcal{Z}, \forall y^1 \in \mathcal{Y}, \forall y^0 \in \mathcal{Y}, \forall v \in \mathcal{V} : F_{Y^1,Y^0,V}(y^1, y^0, v) \cdot F_Z(z) = F_{Y^1,Y^0,V,Z}(y^1, y^0, v, z)$.

We set $\forall k = 1, \ldots, n : v = v_k^*$ and then from equation A.1 we get that

$$F_Z^*(z) = F_Z(z),$$
$$F^*_{Y^1,Y^0,V^*}(y^1, y^0, v^*) = F_{Y^1,Y^0,V}(y^1, y^0, v),$$
$$F^*_{Y^1,Y^0,V^*,Z}(y^1, y^0, v^*, z) = F_{Y^1,Y^0,V,Z}(y^1, y^0, v, z),$$

so that condition (C3) holds.

(C4) $\implies$ (D4)
This holds trivially as these conditions do not involve V.

(C5) $\implies$ (D5)
We need to show that $V \sim \text{Unif}(0, 1) \implies \forall v^* \in \mathcal{V}^*, F^*_{V^*}(v^*) = v^*$. We get this directly by summing using equation A.1.

$$\forall v_K^* \in \mathcal{V}^* : \;\; F^*_{V^*}(v_K^*) = \sum_{j=1}^{m} \sum_{k=1}^{K} \pi^*_{j,k} = \sum_{j=1}^{m} \sum_{k=1}^{K} \int_{v_{k-1}^*}^{v_k^*} \pi_j(v) \mathrm{d}v = F_V(v_K^*) = v_K^*.$$

(C6) $\implies$ (D6)
We need to prove that $\forall z \in \mathcal{Z}, \forall w \in \mathcal{N} : V < p(w, z) \implies V^* \leq p(w, z)$. This can be done substituting $w_{\max}$ with w in the proof of (C2) $\implies$ (D2). Trivially $Z = z \implies Z = z$. Then assuming eqiuvalence (C6) we get $\Big(V^* \leq p(w, z), Z = z \implies N \leq w, Z = z\Big)$.

It is left is to show that ATEs are the same: $\mathbb{E}^*\Big(Y^1 - Y^0 | V^* \leq \overline{p}\Big) = \mathbb{E}\Big(Y^1 - Y^0 | V < \overline{p}\Big)$. It is sufficient to show that $\pi\Big(y^1 | V^* \leq \overline{p}\Big) = \pi^*\Big(y^1 | V < \overline{p}\Big)$ and $\pi\Big(y^0 | V^* \leq \overline{p}\Big) = \pi^*\Big(y^0 | V < \overline{p}\Big)$.
From equation (A.1) we have that $\forall z \in \mathcal{Z}, \pi\Big(y^z, V < \overline{p}\Big) = \pi^*\Big(y^z, V^* \leq \overline{p}\Big)$ and $\pi\Big(V < \overline{p}\Big) = \pi^*\Big(V^* \leq \overline{p}\Big)$, where $\overline{p} = \min\{p(w_{\max}, 0), p(w_{\max}, 1)\}$, which proves the previous assertion.

Part "($\impliedby$)":

Now we show that given $\pi^* \in \Phi^*$ we can find π that satisfies restrictions Φ and that ATE is equal. We construct the continuous distribution π in the following way:

$\forall j \in \{1, \ldots, m\}, \forall k \in \{1, \ldots, n\}, \forall v \in \langle v_{k-1}^*, v_k^*)$ let

$$\pi_j(v) = \frac{\pi^*_{j,k}}{\pi_k^{*,V^*}}. \tag{A.2}$$

From equation A.2 and condition (C5), we see that marginal distributions π_j^{-V} and $\pi_j^{-V^*}$ are identical

$$\pi_j^{-V} = \int_0^1 \pi_j(v)\mathrm{d}v = \int_{v_0^*}^{v_n^*} \pi_j(v)\mathrm{d}v = \sum_{k=1}^{n} \int_{v_{k-1}^*}^{v_k^*} \frac{\pi_{j,k}^*}{\pi_k^{*,V^*}}\mathrm{d}v = \sum_{k=1}^{n} \frac{\pi_{j,k}^*}{\pi_k^{*,V^*}} \int_{v_{k-1}^*}^{v_k^*} \mathrm{d}v =$$

$$= \sum_{k=1}^{n} \frac{\pi_{j,k}^*}{\pi_k^{*,V^*}}\Big(F_{V^*}^*(v_k^*) - F_{V^*}^*(v_{k-1}^*)\Big) = \sum_{k=1}^{n} \frac{\pi_{j,k}^*}{\pi_k^{*,V^*}}\pi_k^{*,V^*} = \sum_{k=1}^{n} \pi_{j,k}^* = \pi_j^{-V^*}.$$

This means that condition (C1) is trivially implied by (D1) as in part "($\implies$)".

