On Quantile Treatment Effects, Rank Similarity, and the Variation of Instrumental Variables*

Sukjin Han

School of Economics

University of Bristol

sukjin.han@gmail.com

Haiqing Xu

Department of Economics

University of Texas at Austin

h.xu@austin.utexas.edu

February 22, 2023

Abstract

This paper investigates how certain relationship between observed and counterfactual distributions plays a role in the identification of distributional treatment effects under endogeneity, and shows that this relationship holds in a range of nonparametric models for treatment effects. To motivate the new identifying condition, we first provide a novel characterization of popular assumptions restricting treatment heterogeneity in the literature, specifically rank similarity. We show the stringency of this type of assumptions and propose to relax them in economically meaningful ways. This relaxation will justify certain parameters (e.g., treatment effects on the treated) against others (e.g., treatment effects for the entire population). It will also justify the quest of richer exogenous variation in the data (e.g., the use of multi-valued or multiple instrumental variables). The prime goal of this investigation is to provide empirical researchers with tools for identifying and estimating treatment effects that are flexible

1

enough to allow for treatment heterogeneity but still yield tight policy evaluation and are easy to implement.

1   Introduction

This paper investigates how certain relationship between observed and counterfactual distributions plays a role in the identification of distributional treatment effects under endogeneity and shows that this relationship holds in a range of nonparametric models for treatment effects. To motivate the new identifying condition, we first provide a novel characterization of popular assumptions restricting treatment heterogeneity in the literature, specifically the rank similarity assumptions. This characterization allows us to show the stringency of the assumptions and motivates us ways of relaxing them in economically meaningful ways. This relaxation will justify the quest of richer exogenous variation in the data (e.g., the use of multi-valued or multiple instrumental variables) for identification.

The prime goal of this investigation is to provide empirical researchers with (i) a framework where validity of identifying conditions prescribes the parameters of interest, (ii) tools for identifying and estimating treatment effects that are flexible enough to allow for treatment heterogeneity, but that still yield tight policy evaluation and are easy to implement, and (iii) guidance on data collection that leads to drawing meaningful causal conclusions.

We consider a mapping between observed and counterfactual distributions. A version of this mapping can be informally stated in terms of the following preservation of first order stochastic dominance (FOSD): for arbitrary compliance types $t, t' \in \mathcal{T}$ induced by instrumental variables (IVs), if

$$Y_1|t \prec_{FOSD} Y_1|t' \tag{1.1}$$

then

$$Y_0|t \prec_{FOSD} Y_0|t', \tag{1.2}$$

where Y_d denotes the counterfactual outcome given treatment $D = d.$[1] This condition relates the observed distributions to the counterfactual distributions. For the sake of illustration, consider binary instrument $Z \in \{0, 1\}$ and $\mathcal{T} = \{C, AT, NT\}$ where C, AT, and NT stand for compliers, always-takers, and never-takers, respectively. Let Y be the observed outcome given by $Y \equiv DY_1 + (1 - D)Y_0$. A simple algebra shows that $Y_1|AT \prec_{FOSD} Y_1|C$ can be expressed as

$$P[Y \leq y|D = 1, Z = 0] \leq P[Y \leq y|D = 1, Z = 1] \quad \text{for all } y \tag{1.3}$$

and $Y_0|AT \prec_{FOSD} Y_0|C$ can be expressed as

$$P[Y_0 \leq y|D = 1] \leq \frac{P[Y \leq y, D = 0|Z = 1] - P[Y \leq y, D = 0|Z = 0]}{P[D = 1|Z = 0] - P[D = 1|Z = 1]} \quad \text{for all } y. \tag{1.4}$$

Then, (1.4) produces an informative upper bound for $P[Y_0 \leq y|D = 1]$ and a symmetric analysis produces a lower bound. Note $P[Y_0 \leq y|D = 1]$ is a key component in calculating the distributional treatment effect, such as the quantile treatment effect on the treated (QTT). We also provide analogous conditions to bound average treatment effects. The proposed bounds can be much tighter than other bounds available in the literature but is not guaranteed without (1.3) (or (1.1)) being satisfied, which can be verified from the data. Although (1.1) may seem too restrictive to hold in the data, this is *not* generally the case when Z departs from a scalar binary variable. One of the main messages we hope to deliver in this paper is precisely that: the need of Z that takes multiple values (or equivalently the need of multi-valued IVs). In that case, (1.1) is likely to satisfy and we can obtain informative bounds

---

[1]For r.v.'s A and B, let $A \prec_{FOSD} B$ denotes $F_B(t) \leq F_A(t)$ where F_A and F_B are CDFs of A and B, respectively.

on the distributional treatment effects under relatively weak assumptions on heterogeneity. Although the procedure may require multi-valued IVs, the benefit of such variables can be manifested without requiring continuous or large support. This is practically relevant as infinitely supported IVs and continuously varying IVs are typically hard to find in practice, whereas discrete IVs are more common in an experimental and observational settings. We also show that, when IVs take more than four values, the preservation of FOSD (i.e., if (1.1) then (1.2)) yields a testable restriction.

Nonparametric identification of treatment effects using IVs with limited support has been a challenging goal even when the focus is on mean treatment effects, such as the average treatment effect (ATE) and the ATE on the treated (ATT). In an influential line of work (Manski (1990), Manski (1997), Manski and Pepper (2000)), Manski proposes the approach of partial identification and shows that sharp bounds on the ATE can be constructed under a set of assumptions relating to the directions of treatment effects and treatment selection while allowing instruments to be invalid in a specific sense. Even with valid instruments, however, bounds on the ATE are typically wide and uninformative to yield precise policy prediction. The local ATE (LATE) (Imbens and Angrist (1994)) and local QTE (Abadie et al. (2002)) have been a popular (point-identified) alternative when researchers are equipped with discrete IVs and impose a monotonicity assumption in the treatment selection. Unfortunately, it has intrinsic limitation that the local group for which the treatment effect is identified may not be the group of policy interest. Correspondingly, the extrapolation of the local parameters becomes an important issue for policy analysis (e.g., treatment allocation), in which case the identification challenge still remains (Mogstad et al. (2018), Han and Yang (2020)).

In another seminal work, Chernozhukov and Hansen (2005) propose assumptions that restrict the degree of treatment heterogeneity, called rank similarity and rank invariance, and point identify the QTE and the ATE. These assumptions (which are in fact observationally equivalent to each other) are shown to have great identifying power and used in various nonparametric contexts implicitly or explicitly (Heckman et al. (1997), Vytlacil and Yildiz (2007), Shaikh and Vytlacil (2011), Vuong and Xu (2017), Han (2019) to name a few). How-

ever, their plausibility can be questionable in many applications (e.g., Maasoumi and Wang (2019)) and testing methods are also proposed as one reaction to the skepticism (Frandsen and Lefgren (2018), Dong and Shen (2018), Kim and Park (2022)).

In this paper, we provide an alternative interpretation of rank similarity to clarify its stringency. We do so by first introducing *rank linearity*, a departure from rank similarity that allows the crossing of ranks, and show it is equivalent to the *strong* preservation of FOSD (i.e., (1.1) *if and only if* (1.2)). This way, we establishes the connection between a structural model under rank similarity and a restriction within Rubin (1974)'s causal framework with counterfactual outcomes. We then further relax the strong preservation of FOSD and propose our identifying conditions (e.g., (1.1) *only if* (1.2) discussed above). We therefore view these conditions as substantial relaxation of rank similarity.

This paper embraces the difficulty of identifying distributional and average effects under arbitrary heterogeneity without requiring IVs to have infinite or continuous support. Our approach is to exploit the trade-off between flexibility of models and data requirements in a way that is empirically relevant. The condition we introduce (i.e., the preservation of stochastic ordering) substantially weakens rank similarity while maintaining strong identifying power and reveals the need of multi-valued IVs. We believe this particular trade-off has not been previously explored. We verify that the preservation of FOSD ordering is satisfied in a range of nonseparable models that nest models under rank similarity as a special case. Researchers may justify these models in their particular applications or directly argue that the preservation condition holds (which is weaker to assume). A prevailing feature in these models is that they contain multi-dimensional unobservables, which is crucial to allow for essential treatment heterogeneity. A notable aspect of our approach is that depending on the direction of the FOSD preservation (i.e., (1.1) only if (1.2) or the converse) the procedure calculates bounds on either QTT or the QTE on the untreated (QTUT). In the context of the proposed models, we provide economic interpretation of a specific direction of preservation and justify why a policymaker wants to consider a particular type of treatment parameters depending on the assumption's plausibility.

We develop a procedure of calculating bounds on the QTT and QTUT, as well as the ATT and ATE on the untreated (ATUT), when the data requirement is met. The bound calculation is straightforward; it only requires to solve a set of linear programs to gather information from the data via (1.1) or (1.2) (or its observable equivalence). The bounds are initially characterized by optimal values of semi-infinite programs, especially with a continuous outcome variable. To tackle the infeasibility of the program, we propose two alternative approaches to transform it into a feasible linear program: (i) randomizing the constraints in the semi-infinite programs; (ii) invoking duality of the semi-infinite programs and approximating the Lagrangian measure using sieve.

The next section formally introduces the main identifying conditions (i.e., the preservation of stochastic ordering) and shows how to construct bounds on treatment effects. Section 3 provides additional interpretation of the identifying conditions, and Section 4 introduces structural models as sufficient conditions for them. Section 5 shows that point identification can be achieved with sufficient (but not infinite) variation of IVs. Section 6 discusses how to systematically calculate bounds using linear programming and, finally, Section 7 presents numerical studies.

2   Key Conditions and Bounds on Treatment Effects

Let $D \in \{0, 1\}$ be the observed treatment indicator, which represents the endogenous decision of an individual responding to IVs Z. We assume Z is either a vector of binary IVs or a multi-valued IV. In either case, we assume that Z takes L distinct values: $Z \in \mathcal{Z} \equiv \{z_1, ..., z_L\}$. Multi-valued or multiple IVs are common in many observational studies (e.g., natural experiments typically provide more than one instrument) and experimental studies (e.g., randomized control trials where multiple treatment arms are implemented either simultaneously or sequentially).[2] One of the main purposes of this paper is to motivate this type of IVs from the perspective of identification analysis. Let Y_1 be the counterfactual outcome of being

---

[2]See Mogstad et al. (2021) for a recent survey.

treated and Y_0 be that of not being treated. They can be either continuously or discretely distributed. The observed outcome $Y \in \mathcal{Y}$ satisfies $Y = DY_1 + (1 - D)Y_0$. Finally, $X \in \mathcal{X}$ denotes other covariates that may be endogenous.

Define QTE and ATE for treated and untreated populations. For $d \in \{0, 1\}$ and $x \in \mathcal{X}$, define

$$QTE_\tau(d, x) = Q_{Y_1|D,X}(\tau|d, x) - Q_{Y_0|D,X}(\tau|d, x)$$

and

$$ATE(d, x) = E[Y_1 - Y_0|D = d, X = x].$$

These parameters are what researchers and policymakers are potentially interested. The unconditional QTE and ATE (with respect to $D = d$) can be recovered when these parameters are identified for all $d \in \{0, 1\}$. Throughout the paper, we maintain that the IVs are valid and satisfy the exclusion restriction.

Assumption Z. *For $d \in \{0, 1\}$ and $z \in \{z_1, ..., z_L\}$, (i) $Y_{d,z} = Y_d$; (ii) $Z \perp Y_d|X$.*

Now we introduce the main identifying condition of this paper that establishes the mapping between observed and counterfactual distributions.

Condition S_1. *For arbitrary non-negative weight vectors $(w_1, ..., w_L)$ and $(\tilde{w}_1, ..., \tilde{w}_L)$ that satisfy $\sum_{\ell=1}^{L} w_\ell = \sum_{\ell=1}^{L} \tilde{w}_\ell = 1$, if*

$$\sum_{\ell=1}^{L} w_\ell P[Y_1 \leq \cdot|D = 1, Z = z_\ell, X = x] \leq \sum_{\ell=1}^{L} \tilde{w}_\ell P[Y_1 \leq \cdot|D = 1, Z = z_\ell, X = x], \quad (2.1)$$

then

$$\sum_{\ell=1}^{L} w_\ell P[Y_0 \leq \cdot|D = 1, Z = z_\ell, X = x] \leq \sum_{\ell=1}^{L} \tilde{w}_\ell P[Y_0 \leq \cdot|D = 1, Z = z_\ell, X = x]. \quad (2.2)$$

Note that the probabilities in (2.1) are observed as $Y_D = Y$. The mapping between observed and counterfactual distributions has been considered in Vuong and Xu (2017), which insights we share. Suppose $Z \perp (Y_d, D_z)|X$ where D_z is the counterfactual treatment that satisfies $D = \sum_{z \in \mathcal{Z}} 1\{Z = z\}D_z$; this assumption is stronger than Assumption Z that we require. Under this assumption, Condition S_1 is equivalent to assuming if

$$\sum_{\ell=1}^{L} w_\ell P[Y \leq \cdot | D_{z_\ell} = 1, X = x] \leq \sum_{\ell=1}^{L} \tilde{w}_\ell P[Y \leq \cdot | D_{z_\ell} = 1, X = x], \qquad (2.3)$$

then

$$\sum_{\ell=1}^{L} w_\ell P[Y_0 \leq \cdot | D_{z_\ell} = 1, X = x] \leq \sum_{\ell=1}^{L} \tilde{w}_\ell P[Y_0 \leq \cdot | D_{z_\ell} = 1, X = x]. \qquad (2.4)$$

Note that the event $\{D_{z_\ell} = 1\}$ (for $z_1, ..., z_L$) captures a type of compliance to a given $Z = z_\ell$. Then, $\sum_{\ell=1}^{L} w_\ell P[Y_d \leq y | D_{z_\ell} = 1, X = x]$ can be viewed as a distribution of Y_d weighted across different compliance types, and thus resulting in a distribution for a hypothetical population with a specific composition of compliance types. Therefore, Condition S_1 posits that the FOSD ordering between the distributions of Y of two compliance compositions is preserved between the distributions of Y_0 of the same pair of compliance compositions. For example, when $L = 2$ and defiers are excluded as a possible compliance type (e.g., the monotonicity assumption for local ATE (LATE) by Imbens and Angrist (1994)), then Condition S_1 simply concerns the stochastic ordering between always-takers and compliers. When $L \geq 3$, the composition becomes more complicated; see Section 3. We provide a condition sufficient to Condition S_1 that may be easier to interpret. To state this sufficient condition, we introduce a very general model for the treatment selection:

Assumption D. *Assume that*

$$D = h(Z, X, \eta), \qquad (2.5)$$

8

where $\eta \in \mathcal{T}$ can be an arbitrary vector.

