Manuscript submitted to *The Econometrics Journal*, pp. 1–24.

Choosing Exogeneity Assumptions in Potential Outcome Models

Matthew A. Masten[†] and Alexandre Poirier[‡]

[†]*Department of Economics, Duke University, Durham, NC*
E-mail: matt.masten@duke.edu
[‡]*Department of Economics, Georgetown University, Washington, DC*
E-mail: alexandre.poirier@georgetown.edu

Summary There are many kinds of exogeneity assumptions. How should researchers choose among them? When exogeneity is imposed on an unobservable like a potential outcome, we argue that the form of exogeneity should be chosen based on the kind of selection on unobservables it allows. Consequently, researchers can assess the plausibility of any exogeneity assumption by studying the distributions of treatment given the unobservables that are consistent with that assumption. We use this approach to study two common exogeneity assumptions: quantile and mean independence. We show that both assumptions require a kind of non-monotonic relationship between treatment and the potential outcomes. We discuss how to assess the plausibility of this kind of treatment selection. We also show how to define a new and weaker version of quantile independence that allows for monotonic selection on unobservables. We then show the implications of the choice of exogeneity assumption for identification. We apply these results in an empirical illustration of the effect of child soldiering on wages.

Keywords: *Selection on Unobservables, Nonparametric Identification, Treatment Effects, Partial Identification, Sensitivity Analysis*

1. INTRODUCTION

Exogeneity is a critical assumption in much structural or causal empirical work. In general, exogeneity refers to assumptions on the statistical dependence between an observable term and an unobservable term.[1] There are many such assumptions, however, including zero correlation, median independence, and full statistical independence. The choice of a formal definition of exogeneity is not innocuous. Different definitions have different substantive interpretations and different implications for identification, rates of convergence, asymptotic distributions, efficiency bounds, and overidentification.

In the context of potential outcome models, there has been debate over the appropriate choice of exogeneity assumption. Heckman, Ichimura, and Todd (1998) assume potential outcomes are mean independent of treatment, conditional on covariates. They justify this focus on mean independence by arguing that "conditional independence assumptions...are far stronger than the mean-independence conditions typically invoked by economists" (page 262). Imbens (2004, page 8) agrees that "this [mean independence] assumption is unquestionably weaker [than full independence]", but argues that "in practice it is rare that a convincing case is made for the weaker [mean independence]

[1]Throughout this paper we use 'exogeneity' in the same sense as the treatment effects literature; for example, see Imbens (2004). This is related to but distinct from the Cowles Commission definition of an exogenous variable as a variable that is determined outside of the model under consideration. See the discussion in Hendry and Morgan (1995, pages 74–75), Imbens (1997, page 93), and Heckman (2000, footnote 11).

assumption 2.3 without the case being equally strong for the stronger [full independence assumption]." The justification he provides is that "the weaker assumption is intrinsically tied to functional-form assumptions, and as a result one cannot identify average effects on transformations of the original outcome (such as logarithms) without the stronger assumption."

In this paper, we contribute to this debate as follows: We recommend that researchers focus directly on the substantive economic interpretation of the assumption, and the plausibility of its restrictions, rather than assess assumptions based on what they can be used to identify (as in the functional form dependency critique of mean independence) or on mathematical orderings of what implies what (as in the observation that statistical independence implies mean independence, but not vice versa). Specifically, we focus on two of the most common forms of exogeneity assumptions used, quantile independence and mean independence. We provide several results to help researchers assess the economic plausibility of these exogeneity assumptions. First, in section 2, we provide a brief informal discussion of the motivations for making exogeneity assumptions. There we note that exogeneity assumptions are typically made on structural unobservables, variables that satisfy some kind of policy or treatment invariance property, like potential outcomes or unobserved ability. In this case, the form of exogeneity depends on the form of treatment selection. Consequently, the plausibility of any exogeneity assumption can be assessed by examining the distributions of treatment given the unobservables that are consistent with that assumption.

Next, in section 3, we characterize these distributions of treatment given the unobservables that are consistent with either quantile independence or mean independence. This characterization shows that both quantile independence and mean independence require specific kinds of non-monotonic treatment selection. Specifically, we define the *latent propensity score* as the probability of treatment conditional on a potential outcome. We say there is *monotonic selection on unobservables* if this function is monotonic. This happens when, for example, units with larger values of potential outcomes are more likely to get treated. Our characterization result shows that this kind of treatment selection is incompatible with quantile and mean independence.

In section 4.1 we show how to modify the quantile independence assumption to create a new, weaker exogeneity assumption which allows for more plausible forms of treatment selection, including monotonic selection on unobservables. We call this assumption $\mathcal{U}$-*independence*. In section 4.2 we derive identified sets for the average effect of treatment for the treated (ATT) and the quantile treatment effect for the treated (QTT) parameters under both quantile independence and $\mathcal{U}$-independence. These identified sets have a simple, closed form characterization, which makes them easy to use in practice. By comparing these identified sets we show that the identifying power of quantile independence comes from the fact that it only allows for a restrictive kind of non-monotonic selection on unobservables. In section 5 we examine these differences in an empirical illustration of the effects of child soldiering on wages based on unconfoundedness. We show that the baseline results are generally robust under quantile independence relaxations of unconfoundedness, but not under $\mathcal{U}$-independence relaxations. This difference highlights both the practical importance of choosing exogeneity assumptions and how our approach can help researchers make this choice. Finally, in section 6 we use a Roy model to further illustrate how researchers can assess the plausibility of non-monotonic selection on unobservables.

2. CHOOSING EXOGENEITY ASSUMPTIONS

There are many different kinds of exogeneity assumptions available to researchers. Manski (1988) and Powell (1994) catalog some of the most common forms, including zero correlation, mean independence, quantile independence, conditional symmetry, statistical independence, and a variety of index conditions. Other kinds of exogeneity assumptions have since been defined, including mean monotonicity (e.g., Manski and Pepper 2000, 2009, chapter 2 of Manski 2003), approximate mean independence (Manski 2003, section 9.4), stochastic dominance assumptions (e.g., Blundell, Gosling, Ichimura, and Meghir 2007), quantile uncorrelation (Komarova, Severini, and Tamer 2012), and partial statistical independence (e.g., Rosenbaum and Rubin 1983, Rosenbaum 1995, 2002, Imbens 2003, Masten and Poirier 2018), among many others. How should researchers choose among these many options? In this section we discuss one approach to answering this question.

The answer depends on whether the exogeneity assumption is made on a structural unobservable or a reduced form unobservable. This distinction goes back to the earliest work on simultaneous equation models in econometrics (see Hausman 1983 for a survey) but has been used in recent work as well (e.g. Blundell and Matzkin 2014). Structural unobservables are variables that satisfy some kind of policy or treatment invariance property, like potential outcomes, unobserved ability, or preferences. Reduced form unobservables are functions of the structural unobservables, and possibly other variables in the model, like realized treatment. Since quantile independence is often imposed on the relationship between treatment variables and structural unobservables, we focus on that case. We briefly discuss exogeneity assumptions for reduced form unobservables in appendix A, along with some additional background and examples.

Let X denote the observed, realized treatment, and let Y_x be a potential outcome where x is a logically possible value of treatment. Since, by definition, potential outcomes have a meaning and interpretation that does not depend on the realized treatment X, any stochastic dependence between them must reflect some form of treatment selection on Y_x. That is, it must reflect some form of selection on unobservables, which is described by the distribution of $X \mid Y_x$. Many exogeneity assumptions, such as quantile independence and mean independence, are defined as constraints on the distribution of $Y_x \mid X$. Consequently, to assess the plausibility of these assumptions, we recommend examining the set of distributions of $X \mid Y_x$ that are consistent with the given constraint on the distribution of $Y_x \mid X$. This allows researchers to use the large literature on treatment selection to assess the plausibility of various exogeneity assumptions (see Heckman and Vytlacil 2007a,b and Abbring and Heckman 2007 for a comprehensive survey). Note that this discussion and recommendation extend to more general structural or causal models where one is considering exogeneity assumptions between a structural unobservable U and an observed variable X; $U = Y_x$ is the specific case we focus on here.

Descriptive Analysis and Causal Models Before proceeding, it is important to emphasize the scope of the process we just described: We are interested in assumptions about the dependence structure between observable and unobservable variables in causal models. This does *not* include research whose end goal is a description of the joint distribution of observed random variables, and which does not aim to make causal statements. Such descriptive research studies the relationship between observed variables. For example, suppose Y is an observed outcome and X is an observed covariate. We might define E

to be the residual from a linear projection of Y onto $(1, X)$. We can then ask about the statistical relationship between E and X: It satisfies zero correlation by construction, but not necessarily other restrictions like mean independence or statistical independence. In this case, however, the joint distribution of (E, X) is always point identified. Hence, at the population level, the precise relationship between these variables is always known. Consequently, there is no need to make or choose assumptions about the stochastic relationship between E and X. In contrast, in causal models, exogeneity assumptions typically have substantial identifying power for causal effects or structural parameters.

3. CHARACTERIZING EXOGENEITY ASSUMPTIONS

In this section, we present our main characterization results. We provide results for two of the most common exogeneity assumptions: quantile independence and mean independence. We focus on binary treatments throughout the paper; we generalize our results to multi-valued discrete and continuous treatments in appendix S1. All of our results also hold if one conditions on an additional vector of observed covariates, as is typically the case in empirical applications, but we omit these for simplicity.

3.1. The Potential Outcomes Model

Let $X \in \{0, 1\}$ be a binary treatment variable. Let (Y_1, Y_0) denote unobserved potential outcomes. We observe the scalar outcome variable

$$Y = XY_1 + (1 - X)Y_0. \tag{3.1}$$

Let $p_x = \mathbb{P}(X = x)$ for $x \in \{0, 1\}$. We impose the following assumption on the joint distribution of (Y_1, Y_0, X).

