University of Zurich

Socioeconomic Institute
Sozialökonomisches Institut

Working Paper No. 0709

Nonparametric Analysis of Treatment Effects in Ordered Response Models

Stefan Boes

July 2007

Socioeconomic Institute
University of Zurich

Working Paper No. 0709

Nonparametric Analysis of Treatment Effects in Ordered Response Models

July 2007

Authors' addresses: Stefan Boes
E-mail: boes@sts.uzh.ch

Publisher Sozialökonomisches Institut
Bibliothek (Working Paper)
Rämistrasse 71
CH-8006 Zürich
Phone: +41-44-634 21 37
Fax: +41-44-634 49 82
URL: www.soi.uzh.ch
E-mail: soilib@soi.uzh.ch

Nonparametric Analysis of Treatment Effects in Ordered Response Models

Stefan Boes*
University of Zurich

July 25, 2007

Abstract

This paper deals with the identification of treatment effects when the outcome variable is ordered. If outcomes are measured ordinally, previously developed methods to investigate the impact of an endogenous binary regressor on average outcomes cannot be applied as the expectation of an ordered variable, in its strict sense, does not exist, and a shift in focus to distributional effects is indispensable. Without imposing a fully fledged parametric model the treatment effects are generally not point-identified. Assuming a threshold crossing model on both the ordered potential outcomes and the binary treatment variable leaving the distribution of error terms and functional forms unspecified, it is discussed how the treatment effects can be bounded and inference on the bounds can be conducted.

JEL Classification: C14, C25, C35

Keywords: Nonparametric bounds, causal effects, potential outcomes, latent variables.

*Address for correspondence: University of Zurich, Socioeconomic Institute, Zuerichbergstrasse 14, CH-8032 Zurich, Switzerland, phone: +41 44 634 2301, email: boes@sts.uzh.ch.

1 Introduction

Suppose one is interested in the effect of a binary treatment D on an ordered response Y. The treatment variable is such that $D = 1$ whenever the treatment is received, and $D = 0$ otherwise. It is often useful to think of D as a dummy endogenous variable in the model for Y, provided that the treatment status is determined by self-selected individuals rather than randomly assigned treatment groups. In terms of potential outcomes (Neyman 1923, Rubin 1974), let Y_1 denote the potential outcome with treatment, and let Y_0 denote the potential outcome without treatment. The measured outcome Y is related to potential outcomes (Y_1, Y_0) so that

$$Y = DY_1 + (1 - D)Y_0 \tag{1}$$

Assume that the total number of categories is independent of the treatment status, i.e., irrespective of being treated or not the individual will face the same set of mutually exclusive and exhaustive ordered categories. Without loss of generality, let $\mathcal{Y} = \{1, 2, \ldots, J\}$ denote the set of possible outcomes of Y, where "1" < "2" < ... < "J". The assigned values in $\mathcal{Y}$ are entirely meaningless, as long as they keep the ordering, and are just for notational convenience.

One can think of a number of applications with a binary treatment and an ordered response. For example, in medical research the effectiveness of a new drug may be evaluated regarding the patient's health status, in educational economics one may be interested in the effect of out-of-school training programs on exam grades, and in labor economics the sorting of workers into public and private sector jobs may be analyzed with respect to their economic performance, the latter measured as promotion, lateral move, or demotion.

The ordinal nature of Y needs to be taken into account when defining treatment effects. With quantitative and binary outcomes, the individual treatment effect $Y_1 - Y_0$ has potential interest. For example, if Y measures wages and D is participation in job training, then $Y_1 - Y_0$ gives the wage difference with and without the training program. If Y indicates, for example, one-year survival after cardiac surgery and D indicates medical treatment, then $Y_1 - Y_0$ shows whether the individual would survive with medication and die without (1), is not affected by medication (0), or would survive without and die with medication (-1). For ordinal variables such an interpretation does not exist because the distance between outcomes is not defined.

In practice, only one of two potential outcomes Y_1 or Y_0 can be observed because each individual either receives the treatment, or does not. Thus, it is impossible to recover the individual treatment effect and the literature typically focuses on averages of $Y_1 - Y_0$, such as the average treatment effect, $E(Y_1 - Y_0)$, or the average treatment effect on the treated, $E(Y_1 - Y_0|D = 1)$. Under certain assumptions, these parameters can at least partly be recovered from observed data (Heckman and Robb 1985, 1986, Manski 1990, 1994, 1995, Imbens and Angrist 1994, Angrist *et al.* 1996, Heckman *et al.* 1999, among many others). With ordinal data, again, the case is different: Any rank preserving recoding of the elements in $\mathcal{Y}$ should not affect the parameters of interest. $E(Y_1)$ and $E(Y_0)$, however, will be affected by such a value conversion, so that the concept of averages needs to be replaced by a concept insensitive to the definition of $\mathcal{Y}$.

For these reasons I propose to analyze treatment effects for ordinal outcomes in terms of probabilities rather than expectations. Investigating treatment effects in terms of probabilities is particularly attractive for discrete responses as each outcome occurs with a positive probability, and analyzing probability effects is thus of interest on its own. Let the "average" treatment effect (ATE) be defined as the probability difference of observing a particular outcome with and without the treatment, formally

$$\Delta_y^{ATE} \equiv P(Y_1 = y) - P(Y_0 = y) \qquad y = 1, \ldots, J \tag{2}$$

Note that there are indeed J effects, one for each outcome of Y. If the treatment affects responses positively — adopting the convention that higher outcomes of Y are in some way "better" —, then one would expect Δ_y^{ATE} negative for low y and positive for high y. In practice, there may not exist such a clear systematic indicating whether the treatment has a positive or a negative effect, but the shift in focus to probability effects allows for a detailed analysis of the effects of the treatment in all parts of the outcome distribution.

Analogously, the effect on outcome probabilities for individuals who actually received the treatment can be defined as treatment on the treated parameter (TT)

$$\Delta_y^{TT} \equiv P(Y_1 = y|D = 1) - P(Y_0 = y|D = 1) \qquad y = 1, \ldots, J \tag{3}$$

Both treatment parameters are robust against the particular values assigned to outcomes, but rely on the "same scale" assumption. Yet this assumption is not overly restrictive, as otherwise it would be difficult to compare the Y_1 and the Y_0 distribution. One may also define other treatment effect parameters in terms of probabilities rather than expectations, such as the local average treatment effect (LATE) of Imbens and Angrist (1994), or the marginal treatment effect (MTE) of Heckman (1997). In this paper, I will confine myself on the parameters in (2) and (3), but some remarks on other parameters will be given below.

Δ_y^{ATE} and Δ_y^{TT} are not immediately identified from the population distribution of (Y, D). To see why, consider the average treatment effect and $P(Y_1)$.[1] By the law of total probability,

$$P(Y_1) = P(Y_1|D = 1)P(D = 1) + P(Y_1|D = 0)P(D = 0)$$

The sampling process identifies the probability of treatment selection, $P(D = 1)$, and the outcome probability with treatment given treatment has been received, $P(Y_1|D = 1) = P(Y|D = 1)$. The sampling process is uninformative, however, regarding $P(Y_1|D = 0)$, which is the outcome probability with treatment, given the treatment has not been received. In the common terminology such a term is referred to as counterfactual probability. $P(Y_0)$ is not identified either, because the sampling process does not reveal $P(Y_0|D = 1)$, and therefore the average treatment effect is not identified. Lack of observability of the counterfactual $P(Y_0|D = 1)$ also makes identification of the treatment on the treated parameter fail.

The aim of this paper is to find reasonable bounds on counterfactual probabilities in a setting with ordinal outcomes and binary treatment, and thus to bound the treatment effect parameters. As a starting point and without imposing any assumptions on the data-generating process, it must certainly hold that both counterfactuals, $P(Y_1|D = 0)$ and $P(Y_0|D = 1)$, are bounded by zero and one. The average treatment effect is thus bounded by

$$\Delta_y^{ATE} \in \left[\mathcal{LB}1_y^{ATE}, \mathcal{UB}1_y^{ATE}\right] \qquad \text{with} \tag{4}$$

$$\begin{aligned} \mathcal{LB}1_y^{ATE} &= P(D = 1, Y = y) - P(D = 1) - P(D = 0, Y = y) \\ \mathcal{UB}1_y^{ATE} &= P(D = 1, Y = y) + P(D = 0) - P(D = 0, Y = y) \end{aligned}$$

[1] From now on, I will drop the y argument in the probability statements if possible to save some notation, e.g., $P(Y_1)$ will be shorthand notation for $P(Y_1 = y)$, or $P(Y_1|D = 1)$ will be shorthand for $P(Y_1 = y|D = 1)$. If not mentioned otherwise, the equations will hold for all $y = 1, \dots, J$.

Analogously, the average treatment effect on the treated is restricted to the interval

$$\Delta_y^{TT} \in \left[L\mathcal{B}1_y^{TT}, U\mathcal{B}1_y^{TT}\right] \qquad \text{with} \tag{5}$$

$$L\mathcal{B}1_y^{TT} = P(Y = y|D = 1) - 1, \qquad U\mathcal{B}1_y^{TT} = P(Y = y|D = 1)$$

The intervals in (4) and (5) define identification regions for the treatment parameters since all valid probability distributions $P(Y_1|D = 0)$ and $P(Y_0|D = 1)$ necessarily yield treatment effects within the stated bounds (Manski 2000, 2003). Note that the width of the regions is one, which is the logical maximum for a probability effect. Note too that the bounds are not informative regarding the sign of both treatment parameters as zero is included in the range of possible values. The question to be investigated in the following sections is how further assumptions on the sampling process do narrow these bounds.

More specifically, the paper will explore a nonparametric threshold crossing model on both the ordered potential outcomes and the treatment selection. Ordinal data modeling is traditionally based on latent variables and threshold crossing mechanisms. For example, parametric models like the ordered probit and the ordered logit model follow this structure (McKelvey and Zavoina 1975, McCullagh 1980), but also semiparametric approaches like Klein and Sherman (2002), Bellemare *et al.* (2002), Coppejans (2007), Lewbel (1997, 2003), and Stewart (2004) impose a threshold crossing model to generate ordinality in the response variable. It therefore seems natural to analyze the implications of such a model structure in a nonparametric bounding analysis. The model is nonparametric in the sense that no distributional assumptions, and no functional form assumptions will be imposed other than the threshold mechanism.

Three recent papers are related to mine. First, Shaikh and Vytlacil (2005) discuss treatment effect bounds with a binary response variable and a binary treatment. They impose nonparametric threshold crossing models on both the treatment selection and the binary potential outcomes, whereas the model here assumes ordinal potential outcomes. As it will be worked out below, this requires a slightly different bounding strategy, and supplemental interpretations can be given in the extended setting. Second, Scharfstein *et al.* (2004) analyze bounds on the distribution of ordinal outcomes, but their model setup is different from mine because they consider two outcome variables where the first is always observed and the second (sequentially following the first) is potentially missing so that the joint distribution of the two outcomes is not identified. Third, Li and Tobias (2007) describe Bayesian estimation of treatment effects for ordinal outcomes. They impose more structure on the model than it is imposed here and focus on mean treatment effect parameters (and therefore require additional implicit assumptions on the kind of ordinal response variable that is analyzed).

2 Model and Assumptions

The model for the treatment status and the potential outcomes is a version of the model in Shaikh and Vytlacil (2005) generalized to the case of ordinal outcomes and defined as

$$\begin{aligned} D^* &= s(Z) - \nu \qquad D = \mathbf{1}(D^* \geq 0) \\ Y_0^* &= r_0(X) + \varepsilon_0 \qquad Y_0 = \sum_{y=1}^{J} y\mathbf{1}(\kappa_{0y-1} < Y_0^* \leq \kappa_{0y}) \\ Y_1^* &= r_1(X) + \varepsilon_1 \qquad Y_1 = \sum_{y=1}^{J} y\mathbf{1}(\kappa_{1y-1} < Y_1^* \leq \kappa_{1y}) \end{aligned} \tag{6}$$

where (X, Z) is a random vector of observed covariates, ν, ε_0, and ε_1 are unobserved random variables, and $\mathbf{1}(\cdot)$ is the logical indicator function. The model is a latent index model with latent variables D^*, Y_0^*, and Y_1^*, and a threshold crossing mechanism that generates the treatment status D and the potential outcomes Y_0 and Y_1. The model is nonparametric in the sense that the functional forms of $s(Z)$, $r_0(X)$, and $r_1(X)$ are left unspecified and no parametric assumption on the distribution of $(\varepsilon_0, \varepsilon_1, \nu)$ is made. The model presumes that the error terms and the functions of observable factors are additively separable; see Vytlacil (2002, 2006) for a discussion of this property in latent index threshold crossing models. Finally, the observed outcome Y is generated according to (1), completing the model.

The definition of treatment parameters and the identification regions stated in the introduction still hold conditional on the vector of observed covariates X. In this case, the average treatment effect and the average treatment effect on the treated are local (conditional on X), and unconditional treatment effects may be obtained as weighted averages. The model as presented above also includes a vector Z that affects the treatment selection. Z may contain all elements of X, and additional elements in Z will generally be referred to as instrumental variables. X may or may not contain an element that is not included in Z. If such an element exists, then this information can be gainfully employed in the bounding analysis. Let $\mathcal{X}$ denote the support of the random vector X, and let $\mathcal{Z}$ denote the support of the random vector Z.

The assumptions imposed on the model (extending Shaikh and Vytlacil 2005) are:

(A1) The threshold parameters κ_{0j}, κ_{1j}, $j = 0, \ldots, J$ are fixed and fulfill the order requirement $-\infty = \kappa_{00} < \kappa_{01} < \ldots < \kappa_{0J} = \infty$, and $-\infty = \kappa_{10} < \kappa_{11} < \ldots < \kappa_{1J} = \infty$.

(A2) For some $x_0 \in \mathcal{X}$ let $r_0(x_0) = 0$, and for some $x_1 \in \mathcal{X}$ let $r_1(x_1) = 0$.

(A3) The distribution of ν is absolutely continuous with respect to Lebesgue measure.

(A4) $(\varepsilon_0, \varepsilon_1, \nu) \perp\!\!\!\perp (Z, X)$.

(A5) $\varepsilon_j | \nu \sim \varepsilon | \nu$, $j = 0, 1$.

(A6) The distribution of $\varepsilon_j | \nu$ has strictly positive density w.r.t. Lebesque measure on $\mathbf{R}$, $j = 0, 1$.

(A7) $s(Z)$ is non-degenerate conditional on X.

(A8) The support of the distribution of (X, Z) is compact, and $r_0(\cdot)$, $r_1(\cdot)$, $s(\cdot)$ are continuous.