(D2) $\implies$ (C2)
From definition of V^* have that $\forall z \in \mathcal{Z}$, $\exists v_K^* \in \mathcal{V}^* : v_K^* = p(w_{\max}, z)$. If $V^* \leq v_K^*$ then $V^* \in \{v_1^*, \dots, v_K^*\}$ and $V \in \bigcup_{k=1}^{n} \langle v_{k-1}^*, v_k^*)$, thus $V < v_K^* = p(w_{\max}, z)$ because v_k^*'s are ordered.

(D3) $\implies$ (C3)
$\forall z \in \mathcal{Z}, \forall y^1 \in \mathcal{Y}, \forall y^0 \in \mathcal{Y}$, $\forall v \in \mathcal{V}, \exists k \in \{1, \dots, n\}$, $v \in \langle v_{k-1}^*, v_k^*)$, then summing using equation A.2 we get:

$$\pi_Z(z) = \pi_Z^*(z),$$

$$\frac{\pi_{Y^1,Y^0,V,Z}(y^1, y^0, v, z)}{\pi_{Y^1,Y^0,V}(y^1, y^0, v)} = \frac{\pi_{Y^1,Y^0,V^*,Z}^*(y^1, y^0, v_k^*, z)/\pi_{V^*}^*(v_k^*)}{\pi_{Y^1,Y^0,V^*}^*(y^1, y^0, v_k^*)/\pi_{V^*}^*(v_k^*)} =$$

$$= \frac{\pi_{Y^1,Y^0,V^*,Z}^*(y^1, y^0, v_k^*, z)}{\pi_{Y^1,Y^0,V^*}^*(y^1, y^0, v_k^*)} = \pi_Z^*(z),$$

where π_X, π_X^* are marginal distributions for random vector X.

(D4) $\implies$ (C4)
Function p has the same definition in both problems.

(D5) $\implies$ (C5)
$\forall v \in \mathcal{V}, \exists K \in \{1, \dots, n\}$, $v \in \langle v_{K-1}^*, v_K^*)$, then summing using equation A.2:

$$F_V(v) = \int_{v_0^*}^{v} \pi^V(v)\mathrm{d}v = \int_{v_0^*}^{v_{K-1}^*} \pi^V(v)\mathrm{d}v + \int_{v_{K-1}^*}^{v} \pi^V(v)\mathrm{d}v =$$

$$= \sum_{j=1}^{m}\sum_{k=1}^{K-1} \int_{v_{k-1}^*}^{v_k^*} \pi_j(v)\mathrm{d}v + \sum_{j=1}^{m} \int_{v_{K-1}^*}^{v} \pi_j(v)\mathrm{d}v = \sum_{j=1}^{m}\sum_{k=1}^{K-1} \int_{v_{k-1}^*}^{v_k^*} \frac{\pi_{j,k}^*}{\pi_k^{*,V^*}}\mathrm{d}v + \sum_{j=1}^{m} \int_{v_{K-1}^*}^{v} \frac{\pi_{j,K}^*}{\pi_K^{*,V^*}}\mathrm{d}v =$$

$$= \sum_{j=1}^{m}\sum_{k=1}^{K-1} \frac{\pi_{j,k}^*}{\pi_k^{*,V^*}}\Big(v_k^* - v_{k-1}^*\Big) + \sum_{j=1}^{m} \frac{\pi_{j,K}^*}{\pi_K^{*,V^*}}\Big(v - v_{K-1}^*\Big) = \sum_{j=1}^{m}\sum_{k=1}^{K-1} \frac{\pi_{j,k}^*}{\pi_k^{*,V^*}}\Big(F_{V^*}^*(v_k^*) - F_{V^*}^*(v_{k-1}^*)\Big) +$$

$$+ \frac{\Big(v - v_{K-1}^*\Big)}{\pi_K^{*,V^*}} \sum_{j=1}^{m} \pi_{j,K}^* = \sum_{j=1}^{m}\sum_{k=1}^{K-1} \pi_{j,k}^* + \frac{\Big(v - v_{K-1}^*\Big)}{\pi_K^{*,V^*}}\pi_K^{*,V^*} = v_{K-1}^* + v - v_{K-1}^* = v.$$

From $F_V(v) = v$ we have $V \sim \text{Unif}(0,1)$.

(D6) $\implies$ (C6)
Proving this implication requires proving that $\forall z \in \mathcal{Z}, \forall w \in \mathcal{N} : V^* < p(w,z) \implies V \leq p(w,z)$. This can be done substituting $w_{\max}$ with w in the proof of (D2) $\implies$ (C2). While $Z = z \implies Z = z$, from eqiuvalence (D6) we have $\Big(V^* \leq p(w,z), Z = z \implies N \leq w, Z = z\Big)$.

