Quantitative Economics 11 (2020), 41–111 1759-7331/20200041

Inference on breakdown frontiers

Matthew A. Masten
Department of Economics, Duke University

Alexandre Poirier
Department of Economics, Georgetown University

Given a set of baseline assumptions, a breakdown frontier is the boundary between the set of assumptions which lead to a specific conclusion and those which do not. In a potential outcomes model with a binary treatment, we consider two conclusions: First, that ATE is at least a specific value (e.g., nonnegative) and second that the proportion of units who benefit from treatment is at least a specific value (e.g., at least 50%). For these conclusions, we derive the breakdown frontier for two kinds of assumptions: one which indexes relaxations of the baseline random assignment of treatment assumption, and one which indexes relaxations of the baseline rank invariance assumption. These classes of assumptions nest both the point identifying assumptions of random assignment and rank invariance and the opposite end of no constraints on treatment selection or the dependence structure between potential outcomes. This frontier provides a quantitative measure of the robustness of conclusions to relaxations of the baseline point identifying assumptions. We derive $\sqrt{N}$-consistent sample analog estimators for these frontiers. We then provide two asymptotically valid bootstrap procedures for constructing lower uniform confidence bands for the breakdown frontier. As a measure of robustness, estimated breakdown frontiers and their corresponding confidence bands can be presented alongside traditional point estimates and confidence intervals obtained under point identifying assumptions. We illustrate this approach in an empirical application to the effect of child soldiering on wages. We find that sufficiently weak conclusions are robust to simultaneous failures of rank invariance and random assignment, while some stronger conclusions are fairly robust to failures of rank invariance but not necessarily to relaxations of random assignment.

Matthew A. Masten: matt.masten@duke.edu
Alexandre Poirier: alexandre.poirier@georgetown.edu
This paper was presented at the 2017 IAAE Conference, the 2017 CEME Conference, the 2017 U Chicago Interactions Conference, the 2017 Midwest Econometrics Group, the 2017 SEA Meetings, the 2018 Winter Meeting of the Econometric Society, the 2018 NU Econometrics Alumni Conference, Duke, UC Irvine, U Penn, UC Berkeley, Yale, UGA, Auburn, UC Boulder, Georgetown, UNC-Chapel Hill, Rice, Boston University, UCL, University of Cambridge, and the LSE. We thank audiences at those conferences and seminars, the referees, as well as Federico Bugni, Ivan Canay, Joachim Freyberger, Guido Imbens, Ying-Ying Lee, Chuck Manski, Anna Mikusheva, Jim Powell, Pedro Sant'Anna, and Andres Santos for helpful conversations and comments. We thank Margaux Luflade for excellent research assistance. This paper uses data from phase 1 of SWAY, the Survey of War Affected Youth in northern Uganda. We thank the SWAY principal researchers Jeannie Annan and Chris Blattman for making their data publicly available and for answering our questions.

Available at http://qeconomics.org. https://doi.org/10.3982/QE1288

Keywords. Nonparametric identification, partial identification, sensitivity analysis, selection on unobservables, rank invariance, treatment effects, directional differentiability.

JEL classification. C14, C18, C21, C25, C51.

1. Introduction

Traditional empirical analysis combines the observed data with a set of assumptions to draw conclusions about a parameter of interest. *Breakdown frontier* analysis reverses this ordering. It begins with a fixed conclusion and a set of baseline assumptions and asks, "What are the weakest assumptions needed to draw that conclusion, given the observed data?" For example, consider the impact of a binary treatment on some outcome variable. The traditional approach might assume random assignment, point identify the average treatment effect (ATE), and then report the obtained value. The breakdown frontier approach instead begins with a conclusion about ATE, like "ATE is positive," and reports the weakest assumption—relative to random assignment—on the relationship between treatment assignment and potential outcomes needed to obtain this conclusion, when such an assumption exists. When more than one kind of assumption is considered, this approach leads to a curve, representing the weakest combinations of assumptions which lead to the desired conclusion. This curve is the breakdown frontier.

At the population level, the difference between the traditional approach and the breakdown frontier approach is a matter of perspective: an answer to one question is an answer to the other. This relationship has long been present in the literature initiated by Manski on partial identification (e.g., see Manski (2007) or Section 3 of Manski (2013)). In finite samples, however, which approach one chooses has important implications for how one does statistical inference. Specifically, the traditional approach estimates the parameter or its identified set. Here we instead estimate the breakdown frontier. The traditional approach then performs inference on the parameter or its identified set. Here we instead perform inference on the breakdown frontier. Thus the breakdown frontier approach puts the weakest assumptions necessary to draw a conclusion at the center of attention. Consequently, by construction, this approach avoids the nontight bounds critique of partial identification methods (e.g., see Section 7.2 of Ho and Rosen (2017)). One distinction is that the traditional approach may require inference on a partially identified parameter. The breakdown frontier approach, however, only requires inference on a point identified object.

The breakdown frontier we study generalizes the concept of an "identification breakdown point" introduced by Horowitz and Manski (1995), a one-dimensional breakdown frontier.[1] Their breakdown point was further studied and generalized by Stoye (2005, 2010). Our emphasis on inference on the breakdown frontier follows Kline and Santos

[1]The *identification* breakdown point is distinct from the breakdown point introduced earlier by Hampel (1968, 1971) in the robust statistics literature that began with Huber (1964); also see Donoho and Huber (1983). Horowitz and Manski (1995) gave a detailed comparison of the two concepts. Throughout this paper we use the term "breakdown" in the same sense as Horowitz and Manski's identification breakdown point.

(2013), who proposed doing inference on a breakdown point. Finally, our focus on multidimensional frontiers builds on the graphical sensitivity analysis of Imbens (2003) and the multidimensional sensitivity analysis of Manski and Pepper (2018). We discuss these papers and others in detail in Appendix A.

The breakdown frontier approach

The breakdown frontier approach requires six main steps: (a) specify a parameter of interest, (b) specify a set of baseline assumptions, (c) define a class of assumptions indexed by a sensitivity parameter which deliver a nested sequence of identified sets, with the baseline assumptions obtained at one extreme and the no assumptions bounds obtained at the other, (d) characterize identified sets for the parameter of interest as a function of the sensitivity parameter, (e) use those identified sets to define the breakdown frontier for a conclusion of interest, and (f) develop estimation and inference procedures for that frontier based on its characterization.

In principle this analysis can be done for a general class of models, for example, by using the general identification analysis in Chesher and Rosen (2017) or Torgovitsky (2019) for step (d), and then applying general tools for nonparametric estimation and inference for step (f). While such a general analysis is an important next step for future work, in this paper we focus on just one important and widely used model: the potential outcomes model with a binary treatment. By focusing on a single concrete model, we can clearly illustrate how to do the six main steps (a)–(f) required for a breakdown frontier analysis in any model. While the mathematical analysis will differ from model to model, the general ideas and approach do not.

In the rest of this section and this paper, we use the binary treatment potential outcomes model to explain and illustrate the breakdown frontier approach. Our main parameter of interest is the proportion of units who benefit from treatment. Under random assignment of treatment and rank invariance, this parameter is point identified. One may be concerned, however, that these two assumptions are too strong. We relax rank invariance by supposing that there are two types of units in the population: one type for which rank invariance holds and another type for which it may not. The proportion t of the second type measures the relaxation of rank invariance. We relax random assignment using a propensity score distance $c \geq 0$ as in our previous work, Masten and Poirier (2018a). We give more details on both of these relaxations in Section 2. We derive the identified set for $\mathbb{P}(Y_1 > Y_0)$ as a function of (c, t). For a specific conclusion, such as $\mathbb{P}(Y_1 > Y_0) \geq 0.5$, this identification result defines a breakdown frontier.

Figure 1 illustrates this breakdown frontier. The horizontal axis measures c, the relaxation of the random assignment assumption. The vertical axis measures t, the relaxation of rank invariance. The origin represents the baseline point identifying assumptions of random assignment and rank invariance. Points along the vertical axis represent random assignment paired with various relaxations of rank invariance. Points along the horizontal axis represent rank invariance paired with various relaxations of random as-

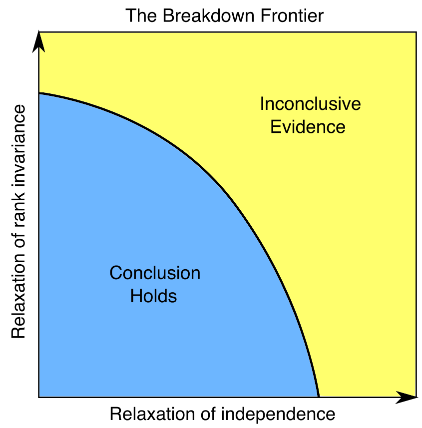

FIGURE 1. An example breakdown frontier, partitioning the space of assumptions into the set for which our conclusion of interest holds (the robust region) and the set for which our evidence is inconclusive.

signment. Points in the interior of the box represent relaxations of both assumptions. The points in the lower left region are pairs of assumptions (c, t) such that the data allow us to draw our desired conclusion: $\mathbb{P}(Y_1 > Y_0) \geq 0.5$. We call this set the *robust region*. Specifically, no matter what value of (c, t) we choose in this region, the identified set for $\mathbb{P}(Y_1 > Y_0)$ always lies completely above 0.5. The points in the upper right region are pairs of assumptions that do not allow us to draw this conclusion. For these pairs (c, t), the identified set for $\mathbb{P}(Y_1 > Y_0)$ contains elements smaller than 0.5. The boundary between these two regions is precisely the breakdown frontier. The area under the breakdown frontier—the robust region—is a quantitative measure of robustness.

The breakdown frontier is similar to the classical Pareto frontier from microeconomic theory: It shows the trade-offs between different assumptions in drawing a specific conclusion. For example, consider Figure 1. If we are at the top left point, where the breakdown frontier intersects the vertical axis, then any relaxation of random assignment requires strengthening the rank invariance assumption in order to still be able to draw our desired conclusion. The breakdown frontier specifies the precise marginal rate of substitution between the two assumptions.

Figure 2 illustrates how the breakdown frontier changes as our conclusion of interest changes. Specifically, consider the conclusion that

$$\mathbb{P}(Y_1 > Y_0) \geq \underline{p}$$

for five different values for $\underline{p}$. The figure shows the corresponding breakdown frontiers. As $\underline{p}$ increases toward one, we are making a stronger claim about the true parameter, and hence the set of assumptions for which the conclusion holds shrinks. For strong enough claims, the claim may be refuted even with the strongest assumptions possible. Conversely, as $\underline{p}$ decreases toward zero, we are making progressively weaker claims about the true parameter, and hence the set of assumptions for which the conclusion holds grows larger.

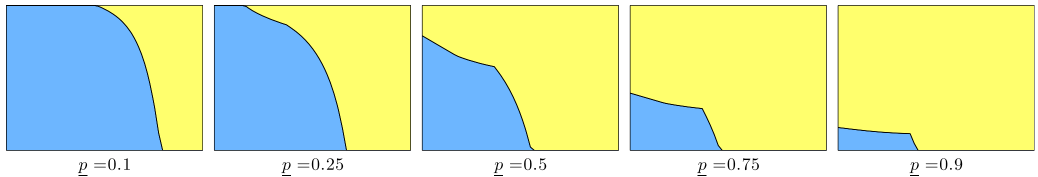

FIGURE 2. Example breakdown frontiers for the claim $\mathbb{P}(Y_1 > Y_0) \geq \underline{p}$, for five different values of $\underline{p}$.

Under the strongest assumptions of $(c, t) = (0, 0)$, the parameter $\mathbb{P}(Y_1 > Y_0)$ is point identified. Let $p_{0,0}$ denote its value. The value $p_{0,0}$ is often strictly less than 1. In this case, any $\underline{p} \in (p_{0,0}, 1]$ yields a degenerate breakdown frontier: This conclusion is refuted under the point identifying assumptions. Even if $p_{0,0} < \underline{p}$, the conclusion $\mathbb{P}(Y_1 > Y_0) \geq \underline{p}$ may still be correct. This follows since, for strictly positive values of c and t, the identified sets for $\mathbb{P}(Y_1 > Y_0)$ do contain values larger than $p_{0,0}$. But they also contain values smaller than $p_{0,0}$. Hence there do not exist any assumptions for which we can draw the desired conclusion.

We provide additional interpretation of population breakdown frontiers in Section 2, Section 4, and Appendix E in the Online Supplemental Material (Masten and Poirier (2020)). We discuss estimation and inference in Section 3 and Appendix D in the Online Supplemental Material. Finally, we illustrate how our results can be used as a sensitivity analysis within a larger empirical study in Section 4. In that section we also provide a detailed and largely self-contained guide to implementing our approach.

2. Model and identification

In this section we study identification of the standard potential outcomes model with a binary treatment. We focus on two key assumptions: random assignment and rank invariance. We discuss how to relax these two assumptions and derive identified sets for various parameters under these relaxations. While there are many different ways to relax these assumptions, our goal is to illustrate the breakdown frontier methodology, and hence we focus on just one kind of relaxation for each assumption.

Setup

Let Y_1 and Y_0 denote the unobserved potential outcomes. Let $X \in \{0, 1\}$ be an observed binary treatment. Let $W \in \operatorname{supp}(W)$ be a vector of observed covariates. This vector may contain discrete covariates, continuous covariates, or both. We observe the scalar outcome variable

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

Let $p_{x|w} = \mathbb{P}(X = x \mid W = w)$ denote the observed propensity score. We maintain the following assumption on the joint distribution of (Y_1, Y_0, X, W) throughout.

Assumption A1. *For each* $x, x' \in \{0, 1\}$ *and* $w \in \operatorname{supp}(W)$:

1. $Y_x \mid X = x', W = w$ *has a strictly increasing and continuous distribution function on its support,* $\operatorname{supp}(Y_x \mid X = x', W = w)$.

2. $\operatorname{supp}(Y_x \mid X = x', W = w) = \operatorname{supp}(Y_x \mid W = w) = [\underline{y}_x(w), \overline{y}_x(w)]$ *where* $-\infty \leq \underline{y}_x(w) < \overline{y}_x(w) \leq \infty$.

3. $p_{x|w} > 0$.

Via Assumption A1.1, we restrict attention to continuously distributed potential outcomes. Assumption A1.2 states that the support of $Y_x \mid X = x', W = w$ does not depend on x', and is a possibly infinite closed interval. This assumption implies that the endpoints $\underline{y}_x(w)$ and $\overline{y}_x(w)$ are point identified. We maintain Assumption A1.2 for simplicity, but it can be relaxed using similar derivations as in Masten and Poirier (2016). Assumption A1.3 is an overlap assumption.