ASSUMPTION 3.1. *For each $x, x' \in \{0, 1\}$:*

1 $Y_x \mid X = x'$ *has a strictly increasing and continuous distribution function on its support,* $\text{supp}(Y_x \mid X = x')$.
2 $\text{supp}(Y_x \mid X = x') = \text{supp}(Y_x) = [\underline{y}_x, \overline{y}_x]$ *where* $-\infty \leq \underline{y}_x < \overline{y}_x \leq \infty$.
3 $p_x > 0$.

Via assumption 3.1.1, we restrict attention to continuously distributed potential outcomes. Assumption 3.1.2 states that the unconditional and conditional supports of Y_x are equal, and are a possibly infinite closed interval. We maintain assumption 3.1.2 for simplicity, but it can be relaxed using similar derivations as in Masten and Poirier (2016). Assumption 3.1.3 is an overlap assumption.

In potential outcome models, statistical independence between potential outcomes and treatment is sometimes assumed:

$$Y_x \perp\!\!\!\perp X. \tag{3.2}$$

This assumption, equivalent to random assignment of treatment, can be made for $x = 0$, $x = 1$, or both. When this assumption is made conditional on covariates, it is often called *unconfoundedness*. As discussed in section 2, however, there are many other kinds of exogeneity assumptions available in the literature, which are all weaker than full statistical independence. The purpose of this paper is to help researchers choose between these different kinds of exogeneity assumptions. To that end, in the next two subsections

we provide characterization results that can help researchers assess the plausibility of these exogeneity assumptions. We then study the identifying power of these different assumptions in section 4.

3.2. A Class of Quantile Independence Assumptions

Quantile independence of the potential outcome Y_x from X at quantile τ holds if

$$Q_{Y_x|X}(\tau \mid 0) = Q_{Y_x|X}(\tau \mid 1). \tag{3.3}$$

This assumption is often imposed at a single quantile. For example, imposing (3.3) at $\tau = 0.5$ yields median independence. If this holds for all $\tau \in (0, 1)$, then Y_x and X are statistically independent. Therefore quantile independence is a relaxation of independence.

It is often more natural to work with cdfs, an inverse of the quantile function.[2] Say Y_x is τ-cdf independent of X if

$$F_{Y_x|X}(\tau \mid 0) = F_{Y_x|X}(\tau \mid 1). \tag{3.4}$$

Note that Y_x is quantile independent of X at quantile τ if and only if Y_x is $Q_{Y_x}(\tau)$-cdf independent of X for continuously distributed Y_x. This motivates the following definition.[3]

Definition 3.1. *Let $\mathcal{T}$ be a subset of $\mathbb{R}$. Say Y_x is* $\mathcal{T}$-independent *of X if the cdf independence condition* (3.4) *holds for all $\tau \in \mathcal{T}$.*

With binary treatments, the dependence structure between X and the potential outcome Y_x is fully characterized by the function

$$p(y_x) = \mathbb{P}(X = 1 \mid Y_x = y_x),$$

which we call the latent propensity score. Full statistical independence of Y_x and X, or $\mathcal{T}$-independence with $\mathcal{T} = [\underline{y}_x, \overline{y}_x]$, is equivalent to this latent propensity score being constant:

$$p(y_x) = \mathbb{P}(X = 1)$$

for almost all $y_x \in \text{supp}(Y_x)$. Analogously, $\mathcal{T}$-independence for $\mathcal{T} \subsetneq [\underline{y}_x, \overline{y}_x]$ is weaker than full independence, and thus partially restricts the shape of $p(y_x)$. The following theorem characterizes the set of latent propensity scores consistent with $\mathcal{T}$-independence.

Theorem 3.1. (Average value characterization) *Suppose assumption 3.1.1 holds. Then Y_x is $\mathcal{T}$-independent of X if and only if*

$$\mathbb{E}\big(p(Y_x) \mid Y_x \in (t_1, t_2)\big) = \mathbb{P}(X = 1) \tag{3.5}$$

for all $t_1, t_2 \in \mathcal{T} \cup \{\underline{y}_x, \overline{y}_x\}$ with $t_1 < t_2$.

The proof, along with all others, is in appendix S3. Theorem 3.1 says that $\mathcal{T}$-independence

[2]For example, see assumption QI on page 731 of Manski (1988) or equation (1.7) on page 2452 of Powell (1994). Definitions using cdfs and those using quantiles directly (e.g., via equation (3.3)) are often equivalent. Throughout this paper we use "quantile independence" to mean the cdf-based definition, as is common in the literature.

[3]See Belloni et al. (2017) and Zhu et al. (2018) for similar generalizations of quantile independence.

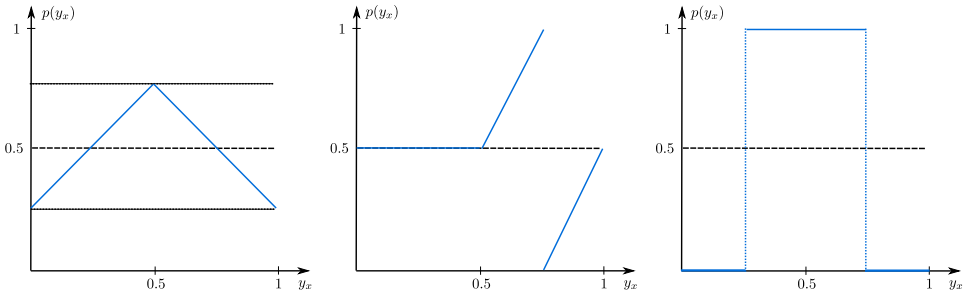

Figure 1. Various latent propensity scores consistent with $\mathcal{T} = \{0.5\}$-independence, when $\mathbb{P}(X = 1) = 0.5$.

holds if and only if for every interval with endpoints in $\mathcal{T} \cup \{\underline{y}_x, \overline{y}_x\}$ the average latent propensity score over $Y_x \in (t_1, t_2)$ equals the overall average of the latent propensity score, which is $\mathbb{P}(X = 1)$. Also note that $\mathbb{E}(p(Y_x) \mid Y_x \in (t_1, t_2)) = \mathbb{P}(X = 1 \mid Y_x \in (t_1, t_2))$.

To illustrate theorem 3.1, suppose $\mathcal{T} = \{0.5\}$ and $\mathbb{P}(X = 1) = 0.5$. Further suppose that Y_x is uniformly distributed on $[0, 1]$ to simplify the figures. Here we have a single nontrivial cdf independence condition: median independence. Figure 1 plots three different latent propensity scores which are consistent with $\mathcal{T}$-independence under this choice of $\mathcal{T}$; that is, which are consistent with median independence. This figure illustrates several features of such latent propensity scores: The value of $p(y_x)$ may vary over the entire range $[0, 1]$. p does not need to be symmetric about $y_x = 0.5$, nor does it need to be continuous. It does need to satisfy equation (3.5) over the intervals $(t_1, t_2) = (0, 0.5)$ and $(t_1, t_2) = (0.5, 1)$. Finally, as suggested by the pictures, p must actually be nonmonotonic; we show this in corollary 3.1 next.

COROLLARY 3.1. *Suppose assumption 3.1.1 holds. Suppose the latent propensity score p is weakly monotonic and not constant on $(\underline{y}_x, \overline{y}_x)$. Then, for all $\tau \in (\underline{y}_x, \overline{y}_x)$, Y_x is not τ-cdf independent of X.*

Corollary 3.1 shows that any quantile independence assumption rules out all types of monotonic selection on unobservables, except for the trivially monotonic constant $p(y_x) = \mathbb{P}(X = 1)$ implied by full statistical independence.

Next we show that imposing multiple quantile independence conditions imposes further non-monotonicity. Say that a function f *changes direction at least K times* if there exists a partition of its domain into K intervals such that f is not monotonic on each interval.

COROLLARY 3.2. *Suppose assumption 3.1.1 holds. Suppose Y_x is $\mathcal{T}$-independent of X. Suppose there exists a version of p without removable discontinuities. Partition $(\underline{y}_x, \overline{y}_x)$ by the sets $\mathcal{Y}_1 = (t_0, t_1)$, $\mathcal{Y}_k = [t_{k-1}, t_k)$ for $k = 2, \ldots, K$ with $t_0 = \underline{y}_x$, $t_K = \overline{y}_x$, and such that for each k there is a $\tau_k \in \mathcal{T} \cap \mathcal{Y}_k$. Suppose p is not constant over each set $\mathcal{Y}_k$, $k = 1, \ldots, K$. Then p changes direction at least K times.*

This result says that such latent propensity scores must oscillate up and down at least K times (we assume p does not have removable discontinuities to rule out trivial direction changes). For example, as in figure 1, suppose we continue to have $\mathbb{P}(X = 1) = 0.5$ and $Y_x \sim \text{Unif}[0, 1]$ but we add a few more isolated τ's to $\mathcal{T}$. Figure 2 shows several