Now it's left to show that ATEs are the same: $\mathbb{E}^*\Big(Y^1 - Y^0|V^* \leq \overline{p}\Big) = \mathbb{E}\Big(Y^1 - Y^0|V < \overline{p}\Big)$. This is true if $\pi\Big(y^1|V^* \leq \overline{p}\Big) = \pi^*\Big(y^1|V < \overline{p}\Big)$ and $\pi\Big(y^0|V^* \leq \overline{p}\Big) = \pi^*\Big(y^0|V < \overline{p}\Big)$. But $\forall z \in \mathcal{Z}$ we have that $\pi\Big(y^z, V < \overline{p}\Big) = \pi^*\Big(y^z, V^* \leq \overline{p}\Big)$ and $\pi\Big(V < \overline{p}\Big) = \pi^*\Big(V^* \leq \overline{p}\Big)$, where $\overline{p} = \min\{p(w_{\max}, 0), p(w_{\max}, 1)\}$, by summing and integrating from equation A.2. ■

Proof of Proposition 3.

Part "($\implies$)":

First we will show that given $\pi \in \Phi_r$ we can find π^* that satisfies the model restrictions Φ_r^* and that ATE_c and ATE_c^* are equal.

For $\forall y^1 \times y^0 \in \mathcal{Y} \times \mathcal{Y}, \forall k \in \{1, \ldots, n\}$, and for an ordered sequence $\Big(v_k^*\Big)_{k=1}^n$ with $v_k^* \in \mathcal{V}^*$ and $v_0^* = 0$, we define

$$\alpha^*_{y^1 y^0 k} \equiv \int_{v^*_{k-1}}^{v^*_k} \alpha_{y^1 y^0}(v) \mathrm{d}v. \tag{A.3}$$

(rC3) $\implies$ (rD3)
From definition of $\mathcal{V}^*$ integrating (rC3) using basic integral properties we know that $\forall y^1 \times y^0 \times v \in \mathcal{Y} \times \mathcal{Y} \times \mathcal{V}, \forall k \in \{1, \ldots, n\}, \forall v \in \langle v^*_{k-1}, v^*_k) :$

$$\Big|\pi_{Y^1,Y^0,V|Z}(Y^1 = y^1, Y^0 = y^0, V = v|Z = 1) - \pi_{Y^1,Y^0,V|Z}(Y^1 = y^1, Y^0 = y^0, V = v|Z = 0)\Big| \leq \alpha_{y^1 y^0}(v),$$

$$\Bigg|\int_{v^*_{k-1}}^{v^*_k} \pi_{Y^1,Y^0,V,Z}(Y^1 = y^1, Y^0 = y^0, V = v, Z = 1)\mathrm{d}v / \pi_Z(Z = 1) -$$

$$- \int_{v^*_{k-1}}^{v^*_k} \pi_{Y^1,Y^0,V,Z}(Y^1 = y^1, Y^0 = y^0, V = v, Z = 0)\mathrm{d}v / \pi_Z(Z = 0)\Bigg| \leq \int_{v^*_{k-1}}^{v^*_k} \alpha_{y^1 y^0}(v)\mathrm{d}v.$$

Now summing using equation A.1 and A.3 we get:

$$\Big|\pi^*_{Y^1,Y^0,V^*|Z}(Y^1 = y^1, Y^0 = y^0, V^* = v^*_k|Z = 1) - \pi^*_{Y^1,Y^0,V^*|Z}(Y^1 = y^1, Y^0 = y^0, V^* = v^*_k|Z = 0)\Big| \leq \alpha^*_{y^1 y^0 k}.$$

By equation A.3 and by using basic integral properties we see that:

$$\sum_{y^1, y^0 \in \mathcal{Y}} \sum_{k=1}^{n} \alpha^*_{y^1 y^0 k} = \sum_{y^1, y^0 \in \mathcal{Y}} \int_0^1 \alpha_{y^1 y^0}(v)\mathrm{d}v = \alpha_E,$$

$$\text{and } \forall y^1 \times y^0 \times v \in \mathcal{Y} \times \mathcal{Y} \times \langle v^*_{k-1}, v^*_k), \forall k \in \{1, \dots, n\}$$

$$\alpha_{y^1 y^0}(v) \geq 0,$$

$$\int_{v^*_{k-1}}^{v^*_k} \alpha_{y^1 y^0}(v) \mathrm{d}v \geq 0,$$

$$\alpha^*_{y^1 y^0 k} \geq 0.$$

Part "($\Longleftarrow$)":

Given $\pi^* \in \Phi^*_r$ we can find π that satisfies the model restrictions Φ_r and that ATE^*_c and ATE_c are equal.