Although Assumption D is not necessary for our main procedure, it is useful in capturing the types of compliance behavior via the unobservable η. Note that (2.5) permits a more general compliance behavior than what a weakly separable model $D = 1\{\eta \leq h(Z, X)\}$ does (or equivalently, what the LATE monotonicity does). This generalized selection equation is introduced to state the following sufficient condition and is not necessary for the main results of our identification analysis. Let $F_{Y_d|\eta,X}(y|t, x) \equiv P[Y_d \leq y|\eta = t, X = x]$.

Condition S_1^*. *Fix $x \in \mathcal{X}$. For arbitrary weight functions $w : \mathcal{T} \times \mathcal{X} \to \mathbb{R}_+$ and $\tilde{w} : \mathcal{T} \times \mathcal{X} \to \mathbb{R}_+$ such that $\int w(t, x)dt = \int \tilde{w}(t, x)dt = 1$, if*

$$\int w(t, x)F_{Y_1|\eta,X}(\cdot|t, x)dt \leq \int \tilde{w}(t, x)F_{Y_1|\eta,X}(\cdot|t, x)dt,$$

then

$$\int w(t, x)F_{Y_0|\eta,X}(\cdot|t, x)dt \leq \int \tilde{w}(t, x)F_{Y_0|\eta,X}(\cdot|t, x)dt.$$

Because $\int w(t, x)dt = 1$, note that $\int w(t, x)F_{Y_d|\eta,X}(\cdot|t, x)dt$ is a mixture of conditional CDFs (with $w(t, x)$ being the mixing distribution) and thus itself a CDF. In other words, defining a type distribution $W_x(t) = \int^t w(\eta, x)d\eta$, we can write $\int w(t, x)F_{Y_d|\eta,X}(\cdot|t, x)dt = \int F_{Y_d|\eta,X}(\cdot|t, x)dW_x(t)$.[3] Therefore, Condition S_1^* assumes that the FOSD ordering of Y_1 distributions conditional on η conforming to two different type distributions is preserved in the ordering of Y_0 distributions conditional on the same type distributions. Note that Condition S_1^* (or Condition S_1) is *not* an "if and only if" statement. If a condition imposes that the preservation of ordering holds in both directions, it would be a very stringent one. In fact, such a condition is closely related to the rank similarity condition (Chernozhukov and Hansen (2005)); see Section 4 for full discussions. The following lemma establishes the sufficiency of Condition S_1^* for Condition S_1.

Lemma 2.1. *Under Assumption D, Condition S_1^* implies Condition S_1.*

---

[3]Since η has arbitrary dimensions, the integral with respect to t is understood to be a multivariate integral.

The proof of this lemma and most of other proofs are contained in the appendix. Now, we show that Condition S_1 is useful in constructing bounds on $F_{Y_0|D,X}(\cdot|1,x)$ and subsequently on $QTE_\tau(1,x)$. Let

$$\Gamma_p(x) \equiv \left\{ (\gamma_1(x),...,\gamma_L(x)) \in \mathbb{R}^L : \sum_{\ell=1}^{L} \gamma_\ell(x) = 0 \text{ and } \sum_{\ell=1}^{L} \gamma_\ell(x)p(z_\ell,x) = 1 \right\}.$$

Theorem 2.1. *Suppose that Assumption Z and Condition S_1 hold. Fix $x \in \mathcal{X}$. For $\gamma(x) \equiv (\gamma_1(x),...,\gamma_L(x))$ and $\tilde{\gamma}(x) \equiv (\tilde{\gamma}_1(x),...,\tilde{\gamma}_L(x))$ in $\Gamma_p(x)$, suppose*

$$P[Y \leq y|D=1, X=x] \leq \sum_{\ell=1}^{L} \gamma_\ell(x)P[Y \leq y, D=1|Z=z_\ell, X=x], \quad \forall y \in \mathcal{Y} \qquad (2.6)$$

$$\sum_{\ell=1}^{L} \tilde{\gamma}_\ell(x)P[Y \leq y, D=1|Z=z_\ell, X=x] \leq P[Y \leq y|D=1, X=x], \quad \forall y \in \mathcal{Y} \qquad (2.7)$$

Then $F_{Y_0|D,X}(\cdot|1,x)$ is bounded by

$$-\sum_{\ell=1}^{L} \tilde{\gamma}_\ell(x)P[Y \leq \cdot, D=0|Z=z_\ell, X=x] \qquad (2.8)$$

$$\leq P[Y_0 \leq \cdot|D=1, X=x]$$

$$\leq -\sum_{\ell=1}^{L} \gamma_\ell(x)P[Y \leq \cdot, D=0|Z=z_\ell, X=x] \qquad (2.9)$$

Proof. We suppress X for simplicity and prove the upper bound; the lower bound can be analogously derived. Without loss of generality, for some $\ell^* \leq L$, let $\gamma_\ell \leq 0$ for $\ell \leq \ell^*$ and $\gamma_\ell > 0$ for $\ell > \ell^*$. Let $q(z_\ell) \equiv P[Z=z_\ell|D=1]$. Then, (2.6) can be rewritten as

$$\sum_{\ell=1}^{L} q(z_\ell) \times P[Y \leq y|D=1, Z=z_\ell] - \sum_{\ell=1}^{\ell^*} \gamma_\ell p(z_\ell) \times P[Y \leq y|D=1, Z=z_\ell]$$

$$\leq \sum_{\ell=\ell^*+1}^{L} \gamma_\ell p(z_\ell) \times P[Y \leq y|D=1, Z=z_\ell].$$

Let $a \equiv 1 - \sum_{\ell=1}^{\ell^*} \gamma_\ell p(z_\ell)$. By definition and that $\sum_{\ell=1}^{L} \gamma_\ell p(z_\ell) = 1$, we have $a = \sum_{\ell=\ell^*+1}^{L} \gamma_\ell p(z_\ell)$.

Therefore, we have

$$\sum_{\ell=1}^{\ell^*} \frac{q(z_\ell) - \gamma_\ell p(z_\ell)}{a} \times P[Y_1 \leq y | D = 1, Z = z_\ell] + \sum_{\ell=\ell^*+1}^{L} \frac{q(z_\ell)}{a} \times P[Y_1 \leq y | D = 1, Z = z_\ell]$$

$$\leq \sum_{\ell=\ell^*+1}^{L} \frac{\gamma_\ell p(z_\ell)}{a} \times P[Y_1 \leq y | D = 1, Z = z_\ell],$$

where $\sum_{\ell=1}^{\ell^*} \frac{q(z_\ell) - \gamma_\ell p(z_\ell)}{a} + \sum_{\ell=\ell^*+1}^{L} \frac{q(z_\ell)}{a} = 1$ and $\sum_{\ell=\ell^*+1}^{L} \frac{\gamma_\ell p(z_\ell)}{a} = 1$. Therefore, by Condition S_1, we have

$$\sum_{\ell=1}^{k} \frac{q(z_\ell) - \gamma_\ell p(z_\ell)}{a} \times P[Y_0 \leq y | D = 1, Z = z_\ell] + \sum_{\ell=\ell^*+1}^{L} \frac{q(z_\ell)}{a} \times P[Y_0 \leq y | D = 1, Z = z_\ell]$$

$$\leq \sum_{\ell=\ell^*+1}^{L} \frac{\gamma_\ell p(z_\ell)}{a} \times P[Y_0 \leq y | D = 1, Z = z_\ell].$$

Equivalently, we have

$$\begin{aligned}
P[Y_0 \leq y | D = 1] \quad &\leq \quad \sum_{\ell=1}^{L} \gamma_\ell \times P[Y_0 \leq y, D = 1 | Z = z_\ell] \\
&= \quad \sum_{\ell=1}^{L} \gamma_\ell \times \{P[Y_0 \leq y | Z = z_\ell] - P[Y_0 \leq y, D = 0 | Z = z_\ell]\} \\
&= \quad \sum_{\ell=1}^{L} \gamma_\ell P[Y_0 \leq y | Z = z_\ell] - \sum_{\ell=1}^{L} \gamma_\ell \times P[Y_0 \leq y, D = 0 | Z = z_\ell] \\
&= \quad P[Y_0 \leq y] \times \sum_{\ell=1}^{L} \gamma_\ell - \sum_{\ell=1}^{L} \gamma_\ell \times P[Y \leq y, D = 0 | Z = z_\ell] \\
&= \quad -\sum_{\ell=1}^{L} \gamma_\ell \times P[Y \leq y, D = 0 | Z = z_\ell],
\end{aligned}$$

where the last equality is by $\sum_{\ell=1}^{L} \gamma_\ell = 0$. $\qquad \square$

Note that there can be multiple $\gamma(x)$ and $\tilde{\gamma}(x)$ in $\Gamma_p(x)$ that satisfy (2.6) and (2.7), respectively. Therefore, we can further tighten the bounds as follows.

Corollary 2.1. *Suppose that Assumption Z and Condition S_1 hold. Fix $x \in \mathcal{X}$. Then,*

$F_{Y_0|D,X}(\cdot|1,x)$ is upper and lower bounded by

$$F_{Y_0|D,X}^{UB}(y|1,x) \equiv \min_{\gamma(x)\in\Gamma_p(x):(2.6)\ holds} -\sum_{\ell=1}^{L}\gamma_\ell(x)P[Y\le y, D=0|Z=z_\ell],$$

$$F_{Y_0|D,X}^{LB}(y|1,x) \equiv \max_{\tilde\gamma(x)\in\Gamma_p(x):(2.7)\ holds} -\sum_{\ell=1}^{L}\tilde\gamma_\ell(x)P[Y\le y, D=0|Z=z_\ell].$$

Theorem 2.1 and Corollary 2.1 highlight the usefulness of multi-valued IVs. The key to calculating bounds is to find $\gamma(x)$ (and $\tilde\gamma(x)$) in $\Gamma_p(x)$ that satisfies (2.6) (and (2.7)). The inequality is more likely to hold when instruments take more values (i.e., when L is large) because, in this case, the degree of freedom in Γ_p increases. The corollary additionally shows how the bounds can be further tightened if the set of $\gamma(x)\in\Gamma_p(x)$ that satisfies (2.6) becomes "large" with large L. Sections 3 and 7 contain related discussions. Note that (2.6)–(2.7) are testable in the data.

Finally, note that

$$QTE_\tau(1,x) = Q_{Y|D,X}(\tau|1,x) - Q_{Y_0|D,X}(\tau|1,x)$$

and the bounds on the second quantity on the right-hand side can be calculated using the worst case bounds for the conditional quantile (Manski (1994), Blundell et al. (2007)):

$$Q_{Y_0|D,X}^{LB}(\tau|1,x) \le Q_{Y_0|D,X}(\tau|1,x) \le Q_{Y_0|D,X}^{UB}(\tau|1,x)$$

with $Q_{Y_0|D,X}^{LB}(\tau|1,x)$ and $Q_{Y_0|D,X}^{UB}(\tau|1,x)$ being the solutions to

$$\tau = F_{Y_0|D,X}^{UB}(y|1,x),$$
$$\tau = F_{Y_0|D,X}^{LB}(y|1,x),$$

respectively, using Corollary 2.1. Although the bounds on $ATE(1,x) = E[Y|D=1,X=x] - E[Y_0|D=1,X=x]$ can be calculated based on $E[Y_0|D=1,X=x] = \int_0^1 Q_{Y_0|D,X}(\tau|1,x)d\tau$,

we present later how the bounds on the $ATE(d, x)$ can be calculated under a weaker condition than Condition S_1.

Remark 2.1 (Multi-Valued IVs). *In the introduction we claim that, when $L = 2$ and $(\gamma_1, \gamma_2) = \left(\frac{1}{p(z_1) - p(z_2)}, -\frac{1}{p(z_1) - p(z_2)}\right)$, (2.6) can be expressed as $P[Y \leq y | D = 1, Z = 0] \leq P[Y \leq y | D = 1, Z = 1]$ for all y, and that the latter is in turn establishes FOSD of Y_1 between always takers and compliers, assuming there is no defiers. In the next section, we formally prove this claim and provide further insights by giving interpretations of (2.6) for $L = 3$.*

Remark 2.2 (More on $\gamma(x)$). *The restrictions on $\gamma(x)$ (i.e., $\sum_{\ell=1}^{L} \gamma_\ell(x) = 0$ and $\sum_{\ell=1}^{L} \gamma_\ell p(z_\ell, x) = 1$) are worth discussing. First, note that the existence of such a sequence requires the relevance of the IV: $p(z_\ell, x) \neq p(z_{\ell'}, x)$ for some $z_\ell, z_{\ell'}$. Moreover, any sequence $\gamma(x)$ satisfying $\sum_{\ell=1}^{L} \gamma_\ell(x) = 0$ and $\sum_{\ell=1}^{L} \gamma_\ell(x) p(z_\ell, x) \neq 0$ is enough. We can always rescale $\gamma_\ell(x)$'s as $\gamma_\ell^*(x) = \frac{\gamma_\ell(x)}{\sum_{\ell=1}^{L} \gamma_\ell(x) p(z_\ell, x)}$, then $\gamma^*(x) = (\gamma_1^*(x), ..., \gamma_L^*(x))$ would satisfy (2.6) and (2.9) in Theorem 2.1.*

If we assume the converse of Condition S_1, we can calculate bounds on the $QTE(0, x)$.