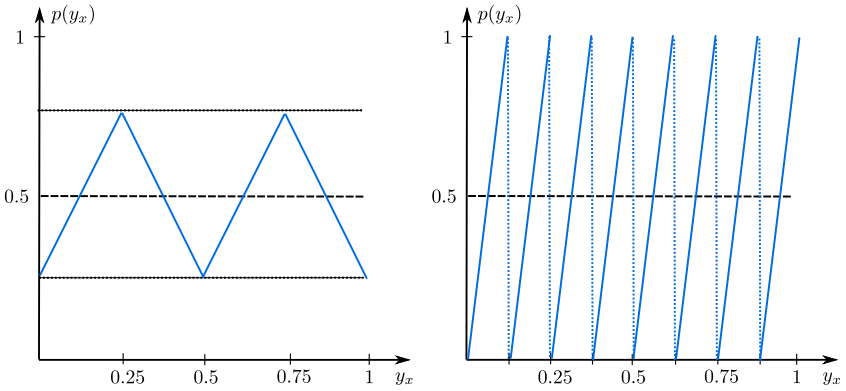

Figure 2. Some latent propensity scores consistent with $\mathcal{T}$-independence when $\mathbb{P}(X = 1) = 0.5$. Left: $\mathcal{T} = \{0.25, 0.5, 0.75\}$. Right: $\mathcal{T} = \{0.125,\ 0.25,\ 0.375,\ 0.5,\ 0.625,\ 0.75,\ 0.875\}$.

latent propensity scores consistent with $\mathcal{T}$-independence when $\mathcal{T}$ has several isolated elements. Consider the figure on the left, with $\mathcal{T} = \{0.25, 0.5, 0.75\}$. Partition $(0, 1) = (0, 0.4) \cup [0.4, 0.6) \cup [0.6, 1)$. Then p is not monotonic over each partition set, and each partition set contains one element of $\mathcal{T}$: $0.25 \in (0, 0.4)$, $0.5 \in [0.4, 0.6)$, and $0.75 \in [0.6, 1)$. There are $K = 3$ partition sets, and hence the corollary says p must change direction at least 3 times. We see this in the figure since there are 3 interior local extrema. A similar analysis holds for the figure on the right. Overall, these triangular and sawtooth latent propensity scores illustrate the oscillation required by corollary 3.2.

We document one more feature: As long as there is some interval that is not in $\mathcal{T}$ then there is a latent propensity score that takes the most extreme values possible, 0 and 1.

COROLLARY 3.3. *Suppose assumptions 3.1.1 and 3.1.3 hold. Suppose* $[\underline{y}_x, \overline{y}_x] \backslash \mathcal{T}$ *contains a non-degenerate interval. Then there exists a latent propensity score which is consistent with* $\mathcal{T}$*-independence of* Y_x *from* X *and for which the sets*

$$\{y_x \in [\underline{y}_x, \overline{y}_x] : p(y_x) = 0\} \qquad \textit{and} \qquad \{y_x \in [\underline{y}_x, \overline{y}_x] : p(y_x) = 1\}$$

have positive Lebesgue measure.

Consequently, $\mathcal{T}$-independence allows for a kind of extreme imbalance, where there is a positive mass of potential outcome values that only appear in the treatment group and another positive mass of potential outcome values that only appear in the control group.

3.3. Characterizing Mean Independence

Mean independence is another commonly used exogeneity assumption. For example, Heckman et al. (1998) assume potential outcomes are mean independent of treatments, conditional on covariates. As in the previous section, we characterize the constraints this assumption places on the conditional distribution of X given Y_x.

DEFINITION 3.2. *Say* Y_x *is* mean independent *of* X *if* $\mathbb{E}(Y_x \mid X = 0) = \mathbb{E}(Y_x \mid X = 1)$.

From definition 3.2 and Bayes' rule, it immediately follows that

$$\mathbb{E}\left(\frac{Y_x}{\mathbb{E}(Y_x)}p(Y_x)\right) = \mathbb{P}(X=1), \tag{3.6}$$

assuming $\mathbb{E}(Y_x) \neq 0$. Theorem 3.1 showed that quantile independence constrains the *unweighted* average value of the latent propensity score over certain *subintervals* of its domain. In contrast, equation (3.6) shows that mean independence constrains a *weighted* average value of the latent propensity score over its entire domain. Equation (3.6) can be extended to multi-valued and continuous X as in our analysis of quantile independence in appendix S1; we omit this extension for brevity.

Although mean independence imposes a different constraint on the latent propensity score than quantile independence, it also requires non-constant latent propensity scores to be non-monotonic.

PROPOSITION 3.1. *Suppose assumption 3.1.1 holds. Suppose* $\mathbb{E}(|Y_x|) < \infty$. *Suppose the latent propensity score* p *is weakly monotonic and not constant on the interior of its domain. Then* Y_x *is not mean independent of* X.

Hence mean independence rules out all types of monotonic selection, except for the trivially monotonic constant $p(y_x) = \mathbb{P}(X=1)$ implied by full statistical independence.

3.4. Discussion

To place these results in context, consider the case where X is an indicator for completing college and Y_0 denotes a person's earnings if they do not complete college. X is often thought to be endogenous due to its relationship with ability, which is captured by Y_0. In this example, $p(y_0)$ is the proportion of people who complete college, among those with a fixed level of non-graduate earnings. Corollary 3.1 states that any quantile independence condition rules out nonconstant, weakly monotonic selection on unobservables. Similarly, proposition 3.1 implies that mean independence rules out this monotonic selection on unobservables. They would thus rule out that the proportion who attend college is weakly increasing in the level of non-graduate earnings, unless we assume that college attendance and non-graduate earnings are statistically independent, in which case $p(y_0)$ is constant.

Corollary 3.2 would require the probability of attending college to oscillate (or be constant) to accommodate multiple quantile independence conditions. For example, if K quantile independence conditions hold, there must exist K non-graduate earnings thresholds where the effect of non-graduate earnings on college attendance changes sign. Alternatively, college attendance can again be statistically independent of non-graduate earnings. We discuss the plausibility of these oscillations in the context of a Roy model in section 6.

Corollary 3.3 characterizes another feature of the latent propensity scores allowed by quantile independence. In our returns to schooling example, a finite number of quantile independence restrictions allow for a strictly positive proportion of people with non-graduate earnings levels for which nobody attends college ($p(y_0) = 0$), and for which everyone attends college ($p(y_0) = 1$). Depending on the context, the existence of these two groups may or may not appear plausible. If it appears implausible, using an assumption that allows for their existence is unnecessarily conservative compared to an assumption that rules out their existence. Since ruling out their existence is a stronger assumption,

it would result in weakly narrower identified sets compared to the overly conservative set obtained under quantile independence alone. To rule out their existence, one could add a constraint such as $p(y_0) \in [c_1, c_2]$ for prespecified $0 < c_1 \leq c_2 < 1$ and derive the identified set for the relevant parameter under quantile independence and this added constraint. Throughout this paper, we emphasize this approach of tailoring exogeneity conditions to the likely and unlikely features of the treatment selection function $p(y_0)$ in the given application. Precisely finding the assumption that best captures the nonexistence of those groups is beyond the scope of this paper, however, since it is application dependent.

4. THE IDENTIFYING POWER OF DIFFERENT EXOGENEITY ASSUMPTIONS

In this section we study the implications of the choice of exogeneity assumption for identification. Our first result in section 3 shows that quantile independence imposes a constraint on the average value of a latent propensity score. We use this characterization to motivate an assumption weaker than quantile independence, which we call $\mathcal{U}$*-independence*. The difference between these two assumptions is that quantile independence imposes some additional average value constraints on $p(y_x)$ that $\mathcal{U}$-independence does not. In particular, $\mathcal{U}$-independence allows for monotonic selection on unobservables. Hence the difference between identified sets obtained under these two assumptions is a measure of the identifying power of these additional average value constraints, which are the features of quantile independence that require the latent propensity score to be non-monotonic. Unlike quantile independence, it is not clear to us how to naturally weaken mean independence to allow for monotonic latent propensity scores while still retaining an interpretable assumption with identifying power. For that reason, in this section we only study identification under quantile independence and its weaker version, $\mathcal{U}$-independence.

We focus on two parameters: The average treatment effect for the treated,

$$\begin{aligned} \text{ATT} &= \mathbb{E}(Y_1 - Y_0 \mid X = 1) \\ &= \mathbb{E}(Y \mid X = 1) - \mathbb{E}(Y_0 \mid X = 1), \end{aligned}$$

and the quantile treatment effect for the treated,

$$\begin{aligned} \text{QTT}(q) &= Q_{Y_1|X}(q \mid 1) - Q_{Y_0|X}(q \mid 1) \\ &= Q_{Y|X}(q \mid 1) - Q_{Y_0|X}(q \mid 1), \end{aligned}$$

for $q \in (0, 1)$. To analyze treatment on the treated parameters, we only need to make assumptions on the relationship between Y_0 and X. Our analysis can easily be extended to parameters like ATE by imposing $\mathcal{T}$- or $\mathcal{U}$-independence between Y_1 and X as well as between Y_0 and X.

Under statistical independence $Y_0 \perp\!\!\!\perp X$, both the ATT and QTT are point identified. Under $\mathcal{T}$- or $\mathcal{U}$-independence, however, they are generally partially identified. We derive identified sets for ATT and QTT under both classes of exogeneity assumptions in this section. These sets have simple explicit expressions which make them easy to use in practice. In section 5 we compare these identified sets in an empirical application. In this application, the estimated identified sets are substantially larger under $\mathcal{U}$-independence, implying that the additional average value constraints inherent in $\mathcal{T}$-independence have substantial identifying power.

4.1. Weakening Quantile Independence

Throughout this section, we focus on the case where $\mathcal{T}$ is an interval. In this case, we show that latent propensity scores consistent with $\mathcal{T}$-independence have two features: (a) they are flat on $\mathcal{T}$ and (b) they are non-monotonic outside the flat regions, such that the average value constraint (3.5) is satisfied. We use this finding to motivate a weaker assumption which retains feature (a) but drops feature (b). We call this assumption $\mathcal{U}$-*independence*. This new weaker assumption has two uses: First, we can use it as a tool for understanding quantile independence. Specifically, by comparing identified sets under quantile independence and under the weaker $\mathcal{U}$-independence we will learn the identifying power of the average value constraints on the latent propensity score. Second, $\mathcal{U}$-independence can be used by itself as a method for relaxing statistical independence and performing sensitivity analysis. We illustrate both of these uses below and in our empirical analysis of section 5.