Define the *conditional rank* random variables $U_0 = F_{Y_0|W}(Y_0 \mid W)$ and $U_1 = F_{Y_1|W}(Y_1 \mid W)$. Since $F_{Y_1|W}(\cdot \mid w)$ and $F_{Y_0|W}(\cdot \mid w)$ are strictly increasing (by Assumption A1.1), $U_0 \mid W$ and $U_1 \mid W$ are uniformly distributed on $[0, 1]$. The value of unit i's conditional rank random variable U_x tells us where unit i lies in the distribution of $Y_x \mid W$.

Identifying assumptions

It is well known that the joint conditional distribution of potential outcomes $(Y_1, Y_0) \mid W$ is point identified under two assumptions:

1. Conditional random assignment of treatment: $X \perp\!\!\!\perp Y_1 \mid W$ and $X \perp\!\!\!\perp Y_0 \mid W$.
2. Conditional rank invariance: $U_1 = U_0$ almost surely.

Note that the joint conditional independence assumption $X \perp\!\!\!\perp (Y_1, Y_0) \mid W$ provides no additional identifying power beyond the marginal conditional independence assumption stated above. Any functional of $F_{Y_1,Y_0|W}$ is point identified under these random assignment and rank invariance assumptions. The goal of our identification analysis is to study what can be said about such functionals when one or both of these point identifying assumptions fails. To do this, we define two classes of assumptions: one which indexes the relaxation of random assignment of treatment, and one which indexes the relaxation of rank invariance. These classes of assumptions nest both the point identifying assumptions of random assignment and rank invariance and the opposite end of no constraints on treatment selection or the dependence structure between potential outcomes.

A key feature of the relaxations we use is that they are "orthogonal" in the sense that we can relax each of the two assumptions separately: The amount by which we relax one assumption does not constrain the amount by which we can relax the other assumption. This feature is important since a key goal of our analysis is to quantify the trade-off between relaxations of these two assumptions.

We begin with our measure of distance from conditional independence.

Definition 1. Let c be a scalar between 0 and 1. Say X is *conditionally c-dependent* with Y_x given W if

$$\sup_{y \in \operatorname{supp}(Y_x \mid W=w)} \left| \mathbb{P}(X = x \mid Y_x = y, W = w) - \mathbb{P}(X = x \mid W = w) \right| \leq c \tag{2}$$

for $x \in \{0, 1\}$ and $w \in \operatorname{supp}(W)$.

For $c = 0$, conditional c-dependence implies $X \perp\!\!\!\perp Y_1 \mid W$ and $X \perp\!\!\!\perp Y_0 \mid W$. For $c > 0$, however, it allows for some deviations from conditional independence. Specifically, it allows the unobserved treatment assignment probability $\mathbb{P}(X = 1 \mid Y_x = y, W = w)$ to be at most c probability units away from the observed propensity score $p_{1|w}$. We discuss one way to interpret the magnitude of c on page 68. We give further discussion in our previous paper, Masten and Poirier (2018a).

Our second class of assumptions constrains the dependence structure between Y_1 and Y_0, conditional on W. By Sklar's theorem (Sklar (1959)), write

$$F_{Y_1, Y_0 \mid W}(y_1, y_0 \mid w) = C\bigl(F_{Y_1 \mid W}(y_1 \mid w), F_{Y_0 \mid W}(y_0 \mid w) \mid w\bigr),$$

where $C(\cdot, \cdot \mid w)$ is a unique conditional copula function. See Nelsen (2006) for an overview of copulas and Fan and Patton (2014) for a survey of their use in econometrics. Restrictions on C constrain the dependence between potential outcomes. For example, if

$$C(u_1, u_0 \mid w) = \min\{u_1, u_0\},$$

then $U_1 = U_0$ almost surely. Thus conditional rank invariance holds. In this case the potential outcomes Y_1 and Y_0 are sometimes called *conditionally comonotonic* and $\min\{\cdot, \cdot\}$ is called the *comonotonicity* copula. At the opposite extreme, when C is an arbitrary copula, the dependence between $Y_1 \mid W$ and $Y_0 \mid W$ is constrained only by the Fréchet–Hoeffding bounds, which state that

$$\max\{u_1 + u_0 - 1, 0\} \leq C(u_1, u_0 \mid w) \leq \min\{u_1, u_0\}.$$

We next define a class of assumptions which includes both conditional rank invariance and no assumptions on the dependence structure as special cases.

Definition 2. The potential outcomes (Y_1, Y_0) satisfy *$(1-t)$-percent conditional rank invariance* given W if for all $w \in \operatorname{supp}(W)$ their conditional copula C satisfies

$$C(u_1, u_0 \mid w) = (1 - t)\min\{u_1, u_0\} + tH(u_1, u_0 \mid w), \tag{3}$$

where H is some conditional copula.

This assumption says that within each covariate cell the population is a mixture of two parts: In one part, rank invariance holds. This part contains $100(1 - t)\%$ of the overall population in that cell. In the second part, rank invariance may fail in an arbitrary, unknown way. Hence, for this part, the dependence structure is unconstrained beyond the Fréchet–Hoeffding bounds. This part contains $100 \cdot t\%$ of the overall population in that cell. Thus for $t = 0$ the usual conditional rank invariance assumption holds, while for $t = 1$ no assumptions are made about the dependence structure. For $t \in (0, 1)$, we obtain a kind of partial conditional rank invariance. Note that by exercise 2.3 on page 14 of Nelsen (2006), a mixture of copulas like that in equation (3) is also a copula.

To see this mixture interpretation formally, let T follow a Bernoulli distribution with $\mathbb{P}(T = 1 \mid W = w) = t$, where $T \perp\!\!\!\perp Y_1 \mid W$ and $T \perp\!\!\!\perp Y_0 \mid W$, but T is not independent of $(Y_1, Y_0) \mid W$ jointly. This implies that T has an effect on the dependence structure of $(Y_1, Y_0) \mid W$ but not on their conditional marginal distributions. Suppose that individuals for whom $T_i = 1$ have an arbitrary dependence structure, while those with $T_i = 0$ have conditionally rank invariant potential outcomes. Then by the law of total probability,

$$\begin{aligned}
&F_{Y_1,Y_0|W}(y_1, y_0 \mid w) \\
&\quad = (1 - t)F_{Y_1,Y_0|W,T}(y_1, y_0 \mid w, 0) + tF_{Y_1,Y_0|W,T}(y_1, y_0 \mid w, 1) \\
&\quad = (1 - t)C\big(F_{Y_1|W,T}(y_1 \mid w, 0), F_{Y_0|W,T}(y_0 \mid w, 0) \mid w, 0\big) \\
&\qquad + tC\big(F_{Y_1|W,T}(y_1 \mid w, 1), F_{Y_0|W,T}(y_0 \mid w, 1) \mid w, 1\big) \\
&\quad = (1 - t) \min\big\{F_{Y_1|W}(y_1 \mid w), F_{Y_0|W}(y_0 \mid w)\big\} + tH\big(F_{Y_1|W}(y_1 \mid w), F_{Y_0|W}(y_0 \mid w) \mid w\big).
\end{aligned}$$

Our approach to relaxing rank invariance is an example of a more general approach. In this approach we take a weak assumption and a stronger assumption and use them to define a continuous class of assumptions by considering the population as a mixture of two subpopulations. The weak assumption holds in one subpopulation while the stronger assumption holds in the other subpopulation. The mixing proportion t continuously spans the two distinct assumptions we began with. This approach was used earlier by Horowitz and Manski (1995) in their analysis of the contaminated sampling model. While this general approach may not always be the most natural way to relax an assumption, it is often available and hence can be used to facilitate breakdown frontier analyses.

Throughout the rest of this section we impose both conditional c-dependence and $(1-t)$-percent conditional rank invariance given W.

Assumption A2. *The following hold*:

1. *X is conditionally c-dependent with the potential outcomes Y_x given W, where for all $w \in \operatorname{supp}(W)$ we have $c < \min\{p_{1|w}, p_{0|w}\}$.*

2. *(Y_1, Y_0) satisfies $(1-t)$-percent conditional rank invariance given W, where $t \in [0, 1]$.*

For brevity we focus on the case $c < \min\{p_{1|w}, p_{0|w}\}$ throughout this paper. This allows us to explain the key ideas while keeping the notation and derivations relatively simple. All of our results, however, can be relaxed to the general case where $c \in [0, 1]$.

Partial identification under relaxations of independence and rank invariance

We next study identification under the relaxations of full independence and rank invariance defined above. We begin by briefly recalling results from Masten and Poirier (2018a) on identification of the conditional quantile treatment effect $\text{CQTE}(\tau \mid w) = Q_{Y_1|W}(\tau \mid w) - Q_{Y_0}(\tau \mid w)$, the conditional average treatment effect $\text{CATE}(w) = \mathbb{E}(Y_1 - Y_0 \mid W = w)$, and the conditional marginal cdfs of potential outcomes under c-dependence. We then derive new identification results for the distribution of treatment effects (DTE), $F_{Y_1-Y_0}(z)$, and its related parameter $\mathbb{P}(Y_1 > Y_0)$.

In Masten and Poirier (2018a), we showed that Assumption A1 and Assumption A2.1 imply that

$$\overline{F}^c_{Y_x|W}(y \mid w) = \min\left\{\frac{F_{Y|X,W}(y \mid x, w)p_{x|w}}{p_{x|w} - c}, \frac{F_{Y|X,W}(y \mid x, w)p_{x|w} + c}{p_{x|w} + c}\right\} \tag{4}$$

and

$$\underline{F}^c_{Y_x|W}(y \mid w) = \max\left\{\frac{F_{Y|X,W}(y \mid x, w)p_{x|w}}{p_{x|w} + c}, \frac{F_{Y|X,W}(y \mid x, w)p_{x|w} - c}{p_{x|w} - c}\right\} \tag{5}$$

are functionally sharp bounds on the cdf $F_{Y_x|W}$. Let

$$\overline{Q}^c_{Y_x|W}(\tau \mid w) = Q_{Y|X,W}\left(\tau + \frac{c}{p_{x|w}}\min\{\tau, 1-\tau\} \,\middle|\, x, w\right) \tag{6}$$

denote the inverse of the cdf bound (5) and

$$\underline{Q}^c_{Y_x|W}(\tau \mid w) = Q_{Y|X,W}\left(\tau - \frac{c}{p_{x|w}}\min\{\tau, 1-\tau\} \,\middle|\, x, w\right) \tag{7}$$

denote the inverse of the cdf bound (4). We showed that the identified set for $\text{CQTE}(\tau \mid w)$ is[2]

$$\begin{aligned} &\big[\underline{\text{CQTE}}(\tau, c \mid w), \overline{\text{CQTE}}(\tau, c \mid w)\big] \\ &\quad \equiv \big[\underline{Q}^c_{Y_1|W}(\tau \mid w) - \overline{Q}^c_{Y_0|W}(\tau \mid w), \overline{Q}^c_{Y_1|W}(\tau \mid w) - \underline{Q}^c_{Y_0|W}(\tau \mid w)\big]. \end{aligned} \tag{8}$$

We further showed that, assuming $\mathbb{E}(|Y| \mid X = x, W = w) < \infty$ for $x \in \{0, 1\}$ and $w \in \text{supp}(W)$, the identified set for $\text{CATE}(w)$ is

$$\big[\underline{\text{CATE}}(c \mid w), \overline{\text{CATE}}(c \mid w)\big] \equiv \left[\int_0^1 \underline{\text{CQTE}}(\tau, c \mid w)\, d\tau, \int_0^1 \overline{\text{CQTE}}(\tau, c \mid w)\, d\tau\right]. \tag{9}$$

We use these conditional cdf bounds in our DTE bounds. Bounds on the corresponding unconditional parameters, like $\text{ATE} = \mathbb{E}[\text{CATE}(W)]$, can be obtained by integrating the conditional bounds over the marginal distribution of W. These results are unchanged if we further impose Assumption A2.2. That is, assumptions on rank invariance

[2] In that paper we also extended the bounds in equations (4)–(9) to the $c \geq \min\{p_{1|w}, p_{0|w}\}$ case. As noted earlier, here we focus on the $c < \min\{p_{1|w}, p_{0|w}\}$ case for brevity.

have no identifying power for functionals of the marginal distributions of potential outcomes.

We next derive the identified set for the distribution of treatment effects, the cdf

$$\mathrm{DTE}(z) = \mathbb{P}(Y_1 - Y_0 \leq z).$$

To do this, we first derive the identified set for the conditional distribution of treatment effects (CDTE), the cdf

$$\mathrm{CDTE}(z \mid w) = \mathbb{P}(Y_1 - Y_0 \leq z \mid W = w).$$

By the law of iterated expectations,

$$\mathrm{DTE}(z) = \mathbb{E}\big[\mathrm{CDTE}(z \mid W)\big].$$

Thus we will obtain the identified set for the DTE by averaging bounds for the CDTE. While the ATE only depends on the conditional marginal distributions of potential outcomes, the CDTE depends on the joint distribution of $(Y_1, Y_0) \mid W$. Consequently, as we will see below, the identified set for the CDTE depends on the value of t.

For any $z \in \mathbb{R}$, define $\mathcal{Y}_z(w) = [\underline{y}_1(w), \overline{y}_1(w)] \cap [\underline{y}_0(w) + z, \overline{y}_0(w) + z]$. Note that $\operatorname{supp}(Y_1 - Y_0 \mid W = w) \subseteq [\underline{y}_1(w) - \overline{y}_0(w), \overline{y}_1(w) - \underline{y}_0(w)]$. Let z be an element of $[\underline{y}_1(w) - \overline{y}_0(w), \overline{y}_1(w) - \underline{y}_0(w)]$ such that $\mathcal{Y}_z(w)$ is nonempty. If z is such that $\mathcal{Y}_z(w)$ is empty, then the CDTE is either 0 or 1 depending solely on the relative location of the two supports, which is point identified by Assumption A1.2. In this case, define $\overline{\mathrm{CDTE}}(z, c, t \mid w)$ and $\underline{\mathrm{CDTE}}(z, c, t \mid w)$ to equal this point identified value. If $z > \overline{y}_1(w) - \underline{y}_0(w)$, define these CDTE bounds to equal 1. If $z < \underline{y}_1(w) - \overline{y}_0(w)$, define these CDTE bounds to equal 0.