Condition S_0. *For arbitrary non-negative weight vectors $(w_1, ..., w_L)$ and $(\tilde{w}_1, ..., \tilde{w}_L)$ that satisfy $\sum_{\ell=1}^{L} w_\ell = \sum_{\ell=1}^{L} \tilde{w}_\ell = 1$, if*

$$\sum_{\ell=1}^{L} w_\ell P[Y_0 \leq \cdot | D = 0, Z = z_\ell, X = x] \leq \sum_{\ell=1}^{L} \tilde{w}_\ell P[Y_0 \leq \cdot | D = 0, Z = z_\ell, X = x], \quad (2.10)$$

then

$$\sum_{\ell=1}^{L} w_\ell P[Y_1 \leq \cdot | D = 0, Z = z_\ell, X = x] \leq \sum_{\ell=1}^{L} \tilde{w}_\ell P[Y_1 \leq \cdot | D = 0, Z = z_\ell, X = x]. \quad (2.11)$$

Theorem 2.2. *Suppose that Assumption Z and Condition S_0 hold. Fix $x \in \mathcal{X}$. For $\gamma(x) \equiv$*

$(\gamma_1(x), ..., \gamma_L(x))$ *and* $\tilde{\gamma}(x) \equiv (\tilde{\gamma}_1(x), ..., \tilde{\gamma}_L(x))$ *in* $\Gamma_p(x)$*, suppose*

$$P[Y \leq y | D = 0, X = x] \leq \sum_{\ell=1}^{L} \gamma_\ell(x) P[Y \leq y, D = 0 | Z = z_\ell, X = x], \quad \forall y \in \mathcal{Y} \qquad (2.12)$$

$$\sum_{\ell=1}^{L} \tilde{\gamma}_\ell(x) P[Y \leq y, D = 0 | Z = z_\ell, X = x] \leq P[Y \leq y | D = 0, X = x], \quad \forall y \in \mathcal{Y} \qquad (2.13)$$

Then $F_{Y_1|D,X}(\cdot|0, x)$ *is bounded by*

$$- \sum_{\ell=1}^{L} \tilde{\gamma}_\ell(x) P[Y \leq \cdot, D = 1 | Z = z_\ell, X = x] \qquad (2.14)$$

$$\leq P[Y_1 \leq \cdot | D = 0, X = x]$$

$$\leq - \sum_{\ell=1}^{L} \gamma_\ell(x) P[Y \leq \cdot, D = 1 | Z = z_\ell, X = x] \qquad (2.15)$$

The proof of this theorem is analogous to that of Theorem 2.1. Also the derivation of the bounds on $QTE_\tau(0, x)$ is symmetric to the case of $QTE_\tau(1, x)$ so is omitted. Notably, which treatment parameter we can obtain bounds for is determined by which identifying condition we impose. In Section 4, we investigate this aspect within a structural model. Finally, in the Appendix, we introduce weaker conditions to bound average treatment effects instead of quantile effects.

3 Understanding Key Conditions

We further explore Conditions S_1 and S_0 to give additional interpretation and discuss testability. Suppress $X = x$ to simplify our discussions. Under $Z \perp (Y_d, D_z)$, the inequalities for FOSD in Conditions S_1 and S_0 can be rewritten as

$$\sum_{\ell=1}^{L} w_\ell P[Y_d \leq y | D_{z_\ell} = 1] \leq \sum_{\ell=1}^{L} \tilde{w}_\ell P[Y_d \leq y | D_{z_\ell} = 1]. \qquad (3.1)$$

Recall, Theorem 2.1 relies on the existence of a sequence $\gamma = (\gamma_1, ..., \gamma_L)$ satisfying $\sum_{\ell=1}^{L} \gamma_\ell = 0$, $\sum_{\ell=1}^{L} \gamma_\ell p(z_\ell, x) = 1$, and the inequality (2.6), that is, $P[Y \leq y | D = 1] \leq \sum_{\ell=1}^{L} \gamma_\ell P[Y \leq y, D = 1 | Z = z_\ell]$ for all y. Note that (2.6) is a special case of (3.1) with $d = 1$, which is the "if" part of Condition S_1. To understand (3.1) better, we consider simple cases of binary and trinary IVs. Let $p(z) \equiv (D = 1 | Z = z)$. Assume weak separability in the selection equation for the sake of discussion:

Assumption D*. *Assume that* $D = h(Z, \eta) = 1\{\eta \leq p(Z)\}$ *with* $\eta \sim U[0, 1]$.

Then, $\{D_{z_\ell} = 1\}$ are a mix of individuals who are compliers (C) and always-takers (AT). This observation is detailed in the two cases of IV.

Lemma 3.1. *Suppose* $L = 2$ *with* $\mathcal{Z} = \{0, 1\}$ *and* $D = 1\{\eta \leq p(Z)\}$. *Also suppose* $Z \perp (Y_d, D_z)$ *and* $0 < p(z) < 1$ *for* $z = 0, 1$. *Then,* (2.6) *uniquely holds with* $(\gamma_1, \gamma_2) = \left(\frac{1}{p(1) - p(0)}, -\frac{1}{p(1) - p(0)}\right)$, *which is equivalent to* (1.3) *in the introduction and is in turn equivalent to* $Y_1 | AT \prec_{FOSD} Y_1 | C$.

Now, consider the case of $L = 3$. It is more convenient to focus on (3.1) rather than its special case, (2.6). Suppose for $z, z' \in \mathcal{Z} = \{0, 1, 2\}$ such that $z < z'$, we have $p(z) < p(z')$ in Assumption D. Following the language of Mogstad et al. (2021), let $\{i : D_{0,i} = D_{1,i} = D_{2,i} = 1\}$ be always-takers (AT), $\{i : D_{0,i} = 0, D_{1,i} = D_{2,i} = 1\}$ be eager compliers (E-C), $\{i : D_{0,i} = D_{1,i} = 0, D_{2,i} = 1\}$ be reluctant compliers (R-C), and $\{i : D_{0,i} = D_{1,i} = D_{2,i} = 0\}$ be never-takers (NT). Then, for example

$$\{D_1 = 1\} = \{D_0 = 1, D_1 = 1, D_2 = 1\} \cup \{D_0 = 0, D_1 = 1, D_2 = 1\} = \{AT\} \cup \{E\text{-}C\},$$

because $\{D_0 = 1, D_1 = 1, D_2 = 0\} = \emptyset$ and $\{D_0 = 0, D_1 = 1, D_2 = 0\} = \emptyset$. Also, $\{D_2 = 1\} = \{AT\} \cup \{E\text{-}C\} \cup \{R\text{-}C\}$ and $\{D_0 = 1\} = \{AT\}$. Therefore, (3.1) establishes the FOSD relationship between the mixtures of observed distributions of Y conditional on various always-takers and compliers groups.

Remark 3.1 (Testability of the Conditions). *Conditions S_1 and S_0 can be tested from the data when IVs take many values and Assumption D^*. The intuition is as follows. With a binary IV, the marginal distributions of Y_1 and Y_0 are identified for compliers (Abadie et al. (2002)). This result holds for any complier group defined by a pair of instrument values. When $L = 3$, for example, $F_{Y_d}(\cdot|E\text{-}C)$ can be identified by $Z = 0$ and $Z = 1$ because $\{D_0 = 0, D_1 = 1, D_2 = 1\} = \{D_0 = 0, D_1 = 1\}$ by Assumption D^*. Similarly, $F_{Y_d}(\cdot|R\text{-}C)$ can be identified by $Z = 1$ and $Z = 2$. Then, when $L \geq 4$, we can find vectors w and $\tilde{w}$ in $\mathbb{R}_+^L$ that assign zero weights to the distributions for AT and still make (3.1) a non-trivial inequality with all the associated distributions for compliers are identified for all $d = 1, 0$. The formal proof is in the Appendix. Therefore, Conditions S_1 and S_0 are refutable by the data.*

Remark 3.2 (Conditions w.r.t. Compliance Types). *Motivated from the discussion of this section, we can rewrite Condition S_1 (and all the relevant conditions) without even invoking Assumption D. This modification will provide an interpretation of the condition that solely relies on compliance types. Let $T \equiv \{D(z_1), ..., D(z_L)\}$ be a random vector that indicates a particular compliance type with its realized value in $\{0, 1\}^L \equiv \mathcal{T}$. For example, when $L = 2$ (i.e., binary IV), $T = (D(0), D(1)) \in \{(0, 0), (1, 0), (0, 1), (1, 1)\} \equiv \mathcal{T}$. Since D and Z are discrete, T is naturally a discrete random vector. Note that this framework do not rely on any selection models, and therefore T captures all possible compliance types given D and Z. Then Condition S_1 can be modified as follows:*

Fix $x \in \mathcal{X}$. For arbitrary weight functions $w : \mathcal{T} \times \mathcal{X} \to \mathbb{R}_+$ and $\tilde{w} : \mathcal{T} \times \mathcal{X} \to \mathbb{R}_+$ such that $\sum_{t \in \mathcal{T}} w(t, x) = \sum_{t \in \mathcal{T}} \tilde{w}(t, x) = 1$, if

$$\sum_{t \in \mathcal{T}} w(t, x) F_{Y_1|T,X}(\cdot|t, x) \leq \sum_{t \in \mathcal{T}} \tilde{w}(t, x) F_{Y_1|T,X}(\cdot|t, x),$$

then

$$\sum_{t \in \mathcal{T}} w(t, x) F_{Y_0|T,X}(\cdot|t, x) \leq \sum_{t \in \mathcal{T}} \tilde{w}(t, x) F_{Y_0|T,X}(\cdot|t, x).$$

Then, the weighted sum in each inequality can be interpreted as the distribution of Y_d weighted across all compliance types.

4 Structural Models and Policymaker's Problems

We show that Conditions S_1 and S_0 can be justified in a range of nonparametric structural models for the counterfactual outcomes. By relating the conditions with the structural models, we provide additional intuitions for the conditions. We present a leading model here and the rest in the Appendix. For arbitrary r.v.'s A and $\tilde{A}$, let $A \overset{d}{=} \tilde{A}$ denote $F_A = F_{\tilde{A}}$.

Model 1. (i) We have

$$Y_d = q(d, X, U_d) \qquad \text{for } d \in \{0, 1\}, \tag{4.1}$$

where $q(d, x, \cdot)$ is continuous and monotone increasing, (ii) $D = h(Z, X, \eta)$ (i.e., Assumption D), (iii) conditional on (η, X, Z), $U_d \overset{d}{=} U + \xi_d$ where $\xi_d \perp (\eta, U)$, (iv) conditional on (X, Z), ξ_0 is (weakly) more or less noisy than ξ_1, that is, $\xi_0 \overset{d}{=} \xi_1 + V$ for some V independent of ξ_1.

Note that U is the source of endogeneity in that it allowed to be dependent on η. Model 1(iii)–(iv) implies that $U_0 \overset{d}{=} U_1 + V$ conditional on (η, X, Z). Importantly, Model 1 nests the model in Chernozhukov and Hansen (2005) as a special case. This can be shown as follows. First, Chernozhukov and Hansen (2005) assume Model 1(i)–(ii). Additionally, they assume, conditional on (X, Z), either rank similarity ($F_{U_0|\eta} = F_{U_1|\eta}$) or rank invariance ($U_0 = U_1$).[4] Then, by taking $\xi_d = 0$ for all d in Model 1(iii), we have $U_0 \overset{d}{=} U_1 \overset{d}{=} U$ conditional on (η, X, Z), which proves the claim.

Model 1(iv) assumes that the unobservable under the counterfactual status of being treated are more (or less) dispersed than that under the counterfactually untreated status. Although this may seem stringent, it is substantially weaker than rank similarity (or invariance) and can be plausible in various scenarios. Before providing examples of these

---

[4]Note that rank similarity and rank invariance are observationally equivalent under Model 1(i)–(ii) in that they produce the same distribution of observables (Chernozhukov and Hansen (2013)).

scenarios, we first establish the connection between Model 1 and Condition S_1^* (and thus Condition S_1 by Lemma 2.1).

Theorem 4.1. *Under Assumptions Z, Model 1 (with ξ_0 being weakly more noisy than ξ_1) implies Condition S_1^*.*

Analogous to Theorem 4.1, one can readily show that Model 1 with ξ_0 being weakly less noisy than ξ_1 implies Condition S_0.

Now we provide examples that are consistent with Model 1.

Example 1 (Auction)**.** *Consider online and offline auctions. Let Y be the bid (which subsequently forms revenue) and D be participating in an auction with different format ($D = 1$ if online and $= 0$ if offline). Let $U_d \stackrel{d}{=} U + \xi_d$ be the valuation of the item where U is the common valuation (correlated with D) and ξ_d is format specific random shocks satisfying $\xi_d \perp (\eta, U)$. We assume that bidders have limited information on certain features of the auction that affect valuation (e.g., they know the distribution of ξ_d but not its realization). In this example, what would justify $var(\xi_0) > var(\xi_1)$? It may be the case that, in the offline auction, bidders are more emotionally affected by other bidders, which makes their bids more variable.*

Example 2 (Insurance)**.** *We are interested in the effect of insurance on health outcomes. Let Y be the health outcome and D be the decision of getting insurance ($D = 1$ being insured). Let $U_d \stackrel{d}{=} U + \xi_d$ be underlying health conditions where U captures health conditions known to participant (and thus correlated with D) while ξ_d is health conditions not fully known a priori and thus random. In this example, $var(\xi_0) > var(\xi_1)$ may hold because insurance by definition ensures a certain level of health conditions.*

Example 3 (Vaccination)**.** *Similar to Example 2, suppose that D is instead getting vaccination (of an established vaccine). Again, $U_d \stackrel{d}{=} U + \xi_d$ is health conditions where U captures conditions known to participant (and correlated with D) and ξ_d is vaccination-status-specific health conditions, which are not fully known a priori. Then, similarly as before*

$var(\xi_0) > var(\xi_1)$ *may hold because, when not vaccinated, one is exposed to the risk of a serious illness, while vaccination ensures a certain level of immunity.*

The scenarios in Examples 1–3 justify Condition S_1 via Theorem 4.1. Then, under Condition S_1, Theorem 2.1 and Corollary 2.1 yield bounds on $QTE_\tau(1, x)$, the effects of treatment for those who take the treatment. The final example illustrates the converse case.

Example 4 (Medical Trial)**.** *In contrast to Example 3, suppose the treatment itself is risky. That is, let D be participating in a frontier medical trial ($D = 1$ being participation). In this case, $var(\xi_0) < var(\xi_1)$ is more plausible because, with a newly developed medicine, there is the high risk of unknown side effects.*

The scenario in Example 4 justifies Condition S_0, under which bounds on $QTE_\tau(0, x)$, the effects of treatment for those who abstain from it, can be obtained.