We begin with the following corollary to theorem 3.1.

COROLLARY 4.1. *Suppose assumption 3.1 holds for Y_0. Suppose Y_0 is $\mathcal{T}$-independent from X with $\mathcal{T} = [a, b] \subseteq [\underline{y}_0, \overline{y}_0]$. Then*

$$\mathbb{P}(X = 1 \mid Y_0 = y_0) = \mathbb{P}(X = 1) \tag{4.7}$$

for almost all $y_0 \in \mathcal{T}$.

Corollary 4.1 shows that $\mathcal{T}$-independence requires the latent propensity score to be constant on $\mathcal{T}$ and equal to the overall unconditional probability of being treated. The first property—that the latent propensity score is flat on $\mathcal{T}$—means that random assignment holds within the subpopulation of units whose untreated outcomes are in the set $\mathcal{T}$; that is, $X \perp\!\!\!\perp Y_0 \mid \{Y_0 \in \mathcal{T}\}$. Corollary 4.1 can be generalized to allow $\mathcal{T}$ to be a finite union of intervals, but we omit this for brevity.

This corollary motivates the following definition.

DEFINITION 4.1. *Let $\mathcal{U} \subseteq [\underline{y}_x, \overline{y}_x]$ be an interval. Say that Y_x is* $\mathcal{U}$-independent *of X if $\mathbb{P}(X = 1 \mid Y_x = y_x) = \mathbb{P}(X = 1)$ for almost all $y_x \in \mathcal{U}$.*

Importantly, unlike $\mathcal{T}$-independence, $\mathcal{U}$-independence allows for monotonic selection on unobservables. Corollary 4.1 shows that $\mathcal{T}$-independence implies $\mathcal{U}$-independence with $\mathcal{U} = \mathcal{T}$. The converse does not hold since $\mathcal{T}$-independence requires additional average value constraints to hold, by theorem 3.1. In particular, $\mathcal{U}$-independence implies the average value constraint

$$\mathbb{E}(p(Y_x) \mid Y_x \in (t_1, t_2)) = \mathbb{P}(X = 1) \tag{3.5}$$

for all $t_1, t_2 \in \mathcal{U}$, because it requires that $p(u)$ is constant on $\mathcal{U}$. But $\mathcal{T}$-independence also requires that (3.5) holds for choices of t_1 and t_2 in $\mathcal{U} \cup \{\underline{y}_x, \overline{y}_x\}$. That is, t_1 and t_2 can equal the end points $\underline{y}_x$ or $\overline{y}_x$. Hence it imposes an average value constraint outside of the set $\mathcal{U}$. For example, figure 3 shows two latent propensity scores. One satisfies $\mathcal{T}$-independence, but the other only satisfies $\mathcal{U}$-independence. Finally, note that $\mathcal{U}$-independence is a nontrivial assumption only when $\mathbb{P}(Y_x \in \mathcal{U}) > 0$. Conversely, $\mathcal{T}$-independence is nontrivial even when $\mathcal{T}$ is a singleton.

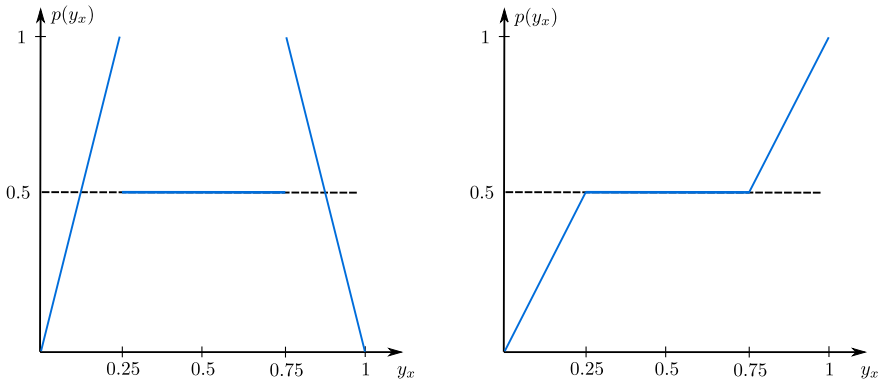

Figure 3. Let $\mathcal{T} = \mathcal{U} = [0.25, 0.75]$ and $\mathbb{P}(X = 1) = 0.5$. Also normalize Y_0 to be uniform on $(0, 1)$ for simplicity. This figure shows a latent propensity score consistent with both $\mathcal{T}$- and $\mathcal{U}$-independence on the left and a latent propensity score consistent with $\mathcal{U}$-, but not $\mathcal{T}$-independence on the right.

4.2. The Identified Sets For ATT and QTT(q)

In this subsection we derive sharp bounds on the ATT and QTT(q) under both $\mathcal{T}$- and $\mathcal{U}$-independence. To do so, it suffices to derive bounds on $Q_{Y_0|X}(q \mid 1)$.

We show the validity of the following bounds in proposition 4.1 below. Let $\mathcal{T} = \mathcal{U} = [Q_{Y_0}(a), Q_{Y_0}(b)]$ for $0 < a \leq b < 1$. The $\mathcal{T}$-independence bounds are then defined by

$$\overline{Q}^{\mathcal{T}}_{Y_0|X}(\tau \mid 1) = \begin{cases} Q_{Y|X}(a \mid 0) & \text{for } \tau \in (0, a] \\ Q_{Y|X}(\tau \mid 0) & \text{for } \tau \in (a, b] \\ Q_{Y|X}(1 \mid 0) & \text{for } \tau \in (b, 1) \end{cases}$$

and

$$\underline{Q}^{\mathcal{T}}_{Y_0|X}(\tau \mid 1) = \begin{cases} Q_{Y|X}(0 \mid 0) & \text{for } \tau \in (0, a] \\ Q_{Y|X}(\tau \mid 0) & \text{for } \tau \in (a, b] \\ Q_{Y|X}(b \mid 0) & \text{for } \tau \in (b, 1). \end{cases}$$

We let $Q_{Y|X}(0 \mid x) = \underline{y}_x$ and $Q_{Y|X}(1 \mid x) = \overline{y}_x$. For $\mathcal{U}$-independence, there are two cases. First consider the lower bound. Recall that $p_x = \mathbb{P}(X = x)$. If $(1 - (b - a))p_1 \leq a$,

$$\underline{Q}^{\mathcal{U}}_{Y_0|X}(\tau \mid 1) = \begin{cases} Q_{Y|X}(0 \mid 0) & \text{for } \tau \in (0, 1 - (b - a)] \\ Q_{Y|X}\left(\tau + \dfrac{b-1}{p_0} \mid 0\right) & \text{for } \tau \in (1 - (b - a), 1). \end{cases}$$

If $(1 - (b - a))p_1 \geq a$,

$$\underline{Q}^{\mathcal{U}}_{Y_0|X}(\tau \mid 1) = \begin{cases} Q_{Y|X}(0 \mid 0) & \text{for } \tau \in \left(0, \dfrac{a}{p_1}\right] \\ Q_{Y|X}\left(\tau - \dfrac{a}{p_1} \mid 0\right) & \text{for } \tau \in \left(\dfrac{a}{p_1}, \dfrac{a}{p_1} + b - a\right] \\ Q_{Y|X}(b - a \mid 0) & \text{for } \tau \in \left(\dfrac{a}{p_1} + b - a, 1\right). \end{cases}$$

Next consider the upper bound. If $(1-(b-a))p_0 \leq a$,

$$\overline{Q}^{\mathcal{U}}_{Y_0|X}(\tau \mid 1) = \begin{cases} Q_{Y|X}(1-(b-a) \mid 0) & \text{for } \tau \in \left(0, 1-(b-a)-\dfrac{1-b}{p_1}\right] \\ Q_{Y|X}\left(\tau + \dfrac{1-b}{p_1} \mid 0\right) & \text{for } \tau \in \left(1-(b-a)-\dfrac{1-b}{p_1}, 1-\dfrac{1-b}{p_1}\right] \\ Q_{Y|X}(1 \mid 0) & \text{for } \tau \in \left(1-\dfrac{1-b}{p_1}, 1\right). \end{cases}$$

If $(1-(b-a))p_0 \geq a$,

$$\overline{Q}^{\mathcal{U}}_{Y_0|X}(\tau \mid 1) = \begin{cases} Q_{Y|X}\left(\tau + \dfrac{a}{p_0} \mid 0\right) & \text{for } \tau \in (0, b-a] \\ Q_{Y|X}(1 \mid 0) & \text{for } \tau \in (b-a, 1). \end{cases}$$

Proposition 4.1. *Let assumption 3.1 hold. Suppose Y_0 is $\mathcal{T}$-independent of X with $\mathcal{T} = [Q_{Y_0}(a), Q_{Y_0}(b)]$, $0 < a \leq b < 1$. Suppose the joint distribution of (Y, X) is known. Let $q \in (0,1)$. Then*

$$Q_{Y_0|X}(q \mid 1) \in \left[\underline{Q}^{\mathcal{T}}_{Y_0|X}(q \mid 1),\, \overline{Q}^{\mathcal{T}}_{Y_0|X}(q \mid 1)\right]. \tag{4.8}$$

Moreover, the interior of the set in equation (4.8) *equals the interior of the identified set. Finally, the proposition also holds if we replace $\mathcal{T}$ with $\mathcal{U}$.*

$\mathcal{T}$-independence of Y_0 from X with $\mathcal{T} = [Q_{Y_0}(a), Q_{Y_0}(b)]$ is equivalent to the quantile independence assumptions $Q_{Y_0|X}(\tau \mid x) = Q_{Y_0}(\tau)$ for all $\tau \in [a, b]$, by assumption 3.1. The bounds (4.8) are also sharp for the function $Q_{Y_0|X}(\cdot \mid 1)$ in a sense similar to that used in proposition S3.1 in the appendix; we omit the formal statement for brevity. This functional sharpness delivers the following result.