If $\mathcal{Y}_z(w)$ is nonempty, define

$$\begin{aligned}
\overline{\mathrm{CDTE}}(z, c, t \mid w) &= (1-t)\mathbb{P}\big(\underline{Q}^c_{Y_1|W}(U \mid w) - \overline{Q}^c_{Y_0|W}(U \mid w) \leq z\big) \\
&\quad + t\Big(1 + \min\Big\{\inf_{y \in \mathcal{Y}_z(w)} \big(\overline{F}^c_{Y_1|W}(y \mid w) - \underline{F}^c_{Y_0|W}(y - z \mid w)\big), 0\Big\}\Big), \\
\underline{\mathrm{CDTE}}(z, c, t \mid w) &= (1-t)\mathbb{P}\big(\overline{Q}^c_{Y_1|W}(U \mid w) - \underline{Q}^c_{Y_0|W}(U \mid w) \leq z\big) \\
&\quad + t\max\Big\{\sup_{y \in \mathcal{Y}_z(w)} \big(\underline{F}^c_{Y_1|W}(y \mid w) - \overline{F}^c_{Y_0|W}(y - z \mid w)\big), 0\Big\},
\end{aligned}$$

where $U \sim \mathrm{Unif}[0, 1]$. The following result shows that (a) these are sharp bounds on the CDTE, and (b) the integral of these bounds over the marginal distribution of W yields sharp bounds on the DTE, defined as $\mathbb{P}(Y_1 - Y_0 \leq z)$.

Theorem 1 (DTE bounds). *Suppose the joint distribution of (Y, X, W) is known. Suppose Assumptions A1 and A2 hold. Let $z \in \mathbb{R}$. Then the identified set for $\mathbb{P}(Y_1 - Y_0 \leq z \mid W = w)$ is*

$$\big[\underline{\mathrm{CDTE}}(z, c, t \mid w), \overline{\mathrm{CDTE}}(z, c, t \mid w)\big].$$

Moreover, the identified set for $\mathbb{P}(Y_1 - Y_0 \leq z)$ *is*

$$\begin{aligned}&\big[\underline{\text{DTE}}(z, c, t), \overline{\text{DTE}}(z, c, t)\big]\\&\quad= \left[\int_{\text{supp}(W)} \underline{\text{CDTE}}(z, c, t \mid w)\, dF_W(w), \int_{\text{supp}(W)} \overline{\text{CDTE}}(z, c, t \mid w)\, dF_W(w)\right].\end{aligned}$$

The bound functions $\underline{\text{DTE}}(z, \cdot, \cdot)$ and $\overline{\text{DTE}}(z, \cdot, \cdot)$ are continuous and monotonic in both arguments. When both conditional random assignment ($c = 0$) and conditional rank invariance ($t = 0$) hold, these bounds collapse to a single point and we obtain point identification. If we impose conditional random assignment ($c = 0$) but allow arbitrary dependence between Y_1 and Y_0 ($t = 1$) then we obtain a conditional version of the well known Makarov (1982) bounds. For example, see equation (2) of Fan and Park (2010). DTE bounds have been studied extensively by Fan and coauthors; see the introduction of Fan, Guerre, and Zhu (2017) for a recent and comprehensive discussion of this literature.

Theorem 1 immediately implies that the identified set for $\mathbb{P}(Y_1 - Y_0 > z)$ is

$$\mathbb{P}(Y_1 - Y_0 > z) \in \big[1 - \overline{\text{DTE}}(z, c, t), 1 - \underline{\text{DTE}}(z, c, t)\big].$$

In particular, setting $z = 0$ yields the proportion who benefit from treatment, $\mathbb{P}(Y_1 > Y_0)$. Thus Theorem 1 allows us to study the sensitivity of this parameter to relaxations of full conditional independence and conditional rank invariance.

Finally, notice that all of the bounds and identified sets discussed in this section are analytically tractable and depend on just three functions identified from the population—the conditional cdf $F_{Y|X,W}$, the propensity scores $p_{x|w}$, and the marginal distribution of covariates F_W. This suggests a plug-in estimation approach which we study in Section 3.

Breakdown frontiers

We now formally define the *breakdown frontier*, which generalizes the scalar breakdown point to multiple assumptions or dimensions. We also define the *robust region*, the area below the breakdown frontier. These objects can be defined for different conclusions about different parameters in various models. For concreteness, however, we focus on just a few conclusions about $\mathbb{P}(Y_1 - Y_0 > z)$ and ATE in the potential outcomes model discussed above.

We begin with the conclusion that $\mathbb{P}(Y_1 - Y_0 > z) \geq \underline{p}$ for a fixed $\underline{p} \in [0, 1]$ and $z \in \mathbb{R}$. For example, if $z = 0$ and $\underline{p} = 0.5$, then this conclusion states that at least 50% of people have higher outcomes with treatment than without. If we impose conditional random assignment and conditional rank invariance, then $\mathbb{P}(Y_1 - Y_0 > z)$ is point identified, and hence we can directly check whether this conclusion holds. But the breakdown frontier approach asks: Relative to these baseline assumptions, what are the *weakest* assumptions that allow us to draw this conclusion, given the observed distribution of (Y, X, W)? Specifically, since larger values of c and t correspond to weaker assumptions, what are the *largest* values of c and t such that we can still definitively conclude that $\mathbb{P}(Y_1 - Y_0 > z) \geq \underline{p}$?

We answer this question in two steps. First, we gather *all* values of c and t such that the conclusion holds. We call this set the *robust region*. Since the lower bound of the identified set for $\mathbb{P}(Y_1 - Y_0 > z)$ is $1 - \overline{\text{DTE}}(z, c, t)$ (by Theorem 1), the robust region for the conclusion that $\mathbb{P}(Y_1 - Y_0 > z) \geq \underline{p}$ is

$$\begin{aligned} \text{RR}(z, \underline{p}) &= \big\{(c, t) \in [0, 1]^2 : 1 - \overline{\text{DTE}}(z, c, t) \geq \underline{p}\big\} \\ &= \big\{(c, t) \in [0, 1]^2 : \overline{\text{DTE}}(z, c, t) \leq 1 - \underline{p}\big\}. \end{aligned}$$

The robust region is simply the set of all (c, t) which deliver an identified set for $\mathbb{P}(Y_1 - Y_0 > z)$ which lies on or above $\underline{p}$. See pages 60–61 of Stoye (2005) for similar definitions in the scalar assumption case in a different model. Since $\overline{\text{DTE}}(z, c, t)$ is increasing in c and t, the robust region will be empty if $\overline{\text{DTE}}(z, 0, 0) > 1 - \underline{p}$, and nonempty if $\overline{\text{DTE}}(z, 0, 0) \leq 1 - \underline{p}$. That is, if the conclusion of interest does not hold under the point identifying assumptions, it certainly will not hold under weaker assumptions. From here on we only consider the first case, where the conclusion of interest holds under the point identifying assumptions. That is, we suppose $\overline{\text{DTE}}(z, 0, 0) \leq 1 - \underline{p}$ so that $\text{RR}(z, \underline{p}) \neq \emptyset$.

Second, the *breakdown frontier* is the set of points (c, t) on the boundary of the robust region. Specifically, for the conclusion that $\mathbb{P}(Y_1 > Y_0) \geq \underline{p}$, this frontier is the set

$$\text{BF}(\underline{p}) = \big\{(c, t) \in [0, 1]^2 : \overline{\text{DTE}}(0, c, t) = 1 - \underline{p}\big\}.$$

Solving for t in the equation $\overline{\text{DTE}}(0, c, t) = 1 - \underline{p}$ yields

$$\text{bf}(c, \underline{p}) = \frac{\text{num}(c, \underline{p})}{\text{denom}(c)}, \tag{10}$$

where

$$\text{num}(c, \underline{p}) = 1 - \underline{p} - \int_{\text{supp}(W)} \mathbb{P}\big(\underline{Q}^c_{Y_1|W}(U \mid w) - \overline{Q}^c_{Y_0|W}(U \mid w) \leq 0\big)\, dF_W(w)$$

and

$$\begin{aligned} \text{denom}(c) = 1 + \int_{\text{supp}(W)} \Big[&\min\Big\{\inf_{y \in \mathcal{Y}_0(w)} \big(\overline{F}^c_{Y_1|W}(y \mid w) - \underline{F}^c_{Y_0|W}(y \mid w)\big), 0\Big\} \\ &- \mathbb{P}\big(\underline{Q}^c_{Y_1|W}(U \mid w) - \overline{Q}^c_{Y_0|W}(U \mid w) \leq 0\big)\Big]\, dF_W(w). \end{aligned}$$

Thus we obtain the following analytical expression for the breakdown frontier as a function of c:

$$\text{BF}(c, \underline{p}) = \min\big\{\max\big\{\text{bf}(c, \underline{p}), 0\big\}, 1\big\}.$$

This frontier provides the largest relaxations c and t which still allow us to conclude that $\mathbb{P}(Y_1 > Y_0) \geq \underline{p}$. It thus provides a quantitative measure of robustness of this conclusion to relaxations of the baseline point identifying assumptions of conditional random assignment and conditional rank invariance. Moreover, the shape of this frontier allows us to understand the trade-off between these two types of relaxations in drawing our conclusion.

Here our two relaxations c and t are measured in different units. In general, we can interpret the trade-off between any two relaxations so long as we can interpret the units of each relaxation separately. It is not necessary to measure the two relaxations in common units, although this may sometimes be helpful. This is a common remark when studying trade-offs. For example, in labor supply models agents trade off leisure hours for consumption, although time and quantities of goods are measured in fundamentally different units. Many common rates outside of economics, like kilometers per hour or beats per minute, also do not have common units.

We next consider breakdown frontiers for ATE. Consider the conclusion that $\text{ATE} \geq \mu$ for some $\mu \in \mathbb{R}$. Analogously to above, the robust region for this conclusion is

$$\text{RR}_{\text{ATE}}(\mu) = \{(c, t) \in [0, 1]^2 : \underline{\text{ATE}}(c) \geq \mu\}$$

and the breakdown frontier is

$$\text{BF}_{\text{ATE}}(\mu) = \{(c, t) \in [0, 1]^2 : \underline{\text{ATE}}(c) = \mu\}.$$

These sets are nonempty if $\underline{\text{ATE}}(0) \geq \mu$; that is, if our conclusion holds under the point identifying assumptions. As we mentioned earlier, rank invariance has no identifying power for ATE, and hence the breakdown frontier is a vertical line at the point

$$c^* = \inf\{c \in [0, 1] : \underline{\text{ATE}}(c) \leq \mu\}.$$

This point c^* is a breakdown point for the conclusion that $\text{ATE} \geq \mu$. Note that continuity of $\underline{\text{ATE}}(\cdot)$ implies $\underline{\text{ATE}}(c^*) = \mu$. Thus we've seen two kinds of breakdown frontiers so far: The first had nontrivial curvature, which indicates a trade-off between the two assumptions. The second was vertical in one direction, indicating a lack of identifying power of that assumption.

We can also derive robust regions and breakdown frontiers for more complicated joint conclusions. For example, suppose we are interested in concluding that both $\mathbb{P}(Y_1 > Y_0) \geq \underline{p}$ and $\text{ATE} \geq \mu$ hold. Then the robust region for this joint conclusion is the intersection of the two individual robust regions:

$$\text{RR}(0, \underline{p}) \cap \text{RR}_{\text{ATE}}(\mu).$$

The breakdown frontier for the joint conclusion is the boundary of this intersected region. Viewing these frontiers as functions mapping c to t, the breakdown frontier for this joint conclusion can be computed as the minimum of the two individual frontier functions.

Above we focused on one-sided conclusions about the parameters of interest. Another natural joint conclusion is the two-sided conclusion that $\mathbb{P}(Y_1 - Y_0 > z) \geq \underline{p}$ and $\mathbb{P}(Y_1 - Y_0 > z) \leq \overline{p}$, for $0 \leq \underline{p} < \overline{p} \leq 1$. No new issues arise here: the robust region for this joint conclusion is still the intersection of the two separate robust regions. Keep in mind, though, that whether we look at a one-sided or a two-sided conclusion is unrelated to the fact that we use *lower* confidence bands in Section 3.

Finally, the bootstrap procedures we propose in Section 3 can also be used to do inference on these joint breakdown frontiers. For simplicity, though, in that section we focus on the case where we are only interested in the conclusion $\mathbb{P}(Y_1 - Y_0 > z) \geq \underline{p}$.

3. Estimation and inference

In this section we study estimation and inference on the breakdown frontier defined above. The breakdown frontier is a known functional of the conditional cdf of outcomes given treatment and covariates, the probability of treatment given covariates, and the marginal distribution of the covariates. Hence we propose nonparametric sample analog estimators of the breakdown frontier. We derive $\sqrt{N}$-consistency and asymptotic distributional results using a delta method for directionally differentiable functionals. We then use a bootstrap procedure to construct asymptotically valid lower confidence bands for the breakdown frontier. We also provide a similar procedure for doing inference on breakdown points for ATE.

Although we focus on inference on the breakdown frontier, one might also be interested in doing inference directly on the parameters of interest. If we fix c and t a priori, then we obtain identified sets for ATE, QTE, and the DTE from Section 2. Our asymptotic results may be used as inputs to traditional inference on partially identified parameters. See Canay and Shaikh (2017) for a survey of this literature. For brevity, in this section we only state the asymptotic results for breakdown points and frontiers. In Appendix B.1, we derive the limiting distribution for the following objects: (1) the bounds on the marginal distributions of potential outcomes conditional on W, (2) the CQTE bounds, (3) the CATE and ATE bounds, (4) the CDTE under conditional rank invariance but without full conditional independence, and finally (5) the DTE without either conditional rank invariance or full conditional independence.

We first suppose we observe a random sample of data.

Assumption A3. *The random variables $\{(Y_i, X_i, W_i)\}_{i=1}^N$ are independently and identically distributed according to the distribution of (Y, X, W).*

We assume the support of W is discrete. We sketch an approach to handling continuous covariates in Appendix B.2. Note that W may still be a vector.

Assumption A4. *The support of W is discrete and finite. Let* $\operatorname{supp}(W) = \{w_1, \ldots, w_K\}$.

All parameters of interest are defined as functionals of the underlying parameters $F_{Y|X,W}(y \mid x, w)$, $p_{x|w} = \mathbb{P}(X = x \mid W = w)$, and $q_w = \mathbb{P}(W = w)$. Let

$$\widehat{F}_{Y|X,W}(y \mid x, w) = \frac{\dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \mathbb{1}(Y_i \leq y)\mathbb{1}(X_i = x, W_i = w)}{\dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \mathbb{1}(X_i = x, W_i = w)}, \tag{11}$$

$$\widehat{p}_{x|w} = \frac{\frac{1}{N}\sum_{i=1}^{N}\mathbb{1}(X_i = x, W_i = w)}{\frac{1}{N}\sum_{i=1}^{N}\mathbb{1}(W_i = w)}, \tag{12}$$

and

$$\widehat{q}_w = \frac{1}{N}\sum_{i=1}^{N}\mathbb{1}(W_i = w) \tag{13}$$

denote the sample analog estimators of these three quantities, which converge uniformly to a Gaussian process at a $\sqrt{N}$-rate; see Lemma C1 in Appendix C.