Model 1 and these examples show how a certain treatment parameter may be more relevant for policy than others depending on the plausibility of assumptions. Consider the problem of a policymaker. Assume that the policymaker concerns risk-averse individuals, *which are typically the majority.* For this policymaker, a candidate policy would aim at providing "insurance," which can be either literally insurance or policies that serve as insurance (e.g., vaccination, subsidies). Therefore, she would be interested in understanding the treatment effects for the target individuals that are risk-averse. Our procedure provides a statistical tool for such a policymaker. That is, under Model 1, our procedure has the ability to bound the treatment effects for individuals with $D = d$ such that $var(\xi_d) < var(\xi_{1-d})$. This is a unique feature of our setting: the plausibility of assumptions dictates the parameters of interest, which then can be terms as *assumption-driven* treatment parameters.

A remaining question one might have is as follows. How much Condition S_1 has to be strengthened to be equivalent to rank similarity? To answer this question, recall that Condition S_1^* is stronger than Condition S_1 (by Lemma 2.1). We strengthen Condition S_1^* further by making it an "if and only if" condition:

Condition S*. *Fix $x \in \mathcal{X}$. For arbitrary weight functions $w : \mathcal{T} \times \mathcal{X} \to \mathbb{R}_+$ and $\tilde{w} : \mathcal{T} \times \mathcal{X} \to \mathbb{R}_+$ such that $\int w(t, x)dt = \int \tilde{w}(t, x) = 1$, it holds that*

$$\int w(t, x) F_{Y_1|\eta, X}(\cdot|t, x)dt \leq \int \tilde{w}(t, x) F_{Y_1|\eta, X}(\cdot|t, x)dt$$

if and only if

$$\int w(t, x) F_{Y_0|\eta, X}(\cdot|t, x)dt \leq \int \tilde{w}(t, x) F_{Y_0|\eta, X}(\cdot|t, x)dt.$$

It turns out that we can establish the following result.

Theorem 4.2. *Model 1(i)–(ii) with $F_{U_0|\eta, X, Z} = F_{U_1|\eta, X, Z}$ (i.e., rank similarity) implies Condition S*.*

This theorem highlights the stringency of rank similarity relative to Condition S_1. The proof is trivial so omitted. It is worth noting that the converse of Theorem 4.2 is *not* true. Here is a counter-example for the converse statement.

Definition 4.1 (Rank Linearity)**.** *Assume Model 1(i)–(ii) and*

$$F_{Y_0|\eta, X, Z}(\cdot|t, x, z) = \lambda(\cdot, x) F_{Y_1|\eta, X, Z}(\psi(\cdot, x)|t, x, z) \tag{4.2}$$

for every $t \in \mathcal{T}$ and $x \in \mathcal{X}$, where $\psi(\cdot, x) : \mathcal{Y} \to \mathcal{Y}$, a one-to-one and onto mapping, is strictly increasing, and $\lambda(\cdot, x) : \mathcal{Y} \to \mathbb{R}_+$ is consistent with $F_{Y_d|\eta, X, Z}$ being a proper CDF.

This *rank linearity* implies Condition S*, which is trivial to show. However, rank linearity is weaker than rank similarity as the latter is a special case of the former. To see this, conditional on $Z = z$ (and suppressing X), (4.2) with Model 1(i) yields $F_{U_0|\eta}(q^{-1}(0, y)|t) = \lambda(y) F_{U_1|\eta}(q^{-1}(1, \psi(y))|t)$. Then, by choosing $\lambda(y) = 1$ and $\psi(y) = q(1, q^{-1}(0, y))$, we have $F_{U_0|\eta}(\cdot|t) = F_{U_1|\eta}(\cdot|t)$. In general, while the ranks between Y_0 and Y_1 should stay the same under rank similarity, they can be arbitrarily different (e.g., the rank can be reversed) under rank linearity because of the multiplying term $\lambda(\cdot)$ in $F_{U_0|\eta}(u|t) = \lambda(q(0, u)) F_{U_1|\eta}(u|t)$.

Interestingly, rank linearity is equivalent to Condition S^*. The following theorem is one of the main contributions of this paper. Suppress (Z, X) for simplicity.

Theorem 4.3. *Suppose for any CDF $F_1(\cdot)$ supported on $\mathbb{R}$, there always exists a function $c : \mathcal{T} \to \mathbb{R}$ such that*

$$F_d(\cdot) = \int c(t) F_{Y_d|\eta}(\cdot|t) dt. \tag{4.3}$$

Then Condition S^* holds if and only if there exits some $\psi(\cdot)$ that is strictly increasing and $\lambda(\cdot) > 0$ such that

$$F_{Y_0|\eta}(\cdot|t) = \lambda(\cdot) F_{Y_1|\eta}(\psi(\cdot)|t) \qquad for \ y \in \mathbb{R} \ and \ t \in \mathcal{T}. \tag{4.4}$$

We prove this equivalence in the Appendix. The proof with continuous Y_d is more involved than that with discrete Y_d; we recommend that the interested reader reads the latter first. The condition (4.3) is only introduced in this theorem to establish the relationship between rank linearity (and hence rank similarity) and the range of identifying conditions of this paper, and it is not necessary for our bound analysis. This condition would be violated when there is no endogeneity (i.e., $Y_d \perp \eta$), which is not our focus.

Condition S^* is crucial in bounding $QTE_\tau(x) = Q_{Y_1|X}(\tau|x) - Q_{Y_0|X}(\tau|x)$ unconditional with respect to $D = d$. The "only if" part (i.e., Condition S_1) will bound $Q_{Y_0|D=1}(\tau)$ and thus $Q_{Y_0}(\tau)$ by Theorem 2.1, while the "if" part (i.e., Condition S_0) will bound $Q_{Y_1|D=0}(\tau)$ and thus $Q_{Y_1}(\tau)$ by the symmetric version of Theorem 2.1. The fact that Condition S^* is weaker than rank similarity illustrates the importance of rank similarity in the identification of the QTE and ATE.

Remark 4.1 (Testability of the Conditions, continued)**.** *It is immediate from Remark 3.1 that Condition S^* can be tested from the data when $L \geq 4$. Given the established connection between Condition S^* and rank similarity, it is not surprising that the testability of the former requires certain variation of IVs as the latter does (Frandsen and Lefgren (2018), Dong and Shen (2018), Kim and Park (2022)).*

Remark 4.2 (Conditions w.r.t. Compliance Types, continued). *In the framework proposed in Remark 3.2, if we additionally impose Assumption D, we have $\sigma(T) \subset \sigma(\eta)$ where $\sigma(A)$ is a σ-field generated by a random vector A. Therefore, given Model 1(i)–(ii), $F_{U_1|T} = F_{U_0|T}$ is weaker than $F_{U_1|\eta} = F_{U_0|\eta}$. This way by using T, we can define a weaker version of rank similarity. Moreover, Condition C.1 can be motivated by this framework as the discrete η can be viewed as T (with $k_\eta = 2^L$).*

5 Point Identification

Point identification of $QTE_\tau(d)$ and $ATE(d)$ can be achieved as long as the stochastic dominance ordering is preserved (i.e., Condition S_1 or S_0) and instruments have sufficient variation in a specific sense. As is clear below, however, we do *not* require $p(z) \to 1$ or 0 (i.e., instruments with large support). In this sense, our approach to point identification complements the approach of identification at infinity. To see this, consider the following theorem.

Theorem 5.1. *Suppose that Assumption Z and Condition S_1 hold. Fix $x \in \mathcal{X}$. For $\gamma(x) \equiv (\gamma_1(x), ..., \gamma_L(x))$ in $\Gamma_p(x)$, suppose*

$$P[Y \leq y | D = 1, X = x] = \sum_{\ell=1}^{L} \gamma_\ell(x) P[Y \leq y, D = 1 | Z = z_\ell, X = x], \quad \forall y \in \mathcal{Y} \qquad (5.1)$$

Then $F_{Y_0|D,X}(\cdot|1,x)$ is identified as

$$P[Y_0 \leq \cdot | D = 1, X = x] = -\sum_{\ell=1}^{L} \gamma_\ell(x) P[Y \leq \cdot, D = 0 | Z = z_\ell, X = x] \qquad (5.2)$$

The key for this point identification result is that there exists $\gamma(x)$ such that (5.1) holds, which is a stronger requirement than the inequality version (2.6). The equation (5.1) is more likely to hold when L is large, that is, when instruments take more values. In particular, when $L \to \infty$ (e.g., continuous Z), we may invoke the notion similar to sieve approximation for $P[Y \leq y | D = 1, X = x]$ with $\{P[Y \leq y | D = 1, Z = z_\ell, X = x]\}_{\ell=1}^{L}$ being the "basis

functions" and assume

$$P[Y \leq y | D = 1, X = x] = \lim_{L \to \infty} \sum_{\ell=1}^{L} \gamma_\ell(x) p(z_\ell, x) P[Y \leq y | D = 1, Z = z_\ell, X = x],$$

where $\sum_{\ell=1}^{L} \gamma_\ell(x) p(z_\ell, x) = 1$ and $\sum_{\ell=1}^{L} \gamma_\ell(x) = 0$. Note that, although this does *not* demand an infinite support for Z, it implicitly assumes that Z sufficiently influences the distribution of Y conditional on $(D, X) = (1, x)$ in a way that the resulting basis functions generate $P[Y \leq y | D = 1, X = x]$. Importantly, whether this is possible or not can be confirmed from the data.

Given Theorem 5.1, we identify $QTE_\tau(1, x) = Q_{Y|D,X}(\tau|1, x) - Q_{Y_0|D,X}(\tau|1, x)$ where $Q_{Y_0|D,X}(\tau|1, x)$ is a solution to $\tau = -\sum_{\ell=1}^{L} \gamma_\ell(x) P[Y \leq \cdot, D = 0 | Z = z_\ell, X = x]$. Similarly, under Condition S_0, we can identify $F_{Y_0|D,X}(\cdot|1, x)$ and thus $QTE_\tau(0, x)$. We omit this result for succinctness.

It is worth comparing the point identification result with that in Chernozhukov and Hansen (2005). While they point identify $QTE_\tau(x)$ with a binary instrument by assuming rank similarity, the result of this section suggests that we may achieve the identification of $QTE_\tau(x)$ under rank linearity if we have a continuous instrument in the sense of (5.1). This is because rank linearity is equivalent to Condition S^* (by Theorem 4.3), but the latter implies Conditions S_1 and S_0 that identify $QTE_\tau(1, x)$ and $QTE_\tau(0, x)$, respectively.

6 Systematic Calculation of Bounds

In Theorem 2.1, there can be many γ's that satisfy the condition (2.6) (and $\tilde{\gamma}$ for (2.7)), especially with $L \geq 3$. This motivates the use of optimization in calculating the bounds via Corollary 2.1. We only focus on the upper bound and suppress X henceforth for brevity.

6.1 Semi-Infinite Programming

To simplify notation, let $\boldsymbol{p}(y,d) \equiv (p(y,d|z_1), ..., p(y,d|z_L))'$ where $p(y,d|z_\ell) \equiv P[Y \leq y, D = d|Z = z_\ell]$ and $p(y|d) \equiv P[Y \leq y|D = d]$. Also, let $\boldsymbol{1} \equiv (1, ..., 1)'$ and $\boldsymbol{p} \equiv (p(z_1), ..., p(z_L))'$ with $p(z) \equiv P[D = 1|Z = z]$ so that

$$\Gamma_p = \{\gamma : \gamma' [\begin{array}{cc} \boldsymbol{1} & \boldsymbol{p} \end{array}] = [\begin{array}{cc} 0 & 1 \end{array}]\} \subset \mathbb{R}^L.$$

Consider the following linear semi-infinite programming for the upper bound on $P[Y_0 \leq \bar{y}|D = 1]$:

$$UB(\bar{y}) = \min_{\gamma \in \Gamma_p} -\boldsymbol{p}(\bar{y}, 0)'\gamma \tag{6.1}$$

$$s.t. \quad \boldsymbol{p}(y, 1)'\gamma \geq p(y|1), \quad \forall y \in \mathcal{Y} \tag{6.2}$$

Note that the condition (2.6) guarantees the feasible set is non-empty. Also note that this condition is allowed to satisfy only almost everywhere (a.e.), which we suppress for simplicity. This program is infeasible to solve in practice as there are infinitely many constraints. We propose two approaches to approximate it with a linear program (LP).

6.2 Linear Program with Randomized Constraints

One approach to the semi-infinite program (6.1)–(6.2) is to approximate (6.2) using an i.i.d. simulated sample $\{Y_i\}_{i=1}^{s_n}$ as is done in the literature (e.g., Calafiore and Campi (2005)). This approach is also reminiscent of Han and Yang (2020). An obvious candidate of this sample would be $\{Y_i\}_{i=1}^{n}$ with $s_n = n$. Consider a sampled LP of the following:

$$\overline{UB}_n(\bar{y}) = \min_{\gamma \in \Gamma_p} -\boldsymbol{p}(\bar{y}, 0)'\gamma \tag{6.3}$$

$$s.t. \quad \boldsymbol{p}(Y_i, 1)'\gamma \geq p(Y_i|1). \quad \forall i = 1, ..., n \tag{6.4}$$

In the appendix, we show that the probability of violating the original constraints (6.2) by using (6.4) can be bounded by $O(1/n)$.