Corollary 4.2. *Suppose the assumptions of proposition 4.1 hold. Let $\mathbb{E}(|Y_0|) < \infty$. Then $\mathbb{E}(Y_0 \mid X = 1)$ lies in the set*

$$\left[\underline{\mathbb{E}}^{\mathcal{T}}(Y_0 \mid X = 1),\, \overline{\mathbb{E}}^{\mathcal{T}}(Y_0 \mid X = 1)\right] \equiv \left[\int_0^1 \underline{Q}^{\mathcal{T}}_{Y_0|X}(q \mid 1)\, dq,\, \int_0^1 \overline{Q}^{\mathcal{T}}_{Y_0|X}(q \mid 1)\, dq\right].$$

Moreover, the interior of this set equals the interior of the identified set for $\mathbb{E}(Y_0 \mid X = 1)$. Finally, the corollary also holds if we replace $\mathcal{T}$ with $\mathcal{U}$.

By proposition 4.1 we have that $\mathcal{T}$-independence implies that $\mathrm{QTT}(q)$ lies in the set

$$\left[Q_{Y|X}(q \mid 1) - \overline{Q}^{\mathcal{T}}_{Y_0|X}(q \mid 1),\; Q_{Y|X}(q \mid 1) - \underline{Q}^{\mathcal{T}}_{Y_0|X}(q \mid 1)\right]$$

and that the interior of this set equals the interior of the identified set for $\mathrm{QTT}(q)$. Likewise for $\mathcal{U}$-independence. If $q \in \mathcal{T}$, then $\mathrm{QTT}(q)$ is point identified under $\mathcal{T}$-independence; this follows immediately from our bound expressions above. This result—that a single quantile independence condition can be sufficient for point identifying a treatment effect—was shown by Chesher (2003). A similar result holds in the instrumental variables model of Chernozhukov and Hansen (2005) and the LATE model of Imbens and Angrist (1994). See the discussion around assumption 4 in section 1.4.3 of Melly and Wüthrich (2017).

By corollary 4.2 we have that $\mathcal{T}$-independence implies that the ATT lies in the set

$$\left[\mathbb{E}(Y \mid X=1) - \overline{\mathbb{E}}^{\mathcal{T}}(Y_0 \mid X=1),\ \mathbb{E}(Y \mid X=1) - \underline{\mathbb{E}}^{\mathcal{T}}(Y_0 \mid X=1)\right]$$

and that the interior of this set equals the interior of the identified set for the ATT. Likewise for $\mathcal{U}$-independence. Furthermore, in appendix S2 we show that these ATT bounds have simple analytical expressions, obtained from integrating our closed form expressions for the bounds on $Q_{Y_0|X}(q \mid 1)$.

Remarks on Estimation and Inference

The bounds on QTT(τ) and ATT depend on the data through two of its features: the conditional quantile function $Q_{Y|X}(\tau \mid x)$ for $(\tau, x) \in [0,1] \times \{0,1\}$ and the marginal probability of treatment p_1. This is the case under both $\mathcal{T}$- and $\mathcal{U}$-independence. One can estimate these bounds using a variety of sample analog estimators. For example, we can estimate p_1 using the sample frequency of treatment. Similarly, for $\tau \in (0,1)$, $Q_{Y|X}(\tau \mid x)$ can be estimated using sample quantiles or by inverting a possibly smoothed estimator of the conditional cdf of $Y \mid X$. These estimators have well known asymptotic properties. However, these bounds depend on $Q_{Y|X}(\tau \mid x)$ for $\tau = 0$ and $\tau = 1$ as well. These quantiles can be estimated using methods for extremal quantiles: see Chernozhukov (2005) and Chernozhukov et al. (2017) for example, and they generally have non-standard asymptotic behavior.

When a vector of covariates W is observed, estimators of $Q_{Y|X,W}$ and $\mathbb{P}(X = x \mid W = w)$ must be obtained for all w in the support of W. These can be estimated nonparametrically, or under flexible parametric assumptions as well: see Masten et al. (2020) for example. We leave a full analysis of estimation and inference under these relaxations of independence for future research.

5. EMPIRICAL ILLUSTRATION: THE EFFECT OF CHILD SOLDIERING ON WAGES

In this section we use our results to study the impact of relaxing the unconfoundedness assumption in an empirical study of the effects of child soldiering on wages. We do this using both the $\mathcal{T}$- and $\mathcal{U}$-independence relaxations of statistical independence. We find that the identified sets are substantially larger under $\mathcal{U}$-independence. This implies that the average value constraints imposed by $\mathcal{T}$-independence have substantial identifying power; recall that these constraints are the features of quantile independence that require the latent propensity score to be non-monotonic. In particular, the baseline empirical results are generally quite robust under the $\mathcal{T}$-independence relaxation of unconfoundedness, but not under $\mathcal{U}$-independence. This difference highlights the importance of the choice of exogeneity assumptions in practice, and how researchers can use their beliefs about the form of latent selection to assist in this choice.

Background

By collecting extensive survey data, Blattman and Annan (2010) study the impact of child abductions during a twenty year war in Uganda, where "an unpopular rebel group has forcibly recruited tens of thousands of youth" (page 882). Although they consider a variety of outcome variables, we focus on the impact of abduction on later life wages.

The main identification problem is that selection into military service is typically nonrandom. They argue, however, that forced recruitment in Uganda led to conditional random assignment of military service. They condition on two variables: (1) Prewar household size, because larger households were less likely to be raided by small bands of rebels, and (2) Year of birth, because abduction levels varied over time, so that some youth ages were more likely to be abducted than others. Hence their identification strategy is based on unconfoundedness, conditioning on these two variables. Although their qualitative evidence supporting unconfoundedness is compelling, this assumption is still nonrefutable. We therefore use our results to assess the sensitivity of relaxing unconfoundedness on their empirical conclusions.

Sample Definition

The data comes from phase 1 of SWAY, the Survey of War Affected Youth in northern Uganda (see Annan et al. 2006 for details of the original data collection; we use the version of this data from Masten and Poirier 2020a,b). This phase has 1216 males born between 1975 and 1991. We look at the subsample of units who (1) have wage data available and (2) earned positive wages. This leaves us with 448 observations. Let Y denote log wage. We define treatment X to be an indicator that the person was *not* abducted. We include the two covariates discussed above, age when surveyed and household size in 1996. We omit other covariates for simplicity.

Age has 17 support points, household size has 21 support points, and treatment has 2 support points. Hence there are 714 total conditioning variable cells, relative to our sample size of 448 observations. To ensure that our conditional quantile estimators are reasonably smooth in the quantile index, we collapse these conditioning variables into 8 cells. Specifically, we replace age with a binary indicator of whether one is above or below the median age. Likewise, we replace household size with a binary indicator of whether one lived in a household with above or below median household size. This gives 8 total conditioning variable cells, with approximately 55 observations each.

Baseline Analysis

First we present estimates of conditional average treatment effects for the treated (CATT) and conditional quantile treatment effects for the treated (CQTTs), under the unconfoundedness assumption. For brevity we focus on the covariate cell $w =$ (age, household size) = (above median, above median). This group has the largest baseline effects of treatment, meaning that being abducted lowered their later life wages by the largest amount. Specifically, our estimate of the CATT for this group is 0.57. Our CQTT estimates are 0.67 for $\tau = 0.25$, 0.54 for $\tau = 0.5$, and 0.56 for $\tau = 0.75$. Note that our sample size is small, with 121 observations in this cell. We omit standard errors here because the purpose of this section is to illustrate the methods developed in our paper.

Sensitivity Analysis

To check the robustness of these baseline point estimates to failure of unconfoundedness, we estimate identified sets for the CATT and CQTT using our results from section 4. To highlight the importance of the choice of relaxation, we consider sets $\mathcal{T} = \mathcal{U}$. In this case, corollary 4.1 shows that $\mathcal{T}$-independence implies $\mathcal{U}$-independence. Hence identified

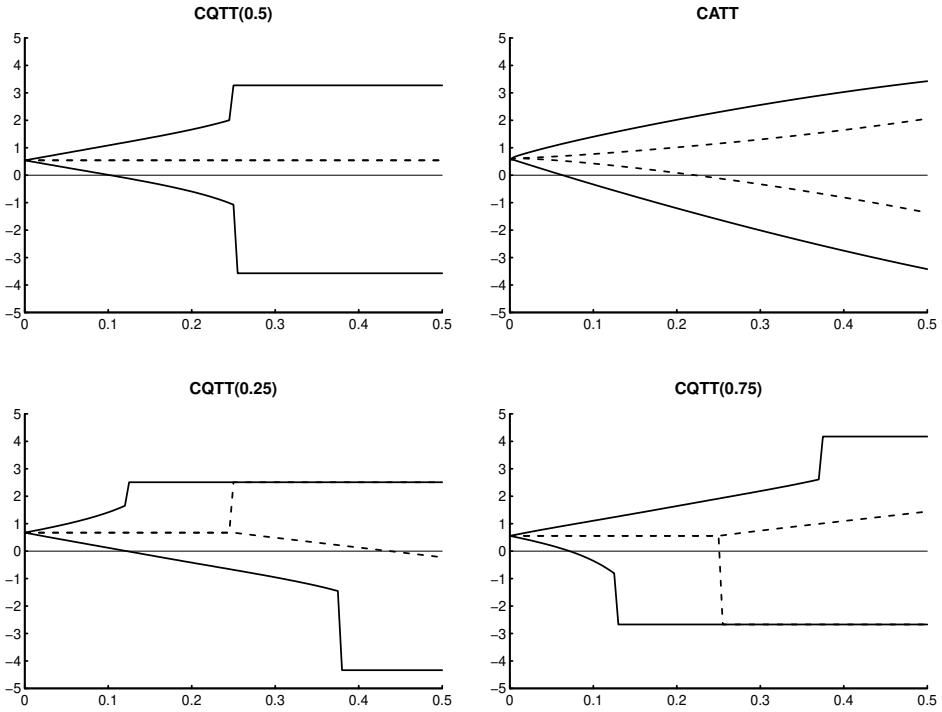

Figure 4. Estimated identified sets for various parameters of interest, for $\mathcal{U} = \mathcal{T} = [\delta, 1-\delta]$ with $\delta \in [0, 0.5]$. Solid: $\mathcal{U}$-independence. Dashed: $\mathcal{T}$-independence. The horizontal axis shows values of δ. All parameters are conditional on the covariate cell $w =$ (age, household size) = (above median, above median).

sets using $\mathcal{T}$-independence must necessarily be weakly contained within identified sets using only $\mathcal{U}$-independence. We explore the magnitude of this difference in the data. Since $\mathcal{T}$-independence is simply $\mathcal{U}$-independence combined with some additional average value constraints, the difference between these identified sets tells us the identifying power of these additional average value constraints.