Next consider the bounds (4) and (5) on the marginal distributions of potential outcomes. These population bounds are a functional ϕ_1 evaluated at $(F_{Y|X,W}(\cdot \mid \cdot, \cdot), p_{(\cdot|\cdot)}, q_{(\cdot)})$ where $p_{(\cdot|\cdot)}$ denotes the probability $p_{x|w}$ as a function of $(x, w) \in \{0, 1\} \times \operatorname{supp}(W)$, and $q_{(\cdot)}$ denotes q_w as a function of $w \in \operatorname{supp}(W)$. We estimate these bounds by a plug-in estimator $\phi_1(\widehat{F}_{Y|X,W}(\cdot \mid \cdot, \cdot), \widehat{p}_{(\cdot|\cdot)}, \widehat{q}_{(\cdot)})$. If this functional is differentiable in an appropriate sense, $\sqrt{N}$-convergence in distribution of its arguments will carry over to the functional by the delta method. The type of differentiability we require is *Hadamard directional differentiability*, first defined by Shapiro (1990) and Dümbgen (1993), and further studied in Fang and Santos (2019). We use the functional delta method for Hadamard directionally differentiable mappings (e.g., Theorem 2.1 in Fang and Santos (2019)) to show convergence in distribution of our estimators. Such convergence is usually to a non-Gaussian limiting process. We do not use this distribution to do inference since obtaining analytical asymptotic confidence bands would be challenging. Instead, we use a bootstrap procedure to obtain asymptotically valid uniform confidence bands for our breakdown frontier and associated estimators.

Returning to our population bounds (4) and (5), we estimate these by

$$\begin{aligned}
\widehat{\overline{F}}^{c}_{Y_x|W}(y \mid w) &= \min\left\{\frac{\widehat{F}_{Y|X,W}(y \mid x, w)\widehat{p}_{x|w}}{\widehat{p}_{x|w} - c}, \frac{\widehat{F}_{Y|X,W}(y \mid x, w)\widehat{p}_{x|w} + c}{\widehat{p}_{x|w} + c}\right\}, \\
\widehat{\underline{F}}^{c}_{Y_x|W}(y \mid w) &= \max\left\{\frac{\widehat{F}_{Y|X,W}(y \mid x, w)\widehat{p}_{x|w}}{\widehat{p}_{x|w} + c}, \frac{\widehat{F}_{Y|X}(y \mid x, w)\widehat{p}_{x|w} - c}{\widehat{p}_{x|w} - c}\right\}.
\end{aligned} \tag{14}$$

Note that these estimators may not perform well when c is close to $p_{x|w}$. In our analysis we assume c is bounded away from $p_{x|w}$. We show that these estimators converge in distribution to a nonstandard distribution in Appendix B.1, Lemma 1.

In addition to Assumption A1, we make the following regularity assumptions.

Assumption A5. *For each $x \in \{0, 1\}$ and $w \in \operatorname{supp}(W)$:*

1. $-\infty < \underline{y}_x(w) < \overline{y}_x(w) < +\infty$.

2. *$F_{Y|X,W}(y \mid x, w)$ is continuously differentiable everywhere. Its density $f_{Y|X,W}(y \mid x, w)$ is uniformly continuous in y, uniformly bounded from above, and uniformly bounded away from zero on* $\operatorname{supp}(Y \mid X = x, W = w)$.

Assumption A5.1 combined with our earlier Assumption A1.2 constrain the potential outcomes to have compact support. This compact support assumption is not used to analyze our cdf bounds estimators (14), but we use it later to obtain estimates of the corresponding quantile function bounds uniformly over their arguments $u \in (0,1)$, which we then use to estimate the bounds on $\mathbb{P}(Q_{Y_1|W}(U \mid w) - Q_{Y_0|W}(U \mid w) \leq z)$. This is a well-known issue when estimating quantile processes; for example, see van der Vaart (2000, Lemma 21.4(ii)). Assumption A5.2 requires the density of $Y \mid X, W$ to be bounded away from zero uniformly. This ensures that conditional quantiles of $Y \mid X, W$ are uniquely defined. It also implies that the limiting distribution of the estimated quantile bounds are well behaved. Uniform continuity of the density implies that the derivatives of the conditional quantile function with respect to τ are uniformly continuous.

For our main result in this section (Theorem 2) along with some of our preliminary results (Lemmas 1, 3, and 4 in Appendix B.1), we establish convergence uniformly over $c \in \mathcal{C}$ for some finite grid $\mathcal{C} = \{c_1, c_2, \dots, c_J\} \subset [0, \overline{C}]$ where $\overline{C} \in (0, \min\{p_{1|w}, p_{0|w}\})$ for all $w \in \text{supp}(W)$. We discuss the choice of these grid points in Appendix B.3. We constrain this grid to be below $\min\{p_{1|w}, p_{0|w}\}$ solely for simplicity, as all our results can be extended to grids $\mathcal{C} \subset [0,1]$ by combining our present bound estimates with estimates based on the $c \geq \min\{p_{1|w}, p_{0|w}\}$ case given in Masten and Poirier (2018a). Weak convergence of the breakdown frontier estimators does not hold uniformly over an interval of c since the associated functional is not Hadamard directionally differentiable when its codomain is a set of functions on that interval. To resolve this issue, we propose two ways of conducting inference on the breakdown frontier uniformly over intervals of c. The first is to use the fixed grid and monotonicity of the breakdown frontier to construct a uniform band. The second is to smooth the population breakdown frontier such that it is Hadamard differentiable when viewed as a function of c. We use the first approach in this section and in the empirical illustration, and discuss the second approach in Appendix G in the Online Supplemental Material.

Next, we consider the conditional quantile bounds (6) and (7), which we estimate by

$$\begin{aligned} \widehat{\overline{Q}}^c_{Y_x|W}(\tau) &= \widehat{Q}_{Y|X,W}\left(\tau + \frac{c}{\widehat{p}_{x|w}} \min\{\tau, 1-\tau\} \,\middle|\, x, w\right), \\ \underline{\widehat{Q}}^c_{Y_x|W}(\tau) &= \widehat{Q}_{Y|X,W}\left(\tau - \frac{c}{\widehat{p}_{x|w}} \min\{\tau, 1-\tau\} \,\middle|\, x, w\right). \end{aligned} \tag{15}$$

In Appendix B.1, Lemma 2, we establish their uniform convergence in distribution to a Gaussian process.

By applying the functional delta method, we can show asymptotic normality of smooth functionals of these conditional quantile bounds. A first set of functionals are the CQTE bounds of equation (8), which are a linear combination of the quantile bounds. Let

$$\underline{\widehat{\text{CQTE}}}(\tau, c \mid w) = \underline{\widehat{Q}}^c_{Y_1|W}(\tau \mid w) - \widehat{\overline{Q}}^c_{Y_0|W}(\tau \mid w) \tag{16}$$

and

$$\widehat{\overline{\text{CQTE}}}(\tau, c \mid w) = \widehat{\overline{Q}}^c_{Y_1|W}(\tau \mid w) - \underline{\widehat{Q}}^c_{Y_0|W}(\tau \mid w). \tag{17}$$

Both of these estimators converge in distribution to Gaussian processes.

A second set of functionals are the CATE bounds from equation (9). These bounds are continuous linear functionals of the CQTE bounds. Therefore the joint asymptotic distribution of these bounds can be established by the continuous mapping theorem. Let

$$\underline{\widehat{\text{CATE}}}(c \mid w) = \int_0^1 \underline{\widehat{\text{CQTE}}}(u, c \mid w)\, du \quad \text{and} \quad \overline{\widehat{\text{CATE}}}(c \mid w) = \int_0^1 \overline{\widehat{\text{CQTE}}}(u, c \mid w)\, du.$$

Then, by the linearity of the integral operator, these estimated CATE bounds converge to their population counterpart at a $\sqrt{N}$-rate. We give their asymptotic distribution in Appendix B.1.

We estimate the unconditional ATE bounds by integrating over the empirical distribution of the covariates W: Let

$$\underline{\widehat{\text{ATE}}}(c) = \frac{1}{N}\sum_{i=1}^{N} \underline{\widehat{\text{CATE}}}(c \mid W_i) \quad \text{and} \quad \overline{\widehat{\text{ATE}}}(c) = \frac{1}{N}\sum_{i=1}^{N} \overline{\widehat{\text{CATE}}}(c \mid W_i). \tag{18}$$

We show that these estimated ATE bounds converge weakly to a Gaussian element in Appendix B.1.

Next consider estimation of the breakdown point for the claim that ATE $\geq \mu$ where $\mu \in \mathbb{R}$. To focus on the nondegenerate case, suppose the population value of ATE obtained under full independence is greater than μ, $\underline{\text{ATE}}(0) > \mu$. This implies $c^* > 0$. Let

$$\widehat{c}^* = \inf\bigl\{c \in [0,1] : \underline{\widehat{\text{ATE}}}(c) \leq \mu\bigr\} \tag{19}$$

be the estimated breakdown point. This is the estimated smallest relaxation of independence such that we cannot conclude that the ATE is strictly greater than μ. By the properties of the quantile bounds as a function of c, the function $\underline{\text{ATE}}(c)$ is nonincreasing and differentiable in c. We now formally present a result about the asymptotic distribution of $\widehat{c}^*$.

Proposition 1. *Suppose Assumptions A1, A3, A4, and A5 hold. Assume $c^* \in (0, \overline{C}]$. Then $\sqrt{N}(\widehat{c}^* - c^*) \rightsquigarrow \mathbf{Z}_{\text{BP}}$, a Gaussian random variable.*

The assumption that $c^* \in (0, \overline{C}]$ can again be relaxed to the general case where $c^* \in (0,1]$ but we maintain the stronger assumption for brevity. We characterize $\mathbf{Z}_{\text{BP}}$, the asymptotic distribution of the estimated breakdown point, in the proof of Proposition 1.

Under conditional rank invariance, we can also establish asymptotic normality of bounds for $\mathbb{P}(Q_{Y_1|W}(U \mid w) - Q_{Y_0|W}(U \mid w) \leq z)$ for a fixed $z \in \mathbb{R}$. These bounds are given by

$$\bigl(\underline{P}(c \mid w), \overline{P}(c \mid w)\bigr) \equiv \bigl(\underline{\text{CDTE}}(z, c, 0 \mid w), \overline{\text{CDTE}}(z, c, 0 \mid w)\bigr).$$

We keep z implicit in the notation for these bounds. Estimates for these quantities are provided by

$$
\begin{aligned}
\widehat{\underline{P}}(c \mid w) &= \int_0^1 \mathbb{1}\big(\widehat{\overline{Q}}^c_{Y_1|W}(u \mid w) - \widehat{\underline{Q}}^c_{Y_0|W}(u \mid w) \le z\big)\, du, \\
\widehat{\overline{P}}(c \mid w) &= \int_0^1 \mathbb{1}\big(\widehat{\underline{Q}}^c_{Y_1|W}(u \mid w) - \widehat{\overline{Q}}^c_{Y_0|W}(u \mid w) \le z\big)\, du.
\end{aligned}
\tag{20}
$$

Their convergence to nonstandard distributions can be established using the Hadamard directional differentiability of the mapping from the differences in quantile bounds to the bounds $(\underline{P}(c \mid w), \overline{P}(c \mid w))$. We do this in Appendix B.1, Lemma 3. There we also discuss an additional technical assumption on the smoothness of the conditional quantile functions, Assumption A6.

If conditional random assignment holds $(c = 0)$ in addition to conditional rank invariance $(t = 0)$, then the CDTE is point identified and Lemma 3 in Appendix B.1 gives the asymptotic distribution of the sample analog CDTE estimator (in this case the upper and lower bound functions are equal). This can be considered an estimator of the CDTE in one of the models of Matzkin (2003).

For c and t values greater than or equal to zero, we estimate the CDTE bounds by

$$
\begin{aligned}
&\widehat{\underline{\text{CDTE}}}(z, c, t \mid w) \\
&\quad = (1-t)\widehat{\underline{P}}(c \mid w) + t \max\Big\{ \sup_{y \in \mathcal{Y}_z(w)} \big(\widehat{\underline{F}}^c_{Y_1|W}(y \mid w) - \widehat{\overline{F}}^c_{Y_0|W}(y - z \mid w)\big), 0\Big\}, \\
&\widehat{\overline{\text{CDTE}}}(z, c, t \mid w) \\
&\quad = (1-t)\widehat{\overline{P}}(c \mid w) + t\Big(1 + \min\Big\{ \inf_{y \in \mathcal{Y}_z(w)} \big(\widehat{\overline{F}}^c_{Y_1|W}(y \mid w) - \widehat{\underline{F}}^c_{Y_0|W}(y - z \mid w)\big), 0\Big\}\Big).
\end{aligned}
$$

We estimate the unconditional DTE bounds by integrating the estimated CDTE bounds over the empirical distribution of the covariates: Let

$$
\begin{aligned}
\widehat{\underline{\text{DTE}}}(z, c, t) &= \frac{1}{N}\sum_{i=1}^N \widehat{\underline{\text{CDTE}}}(z, c, t \mid W_i), \\
\widehat{\overline{\text{DTE}}}(z, c, t) &= \frac{1}{N}\sum_{i=1}^N \widehat{\overline{\text{CDTE}}}(z, c, t \mid W_i).
\end{aligned}
$$

Lemma 3 in Appendix B.1 shows that the terms $\underline{P}(c \mid w)$ and $\overline{P}(c \mid w)$ are estimated at a $\sqrt{N}$-rate by the Hadamard directional differentiability of the mapping linking empirical cdfs and these terms. The second components of the CDTE bounds are a Hadamard directionally differentiable functional as well, leading to the $\sqrt{N}$ joint convergence of the DTE bounds to a tight, random element uniformly in c and t. Lemma 4 in Appendix B.1 shows this formally.

Having established the convergence in distribution of the DTE, we can now show that the breakdown frontier also converges in distribution uniformly over its arguments.

