Statistics and Probability Letters xxx (xxxx) xxx

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Identification of the outcome distribution and sensitivity analysis under weak confounder–instrument interaction

Lu Mao

Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, 610 Walnut St, Room 207A, Madison, WI 53726, USA

ARTICLE INFO

Article history:
Received 18 March 2022
Received in revised form 2 June 2022
Accepted 15 June 2022
Available online xxxx

Keywords:
Causal inference
Efficient influence function
Multiple robustness
Nonignorable noncompliance
Quantile treatment effect

ABSTRACT

Recently, Wang and Tchetgen Tchetgen (2018) showed that the global average treatment effect is identifiable even in the presence of unmeasured confounders so long as they do not modify the instrument's additive effect on the treatment. We use a simple and direct method to show that this no-interaction assumption allows identification of the entire outcome distribution, which leads to multiply robust estimation procedures for nonlinear functionals like the quantile and Mann–Whitney treatment effects. Similarly, we can bound these causal estimands through the outcome distribution in sensitivity analysis against confounder–instrument interaction.

1. Background

Instrumental variables, such as randomized treatment assignment, are indispensable in causal inference with endogenous treatments (Baiocchi et al., 2014). Due to identifiability issues, however, they are mostly used to estimate local treatment effects on the subpopulation of compliers, or those whose treatment varies in lockstep with the instrument (Angrist et al., 1996). The overall treatment effect, while substantively more relevant, is commonly thought of as inaccessible without an intricate model to extrapolate local effects globally (Goldberger, 1972).

Such notion was recently challenged by Wang and Tchetgen Tchetgen (2018), who showed that the global average treatment effect can be identified model-free if the confounders do not modify the instrument's additive effect on the treatment. Because this identifying assumption does not constrain the observed data distribution, the investigator is free to model the latter as they see fit. Given this leeway, the authors went on to construct a multiply robust estimator for the average treatment effect on a binary outcome under three sets of models.

The identifying power of this no-interaction assumption goes beyond the average treatment effect. In a technical report, Tchetgen Tchetgen et al. (2018) established a general identification result for time-varying treatments, confounders, and instruments. This was later used to make inference in marginal structural mean (Michael et al., 2020) and hazard (Wang et al., 2022) models without the traditional sequential randomization assumption. For a point treatment, a technical reader could specialize Lemma 2 of Tchetgen Tchetgen et al. (2018) to find that the entire outcome distributions are identified under the no-interaction assumption. But it is not immediately clear how to estimate them efficiently and robustly. Besides estimation, little is yet known about the quantitative implications of an imperfect no-interaction assumption on the causal estimand of interest.

In this note, we express the outcome distributions in terms of certain identifiable quantities along with a latent one that measures the extent of confounder–instrument interaction. Not only does the expression readily recovers the

E-mail address: lmao@biostat.wisc.edu.

https://doi.org/10.1016/j.spl.2022.109590

Please cite this article as: L. Mao, Identification of the outcome distribution and sensitivity analysis under weak confounder–instrument interaction. Statistics and Probability Letters (2022) 109590, https://doi.org/10.1016/j.spl.2022.109590.

identification result in the absence of interaction, it provides specific instructions on constructing distributional estimators that are multiply robust in the sense of Wang and Tchetgen Tchetgen (2018) and Tchetgen Tchetgen et al. (2018). It also helps to quantify the stochastic range of the distributions in relation to the degree of interaction, which proves useful in sensitivity analysis of estimands that respect the stochastic order of the distributions (Manski, 2003).

2. Main results

Let $Y(a)$ denote the potential outcome under treatment a, where $a = 1$ and 0 indicate the active treatment and control, respectively. With observed treatment A, the observed outcome is $Y = AY(1)+(1-A)Y(0)$. Let X denote a set of observed covariates and U a set of unmeasured confounders. Suppose that $Z \in \{1, 0\}$ is valid instrument that is unconfounded by U and that influences Y only through A. More formally, we make and list below the same assumptions as in Wang and Tchetgen Tchetgen (2018). In particular, (A4) means that all confounders are captured between X and U.