Specifically, we use the choice $\mathcal{T} = \mathcal{U} = [\delta, 1 - \delta]$ for $\delta \in [0, 0.5]$. For $\delta = 0$, this choice corresponds to full conditional independence $Y_0 \perp\!\!\!\perp X \mid W = w$ under both classes of assumptions. For $\delta = 0.5$, this choice corresponds to median independence for $\mathcal{T}$-independence, and no assumptions for $\mathcal{U}$-independence. Values of δ between 0 and 0.5 yield conditional partial independence between Y_0 and X for both classes of assumptions.

Figure 4 shows estimated identified sets for both CATT and CQTT(τ) as δ varies from 0 to 0.5, and for $\tau \in \{0.25, 0.5, 0.75\}$. These are sample analog estimates, where $\widehat{Q}_{Y|X,W}(\cdot \mid x, w)$ is estimated by inverting a kernel based estimate of $F_{Y|X,W}(\cdot \mid x, w)$. First consider the plot on the top left, which shows the estimated CQTT(0.5) bounds. The dashed lines are the identified sets under $\mathcal{T}$-independence. Since median independence of Y_0 from X conditional on $W = w$ is sufficient to point identify the conditional median $Q_{Y_0|X,W}(0.5 \mid 1, w)$, median independence is also sufficient to point identify the CQTT at 0.5. Hence the identified set is a singleton for all $\delta \in [0, 0.5]$. This singleton equals 0.54,

the baseline estimate. Next consider the solid lines. These are the estimated identified sets under $\mathcal{U}$-independence. When $\delta = 0.5$, $\mathcal{U}$-independence does not impose any constraints on the model, and hence we obtain the no assumption bounds, which are quite wide: $[-3.42, 3.43]$. If we decrease δ by a small amount, thus making the $\mathcal{U}$-independence constraint nontrivial, the estimated identified set does not change. In fact, we can impose random assignment for about the middle 50% of units (i.e., $\mathcal{U} = [0.25, 0.75]$, or $\delta = 0.25$) and still we only obtain the no assumption bounds. Consequently, for intervals $\mathcal{T} \subseteq [0.25, 0.75]$, the point identifying power of $\mathcal{T}$-independence is due solely to the constraint it imposes on the average value of the latent propensity score outside the interval $\mathcal{T}$, rather than the constraint that random assignment holds for units in the middle of the distribution of Y_0.

Next define

$$\delta^{\mathcal{U}}_{\text{bp}}(\tau) = \sup\{\delta \in [0, 0.5] : \text{LB}^{\mathcal{U}}(\tau, \delta) \geq 0\}$$

where $\text{LB}^{\mathcal{U}}(\tau, \delta)$ is the lower bound of the identified set for $\text{CQTT}(\tau)$ under $\mathcal{U}$-independence with $\mathcal{U} = [\delta, 1 - \delta]$. Define $\delta^{\mathcal{T}}_{\text{bp}}(\tau)$ analogously. This value $\delta^{\mathcal{U}}_{\text{bp}}(\tau)$ is a breakdown point: It is the largest amount we can relax full independence while still being able to conclude that the treatment effect is nonnegative. For $\tau = 0.5$, the estimated breakdown point for CQTT(0.5) is 0.103. Thus we can allow randomization to fail for about 20.6% of units while still being able to conclude that CQTT(0.5) is nonnegative. In contrast, as mentioned above, the breakdown point for $\mathcal{T}$-independence is always 0.5.

Next consider the lower two plots of figure 4. These plots show estimated identified sets for CQTT(0.25) on the left and CQTT(0.75) on the right. There are two main differences between these plots and that of CQTT(0.5): First, the $\mathcal{U}$-independence upper and lower bounds are not symmetric. Nonetheless, the qualitative robustness conclusions are similar. For example, $\widehat{\delta}^{\mathcal{U}}_{\text{bp}}(0.25)$ is 0.122 and $\widehat{\delta}^{\mathcal{U}}_{\text{bp}}(0.75)$ is 0.071. So conclusions about smaller quantiles are slightly more robust than conclusions about larger quantiles. Second, the $\mathcal{T}$-independence identified sets are no longer always singletons. In particular, we obtain non-singleton bounds when $\delta > 0.25$. However, conclusions under the $\mathcal{T}$-independence relaxation are substantially more robust than conclusions under the $\mathcal{U}$-independence relaxation. Specifically, $\widehat{\delta}^{\mathcal{T}}_{\text{bp}}(0.25)$ is 0.437. This is about 3.5 times as large as $\widehat{\delta}^{\mathcal{U}}_{\text{bp}}(0.25)$. Similarly, $\widehat{\delta}^{\mathcal{T}}_{\text{bp}}(0.75)$ is 0.25. This is also about 3.5 times as large as $\widehat{\delta}^{\mathcal{U}}_{\text{bp}}(0.75)$.

Finally consider the plot on the top right of figure 4, which shows estimated identified sets for CATT. First consider the $\mathcal{T}$-independence relaxation, the dashed lines. The CATT is no longer point identified under median independence, or any set $\mathcal{T} \subsetneq (0, 1)$ of quantile independence conditions; that is, the CATT is partially identified for all $\delta > 0$. Nonetheless, even median independence alone has substantial identifying power: For $\delta = 0.5$, the estimated identified set under median independence is $[-1.36, 2.06]$, whereas the no assumption bounds are $[-3.42, 3.43]$. Thus the width of the bounds has been cut in half. For $\delta > 0$, $\mathcal{U}$-independence has non-trivial identifying power, as shown in the solid lines. However, comparing the length of these bounds to the length to the $\mathcal{T}$-independence bounds, we see that imposing the average value constraint outside the interval $[\delta, 1 - \delta]$ again has substantial identifying power: the $\mathcal{T}$-independence bounds are anywhere from 50% ($\delta = 0.5$) to almost 100% (arbitrarily small δ) smaller than the $\mathcal{U}$-independence bounds. That is, the difference in lengths increases as we get closer to independence (as δ gets smaller). Thus conclusions about CATT are substantially more sensitive to small deviations from independence which do not impose the average value constraint outside the interval $[\delta, 1-\delta]$, compared with small deviations which do impose

that constraint. A second way to see this is to compare the breakdown points under the two relaxations. Define

$$\delta^{\mathcal{U}}_{\text{bp}} = \sup\{\delta \in [0, 0.5] : \text{LB}^{\mathcal{U}}(\delta) \geq 0\}$$

where $\text{LB}^{\mathcal{U}}(\delta)$ is the lower bound of the identified set for CATT under $\mathcal{U}$-independence with $\mathcal{U} = [\delta, 1-\delta]$. Define $\delta^{\mathcal{T}}_{\text{bp}}$ analogously. As shown in the plots above, $\widehat{\delta}^{\mathcal{U}}_{\text{bp}} = 0.063$ while $\widehat{\delta}^{\mathcal{T}}_{\text{bp}} = 0.222$. Thus the breakdown point under $\mathcal{T}$-independence is again about 3.5 times as large as the breakdown point under $\mathcal{U}$-independence.

Empirical Conclusions

In this section we used our identification results to study the robustness of conclusions about CATT and CQTTs to failures of unconfoundedness. Our baseline point estimates suggest that child abduction and forced military service has a negative effect on later life wages, for those children who were older when they were abducted and who came from larger households. This holds both on average (from the CATT) and across the distribution of treatment effects (as seen in the CQTTs). We then asked: How sensitive are these conclusions to failures of unconfoundedness? We saw that using the $\mathcal{T}$-independence relaxation, these conclusions are generally robust to large relaxations of unconfoundedness. However, using the $\mathcal{U}$-independence relaxation, these conclusions appear much more sensitive. As we earlier discussed, the difference arises from the additional average value constraints that $\mathcal{T}$-independence imposes. Those constraints are the features of quantile independence that require the latent propensity score to be non-monotonic. Thus it is critical to assess the plausibility of those additional constraints when deciding between these two forms of exogeneity assumptions to use for assessing sensitivity.

In this empirical context, a monotonic latent propensity score arises when youths who have larger potential earnings when they're abducted (larger Y_0) are more likely to be abducted. If youths are targeted for abduction because of their innate or pre-existing skills, which would generally lead to large Y_0, then this would be a form of monotonic selection that would *not* be allowed for by the $\mathcal{T}$-independence relaxation, but *would* be allowed for by the $\mathcal{U}$-independence relaxation. So if we are concerned that unconfoundedness fails due to this kind of non-random selection, then $\mathcal{U}$-independence is a more appropriate choice for assessing sensitivity than $\mathcal{T}$-independence. Given this choice, the baseline results still hold under mild relaxations of unconfoundedness, since we saw that $\mathcal{U}$-independence breakdown points were generally around $\delta = 0.1$. But the baseline results no longer hold for larger relaxations; in this case, the data are inconclusive.