For $\forall y^1 \times y^0 \in \mathcal{Y} \times \mathcal{Y}, \forall k \in \{1, \dots, n\}$, and for an ordered sequence $\left(v^*_k\right)_{k=1}^n$ with $v^*_k \in \mathcal{V}^*$ and $v^*_0 = 0$, we define

$$\alpha_{y^1 y^0}(v) \equiv \frac{\alpha^*_{y^1 y^0 k}}{\pi^{*,V^*}_k}. \tag{A.4}$$

(rD3) $\implies$ (rC3)

From definition of V^*, (rD3), A.4 and by summing in A.2 we get,
$\forall z \in \mathcal{Z}, \forall v \in \mathcal{V}, \exists k \in \{1, \dots, n\}, v \in \langle v^*_{k-1}, v^*_k) :$

$$\Big|\pi^*_{Y^1,Y^0,V^*|Z}(Y^1 = y^1, Y^0 = y^0, V^* = v^*_k | Z = 1)$$

$$-\pi^*_{Y^1,Y^0,V^*|Z}(Y^1 = y^1, Y^0 = y^0, V^* = v^*_k | Z = 0)\Big| \leq \alpha^*_{y^1 y^0 k},$$

$$\left|\pi_{Y^1,Y^0,V|Z}(y^1, y^0, v|1) \cdot \pi^{*,V^*}_k - \pi_{Y^1,Y^0,V|Z}(y^1, y^0, v|0) \cdot \pi^{*,V^*}_k\right| \leq \alpha_{y^1 y^0}(v) \cdot \pi^{*,V^*}_k,$$

By equation A.3, (D5) and by using basic integral properties we see that:

$$\alpha_{y^1 y^0}(v) = \frac{\alpha^*_{y^1 y^0 k}}{\pi^{*,V^*}_k} \geq 0,$$

$$\int_{v^*_{k-1}}^{v^*_k} \alpha_{y^1 y^0}(v) \mathrm{d}v = \int_{v^*_{k-1}}^{v^*_k} \frac{\alpha^*_{y^1 y^0 k}}{\pi^{*,V^*}_k} \mathrm{d}v = \alpha^*_{y^1 y^0 k},$$

$$\sum_{y^1, y^0 \in \mathcal{Y}} \int_0^1 \alpha_{y^1 y^0}(v) \mathrm{d}v = \sum_{y^1, y^0 \in \mathcal{Y}} \sum_{k=1}^n \alpha^*_{y^1 y^0 k} = \alpha_E.$$

Validity of the other conditions and equality ATE's is proved by the same argument as previous proposition as this proof doesn't involve treatment independence. ■

References

Balke, A. and J. Pearl (1997): "Bounds on treatment effects from studies with imperfect compliance," *Journal of the American Statistical Association*, 92, 1171–1176.

Behaghel, L., B. Crépon, M. Gurgand, and T. Le Barbanchon (2015): "Please Call Again: Correcting Nonresponse Bias in Treatment Effect Models," *Review of Economics and Statistics*, 97, 1070–1080.

Berk, R. A., G. K. Smyth, and L. W. Sherman (1988): "When random assignment fails: Some lessons from the Minneapolis spouse abuse experiment," *Journal of Quantitative Criminology*, 4, 209–223.

Freyberger, J. and J. L. Horowitz (2015): "Identification and shape restrictions in nonparametric instrumental variables estimation," *Journal of Econometrics*, 189, 41–53.

Heckman, J. J. (1977): *Sample selection bias as a specification error (with an application to the estimation of labor supply functions)*, National Bureau of Economic Research Cambridge, Mass., USA.

Honoré, B. E. and E. Tamer (2006): "Bounds on parameters in panel dynamic discrete choice models," *Econometrica*, 74, 611–629.

Lafférs, L. (2018): "Bounding average treatment effects using linear programming," *Empirical Economics*.

——— (2019): "Identification in models with discrete variables," *Computational Economics*, 53, 657–698.

Lafférs, L. and R. Nedela Jr (2017): "Sensitivity of the bounds on the ATE in the presence of sample selection," *Economics Letters*, 158, 84–87.

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

Manski, C. F. (2007): "Partial identification of counterfactual choice probabilities," *International Economic Review*, 48, 1393–1410.

Molinari, F. (2008): "Partial identification of probability distributions with misclassified data," *Journal of Econometrics*, 144, 81–117.

Politis, D. N., J. P. Romano, and M. Wolf (1999): *Subsampling*, Springer Science & Business Media.

Torgovitsky, A. (2016): "Nonparametric Inference on State Dependence with Applications to Employment Dynamics," *Available at SSRN 2564305*.

Vytlacil, E. (2002): "Independence, monotonicity, and latent index models: An equivalence result," *Econometrica*, 70, 331–341.