(A1) Exclusion restriction: $Y(z, a) = Y(a)$ almost surely, where $Y(z, a)$ denotes the potential outcome under instrument z and treatment a.
(A2) Independence: $Z \perp\!\!\!\perp U \mid X$.
(A3) Relevance: $Z \not\perp\!\!\!\perp A \mid X$.
(A4) $Y(a) \perp\!\!\!\perp (Z, A) \mid (U, X)$.

Let $\nu_a(\cdot)$ denote the marginal distribution of $Y(a) \in \mathcal{Y}$ $(a = 1, 0)$, where $\mathcal{Y}$ denotes the outcome space. We seek to characterize a generic finite measure μ on $\mathcal{Y}$ through its integrals $\mu(f) = \int f(y)\mu(\mathrm{d}y)$, where $f : \mathcal{Y} \to \mathbb{R}$ is an arbitrary integrable function. Under this notation, let $p_{a,z}(\cdot \mid x)$ denote the conditional measure of Y given $A = a, Z = z$, and $X = x$. Write $\delta_{a,z}(x) = \mathrm{pr}(A = a \mid Z = z, X = x)$. Because $\delta_{a,1}(x) \neq \delta_{a,0}(x)$ by (A3), we can define a finite, possibly signed measure $\nu_a^*(\cdot \mid x)$ by

$$\nu_a^*(\cdot \mid x) = \frac{p_{a,1}(\cdot \mid x)\delta_{a,1}(x) - p_{a,0}(\cdot \mid x)\delta_{a,0}(x)}{\delta_{a,1}(x) - \delta_{a,0}(x)}. \tag{1}$$

Clearly, $\nu_a^*(\cdot \mid x)$ is identifiable as it concerns the observed data (Y, A, Z, X).

Switching to the latent setting, let $\delta_{a,z}(x, U) = \mathrm{pr}(A = a \mid Z = z, X = x, U)$. For shorthand notation, write $\delta_z(x, U) = \delta_{1,z}(x, U)$ and $\delta_z(x) = \delta_{1,z}(x)$. Then $\Delta\delta(x, U) = \delta_1(x, U) - \delta_0(x, U)$ and $\Delta\delta(x) = \delta_1(x) - \delta_0(x)$ can be viewed as the instrument's additive effects on the treatment with and without adjusting for the unknown confounders, respectively. Deviation of $\Delta\delta(x, U)$ from $\Delta\delta(x)$ thus reflects the modification of this effect by the confounders through their interaction with the instrument. In that sense, it is natural to measure the extent of interaction by the relative difference

$$\varepsilon(x, U) = \Delta\delta(x)^{-1}\{\Delta\delta(x; U) - \Delta\delta(x)\}.$$

We call $\varepsilon(x, U)$ the relative interaction function and use it to restate Assumption 5(a) of Wang and Tchetgen Tchetgen (2018).

(A5) There is no confounder–instrument interaction on the treatment, i.e., $\varepsilon(x, U) \equiv 0$ almost surely for every x.

Let $\nu_a(\cdot \mid x)$ denote the conditional measure of $Y(a)$ given $X = x$. The following lemma relates $\nu_a(\cdot \mid x)$ to the identifiable $\nu_a^*(\cdot \mid x)$ up to an error due to a possibly nonzero $\varepsilon(x, U)$. Write $\Delta\delta_a(x) = \delta_{a,1}(x) - \delta_{a,0}(x) = (-1)^{a+1}\Delta\delta(x)$.

Lemma 2.1. *Under Assumptions (A1)–(A4), we have that, for $a = 1, 0$,*

$$\nu_a^*(\cdot \mid x) = \nu_a(\cdot \mid x) + E_{U|x}\{\nu_a(\cdot \mid x, U)\varepsilon(x, U)\}, \tag{2}$$

where $E_{U|x}(\cdot)$ denotes conditional expectation taken over U given $X = x$ and $\nu_a(\cdot \mid x, U)$ is the conditional measure of $Y(a)$ given $X = x$ and U.