6. THE TREATMENT SELECTION IMPLICATIONS OF A ROY MODEL

As we emphasized, there is a direct mapping between exogeneity assumptions and the allowed forms of treatment selection. At one extreme, full independence assumes no selection at all of X on Y_x, and therefore $p(y_x)$ is constant. On the other hand, weaker exogeneity assumptions allow for a class of deviations that one wishes to be robust against. Since this class is often not explicitly specified, we refer to such deviations as *latent selection models*. Our main results in section 3 characterize the set of latent selection models allowed by quantile and mean independence restrictions.

In this section, we consider a class of Roy models and examine the relationship between their implied treatment selection functions and the exogeneity assumptions of section 3.

We discuss different assumptions on the economic primitives which lead these models to be either consistent or inconsistent with quantile or mean independence restrictions. We only consider single-agent models, but similar analyses can likely be done for multi-agent models.

Suppose we are again interested in identifying the average treatment effect for the treated parameter

$$\begin{aligned}\text{ATT} &= \mathbb{E}(Y_1 - Y_0 \mid X = 1) \\ &= \mathbb{E}(Y \mid X = 1) - \mathbb{E}(Y_0 \mid X = 1).\end{aligned}$$

As in section 4, its identification depends on our assumptions about the stochastic relationship between X and Y_0. Suppose agents choose treatment to maximize their outcome:

$$X = \mathbb{1}(\Delta > 0) \tag{6.9}$$

where $\Delta \equiv Y_1 - Y_0$ is the unit level treatment effect. This is the classical Roy model (see Heckman and Vytlacil 2007a). This assumption specifies how treatment X relates to Y_0. Specifically, consider the latent propensity score

$$\begin{aligned}p(y_0) &\equiv \mathbb{P}(X = 1 \mid Y_0 = y_0) \\ &= \mathbb{P}(\Delta > 0 \mid Y_0 = y_0).\end{aligned}$$

The second line follows by our Roy model treatment choice assumption. Thus the shape of p depends on the joint distribution of (Y_1, Y_0) or, equivalently, of (Δ, Y_0). We classify these distributions into two possible cases, based on a concept called regression dependence, which dates back to Tukey (1958) and Lehmann (1966), who give the following definition.

Definition 6.1. *Say X is* positively [negatively] regression dependent *on U if $\mathbb{P}(X > x \mid U = u)$ is weakly increasing [decreasing] in u, for all $x \in \mathbb{R}$. Say X is* regression dependent *on U if it is either positively or negatively regression dependent on U.*

Regression dependence is also known as *stochastic monotonicity*, since it is equivalent to the set of cdfs $\{F_{X|U}(\cdot \mid u) : u \in \text{supp}(U)\}$ being either increasing or decreasing in the first order stochastic dominance ordering. We give additional background on this definition in appendix S1.1.

1 First suppose (Y_1, Y_0) is such that Δ is regression dependent on Y_0. This implies that p is monotonic. Corollary 3.1 and proposition 3.1 therefore imply that no mean or quantile independence conditions of Y_0 on X can hold unless $X \perp\!\!\!\perp Y_0$. This occurs when X is degenerate, as when treatment effects Δ are constant, or more generally when $\Delta \perp\!\!\!\perp Y_0$. In particular, any mean or quantile independence condition of Y_0 on X rules out the bivariate normal distribution for (Δ, Y_0) or (Y_1, Y_0), again unless $X \perp\!\!\!\perp Y_0$.

2 Next suppose (Y_1, Y_0) is such that Δ is not regression dependent on Y_0. For example, let $\Delta = \mu(Y_0) - \varepsilon$ where μ is a deterministic function and $\varepsilon \sim \mathcal{N}(0, 1)$, $\varepsilon \perp\!\!\!\perp Y_0$. Then

$$p(y_0) = \mathbb{P}(X = 1 \mid Y_0 = y_0) = \Phi[\mu(y_0)],$$

where Φ is the standard normal cdf. If μ is non-monotonic then p will also be non-monotonic. For this joint distribution of potential outcomes, the unit level treatment effects Δ conditional on the baseline outcome $Y_0 = y_0$ are distributed

$\mathcal{N}(\mu(y_0), 1)$. Hence non-monotonicity of μ implies that the mean of this distribution of treatment effects is not monotonic. For instance, suppose the outcome is earnings and treatment is completing college. Let

$$\begin{aligned} \mu(y_0) &> 0 \qquad \text{if } y_0 \in (\alpha, \beta) \\ \mu(y_0) &\leq 0 \qquad \text{if } y_0 \in (-\infty, \alpha] \cup [\beta, \infty) \end{aligned}$$

for $-\infty < \alpha < \beta < \infty$. Then people with sufficiently small or sufficiently large earnings when they do not complete college do not benefit from completing college, on average. People with moderate earnings when they do not complete college, on the other hand, do typically benefit from completing college. This kind of joint distribution of potential outcomes combined with the Roy model assumption (6.9) on treatment selection produces non-monotonic latent propensity scores.

We just gave one example joint distribution of (Y_1, Y_0) where regression dependence fails. More generally, theorem 5.2.10 on page 196 of Nelsen (2006) characterizes the set of copulas for which Δ is regression dependent on Y_0, when both are continuously distributed. This result therefore also tells us the set of copulas where Δ is *not* regression dependent on Y_0. Among these copulas, $\mathcal{T}$-independence (or, analogously, mean independence) of Y_0 from X will specify a further subset of allowed dependence structures. The precise set is given by all copulas which lead to latent propensity scores that satisfy the average value constraint.

Which of these cases is plausible depends on the specific application at hand.

7. CONCLUSION

In this paper we gave several results to help researchers assess the plausibility of quantile and mean independence assumptions on structural unobservables like potential outcomes. Keep in mind, however, that when doing identification analysis it is not necessary to choose a single exogeneity assumption. For example, researchers may want to consider a variety of exogeneity assumptions in this step, as part of a sensitivity analysis. We illustrated this in sections 4 and 5. The choice of which exogeneity assumptions to consider is still determined by considering the kinds of treatment selection we want to allow for, as discussed in section 2. Conversely, there may be situations where researchers do not find it plausible to impose *any* kind of exogeneity assumption. In this case we often can still learn something about the parameters of interest, as in the classical no assumption bounds of Manski (1990). In this paper we focused on the case where the researcher does want to impose some kind of exogeneity assumption, however. In this case, we hope that our results can help researchers better select the most appropriate exogeneity assumptions for their settings.

ACKNOWLEDGMENTS

This paper supersedes our previous paper titled "Interpreting Quantile Independence" (arXiv:1804.10957), as well as the unpublished sections 3.2, 3.4, and 4.1 of our inactive working paper Masten and Poirier (2016). We thank the editor, the referees, audiences at various seminars and conferences, as well as Federico Bugni, Ivan Canay, Andrew Chesher, Joachim Freyberger, Bo Honoré, Joel Horowitz, Shakeeb Khan, Ivana Komunjer, Roger Koenker, Chuck Manski, Francesca Molinari, Jim Powell, and Adam Rosen for helpful conversations and comments. We thank Hongchang Guo for excellent

research assistance. Masten thanks the National Science Foundation for research support under Grant No. 1943138.

REFERENCES

Abbring, J. H. and J. J. Heckman (2007). Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation. *Handbook of Econometrics 6*, 5145–5303.

Angrist, J. D. (2001). Estimation of limited dependent variable models with dummy endogenous regressors: Simple strategies for empirical practice. *Journal of Business & Economic Statistics 19*(1), 2–28.

Annan, J., C. Blattman, and R. Horton (2006). The state of youth and youth protection in northern uganda. *Final Report for UNICEF Uganda*.

Belloni, A., M. Chen, and V. Chernozhukov (2017). Quantile graphical models: Prediction and conditional independence with applications to systemic risk. *Working paper*.

Blattman, C. and J. Annan (2010). The consequences of child soldiering. *The Review of Economics and Statistics 92*(4), 882–898.

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*(2), 323–363.

Blundell, R. and R. L. Matzkin (2014). Control functions in nonseparable simultaneous equations models. *Quantitative Economics 5*(2), 271–295.

Chernozhukov, V. (2005). Extremal quantile regression. *The Annals of Statistics 33*(2), 806–839.

Chernozhukov, V., I. Fernández-Val, and T. Kaji (2017). Extremal quantile regression. *Handbook of Quantile Regression*.

Chernozhukov, V. and C. Hansen (2005). An IV model of quantile treatment effects. *Econometrica 73*(1), 245–261.

Chesher, A. (2003). Identification in nonseparable models. *Econometrica 71*(5), 1405–1441.

Fox, J. T. (2007). Semiparametric estimation of multinomial discrete-choice models using a subset of choices. *The RAND Journal of Economics 38*(4), 1002–1019.

Hausman, J. A. (1983). Specification and estimation of simultaneous equation models. *Handbook of Econometrics 1*, 391–448.

Heckman, J. J. (2000). Causal parameters and policy analysis in economics: A twentieth century retrospective. *The Quarterly Journal of Economics 115*(1), 45–97.

Heckman, J. J., H. Ichimura, and P. Todd (1998). Matching as an econometric evaluation estimator. *The Review of Economic Studies 65*(2), 261–294.

Heckman, J. J. and E. J. Vytlacil (2007a). Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation. *Handbook of Econometrics 6*, 4779–4874.