6.3 Dual Program and Sieve Approximation

Another approach to the semi-infinite program (6.1)–(6.2) is to invoke its dual and approximate the Lagrangian measure using sieve. With the constraint $p(y|1) - \boldsymbol{p}(y,1)'\gamma \leq 0$, the Lagrangian for (6.1)–(6.2) is

$$
\begin{aligned}
\mathcal{L}(\gamma, \Lambda, \lambda) &= -\boldsymbol{p}(\bar{y},0)'\gamma + \int_{\mathcal{Y}} [p(y|1) - \boldsymbol{p}(y,1)'\gamma] d\Lambda(y) + \lambda'([~\mathbf{1}~~\boldsymbol{p}~]'\gamma - [~0~~1~]') \\
&= \int_{\mathcal{Y}} p(y|1) d\Lambda(y) - [~0~~1~]\lambda + (\lambda'[~\mathbf{1}~~\boldsymbol{p}~]' - \int_{\mathcal{Y}} \boldsymbol{p}(y,1)' d\Lambda(y) - \boldsymbol{p}(\bar{y},0)')\gamma
\end{aligned}
$$

and

$$
UB(\bar{y}) = \min_{\gamma \in \mathbb{R}^L} \sup_{\Lambda \succeq 0, \lambda \in \mathbb{R}^2} \mathcal{L}(\gamma, \Lambda, \lambda),
$$

where Λ is a non-negative (not necessarily probability) measure (i.e., $\Lambda \succeq 0$) that assigns weights to binding constraints. Therefore, the dual problem to (6.1)–(6.2) is

$$
\widetilde{UB}(\bar{y}) = \sup_{\Lambda \succeq 0, \lambda \in \mathbb{R}^2} \int_{\mathcal{Y}} p(y|1) d\Lambda(y) - [~0~~1~]\lambda \tag{6.5}
$$

$$
s.t. \quad [~\mathbf{1}~~\boldsymbol{p}~]\lambda - \int_{\mathcal{Y}} \boldsymbol{p}(y,1) d\Lambda(y) - \boldsymbol{p}(\bar{y},0) = \mathbf{0}, \tag{6.6}
$$

which now has a finite number of constraints (i.e., L constraints). It is trivial to show weak duality, $\widetilde{UB}(\bar{y}) \leq UB(\bar{y})$.[5] It is conjectured that strong duality may also hold because of

---

[5]This is because, by (6.1)–(6.2) and (6.5)–(6.6), we have

$$
\begin{aligned}
-\boldsymbol{p}(\bar{y},0)'\gamma &= \left\{ \int_{\mathcal{Y}} \boldsymbol{p}(y,1) d\Lambda(y) - [~\mathbf{1}~~\boldsymbol{p}~]\lambda \right\}' \gamma = \int_{\mathcal{Y}} \boldsymbol{p}(y,1)'\gamma d\Lambda(y) - \lambda'[~\mathbf{1}~~\boldsymbol{p}~]'\gamma \\
&\geq \int_{\mathcal{Y}} p(y|1) d\Lambda(y) - \lambda'[~0~~1~]'.
\end{aligned}
$$

25

the structure of the problem (e.g., linearity, continuity of $\boldsymbol{p}(\cdot, d)$ and $p(\cdot|d)$, and etc.), which is yet to be explored. Note that $\Lambda(y)$ is smooth as the feasible set of the primal problem is smooth due to the smoothness of $p(y|d)$ and $\boldsymbol{p}(y, d)$, which are CDFs. This motivates us to use sieve approximation for $\Lambda(y)$ to turn the dual into a linear programming problem. The smoothness class for $\Lambda(y)$ will be determined by the smoothness class of CDFs. Let $\mathcal{Y}$ is normalized to be $[0, 1]$ and $\lambda(y) \equiv d\Lambda(y)/dy$. Consider the following sieve approximation:

$$\lambda(y) \approx \sum_{j=1}^{J} \theta_j b_j(y),$$

where $b_j(y) \equiv b_{j,J}(y)$ is a Bernstein basis function. Then, the LP can be written as

$$\widetilde{UB}_J(\bar{y}) = \max_{\theta \in \mathbb{R}_+^J, \lambda \in \mathbb{R}^2} \sum_{j=1}^{J} \theta_j b_j^1 - [\ 0\ \ 1\ ]\lambda$$

$$s.t. \quad [\ \mathbf{1}\ \ \boldsymbol{p}\ ]\lambda - \sum_{j=1}^{J} \theta_j \boldsymbol{b}_{1,j} - \boldsymbol{p}(\bar{y}, 0) = \mathbf{0},$$

or equivalently,

$$\widetilde{UB}_J(\bar{y}) = \max_{\theta \in \mathbb{R}_+^J, \lambda \in \mathbb{R}^2} \theta' b^1 - [\ 0\ \ 1\ ]\lambda \tag{6.7}$$

$$s.t. \quad [\ \mathbf{1}\ \ \boldsymbol{p}\ ]\lambda - B_1\theta - \boldsymbol{p}(\bar{y}, 0) = \mathbf{0}, \tag{6.8}$$

where $\theta \equiv (\theta_1, ..., \theta_J)'$, $b^d \equiv (b_1^d, ..., b_J^d)'$ with $b_j^d \equiv \int_{\mathcal{Y}} b_j(y)p(y|d)dy$, $\boldsymbol{b}_{d,j} \equiv (b_{d,j,1}, ..., b_{d,j,L})'$ with $b_{d,j,\ell} \equiv \int_{\mathcal{Y}} b_j(y)p(y, d|z_\ell)dy$, and $B_d \equiv [\ \boldsymbol{b}_{d,1}\ \ \cdots\ \ \boldsymbol{b}_{d,J}\ ]$ is an $L \times J$ matrix. Using Bernstein polynomials to approximate infinite-dimensional decision variables is also used in Han and Yang (2020).

Remark 6.1 (Local Approximation). *The LP* (6.7)–(6.8) *may be more stable than the LP* (6.3)–(6.4). *In terms of dual, the latter approach is equivalent to having* $\sum_{i=1}^{n} p(Y_i|1)\lambda_i$ *as an approximation for* $\int_{\mathcal{Y}} p(y|1)\lambda(y)dy$. *This can be viewed as a crude local approximation that involves a uniform kernel.*

26

7 Numerical Studies

To illustrate the importance of multiple IVs and the informativeness of resulting bounds, we conduct numerical exercises. We generate the data so that they are consistent with Model 1 and hence satisfy Condition S_1^*. The variables (Y, D, Z) are generated in the following fashion:

- $Y_d = q(d, U_d) = 1 - d + (d+1)U_d$ for $\mathcal{Y} = \mathbb{R}$, that is, $Y_1 = 2U_1$ and $Y_0 = 1 + U_0$

- $(U, \eta) \sim BVN((0,0)', \Sigma)$

- $V \sim N(0, \sigma_V^2)$ and $\xi_1 \sim N(0, \sigma_\xi^2)$

- $\xi_0 = \xi_1 + V$

- $U_d = U + \xi_d$

- $Z \sim Bin(L-1, p)/(L-1) \in [0,1]$ with $L \in \{2, 3, 4, 5, 6, 7, 8\}$

- $D = 1\{\pi_0 + \pi_1 Z \geq \eta\}$

- $Y = DY_1 + (1 - D)Y_0$

Here, Z is normalized so that the endpoints of the support are invariant regardless of the value of L. This is intended to understand the role of the number of values Z takes while fixing the role of instrument strength. Figures 1–3 presents the bounds on $\Pr[Y_0 \leq y | D = 1]$ while varying L. The bounds are calculated using the approach proposed in Section 6.2. We only report $L \in \{2, 5, 6\}$ for succinctness. In these figures, the black solid line indicates the true value of $\Pr[Y_0 \leq y | D = 1]$ and the red and blue crosses depict the upper and lower bounds. Although the upper bound is a trivial upper bound for the CDF when $L = 2$, it quickly becomes informative as L increases beyond 5. To put this in a context, this corresponds to the number of instrument values that three binary IVs can easily surpass or a single continuous IV.
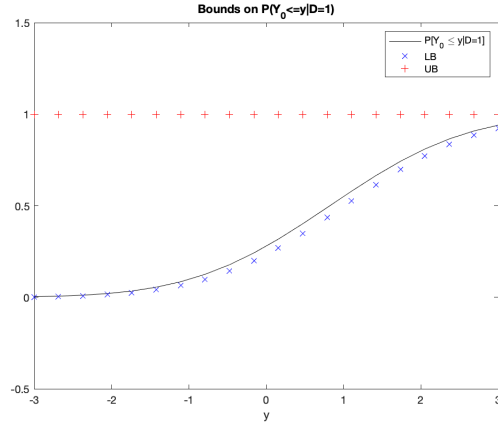
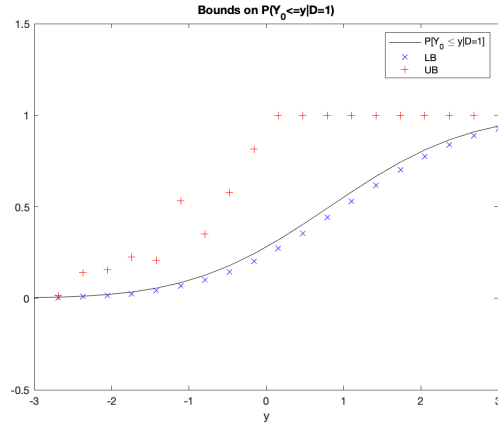Figure 1: Bounds on $\Pr[Y_0 \leq y | D = 1]$ When $L = 2$

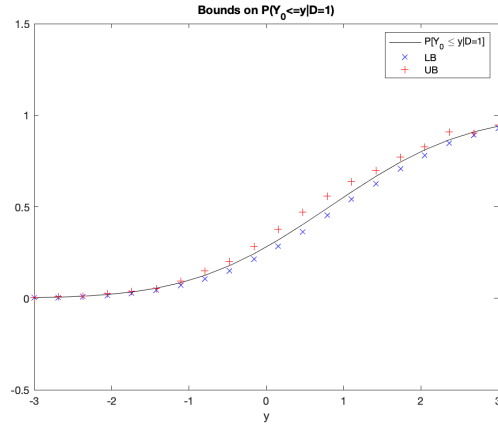Figure 2: Bounds on $\Pr[Y_0 \leq y | D = 1]$ When $L = 5$

Figure 3: Bounds on $\Pr[Y_0 \leq y | D = 1]$ When $L = 6$

A   Conditions for Average Treatment Effects

To calculate bounds on $ATE(1,x)$ and $ATE(0,x)$, we introduce conditions that are weaker that Conditions S_1 and S_0.

Condition S$_1'$. *For arbitrary non-negative weight vectors $(w_1,...,w_L)$ and $(\tilde{w}_1,...,\tilde{w}_L)$ that satisfy $\sum_{\ell=1}^{L} w_\ell = \sum_{\ell=1}^{L} \tilde{w}_\ell = 1$, if*

$$\sum_{\ell=1}^{L} w_\ell P[Y_1 \leq \cdot | D = 1, Z = z_\ell, X = x] \leq \sum_{\ell=1}^{L} \tilde{w}_\ell P[Y_1 \leq \cdot | D = 1, Z = z_\ell, X = x], \quad (A.1)$$

then

$$\sum_{\ell=1}^{L} w_\ell E[Y_0 | D = 1, Z = z_\ell, X = x] \leq \sum_{\ell=1}^{L} \tilde{w}_\ell E[Y_0 | D = 1, Z = z_\ell, X = x]. \quad (A.2)$$

Condition S_1' can be used to bound the $ATE(1,x)$. An analogous condition can be imposed to bound $ATE(0,x)$.

Condition S$_0'$. *For arbitrary non-negative weight vectors $(w_1,...,w_L)$ and $(\tilde{w}_1,...,\tilde{w}_L)$ that satisfy $\sum_{\ell=1}^{L} w_\ell = \sum_{\ell=1}^{L} \tilde{w}_\ell = 1$, if*

$$\sum_{\ell=1}^{L} w_\ell P[Y_0 \leq \cdot | D = 1, Z = z_\ell, X = x] \leq \sum_{\ell=1}^{L} \tilde{w}_\ell P[Y_0 \leq \cdot | D = 1, Z = z_\ell, X = x], \quad (A.3)$$

then

$$\sum_{\ell=1}^{L} w_\ell E[Y_1 | D = 1, Z = z_\ell, X = x] \leq \sum_{\ell=1}^{L} \tilde{w}_\ell E[Y_1 | D = 1, Z = z_\ell, X = x]. \quad (A.4)$$

B   Other Structural Models as Sufficient Conditions

We present two more structural models that are not nested to Model 1 in the text. Model1(i)–(ii) are maintained in these models, that is, (i) $Y_d = q(d, X, U_d)$ for $d \in \{0, 1\}$ where $q(d, x, \cdot)$ is continuous and monotone increasing and (ii) $D = h(Z, X, \eta)$.

Model 2. (iii) $U_0 \stackrel{d}{=} \phi(U_1, V)$ conditional on (η, X) where $V \perp (U_1, \eta)|X$ and $\phi(\cdot, v)$ is strictly increasing for all v.

Model 2(iii) defines that U_0 is "noisier" than U_1. Therefore, Model 2 is weaker than the model in Chernozhukov and Hansen (2005). Model 2 and Model 1 are not nested because, in $U_0 = U_1 + V$ of Model 1, V is not independent of U_1. We show below that Model 2 implies Condition S_1. Interestingly, Model 2(iii) with $U_0 = \phi(U_1, V)$ (instead of "$\stackrel{d}{=}$") is a generalization of the definition that U_0 is "noisier" than U_1 if $U_0 = U_1 + V$ with $U_1 \perp V$ in Pomatto et al. (2020, p. 1880).

Model 3. (iii) $U_0 \stackrel{d}{=} \max\{\phi(U_1), V\}$ conditional on (η, X) where $V \perp (U_1, \eta)|X$ and $\phi(\cdot)$ is strictly increasing.

We show below that Model 3 implies rank linearity. Model 3 can alternatively be defined as follows: $Y_0 \stackrel{d}{=} \max\{\phi(Y_1), V\}$ conditional on (η, X) where $V \perp (Y_1, \eta)|X$ and $\phi(\cdot)$ is strictly increasing. Then, this model also implies rank linearity with $\psi(\cdot) = \phi^{-1}(\cdot)$ because

$$\Pr[Y_0 \leq y|\eta, X] = \Pr[\phi(Y_1) \leq y, V \leq y|\eta, X] = \Pr[Y_1 \leq \phi^{-1}(y)|\eta, X]\Pr[V \leq y|X].$$

This model provides another interpretation of an insurance policy $(D = 1)$ as $Y_1 = \max\{Y_0, V\}$ guarantees at least Y_0. Models 2 and 3 are not nested.

Lemma B.1. *(i) Model 2 implies Condition S_1^*; (ii) Model 3 implies rank linearity.*

The proof of this lemma is contained in the next section.