Proof. For $z = 1, 0$, $E\{f(Y)I(A = a) \mid Z = z, X = x, U\} = E[f\{Y(a)\}I(A = a) \mid Z = z, X = x, U] = E[f\{Y(a)\} \mid Z = z, X = x, U]\mathrm{pr}(A = a \mid Z = z, X = x, U) = \nu_a(f \mid x, U)\delta_{a,z}(x; U)$, where the first equality follows by (A1), and the second and third by (A4). Thus,

$$\begin{aligned} &E\{f(Y)I(A = a) \mid Z = 1, X = x, U\} - E\{f(Y)I(A = a) \mid Z = 0, X = x, U\} \\ =&\nu_a(f \mid x, U)\{\delta_{a,1}(x; U) - \delta_{a,0}(x; U)\} \\ =&\Delta\delta_a(x)\{\nu_a(f \mid x, U) + \nu_a(f \mid x, U)\varepsilon(x; U)\}. \end{aligned} \tag{3}$$

By (A2), taking $E_{U|x}(\cdot)$ on both sides of (3) yields $E\{f(Y)I(A = a) \mid Z = 1, X = x\} - E\{f(Y)I(A = a) \mid Z = 0, X = x\} = p_{a,1}(f \mid x)\delta_{a,1}(x) - p_{a,0}(f \mid x)\delta_{a,0}(x)$ and $\Delta\delta_a(x)[\nu_a(f \mid x) + E_{U|x}\{\nu_a(f \mid x, U)\varepsilon(x; U)\}]$, respectively. Divide both sides by $\Delta\delta_a(x) \neq 0$ to obtain (2). □

Lemma 2.1 immediately implies identification of $\nu_a(\cdot)$ under (A5).

Proposition 2.1. *Under Assumptions (A1)–(A5), $\nu_a(\cdot \mid x) = \nu_a^*(\cdot \mid x)$ so that $\nu_a(\cdot)$ is identifiable through $\nu_a(\cdot) = E\{\nu_a^*(\cdot \mid X)\}$ $(a = 1, 0)$.*

The identification of $E\{Y(1) - Y(0)\}$ (Wang and Tchetgen Tchetgen, 2018) is implied by inserting the identity function in the $\nu_a(\cdot)$. In full strength, Proposition 2.1 also implies identification of other functionals of the outcome distributions like the quantile and Mann–Whitney treatment effects (Mao, 2018).

3. Multiply robust estimation under no interaction

The first step to estimate the various functionals is to estimate the $\nu_a(\cdot)$ themselves. This can be done in a multiply robust way under the general theory developed in Section 4.2 of Tchetgen Tchetgen et al. (2018) for marginal structural models with time-varying treatments. We provide a specific route by mimicking the construction for the average treatment effect (Wang and Tchetgen Tchetgen, 2018) given $\nu_a^*(\cdot \mid x) = \nu_a(\cdot \mid x)$ under (A5).

With a slight abuse of notation, we use $\nu_a(y)$ to denote $\nu_a\{I(\cdot \le y)\}$, i.e., $\nu_a(y) = \text{pr}\{Y(a) \le y\}$. A similar convention applies to other measures. Let $q(z \mid x) = \text{pr}(Z = z \mid X = x)$ denote the propensity score for the instrument. Consider three classes of models: $\mathcal{M}_1$, consisting of $\nu_a(\cdot \mid x) = \nu_a(\cdot \mid x; \alpha)$, $\Delta\delta(x) = \Delta\delta(x; \beta)$, $\delta_0(x) = \delta_0(x; \zeta)$, and $p_{a,0}(y \mid x) = p_{a,0}(y \mid x; \theta)$; $\mathcal{M}_2$, consisting of $q(z \mid x) = q(z \mid x; \gamma)$ and $\Delta\delta(x) = \Delta\delta(x; \beta)$; and $\mathcal{M}_3$, consisting of $q(z \mid x) = q(z \mid x; \gamma)$ and $\nu_a(\cdot \mid x) = \nu_a(\cdot \mid x; \alpha)$, where $\alpha, \beta, \gamma, \theta$, and ζ are parameters indexing the corresponding distributions. Following Wang and Tchetgen Tchetgen (2018), we first derive the efficient influence function for $\nu_a(y)$ in the union model, with details relegated to the supplementary material.