Heckman, J. J. and E. J. Vytlacil (2007b). Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. *Handbook of Econometrics 6*, 4875–5143.

Hendry, D. F. and M. S. Morgan (1995). *The Foundations of Econometric Analysis*. Cambridge University Press.

Imbens, G. W. (1997). Book review: 'The Foundations of Econometric Analysis' by David F. Hendry and Mary S. Morgan. *Journal of Applied Econometrics 12*(1), 91–94.

Imbens, G. W. (2003). Sensitivity to exogeneity assumptions in program evaluation. *American Economic Review P&P 93*(2), 126–132.

Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics 86*(1), 4–29.

Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

Komarova, T., T. A. Severini, and E. T. Tamer (2012). Quantile uncorrelation and instrumental regressions. *Journal of Econometric Methods 1*(1), 2–14.

Lehmann, E. L. (1966). Some concepts of dependence. *The Annals of Mathematical Statistics 37*(5), 1137–1153.

Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics 3*(3), 205–228.

Manski, C. F. (1977). The structure of random utility models. *Theory and Decision 8*(3), 229–254.

Manski, C. F. (1988). Identification of binary response models. *Journal of the American Statistical Association 83*(403), 729–738.

Manski, C. F. (1990). Nonparametric bounds on treatment effects. *American Economic Review P&P 80*(2), 319–323.

Manski, C. F. (2003). *Partial Identification of Probability Distributions*. Springer.

Manski, C. F. and J. V. Pepper (2000). Monotone instrumental variables: With an application to the returns to schooling. *Econometrica 68*(4), 997–1010.

Manski, C. F. and J. V. Pepper (2009). More on monotone instrumental variables. *The Econometrics Journal 12*(s1), S200–S216.

Masten, M. A. and A. Poirier (2016). Partial independence in nonseparable models. *Working paper*.

Masten, M. A. and A. Poirier (2018). Identification of treatment effects under conditional partial independence. *Econometrica 86*(1), 317–351.

Masten, M. A. and A. Poirier (2020a). Inference on breakdown frontiers. *Quantitative Economics 11*(1), 41–111.

Masten, M. A. and A. Poirier (2020b). Replication package of "Inference on breakdown frontiers". *Quantitative Economics 11*(1), 41–111. Data deposited at `https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.3982%2FQE1288&file=quan200069-sup-0002-dataandprograms.zip`.

Masten, M. A., A. Poirier, and L. Zhang (2020). Assessing sensitivity to unconfoundedness: Estimation and inference. *Working paper*.

Melly, B. and K. Wüthrich (2017). Local quantile treatment effects. *Handbook of Quantile Regression*.

Nelsen, R. B. (2006). *An Introduction to Copulas* (Second ed.). Springer.

Powell, J. L. (1994). Estimation of semiparametric models. *Handbook of Econometrics 4*, 2443–2521.

Rosenbaum, P. R. (1995). *Observational Studies*. Springer.

Rosenbaum, P. R. (2002). *Observational Studies* (Second ed.). Springer.

Rosenbaum, P. R. and D. B. Rubin (1983). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society, Series B 45*(2), 212–218.

Torgovitsky, A. (2019). Partial identification by extending subdistributions. *Quantitative Economics 10*(1), 105–144.

Tukey, J. W. (1958). A problem of Berkson, and minimum variance orderly estimators. *The Annals of Mathematical Statistics 29*(2), 588–592.

Zhu, L., Y. Zhang, and K. Xu (2018). Measuring and testing for interval quantile dependence. *The Annals of Statistics 46*(6A), 2683–2710.

APPENDIX A: STRUCTURAL AND REDUCED FORM UNOBSERVABLES

The approach we recommend in section 2 begins by distinguishing between two kinds of unobservables: (1) Structural unobservables and (2) Reduced form unobservables. In this appendix we discuss a specific example to clarify the distinction between these two variables. We also use this example to discuss the difference between exogeneity assumptions involving reduced form unobservables and those involving structural unobservables.

Example: The Binary Response Random Coefficients Model Let X be a scalar observed random variable. Let $Y^*(x)$ denote a latent potential outcome for $x \in \mathbb{R}$. Let

$$Y(x) = \mathbb{1}[Y^*(x) \geq 0]$$

denote the usual potential outcomes. Let $Y = Y(X)$ denote the observed outcome. Suppose latent potential outcomes satisfy the linear random coefficient model

$$Y^*(x) = A + Bx, \tag{A.1}$$

where A and B are structural unobserved random variables. Suppose we impose the following constraint on the tails of A, B, and X.

Assumption A.1. *$\mathbb{E}(A)$, $\mathbb{E}(B)$, and $\mathbb{E}(X)$ exist and are finite.*

In this model, (A, B) are structural unobservables. An exogeneity assumption about the relationship between (A, B) and X is therefore a statement about treatment selection: How does assigned treatment depend on these structural unobservables? This is the kind of exogeneity assumption we focus on in this paper. In this appendix, however, we will discuss a second kind of exogeneity condition, which constrains the relationship between treatment and a reduced form unobservable. This is a *derived* exogeneity condition: It is not directly imposed but rather is a consequence of a choice of (1) an exogeneity assumption involving the structural unobservables and (2) the functional form of the reduced form unobservables.

To illustrate this kind of derived exogeneity condition, we'll assume treatment is randomly assigned.

Assumption A.2. *$(A, B) \perp\!\!\!\perp X$.*

Next write the equation for realized latent potential outcomes as

$$\begin{aligned} Y^*(X) &= \mathbb{E}(A) + \mathbb{E}(B)X + \big([A - \mathbb{E}(A)] + [B - \mathbb{E}(B)]X\big) \\ &\equiv \mathbb{E}(A) + \mathbb{E}(B)X + V. \end{aligned} \tag{A.2}$$

V is a reduced form unobservable. It is a function of the structural unobservables (A, B) as well as the realized treatment X. Consequently, it is not invariant to changes in the distribution of X. The coefficients $\mathbb{E}(A)$ and $\mathbb{E}(B)$ in equation (A.2) do have structural interpretations, however.

By definition, V depends on X, and so typically V is not independent of X. Nonetheless, we can derive some restrictions on the distribution of $V \mid X$. Specifically, suppose we also make the following assumption.

ASSUMPTION A.3. *$Y^*(x)$ is symmetrically distributed about $\mathbb{E}[Y^*(x)]$ for all $x \in \text{supp}(X)$.*

Under this additional assumption, Manski (1975, page 220) showed the following result; also see Manski (1977, pages 247–249) and Fox (2007, pages 1007–1008).

PROPOSITION A.1. *Suppose assumptions A.1, A.2, and A.3 hold. Then $\mathbb{P}(V \leq 0 \mid X = x) = \mathbb{P}(V \leq 0) = 0.5$ for all $x \in \text{supp}(X)$. That is, V is median independent of X.*

Thus we have derived a median independence restriction on the reduced form unobservable V as a consequence of (1) the definition of V and (2) the exogeneity assumption about the relationship between X and the structural unobservables.

Discussion In this example there are two kinds of unobservables: The structural unobservables (A, B) and the reduced form unobservable V. We made an exogeneity assumption about the relationship between (A, B) and X based on our beliefs about treatment assignment. We then *derived* an exogeneity condition on the relationship between the reduced form unobservable V and X. More generally, for researchers interested in choosing an exogeneity assumption that relates treatment to reduced form unobservables, our recommendation is that this assumption be derived from a more primitive exogeneity assumption about the structural unobservables, as in the example. The analysis in sections 2–6 of our paper can then be used to assess the plausibility of these more primitive exogeneity assumptions.

Researchers may sometimes prefer to make an exogeneity assumption on the reduced form unobservable directly, without directly deriving it from a more primitive model like we did above. There are a few concerns with this approach, however:

1 As Angrist (2001) argues, we often care about parameters like average structural functions (ASFs) and average treatment effects (ATEs). In nonseparable models like the example above, however, the ASF depends on the distribution of the structural unobservables. In particular, for that example,

$$\begin{aligned} \text{ASF}(x) &= \mathbb{E}[Y(x)] \\ &= \mathbb{P}_{A,B}(A + Bx \geq 0) \\ &\neq \mathbb{P}_V(\mathbb{E}(A) + \mathbb{E}(B)x + V \geq 0). \end{aligned}$$

The true ASF does *not* generally equal the parameter that you would compute if you worked with equation (A.2), but incorrectly treated V as a structural unobservable in an ASF calculation. Thus, in nonseparable models, it is generally not possible to avoid working with structural unobservables if we are interested in ASFs and ATEs.

For example, among many other derivations, Torgovitsky (2019) computes identified sets for ASFs and ATEs in a binary response model with constant coefficients and median independence. For these identified sets to have the correct interpretation, the unobservable V in his model must be structural, rather than a reduced

form. Consequently, the fact that we are only imposing median independence of these unobservables from treatment implies that we are concerned with a particular kind of non-random treatment assignment. Our results in section 3 characterize the kind of non-random treatment assignment consistent with median independence assumptions.

2 If we only work with the reduced form unobservables, then we might be ignoring a lot of useful information. Consider the example again: Relative to the assumption that V is median independent of X, the stronger assumptions A.2 and A.3 have potentially different implications for falsification, identification, rates of convergence, and efficiency. For example, in the model we know that the ASF is point identified:

$$\mathbb{E}[Y(x)] = \mathbb{P}(Y = 1 \mid X = x).$$

But if we only impose median independence of V from X then the ASF is generally only partially identified.

For these reasons, we recommend working directly with the structural unobservables. This does not require that researchers make strong assumptions like statistical independence on these structural unobservables, however. Instead, they can use the methods discussed in this paper to think about the form of exogeneity they want to impose on the relationship between the structural unobservables and the treatment variables.