For a detailed discussion of assumptions (A3) to (A8) see Shaikh and Vytlacil (2005). Crucial in the following analysis are the independence assumption (A4) and the restriction to equal distributions of ε_1 and ε_0 conditional on ν (A5). The additional assumptions (A1) and (A2) are imposed due to the ordinal nature of Y. (A1) in combination with the model equation explicitly accounts for the order information. The threshold parameters are assumed to be unknown, although the extension to known thresholds (interval data) is possible. In the latter case, knowledge of thresholds in both treatment statuses is required, unless they are independent of treatment and thus equal. Knowledge of κ_0 and κ_1 will considerably simplify the analysis, and remarks will be given at the appropriate places when the additional information can be used.

The model allows for much flexibility in the threshold mechanism since no distributional or functional form assumptions are imposed and the (unknown) threshold parameters are allowed to vary by the treatment status. In particular, the model does not restrict the shape of treatment effects in a way similar to the single crossing property of probability effects in standard parametric ordered probit and logit models (Boes and Winkelmann 2006), nor does it require a specific model for the threshold parameters in order to relax this property.

Assumption (A2) is an identifying assumption that simplifies exposition and is standard in parametric models. If (A2) is not met, then parametric ordinal response models may only identify location-normalized instead of absolute threshold parameters, i.e., κ_0, κ_1 will be replaced by $\kappa_0 - r_0$ and $\kappa_1 - r_1$, respectively, where r_0, r_1 denote the constant terms in $r_0(X), r_1(X)$. As it is irrelevant for the following analysis if all thresholds are shifted equally to the right or to the left, (A2) is purely simplifying and does not restrict the analysis in any way.

3 Bounds on Treatment Effects

For the ease of exposition, I will first consider bounds on the treatment effect parameters when no X covariates are available (Sections 3.1 and 3.2). In this case, the latent potential outcome equations of the model simplify to $Y_1^* = \varepsilon_1$ and $Y_0^* = \varepsilon_0$. The extension to the case when X covariates are present will be separately discussed below (Section 3.3).

3.1 Bounds under the Independence Assumption

The first bounding strategy follows Manski (1990, 1994). Assume that potential outcomes (Y_0, Y_1) are independent of Z, but that treatment selection D varies with Z. One may interpret such a condition as exclusion restriction, and Z is an instrumental variable. It is easy to verify that the model assumptions in Section 2 imply this condition but not vice versa, i.e., the assumptions imposed by the model are stronger than the exclusion restriction alone. Given independence, it must hold that $P(Y_1|Z = z) = P(Y_1)$ for all $z \in \mathcal{Z}$.[2] Moreover, write

$$P(Y_1|Z) = P(Y_1|D = 1, Z)P(D = 1|Z) + P(Y_1|D = 0, Z)P(D = 0|Z)$$

[2] In order to save some notation, I will drop the particular value z (or later on x) that is conditioned on if it is not critical in the given context. It will be implicitly assumed that all expressions are only evaluated over the appropriate support, i.e., at all evaluation points the conditional probabilities exist and are well-defined.

In this expression, all probabilities but the counterfactual $P(Y_1|D=0,Z)$ are identified from the population distribution (Y,D,Z). The unidentified probability is bounded by zero and one which in turn imposes upper and lower bounds on $P(Y_1|Z)$. Due to the independence assumption, the smallest of $P(D=1,Y|Z=z)+P(D=0|Z=z)$ — which is the upper bound of $P(Y_1|Z=z)$ — over all $z \in \mathcal{Z}$ may be used as a new upper bound for $P(Y_1)$, and the largest of $P(D=1,Y|Z=z)$ — which is the lower bound of $P(Y_1|Z=z)$ — over all $z \in \mathcal{Z}$ may be used as a new lower bound for $P(Y_1)$. Analogously, new upper and lower bounds for $P(Y_0)$ may be obtained and the average treatment parameter can be bounded by

$$\Delta_y^{ATE} \in \left[L\mathcal{B}2_y^{ATE}, U\mathcal{B}2_y^{ATE}\right] \qquad \text{with} \tag{7}$$

$$\begin{aligned} L\mathcal{B}2_y^{ATE} &= \sup_{z\in\mathcal{Z}}\{P(D=1,Y=y|Z=z)\} \\ &\quad -\inf_{z\in\mathcal{Z}}\{P(D=1|Z=z)+P(D=0,Y=y|Z=z)\} \\ U\mathcal{B}2_y^{ATE} &= \inf_{z\in\mathcal{Z}}\{P(D=1,Y=y|Z=z)+P(D=0|Z=z)\} \\ &\quad -\sup_{z\in\mathcal{Z}}\{P(D=0,Y=y|Z=z)\} \end{aligned}$$

where $\sup\{\cdot\}$ denotes the supremum and $\inf\{\cdot\}$ the infimum of the argument in curly brackets over the values indicated in the subscript.

For the treatment on the treated effect note that in general $P(Y_0|D=1,Z) \neq P(Y_0|D=1)$, i.e., $Y_0|D=1$ is not independent of Z, as the instrument does affect the treatment status. One option to proceed would be to re-define the treatment on the treated parameter conditional on Z, or conditional on $P(D=1|Z)$, and then obtain the unconditional parameter by integration. I follow an alternative strategy and rewrite the counterfactual $P(Y_0|D=1)$ in terms of an identified probability and a probability that can be bounded under independence. It must hold that

$$\begin{aligned} P(Y_0=y|D=1) &= P(D=1,Y_0=y)/P(D=1) \\ &= [P(Y_0=y)-P(D=0,Y_0=y)]/P(D=1) \end{aligned}$$

by Bayes' theorem and the law of total probability. The sampling process identifies $P(D=1)$ and $P(D=0,Y_0)=P(D=0,Y)$, but only partially identifies $P(Y_0)$. Given $Y_0 \perp\!\!\!\perp Z$, one may construct upper and lower bounds on $P(Y_0)$ in the same manner as above. Rewrite the treatment on the treated parameter as

$$\begin{aligned}\Delta_y^{TT} &= [P(D=1,Y_1=y)-P(D=1,Y_0=y)]/P(D=1)\\ &= [P(D=1,Y=y)-P(Y_0=y)+P(D=0,Y=y)]/P(D=1)\\ &= [P(Y=y)-P(Y_0=y)]/P(D=1) \qquad (8)\end{aligned}$$

so that

$$\Delta_y^{TT} \in \left[L\mathcal{B}2_y^{TT}, U\mathcal{B}2_y^{TT}\right] \qquad \text{with} \qquad (9)$$

$$\begin{aligned}L\mathcal{B}2_y^{TT} &= \left[P(Y=y)-\inf_{z\in\mathcal{Z}}\{P(D=1|Z=z)+P(D=0,Y=y|Z=z)\}\right]/P(D=1)\\ U\mathcal{B}2_y^{TT} &= \left[P(Y=y)-\sup_{z\in\mathcal{Z}}\{P(D=0,Y=y|Z=z)\}\right]/P(D=1)\end{aligned}$$

Note that the bounds in (7) and (9) do not exploit the ordinal nature of the response variable, nor do they exploit the threshold crossing structure of the model. The analysis may therefore be applied to any nominal response Y and binary treatment D. The question to be investigated in the following is how such additional assumptions on the structure of the data can be used to improve upon (7) and (9).

3.2 Bounds Under the Threshold Crossing Model Structure

The bounding strategy of this section generalizes Heckman and Vytlacil (2001) and Shaikh and Vytlacil (2005) to the case of ordinal potential outcomes. Given the threshold crossing structure of the treatment selection equation and the independence assumption, it follows that for any two evaluation points $z_1, z_0 \in \mathcal{Z}$

$$\begin{aligned}P(D=1|Z=z_1) > P(D=1|Z=z_0) &\Leftrightarrow P(s(z_1)\geq\nu) > P(s(z_0)\geq\nu)\\ &\Leftrightarrow s(z_1) > s(z_0)\end{aligned}$$

Furthermore, let

$$\begin{aligned}z^u &= \arg\sup_{z\in\mathcal{Z}} P(D=1|Z=z)\\ z^l &= \arg\inf_{z\in\mathcal{Z}} P(D=1|Z=z)\end{aligned}$$

This information can be used in two ways. First, by definition of z^u and z^l it must hold that $s(z^u) \geq s(z)$ and $s(z^l) \leq s(z)$ for all $z \in \mathcal{Z}$. The following lemma then simplifies the supremum and infimum expressions in the bounds on the average treatment and the treatment on the treated parameters as stated in (7) and (9):

Lemma 1 *Assume that* (Y_0, Y_1, D) *are generated according to model (6), and assume that conditions (A1)-(A4) and (A7)-(A8) are fulfilled. Then,*

$$
\begin{aligned}
(a) \quad & \sup_{z\in\mathcal{Z}}\{P(D=1,Y=y|Z=z)\} = P(D=1,Y=y|Z=z^u) \\
(b) \quad & \sup_{z\in\mathcal{Z}}\{P(D=0,Y=y|Z=z)\} = P(D=0,Y=y|Z=z^l)
\end{aligned}
$$

$$
\begin{aligned}
(c) \quad & \inf_{z\in\mathcal{Z}}\{P(D=1,Y=y|Z=z)+P(D=0|Z=z)\} \\
& \qquad = P(D=1,Y=y|Z=z^u)+P(D=0|Z=z^u)
\end{aligned}
$$

$$
\begin{aligned}
(d) \quad & \inf_{z\in\mathcal{Z}}\{P(D=1|Z=z)+P(D=0,Y=y|Z=z)\} \\
& \qquad = P(D=1|Z=z^l)+P(D=0,Y=y|Z=z^l)
\end{aligned}
$$

Proof. First consider part (a) of the lemma. Recall that at all evaluation points the conditional probabilities exist and are well-defined. The assumptions of the lemma ensure that

$$
\begin{aligned}
& P(D=1,Y|Z=z^u) - P(D=1,Y|Z=z) \qquad (10) \\
= \;& P(\nu \leq s(z^u), Y_1) - P(\nu \leq s(z), Y_1) \\
= \;& P(s(z) < \nu \leq s(z^u), Y_1) \geq 0
\end{aligned}
$$

where the weak inequality follows by definition of z^u. The supremum of $P(D=1,Y=y|Z=z)$ over z is equivalent to the infimum of (10) over z. As (10) must be non-negative, necessary and sufficient condition for an infimum of (10) is that $z = z^u$. Analogously,

$$
\begin{aligned}
& P(D=0,Y|Z=z^l) - P(D=0,Y|Z=z) \qquad (11) \\
= \;& P(\nu > s(z^l), Y_0) - P(\nu > s(z), Y_0) \\
= \;& P(\nu \leq s(z), Y_0) - P(\nu \leq s(z^l), Y_0) \\
= \;& P(s(z^l) < \nu \leq s(z), Y_0) \geq 0
\end{aligned}
$$

where the weak inequality follows by definition of z^l. The supremum in part (b) of the lemma is equivalent to the infimum of (11) over z, and, by the assumptions of the model and given the weak inequality, $z = z^l$ is necessary and sufficient for a supremum of $P(D = 0, Y = y|Z = z)$. In order to show part (c) of the lemma, write

$$\begin{aligned} & P(D=1,Y|Z=z^u)+P(D=0|Z=z^u)-P(D=1,Y|Z=z)-P(D=0|Z=z) \qquad (12) \\ = \;& P(D=1,Y|Z=z^u)-P(D=1,Y|Z=z) \\ & -[P(D=1|Z=z^u)-P(D=1|Z=z)] \\ = \;& P(s(z)<\nu\leq s(z^u),Y_1)-P(s(z)<\nu\leq s(z^u))\leq 0 \end{aligned}$$

where the weak inequality follows by definition of z^u and the law of total probability. The infimum of $P(D = 1, Y = y|Z = z) + P(D = 0|Z = z)$ is equivalent to the supremum of (12) both over z. As (12) must be non-positive, necessary and sufficient condition for a supremum of (12) is that $z = z^u$. Analogous arguments prove part (d) of the lemma. □

A direct implication of Lemma 1 is that the bounds on the average treatment and the treatment on the treated parameters as stated in Section 3.1 simplify to

$$\Delta_y^{ATE} \in \left[L\mathcal{B}3_y^{ATE}, U\mathcal{B}3_y^{ATE}\right] \quad \text{with} \qquad (13)$$

$$\begin{aligned} L\mathcal{B}3_y^{ATE} &= P(D=1,Y=y|Z=z^u)-P(D=1|Z=z^l)-P(D=0,Y=y|Z=z^l) \\ U\mathcal{B}3_y^{ATE} &= P(D=1,Y=y|Z=z^u)+P(D=0|Z=z^u)-P(D=0,Y=y|Z=z^l) \end{aligned}$$

and

$$\Delta_y^{TT} \in \left[L\mathcal{B}3_y^{TT}, U\mathcal{B}3_y^{TT}\right] \quad \text{with} \qquad (14)$$

$$\begin{aligned} L\mathcal{B}3_y^{TT} &= [P(Y=y)-P(D=1|Z=z^l)-P(D=0,Y=y|Z=z^l)]/P(D=1) \\ U\mathcal{B}3_y^{TT} &= [P(Y=y)-P(D=0,Y=y|Z=z^l)]/P(D=1) \end{aligned}$$

Compared to the bounds in (7) and (9), the bounds in (13) and (14) can be readily evaluated once z^u and z^l are determined. It is also possible to calculate their width; for the average treatment effect the width is given by $P(D = 0|Z = z^u) + P(D = 1|Z = z^l)$, and for the treatment on the treated parameter the width is given by $P(D = 1|Z = z^l)/P(D = 1)$. Both are smaller than one given that treatment selection varies with Z, i.e., for both treatment parameters the independence assumption together with the threshold crossing treatment selection is informative and yields narrower bounds than the identification regions stated in the introduction. Note however that the bounds in (13) and (14) do not yield tighter bounds than those in (7) and (9), because the former are simply a special case of the latter, but the imposed model structure considerably simplifies the form and the calculation of the bounds.

The second implication of the threshold crossing treatment selection can be derived in combination with the threshold model for the potential outcomes. Let

$$sgn(a) = \begin{cases} -1 & \text{if } a < 0 \\ 0 & \text{if } a = 0 \\ 1 & \text{if } a > 0 \end{cases}$$

denote the sign function, and consider the following lemma:

Lemma 2 *Assume that (Y_0, Y_1, D) are generated according to model (6), and assume that conditions (A1)-(A8) are fulfilled. Then for any two evaluation points z_1, z_0 with $P(D = 1|Z = z_1) > P(D = 1|Z = z_0)$,*

$$sgn[P(Y \leq y|Z = z_1) - P(Y \leq y|Z = z_0)] = sgn(\kappa_{1y} - \kappa_{0y}) \equiv \delta_y$$

so that δ_y can take three values -1,0,1 depending on whether the difference $\kappa_{1y} - \kappa_{0y}$ is negative, zero, or positive, respectively.