Theorem 3.1. *Under Assumptions (A1)–(A5), the nonparametric efficient influence function for $\nu_a(y)$ is*

$$EIF_a(y) = \frac{2Z-1}{q(Z \mid X)} \Delta\delta_a(X)^{-1}\big[I(Y \le y, A = a) - I(A = a)\nu_a(y \mid X) - \delta_{a,0}(X)\{p_{a,0}(y \mid X) - \nu_a(y \mid X)\}\big] + \nu_a(y \mid X) - \nu_a(y),$$

which also coincides with the efficient influence function in $\mathcal{M}_1 \cup \mathcal{M}_2 \cup \mathcal{M}_3$.

Given a random n-sample of (Y, A, Z, X), let $\hat\gamma$, $\hat\theta$, and $\hat\zeta$ denote the estimators for γ, θ, and ζ by regressing Z against X, Y against X for those with $A = a$ and $Z = 0$, and A against X for those with $Z = 0$, respectively. With $\mathbb{P}_n$ denoting the empirical measure, let $\hat\beta_{\text{dr}}$ and $\hat\alpha_{\text{dr}}$ solve $\mathbb{P}_n[h(X)(2Z - 1)q(Z \mid X; \hat\gamma)^{-1}\{A - \Delta\delta(X; \hat\beta_{\text{dr}})Z - \delta_0(X; \hat\zeta)\}] = 0$ and $\mathbb{P}_n(g(f; X)(2Z-1)q(Z \mid X; \hat\gamma)^{-1}[I(A = a)\{f(Y) - \nu_a(f \mid X; \hat\alpha_{\text{dr}})\} - \delta_{a,0}(X; \hat\zeta)\{p_{a,0}(f \mid X; \hat\theta) - \nu_a(f \mid X; \hat\alpha_{\text{dr}})\}]) = 0$, respectively, where $h(X)$ and $g(\cdot; X)$ are weight functions commensurate with β and α, respectively. For finite-dimensional α, moment-based weights with $f(y) = y, y^2, \ldots$ will probably suffice. When it has infinite-dimensional components, however, we may need to set $f(\cdot) = I(\cdot \le y)$ for $y \in \mathcal{Y}$ and solve a correspondingly infinite-dimensional estimating equation (van der Vaart and Wellner, 1996, Ch. 3.3).

The initial estimators $\hat\beta_{\text{dr}}$ and $\hat\alpha_{\text{dr}}$ are doubly robust in $\mathcal{M}_1 \cup \mathcal{M}_2$ and $\mathcal{M}_1 \cup \mathcal{M}_3$, respectively. We use them to construct a triply robust estimator for $\nu_a(\cdot)$ in the next proposition (proved in the supplementary material). Write $\Delta\delta_a(x; \beta) = (-1)^{a+1}\Delta\delta(x; \beta)$.

Proposition 3.1. *Under Assumptions (A1)–(A5), the estimator*

$$\hat\nu_{a,\text{mr}}(y) = \mathbb{P}_n\bigg(\frac{2Z-1}{q(Z \mid X; \hat\gamma)} \Delta\delta_a(X; \hat\beta_{\text{dr}})^{-1}\big[I(Y \le y, A = a) - I(A = a)\nu_a(y \mid X; \hat\alpha_{\text{dr}}) - \delta_{a,0}(X; \hat\zeta)\{p_{a,0}(y \mid X; \hat\theta) - \nu_a(y \mid X; \hat\alpha_{\text{dr}})\}\big] + \nu_a(y \mid X; \hat\alpha_{\text{dr}})\bigg)$$

is consistent to $\nu_a(y)$ in $\mathcal{M}_1 \cup \mathcal{M}_2 \cup \mathcal{M}_3$ and locally efficient in $\mathcal{M}_1 \cap \mathcal{M}_2 \cap \mathcal{M}_3$.

Now by plugging in the $\hat\nu_{a,\text{mr}}(\cdot)$, we can estimate any treatment effect in the form of a contrast $\mathcal{T}(\nu_1, \nu_0)$. This applies to the average treatment effect with $\mathcal{T}(\nu_1, \nu_0) = \int y\{\nu_1(\text{d}y) - \nu_0(\text{d}y)\}$, the τ-quantile treatment effect with $\mathcal{T}(\nu_1, \nu_0) = \nu_1^{-1}(\tau) - \nu_0^{-1}(\tau)$ $(0 < \tau < 1)$, and the Mann–Whitney stochastic shift $\mathcal{T}(\nu_1, \nu_0) = \int \nu_0(y)\nu_1(\text{d}y) - \int \nu_1(y)\nu_0(\text{d}y)$ (Mao, 2018), the last of which measures the net probability of a greater outcome from the treatment as compared to the control. For these smooth $\mathcal{T}$, $\mathcal{T}(\hat\nu_{1,\text{mr}}, \hat\nu_{0,\text{mr}})$ inherits the robustness and efficiency of the $\hat\nu_{a,\text{mr}}(y)$ established in Proposition 3.1 (Bickel et al., 1993).