Denote the estimated breakdown frontier for the conclusion that $\mathbb{P}(Y_1 > Y_0) \geq \underline{p}$ by

$$\widehat{\mathrm{BF}}(c, \underline{p}) = \min\{\max\{\widehat{\mathrm{bf}}(c, \underline{p}), 0\}, 1\}, \tag{21}$$

where

$$\widehat{\mathrm{bf}}(c, \underline{p}) = \frac{\widehat{\mathrm{num}}(c, \underline{p})}{\widehat{\mathrm{denom}}(c)} \tag{22}$$

with

$$\widehat{\mathrm{num}}(c, \underline{p}) = 1 - \underline{p} - \frac{1}{N}\sum_{i=1}^{N} \widehat{\overline{P}}(c \mid W_i),$$

$$\widehat{\mathrm{denom}}(c) = 1 + \frac{1}{N}\sum_{i=1}^{N}\Big[\min\Big\{\inf_{y \in \mathcal{Y}_0(W_i)}\big(\widehat{\overline{F}}{}^c_{Y_1|W}(y \mid W_i) - \widehat{\underline{F}}{}^c_{Y_0|W}(y \mid W_i)\big), 0\Big\} - \widehat{\overline{P}}(c \mid W_i)\Big].$$

We next show that the estimated breakdown frontier converges in distribution.

Theorem 2. *Suppose Assumptions A1, A3, A4, A5, and A6 hold. Let $\mathcal{P} \subset [0, 1]$ be a finite grid of points. Then*

$$\sqrt{N}\big(\widehat{BF}(c, \underline{p}) - BF(c, \underline{p})\big) \rightsquigarrow \mathbf{Z}_{\mathrm{BF}}(c, \underline{p}),$$

a tight random element of $\ell^\infty(\mathcal{C} \times \mathcal{P})$.

This result essentially follows from the convergence of the preliminary estimators established in Lemma C1 in Appendix C and by showing that the breakdown frontier is a composition of a number of Hadamard differentiable and Hadamard directionally differentiable mappings, implying convergence in distribution of the estimated breakdown frontier.

Breakdown frontiers for more complex conclusions can typically be constructed from breakdown frontiers for simpler conclusions. For example, consider the breakdown frontier for the joint conclusion that $\mathbb{P}(Y_1 > Y_0) \geq \underline{p}$ and $\mathrm{ATE} \geq \mu$. Then the breakdown frontier for this joint conclusion is the minimum of the two individual frontier functions. Alternatively, consider the conclusion that $\mathbb{P}(Y_1 > Y_0) \geq \underline{p}$ *or* $\mathrm{ATE} \geq \mu$, or both, hold. Then the breakdown frontier for this joint conclusion is the maximum of the two individual frontier functions. Since the minimum and maximum operators are Hadamard directionally differentiable, the samaple analog estimators of these joint breakdown frontiers will also converge in distribution.

Since the limiting process is non-Gaussian, inference on the breakdown frontier is not based on standard errors as with Gaussian limiting theory. Our processes' distribution is characterized fully by the expressions in Appendices B.1 and C, but obtaining analytical estimates of quantiles of functionals of these processes would be challenging. In the next subsection we give details on the bootstrap procedure we use to construct confidence bands for the breakdown frontier.

Bootstrap inference

As mentioned earlier, we use a bootstrap procedure to do inference on the breakdown frontier rather than directly using its limiting process. In this subsection we discuss how to use the bootstrap to approximate this limiting process. In the next subsection we discuss its application to constructing uniform confidence bands.

First we define some general notation. Let $Z_i = (Y_i, X_i, W_i)$ and $Z^N = \{Z_1, \dots, Z_N\}$. Let θ_0 denote some parameter of interest and let $\widehat{\theta}$ be an estimator of θ_0 based on the data Z^N. Let $\mathbf{A}_N^*$ denote $\sqrt{N}(\widehat{\theta}^* - \widehat{\theta})$ where $\widehat{\theta}^*$ is a draw from the nonparametric bootstrap distribution of $\widehat{\theta}$. Suppose $\mathbf{A}$ is the tight limiting process of $\sqrt{N}(\widehat{\theta} - \theta_0)$. Denote bootstrap consistency by $\mathbf{A}_N^* \overset{P}{\rightsquigarrow} \mathbf{A}$ where $\overset{P}{\rightsquigarrow}$ denotes weak convergence in probability, conditional on the data Z^N. Weak convergence in probability conditional on Z^N is defined as

$$\sup_{h \in \mathrm{BL}_1} \left| \mathbb{E}\left[h\left(\mathbf{A}_N^*\right) \mid Z^N\right] - \mathbb{E}\left[h(\mathbf{A})\right] \right| = o_p(1),$$

where BL_1 denotes the set of Lipschitz functions into $\mathbb{R}$ with Lipschitz constant no greater than 1.

We focus on the following specific choices of θ_0 and $\widehat{\theta}$:

$$\theta_0 = \begin{pmatrix} F_{Y|X,W}(\cdot \mid \cdot, \cdot) \\ p_{(\cdot|\cdot)} \\ q_{(\cdot)} \end{pmatrix} \quad \text{and} \quad \widehat{\theta} = \begin{pmatrix} \widehat{F}_{Y|X,W}(\cdot \mid \cdot, \cdot) \\ \widehat{p}_{(\cdot|\cdot)} \\ \widehat{q}_{(\cdot)} \end{pmatrix}.$$

For these choices, let $\mathbf{Z}_N^* = \sqrt{N}(\widehat{\theta}^* - \widehat{\theta})$. Let $\mathbf{Z}_1$ denote the limiting distribution of $\sqrt{N}(\widehat{\theta} - \theta_0)$; see Lemma C1 in Appendix C. Theorem 3.6.1 of van der Vaart and Wellner (1996) implies that $\mathbf{Z}_N^* \overset{P}{\rightsquigarrow} \mathbf{Z}_1$. Our parameters of interest are all functionals ϕ of θ_0. For Hadamard differentiable functionals ϕ, the nonparametric bootstrap is consistent. For example, see Theorem 3.1 of Fang and Santos (2019). They further show that ϕ is Hadamard differentiable *if and only if*

$$\sqrt{N}\left(\phi(\widehat{\theta}^*) - \phi(\widehat{\theta})\right) \overset{P}{\rightsquigarrow} \phi'_{\theta_0}(\mathbf{Z}_1),$$

where ϕ'_{θ_0} denotes the Hadamard derivative at θ_0. This implies that the nonparametric bootstrap can be used to do inference on the QTE and ATE bounds since they are Hadamard differentiable functionals of θ_0. A second implication is that the nonparametric bootstrap is not consistent for the DTE or for the breakdown frontier for claims about the DTE since they are Hadamard directionally differentiable mappings of θ_0, but they are not ordinary Hadamard differentiable.

In such cases, Fang and Santos (2019) showed that a different bootstrap procedure is consistent. Specifically, let $\widehat{\phi}'_{\theta_0}$ be a consistent estimator of ϕ'_{θ_0}. Under some regularity conditions, their results imply that

$$\widehat{\phi}'_{\theta_0}\left(\mathbf{Z}_N^*\right) \overset{P}{\rightsquigarrow} \phi'_{\theta_0}(\mathbf{Z}_1).$$

Analytical consistent estimates of ϕ'_{θ_0} are often difficult to obtain, so Dümbgen (1993) and Hong and Li (2018) proposed using a numerical derivative estimate of ϕ'_{θ_0}. Their estimate of the limiting distribution of $\sqrt{N}(\phi(\widehat{\theta}) - \phi(\theta_0))$ is given by the distribution of

$$\widehat{\phi}'_{\theta_0}\big(\sqrt{N}(\widehat{\theta}^* - \widehat{\theta})\big) = \frac{\phi\big(\widehat{\theta} + \varepsilon_N \sqrt{N}(\widehat{\theta}^* - \widehat{\theta})\big) - \phi(\widehat{\theta})}{\varepsilon_N} \tag{23}$$

across the bootstrap estimates $\widehat{\theta}^*$. Under the rate constraints $\varepsilon_N \to 0$ and $\sqrt{N}\varepsilon_N \to \infty$, and some measurability conditions stated in their Appendix, Hong and Li (2018) showed

$$\widehat{\phi}'_{\theta_0}\big(\sqrt{N}(\widehat{\theta}^* - \widehat{\theta})\big) \overset{P}{\rightsquigarrow} \phi'_{\theta_0}(\mathbf{Z}_1),$$

where the left-hand side is defined in equation (23).

This bootstrap procedure requires evaluating ϕ at two values, which is computationally simple. It also requires selecting the tuning parameter ε_N, which we discuss in Appendix B.4. Note that the standard, or naive, bootstrap is a special case of this numerical delta method bootstrap where $\varepsilon_N = N^{-1/2}$.

Uniform confidence bands for the breakdown frontier

In this subsection, we combine all of our asymptotic results thus far to construct uniform confidence bands for the breakdown frontier. As in Section 2, we use the function $\text{BF}(\cdot, \underline{p})$ to characterize this frontier. We specifically construct one-sided *lower* uniform confidence bands. That is, we will construct a lower band function $\widehat{\text{LB}}(c)$ such that

$$\lim_{N\to\infty} \mathbb{P}\big(\widehat{\text{LB}}(c) \le \text{BF}(c, \underline{p}) \text{ for all } c \in [0, 1]\big) = 1 - \alpha.$$

We use a one-sided lower uniform confidence band because this gives us an *inner* confidence set for the robust region. Specifically, define the set

$$\text{RR}_L = \big\{(c, t) \in [0, 1]^2 : t \le \widehat{\text{LB}}(c)\big\}.$$

Then validity of the confidence band $\widehat{\text{LB}}$ implies

$$\lim_{N\to\infty} \mathbb{P}\big(\text{RR}_L \subseteq \text{RR}(0, \underline{p})\big) = 1 - \alpha.$$

Thus the area underneath our confidence band, RR_L, is interpreted as follows: Across repeated samples, approximately $100(1 - \alpha)$% of the time, every pair $(c, t) \in \text{RR}_L$ leads to a population level identified set for the parameter $\mathbb{P}(Y_1 > Y_0)$ which lies weakly above $\underline{p}$. Put differently, approximately $100(1 - \alpha)$% of the time, every pair $(c, t) \in \text{RR}_L$ still lets us draw the conclusion we want at the population level. Hence the size of this set RR_L is a finite sample measure of robustness of our conclusion to failure of the point identifying assumptions. We discuss an alternative testing-based interpretation in Appendix D in the Online Supplemental Material.

One might be interested in constructing one-sided *upper* confidence bands if the goal was to do inference on the set of assumptions for which we *cannot* come to the conclusion of interest. This might be useful in situations where two opposing sides are debating a conclusion. But since our focus is on trying to determine when we can come to the desired conclusion, rather than looking for when we cannot, we only describe the one-sided lower confidence band case.

When studying inference on scalar breakdown points, Kline and Santos (2013) constructed one-sided lower confidence intervals. Unlike for breakdown frontiers, uniformity over different points in the assumption space is not a concern for inference on breakdown points. See Appendix D in the Online Supplemental Material for more discussion.

We consider bands of the form

$$\widehat{\mathrm{LB}}(c) = \widehat{\mathrm{BF}}(c, \underline{p}) - \widehat{k}(c) \tag{24}$$

for some function $\widehat{k}(\cdot) \geq 0$. This band is an asymptotically valid lower uniform confidence band of level $1-\alpha$ if

$$\lim_{N\to\infty} \mathbb{P}\big(\widehat{\mathrm{BF}}(c, \underline{p}) - \widehat{k}(c) \leq \mathrm{BF}(c, \underline{p}) \text{ for all } c \in [0,1]\big) = 1-\alpha,$$

or, equivalently, if

$$\lim_{N\to\infty} \mathbb{P}\Big(\sup_{c\in[0,1]} \sqrt{N}\big(\widehat{\mathrm{BF}}(c, \underline{p}) - \mathrm{BF}(c, \underline{p}) - \widehat{k}(c)\big) \leq 0\Big) = 1-\alpha.$$

In our theoretical analysis, we consider $\widehat{k}(c) = \widehat{z}_{1-\alpha}\sigma(c)$ for a scalar $\widehat{z}_{1-\alpha}$ and a function σ. We focus on known σ for simplicity. We start by deriving a uniform band over a grid $\mathcal{C}$, then extend it over an interval using monotonicity of the breakdown frontier. As discussed earlier, we only derive uniformity of the band over $c \in [0, \overline{C}]$ rather than over $c \in [0,1]$, but this is also for brevity and can be relaxed. The choice of σ affects the shape of the confidence band, and there are many possible choices of the function σ which yield valid level $1-\alpha$ uniform confidence bands. See Freyberger and Rai (2018) for a detailed analysis. A simple choice of σ is the constant function: $\sigma(c) = 1$, which delivers an equal width uniform band. Alternatively, as we do below, one could choose $\sigma(c)$ to construct a minimum width confidence band (equivalently, maximum area of RR_L).

Proposition 2. *Suppose Assumptions A1, A3, A4, A5, and A6 hold. Define* $\phi : \ell^\infty(\mathbb{R} \times \{0,1\} \times \mathrm{supp}(W)) \times \ell^\infty(\{0,1\} \times \mathrm{supp}(W)) \times \ell^\infty(\mathrm{supp}(W)) \to \ell^\infty(\mathcal{C})$ *such that*

$$\widehat{BF}(c, \underline{p}) = \big[\phi(\widehat{\theta})\big](c).$$

Then ϕ is Hadamard directionally differentiable. Suppose that $\varepsilon_N \to 0$ and $\sqrt{N}\varepsilon_N \to \infty$. Let $\widehat{\theta}^*$ denote a draw from the nonparametric bootstrap distribution of $\widehat{\theta}$. Then

$$\big[\widehat{\phi}'_{\theta_0}\big(\sqrt{N}(\widehat{\theta}^* - \widehat{\theta})\big)\big] \overset{P}{\rightsquigarrow} \big[\phi'_{\theta_0}(\mathbf{Z}_1)\big] \equiv \mathbf{Z}_{\mathrm{BF}}. \tag{25}$$

For a given function $\sigma(\cdot)$ such that $\inf_{c\in\mathcal{C}} \sigma(c) > 0$, define

$$\widehat{z}_{1-\alpha} = \inf\left\{z \in \mathbb{R} : \mathbb{P}\left(\sup_{c\in\mathcal{C}} \frac{\big[\widehat{\phi}'_{\theta_0}\big(\sqrt{N}(\widehat{\theta}^* - \widehat{\theta})\big)\big](c, \underline{p})}{\sigma(c)} \leq z \;\Big|\; Z^N\right) \geq 1-\alpha\right\}. \tag{26}$$

Finally, suppose also that the cdf of

$$\sup_{c\in\mathcal{C}} \frac{[\phi'_{\theta_0}(\mathbf{Z}_1)](c, \underline{p})}{\sigma(c)} = \sup_{c\in\mathcal{C}} \frac{\mathbf{Z}_{\mathrm{BF}}(c, \underline{p})}{\sigma(c)}$$

is continuous and strictly increasing at its $1-\alpha$ *quantile, denoted* $z_{1-\alpha}$. *Then* $\widehat{z}_{1-\alpha} = z_{1-\alpha} + o_p(1)$.