Proof. Consider the cumulative outcome probability conditional on the instrument

$$\begin{aligned} P(Y \leq y|Z) &= P(D = 1, Y \leq y|Z) + P(D = 0, Y \leq y|Z) \\ &= P(D = 1, Y_1 \leq y|Z) + P(D = 0, Y_0 \leq y|Z) \\ &= P(\nu \leq s(z), \varepsilon_1 \leq \kappa_{1y}) + P(\nu > s(z), \varepsilon_0 \leq \kappa_{0y}) \\ &= P(\nu \leq s(z), \varepsilon \leq \kappa_{1y}) + P(\nu > s(z), \varepsilon \leq \kappa_{0y}) \end{aligned}$$

where the first equality follows by the law of total probability, the second equality follows by (1), the third equality follows by the model and the independence assumption, and the last equality follows by assumption (A5). Now take the difference of the cumulative outcome probabilities evaluated at any two evaluation points z_1, z_0 with $P(D = 1|Z = z_1) > P(D = 1|Z = z_0)$ such that $s(z_1) > s(z_0)$. Then,

$$\begin{aligned}
&P(Y \leq y|Z = z_1) - P(Y \leq y|Z = z_0) \\
&= P(s(z_0) < \nu \leq s(z_1), \varepsilon \leq \kappa_{1y}) - P(s(z_0) < \nu \leq s(z_1), \varepsilon \leq \kappa_{0y}) \\
&= \begin{cases} P(s(z_0) < \nu \leq s(z_1), \kappa_{0y} < \varepsilon \leq \kappa_{1y}) & \text{iff} \quad \kappa_{1y} > \kappa_{0y} \\ 0 & \text{iff} \quad \kappa_{1y} = \kappa_{0y} \\ -P(s(z_0) < \nu \leq s(z_1), \kappa_{1y} < \varepsilon \leq \kappa_{0y}) & \text{iff} \quad \kappa_{1y} < \kappa_{0y} \end{cases}
\end{aligned}$$

Thus, the sign of the difference in the cumulative probabilities can be used to identify the relative magnitude of threshold parameters. More precisely, the difference will be positive if and only if the difference between upper treated and upper nontreated threshold parameters is positive. The difference will be zero if and only if the upper thresholds are equal, and negative if and only if the difference between upper treated and upper non-treated thresholds is negative. □

Lemma 2 is analogous to Lemma 4.2 of Shaikh and Vytlacil (2005), but now with respect to the properties of ordinal potential outcomes. Information on the relative magnitude of threshold parameters can be used to tighten the bounds on the unidentified probabilities $P(Y_0|D = 1, Z)$ and $P(Y_1|D = 0, Z)$. Consider $P(Y_1|D = 0, Z)$ and recall that so far it was assumed that this probability was bounded by zero and one. Now write

$$P(Y_1 = y|D = 0, Z) = P(Y_1 \leq y|D = 0, Z) - P(Y_1 \leq y - 1|D = 0, Z)$$

which follows from the ordinal nature of Y. Furthermore, the difference

$$\begin{aligned}
&P(Y_1 \leq y|D = 0, Z) - P(Y_0 \leq y|D = 0, Z) \qquad (15) \\
&= P(\varepsilon_1 \leq \kappa_{1y}|\nu > s(z)) - P(\varepsilon_0 \leq \kappa_{0y}|\nu > s(z)) \\
&= P(\varepsilon \leq \kappa_{1y}|\nu > s(z)) - P(\varepsilon \leq \kappa_{0y}|\nu > s(z))
\end{aligned}$$

has the same sign as $\kappa_{1y} - \kappa_{0y}$, and $\delta_y \equiv sgn(\kappa_{1y} - \kappa_{0y})$ is identified by Lemma 2. This must hold for all possible outcomes y, so that by the model assumptions, the sign of the difference

$$P(Y_1 \leq y - 1|D = 0, Z) - P(Y_0 \leq y - 1|D = 0, Z) \qquad (16)$$

$$= \quad P(\varepsilon_1 \leq \kappa_{1y-1} | \nu > s(z)) - P(\varepsilon_0 \leq \kappa_{0y-1} | \nu > s(z))$$

$$= \quad P(\varepsilon \leq \kappa_{1y-1} | \nu > s(z)) - P(\varepsilon \leq \kappa_{0y-1} | \nu > s(z))$$

equals $\delta_{y-1} \equiv sgn(\kappa_{1y-1} - \kappa_{0y-1})$. The strategy to bound the unidentified probabilities is a pairwise comparison of terms in the difference

$$\begin{aligned} & P(Y_1 = y | D = 0, Z) - P(Y_0 = y | D = 0, Z) \\ = \quad & [P(Y_1 \leq y | D = 0, Z) - P(Y_1 \leq y - 1 | D = 0, Z)] \\ & -[P(Y_0 \leq y | D = 0, Z) - P(Y_0 \leq y - 1 | D = 0, Z)] \\ = \quad & [P(Y_1 \leq y | D = 0, Z) - P(Y_0 \leq y | D = 0, Z)] \\ & -[P(Y_1 \leq y - 1 | D = 0, Z) - P(Y_0 \leq y - 1 | D = 0, Z)] \end{aligned} \tag{17}$$

With three different outcomes of both δ_y and δ_{y-1} there are in total nine possibilities to consider. The following lemma states and summarizes the results for both unidentified probabilities:

Lemma 3 *Assume that* (Y_0, Y_1, D) *are generated according to model (6), and assume that conditions (A1)-(A8) are fulfilled. Then,*

$$\begin{aligned} \delta_y \quad > \quad & \delta_{y-1} \\ \Leftrightarrow \quad & P(Y_1 = y | D = 0, Z) > P(Y_0 = y | D = 0, Z) = P(Y = y | D = 0, Z) \\ & P(Y_0 = y | D = 1, Z) < P(Y_1 = y | D = 1, Z) = P(Y = y | D = 1, Z) \\ \delta_y \quad = \quad & \delta_{y-1} \ = \ 0 \\ \Leftrightarrow \quad & P(Y_1 = y | D = 0, Z) = P(Y_0 = y | D = 0, Z) = P(Y = y | D = 0, Z) \\ & P(Y_0 = y | D = 1, Z) = P(Y_1 = y | D = 1, Z) = P(Y = y | D = 1, Z) \\ \delta_y \quad < \quad & \delta_{y-1} \\ \Leftrightarrow \quad & P(Y_1 = y | D = 0, Z) < P(Y_0 = y | D = 0, Z) = P(Y = y | D = 0, Z) \\ & P(Y_0 = y | D = 1, Z) > P(Y_1 = y | D = 1, Z) = P(Y = y | D = 1, Z) \end{aligned}$$

If $\delta_y = \delta_{y-1} = \pm 1$*, then the sign of the difference* $P(Y_1 = y | D = 0, Z) - P(Y_0 = y | D = 0, Z)$ *and the sign of the difference* $P(Y_0 = y | D = 1, Z) - P(Y_1 = y | D = 1, Z)$ *are indeterminate.*

Proof. Immediately follows by application of Lemma 2, (15), (16) and (17). Note that the case $\delta_y > \delta_{y-1}$ includes possibilities $(1,0)$, $(1,-1)$, and $(0,-1)$ for pairs (δ_y, δ_{y-1}), and $\delta_y < \delta_{y-1}$ includes possibilities $(0,1)$, $(-1,1)$, and $(-1,0)$. □

Lemma 2 identifies $\delta_y \equiv sgn(\kappa_{1y} - \kappa_{0y})$ for all $y \in \mathcal{Y}$. Lemma 3 then uses the information to impose bounds on counterfactual probabilities tighter than the logical unit range. Without loss of generality, take the two evaluation points z^l and z^u with $s(z^u) > s(z^l)$, and apply Lemma 2 to identify the relative magnitude of threshold parameters. Suppose, the information is revealed that $\delta_y > \delta_{y-1}$. Then $P(Y = y|D = 0, Z)$ can be used as a lower bound for $P(Y_1 = y|D = 0, Z)$ instead of zero, and $P(Y = y|D = 1, Z)$ can be used as an upper bound for $P(Y_0 = y|D = 1, Z)$ instead of one. Bounds on $P(Y_1|Z)$ and $P(Y_0|Z)$ are thus given by

$$\begin{aligned} P(Y = y|Z) &\leq P(Y_1 = y|Z) \leq P(D = 1, Y = y|Z) + P(D = 0|Z) \\ P(D = 0, Y = y|Z) &\leq P(Y_0 = y|Z) \leq P(Y = y|Z) \end{aligned}$$

If alternatively the information is revealed that $\delta_y < \delta_{y-1}$, then the bounds on $P(Y_1|Z)$, $P(Y_0|Z)$ can be derived as

$$\begin{aligned} P(D = 1, Y = y|Z) &\leq P(Y_1 = y|Z) \leq P(Y = y|Z) \\ P(Y = y|Z) &\leq P(Y_0 = y|Z) \leq P(D = 1|Z) + P(D = 1, Y = y|Z) \end{aligned}$$

If upper and lower treated and non-treated thresholds are equal, then the outcome of Y does not vary with the treatment status because the cumulative probabilities are unchanged, and the unidentified probabilities become identified, i.e., $P(Y_1|Z) = P(Y|Z) = P(Y_0|Z)$. The bounds imposed by Lemma 3 thus depend on the category under consideration, i.e., one may have $\delta_y > \delta_{y-1}$, but $\delta_{y+1} < \delta_y$, such that the restrictions on counterfactual probabilities in category y are different from the restrictions in category $y + 1$. If Lemmas 2 and 3 do not reveal further information on the counterfactual probabilities, then the lower bound zero and the upper bound one on $P(Y_1|D = 0, Z)$ and $P(Y_0|D = 1, Z)$ still apply.

As argued above in the derivation of bounds under independence, the model assumptions imply that $P(Y_1|Z) = P(Y_1)$ and $P(Y_0|Z) = P(Y_0)$. $P(Y_1)$ and $P(Y_0)$ must therefore necessarily lie within the intersection over all possible z so that lower bounds can be replaced by supremum expressions, and upper bounds can be replaced by infimum expressions. With the exception of $\sup_{z\in\mathcal{Z}}\{P(Y=y|Z=z)\}$ and $\inf_{z\in\mathcal{Z}}\{P(Y=y|Z=z)\}$, all terms reduce according to Lemma 1. Simplification of the former is possible as well:

Lemma 4 *Assume that* (Y_0, Y_1, D) *are generated according to model (6), and assume that conditions (A1)-(A8) are fulfilled. Then,*

$$
\begin{array}{llcll}
\textit{(a1)} & \sup\limits_{z\in\mathcal{Z}}\{P(Y=y|Z=z)\} & = & P(Y=y|Z=z^u) & \textit{if } \delta_y > \delta_{y-1} \\
\textit{(a2)} & \inf\limits_{z\in\mathcal{Z}}\{P(Y=y|Z=z)\} & = & P(Y=y|Z=z^u) & \textit{if } \delta_y < \delta_{y-1} \\
\textit{(b1)} & \inf\limits_{z\in\mathcal{Z}}\{P(Y=y|Z=z)\} & = & P(Y=y|Z=z^l) & \textit{if } \delta_y > \delta_{y-1} \\
\textit{(b2)} & \sup\limits_{z\in\mathcal{Z}}\{P(Y=y|Z=z)\} & = & P(Y=y|Z=z^l) & \textit{if } \delta_y < \delta_{y-1}
\end{array}
$$

Proof. Consider part (a1) and recall that $s(z^u) \geq s(z)$ for all z. The assumptions ensure that

$$
\begin{aligned}
& P(Y=y|Z=z^u) - P(Y=y|Z=z) \qquad (18) \\
&= P(D=0, Y=y|Z=z^u) + P(D=1, Y=y|Z=z^u) \\
&\quad -P(D=0, Y=y|Z=z) - P(D=1, Y=y|Z=z) \\
&= P(\nu > s(z^u), \kappa_{0y-1} < \varepsilon \leq \kappa_{0y}) + P(\nu \leq s(z^u), \kappa_{1y-1} < \varepsilon \leq \kappa_{1y}) \\
&\quad -P(\nu > s(z), \kappa_{0y-1} < \varepsilon \leq \kappa_{0y}) - P(\nu \leq s(z), \kappa_{1y-1} < \varepsilon \leq \kappa_{1y}) \\
&= P(s(z) < \nu \leq s(z^u), \kappa_{1y-1} < \varepsilon \leq \kappa_{1y}) - P(s(z) < \nu \leq s(z^u), \kappa_{0y-1} < \varepsilon \leq \kappa_{0y}) \geq 0
\end{aligned}
$$

where the last inequality follows by definition of $s(z^u)$ and $\delta_y > \delta_{y-1}$. Since the supremum in part (a1) of the lemma is equivalent to the infimum of (18) over z and (18) must be non-negative, necessary and sufficient condition for an infimum of (18) is that $z = z^u$.

If $\delta_y < \delta_{y-1}$, then (18) holds under the weak inequality ≤ 0, and the infimum in part (a2) of the lemma is equivalent to the supremum of (18) over z. As (18) must be non-positive, necessary and sufficient condition for a supremum is that $z = z^u$. Following analogous arguments for the infimum in the case $\delta_y > \delta_{y-1}$ and the supremum in the case $\delta_y < \delta_{y-1}$ proves parts (b1) and (b2) of the lemma. □

The following proposition uses the bounds on $P(Y_0)$ and $P(Y_1)$ under the threshold crossing model structure of treatment selection and potential outcomes to bound the average treatment and the treatment on the treated parameters:

Proposition 1 *Assume that (Y_0, Y_1, D) are generated according to model (6), and assume that conditions (A1)-(A8) are fulfilled. Then,*

$$\Delta_y^{ATE} \in \left[LB4_y^{ATE}, UB4_y^{ATE} \right] \qquad \textit{with} \tag{19}$$

$$LB4_y^{ATE} = \begin{cases} P(Y=y|Z=z^u) - P(Y=y|Z=z^l) & \textit{if } \delta_y > \delta_{y-1} \\ 0 & \textit{if } \delta_y = \delta_{y-1} = 0 \\ LB3_y^{ATE} & \textit{if } \delta_y < \delta_{y-1} \\ LB3_y^{ATE} & \textit{if } \delta_y = \delta_{y-1} = \pm 1 \end{cases}$$

$$UB4_y^{ATE} = \begin{cases} UB3_y^{ATE} & \textit{if } \delta_y > \delta_{y-1} \\ 0 & \textit{if } \delta_y = \delta_{y-1} = 0 \\ P(Y=y|Z=z^u) - P(Y=y|Z=z^l) & \textit{if } \delta_y < \delta_{y-1} \\ UB3_y^{ATE} & \textit{if } \delta_y = \delta_{y-1} = \pm 1 \end{cases}$$

and

$$\Delta_y^{TT} \in \left[LB4_y^{TT}, UB4_y^{TT} \right] \qquad \textit{with} \tag{20}$$

$$LB4_y^{TT} = \begin{cases} [P(Y=y) - P(Y=y|Z=z^l)]/P(D=1) & \textit{if } \delta_y > \delta_{y-1} \\ 0 & \textit{if } \delta_y = \delta_{y-1} = 0 \\ LB3_y^{TT} & \textit{if } \delta_y < \delta_{y-1} \\ LB3_y^{TT} & \textit{if } \delta_y = \delta_{y-1} = \pm 1 \end{cases}$$

$$UB4_y^{TT} = \begin{cases} UB3_y^{TT} & \textit{if } \delta_y > \delta_{y-1} \\ 0 & \textit{if } \delta_y = \delta_{y-1} = 0 \\ [P(Y=y) - P(Y=y|Z=z^l)]/P(D=1) & \textit{if } \delta_y < \delta_{y-1} \\ UB3_y^{TT} & \textit{if } \delta_y = \delta_{y-1} = \pm 1 \end{cases}$$

For known threshold parameters (interval data), (19) and (20) still hold, but δ_y and δ_{y-1} can a-priori be determined and there is no uncertainty about the four cases.