4. Sensitivity analysis against interaction

All such plug-in estimators rely on (A5) for validity. To assess their sensitivity to the presence of interaction, consider a relaxed form of (A5).

(A5*) The relative interaction function satisfies

$$-\xi^-(x) \le \varepsilon(x,U) \le \xi^+(x) \tag{4}$$

almost surely with some $0 \le \xi^-(x) \le |\Delta\delta(x)|^{-1} + 1$ and $0 \le \xi^+(x) \le |\Delta\delta(x)|^{-1} - 1$.

Remark 4.1. The constraints on $\xi^-(x)$ and $\xi^+(x)$ arise from the natural bounds $-|\Delta\delta(x)|^{-1} - 1 \le \varepsilon(x,U) \le |\Delta\delta(x)|^{-1} - 1$ due to $|\Delta\delta(x,U)| \le 1$, which form the identifiable region for $\varepsilon(x,U)$. As a result, tighter bounds are untestable because they have no empirical implications for the data at hand. If circumstances allow, it would be ideal to estimate $\xi^-(x)$ and $\xi^+(x)$ from a previous study where confounders are more extensively measured. Without such external data, one can use the natural bounds themselves, which, as will be seen later, may still bound the treatment effect meaningfully provided that the instrument-treatment association is sufficiently strong.

Similarly to estimation under (A5), we can bound a suitable treatment effect under (A5*) by bounding the distribution functions $\nu_a(\cdot)$ via the $\nu_a(\cdot \mid x)$. First, consider bounding $\nu_a(B \mid x)$ for some measurable $B \subset \mathcal{Y}$, where $\mu(B) = \mu\{I(\cdot \in B)\}$. By (2) of Lemma 2.1, the error between $\nu_a(B \mid x)$ and the identifiable $\nu_a^*(B \mid x)$ is $-E_{U|x}\{\nu_a(B \mid x, U)\varepsilon(x,U)\}$, which by $0 \le \nu_a(B \mid x, U) \le 1$ is bracketed between $-\xi^+(x)$ and $\xi^-(x)$ under (4), leading to

$$\nu_a^*(B \mid x) - \xi^+(x) \le \nu_a(B \mid x) \le \nu_a^*(B \mid x) + \xi^-(x). \tag{5}$$

These bounds can be improved since they have not used the fact that $E_{U|x}\{\varepsilon(x,U)\} = 0$ or that $E_{U|x}\{\nu_a(B \mid x, U)\} = \nu_a(B \mid x)$.

The following lemma tightens the bounds by exploiting these constraints. Write $\xi(x) = \{\xi^-(x), \xi^+(x)\}$.

Lemma 4.1. *Under Assumptions (A1)–(A4) and (A5*), we have that, for any measurable $B \subset \mathcal{Y}$,*

$$-\mathcal{E}^{\mathrm{L}}\{\nu_a(B \mid x); \xi(x)\} \le E_{U|x}\{\nu_a(B \mid x, U)\varepsilon(x,U)\} \le \mathcal{E}^{\mathrm{U}}\{\nu_a(B \mid x); \xi(x)\}, \tag{6}$$

where

$$\mathcal{E}^{\mathrm{L}}(\pi;\xi) = \begin{cases} \xi^-\pi, & 0 \le \pi \le \frac{\xi^+}{\xi^- + \xi^+} \\ \xi^+(1-\pi), & \frac{\xi^+}{\xi^- + \xi^+} < \pi \le 1 \end{cases}$$

and $\mathcal{E}^{\mathrm{U}}(\pi;\xi) = \begin{cases} \xi^+\pi, & 0 \le \pi \le \frac{\xi^-}{\xi^- + \xi^+} \\ \xi^-(1-\pi), & \frac{\xi^-}{\xi^- + \xi^+} < \pi \le 1 \end{cases}$

for $\pi \in [0,1]$ and $\xi = (\xi^-, \xi^+) \in \mathbb{R}^{+\otimes 2}$.