This proposition is a variation of Corollary 3.2 in Fang and Santos (2015). Note that this result is pointwise in the underlying dgp; we are unsure if it can be extended to hold uniformly over the dgp and leave this question to future work. Proposition 2 implies that the lower $1-\alpha$ band $\widehat{\mathrm{LB}}(c) = \widehat{\mathrm{BF}}(c, \underline{p}) - \widehat{z}_{1-\alpha}\sigma(c)$ is valid uniformly on the grid $\mathcal{C}$. To extend the uniformity to all of $[0, \overline{C}]$, we propose the following lower confidence band:

$$\widetilde{\mathrm{LB}}(c) = \begin{cases} \widehat{\mathrm{LB}}(c_1) & \text{if } c \in [0, c_1], \\ \vdots & \\ \widehat{\mathrm{LB}}(c_j) & \text{if } c \in (c_{j-1}, c_j], \text{ for } j = 2, \dots, J, \\ \vdots & \\ 0 & \text{if } c \in (c_J, \overline{C}]. \end{cases}$$

This band is a step function which interpolates between grid points using the least monotone interpolation. The following result shows its validity.

COROLLARY 1. *Let the assumptions of Proposition 2 hold. Then* $\widetilde{LB}(c)$ *is a uniform lower* $1-\alpha$ *band for* $BF(c, \underline{p})$ *over* $c \in [0, \overline{C}]$.

Corollary 1 shows that, for any fixed $J \geq 1$, the interpolated lower confidence band preserves the exact $1-\alpha$ coverage on the grid points. This follows by monotonicity of the breakdown frontier; see Lemma C6 in Appendix C. That said, this interpolated band might not be taut, in the sense that there may exist other lower bands with $1-\alpha$ coverage that are weakly larger than $\widetilde{\mathrm{LB}}(c)$ for all c and strictly larger at some values of c. See Freyberger and Rai (2018) for further discussion of taut confidence bands.

Proposition 2 can be extended to estimated functions σ, although we leave the details for future work. We use an estimated σ in our application, as described next. When both $z_{1-\alpha}$ and σ are estimated, we work directly with $\widehat{k}(c) = \widehat{z}_{1-\alpha}\widehat{\sigma}(c)$. We choose $\widehat{k}(c)$ to minimize an approximation to the area between the confidence band and the estimated function; equivalently, to maximize the area of RR_L. Specifically, we let $\widehat{k}(c_1), \dots, \widehat{k}(c_J)$ solve

$$\min_{k(c_1),\dots,k(c_J)\geq 0} \sum_{j=2}^{J} k(c_j)(c_j - c_{j-1}) \tag{27}$$

subject to

$$\mathbb{P}\Big(\sup_{c\in\{c_1,\dots,c_J\}} \sqrt{N}\big(\widehat{\mathrm{BF}}(c, \underline{p}) - \mathrm{BF}(c, \underline{p}) - k(c)\big) \leq 0\Big) = 1-\alpha, \tag{28}$$

where we approximate the left-hand side probability via the numerical delta method bootstrap. The criterion function here is just a right Riemann sum over the grid points. This optimization is not computationally costly: It is only performed once per value of $\underline{p}$ and ε_N. Moreover, in our empirical illustration it takes an average of 15 seconds per run on a mid-2013 MacBook Air.

4. Empirical illustration: The effects of child soldiering

In this section we use our results to examine the impact of assumptions in determining the effects of child soldiering on wages. We first briefly discuss the background and motivation and then we present our analysis.

Background

We use data from phase 1 of SWAY, the Survey of War Affected Youth in northern Uganda, conducted by principal researchers Jeannie Annan and Chris Blattman (see Annan, Blattman, and Horton (2006)). As Blattman and Annan (2010) discuss on page 882, a primary goal of this survey was to understand the effects of a twenty year war in Uganda, where "an unpopular rebel group has forcibly recruited tens of thousands of youth." In their paper, they use this data to examine the impacts of abduction on educational, labor market, psychosocial, and health outcomes. In our illustration, we focus solely on the impact of abduction on wages.

Blattman and Annan noted that self-selection into the military is a common problem in the literature studying the effects of military service on outcomes. They argue that forced recruitment in Uganda led to random assignment of military service in their data. They first provide qualitative evidence for this, based on interviews with former rebels who led raiding parties. After murdering and mutilating civilians, the rebels had no public support, making abduction the only means of recruitment. Youths were generally taken during nighttime raids on rural households. According to the former rebel leaders, "targets were generally unplanned and arbitrary; they raided whatever homesteads they encountered, regardless of wealth or other traits."

This qualitative evidence is supported by their survey data, where Blattman and Annan show that most pretreatment covariates are balanced across the abducted and nonabducted groups (see their Table 2). Only two covariates are not balanced: year of birth and prewar household size. They say this is unsurprising because

> "a youth's probability of ever being abducted depended on how many years of the conflict he fell within the [rebel group's] target age range. Moreover, abduction levels varied over the course of the war, so youth of some ages were more vulnerable to abduction than others. The significance of household size, meanwhile, is driven by households greater than 25 in number. We believe that rebel raiders, who traveled in small bands, were less likely to raid large, difficult-to-control households." (page 887)

Hence they use a selection-on-observables identification strategy, conditioning on these two variables.

While their evidence supporting the full conditional independence assumption is compelling, this assumption is still nonrefutable. Hence they apply the methods of Imbens (2003) to analyze the sensitivity to this assumption. In this analysis, they only consider one outcome variable, years of education. Likewise, as in Imbens (2003), they only look at one parameter, the constant treatment effect in a fully parametric model.

We complement their results by applying the breakdown frontier methods we develop in this paper. We focus on the log-wage outcome variable. We look at both the average treatment effect and $\mathbb{P}(Y_1 > Y_0)$, which was not studied in Blattman and Annan (2010).

Motivation for the parameter $\mathbb{P}(Y_1 > Y_0)$

Blattman and Annan (2010) is part of a large literature on the impact of compulsory military service on wages. As Card and Cardoso (2012, pages 57–58) note,

> "Angrist (1990) showed that Vietnam-era draftees had lower earnings than non-draftees, a finding he attributed to the low value of military experience in the civilian labor market. Subsequent research in the United States and other countries, however, has uncovered a surprisingly mixed pattern of impacts."

On one hand, enlistees learn basic skills and receive occupational training in the military. On the other hand, they forgo civilian schooling and work experience, and may experience debilitating psychological trauma. Which of these two explanations is correct? Most likely, both. By using models where the treatment effects $Y_1 - Y_0$ are heterogeneous, we allow some people to have overall positive effects, and thus satisfy the first explanation, and some people to have overall negative effects, and thus satisfy the second explanation. The parameter $\mathbb{P}(Y_1 > Y_0)$ tells us precisely which proportion of the population primarily satisfies the first versus the second explanation. Thus it gives researchers a more nuanced understanding of treatment effects, by quantitatively measuring the prevalence of two opposing explanations of the impact of treatment on outcomes. We suspect this parameter can be particularly helpful in literatures which find a "mixed pattern of impacts," like the work on compulsory military service and wages discussed here.

That said, one may be concerned that the rank invariance assumption used to point identify this parameter is too strong. As in Heckman, Smith, and Clements (1997), this concern motivates our sensitivity analysis. For further motivation of this parameter in various settings, see Bedoya, Bittarello, Davis, and Mittag (2017) and Mullahy (2018).

Sample and summary statistics

The original phase 1 SWAY data has 1216 males born between 1975 and 1991. Of these, wage data is available for 504 observations. 56 of these earned zero wages; we drop these and only look at people who earned positive wages. This leaves us with our main sample of 448 observations. In addition to this outcome variable, we let our treatment variable be an indicator that the person was *not* abducted. We include the two covariates discussed above, age when surveyed and household size in 1996. Additional covariates can be included, but we focus on just these two for simplicity.

Table 1. Summary statistics.

Variable Name	Mean	Median	Stddev	Min	Max
Daily wage in Uganda shillings	2957.60	1400.00	6659.76	35.71	83,333.34
Log wage	7.23	7.24	1.18	3.58	11.33
Not abducted?	0.36	0.00	0.48	0.00	1.00
Age when surveyed	22.11	22.00	4.88	14.00	30.00
Household size in 1996	8.31	8.00	4.19	2.00	28.00

Note: Sample size is 448. 1 USD is approximately 1800 Uganda shillings (Exchange rate at time of survey, 2005–2006; source: World Bank).

Table 1 shows summary statistics for these four variables. 36% of our sample were not abducted. Age ranges from 14 years old to 30 years old, with a median of 22 years old. Household size ranges from 2 people to 28, with a median of 8 people. Wages range from as low as 36 shillings to as high as about 83,300 shillings, with a median of 1400 shillings.

Age has 17 support points and household size has 21 support points. Hence there are 357 total covariate cells. Including the treatment variable, this yields 714 total cells, compared to our sample size of 448 observations. Since we focus on unconditional parameters, having small or zero observations per cell is not a problem in principle. However, in the finite sample we have, to ensure that our estimates of the cdf bounds $\overline{F}^c_{Y_x|W}(y \mid w)$ and $\underline{F}^c_{Y_x|W}(y \mid w)$ are reasonably smooth in y, we collapse our covariates as follows. 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 reduces the number of covariate cells to 4, giving 8 total cells including the treatment variable. This yields approximately 55 observations per cell. While this crude approach suffices for our illustration, in more extensive empirical analyses one may want to use more sophisticated methods. For example, we could use discrete kernel smoothing, as discussed in Li and Racine (2008), who also provide additional references. We also consider alternative coarsenings in Appendix H in the Online Supplemental Material.

Baseline analysis for ATE

Table 2 shows unconditional comparisons of means of the outcome and the original covariates across the treatment and control groups. Wages for people who were not abducted are 702 shillings larger on average. People who were not abducted are also about 1.4 years younger than those who were abducted. People who were not abducted also had a slightly larger household size than those who were abducted. Only the difference in ages is statistically significant at the usual levels, but as in Tables 2 and 3 of Blattman and Annan (2010) the standard errors can be decreased by including additional controls. These extra covariates are not essential for illustrating our breakdown frontier methods, however.

The point estimates in Table 2 do not condition on the two covariates. Next we consider the conditional independence assumption, with age and household size in 1996 as

Table 2. Comparison of means.

Variable Name	Not Abducted	Abducted	Difference
Daily wage in Uganda shillings	3409.12	2706.75	702.36 [725.49]
Log wage	7.33	7.18	0.15 [0.12]
Age when surveyed	21.23	22.60	−1.37 [0.48]
Household size in 1996	8.53	8.19	0.34 [0.42]
Observations	160	288	

Note: Sample size is 448. 1 USD is approximately 1800 Uganda shillings (Exchange rate at time of survey, 2005–2006; source: World Bank). Standard errors in brackets.

our covariates. Under this assumption, our estimate of ATE is 890 [726.13] shillings when the outcome variable is level of wages, and is 0.21 [0.11] when the outcome variable is log wage. Using their full set of control variables, Blattman and Annan (2010) estimated ATE to be 0.33 [0.15] when the outcome is log wage. See column 1 of their Table 3.

Breakdown point for ATE: Implementation

To check the robustness of these baseline point estimates to failure of conditional independence, we estimate the breakdown point c^* for the conclusion ATE ≥ 0, where we use log wage as our outcome variable. We measure relaxations of conditional independence by our conditional c-dependence distance. Here, we describe the mechanics of estimating the breakdown point. In the next subsection, we describe the empirical results.

Since our estimated ATE is positive, the estimated breakdown point depends only on the estimated ATE lower bound function, as defined by equation (19). Computing $\widehat{c}^*$ using this equation requires two steps:

1. Define a function to compute $\underline{\widehat{\text{ATE}}}(c)$ for any $c \in [0, 1]$. This can be done in several ways. One is to directly compute the estimators in equation (18). That approach averages over the covariate cells in the final step. Here we average over them in the first step. Either approach works for the ATE, but averaging over the covariates early leads to simpler code.

(a) Define two functions to compute the conditional cdf bounds estimators in equation (14). At this step we use two preliminary estimators:

i. Cell means to compute the propensity scores $\widehat{p}_{x|w}$, as defined in equation (12).

ii. A smoothed empirical cdf $\widehat{F}_{Y|X,W}(\cdot \mid x, w)$. For example, see Hansen (2004). We use the standard logistic kernel. We use Hansen's method for selecting the bandwidth and then divide by two, as manual undersmoothing.

Average this function over the covariate cells (using $\widehat{q}_w$ from equation (13) as weights) to get estimates of the unconditional cdf bounds $\widehat{\overline{F}}^c_{Y_x}(y)$ and $\widehat{\underline{F}}^c_{Y_x}(y)$. Note that equation (14) requires $c < \min_{x,w} p_{x|w}$. For larger c values we use the sample analogs of equations (6) and (7) in Masten and Poirier (2018a), which hold for all values of $c \in [0, 1]$.

(b) Given these functions, define a function to compute the unconditional QTE bound

$$\underline{\widehat{\text{QTE}}}(\tau, c) = \big(\widehat{\overline{F}}^c_{Y_1}\big)^{-1}(\tau) - \big(\widehat{\underline{F}}^c_{Y_0}\big)^{-1}(\tau).$$

This is the unconditional version of equation (16). To compute the inverse cdfs, we do the following: Define a uniform grid from the smallest to largest observed outcome value, with 500 grid points. Evaluate the cdf at each point. This gives 500 pairs of the form $(y, F(y))$. Flip them around to get $(F(y), y)$ and we have the quantile function evaluated at 500 points. Then use linear interpolation to compute the quantile function at any value τ.

(c) The last step is to numerically integrate these QTE bounds to get

$$\underline{\widehat{\text{ATE}}}(c) = \int_0^1 \underline{\widehat{\text{QTE}}}(u, c)\, du.$$

Any numerical integration method can be used here. We use a simple approach: Approximate the intergral by an average over 1000 uniformly spaced grid points between 0 and 1. That is, we evaluate the integrand at each of the grid points and then take the mean.

2. The function from step 1 is weakly decreasing in c. Given this function, we just need to solve for roots of the equation $\underline{\widehat{\text{ATE}}}(c) = 0$. This can be done using standard root finding algorithms, like bisection. We use Matlab's `fzero` function, which uses a more sophisticated algorithm than simple bisection, but is still guaranteed to converge. See Chapter 4 of Brent (1973). The solution is our estimate $\widehat{c}^*$.