C Proofs

C.1 Proof of Lemma 2.1

Let $p(z,x) \equiv P[D = 1|Z = z, X = x]$ and let $H(z,x) \equiv \{\eta : h(z,x,\eta) = 1\}$ be a level set. Then,

$$\sum_\ell w_\ell P[Y_1 \leq y|D = 1, Z = z_\ell, X = x] = \sum_\ell w_\ell P[Y_1 \leq y|\eta \in H(z_\ell, x), X = x]$$

$$= \int \frac{\sum_\ell w_\ell 1[t \in H(z_\ell, x)]}{p(z_\ell, x)} P[Y_1 \leq y|\eta = t, X = x]dt.$$

Take $w(t,x) = \frac{\sum_\ell w_\ell 1[t \in H(z_\ell,x)]}{p(z_\ell,x)}$. Then, $w(t,x)$ satisfies

$$\int \frac{\sum_\ell w_\ell 1[t \in H(z_\ell, x)]}{p(z_\ell, x)} dt = 1.$$

The same argument applies to $\tilde{w}$ and $\tilde{w}(t,x)$, and also for the distribution of Y_0. $\square$

C.2 Proof of Lemma 3.1

Let $p(1) > p(0)$ without loss of generality. Because

$$P[Y \leq y|D = 1]$$

$$= P[Y \leq y|D = 1, Z = 1]P[Z = 1|D = 1] + P[Y \leq y|D = 1, Z = 0]P[Z = 0|D = 1]$$

$$= P[Y \leq y|D = 1, Z = 1]\frac{p(1)P[Z = 1]}{P[D = 1]} + P[Y \leq y|D = 1, Z = 0]\frac{p(0)P[Z = 0]}{P[D = 1]},$$

the inequality (2.6) with $(\gamma_1, \gamma_2) = \left(\frac{1}{p(1)-p(0)}, -\frac{1}{p(1)-p(0)} \right)$ is equivalent to

$$P[Y \le y|D = 1, Z = 1]\frac{[p(1) - p(0)]p(1)P[Z = 1]}{P[D = 1]}$$
$$+ P[Y \le y|D = 1, Z = 0]\frac{[p(1) - p(0)]p(0)P[Z = 0]}{P[D = 1]}$$
$$\le P[Y \le y|D = 1, Z = 1]p(1) - P[Y \le y|D = 1, Z = 0]p(0).$$

Hence,

$$P[Y \le y|D = 1, Z = 0]p(0) + P[Y \le y|D = 1, Z = 0]\frac{[p(1) - p(0)]p(0)P[Z = 0]}{P[D = 1]}$$
$$\le P[Y \le y|D = 1, Z = 1]p(1) - P[Y \le y|D = 1, Z = 1]\frac{[p(1) - p(0)]p(1)P[Z = 1]}{P[D = 1]},$$

that is,

$$P[Y \le y|D = 1, Z = 0]p(0)\left[1 + \frac{[p(1) - p(0)]P[Z = 0]}{P[D = 1]}\right]$$
$$\le P[Y \le y|D = 1, Z = 1]p(1)\left[1 - \frac{[p(1) - p(0)]P[Z = 1]}{P[D = 1]}\right].$$

Note that $P[D = 1] = p(1)P[Z = 1] + p(0)P[Z = 0]$. Therefore,

$$P[Y \le y|D = 1, Z = 0]p(0)p(1) \le P[Y \le y|D = 1, Z = 1]p(1)p(0).$$

Note that $p(z) \ne 0$ for $z = 0, 1$, and therefore (γ_1, γ_2) are well-defined. Then we have

$$P[Y \le y|D = 1, Z = 0] \le P[Y \le y|D = 1, Z = 1],$$

which proves the first claim of the lemma.

Next, the inequality (1.3) can equivalently be written as

$$P[Y \leq y, D = 1|Z = 1]/p(1) \leq P[Y \leq y, D = 1|Z = 0]/p(0).$$

Then, by the assumed selection equation which excludes defiers, note that $P[Y \leq y, D = 1|Z = z] = P[Y_1 \leq y, D_z = 1]$ by the joint independence assumption and that $P[Y_1 \leq y, D_1 = 1] = P[Y_1 \leq y, AT] + P[Y_1 \leq y, C]$ and $P[Y_1 \leq y, D_0 = 1] = P[Y_1 \leq y, AT]$. Also, note that $p(1) = P(C) + P(AT)$ and $p(0) = P(AT)$ by the selection equation. Then, by a simple algebra, we can show that (1.3) is equivalent to

$$P[Y_1 \leq y|C] \leq P[Y_1 \leq y|AT], \quad \forall y. \tag{C.1}$$

□

C.3 Proof of the Argument in Remark 3.1

Let $\mathcal{Z} = \{z_1, ..., z_L\} = \{0, ..., L - 1\}$ and let p_j for $j = \{1, ..., L - 1\}$ is the proportion of j-complier. When $L = 3$, for example, $p_1 = P[1\text{-C}] = P[\text{E-C}]$ and $p_2 = P[2\text{-C}] = P[\text{R-C}]$. Also, let $p_0 = P[AT]$. Then, for $j = \{1, ..., L - 1\}$,

$$P[Y_d \leq y|D_j = 1] = P\left[Y_d \leq y| \text{AT} \cup \bigcup_{j'=1}^{j} j'\text{-C}\right]$$

by Assumption D^*. In (3.1), we first set $w_1 = \tilde{w}_1 = 0$ to remove the distribution for AT, that is, $P[Y_d \leq y|D_0 = 1] = P[Y_d \leq y|\text{AT}]$ which is not identified. Then, we have

$$\sum_{\ell=2}^{L} w_\ell P[Y_d \leq y|D_{z_\ell} = 1] = \sum_{\ell=2}^{L} \frac{w_\ell}{\sum_{j=0}^{\ell-1} p_j} \left(p_0 P[Y_d \leq y|\text{AT}] + \sum_{j=1}^{\ell-1} p_j P[Y_d \leq y|j\text{-C}]\right)$$

33

and an analogous expression for $\sum_{\ell=2}^{L} \tilde{w}_\ell P[Y_d \leq y | D_{z_\ell} = 1]$. Therefore, to remove the remaining distribution for AT in (3.1), we set

$$\sum_{\ell=2}^{L} \frac{w_\ell}{\sum_{j=0}^{\ell-1} p_j} = \sum_{\ell=2}^{L} \frac{\tilde{w}_\ell}{\sum_{j=0}^{\ell-1} p_j}. \tag{C.2}$$

Note that when $L \geq 4$, $w \neq \tilde{w}$ even if w and $\tilde{w}$ satisfy (C.2). Then, the resulting (3.1) is the dominance between the two distinct weight sums of $P[Y_d \leq y | j\text{-C}]$'s, which are identified for all $d = 1, 0$. $\square$

C.4 Proof of Theorem 4.1

We suppress X for simplicity. For an arbitrary r.v. A, let $F_A^w(\cdot) \equiv \int w(t) F_{A|\eta}(\cdot|t) dt$, which itself is a CDF. By Model 1(i), $F_{Y_d}^w \leq F_{Y_d}^{\tilde{w}}$ if and only if $F_{U_d}^w \leq F_{U_d}^{\tilde{w}}$. So it suffices to show that, if $F_{U_1}^w \leq F_{U_1}^{\tilde{w}}$, then $F_{U_0}^w \leq F_{U_0}^{\tilde{w}}$.

Let $G(\cdot)$ be an arbitrary monotone increasing function and $g(\cdot) \equiv G'(\cdot)$. Note that

$$\int G dF_{U_0}^w - \int G dF_{U_0}^{\tilde{w}} = \int [\int \tilde{w}(t) F_{U_0|\eta}(u|t) dt - \int w(t) F_{U_0|\eta}(u|t) dt] g(u) du$$

$$= \int [\int \tilde{w}(t) \int F_{U|\eta}(u - s|t) f_{\xi_0}(s) ds dt - \int w(t) \int F_{U|\eta}(u - s|t) f_{\xi_0}(s) ds dt] g(u) du$$

$$= \int \int \int [\tilde{w}(t) - w(t)] F_{U|\eta}(u|t) f_{\xi_0}(s) g(u + s) du ds dt,$$

where the first eq. is due to the integration by part, the second eq. is by $F_{U_d|\eta}(u|t) = \int F_{U|\eta}(u - s|t) f_{\xi_0|\eta}(s|t) ds = \int F_{U|\eta}(u - s|t) f_{\xi_0}(s) ds$ under Model 1(iii), and the last eq. is by change of variables. By Model 1(iv), $f_{\xi_0}(s) = \int f_{\xi_1}(s - v) f_V(v) dv = \int f_{\xi_1}(v) f_V(s - v) dv$

where $f_A(\cdot)$ is the PDF of an arbitrary r.v. A. Therefore,

$$\int G dF_{U_0}^w - \int G dF_{U_0}^{\tilde{w}}$$
$$= \int \int \int [\tilde{w}(t) - w(t)] F_{U|\eta}(u|t) \int f_{\xi_1}(v) f_V(s - v) g(u + s) dv du ds dt$$
$$= \int \int [\tilde{w}(t) - w(t)] F_{U|\eta}(u|t) \int f_{\xi_1}(v) [\int f_V(s) g(u + s + v) ds] dv du dt.$$

Let $\psi(s) \equiv \int f_V(t) g(t + s) dt$. By definition, $\psi \geq 0$ since $g \geq 0$. Therefore,

$$\int G dF_{U_0}^w - \int G dF_{U_0}^{\tilde{w}}$$
$$= \int \int [\tilde{w}(t) - w(t)] F_{U|\eta}(u|t) \int f_{\xi_1}(v) \psi(u + v) dv du dt$$
$$= \int \int [\tilde{w}(t) - w(t)] \int F_{U|\eta}(u - v|t) f_{\xi_1}(v) dv \psi(u) du dt$$
$$= \int \int [\tilde{w}(t) - w(t)] \int F_{U_1|\eta}(u|t) \psi(u) du dt$$
$$= \int [\int \tilde{w}(t) F_{U_1|\eta}(u|t) dt - \int w(t) F_{U_1|\eta}(u|t) dt] \psi(u) du \geq 0,$$

where the last ineq. is by $F_{U_1}^w \leq F_{U_1}^{\tilde{w}}$. Because $G(\cdot)$ is arbitrary, then $F_{U_0}^w$ is first order stochastic dominant over $F_{U_0}^{\tilde{w}}$. $\square$

C.5   Proof of Theorem 4.3: Equivalence Between Rank Linearity and Condition S*

The "if" part is trivial. We prove "only if" part. Suppose Condition S* holds. Let $\mathcal{Y}_\infty \equiv \{y_k \in \mathbb{R} : k = 1, \cdots, \infty\}$ be a sequence that is dense on $\mathbb{R}$. Denote $\mathcal{Y}_n \equiv \{y_k \in \mathbb{R} : k = 1, \cdots, n\}$. Because $\mathcal{Y}_\infty$ is dense in $\mathbb{R}$ and CDFs are right-continuous, it suffices to show the existence of $\lambda(\cdot)$ and $\psi(\cdot)$ on $\mathcal{Y}_\infty$ such that

$$F_{Y_0|\eta}(\psi(y)|t) = \lambda(y) F_{Y_1|\eta}(y|t) \tag{C.3}$$

holds for all $t \in \mathcal{T}$ and $y \in \mathcal{Y}_\infty$.

Fix $n \in \mathbb{N}$. Let $G_{1,k} : \mathbb{R} \to \{0,1\}$ be a simple function defined as $G_{1,k}(\cdot) \equiv 1\{y_k \le \cdot\}$ for $k = 1, \cdots, n$. By the full rank condition (4.3), for each $1 \le k \le n$, there exists a function $c_k : \mathcal{T} \to \mathbb{R}$ such that

$$G_{1,k}(\cdot) = \int c_k(t) F_{Y_1|\eta}(\cdot|t) dt.$$

Define $G_{0,k} : \mathbb{R} \to [0,1]$ as

$$G_{0,k}(\cdot) \equiv \int c_k(t) F_{Y_0|\eta}(\cdot|t) dt.$$

Note that $G_{0,k}$ is a proper CDF. Now, for any vectors $\pi \equiv (\pi_1, \cdots, \pi_n)$ and $\tilde{\pi} \equiv (\tilde{\pi}_1, \cdots, \tilde{\pi}_n)$ such that $\sum_{k=1}^n \pi_k = \sum_{k=1}^n \tilde{\pi}_k = 1$, suppose

$$\sum_{k=1}^n \pi_k G_{1,k}(\cdot) \le \sum_{k=1}^n \tilde{\pi}_k G_{1,k}(\cdot).$$

It follows that

$$\int b_n(t) F_{Y_1|\eta}(\cdot|t) dt \le \int \tilde{b}_n(t) F_{Y_1|\eta}(\cdot|t) dt,$$

where $b_n(t) \equiv \sum_{k=1}^n \pi_k c_k(t)$ and $\tilde{b}_n(t) \equiv \sum_{k=1}^n \tilde{\pi}_k c_k(t)$. Let $b_n^+(t) = \max\{b_n(t), 0\}$ and $b_n^-(t) = \min\{b_n(t), 0\}$ and similarly define $\tilde{b}_n^+(t)$ and $\tilde{b}_n^-(t)$. Then, the above inequality can be written as

$$\int \{b_n^+(t) - \tilde{b}_n^-(t)\} F_{Y_1|\eta}(\cdot|t) dt \le \int \{\tilde{b}_n^+(t) - b_n^-(t)\} F_{Y_1|\eta}(\cdot|t) dt,$$

where the resulting weight functions on both sides are non-negative. Then, by Condition S^*, we have

$$\sum_{k=1}^n \pi_k G_{0,k}(\cdot) \le \sum_{k=1}^n \tilde{\pi}_k G_{0,k}(\cdot)$$

By a similar argument, the converse is also true and thus we have

$$\sum_{k=1}^n \pi_k G_{1,k}(\cdot) \le \sum_{k=1}^n \tilde{\pi}_k G_{1,k}(\cdot).$$

if and only if

$$\sum_{k=1}^{n} \pi_k G_{0,k}(\cdot) \leq \sum_{k=1}^{n} \tilde{\pi}_k G_{0,k}(\cdot)$$

for any non-negative weights π and $\tilde{\pi}$. Therefore, it follows that

$$\sum_{k=1}^{n} \delta_k G_{1,k}(\cdot) \leq 0 \text{ if and only if } \sum_{k=1}^{n} \delta_k G_{0,k}(\cdot) \leq 0 \qquad (\text{C}.4)$$

for any n-dimensional vector $\delta \equiv (\delta_1, \cdots, \delta_n)$ that satisfies $\sum_{k=1}^{n} \delta_k = 0$.