Proof. Obviously, $-\xi^-(x)\nu_a(B \mid x) \le E_{U|x}\{\nu_a(B \mid x, U)\varepsilon(x,U)\} \le \xi^+(x)\nu_a(B \mid x)$. Because $E_{U|x}\{\varepsilon(x,U)\} = 0$, we have that $-E_{U|x}\{\nu_a(B \mid x, U)\varepsilon(x,U)\} = E_{U|x}[\{1 - \nu_a(B \mid x, U)\}\varepsilon(x,U)]$, which is similarly bounded between $-\xi^-(x)\{1 - \nu_a(B \mid x)\}$ and $\xi^+(x)\{1 - \nu_a(B \mid x)\}$. Composing the two sets of bounds produces (6). □

The bounds in (6) are superior to those in (5) as both $\mathcal{E}^{\mathrm{L}}\{\nu_a(B \mid x); \xi(x)\}$ and $\mathcal{E}^{\mathrm{U}}\{\nu_a(B \mid x); \xi(x)\}$ are always less than or equal to $\min\{\xi^-(x), \xi^+(x)\}$. What is more, they duly shrink to zero when $\nu_a(B \mid x)$ approaches 1 or 0, or when $|\Delta\delta(x)|$ approaches 1. To see the latter, recall that $\xi^+(x) \le |\Delta\delta(x)|^{-1} - 1$ by the natural upper bound on $\varepsilon(x,U)$. Therefore, $|\Delta\delta(x)| \uparrow 1$ implies that $\xi^+(x) \downarrow 0$, which in turn pushes both $\mathcal{E}^{\mathrm{L}}\{\nu_a(B \mid x); \xi(x)\}$ and $\mathcal{E}^{\mathrm{U}}\{\nu_a(B \mid x); \xi(x)\}$ down to 0. This is unsurprising since $|\Delta\delta(x)| \approx 1$ means strong concordance between the instrument and treatment. With the unconfounded instrument as a good proxy for the treatment, the outcome distributions are naturally near identifiable.

Now we can combine (6) with (2) to obtain bounds of $\nu_a(B \mid x)$ as piecewise linear functions of $\nu_a^*(B \mid x)$.

Proposition 4.1. *Under Assumptions (A1)–(A4) and (A5*), we have that, for any measurable $B \subset \mathcal{Y}$,*

$$\mathcal{L}\{\nu_a^*(B \mid x); \xi(x)\} \le \nu_a(B \mid x) \le \mathcal{H}\{\nu_a(B \mid x); \xi(x)\}, \tag{7}$$

where $\mathcal{L}(\pi^*;\xi)$ and $\mathcal{H}(\pi^*;\xi)$ are piecewise linear functions of π^* for any $\xi = (\xi^-, \xi^+) \in \mathbb{R}^{+\otimes 2}$ and are expressed in the supplementary material.

In particular, we can bound the cumulative distribution function $\nu_a(y \mid x)$ by taking $B = \{y' \in \mathcal{Y} : y' \le y\}$, yielding

$$\mathcal{L}\{\nu_a^*(y \mid x); \xi(x)\} \le \nu_a(y \mid x) \le \mathcal{H}\{\nu_a^*(y \mid x); \xi(x)\}. \tag{8}$$

According to the expressions of $\mathcal{L}(\pi^*;\xi)$ and $\mathcal{H}(\pi^*;\xi)$ in the supplementary material, an inflection point is found at $\xi^-(x) = 1$. To illustrate, consider the case with $\delta_1(x) = 0.9$, $\delta_0(x) = 0.1$, and set $\xi^+(x)$ at the natural bound $|\Delta\delta(x)|^{-1} - 1 = 0.25$. For $\xi^-(x) \in \{|\Delta\delta(x)|^{-1} + 1 = 2.25, 1, 0.5\}$, we plot the range of $\nu_a(y \mid x)$ as a function of $\nu_a^*(y \mid x)$ in Fig. 1. Because $\xi^-(x) = 1$ implies that $\delta_1(x,U) \ge \delta_0(x,U)$ almost surely, it is roughly equivalent to a stochastic version of the monotonicity condition under a causal instrument (Angrist et al., 1996). In such cases, a fraction of the population, namely, the noncompliers, are completely unidentified (Imbens and Rubin, 1997), which explains the truncation of the