Proof. Follows directly by Lemmas 1, 2, 3, and 4, and the discussion preceding Lemma 4. For known threshold parameters the identification strategy of Lemma 2 becomes redundant. Given the additional information, bounds on the unidentified counterfactuals $P(Y_0|D = 1, Z)$ and $P(Y_1|D = 0, Z)$ can be directly imposed as described in Lemma 3 with δ_y, δ_{y-1} known. □

Note that the width of the bounds in (19) and (20) is at maximum the same and in many cases smaller than the width of the bounds in (13) and (14). If $\delta_y > \delta_{y-1}$, then the upper bound in (19) corresponds to the upper bound in (13), but the lower bound in (19) is larger than the lower bound in (13), since

$$\begin{aligned} LB4_y^{ATE} - LB3_y^{ATE} &= P(D = 0, Y = y|Z = z^u) \\ &\quad - P(D = 1, Y = y|Z = z^l) + P(D = 1|Z = z^l) > 0 \end{aligned}$$

With the same argument, if $\delta_y < \delta_{y-1}$, then the lower bounds in (19) and (13) are the same, but the upper bound in (19) is lower than the upper bound in (13), i.e., $UB4_y^{ATE} - UB3_y^{ATE} < 0$.

Analogously, for the treatment on the treated parameter and a positive sign of the difference $\delta_y - \delta_{y-1}$, the lower bound in (20) is larger than the lower bound in (14), i.e., $LB4_y^{TT} - LB3_y^{TT} > 0$, with the upper bounds unchanged, and if $\delta_y - \delta_{y-1}$ is negative, then the upper bound in (20) is lower than the upper bound in (14), i.e., $UB4_y^{TT} - UB3_y^{TT} < 0$, with the lower bounds unchanged. If $\delta_y = \delta_{y-1} = 0$, then both treatment parameters become point-identified to be zero. Only if $\delta_y = \delta_{y-1} = \pm 1$, then the width of the bounds does not change and the threshold mechanism is uninformative on the treatment parameters.

Note that unlike for the bounds constructed before, the sign of Δ_y^{ATE} and Δ_y^{TT} as bounded by Proposition 1 can be identified if $\delta_y \lesseqgtr \delta_{y-1}$ or $\delta_y = \delta_{y-1} = 0$. This follows because the lower bounds $LB4_y^{ATE}$ and $LB4_y^{TT}$ of both treatment parameters are positive in the case $\delta_y > \delta_{y-1}$, and in the case $\delta_y < \delta_{y-1}$ the upper bounds $UB4_y^{ATE}$ and $UB4_y^{TT}$ are negative. Finally, if $\delta_y = \delta_{y-1} = 0$, then the sign of the treatment effects is point-identified to be zero.

The final remark on (19) and (20) is related to the case of known thresholds. Given the assumptions of the model and provided that no X covariates are available, the only way that treated and non-treated individuals may differ are the threshold parameters. If the thresholds do not vary by the treatment status, and are thus equal, then $\delta_y = \delta_{y-1} = 0$ in all cases and the treatment parameters are point-identified to be zero, as predicted by Proposition 1.

3.3 Including Covariates

I now turn to the case when X covariates are available and to the full model (6). The treatment parameters conditional on X are defined as

$$\Delta_y^{ATE}(x) = P(Y_1 = y|X = x) - P(Y_0 = y|X = x) \tag{21}$$

$$\begin{aligned} \Delta_y^{TT}(x) &= P(Y_1 = y|D = 1, X = x) - P(Y_0 = y|D = 1, X = x) \\ &= [P(Y = y|X = x) - P(Y_0 = y|X = x)]/P(D = 1|X = x) \end{aligned} \tag{22}$$

By the preceding discussion, it is straightforward to show that $P(Y_1|X)$ and $P(Y_0|X)$ are only partially identified, and so are the treatment parameters. The offending terms are, as before, the counterfactuals $P(Y_1|D = 0, X)$ and $P(Y_0|D = 1, X)$, respectively. All the results derived before in (7) and (9), Lemma 1, and (13) and (14) are trivially extended to X conditioned on.

In principle, the same holds true for the whole discussion in the preceding section, i.e., Lemmas 2, 3, 4, and Proposition 1 may easily be extended to hold conditional on X. There is, however, a potential source of narrowing the bounds, given that X varies conditional on Z, i.e., there exists at least one element in X that is not included in Z. This extra variation can be explored as follows. Consider a modified version of Lemma 2:

Lemma 5 *Assume that (Y_0, Y_1, D) are generated according to model (6), and assume that conditions (A1)-(A8) are fulfilled. Then for any evaluation points x_0, x_1, z_0, z_1 with $P(D = 1|X = x_j, Z = z_1) > P(D = 1|X = x_j, Z = z_0)$, $j = 0, 1$,*

$$\begin{aligned} sgn\Big\{&[P(D = 1, Y \le y|X = x_1, Z = z_1) - P(D = 1, Y \le y|X = x_1, Z = z_0)] \\ &-[P(D = 0, Y \le y|X = x_0, Z = z_0) - P(D = 0, Y \le y|X = x_0, Z = z_1)]\Big\} \\ &= sgn(\kappa_{1y}(x_1) - \kappa_{0y}(x_0)) \equiv \delta_y(x_1, x_0) \end{aligned}$$

so that $\delta_y(x_1, x_0)$ can take three values $-1, 0, 1$ depending on whether the difference between $\kappa_{1y}(x_1) \equiv \kappa_{1y} - r_1(x_1)$ and $\kappa_{0y}(x_0) \equiv \kappa_{0y} - r_0(x_0)$ is negative, zero, or positive, respectively.

Proof. Consider the probability differences in the sign function separately:

$$
\begin{aligned}
&P(D=1,Y\leq y|X=x_1,Z=z_1)-P(D=1,Y\leq y|X=x_1,Z=z_0) &(23)\\
&= P(D=1,Y_1\leq y|X=x_1,Z=z_1)-P(D=1,Y_1\leq y|X=x_1,Z=z_0)\\
&= P(\nu\leq s(z_1),\varepsilon_1\leq \kappa_{1y}-r_1(x_1))-P(\nu\leq s(z_0),\varepsilon_1\leq \kappa_{1y}-r_1(x_1))\\
&= P(s(z_0)<\nu\leq s(z_1),\varepsilon_1\leq \kappa_{1y}(x_1))=P(s(z_0)<\nu\leq s(z_1),\varepsilon\leq \kappa_{1y}(x_1))
\end{aligned}
$$

and

$$
\begin{aligned}
&P(D=0,Y\leq y|X=x_0,Z=z_0)-P(D=0,Y\leq y|X=x_0,Z=z_1) &(24)\\
&= P(D=0,Y_0\leq y|X=x_0,Z=z_0)-P(D=0,Y_0\leq y|X=x_0,Z=z_1)\\
&= P(\nu> s(z_0),\varepsilon_0\leq \kappa_{0y}-r_0(x_0))-P(\nu> s(z_1),\varepsilon_0\leq \kappa_{0y}-r_0(x_0))\\
&= P(s(z_0)<\nu\leq s(z_1),\varepsilon_0\leq \kappa_{0y}(x_0))=P(s(z_0)<\nu\leq s(z_1),\varepsilon\leq \kappa_{0y}(x_0))
\end{aligned}
$$

by the assumptions of the lemma, and $\kappa_{1y}(x_1)\equiv\kappa_{1y}-r_1(x_1)$ and $\kappa_{0y}(x_0)\equiv\kappa_{0y}-r_0(x_0)$. Taking the difference between (23) and (24) yields

$$
\begin{aligned}
&P(D=1,Y\leq y|X=x_1,Z=z_1)-P(D=1,Y\leq y|X=x_1,Z=z_0)\\
&\quad -[P(D=0,Y\leq y|X=x_0,Z=z_0)-P(D=0,Y\leq y|X=x_0,Z=z_1)]\\
&= \begin{cases} P(s(z_0)<\nu\leq s(z_1),\kappa_{0y}(x_0)<\varepsilon\leq\kappa_{1y}(x_1)) & \text{iff} \quad \kappa_{1y}(x_1)>\kappa_{0y}(x_0)\\ 0 & \text{iff} \quad \kappa_{1y}(x_1)=\kappa_{0y}(x_0)\\ -P(s(z_0)<\nu\leq s(z_1),\kappa_{1y}(x_1)<\varepsilon\leq\kappa_{0y}(x_0)) & \text{iff} \quad \kappa_{1y}(x_1)<\kappa_{0y}(x_0)\end{cases}
\end{aligned}
$$

Thus, the sign of the double difference in the cumulative probabilities can be used to identify the relative magnitude of $\kappa_{1y}(x_1)$ and $\kappa_{0y}(x_0)$. More precisely, the double difference will be positive if and only if the difference between $\kappa_{1y}(x_1)\equiv\kappa_{1y}-r_1(x_1)$ and $\kappa_{0y}(x_0)\equiv\kappa_{0y}-r_0(x_0)$ is positive. It will be zero if and only if the indices, accounting for the upper bound of the threshold mechanism, are equal, and negative if and only if $\kappa_{1y}(x_1)-\kappa_{0y}(x_0)$ is negative. □

Lemma 5 can be used to obtain bounds on the counterfactuals $P(Y_1=y|D=0,X,Z)$ and $P(Y_0=y|D=1,X,Z)$ tighter than the logical unit range. Consider the former counterfactual probability, and recall that

$$P(Y_1 = y|D = 0, X, Z) = P(Y_1 \leq y|D = 0, X, Z) - P(Y_1 \leq y - 1|D = 0, X, Z)$$

by the ordinal nature of Y. Take the first cumulative probability, evaluated at x_1, and subtract the identified probability $P(Y \leq y|D = 0, X, Z)$ evaluated at x_0 to obtain

$$\begin{aligned} & P(Y_1 \leq y|D = 0, X = x_1, Z) - P(Y_0 \leq y|D = 0, X = x_0, Z) \\ = \;& P(\varepsilon \leq \kappa_{1y}(x_1)|\nu > s(z)) - P(\varepsilon \leq \kappa_{0y}(x_0)|\nu > s(z)) \end{aligned}$$

The sign of the (unidentified) difference only depends on the sign of the difference $\kappa_{1y}(x_1) - \kappa_{0y}(x_0)$, which is identified by Lemma 5. Thus, if $\delta_y(x_1, x_0) > 0$, and hence $\kappa_{1y}(x_1) > \kappa_{0y}(x_0)$, then the above difference will be positive. If $\delta_y(x_1, x_0) < 0$, then the above difference will be negative, and if $\delta_y(x_1, x_0) = 0$, then $P(Y_1 \leq y|D = 0, X, Z) = P(Y_0 \leq y|D = 0, X, Z)$ becomes point-identified. Since Lemma 5 holds for all $y \in \mathcal{Y}$, analogous arguments prove that the difference

$$\begin{aligned} & P(Y_1 \leq y - 1|D = 0, X = x_1, Z) - P(Y_0 \leq y - 1|D = 0, X = x_0, Z) \\ = \;& P(\varepsilon \leq \kappa_{1y-1}(x_1)|\nu > s(z)) - P(\varepsilon \leq \kappa_{0y-1}(x_0)|\nu > s(z)) \end{aligned}$$

has the same sign as $\delta_{y-1}(x_1, x_0)$. A pairwise comparison of terms in the difference

$$\begin{aligned} & P(Y_1 = y|D = 0, X = x_1, Z) - P(Y_0 = y|D = 0, X = x_0, Z) \qquad (25) \\ = \;& P(Y_1 \leq y|D = 0, X = x_1, Z) - P(Y_1 \leq y - 1|D = 0, X = x_1, Z) \\ & -[P(Y_0 \leq y|D = 0, X = x_0, Z) - P(Y_0 \leq y - 1|D = 0, X = x_0, Z)] \\ = \;& P(Y_1 \leq y|D = 0, X = x_1, Z) - P(Y_0 \leq y|D = 0, X = x_0, Z) \\ & -[P(Y_1 \leq y - 1|D = 0, X = x_1, Z) - P(Y_0 \leq y - 1|D = 0, X = x_0, Z)] \\ = \;& P(\varepsilon \leq \kappa_{1y}(x_1)|\nu > s(z)) - P(\varepsilon \leq \kappa_{0y}(x_0)|\nu > s(z)) \\ & -[P(\varepsilon \leq \kappa_{1y-1}(x_1)|\nu > s(z)) - P(\varepsilon \leq \kappa_{0y-1}(x_0)|\nu > s(z))] \end{aligned}$$

may thus be used to obtain bounds on the unidentified counterfactual probabilities. For example, if Lemma 5 reveals the information that $\delta_y(x_1, x_0) > \delta_{y-1}(x_1, x_0)$, then the difference between the former two probabilities after the last equality in (25) must be larger than the difference between the latter two, so that the overall sign is positive, and $P(Y_0 = y|D = 0, X = x_0, Z)$ can be used as lower bound for $P(Y_1 = y|D = 0, X = x_1, Z)$ instead of zero. By the same arguments, bounds on the counterfactual probability $P(Y_0 = y|D = 1, X, Z)$ can be obtained. The following lemma summarizes and states the results:

Lemma 6 *Assume that* (Y_0, Y_1, D) *are generated according to model (6), and assume that conditions (A1)-(A8) are fulfilled. Then,*

$$\begin{aligned}
(a) \quad \delta_y(x, \tilde{x}) &> \delta_{y-1}(x, \tilde{x}) \\
&\Leftrightarrow P(Y_1 = y|D = 0, X = x, Z) > P(Y = y|D = 0, X = \tilde{x}, Z) \\
\delta_y(x, \tilde{x}) &= \delta_{y-1}(x, \tilde{x}) = 0 \\
&\Leftrightarrow P(Y_1 = y|D = 0, X = x, Z) = P(Y = y|D = 0, X = \tilde{x}, Z) \\
\delta_y(x, \tilde{x}) &< \delta_{y-1}(x, \tilde{x}) \\
&\Leftrightarrow P(Y_1 = y|D = 0, X = x, Z) < P(Y = y|D = 0, X = \tilde{x}, Z)
\end{aligned}$$