Breakdown point for ATE: Results

The estimated breakdown point is $\widehat{c}^* = 0.041$. Based on this point estimate, for all $x \in \{0, 1\}$ and $w \in \text{supp}(W)$ we can allow the conditional propensity scores $\mathbb{P}(X = x \mid Y_x = y, W = w)$ to vary ± 4 percentage points around the observed propensity scores $\mathbb{P}(X = x \mid W = w)$ without changing our conclusion that ATE ≥ 0.

Is this a big or small amount of variation? Well, as a baseline, the upper bound on c is about 0.73. This is an estimate of

$$\max_{w \in \text{supp}(W)} \max\big\{\mathbb{P}(X = 1 \mid W = w), \mathbb{P}(X = 0 \mid W = w)\big\}.$$

Any $c \geq 0.73$ would lead to the no assumptions identified set for ATE. In this sense, 0.041 is quite small, which would suggest that our results are quite fragile. Next we examine variation in the observed propensity scores as we suggested in Masten and Poirier (2018a). Specifically, we consider the difference between the "full" propensity score and the "leave out variable k" propensity score which omits variable k: Define

$$\overline{c}_{\texttt{age}} = \sup_{s=0,1} \sup_{a=0,1} \big|\widehat{\mathbb{P}}(X = 1 \mid \texttt{age} = a, \texttt{hhSize} = s) - \widehat{\mathbb{P}}(X = 1 \mid \texttt{hhSize} = s)\big|$$

and

$$\overline{c}_{\texttt{hhSize}} = \sup_{a=0,1} \sup_{s=0,1} \left|\widehat{\mathbb{P}}(X = 1 \mid \texttt{age} = a, \texttt{hhSize} = s) - \widehat{\mathbb{P}}(X = 1 \mid \texttt{age} = a)\right|.$$

Using these numbers as a reference, a robust result would have a breakdown point above one or both of the $\overline{c}$'s. In the data, we obtain $\overline{c}_{\texttt{age}} = 0.0625$ and $\overline{c}_{\texttt{hhSize}} = 0.0403$. The estimated breakdown point $\widehat{c}^* = 0.041$ is below $\overline{c}_{\texttt{age}}$ and approximately equal to $\overline{c}_{\texttt{hhSize}}$. This suggests that perhaps our conclusion could be considered somewhat robust. Accounting for sampling uncertainty in the breakdown point, however, shows that the true breakdown point may be less than $\overline{c}_{\texttt{hhSize}}$. Overall, this suggests that our conclusion that ATE ≥ 0 is not robust to relaxations of full conditional independence.

This argument for judging the plausibility of specific values of c relies on using variation in the observed propensity score to ground our beliefs about reasonable variation in the unobserved propensity scores. The general question here is how one should quantitatively distinguish "large" and "small" relaxations of an assumption. This is an old and ongoing question in the sensitivity analysis literature, and much work remains to be done. For discussions on this point for different measures of deviations or relaxations from independence in various settings, see Rotnitzky, Robins, and Scharfstein (1998), Robins (1999), Imbens (2003), Altonji, Elder, and Taber (2005, 2008), and Oster (2019).

Baseline analysis for $\mathbb{P}(Y_1 > Y_0)$

Next consider the parameter $\mathbb{P}(Y_1 > Y_0)$. Since we define treatment as *not* being abducted, this parameter measures the proportion of people who earn higher wages when they are not abducted, compared to when they are abducted. For this parameter, we must make both the full conditional independence assumption and the conditional rank invariance assumption to obtain point identification. Under these assumptions, our point estimate is 0.67 with a one-sided lower 95% CI of [0.48, 1]. Is this point estimate robust to failures of full conditional independence and conditional rank invariance? We address this in the next two subsections.

Breakdown frontier for $\mathbb{P}(Y_1 > Y_0)$*: Implementation*

To check the robustness of our baseline point estimate, we estimate breakdown frontiers and corresponding confidence bands for the conclusion that $\mathbb{P}(Y_1 > Y_0) \geq \underline{p}$. We do this for $\underline{p} = 0.1$, 0.25, 0.5 as in our Monte Carlo simulations in Appendix F. We do not consider $\underline{p} = 0.75$ or 0.9 since these values are larger than our point estimate under the baseline assumptions; they yield empty estimated robust regions. Besides picking a grid of $\underline{p}$'s a priori, one could let $\underline{p} = \widehat{p}_{0,0}/2$, half the value of the parameter estimated under the baseline point identifying assumptions. In our application this is 0.34; we omit this choice for brevity. Imbens (2003) suggested a similar choice of cutoff in his approach.

We next explain the mechanics of how to compute the breakdown frontier estimate and its confidence bands. In the following subsection, we describe the empirical results. We start with the breakdown frontier point estimate, as defined by equation (21). There are two components to computing this equation:

1. The term $\widehat{\overline{P}}(c \mid w)$. This estimate is defined as an integral in equation (20).

(a) To compute the integrand, we use equation (15). We compute the conditional quantile functions by inverting the conditional cdfs, as described in the ATE analysis above.

(b) To compute the integral we use proposition 1(ii) of Chernozhukov, Fernández-Val, and Galichon (2010). This proposition uses the special structure of this integral—which is called the *pre-rearrangement operator*—to write the integral in terms of the roots of the function inside the indicator function. This allows us to quickly and accurately compute this integral. It converts the problem of numerical integration into a problem of root finding. To compute these roots, we again use Matlab's `fzero`.

2. The infimum term in $\widehat{\text{denom}}(c)$.

(a) We compute the objective function as described in our ATE breakdown analysis above, using smoothed empirical cdfs.

(b) We then use a combination of grid search and Newton's method to minimize this objective function. Here we search over y values between the smallest and largest observed outcome variable.

This shows how to estimate the breakdown frontier for any c and $\underline{p}$. Later when we plot our estimate we use a finite grid of points c. Specifically, we use 25 values of c, equally spaced between 0 and $0.73 = \max_{x,w} \widehat{p}_{x|w}$. We use the same grid for our bootstrap procedure. The rule of thumb suggested in Appendix B.3 suggests using 22 grid points. We use a few more since the estimated breakdown frontier is zero for most of these values anyway, even for the smallest $\underline{p}$ value. So the number of relevant grid points is fairly small.

Next we describe how we implement the bootstrap confidence bands developed in Section 3. This bootstrap requires that we choose a tuning parameter, ε_N. To choose this parameter, we first restrict attention to a grid of seven possible values: For $\ell = 1, \dots, 7$ we let $\varepsilon_\ell = r \cdot \varepsilon_N^{\text{naive}}$ where $r \in \{1, 1.5, 2, 4, 6, 8, 10\}$ and $\varepsilon_N^{\text{naive}} = 1/\sqrt{N}$. We use the same grid in our Monte Carlo simulations in Appendix F (in the simulations we also use the value $r = 0.5$). Given this grid, we do the following:

1. Draw and store 1000 bootstrap samples of the data. Here we use the standard nonparametric bootstrap.

2. For each ε_ℓ and each bootstrap sample:

(a) Compute the perturbed parameter $\widehat{\theta} + \varepsilon_\ell \sqrt{N}(\widehat{\theta}^* - \widehat{\theta})$ in equation (23). For the perturbed cdf, we form an equally spaced grid of 10,000 values of y between the smallest and largest observed outcome variables. We then compute and store

$$\widehat{F}_{Y|X,W}(y \mid x, w) + \varepsilon_\ell \sqrt{N}\big(\widehat{F}^*_{Y|X,W}(y \mid x, w) - \widehat{F}_{Y|X,W}(y \mid x, w)\big)$$

for all y values on the grid, all $x \in \{0, 1\}$, and all w in the support of the covariates. Here $\widehat{F}_{Y|X,W}$ is the empirical cdf from the original data (equation (11)) and $\widehat{F}^*_{Y|X,W}$ denotes the empirical cdf using the bootstrap data. We use empirical cdfs here since they are substantially faster to compute than smoothed cdfs. For each x and w value, we then monotonize this perturbed cdf and truncate it to lie between 0 and 1, to ensure that it is a valid cdf. This is not necessary asymptotically, but it is helpful in finite samples. Moreover, this operation is Hadamard directionally differentiable and hence does not affect the asymptotic theory.

(b) Next we compute the breakdown frontier at all grid points c and all $\underline{p}$ using these perturbed values as the first step plug-in estimators. This gives us $\phi(\widehat{\theta} + \varepsilon_\ell \sqrt{N}(\widehat{\theta}^* - \widehat{\theta}))$ from equation (23). At this step we compute the integral defining the term $\widehat{\overline{P}}(c \mid w)$ differently than above. The root-based approach is not always reliable in these bootstrap perturbed samples since the term inside the indicator function can oscillate dramatically. So here we instead take a grid of 500 equally spaced values between 0 and 1, evaluate the integrand at all values and take the average. We also use a slightly smoothed indicator function for the integrand.

(c) Continuing from the previous step, subtract the estimated breakdown frontier using the original data and divide by ε_ℓ to finish computing equation (23).

3. Fixing an ε_ℓ and a c, the distribution of the term from the previous step across the bootstrap samples approximates the sampling distribution of

$$\sqrt{N}\big(\widehat{\mathrm{BF}}(c, \underline{p}) - \mathrm{BF}(c, \underline{p})\big).$$

This continues to hold if we look at the vector of these terms for all c in the grid. Thus, for any given vector of constants $k(c)$, we can approximate the left-hand side of equation (28) by looking at the distribution of the supremum over c of equation (23) from the previous step minus $\sqrt{N}k(c)$. Specifically, note that the inequality inside the probability in equation (28) can be written as a product of indicator functions for each grid point. Since we will optimize subject to this constraint, it is helpful to smooth it. So we approximate each indicator function by a smoothed indicator. We use the logitstic kernel and for each c we pick the bandwidth by Hansen's (2004) method, using the distribution of the term from equation (23) across the bootstrap samples as data. For a given bootstrap dataset, we compute the product of these indicators. We then average that product over all bootstrap datasets to get an estimate of the probability on the left-hand side of equation (28).

4. Solve equation (27) subject to equation (28). Since we smoothed the constraint function this can be done using any nonlinear optimizer. At this step we must pick our desired coverage probability. We compute 95% confidence bands. Denote the solution vector by $\widehat{k}(c)$.

5. Finally, we compute the lower confidence band using equation (24). For certain values of ε_ℓ, this band can be nonmonotonic; in these cases we monotonize the band.

This procedure produces seven different confidence bands—one for each ε_ℓ. To pick our preferred band, we use the double bootstrap procedure described in Appendix B.4. We implement this as follows:

1. Draw and store 500 bootstrap samples of the data. For this we use the smoothed bootstrap. Specifically, for each draw $i = 1, \ldots, N$,

(a) Draw a value of the covariate vector from its empirical distribution.

(b) Draw a treatment value from the Bernoulli distribution with success parameter equal to the estimated propensity score evaluated at the covariate draw from the previous step.

(c) Draw from the Unif[0, 1] distribution. Evaluate the smoothed quantile function $\widehat{F}^{-1}_{Y|X,W}(\cdot \mid x, w)$ at this draw, where w is from the first step and x is from the second step. This defines the outcome variable value for this draw. We compute the smoothed quantile by linear interpolation of the smoothed cdf, as described earlier.

2. For each ε_ℓ,

(a) For each of the 500 bootstrap samples, treat it as if it was the true data and compute a confidence band as we described above using this choice of ε_ℓ. Check whether this confidence band lies under the breakdown frontier estimated from the original data. Note that this estimate corresponds to the true breakdown frontier in the "bootstrap world" where we treat the smoothed empirical cdf $\widehat{F}_{Y|X,W}$ as the true distribution of outcomes given treatment and covariates.

(b) Compute the estimated coverage probability for this choice of ε_ℓ by averaging the indicators of coverage for each of the 500 samples.

3. Pick the ε_ℓ with estimated coverage probability closest to 95%.

Part 2(a) can be done in parallel on a computing cluster.

Breakdown frontier for $\mathbb{P}(Y_1 > Y_0)$: Results

Figure 3 shows the results of the procedure we just described. As in our earlier plots, the horizontal axis plots c, relaxations of full conditional independence, while the vertical

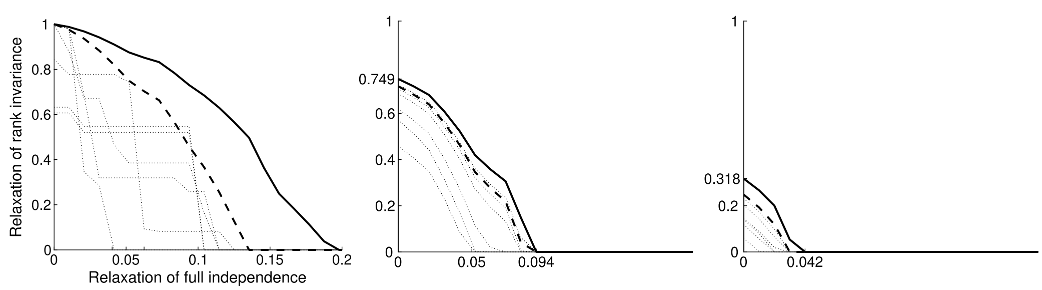

FIGURE 3. Estimated breakdown frontiers (solid lines) and confidence bands (dotted/dashed lines) for the claim $\mathbb{P}(Y_1 > Y_0) \geq \underline{p}$. Left to right: $\underline{p} = 0.1, 0.25, 0.5$. The light dotted lines are confidence bands for all seven values of ε_N considered. The darker dashed line is the band selected by the double bootstrap procedure described in Appendix B.4. See text for additional implementation details.

axis plots t, relaxations of conditional rank invariance. As mentioned earlier, the natural upper bound for c is about 0.73. Since all of the breakdown frontiers intersect the horizontal axis at much smaller values, we have cut off the part of the overall assumption space with $c \geq 0.2$. Remember that, for the following analysis, it is valid to examine various (c, t) combinations since we use uniform confidence bands.