For $d \in \{0, 1\}$, define

$$\Delta_d^G \equiv \left\{ \delta \in \mathbb{R}^n : \sum_{k=1}^{n} \delta_k G_{d,k}(y) \leq 0 \ \forall y \in \mathbb{R}; \sum_{k=1}^{n} \delta_k = 0 \right\}.$$

Note that $\{(G_{1,1}(y), \cdots, G_{1,n}(y)) : y \in \mathbb{R}\} = \{(G_{1,1}(y), \cdots, G_{1,n}(y)) : y \in \mathcal{Y}_n\}$ by definition. Therefore, Δ_1^G is a *finite* cone and its dimension is $n - 1$. Define the polar cone of Δ_d^G as $\Delta_d^{G*} \equiv \{G_d \in \mathbb{R}^n : G_d'\delta \leq 0, \forall \delta \in \Delta_d^G\}$. Note that by definition, $(G_{1,1}(y), \cdots, G_{1,n}(y))$ for $y \in \mathcal{Y}_n/\{y_n\}$ are $n - 1$ linearly independent vectors and therefore generate extreme rays of Δ_1^{G*}. Also note that any element in Δ_0^{G*} is written as $(G_{0,1}(y), \cdots, G_{0,n}(y))$ for some $y \in \mathbb{R}$, and so is a vector that generates its extreme ray. But by (C.4), we have that $\Delta_1^G = \Delta_0^G$ and thus $\Delta_1^{G*} = \Delta_0^{G*}$, and therefore, for each $y_k \in \mathcal{Y}_n/\{y_n\}$, there exists $y_k^* \in \mathbb{R}$ and $\lambda_n(\cdot) > 0$ such that

$$(G_{0,1}(y_k^*), \cdots, G_{0,n}(y_k^*)) = \lambda_n(y_k) \times (G_{1,1}(y_k), \cdots, G_{1,n}(y_k)). \qquad (\text{C}.5)$$

If there exists multiple values of y_k^* satisfying (C.5), we define y_k^* as the infimum of $\{\tilde{y}_k^* : (G_{0,1}(\tilde{y}_k^*), \cdots, G_{0,n}(\tilde{y}_k^*)) = \lambda_n(y_k) \times (G_{1,1}(y_k), \cdots, G_{1,n}(y_k))\}$. Because CDFs are right-continuous function, the infimum should also satisfy (C.5).

For any $j, k = 1, ..., n$, if $G_{1,j}(y_k) = 0$ then $G_{0,j}(y_k^*) = 0$ by (C.5), which further implies that $G_{0,j}(y^*) = 0$ for all $y^* \leq y_k^*$ because $G_{0,j}(\cdot)$ is monotone increasing. Let $\{j_1, \cdots, j_n\}$ be a permutation of $\{1, \cdots, n\}$ such that $y_{j_1} < y_{j_2} < \cdots < y_{j_n}$. Note that $G_{1,j_1}(y_{j_1})$ is the only non-zero component in the set $\{G_{1,k}(y_{j_1}) : k = 1, \cdots, n\}$. Then, by (C.5), $G_{0,j_1}(y_{j_1}^*) \neq 0$ and

$G_{0,j_k}(y^*_{j_1}) = 0$ for $k \geq 2$. Similarly, there are two elements of $\{G_{0,k}(y^*_{j_2}) : k = 1, \cdots, n\}$ which are non-zero, namely, $G_{0,j_1}(y^*_{j_2})$ and $G_{0,j_2}(y^*_{j_2})$. Therefore, by $G_{0,j_2}(y^*_{j_1}) = 0$ and $G_{0,j_2}(y^*_{j_2}) \neq 0$ and the fact that $G_{0,j_2}(\cdot)$ is monotone increasing, we can conclude $y^*_{j_1} < y^*_{j_2}$. Continuing this argument, we can conclude that

$$y^*_{j_1} < y^*_{j_2} < \cdots < y^*_{j_n}.$$

Define a function $\psi_n : y_k \mapsto y^*_k$ for $k = 1, \cdots, n$. By the above analysis, $\psi_n(\cdot)$ is a monotone increasing function. Note that the support of ψ_n is $\mathcal{Y}_n$, which we extend to $\mathbb{R}$ as follows: for any $y \in \mathbb{R}$,

$$\psi_n(y) = \begin{cases} \max\{\psi_n(y_k) : y_k \leq y, \ k = 1, \cdots, n\} & \text{if } y \geq \min\{y_1, \cdots, y_n\} \\ -\infty & \text{otherwise} \end{cases}$$

Then, $\psi_n : \mathbb{R} \to \mathbb{R}$ is still a monotone increasing function.

We now consider increasing n to $n + 1$. By a similar argument, there exists a sequence $\{y^\dagger_1, \cdots, y^\dagger_n, y^\dagger_{n+1}\}$ and $\lambda_{n+1}(\cdot) > 0$ such that for $k = 1, \cdots, n + 1$, we have

$$\left(G_{0,1}(y^\dagger_k), \cdots, G_{0,n}(y^\dagger_k), G_{0,n+1}(y^\dagger_k)\right) = \lambda_{n+1}(y_k) \times \left(G_{1,1}(y_k), \cdots, G_{1,n}(y_k), G_{1,n+1}(y_k)\right), \quad \text{(C.6)}$$

If there exists multiple values of $y^\dagger_k$, we define $y^\dagger_k$ as the infimum of them. Note that, by (C.6) and (C.5), $y^\dagger_k$ is one of the candidates $\tilde{y}^*_k$'s that make $\left(G_{0,1}(\tilde{y}^*_k), \cdots, G_{0,n}(\tilde{y}^*_k)\right)$ proportional to $\left(G_{1,1}(y_k), \cdots, G_{1,n}(y_k)\right)$ satisfy (C.5). While y^*_k is the infimum of those candidates, $y^\dagger_k$ cannot reach that infimum because it has to satisfies the additional restriction, $G_{0,n+1}(y^\dagger_k) = \lambda_{n+1}(y_k)G_{1,n+1}(y_k)$. Therefore, we can conclude that $y^\dagger_k \geq y^*_k$ for $k = 1, \cdots, n$. Using $\{y_1, \cdots, y_n, y_{n+1}\}$ and $\{y^\dagger_1, \cdots, y^\dagger_n, y^\dagger_{n+1}\}$, define $\psi_{n+1}(\cdot)$ analogous to $\psi_n(\cdot)$ above. Then, $\psi_{n+1}(y_k) = y^\dagger_k \geq y^*_k = \psi_n(y_k)$ for $k = 1, \cdots, n$. Furthermore, by definition, $\psi_{n+1}(y_{n+1}) \geq$

$\psi_n(y_{n+1})$ regardless of the rank order of y_{n+1} in $\mathcal{Y}_{n+1}$. Therefore, for any $y \in \mathbb{R}$,

$$\psi_{n+1}(y) \geq \psi_n(y),$$

and thus the limit of the sequence of functions $\psi_n(\cdot)$ exists as $n \to \infty$, which we denote as $\psi_\infty(\cdot)$. Recall each $\psi_n(\cdot)$ is weakly increasing. It is easy to prove by contradiction that $\psi_\infty(\cdot)$ is strictly increasing. Fix $y_k \in \mathcal{Y}_\infty$. For any $n \geq k$, $\big(G_{0,1}(\psi_\infty(y_k)), \cdots, G_{0,n}(\psi_\infty(y_k))\big)$ is proportional to $\big(G_{1,1}(y_k), \cdots, G_{1,n}(y_k)\big)$ and therefore there exists $\lambda_\infty(y_k)$ such that

$$\big(G_{0,1}(\psi_\infty(y_k)), \cdots, G_{0,n}(\psi_\infty(y_k))\big) = \lambda_\infty(y_k) \times \big(G_{1,1}(y_k), \cdots, G_{1,n}(y_k)\big) \qquad (\text{C.7})$$

for any $n \in \mathbb{N}$. Moreover, because $\mathcal{Y}_\infty$ is dense in $\mathbb{R}$ and ψ_∞ and $G_{d,k}$ are right-continuous functions, the above condition holds for all $y \in \mathbb{R}$.

Note $\{G_{1,k}(\cdot) : k = 1, \cdots, \infty\}$ is a class of simple functions. Therefore, any $F_{Y_1|\eta}(\cdot|t)$ can be written as

$$F_{Y_1|\eta}(\cdot|t) = \lim_{K\to\infty} \sum_{k=1}^{K} a_{K,k}(t) G_{1,k}(\cdot)$$

for some triangular array $\{a_{Kk}(t) : 1 \leq k \leq K, K = 1, 2, \cdots, \infty\}$. By the definition of $G_{1,k}(\cdot)$, it follows that

$$F_{Y_1|\eta}(\cdot|t) = \lim_{K\to\infty} \sum_{k=1}^{K} a_{K,k}(t) \int w_k(s) F_{Y_1|\eta}(\cdot|s) ds = \int \lim_{K\to\infty} \sum_{k=1}^{K} a_{K,k}(t) w_k(s) F_{Y_1|\eta}(\cdot|s) ds$$

$$\equiv \int \kappa(t,s) F_{Y_1|\eta}(\cdot|s) ds, \qquad (\text{C.8})$$

where $\kappa(t,s) \equiv \lim_{K\to\infty} \sum_{k=1}^{K} a_{K,k}(t) w_k(s)$ serves as a Dirac delta function. Because $F_{Y_1|\eta}(\cdot|t) = \int \kappa(t,s) F_{Y_1|\eta}(\cdot|s) ds$ if and only if $F_{Y_1|\eta}(\cdot|t) \leq \int \kappa(t,s) F_{Y_1|\eta}(\cdot|s) ds$ and $F_{Y_1|\eta}(\cdot|t) \geq \int \kappa(t,s) F_{Y_1|\eta}(\cdot|s) ds$, we have, by Condition S*,

$$F_{Y_0|\eta}(\cdot|t) = \int \kappa(t,s) F_{Y_0|\eta}(\cdot|s) ds = \lim_{K\to\infty} \sum_{k=1}^{K} a_{K,k}(t) G_{0,k}(\cdot) \qquad (\text{C.9})$$

using the definition of $G_{0,k}(\cdot)$. Combining (C.8), (C.9) and (C.7), for any $y \in \mathbb{R}$ and $t \in \mathcal{T}$, we have

$$F_{Y_0|\eta}(\psi_\infty(y)|t) = \lim_{K\to\infty} \sum_{k=1}^{K} a_{K,k}(t)G_{0,k}(\psi_\infty(y)) = \lim_{K\to\infty} \sum_{k=1}^{K} a_{K,k}(t)\lambda_\infty(y)G_{1,k}(y)$$

$$= \lambda_\infty(y)F_{Y_1|\eta}(y|t),$$

which completes the proof. $\square$

C.6 Equivalence Between Rank Linearity and Condition $\mathbf{S}^*$: Discrete Y_d

For $d \in \{0, 1\}$, suppose Y_d and η are discretely distributed. Specifically, let $\mathcal{Y}_d \equiv \{y_{d,1}, \cdots, y_{d,K_d}\}$ and $\mathcal{T} \equiv \{t_1, \cdots, t_{K_\eta}\}$ be the support of Y_d and η, respectively. Note that even with $K_0 = K_1$, we allow that Y_0 and Y_1 have different supports (i.e., there can be a "drift"). Suppress (Z, X) for simplicity.

Condition C.1. *For arbitrary non-negative weights $\{w_1, \cdots, w_{K_\eta}\}$ and $\{\tilde{w}_1, \cdots, \tilde{w}_{K_\eta}\}$ such that $\sum_{k=1}^{K_\eta} w_k = 1$ and $\sum_{k=1}^{K_\eta} \tilde{w}_k = 1$, it holds that*

$$\sum_{k=1}^{K_\eta} w_k F_{Y_1|\eta}(\cdot|t_k) \leq \sum_{k=1}^{K_\eta} \tilde{w}_k F_{Y_1|\eta}(\cdot|t_k)$$

if and only if

$$\sum_{k=1}^{K_\eta} w_k F_{Y_0|\eta}(\cdot|t_k) \leq \sum_{k=1}^{K_\eta} \tilde{w}_k F_{Y_0|\eta}(\cdot|t_k).$$

Theorem C.1. *For any probability distribution function $\tilde{F}_d$ supported on $\mathcal{Y}_d \equiv \{y_{d,1}, \cdots, y_{d,K_d}\}$, suppose there always exists a sequence $\{c_{d,1}, \cdots, c_{d,K_\eta}\}$ such that*

$$\tilde{F}_d(\cdot) = \sum_{k=1}^{k_\eta} c_{d,k} F_{Y_d|\eta}(\cdot|t_k), \tag{C.10}$$

Then, Condition C.1 holds if and only if (i) $K_0 = K_1$ and (ii) for some strictly increasing mapping $\psi : \{y_{0,1}, \cdots, y_{0,K_0}\} \to \{y_{1,1}, \cdots, y_{1,K_1}\}$ and some $\lambda : \{y_{0,1}, \cdots, y_{0,K_0}\} \to \mathbb{R}_+$,

$$F_{Y_0|\eta}(y|t_k) = \lambda(y) F_{Y_1|\eta}(\psi(y)|t_k), \qquad \text{for } y \in \mathcal{Y}_0, k = 1, \cdots, K_\eta. \tag{C.11}$$

The condition C.10 is a rank condition as the rank of matrix $\{F_{Y_d|\eta}(y_{d,j}|t_{j'}) : j = 1, ..., K_d, \quad j' = 1, \cdots, k_\eta\}$ should be no smaller than K_d. A necessary condition is $K_\eta \geq K_d$, namely, the support of η is no coarser than the support of Y_d. The rank condition would be violated when there is no endogeneity (i.e., $Y_d \perp \eta$), which is not our focus. Again, the rank condition is only introduced in this theorem to establish the relationship between rank linearity (and hence rank similarity) and the range of identifying conditions of this paper, and it is not necessary for our bound analysis.