If $\delta_y(x, \tilde{x}) = \delta_{y-1}(x, \tilde{x}) = \pm 1$*, then the sign of the difference* $P(Y_1 = y|D = 0, X = x, Z) - P(Y_0 = y|D = 0, X = \tilde{x}, Z)$ *is indeterminate. And,*

$$\begin{aligned}
(b) \quad \delta_y(\tilde{x}, x) &> \delta_{y-1}(\tilde{x}, x) \\
&\Leftrightarrow P(Y_0 = y|D = 1, X = x, Z) < P(Y = y|D = 1, X = \tilde{x}, Z) \\
\delta_y(\tilde{x}, x) &= \delta_{y-1}(\tilde{x}, x) = 0 \\
&\Leftrightarrow P(Y_0 = y|D = 1, X = x, Z) = P(Y = y|D = 1, X = \tilde{x}, Z) \\
\delta_y(\tilde{x}, x) &< \delta_{y-1}(\tilde{x}, x) \\
&\Leftrightarrow P(Y_0 = y|D = 1, X = x, Z) > P(Y = y|D = 1, X = \tilde{x}, Z)
\end{aligned}$$

If $\delta_y(\tilde{x}, x) = \delta_{y-1}(\tilde{x}, x) = \pm 1$*, then the sign of the difference* $P(Y_0 = y|D = 1, X = x, Z) - P(Y_1 = y|D = 1, X = \tilde{x}, Z)$ *is indeterminate.*

Proof. Part (a) follows directly by application of Lemma 5 and (25). Part (b) follows by analogous arguments applying Lemma 5 and $P(Y_0 = y|D = 1, X = x_0, Z) - P(Y_1 = y|D = 1, X = x_1, Z)$ replacing the probability difference in (25). □

Lemma 6 holds for all evaluation points $\tilde{x}$ in the support of X. Clearly, there might be some evaluation points $\tilde{x}$ for that $\delta_y(x,\tilde{x}) > \delta_{y-1}(x,\tilde{x})$, and other evaluation points $\tilde{x}$ for that $\delta_y(x,\tilde{x}) < \delta_{y-1}(x,\tilde{x})$, or $\delta_y(x,\tilde{x}) = \delta_{y-1}(x,\tilde{x}) = 1$, for example. In order to use the full information, let

$$\begin{aligned}\mathcal{X}_0^l(x_1) &= \{x_0 : \ \delta_y(x_1,x_0) > \delta_{y-1}(x_1,x_0)\}\\ \mathcal{X}_0^u(x_1) &= \{x_0 : \ \delta_y(x_1,x_0) < \delta_{y-1}(x_1,x_0)\}\end{aligned}$$

and

$$\begin{aligned}\mathcal{X}_1^l(x_0) &= \{x_1 : \ \delta_y(x_1,x_0) < \delta_{y-1}(x_1,x_0)\}\\ \mathcal{X}_1^u(x_0) &= \{x_1 : \ \delta_y(x_1,x_0) > \delta_{y-1}(x_1,x_0)\}\end{aligned}$$

It is made explicit in the definition of sets that these are either over x_0 for x_1 fixed (and thus are a function of x_1), or over x_1 for x_0 fixed (and thus are a function of x_0). Bounds on the counterfactual probability $P(Y_1 = y|D = 0, X, Z)$, conditional on all values z in the support of Z can then be derived as

$$\begin{aligned}&\sup_{\tilde{x}\in\mathcal{X}_0^l(x)} \{P(Y = y|D = 0, X = \tilde{x}, Z)\}\\ &\qquad \le P(Y_1 = y|D = 0, X = x, Z) \le \inf_{\tilde{x}\in\mathcal{X}_0^u(x)} \{P(Y = y|D = 0, X = \tilde{x}, Z)\}\end{aligned}$$

If there exists $\tilde{x}$ such that $\delta_y(x,\tilde{x}) = \delta_{y-1}(x,\tilde{x}) = 0$ (for x fixed), then point-identification of the counterfactual probability follows, i.e., $P(Y_1 = y|D = 0, X = x, Z) = P(Y = y|D = 0, X = \tilde{x}, Z)$. If no such $\tilde{x}$ exists, and no $\tilde{x}$ for that Lemma 6 yields tighter bounds than the unit range, then $\mathcal{X}_0^l$ and $\mathcal{X}_0^u$ are empty and it is understood that the bounds zero and one still apply. Analogously, for $P(Y_0 = y|D = 1, X, Z)$ the bounds can be derived as

$$\begin{aligned}&\sup_{\tilde{x}\in\mathcal{X}_1^l(x)} \{P(Y = y|D = 1, X = \tilde{x}, Z)\}\\ &\qquad \le P(Y_0 = y|D = 1, X = x, Z) \le \inf_{\tilde{x}\in\mathcal{X}_1^u(x)} \{P(Y = y|D = 1, X = \tilde{x}, Z)\}\end{aligned}$$

with point-identification $P(Y_0 = y|D = 1, X = x, Z) = P(Y = y|D = 1, X = \tilde{x}, Z)$ if there exists $\tilde{x}$ such that $\delta_y(\tilde{x}, x) = \delta_{y-1}(\tilde{x}, x) = 0$, and bounds zero and one if $\mathcal{X}_1^l$ and $\mathcal{X}_1^u$ are empty.

Replacing the bounds for the counterfactual probabilities in the expressions for $P(Y_1|X, Z)$ and $P(Y_0|X, Z)$ and following the same arguments as under the independence assumption, the bounds on $P(Y_1|X)$ and $P(Y_0|X)$ are given by

$$
\begin{aligned}
L\mathcal{B}_y^1(x) \equiv \sup_{z\in\mathcal{Z}} \Big\{ & P(D=1, Y=y|X=x, Z=z) \\
& + \sup_{\tilde{x}\in\mathcal{X}_0^l(x)} \{P(Y=y|D=0, X=\tilde{x}, Z=z)\} P(D=0|X=x, Z=z) \Big\} \\
\leq P(Y_1 = y|X=x) \leq & \qquad (26) \\
U\mathcal{B}_y^1(x) \equiv \inf_{z\in\mathcal{Z}} \Big\{ & P(D=1, Y=y|X=x, Z=z) \\
& + \inf_{\tilde{x}\in\mathcal{X}_0^u(x)} \{P(Y=y|D=0, X=\tilde{x}, Z=z)\} P(D=0|X=x, Z=z) \Big\}
\end{aligned}
$$

and

$$
\begin{aligned}
L\mathcal{B}_y^0(x) \equiv \sup_{z\in\mathcal{Z}} \Big\{ & \sup_{\tilde{x}\in\mathcal{X}_1^l(x)} \{P(Y=y|D=1, X=\tilde{x}, Z=z)\} P(D=1|X=x, Z=z) \\
& + P(D=0, Y=y|X=x, Z=z) \Big\} \\
\leq P(Y_0 = y|X=x) \leq & \qquad (27) \\
U\mathcal{B}_y^0(x) \equiv \inf_{z\in\mathcal{Z}} \Big\{ & \inf_{\tilde{x}\in\mathcal{X}_1^u(x)} \{P(Y=y|D=1, X=\tilde{x}, Z=z)\} P(D=1|X=x, Z=z) \\
& + P(D=0, Y=y|X=x, Z=z) \Big\}
\end{aligned}
$$

The following proposition uses the bounds in (26) and (27) under the threshold crossing model structure and the full model to impose bounds on the average treatment effect and the average treatment effect on the treated conditional on X:

Proposition 2 *Assume that (Y_0, Y_1, D) are generated according to model (6), and assume that conditions (A1)-(A8) are fulfilled. Then,*

$$\Delta_y^{ATE}(x) \in \left[L\mathcal{B}5_y^{ATE}(x), U\mathcal{B}5_y^{ATE}(x)\right] \qquad \textit{with} \qquad (28)$$

$$
\begin{aligned}
L\mathcal{B}5_y^{ATE}(x) &= L\mathcal{B}_y^1(x) - U\mathcal{B}_y^0(x) \\
U\mathcal{B}5_y^{ATE}(x) &= U\mathcal{B}_y^1(x) - L\mathcal{B}_y^0(x)
\end{aligned}
$$

and

$$\Delta_y^{TT}(x) \in \left[L\mathcal{B}5_y^{TT}(x), U\mathcal{B}5_y^{TT}(x)\right] \qquad \textit{with} \qquad (29)$$

$$LB5_y^{TT}(x) = [P(Y = y|X = x) - UB_y^0(x)]/P(D = 1|X = x)$$
$$UB5_y^{TT}(x) = [P(Y = y|X = x) - LB_y^0(x)]/P(D = 1|X = x)$$

Proof. Follows directly by Lemmas 5, 6, and the discussion preceding the proposition. □

The bounds imposed by Proposition 2 depend on the amount of variation in X conditional on Z, and therefore it is difficult to make a general statement about their properties. However, two important conclusions can be drawn. First, if X does not vary conditional on Z, then the bounds in (28) and (29) simplify to the bounds in (19) and (20) with X conditioned on, but there is no possibility to further narrow the bounds. The reason is that if X is degenerate conditional on Z, then there exists only one $\tilde{x} = x$ in Lemma 6, which then becomes equivalent to Lemma 3 conditional on X. Thus, the cases $\delta_y(x, x) \lessgtr \delta_{y-1}(x, x)$, $\delta_y(x, x) = \delta_{y-1}(x, x) = 0$ allow to impose new upper or / and lower bounds on the counterfactual probabilities, if $\delta_y(x, x) = \delta_{y-1}(x, x) \pm 1$, then the bounds zero and one still apply, and as a consequence, the bounds in Proposition 2 collapse to those in Proposition 1 (conditional on X).

Second, the sign of the treatment effects is always identified by the bounds in Proposition 2. First consider the bounds in (28) and assume that the true average treatment effect is positive, i.e., $P(Y_1 = y|X = x) > P(Y_0 = y|X = x)$. Then $\delta_y(x, x) > \delta_{y-1}(x, x)$ so that $x \in \mathcal{X}_0^l(x)$ and $x \in \mathcal{X}_1^u(x)$ by Lemma 5. Thus, for the lower bound it must hold that

$$\begin{aligned} & LB_y^1(x) - UB_y^0(x) \\ &= \sup_{z \in Z} \{P(D = 1, Y = y|X = x, Z = z) + P(D = 0, Y = y|X = x, Z = z)\} \\ &\quad - \inf_{z \in Z} \{P(D = 1, Y = y|X = x, Z = z) + P(D = 0, Y = y|X = x, Z = z)\} \\ &= P(Y = y|X = x, Z = z^u) - P(Y = y|X = x, Z = z^l) > 0 \end{aligned}$$

which follows by Lemma 4 conditional on X, Lemma 6, and the definition of z^u and z^l. The inequality holds for $\tilde{x} = x$, if other $\tilde{x} \in \mathcal{X}_0^l(x)$ and $\tilde{x} \in \mathcal{X}_1^u(x)$ exist, then $LB_y^1(x)$ may get larger but never can get smaller by the supremum condition, and $UB_y^0(x)$ may get smaller but never can get larger by the infimum condition, so that the inequality will still hold, and the lower bound in (28) will strictly be positive. By similar arguments, one can show that for the upper bound in the case of $P(Y_1 = y|X = x) < P(Y_0 = y|X = x)$, $U\mathcal{B}_y^1(x) - L\mathcal{B}_y^0(x)$ is negative for $\tilde{x} = x$, and will always be negative for all $\tilde{x} \in \mathcal{X}_0^u(x)$ and $\tilde{x} \in \mathcal{X}_1^l(x)$ other than x. If $P(Y_1 = y|X = x) = P(Y_0 = y|X = x)$, then $\delta_y(x, x) = \delta_{y-1}(x, x) = 0$, so that the counterfactual probabilities become identified by Lemma 6, and the average treatment effect is point-identified to be zero.

Next consider the treatment on the treated parameter and assume that the true parameter is positive. Then $\delta_y(x, x) > \delta_{y-1}(x, x)$ by Lemma 5 so that $x \in \mathcal{X}_1^u(x)$. The sign of $L\mathcal{B}5_y^{TT}(x)$ is determined by the sign of $P(Y = y|X = x) - U\mathcal{B}_y^0(x)$. Simplifying terms yields

$$P(Y = y|X = x) - P(Y = y|X = x, Z = z^l) > 0$$

by Lemma 4 conditional on X, Lemma 6, and the definition of z^l. Thus, the lower bound is positive for $\tilde{x} = x$, and will always be positive for all $\tilde{x} \in \mathcal{X}_1^u(x)$ other than x due to the infimum condition. By analogous steps, one can show that the upper bound of the treatment on treated parameter will always be negative if the true parameter is negative, and the bounds collapse to zero and thus provide point-identification if the true parameter is zero.

4 Inference

Shaikh and Vytlacil (2005) describe the construction of confidence sets given a discontinuity in the form of the bounds. Special attention to inference is necessary in this case because the usual approach of estimating probabilities by relative frequencies (or replacing population features by sample counterparts) will be inconsistent at the jump points. Their approach is based on the construction of a random set $\mathcal{CI}$ that will asymptotically cover, with probability at least $1 - \alpha$ for fixed $\alpha \in (0, 1)$, all treatment effects as identified by the population bounds.