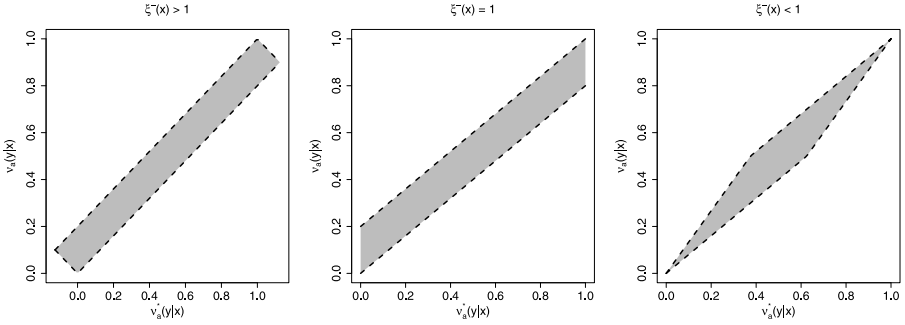

Fig. 1. Range of $\nu_a(y \mid x)$ as a function of $\nu_a^*(y \mid x)$ under Assumption (A5*) based on an instance with $\delta_1(x) = 0.9$ and $\delta_0(x) = 0.1$.

distribution function in the middle panel of Fig. 1. Interestingly, such truncation disappears when $\xi^-(x)$ is strictly less than 1. More discussions in this regard are provided in the supplementary material.

In general, we can always find a least and a greatest element for $\nu_a(\cdot \mid x)$ in terms of stochastic order, which help bound certain treatment effects of interest.

Proposition 4.2. *For all $\nu_a(\cdot \mid x)$ satisfying (8), a stochastically least and a greatest element exist, denoted by $\underline{\nu}_a(\cdot \mid x)$ and $\overline{\nu}_a(\cdot \mid x)$, respectively. Then the stochastically least and greatest $\nu_a(\cdot)$ are $\underline{\nu}_a(\cdot) = E\{\underline{\nu}_a(\cdot \mid X)\}$ and $\overline{\nu}_a(\cdot) = E\{\overline{\nu}_a(\cdot \mid X)\}$, respectively. Moreover, if the treatment effect is defined by $\theta = \mathcal{T}(\nu_1, \nu_0)$ with $\mathcal{T}(\nu_1, \nu_0)$ nondecreasing in ν_1 and nonincreasing in ν_0 with respect to the stochastic order, then $\mathcal{T}(\underline{\nu}_1, \overline{\nu}_0) \le \theta \le \mathcal{T}(\overline{\nu}_1, \underline{\nu}_0)$.*

The existence of $\underline{\nu}_a(\cdot \mid x)$ and $\overline{\nu}_a(\cdot \mid x)$ is ensured by the closure of the space of distribution functions satisfying (8) under pointwise maximum and minimum. To estimate the bounds from empirical data, we first estimate $\nu_a^*(\cdot \mid x)$, e.g., using the procedures described in Section 3, and the compute $\underline{\nu}_a(y \mid x)$ and $\overline{\nu}_a(y \mid x)$ by applying isotonic methods to find the greatest and smallest monotone functions below $\mathcal{H}\{\nu_a^*(y \mid x); \xi(x)\}$ and above $\mathcal{L}\{\nu_a^*(y \mid x); \xi(x)\}$, respectively (see, e.g., Groeneboom and Jongbloed, 2014). Finally, taking the empirical averages of $\underline{\nu}_a(\cdot \mid X)$ and $\overline{\nu}_a(\cdot \mid X)$ over X gives us estimates of $\underline{\nu}_a(\cdot)$ and $\overline{\nu}_a(\cdot)$, respectively. These estimates can then be used to bound the quantile treatment effect, Mann–Whitney stochastic shift, and average treatment effect, all respecting the stochastic order as required in Proposition 4.2 (Manski, 2003). An illustration using the U.S. National Job Training Partnership Act (JTPA) Study (Abadie et al., 2002) is provided in the supplementary material.