Proof. By Condition C.1, we have

$$\sum_{k=1}^{K_\eta} \delta_k F_{Y_1|\eta}(\cdot|t_k) \leq 0 \text{ if and only if } \sum_{k=1}^{K_\eta} \delta_k F_{Y_0|\eta}(\cdot|t_k) \leq 0 \tag{C.12}$$

for any K_η-dimensional vector $\delta \equiv (\delta_1, \cdots, \delta_n)$ that satisfies $\sum_{k=1}^{K_\eta} \delta_k = 0$.

Note that $\big(F_{Y_1|\eta}(y|t_1), \cdots, (F_{Y_1|\eta}(y|t_{K_\eta}))\big)$ for each $y \in \mathcal{Y}_1/\{y_{K_1}\}$ generates an extreme ray of the $(K_\eta - 1)$-dimensional polar cone of a cone

$$\left\{ \delta \in \mathbb{R}^n : \sum_{k=1}^{K_\eta} \delta_k F_{Y_1|\eta}(\cdot|t_k) \leq 0; \sum_{k=1}^{K_\eta} \delta_k = 0 \right\}.$$

A similar argument holds for $\big(F_{Y_0|\eta}(\cdot|t_1), \cdots, (F_{Y_0|\eta}(\cdot|t_{K_\eta}))\big)$. By (C.12), these two polar cones are the same. Therefore, for each $y_k \in \mathcal{Y}_1/\{y_{K_1}\}$, there exists a $y_k^* \in \mathcal{Y}_0/\{y_{K_0}\}$ such that

$$\big(F_{Y_0|\eta}(y_k^*|t_1), \cdots, F_{Y_0|\eta}(y_k^*|t_{K_0}))\big) = \lambda(y_k) \times \big(F_{Y_1|\eta}(y_k|t_1), \cdots, F_{Y_1|\eta}(y_k|t_{K_1}))\big).$$

Finally it is easy to show that if $y_k < y_{k'}$ then $y_k^* < y_{k'}^*$ and thus $\psi(y_k) = y_k^*$ is a strictly

increasing function.

C.7 Proof of Theorem 5.1

We suppress X for simplicity. The proof is analogous to that of Theorem 2.1. Using the same notation as the earlier proof, (5.1) can be rewritten as

$$\sum_{\ell=1}^{\ell^*} \frac{q(z_\ell) - \gamma_\ell p(z_\ell)}{a} \times P[Y_1 \leq y | D = 1, Z = z_\ell] + \sum_{\ell=\ell^*+1}^{L} \frac{q(z_\ell)}{a} \times P[Y_1 \leq y | D = 1, Z = z_\ell]$$

$$= \sum_{\ell=\ell^*+1}^{L} \frac{\gamma_\ell p(z_\ell)}{a} \times P[Y_1 \leq y | D = 1, Z = z_\ell].$$

The above equation being satisfied as equality can be viewed as being satisfied as inequalities "$\leq$" and "$\geq$." Therefore, by Condition S_1 applied for both inequalities, we have

$$\sum_{\ell=1}^{k} \frac{q(z_\ell) - \gamma_\ell p(z_\ell)}{a} \times P[Y_0 \leq y | D = 1, Z = z_\ell] + \sum_{\ell=\ell^*+1}^{L} \frac{q(z_\ell)}{a} \times P[Y_0 \leq y | D = 1, Z = z_\ell]$$

$$= \sum_{\ell=\ell^*+1}^{L} \frac{\gamma_\ell p(z_\ell)}{a} \times P[Y_0 \leq y | D = 1, Z = z_\ell].$$

Equivalently, we have

$$P[Y_0 \leq y | D = 1] = P[Y_0 \leq y] \times \sum_{\ell=1}^{L} \gamma_\ell - \sum_{\ell=1}^{L} \gamma_\ell \times P[Y \leq y, D = 0 | Z = z_\ell]$$

$$= -\sum_{\ell=1}^{L} \gamma_\ell \times P[Y \leq y, D = 0 | Z = z_\ell]$$

by $\sum_{\ell=1}^{L} \gamma_\ell = 0$. □

C.8 Proof of Lemma B.1

Part (i) can be shown analogous to the proof of Theorem 4.1. Suppose

$$\int w(t,x)F_{Y_1|\eta,X}(\cdot|t,x)dt \leq \int \tilde{w}(t,x)F_{Y_1|\eta,X}(\cdot|t,x)dt$$

holds for some w and $\tilde{w}$. We want to show that

$$\int w(t,x)F_{Y_0|\eta,X}(\cdot|t,x)dt \leq \int \tilde{w}(t,x)F_{Y_0|\eta,X}(\cdot|t,x)dt.$$

First, because of the strict monotonicity of $q(d,x,\cdot)$, we have

$$\int w(t,x)F_{U_1|\eta,X}(\cdot|t,x)dt \leq \int \tilde{w}(t,x)F_{U_1|\eta,X}(\cdot|t,x)dt$$

and it suffices to show

$$\int w(t,x)F_{U_0|\eta,X}(\cdot|t,x)dt \leq \int \tilde{w}(t,x)F_{U_0|\eta,X}(\cdot|t,x)dt.$$

Second, for any $v \in \text{Supp}(V|X = x)$, because of the strict monotonicity of $\phi(\cdot,v)$, we have $1(U_1 \leq u_1) \overset{a.s.}{=} 1(\phi(U_1,v) \leq \phi(u_1,v))$. Because $V \perp (U_1,\eta)|X$, we have

$$\int w(t,x)F_{\phi(U_1,V)|\eta,X,V}(\phi(\cdot,v)|t,x,v)dt \leq \int \tilde{w}(t,x)F_{\phi(U_1,V)|\eta,X,V}(\phi(\cdot,v)|t,x,v)dt$$

and thus,

$$\int w(t,x)F_{U_0|\eta,X,V}(\phi(\cdot,v)|t,x,v)dt \leq \int \tilde{w}(t,x)F_{U_0|\eta,X,V}(\phi(\cdot,v)|t,x,v)dt.$$

Conditional on (η, X, V), $\mathrm{Supp}(\phi(U_1, v)) = \mathrm{Supp}(\phi(U_1, V)) = \mathrm{Supp}(U_0)$. Therefore, for u_0 in that support,

$$\int w(t, x) F_{U_0|\eta, X, V}(u_0|t, x, v) dt \le \int \tilde{w}(t, x) F_{U_0|\eta, X, V}(u_0|t, x, v) dt.$$

It follows that

$$\int \int w(t, x) F_{U_0|\eta, X, V}(u_0|t, x, v) f_{V|X}(v|x) dt dv$$
$$\le \int \int \tilde{w}(t, x) F_{U_0|\eta, X, V}(u_0|t, x, v) f_{V|X}(v|x) dt dv$$

Note that $f_{V|X} = f_{V|\eta, X}$. Then, by the law of iterated expectation, we have

$$\int w(t, x) F_{U_0|\eta, X}(u_0|t, x) dt \le \int \tilde{w}(t, x) F_{U_0|\eta, X}(u_0|t, x) dt.$$

Next, we prove part (ii) by first noting that

$$\Pr[U_0 \le u|\eta, X] = \Pr[\phi(U_1) \le u, V \le u|\eta, X] = \Pr[\phi(U_1) \le u|\eta, X] \Pr[V \le u|X].$$

Therefore,

$$\begin{aligned}
F_{Y_0|\eta, X}(y|t, x) &= \Pr[g(0, x, U_0) \le y|\eta = t, X = x] = \Pr[U_0 \le g^{-1}(0, x, y)|\eta = t, X = x] \\
&= \Pr[\phi(U_1) \le g^{-1}(0, x, y)|\eta = t, X = x] \Pr[V \le g^{-1}(0, x, y)] \\
&= \Pr[Y_1 \le g(1, x, \phi^{-1}(g^{-1}(0, x, y)))|\eta = t, X = x] \Pr[V \le g^{-1}(0, x, y)] \\
&= F_{Y_1|\eta, X}(\psi(y, x)|t, x) \lambda(y, x),
\end{aligned}$$

where $\psi(y, x) \equiv g(1, x, \phi^{-1}(g^{-1}(0, x, y)))$ and $\lambda(y, x) \equiv F_V(g^{-1}(0, x, y))$. $\square$

D Bounding Violation Probability in Linear Program with Randomized Constraints

Let $h(\gamma, y) \equiv p(y|1) - \boldsymbol{p}(y, 1)'\gamma$. Following Calafiore and Campi (2005), define a violation probability and a robustly feasible solution.

Definition D.1 (Violation probability)**.** *Let $\gamma \in \Gamma$ be a candidate solution for* (6.1)–(6.4)*. The probability of violation of γ is defined as*

$$V(\gamma) = \mathbb{P}\{Y \in \mathcal{Y} : h(\gamma, Y) > 0\},$$

where $\{Y \in \mathcal{Y} : h(\gamma, Y) > 0\}$ is assumed to be measurable.

Note that $V(\gamma^*) = 0$ where γ^* is the solution to (6.1)–(6.4).

Definition D.2 (ϵ-level solution)**.** *For $\epsilon \in [0, 1]$, $\gamma \in \Gamma$ is an ϵ-level robustly feasible solution if $V(\gamma) \leq \epsilon$.*

Then, we can show that the violation probability at the solution, denoted as $\bar{\gamma}_n$, to (6.3)–(6.4) is on average bounded by $1/n$.

Proposition D.1. *Let $\bar{\gamma}_n$ be the solution to* (6.3)–(6.4)*. Then,*

$$\mathbb{E}_{P^n}[V(\bar{\gamma}_n)] \leq \frac{1}{n+1},$$

where P^n is the probability measure in the space $\mathcal{Y}^n$ of the multi-sample extraction $Y_1, ..., Y_n$.

Corollary D.1. *Fix $\epsilon \in [0, 1]$ and $\beta \in [0, 1]$ and let*

$$n \geq \frac{1}{\epsilon\beta} - 1.$$

Then, with probability no smaller than $1 - \beta$, the sampled LP (6.3)–(6.4) *returns an optimal solution $\hat{\gamma}_n$ which is ϵ-level robustly feasible.*

The above results implicitly assume a particular rule of tie-breaking when there are multiple solutions in the sampled LP (see Theorem 3 in Calafiore and Campi (2005)). There is also discussions on no solution in the paper.

References

ABADIE, A., J. ANGRIST, AND G. IMBENS (2002): "Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings," *Econometrica*, 70, 91–117. 1, 3.1

BLUNDELL, R., A. GOSLING, H. ICHIMURA, AND C. MEGHIR (2007): "Changes in the distribution of male and female wages accounting for employment composition using bounds," *Econometrica*, 75, 323–363. 2

CALAFIORE, G. AND M. C. CAMPI (2005): "Uncertain convex programs: randomized solutions and confidence levels," *Mathematical Programming*, 102, 25–46. 6.2, D, D

CHERNOZHUKOV, V. AND C. HANSEN (2005): "An IV model of quantile treatment effects," *Econometrica*, 73, 245–261. 1, 2, 4, 5, B

——— (2013): "Quantile models with endogeneity," *Annu. Rev. Econ.*, 5, 57–81. 4

DONG, Y. AND S. SHEN (2018): "Testing for rank invariance or similarity in program evaluation," *Review of Economics and Statistics*, 100, 78–85. 1, 4.1

FRANDSEN, B. R. AND L. J. LEFGREN (2018): "Testing rank similarity," *Review of Economics and Statistics*, 100, 86–91. 1, 4.1

HAN, S. (2019): "Identification in Nonparametric Models for Dynamic Treatment Effects," *UT Austin*. 1

HAN, S. AND S. YANG (2020): "Sharp Bounds on Treatment Effects for Policy Evaluation," *arXiv preprint arXiv:2009.13861*. 1, 6.2, 6.3

HECKMAN, J. J., J. SMITH, AND N. CLEMENTS (1997): "Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts," *The Review of Economic Studies*, 64, 487–535. [1]

IMBENS, G. W. AND J. D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475. [1], [2]

KIM, J. H. AND B. G. PARK (2022): "Testing rank similarity in the local average treatment effects model," *Econometric Reviews*, 1–22. [1], [4.1]

MAASOUMI, E. AND L. WANG (2019): "The gender gap between earnings distributions," *Journal of Political Economy*, 127, 2438–2504. [1]

MANSKI, C. F. (1990): "Nonparametric bounds on treatment effects," *The American Economic Review*, 80, 319–323. [1]

———— (1994): "The selection problem," in *Advances in Econometrics, Sixth World Congress,* ed. by C. Sims, vol. 1, 143–70. [2]

———— (1997): "Monotone treatment response," *Econometrica: Journal of the Econometric Society*, 1311–1334. [1]

MANSKI, C. F. AND J. V. PEPPER (2000): "Monotone instrumental variables: With an application to the returns to schooling," *Econometrica*, 68, 997–1010. [1]

MOGSTAD, M., A. SANTOS, AND A. TORGOVITSKY (2018): "Using instrumental variables for inference about policy relevant treatment parameters," *Econometrica*, 86, 1589–1619. [1]

MOGSTAD, M., A. TORGOVITSKY, AND C. R. WALTERS (2021): "The causal interpretation of two-stage least squares with multiple instrumental variables," *American Economic Review*, 111, 3663–98. [2], [3]

POMATTO, L., P. STRACK, AND O. TAMUZ (2020): "Stochastic dominance under independent noise," *Journal of Political Economy*, 128, 1877–1900. B

RUBIN, D. B. (1974): "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of Educational Psychology*, 66, 688. 1

SHAIKH, A. M. AND E. J. VYTLACIL (2011): "Partial identification in triangular systems of equations with binary dependent variables," *Econometrica*, 79, 949–955. 1

VUONG, Q. AND H. XU (2017): "Counterfactual mapping and individual treatment effects in nonseparable models with binary endogeneity," *Quantitative Economics*, 8, 589–610. 1, 2

VYTLACIL, E. AND N. YILDIZ (2007): "Dummy endogenous variables in weakly separable models," *Econometrica*, 75, 757–779. 1