The confidence set approach can also be implemented here. To simplify exposition, let the observed data be n independently and identically distributed drawings (Y_i, D_i, X_i, Z_i) from the population of interest, and let X and Z be discrete random variables. Furthermore, assume that the probability of treatment selection varies with each single outcome of Z, i.e., for all evaluation points $z_1 \neq z_0$ we have that $P(D = 1|X, Z = z_1) \neq P(D = 1|X, Z = z_0)$. In order to illustrate ideas, consider first the construction of confidence sets for the average treatment parameters defined by (13), i.e., in the case of no X covariates, threshold crossing treatment selection, but without further assumptions on the mechanism generating the outcome variable. Let

$$\hat{P}(z) = \frac{1}{|\{i : Z_i = z\}|} \sum_{i:Z_i=z} D_i$$

denote a consistent estimator of $P(D = 1|Z = z)$. The evaluation points z^l and z^u may then be estimated by $\hat{z}^l = \min_z \hat{P}(z)$ and $\hat{z}^u = \max_z \hat{P}(z)$. Given the assumptions and with n large enough, one can show that $plim\, \hat{z}^l = z^l$ and $plim\, \hat{z}^u = z^u$. Consistent estimators of the bounds $L\mathcal{B}3_y^{ATE}$ and $U\mathcal{B}3_y^{ATE}$ can be obtained by

$$\begin{aligned}
\widehat{L\mathcal{B}3}_y^{ATE} &= \frac{1}{|\{i : Z_i = \hat{z}^u\}|} \sum_{i:Z_i=\hat{z}^u} D_i Y_{iy} - \frac{1}{|\{i : Z_i = \hat{z}^l\}|} \sum_{i:Z_i=\hat{z}^l} D_i \\
&\quad - \frac{1}{|\{i : Z_i = \hat{z}^l\}|} \sum_{i:Z_i=\hat{z}^l} (1 - D_i) Y_{iy} \\
\widehat{U\mathcal{B}3}_y^{ATE} &= \frac{1}{|\{i : Z_i = \hat{z}^u\}|} \sum_{i:Z_i=\hat{z}^u} D_i Y_{iy} + \frac{1}{|\{i : Z_i = \hat{z}^u\}|} \sum_{i:Z_i=\hat{z}^u} (1 - D_i) \\
&\quad - \frac{1}{|\{i : Z_i = \hat{z}^l\}|} \sum_{i:Z_i=\hat{z}^l} (1 - D_i) Y_{iy}
\end{aligned}$$

where Y_{iy} is a dummy variable taking the value one if $Y_i = y$, and zero otherwise. Each of these estimators contains sums of means of binary variables, such that large sample theorems can be evoked to establish, for example,

$$\sqrt{n}\left(\widehat{L\mathcal{B}3}_y^{ATE} - L\mathcal{B}3_y^{ATE}\right) \xrightarrow{d} N(0, \sigma^2_{lb3,y})$$

with asymptotic variance $\sigma^2_{lb3,y}$. Analogously, asymptotic normality of the estimated upper bound can be established, and aymptotically valid confidence intervals for $L\mathcal{B}3_y^{ATE}$ and $U\mathcal{B}3_y^{ATE}$ can be found by the estimated lower and upper bounds plus/minus a measure of variation. The confidence intervals for the bounds in turn can be used to construct a random set that will asymptotically cover, with probability at least $1 - \alpha$, the average treatment effects as defined by the population bounds in (13). Let $q_{1-\alpha}$ denote the $(1-\alpha)$-quantile of the standard normal distribution, and let $\hat{\sigma}^2_{lb3,y}$, $\hat{\sigma}^2_{ub3,y}$ denote consistent estimators of the variances in the asymptotic distributions of the estimated lower and upper bounds, respectively. Then,

$$P\left(L\mathcal{B}3_y^{ATE} > \widehat{L\mathcal{B}3}_y^{ATE} - \frac{\hat{\sigma}_{lb3,y} q_{1-\alpha}}{\sqrt{n}}\right)$$

and

$$P\left(\mathcal{UB3}_y^{ATE} < \widehat{\mathcal{UB3}}_y^{ATE} + \frac{\hat{\sigma}_{ub3,y}q_{1-\alpha}}{\sqrt{n}}\right)$$

both converge in probability to $1-\alpha$. For each Δ_y^{ATE} in the interval $\left[\mathcal{LB3}_y^{ATE}, \mathcal{UB3}_y^{ATE}\right]$ it must therefore hold that in the limit the probability of $\widehat{\mathcal{LB3}}_y^{ATE} - \hat{\sigma}_{lb3,y}q_{1-\alpha}/\sqrt{n}$ being smaller than the true average treatment effect, and the probability of $\widehat{\mathcal{UB3}}_y^{ATE} + \hat{\sigma}_{ub3,y}q_{1-\alpha}/\sqrt{n}$ being larger than the true average treatment effect are at least $1-\alpha$, with equality if Δ_y^{ATE} is exactly at the lower (upper) boundary. Thus, with probability at least $1-\alpha$ and for large n, the interval

$$\mathcal{CI3}_y^{ATE} = \left[\widehat{\mathcal{LB3}}_y^{ATE} - \frac{\hat{\sigma}_{lb3,y}q_{1-\alpha}}{\sqrt{n}},\ \widehat{\mathcal{UB3}}_y^{ATE} + \frac{\hat{\sigma}_{ub3,y}q_{1-\alpha}}{\sqrt{n}}\right] \tag{30}$$

will cover the true average treatment effects as defined by (13). For details on this approach see also Imbens and Manski (2004). Alternative approaches of obtaining asymptotically valid confidence sets exist, such as Horowitz and Manski (2000), or Chernozhukov *et al.* (2007), but I will restrict myself to the confidence set approach as outlined above.

The confidence set for the average treatment on the treated parameters, as bounded by (14), can be derived by parallel arguments. Consistent estimators of the lower and the upper bounds of the average treatment on the treated effect can be found by

$$\begin{aligned}
\widehat{\mathcal{LB3}}_y^{TT} &= \left[\frac{1}{n}\sum_{i=1}^{n} Y_{iy} - \frac{1}{|\{i : Z_i = \hat{z}^l\}|}\sum_{i:Z_i=\hat{z}^l} D_i \right.\\
&\qquad \left. - \frac{1}{|\{i : Z_i = \hat{z}^l\}|}\sum_{i:Z_i=\hat{z}^l} (1-D_i)Y_{iy}\right] \Big/ \left(\frac{1}{n}\sum_{i=1}^{n} D_i\right)\\
\widehat{\mathcal{UB3}}_y^{TT} &= \left[\frac{1}{n}\sum_{i=1}^{n} Y_{iy} - \frac{1}{|\{i : Z_i = \hat{z}^l\}|}\sum_{i:Z_i=\hat{z}^l} (1-D_i)Y_{iy}\right] \Big/ \left(\frac{1}{n}\sum_{i=1}^{n} D_i\right)
\end{aligned}$$

Furthermore, let $\varsigma_{lb3,y}^2$, $\varsigma_{ub3,y}^2$ denote the asymptotic variances of the estimated lower and upper bounds of the average treatment on the treated parameter, respectively, and $\hat{\varsigma}_{lb3,y}^2$, $\hat{\varsigma}_{ub3,y}^2$ the corresponding consistent estimators. Then, the random set constructed as

$$\mathcal{CI3}_y^{TT} = \left[\widehat{\mathcal{LB3}}_y^{TT} - \frac{\hat{\varsigma}_{lb3,y}q_{1-\alpha}}{\sqrt{n}},\ \widehat{\mathcal{UB3}}_y^{TT} + \frac{\hat{\varsigma}_{ub3,y}q_{1-\alpha}}{\sqrt{n}}\right] \tag{31}$$

will cover asymptotically the true average treatment on the treated parameter, as defined by the bounds in (14), with probability at least $1-\alpha$.

The construction of confidence sets for the average treatment and average treatment on the treated parameters as bounded by Proposition 1 proceeds in a similar way. Consider first Δ_y^{ATE} and the bounds in (19), and let $A_y^{ATE} \equiv P(Y = y|Z = z^u) - P(Y = y|Z = z^l)$ which can be consistently estimated by

$$\hat{A}_y^{ATE} = \frac{1}{|\{i : Z_i = \hat{z}^u\}|} \sum_{i:Z_i=\hat{z}^u} Y_{iy} - \frac{1}{|\{i : Z_i = \hat{z}^l\}|} \sum_{i:Z_i=\hat{z}^l} Y_{iy}$$

Large sample results ensure that

$$\sqrt{n}\left(\hat{A}_y^{ATE} - A_y^{ATE}\right) \xrightarrow{d} N(0, \sigma_{a,y}^2)$$

where $\sigma_{a,y}^2$ denotes the variance of the asymptotic normal distribution. For the average treatment on the treated parameter and bounds (20), let $A_y^{TT} \equiv [P(Y = y) - P(Y = y|Z = z^l)]/P(D = 1)$ which can be consistently estimated by

$$\hat{A}_y^{TT} = \left[\frac{1}{n}\sum_{i=1}^{n} Y_{iy} - \frac{1}{|\{i : Z_i = \hat{z}^l\}|} \sum_{i:Z_i=\hat{z}^l} Y_{iy}\right] \Big/ \left(\frac{1}{n}\sum_i D_i\right)$$

and again, by large sample arguments

$$\sqrt{n}\left(\hat{A}_y^{TT} - A_y^{TT}\right) \xrightarrow{d} N(0, \varsigma_{a,y}^2)$$

with asymptotic variance $\varsigma_{a,y}^2$. Thus, for each of the terms in (19) and (20) a consistent estimator exists and an asymptotically valid confidence interval can be constructed.

An additional complication arises because the bounds in (19) and (20) are discontinuous functions of δ_y and δ_{y-1}. This discontinuity needs to be taken into account when constructing the random set that will asymptotically cover the true parameter with predefined probability. In order to do that, the uncertainty about δ_y should be considered as well. For an analogous argument in a nonparametric regression context see also Gijbels *et al.* (2004). Recall that δ_y was defined as the sign of the difference between two cumulative probabilities, specifically as the sign of $d_y \equiv P(Y \leq y|Z = z_1) - P(Y \leq y|Z = z_0)$ for any two evaluation points z_1, z_0 with $P(D = 1|Z = z_1) > P(D = 1|Z = z_0)$. A consistent estimator of d_y can be obtained as

$$\hat{d}_y(z_1, z_0) = \frac{1}{|\{i : Z_i = z_1\}|} \sum_{i:Z_i=z_1} \sum_{j=1}^{y} Y_{ij} - \frac{1}{|\{i : Z_i = z_0\}|} \sum_{i:Z_i=z_0} \sum_{j=1}^{y} Y_{ij}$$

with z_1, z_0 such that $\hat{P}(z_1) > \hat{P}(z_0)$. The estimator $\hat{d}_y(z_1, z_0)$ uses the information of only two evaluation points, but it is possible to account for the additional information of *all* combinations z_1, z_0 satisfying the condition $P(D = 1|Z = z_1) > P(D = 1|Z = z_0)$, which will generally improve the precision of the estimator. The modified version

$$\hat{d}_y = \frac{1}{|\{(z_1, z_0) : \hat{P}(z_1) > \hat{P}(z_0)\}|} \sum_{(z_1,z_0):\hat{P}(z_1)>\hat{P}(z_0)} \hat{d}_y(z_1, z_0) \tag{32}$$

will therefore be used in the following. The estimator in (32) can be constructed for each outcome $y \in \mathcal{Y}$, and pairs $\hat{d}_{y,y-1} = (\hat{d}_y, \hat{d}_{y-1})$ will asymptotically be bivariate normally distributed with

$$\sqrt{n}\left(\hat{d}_{y,y-1} - d_{y,y-1}\right) \xrightarrow{d} N(0, \Sigma_{y,y-1})$$

The asymptotic covariance matrix $\Sigma_{y,y-1}$ has $Var(\hat{d}_y)$ and $Var(\hat{d}_{y-1})$ the main diagonal entries, and $Cov(\hat{d}_y, \hat{d}_{y-1})$ the off-diagonal entries. An asymptotic confidence ellipse for $d_{y,y-1}$ can be constructed as

$$(\hat{d}_{y,y-1} - d_{y,y-1})'\hat{\Sigma}^{-1}_{y,y-1}(\hat{d}_{y,y-1} - d_{y,y-1}) \leq \chi^2_{2,1-\alpha} \tag{33}$$

where $\hat{\Sigma}_{y,y-1}$ is a consistent estimator of $\Sigma_{y,y-1}$, and $\chi^2_{2,1-\alpha}$ is the $1-\alpha$ quantile of the Chi-square distribution with two degrees of freedom. For n growing large, the ellipse defined by (33) will cover the true $d_{y,y-1}$ with probability $1 - \alpha$.

The confidence sets for the average treatment effects and the average treatment on the treated effects as defined by Proposition 1 can then be constructed as follows. In the d_y, d_{y-1}-plane (where d_y is on the abscissa and d_{y-1} is on the ordinate), if the confidence ellipse defined by (33)

1. ... lies entirely in the fourth quadrant (d_y positive, d_{y-1} negative), or intersects with the abscissa ($d_{y-1} = 0$) only in the first/fourth quadrant, or intersects with the ordinate ($d_y = 0$) only in the third/fourth quadrant, then use the random set

$$\begin{aligned} \mathcal{CI4}a_y^{ATE} &= \left[\hat{A}_y^{ATE} - \frac{\hat{\sigma}_{a,y}q_{1-\alpha}}{\sqrt{n}},\ \widehat{\mathcal{UB}3}_y^{ATE} + \frac{\hat{\sigma}_{ub3,y}q_{1-\alpha}}{\sqrt{n}}\right] \\ \mathcal{CI4}a_y^{TT} &= \left[\hat{A}_y^{TT} - \frac{\hat{\varsigma}_{a,y}q_{1-\alpha}}{\sqrt{n}},\ \widehat{\mathcal{UB}3}_y^{TT} + \frac{\hat{\varsigma}_{ub3,y}q_{1-\alpha}}{\sqrt{n}}\right] \end{aligned}$$

2. ... intersects with both axes, then use the random set

$$
\begin{aligned}
\mathcal{CI4}\,b_y^{ATE} &= \left[\widehat{\mathcal{LB3}}_y^{ATE} - \frac{\hat{\sigma}_{lb3,y} q_{1-\alpha}}{\sqrt{n}},\ \widehat{\mathcal{UB3}}_y^{ATE} + \frac{\hat{\sigma}_{ub3,y} q_{1-\alpha}}{\sqrt{n}}\right] \\
\mathcal{CI4}\,b_y^{TT} &= \left[\widehat{\mathcal{LB3}}_y^{TT} - \frac{\hat{\varsigma}_{lb3,y} q_{1-\alpha}}{\sqrt{n}},\ \widehat{\mathcal{UB3}}_y^{TT} + \frac{\hat{\varsigma}_{ub3,y} q_{1-\alpha}}{\sqrt{n}}\right]
\end{aligned}
$$

3. ... lies entirely in the second quadrant (d_y negative, d_{y-1} positive), or intersects with the abscissa ($d_{y-1} = 0$) only in the second/third quadrant, or intersects with the ordinate ($d_y = 0$) only in the first/second quadrant, then use the random set

$$
\begin{aligned}
\mathcal{CI4}\,c_y^{ATE} &= \left[\widehat{\mathcal{LB3}}_y^{ATE} - \frac{\hat{\sigma}_{lb3,y} q_{1-\alpha}}{\sqrt{n}},\ \hat{A}_y^{ATE} + \frac{\hat{\sigma}_{a,y} q_{1-\alpha}}{\sqrt{n}}\right] \\
\mathcal{CI4}\,c_y^{TT} &= \left[\widehat{\mathcal{LB3}}_y^{TT} - \frac{\hat{\varsigma}_{lb3,y} q_{1-\alpha}}{\sqrt{n}},\ \hat{A}_y^{TT} + \frac{\hat{\varsigma}_{a,y} q_{1-\alpha}}{\sqrt{n}}\right]
\end{aligned}
$$

4. ... lies entirely in the first quadrant (both d_y and d_{y-1} are positive), or entirely in the third quadrant (both d_y and d_{y-1} are negative), then use the random set

$$
\begin{aligned}
\mathcal{CI4}\,d_y^{ATE} &= \left[\widehat{\mathcal{LB3}}_y^{ATE} - \frac{\hat{\sigma}_{lb3,y} q_{1-\alpha}}{\sqrt{n}},\ \widehat{\mathcal{UB3}}_y^{ATE} + \frac{\hat{\sigma}_{ub3,y} q_{1-\alpha}}{\sqrt{n}}\right] \\
\mathcal{CI4}\,d_y^{TT} &= \left[\widehat{\mathcal{LB3}}_y^{TT} - \frac{\hat{\varsigma}_{lb3,y} q_{1-\alpha}}{\sqrt{n}},\ \widehat{\mathcal{UB3}}_y^{TT} + \frac{\hat{\varsigma}_{ub3,y} q_{1-\alpha}}{\sqrt{n}}\right]
\end{aligned}
$$

One can show that asymptotically the random sets $\mathcal{CI4}_y^{ATE}$, consisting of $\mathcal{CI4}\,a_y^{ATE}$ to $\mathcal{CI4}\,d_y^{ATE}$, and $\mathcal{CI4}_y^{TT}$, consisting of $\mathcal{CI4}\,a_y^{TT}$ to $\mathcal{CI4}\,d_y^{TT}$, cover the true average treatment effect and the average treatment effect on the treated, respectively, with probability at least $1-\alpha$. In order to see why the intervals constructed as such will cover the true parameter with probability at least $1-\alpha$, consider the average treatment effect and assume that $\Delta_y^{ATE} > 0$ such that $\delta_y > \delta_{y-1}$ and $\Delta_y^{ATE} \in \left[A_y^{ATE}, \mathcal{UB3}_y^{ATE}\right]$. With probability approaching one, the confidence interval constructed in (33) will fulfill the conditions to choose $\mathcal{CI4}\,a_y^{ATE}$, and $\mathcal{CI4}\,a_y^{ATE}$ covers all parameters $\Delta_y^{ATE} \in \left[A_y^{ATE}, \mathcal{UB3}_y^{ATE}\right]$ with probability at least $1-\alpha$, as desired. Analogous arguments show that in all other cases the desired coverage probability is obtained.