5. Concluding remarks

There are several ways to extend the sensitivity analysis in Section 4. Because bounds are developed on the $\nu_a(\cdot \mid x)$, it should be straightforward to derive similar results under marginal structural models, with point or even time-varying treatments (Tchetgen Tchetgen et al., 2018). Moreover, in selecting optimal treatment regimes, one typically does not need the full range of the effect size. For example, Cui and Tchetgen Tchetgen (2021) provided a necessary and sufficient condition to identify the sign of the average treatment effect in complete absence of (A5). The distributional bounds developed under (A5*) can shed new light on the sign-identifiability of the average and various other treatment effects.

It may also be possible to improve the bounds on the treatment effects themselves. For instance, those in Proposition 4.2 stem from separate bounds on the outcome distributions under the treatment and control. As is clear from (2), however, $\nu_1(\cdot \mid x)$ and $\nu_0(\cdot \mid x)$ are subject to the same influence of $\varepsilon(x, U)$. Simultaneous bounding of the outcome distributions may further narrow their contrast. Certain monotonicity assumptions on the confounders, if justifiable, may also help, as is shown in Appendix B of Wang and Tchetgen Tchetgen (2018) for the average treatment effect. For starters, a distributional version of that result is derived in the supplementary material.

Acknowledgments

I thank the editor and an anonymous referee for helpful comments. This research was supported by the U.S. National Science Foundation grant DMS2015526 and National Institutes of Health grant R01HL149875.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.spl.2022.109590. Supplementary material online includes technical and numerical results.

References

Abadie, A., Angrist, J., Imbens, G., 2002. Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. Econometrica 70 (1), 91–117.
Angrist, J.D., Imbens, G.W., Rubin, D.B., 1996. Identification of causal effects using instrumental variables. J. Amer. Statist. Assoc. 91 (434), 444–455.
Baiocchi, M., Cheng, J., Small, D.S., 2014. Instrumental variable methods for causal inference. Stat. Med. 33 (13), 2297–2340.
Bickel, P.J., Klaassen, C.A., Bickel, P.J., Ritov, Y., Klaassen, J., Wellner, J.A., Ritov, Y., 1993. Efficient and Adaptive Estimation for Semiparametric Models. Vol. 4. Johns Hopkins University Press, Baltimore.
Cui, Y., Tchetgen Tchetgen, E., 2021. On a necessary and sufficient identification condition of optimal treatment regimes with an instrumental variable. Statist. Probab. Lett. 178, 109180.
Goldberger, A.S., 1972. Structural equation methods in the social sciences. Econometrica 40, 979–1001.
Groeneboom, P., Jongbloed, G., 2014. Nonparametric Estimation under Shape Constraints. Cambridge University Press, Cambridge.
Imbens, G.W., Rubin, D.B., 1997. Estimating outcome distributions for compliers in instrumental variables models. Rev. Econom. Stud. 64 (4), 555–574.
Manski, C.F., 2003. Partial Identification of Probability Distributions. Springer, New York.
Mao, L., 2018. On causal estimation using *U*-statistics. Biometrika 105 (1), 215–220.
Michael, H., Cui, Y., Lorch, S., Tchetgen Tchetgen, E., 2020. Instrumental variable estimation of marginal structural mean models for time-varying treatment. arXiv preprint, arXiv:2004.11769.
Tchetgen Tchetgen, E., Michael, H., Cui, Y., 2018. Marginal structural models for time-varying endogenous treatments: A time-varying instrumental variable approach. arXiv preprint, arXiv:1809.05422.
van der Vaart, A., Wellner, J., 1996. Weak Convergence and Empirical Processes: With Applications to Statistics. Springer, New York.
Wang, L., Tchetgen Tchetgen, E., 2018. Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables. J. R. Stat. Soc. Ser. B Stat. Methodol. 80 (3), 531–550.
Wang, L., Tchetgen Tchetgen, E., Martinussen, T., Vansteelandt, S., 2022. Learning causal hazard ratio with endogeneity. arXiv preprint, arXiv:1807.05313.