The confidence sets for the average treatment and the average treatment on the treated effects in the case of X covariates present, i.e., for the parameters as identified by Proposition 2, can be constructed following a similar strategy as in the case of no X covariates available. The steps involved are as follows. To begin with, let

$$\hat{P}(x,z) = \frac{1}{|\{i : X_i = x, Z_i = z\}|} \sum_{i:X_i=x,Z_i=z} D_i$$

denote a consistent estimator of $P(D = 1|X = x, Z = z)$, and define

$$\begin{aligned} d_y(x_1, x_0; z_1, z_0) \;\equiv& \\ &[P(D=1, Y \leq y|X = x_1, Z = z_1) - P(D=1, Y \leq y|X = x_1, Z = z_0)] \\ &-[P(D=0, Y \leq y|X = x_0, Z = z_0) - P(D=0, Y \leq y|X = x_0, Z = z_1)] \end{aligned}$$

which can be consistently estimated by

$$\begin{aligned} \hat{d}_y(x_1, x_0; z_1, z_0) \;=\;& \frac{1}{|\{i : X_i = x_1, Z_i = z_1\}|} \sum_{i:X_i=x_1,Z_i=z_1} \sum_{j=1}^{y} D_i Y_{ij} \\ &- \frac{1}{|\{i : X_i = x_1, Z_i = z_0\}|} \sum_{i:X_i=x_1,Z_i=z_0} \sum_{j=1}^{y} D_i Y_{ij} \\ &- \left[\frac{1}{|\{i : X_i = x_0, Z_i = z_0\}|} \sum_{i:X_i=x_0,Z_i=z_0} \sum_{j=1}^{y} (1-D_i) Y_{ij} \right. \\ &\left. - \frac{1}{|\{i : X_i = x_0, Z_i = z_1\}|} \sum_{i:X_i=x_0,Z_i=z_1} \sum_{j=1}^{y} (1-D_i) Y_{ij} \right] \end{aligned}$$

with z_1, z_0 such that $\hat{P}(x_j, z_1) > \hat{P}(x_j, z_0)$, $j = 0, 1$. Accounting for the information of all such evaluation points z_1, z_0 yields the estimator

$$\hat{d}_y(x_1, x_0) = \frac{\sum_{(z_1,z_0):\hat{P}(x_j,z_1)>\hat{P}(x_j,z_0),j=0,1} \hat{d}_y(x_1, x_0; z_1, z_0)}{|\{(z_1, z_0) : \hat{P}(x_j, z_1) > \hat{P}(x_j, z_0), j = 0, 1\}|}$$

Then consider the estimator either as a function of x_1 keeping x_0 fixed, $\hat{d}_y(x_1|x_0)$, or as a function of x_0 keeping x_1 fixed, $\hat{d}_y(x_0|x_1)$, and note that the estimators hold for all $y \in \mathcal{Y}$ such that pairs $\left(\hat{d}_y(x_1|x_0), \hat{d}_{y-1}(x_1|x_0)\right)$ or $\left(\hat{d}_y(x_0|x_1), \hat{d}_{y-1}(x_0|x_1)\right)$ can be created. From these pairs, one may construct asymptotically valid confidence ellipses with regions as defined in the construction of $\mathcal{CI}4$ (entirely in each quadrant, and the 5 intersection possibilities with the axes).

The bounds $L\mathcal{B}^1_y(x)$, $U\mathcal{B}^1_y(x)$, $L\mathcal{B}^0_y(x)$, $U\mathcal{B}^0_y(x)$ depend on the sets $\mathcal{X}^l_0(x)$, $\mathcal{X}^u_0(x)$, $\mathcal{X}^l_1(x)$, $\mathcal{X}^u_1(x)$, the latter defined by $\delta_y(x_1, x_0)$ relative to $\delta_{y-1}(x_1, x_0)$. This dependence needs to be taken into account when constructing the confidence sets for the parameters. Let $\mathcal{AX}_0^l(x_1)$ denote an alternative set of all x_0 (given x_1) satisfying that in the $d_y(x_0|x_1), d_{y-1}(x_0|x_1)$-plane the confidence ellipse lies entirely in the fourth quadrant ($d_y(x_0|x_1)$ positive, $d_{y-1}(x_0|x_1)$ negative), or intersects with the abscissa ($d_{y-1}(x_0|x_1) = 0$) only in the first/fourth quadrant, or intersects with the ordinate ($d_y(x_0|x_1) = 0$) only in the third/fourth quadrant. Similarly, define $\mathcal{AX}_0^u(x_1)$ as an alternative set of all x_0 (given x_1) satisfying that in the $d_y(x_0|x_1), d_{y-1}(x_0|x_1)$-plane the confidence ellipse lies entirely in the second quadrant ($d_y(x_0|x_1)$ negative, $d_{y-1}(x_0|x_1)$ positive), or intersects with the abscissa ($d_{y-1}(x_0|x_1) = 0$) only in the second/third quadrant, or intersects with the ordinate ($d_y(x_0|x_1) = 0$) only in the first/second quadrant. Analogously, define sets $\mathcal{AX}_1^l(x_0)$ and $\mathcal{AX}_1^u(x_0)$ in the $d_y(x_1|x_0), d_{y-1}(x_1|x_0)$-plane. These alternative sets can be interpreted as estimators of the population sets $\mathcal{X}_j^k(x)$, $j = 0, 1$, $k = l, u$.

Empirical analogues of the upper and lower bounds on $P(Y_1 = y|X = x)$ and $P(Y_0 = y|X = x)$ can be derived from (26) and (27) replacing the population sets by the alternative sets defined above and the probabilities by the appropriate relative frequencies. From these estimators, one may construct estimators of the upper and lower bounds for the average treatment effect and the average treatment effect on the treated. Because of the dependence on the (estimated) alternative sets, obtaining an upper bound of an one-sided $1-\alpha$ confidence interval for $U\mathcal{B}5_y^{ATE}(x)$, and a lower bound of an one-sided $1-\alpha$ confidence interval for $L\mathcal{B}5_y^{ATE}(x)$ is not straightforward. One option, also referred to by Shaikh and Vytlacil (2005), is subsampling; see Politis *et al.* (1999) for details, in particular Chapter 2. Let $\widehat{L\mathcal{B}5}_{y;1-\alpha^-}^{ATE}(x)$ and $\widehat{U\mathcal{B}5}_{y;1-\alpha^+}^{ATE}(x)$ denote such bounds of the confidence interval, then an asymptotically valid confidence interval for the average treatment effect can be obtained by

$$\mathcal{CI}5_y^{ATE} = \left[\widehat{L\mathcal{B}5}_{y;1-\alpha^-}^{ATE}(x), \widehat{U\mathcal{B}5}_{y;1-\alpha^+}^{ATE}(x)\right] \tag{34}$$

By analogous arguments, an asymptotically valid confidence interval

$$\mathcal{CI}5_y^{TT} = \left[\widehat{L\mathcal{B}5}_{y;1-\alpha^-}^{TT}(x), \widehat{U\mathcal{B}5}_{y;1-\alpha^+}^{TT}(x)\right] \tag{35}$$

for the average treatment on the treated effect can be constructed.

5 Moving Beyond ATE and TT

The previous sections have focused on two treatment parameters, namely the average treatment effect and the average treatment effect on the treated. Both parameters were defined in terms of probabilities rather than expectations to circumvent the problem of ordinal but arbitrary coding of the elements in $\mathcal{Y}$. The term "average" was introduced because the parameters reflect how an individual's probability of responding in each of the J ordinal categories will change with and without the receipt of treatment, and where probability was defined from a frequentist perspective as what would happen on average if the same individual was considered repeatedly.

The average treatment effect and the average treatment effect on the treated certainly are the treatment parameters that occur most often in the literature. The former is defined for an individual that is randomly drawn from the entire population of interest, the latter is defined for an individual randomly drawn from those that actually received the treatment. However, alternative parameters have been considered as well for different subgroups of the population. For example, the local average treatment effect (LATE) of Imbens and Angrist (1994) is defined as the average treatment effect for the subgroup of compliers, i.e., those individuals who would comply with the exogenous modification of instruments. This concept can also be translated to probabilities. Let z_1, z_0 denote two evaluation points with $P(D = 1|Z = z_1) > P(D = 0|Z = z_0)$ such that, by the threshold crossing treatment selection, $s(z_1) > s(z_0)$. Then,

$$\begin{aligned}
& P(Y = y|Z = z_1) - P(Y = y|Z = z_0) \qquad (36)\\
= \; & P(D = 1, Y = y|Z = z_1) + P(D = 0, Y = y|Z = z_1)\\
& -P(D = 1, Y = y|Z = z_0) - P(D = 0, Y = y|Z = z_0)\\
= \; & P(\nu \leq s(z_1), Y_1 = y) + P(\nu > s(z_1), Y_0 = y)\\
& -P(\nu \leq s(z_0), Y_1 = y) - P(\nu > s(z_0), Y_0 = y)\\
= \; & P(s(z_0) < \nu \leq s(z_1), Y_1 = y) - P(s(z_0) < \nu \leq s(z_1), Y_0 = y)\\
= \; & \Big[P(Y_1 = y|s(z_0) < \nu \leq s(z_1)) - P(Y_0 = y|s(z_0) < \nu \leq s(z_1))\Big] P(s(z_0) < \nu \leq s(z_1))
\end{aligned}$$

where the first equality follows by the law of total probability, the second equality follows by the observation rule in (1), the threshold crossing treatment selection, and the independence assumption (A4), the third equality follows by $s(z_1) > s(z_0)$, and the last equality follows by Bayes' theorem.

From (36) define the local average treatment effect as

$$\begin{aligned}\Delta_y^{LATE}(z_1, z_0) &\equiv P(Y_1 = y|s(z_0) < \nu \leq s(z_1)) - P(Y_0 = y|s(z_0) < \nu \leq s(z_1)) \\ &= \frac{P(Y = y|Z = z_1) - P(Y = y|Z = z_0)}{P(s(z_0) < \nu \leq s(z_1))} \\ &= \frac{P(Y = y|Z = z_1) - P(Y = y|Z = z_0)}{P(\nu \leq s(z_1)) - P(\nu \leq s(z_0))} \\ &= \frac{P(Y = y|Z = z_1) - P(Y = y|Z = z_0)}{P(D = 1|Z = z_1) - P(D = 1|Z = z_0)}\end{aligned} \tag{37}$$

where the second equality follows by the derivation above, and the last equalities follow by the assumptions of the treatment selection model. Thus, the local average treatment effect gives the change in the probability distribution for those individuals who would not select into treatment if Z was externally set to z such that $s(z) \leq s(z_0)$, and who would select into treatment if Z was externally set to z such that $s(z) \geq s(z_1)$. An important aspect of the local average treatment effect is that it is identified from the population distribution of (Y, D, Z) for all combinations z_1, z_0 with $P(D = 1|Z = z_1) > P(D = 1|Z = z_0)$, which is made explicit in the definition of $\Delta_y^{LATE}(z_1, z_0)$ including z_1 and z_0 in the argument.

A marginal version of the local average treatment effect has been introduced in Heckman (1997). Consider the limit $s(z_0) \to s(z_1)$ of (37) and define the marginal treatment effect as

$$\Delta_y^{MTE}(z_1) \equiv P(Y_1 = y|\nu = s(z_1)) - P(Y_0 = y|\nu = s(z_1)) \tag{38}$$

Thus, the marginal treatment effect gives the change in the probability distribution for those individuals that would just be indifferent between being selected into or out of the treatment if Z was externally set to z such that $s(z) = s(z_1)$. Starting from (38), one can show that the other treatment parameters, Δ_y^{ATE}, Δ_y^{TT}, and Δ_y^{LATE}, are integrated versions of Δ_y^{MTE} over different intervals and with different weighting functions (Heckman and Vytlacil 2001). An estimator of Δ_y^{MTE} can be obtained by $\partial P(Y = y|Z = z_1)/\partial P(D = 1|Z = z_1)$ given that the derivative exists and is finite in a small neighborhood of z_1. Since both Δ_y^{MTE} and Δ_y^{LATE} are identified, identification of Δ_y^{ATE} and Δ_y^{TT} in principle is possible. However, this requires observability of a sufficiently large support of $P(D = 1|Z = z)$, which must not necessarily hold in practice, and therefore the bounding analysis of Section 3 is more general by imposing identification regions for the treatment parameters.

While the previous treatment parameters were defined for different subgroups of the population, the ordinal nature of the response variable allows for a more thorough analysis of the effect on the outcome distribution, either in the entire population or in the subgroup of treated individuals. In particular, analyzing probabilities rather than expectations provides a much richer set of treatment parameters beyond the common mean effects. For example, consider the concept of stochastic order (SO) in two random variables (Mann and Whitney 1947). Let

$$\Delta_y^{SO} \equiv P(Y_1 \leq y) - P(Y_0 \leq y) \tag{39}$$

If $\Delta_y^{SO} \leq 0$ for all y, then Y_0 is said to be stochastically smaller than Y_1, i.e., Y_0 tends to have higher probability for low y, and smaller probability for high y compared to Y_1. Analogously, if $\Delta_y^{SO} \geq 0$ for all y, then Y_0 is said to be stochastically larger than Y_1, and if $\Delta_y^{S0} = 0$ for all y, then Y_0 and Y_1 are said to be stochastically equivalent. Clearly, one may also analyze the stochastic order of Y_1 and Y_0 in the subgroup of the treated (SOT)

$$\Delta_y^{SOT} \equiv P(Y_1 \leq y|D = 1) - P(Y_0 \leq y|D = 1) \tag{40}$$

where, for example, Y_1 is said to be stochastically larger than Y_0, now conditional on $D = 1$, if $\Delta_y^{SOT} \leq 0$ for all y. If neither of the three cases is true for all y, i.e., Y_1 is not stochastically larger or smaller than, nor equivalent to Y_0, then one may at least analyze the degree of stochastic order starting from $y = 1$ moving to $y = J$, or the other way round.

Yet another way to look at the effect of treatment on the outcome distribution, related to the concept of stochastic ordering, is in terms of the relative odds, specifically,

$$\Omega_y \equiv \frac{P(Y_0 \leq y)/P(Y_0 > y)}{P(Y_1 \leq y)/P(Y_1 > y)} \tag{41}$$

and

$$\Omega_y^T \equiv \frac{P(Y_0 \leq y|D = 1)/P(Y_0 > y|D = 1)}{P(Y_1 \leq y|D = 1)/P(Y_1 > y|D = 1)} \tag{42}$$

These parameters show the factor by which the ratio of the odds $Y_0 \leq y$ relative to $Y_0 > y$ in the non-treatment group change compared to the odds $Y_1 \leq y$ relative to $Y_1 > y$ in the treatment group. With a positive treatment effect, i.e., the probability of higher outcomes increases with the receipt of treatment, this factor should be larger than one. If, on the other hand, the treatment effect is negative, then the odds ratio is smaller than one, and if the treatment effect is zero, then the odds ratio is one. Note that there exist $J - 1$ odds ratios, one for each $y = 1, \ldots, J - 1$.

Neither the stochastic order parameters, nor the odds ratios are immediately identified from the population distribution of (Y, D, Z), by the same argument as the average treatment and the average treatment on the treated are not identified. However, one may impose bounds on the unidentified probabilities and thus impose bounds on the parameters in (39)-(42).

6 Conclusion

The properties of ordinally measured variables, in a strict sense, require the shift in focus from mean treatment effects to probability treatment effects. Parametric ordered response models to estimate such effects already exist and are typically based on threshold crossing mechanisms. This is the first paper, to the best of my knowledge, that discovers the informational content of a threshold crossing mechanism in a nonparametric bounding analysis with ordinal potential outcomes; only Scharfstein *et al.* (2004) consider bounds on treatment effects with ordinal responses, but in a very particular prospective data situation.

The approach taken here is closely related to Shaikh and Vytlacil (2005), who consider a model with binary instead of ordinal outcomes, and the results obtained here therefore complement their work. The extension to ordinal outcomes requires a slightly different identification and bounding strategy, where multiple thresholds need to be taken into account. As a central result, the imposed bounds always identify whether the treatment effect is positive, zero, or negative, although point-identification except for the zero treatment effect fails in the nonparametric setting. It is interesting to note that an additional set of parameters becomes available with ordinal outcomes that might be of interest in evaluating the effect of a treatment.

References

Angrist, J.D., G.W. Imbens, and D. Rubin (1996): "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91, 444-455.

Bellemare, C., B. Melenberg, and A. van Soest (2002): "Semi-parametric models for satisfaction with income," Portuguese Economic Journal, 1, 181-203.

Boes, S., and R. Winkelmann (2006): "Ordered Response Models," *Allgemeines Statistisches Archiv*, 90, 165-180.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, forthcoming.

Coppejans, M. (2007): "On efficient estimation of the ordered response model," *Journal of Econometrics*, 137, 577-614.

Gijbels, I., P. Hall, and A. Kneip (2004): "Interval and band estimation for curves with jumps," *Journal of Applied Probability*, 41A, 65-79.

Heckman, J.J. (1997): "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations," *Journal of Human Resources*, 32, 441-462.

Heckman, J.J., R.J. LaLonde, and J.A. Smith (1999): "The Economics and Econometrics of Active Labor Market Programs," in O. Ashenfelter, and D. Card (eds.), *Handbook of Labor Economics Volume III.* Amsterdam: North Holland, 1865-2097

Heckman, J.J., and R. Robb (1985): "Alternative Methods for Estimating the Impact of Interventions," in J.J. Heckman, and B. Singer (eds.) *Longitudinal Analysis of Labor Market Data*, New York: Cambridge University Press, 156-245.

Heckman, J.J., and R. Robb (1986): "Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes," in J. Wainer (ed.) *Drawing Inference From Self-Selected Samples*, New York: Springer-Verlag, 63-107.

Heckman, J.J., and E. Vytlacil (2001): "Local Instrumental Variables," in C. Hsiao, K. Morimune, and J. Powell (eds.) *Nonlinear Statistical Inference: Essays in Honor of Takeshi Amemiya*, Cambridge: Cambridge University Press.

Horowitz, J., and C.F. Manski (2000): "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data," *Journal of the American Statistical Association*, 95, 77-84.

Imbens, G.W., and J.D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467-476.

Imbens, G.W., and C.F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845-1857.

Klein, R.W., and Sherman, R.P. (2002): "Shift restrictions and semiparametric estimation in ordered response models," *Econometrica*, 70, 663-692.

Lewbel, A. (1997): "Semiparametric estimation of location and other discrete choice moments," *Econometric Theory*, 13, 32-51.

Lewbel, A. (2003): "Ordered response threshold estimation," *unpublished working paper.*

Li, M., and J.L. Tobias (2007): "Bayesian Analysis of Treatment Effects in an Ordered Potential Outcomes Model in D. Millimet, J. Smith, and E. Vytlacil (eds.) *Advances in Econometrics, Volume 21: Estimating and Evaluating Treatment Effects in Econometrics*, forthcoming.

Mann, H.B., and D.R. Whitney (1947): "On a test of whether one of two random variables is stochastically larger than the other," *Annals of Mathematical Statistics*, 18, 50-60.

Manski, C.F. (1990): "Nonparametric Bounds on Treatment Effects," *American Economic Review, Papers and Proceedings*, 80, 319-323.

Manski, C.F. (1994): "The Selection Problem," in C. Manski, and D. McFadden (eds.) *Advances in Econometrics: Sixth World Congress*, Cambridge University Press, Cambridge.

Manski, C.F. (1995): *Identification Problems in the Social Sciences*, Harvard University Press, Cambridge and London.

Manski, C.F. (2000): "Identification Problems and Decisions Under Ambiguity: Empirical Analysis of Treatment Response and Normative Analysis of Treatment Choice," *Journal of Econometrics*, 95, 415-442.

Manski, C.F. (2003): *Partial Identification of Probabilities Distributions*, Springer: New York and Heidelberg.

McCullagh, P. (1980): "Regression Models for Ordinal Data," *Journal of the Royal Statistical Society. Series B (Methodological)*, 42, 109-142.

McKelvey, R.D. and W. Zavoina (1975): "A statistical model for the analysis of ordinal level dependent variables," *Journal of Mathematical Sociology*, 4, 103-120.

Neyman, J. (1923): "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9," translated in *Statistical Science*, 5, 465-480.

Politis, D.N., J.P. Romano, and M. Wolf (1999): *Subsampling*, Springer, New York.

Rubin, D.B. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66, 688-701.

Scharfstein, D.O., C.F. Manski, and J.C. Anthony (2004): "On the Construction of Bounds in Prospective Studies with Missing Ordinal Outcomes: Application to the Good Behavior Game Trial," *Biometrics*, 60, 154-164.

Shaikh, A.M., and E. Vytlacil (2005): "Threshold Crossing Models and Bounds on Treatment Effects: A Nonparametric Analysis," *unpublished working paper.*

Stewart, M.B (2004): "Semi-nonparametric Estimation of Extended Ordered Probit Models," *Stata Journal*, 4, 27-39.

Vytlacil, E. (2002): "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 70, 331-341.

Vytlacil, E. (2006): "A Note on Additive Separability and Latent Index Models of Binary Choice: Representation Results," Oxford Bulletin of Economics and Statistics, 68, 515-518.

Working Papers of the Socioeconomic Institute at the University of Zurich

The Working Papers of the Socioeconomic Institute can be downloaded from http://www.soi.uzh.ch/research/wp/index2.html

0709	Nonparametric Analysis of Treatment Effects in Ordered Response Models, Stefan Boes, July 2007, 42p.
0708	Rationality on the Rise: Why Relative Risk Aversion Increases with Stake Size, Helga Fehr-Duda, Adrian Bruhin, Thomas F. Epper, and Renate Schubert, July 2007, 30p.
0707	I'm not fat, just too short for my weight – Family Child Care and Obesity in Germany, Philippe Mahler, May 2007, 27p.
0706	Does Globalization Create Superstars?, Hans Gersbach and Armin Schmutzler , April 2007, 23 p.
0705	Risk and Rationality: Uncovering Heterogeneity in Probability Distortion, Adrian Bruhin, Helga Fehr-Duda, and Thomas F. Epper, July 2007 (revised version), 29p.
0704	Count Data Models with Unobserved Heterogeneity: An Empirical Likelihood Approach, Stefan Boes, March 2007, 26p.
0703	Risk and Rationality: The Effect of Incidental Mood on Probability Weighting, Helga Fehr, Thomas Epper, Adrian Bruhin, and Renate Schubert, February 2007, 27p.
0702	Happiness Functions with Preference Interdependence and Heterogeneity: The Case of Altruism within the Family, Adrian Bruhin and Rainer Winkelmann, February 2007, 20p.
0701	On the Geographic and Cultural Determinants of Bankruptcy, Stefan Buehler, Christian Kaiser, and Franz Jaeger, June 2007 (revised version), 35 p.
0610	A Product-Market Theory of Industry-Specific Training, Hans Gersbach and Armin Schmutzler , November 2006, 28 p.
0609	Entry in liberalized railway markets: The German experience, Rafael Lalive and Armin Schmutzler, April 2007, revised version, 20 p.
0608	The Effects of Competition in Investment Games, Dario Sacco and Armin Schmutzler, April 2007, revised version, 22 p.
0607	Merger Negotiations and Ex-Post Regret, Dennis Gärtner and Armin Schmutzler, September 2006, 28 p.
0606	Foreign Direct Investment and R&D offshoring, Hans Gersbach and Armin Schmutzler, June 2006, 34 p.
0605	The Effect of Income on Positive and Negative Subjective Well-Being, Stefan Boes and Rainer Winkelmann, May 2006, 23p.
0604	Correlated Risks: A Conflict of Interest Between Insurers and Consumers and Its Resolution, Patrick Eugster and Peter Zweifel, April 2006, 23p.
0603	The Apple Falls Increasingly Far: Parent-Child Correlation in Schooling and the Growth of Post-Secondary Education in Switzerland, Sandra Hanslin and Rainer Winkelmann, March 2006, 24p.
0602	Efficient Electricity Portfolios for Switzerland and the United States, Boris Krey and Peter Zweifel, February 2006, 25p.
0601	Ain't no puzzle anymore: Comparative statics and experimental economics, Armin Schmutzler, February 2006, 45 p.
0514	Money Illusion Under Test, Stefan Boes, Markus Lipp and Rainer Winkelmann, November 2005, 7p.
0513	Cost Sharing in Health Insurance: An Instrument for Risk Selection? Karolin Becker and Peter Zweifel, November 2005, 45p.
0512	Single Motherhood and (Un)Equal EducationalOpportunities: Evidence for Germany, Philippe Mahler and Rainer Winkelmann, September 2005, 23p.

0511 Exploring the Effects of Competition for Railway Markets, Rafael Lalive and Armin Schmutzler, April 2007, revised version, 33p.
0510 The Impact of Aging on Future Healthcare Expenditure; Lukas Steinmann, Harry Telser, and Peter Zweifel, September 2005, 23p.
0509 The Purpose and Limits of Social Health Insurance; Peter Zweifel, September 2005, 28p.
0508 Switching Costs, Firm Size, and Market Structure; Simon Loertscher and Yves Schneider, August 2005, 29p.
0507 Ordered Response Models; Stefan Boes and Rainer Winkelmann, March 2005, 21p.
0506 Merge or Fail? The Determinants of Mergers and Bankruptcies in Switzerland, 1995-2000; Stefan Buehler, Christian Kaiser, Franz Jaeger, March 2005, 18p.
0505 Consumer Resistance Against Regulation: The Case of Health Care Peter Zweifel, Harry Telser, and Stephan Vaterlaus, February 2005, 23p.
0504 A Structural Model of Demand for Apprentices Samuel Mühlemann, Jürg Schweri, Rainer Winkelmann and Stefan C. Wolter, February 2005, 25p.
0503 What can happiness research tell us about altruism? Evidence from the German Socio-Economic Panel Johannes Schwarze and Rainer Winkelmann, February 2005, 26p.
0502 Spatial Effects in Willingness-to-Pay: The Case of Nuclear Risks Peter Zweifel, Yves Schneider and Christian Wyss, January 2005, 37p.
0501 On the Role of Access Charges Under Network Competition Stefan Buehler and Armin Schmutzler, January 2005, 30p.
0416 Social Sanctions in Interethnic Relations: The Benefit of Punishing your Friends Christian Stoff, Dezember 2004, 51p.
0415 Single Motherhood and (Un)equal Educational Opportunities: Evidence from Germany, Philippe Mahler and Rainer Winkelmann, November 2004, 23p.
0414 Are There Waves in Merger Activity After All? Dennis Gärtner and Daniel Halbheer, September 2004, 39p.
0413 Endogenizing Private Information: Incentive Contracts under Learning By Doing Dennis Gärtner, September 2004, 32p.
0412 Validity and Reliability of Willingness-to-pay Estimates: Evidence from Two Overlapping Discrete-Choice Experiments Harry Telser, Karolin Becker and Peter Zweifel. September 2004, 25p.
0411 Willingness-to-pay Against Dementia: Effects of Altruism Between Patients and Their Spouse Caregivers, Markus König und Peter Zweifel, September 2004, 22p.
0410 Age and Choice in Health Insurance: Evidence from Switzerland Karolin Becker and Peter Zweifel, August 2004, 30p.
0409 Vertical Integration and Downstream Investment in Oligopoly Stefan Buehler and Armin Schmutzler, July 2004, 30p.
0408 Mergers under Asymmetric Information – Is there a Lemons Problem? Thomas Borek, Stefan Buehler and Armin Schmutzler, July 2004, 38p.
0407 Income and Happiness: New Results from Generalized Threshold and Sequential Models, Stefan Boes and Rainer Winkelmann, June 2004, 30p.
0406 Optimal Insurance Contracts without the Non-Negativity Constraint on Indemnities Revisited, Michael Breuer, April 2004, 17p.
0405 Competition and Exit: Evidence from Switzerland Stefan Buehler, Christian Kaiser and Franz Jaeger, March 2004, 28p.