First consider the left plot, $\underline{p} = 0.1$. Since this is the weakest conclusion of the five we consider, the estimated breakdown frontier and the corresponding robust region are the largest among the three plots. If we impose full conditional independence, then our estimated frontier suggests that we can completely relax conditional rank invariance and still conclude that at least 10% of people benefit from not being forced into military service. Even accounting for sampling uncertainty, we can still draw this conclusion. Moreover, looking at all choices of ε_N—not just our selected one—the lowest the vertical intercept ever gets is about 61%. Next, suppose we relax full conditional independence. Recall that the maximal relaxation between the observed propensity score and the "leave out variable k" propensity scores gave $\overline{c}_{\texttt{age}} = 0.0625$ and $\overline{c}_{\texttt{hhSize}} = 0.0403$. Both of these numbers are substantially smaller than the horizontal intercept of our selected confidence band. Hence, if we impose full conditional rank invariance, our conclusion that $\mathbb{P}(Y_1 > Y_0) \geq 0.1$ is robust to relaxations of full conditional independence. Suppose instead that we think selection on unobservables is at most the largest $\overline{c}$ value, about 0.06. Then for c's in the range $[0, 0.06]$, and accounting for sampling uncertainty, we can still conclude $\mathbb{P}(Y_1 > Y_0) \geq 0.1$ so long as at least 30% of the population satisfies rank invariance. Thus we can relax full independence within this range without paying too high a cost in terms of requiring stronger rank invariance assumptions.

If we are willing to restrict selection on unobservables to be smaller than the largest $\overline{c}$ value, then we can allow for larger relaxations of conditional rank invariance. To quantify this trade-off, we can compute the difference between the values of the estimated breakdown frontier at two different points. As a starting point we recommend computing

$$\widehat{\mathrm{BF}}(\overline{c}_{(K)}, \underline{p}) - \widehat{\mathrm{BF}}(\overline{c}_{(K-1)}, \underline{p}),$$

where K denotes the number of observed regressors, $\overline{c}_{(K)}$ denotes the largest value of $\overline{c}$, and $\overline{c}_{(K-1)}$ denotes the second largest value. One could also divide this difference by $\overline{c}_{(K)} - \overline{c}_{(K-1)}$ to get a secant line. In our empirical application, $\overline{c}_{(K)} = \overline{c}_{\texttt{age}}$ and $\overline{c}_{(K-1)} = \overline{c}_{\texttt{hhSize}}$. Thus

$$\widehat{\mathrm{BF}}(\overline{c}_{\texttt{age}}, 0.1) - \widehat{\mathrm{BF}}(\overline{c}_{\texttt{hhSize}}, 0.1) = -6\%.$$

The corresponding secant line has slope about -3. Thus if we assume selection on unobservables is at most as large as the *second* largest amount of variation in "leave out variable k" propensity scores, we can allow for an additional 6% of the population to violate conditional rank invariance. Put differently, around these values of c, allowing latent conditional propensity scores to vary an extra 1 percentage point requires us to impose that an additional 3% of the population must satisfy conditional rank invariance. This rate of substitution generally increases as c gets larger. Our ability to quantify this kind of trade-off between assumptions is a primary goal of our breakdown frontier analysis.

Overall, our results from this top left plot suggest that the conclusion that at least 10% of people benefit from not being forced into military service is robust to relaxations of full conditional independence up to twice the size we see between the observed and leave out variable k propensity scores, depending on how much conditional rank invariance failure we allow. For relaxations of full conditional independence up to the largest value of $\overline{c}$, we can allow up to 70% of the population to deviate from conditional rank invariance, accounting for sampling uncertainty.

Next consider the middle plot, $\underline{p} = 0.25$. Since this is a stronger conclusion than the previous one, all the frontiers are shifted toward the origin. Consequently, by construction, this conclusion is not as robust as the previous one. Our qualitative conclusions, however, are similar to those obtained for $\underline{p} = 0.1$. If we impose full conditional independence, we can allow conditional rank invariance to fail for about 70% of the population. Conversely, if we impose full conditional rank invariance, we can allow the latent conditional propensity scores to vary by about 10 percentage points—well beyond the largest observed variation $\overline{c}$. For $\underline{p} = 0.25$, we have

$$\widehat{\mathrm{BF}}(\overline{c}_{\texttt{age}}, 0.25) - \widehat{\mathrm{BF}}(\overline{c}_{\texttt{hhSize}}, 0.25) = -17\%.$$

Hence the slope around our observed maximal $\overline{c}$'s is much larger for $\underline{p} = 0.25$ as compared to $\underline{p} = 0.1$. An important caveat to our conclusions for both $\underline{p} = 0.1$ and $\underline{p} = 0.25$ is that there is substantial variation in confidence bands as ε_N changes. This point underscores the need for future work on the choice of ε_N.

Next consider the right plot, $\underline{p} = 0.5$. Here we consider the conclusion that at least half of people benefit from not being forced into military service. If we impose full conditional independence, and accounting for sampling uncertainty, then we can allow conditional rank invariance to fail for about 25% of the population. This is quite large, but it relies on full conditional independence holding exactly. If we also relax conditional independence to $c = 0.03$, then we need conditional rank invariance to hold for everyone if we still want to conclude that at least 50% of people benefit from not being forced into military service. 0.03 is smaller than both $\overline{c}_{\texttt{age}}$ and $\overline{c}_{\texttt{hhSize}}$. Hence we might not be comfortable with such small values of c. This suggests the data do not definitively support the conclusion $\mathbb{P}(Y_1 > Y_0) \geq 0.5$, even though our point estimate under the baseline assumptions is 0.67.

Empirical conclusions

In this section we used our breakdown frontier methods to study the robustness of conclusions about ATE and $\mathbb{P}(Y_1 > Y_0)$ to failures of conditional independence and conditional rank invariance. We first considered the conclusion that the average treatment effect of not being abducted on log wages is nonnegative. Our point estimates suggest that this conclusion is robust to deviations in unobserved latent propensity scores of up to the same value as $\overline{c}_{\texttt{age}}$, which is also about two-thirds as large as $\overline{c}_{\texttt{hhSize}}$; this robustness does not hold up when accounting for sampling uncertainty, however. We then considered the conclusion that at least $\underline{p}\%$ of people earn higher wages when they are not abducted. This conclusion is robust to large simultaneous relaxations of conditional rank invariance and conditional independence for $\underline{p} = 10\%$. For $\underline{p} = 25\%$, this conclusion continues to be robust to reasonable relaxations, although after accounting for the variation in confidence bands over ε_N, this conclusion appears to be more sensitive to conditional independence than to conditional rank invariance. This robustness to rank invariance matches the findings of Heckman, Smith, and Clements (1997), who imposed full independence and studied deviations from rank invariance. In their table 5B they found that, in their empirical application, one could generally conclude that $\mathbb{P}(Y_1 > Y_0)$ was at least 50%, regardless of the assumption on rank invariance. In our empirical application our results are not quite as robust to rank invariance failures, which could be because we use a different measure of relaxation of rank invariance, and also because of differences in the empirical applications.

5. Conclusion

Summary

In this paper we advocated the *breakdown frontier* approach to sensitivity analysis. Given a set of baseline assumptions, this approach defines the population breakdown frontier as the weakest set of assumptions such that a specific conclusion of interest holds. Sample analog estimates and lower uniform confidence bands allow researchers to do inference on this frontier. The area under the confidence band is a quantitative, finite sample measure of the robustness of a conclusion to relaxations of point identifying assumptions. To examine this robustness, empirical researchers can present these estimated breakdown frontiers and their accompanying confidence bands along with traditional point estimates and confidence intervals obtained under point identifying assumptions. We illustrated this general approach in the context of a treatment effects model, where the robustness of conclusions about ATE and $\mathbb{P}(Y_1 > Y_0)$ to relaxations of random assignment and rank invariance are examined. We applied these results in an empirical study of the effect of child soldiering on wages. We found that weak conclusions about $\mathbb{P}(Y_1 > Y_0)$ are fairly robust to failures of both rank invariance and random assignment, but stronger conclusions are more sensitive to relaxations of random assignment.

Breakdown frontier analysis for other models and other relaxations

As discussed in Section 1, breakdown frontier analysis can in principle be done in most models. In that section we outlined the six main steps required for any breakdown frontier analysis. In this paper we illustrated this general approach by studying a single important and widely used model: the potential outcomes model with a binary treatment. In future work it would be helpful to perform breakdown frontier analyses in other models. In particular, it may be possible to do breakdown frontier analyses in a large class of models by using the general identification analysis in Chesher and Rosen (2017) or Torgovitsky (2019).

A key conceptual step in any breakdown frontier analysis is deciding how to define the indexed classes of assumptions such that the magnitude of the relaxation can be reasonably interpreted. This is not easy, and will generally depend on the model, the specific kind of assumption being relaxed, and the empirical context. Moreover, this choice may affect our findings: A conclusion can be robust with respect to one measure of relaxation but not another. Thus one goal of future research is to explore this space of assumption relaxations, to understand their substantive interpretations, and to chart their implications for the robustness of empirical findings. In Masten and Poirier (2016) we have already compared three different measures of relaxation of the random assignment assumption, including the one used here. We further studied quantile independence, a common relaxation of random assignment, in Masten and Poirier (2018b). In the present paper, we also used a general method for spanning two discrete assumptions by defining a $(1-t)$-percent relaxation, as we did with rank invariance. But much work still remains to be done.

Appendix A: Related literature

We begin with the identification literature on breakdown points; as mentioned earlier, here we use "breakdown" in the same sense as Horowitz and Manski's (1995) identification breakdown point. This breakdown point idea goes back to the one of the earliest sensitivity analyses, performed by Cornfield, Haenszel, Hammond, Lilienfeld, Shimkin, and Wynder (1959). They essentially asked how much correlation between a binary treatment and an unobserved binary confounder must be present to fully explain an observed correlation between treatment and a binary outcome, in the absence of any causal effects of treatment. This level of correlation between treatment and the confounder is a kind of breakdown point for the conclusion that some causal effects of treatment are nonzero. Their approach was substantially generalized by Rosenbaum and Rubin (1983), which is discussed in detail in Chapter 22 of Imbens and Rubin (2015). Neither Cornfield et al. (1959) nor Rosenbaum and Rubin (1983) formally defined breakdown points.

Horowitz and Manski (1995) gave the first formal definition and analysis of breakdown points. They studied a "contaminated sampling" model, where one observes a mixture of draws from the distribution of interest and draws from some other distribution. An upper bound λ on the unknown mixing probability indexes identified sets for functionals of the distribution of interest. They focus on a single conclusion: That this functional is not equal to its logical bounds. They then define the breakdown point λ^* as the largest λ such that this conclusion holds. Put differently, λ^* is the largest mixing probability we can allow while still obtaining a nontrivial identified set for our parameter of interest. They also relate this "identification breakdown point" to the earlier breakdown point concepts studied in the robust statistics literature (e.g., Hampel, Ronchetti, Rousseeuw, and Stahel (1986, pp. 96–98) and Huber and Ronchetti (2009, Section 1.4 and Chapter 11)).

More generally, much work by Manski distinguishes between informative and noninformative bounds (which the literature also sometimes calls tight and nontight bounds; see Section 7.2 of Ho and Rosen (2017)). The breakdown point is the boundary between the informative and noninformative cases. For example, see his analysis of bounds on quantiles with missing outcome data on page 40 of Manski (2007). There the identification breakdown point for the τth quantile occurs when $\max\{\tau, 1-\tau\}$ is the proportion of missing data. Similar discussions are given throughout the book.

Stoye (2005, 2010) generalizes the formal identification breakdown point concept by noting that breakdown points can be defined for any claim about the parameter of interest. He then studies a specific class of relaxations of the missing-at-random assumption in a model of missing data. Kline and Santos (2013) studied a different class of relaxations of the missing-at-random assumption and also define a breakdown point based on that class.

While all of these papers study a scalar breakdown point, Imbens (2003) studied a model of treatment effects where deviations from conditional random assignment are parameterized by two numbers $r = (r_1, r_2)$. His parameter of interest $\theta(r)$ is point identified given a fixed value of r. Imbens' Figures 1–4 essentially plot estimated level sets of this function $\theta(r)$, in a transformed domain. While suggestive, these level sets do not generally have a breakdown frontier interpretation. This follows since nonmonotonicities in the function $\theta(r)$ lead to level sets which do not always partition the space of sensitivity parameters into two connected sets in the same way that our breakdown frontier does.

Manski and Pepper (2018) also studied a model where relaxations of baseline assumptions are parameterized by a vector of numbers r. Unlike Imbens, however, they derive identified sets indexed by r. These sets are weakly increasing (in the set inclusion order) in each component of r, and hence the nonmonotonicity issue does not arise. For a two-dimensional relaxation, their Table 2 presents identified sets as a function of a grid of $r = (r_1, r_2)$ values. The boundary between the italicized identified sets in that table and the nonhighlighted sets is essentially a discrete approximation to the breakdown frontier in their model, for the claim that the parameter of interest is positive. Similarly, the boundary between the bold identified sets in that table and the nonhighlighted sets is essentially a discrete approximation to the breakdown frontier in their model, for the claim that the parameter of interest is negative.

Neither Horowitz and Manski (1995) nor Stoye (2005, 2010) discussed estimation or inference of breakdown points. Imbens (2003) estimated his level sets in an empirical application, but does not discuss inference. Manski and Pepper (2018) also do not discuss estimation of or inference on breakdown frontiers, although inference in their setting is conceptually complicated—see their discussion on pages 234–235. Kline and Santos (2013), on the other hand, is the first and only paper we are aware of that explicitly suggests doing inference on a breakdown point. We build on their work by proposing to do inference on the multidimensional breakdown frontier. This allows us to study the trade-off between different assumptions in drawing conclusions. They do study something they call a "breakdown curve," but this is a collection of scalar breakdown points for many different claims of interest, analogous to the collection of frontiers presented in Figure 2. Inference on a frontier rather than a point also raises additional issues they did not discuss; see our Appendix D in the Online Supplemental Material for more details.

Moreover, we study a model of treatment effects while they look at a model of missing data, hence our identification analysis is different.

Building on Horowitz and Manski (1995), Kreider, Pepper, Gundersen, and Jolliffe (2012) combine a continuous relaxation sensitivity analysis for assumptions regarding measurement error with various discrete relaxations of assumptions regarding treatment selection. This allows them to study the interaction between these two kinds of assumptions in drawing conclusions. For inference, they present confidence intervals for partially identified parameters for a variety of values of the relaxations, rather than doing inference on breakdown frontiers. See Gundersen, Kreider, and Pepper (2012) and Kreider, Pepper, and Roy (2016) for further examples of identification analysis combining discrete and continuous relaxations.

Our breakdown frontier is a known functional of the distribution of outcomes given treatment and covariates and the observed propensity scores. This functional is not Hadamard differentiable, however, which prevents us from applying the standard functional delta method to obtain its asymptotic distribution. Instead, we show that it is Hadamard directionally differentiable, which allows us to apply the results of Fang and Santos (2019). We then use the numerical bootstrap of Dümbgen (1993) and Hong and Li (2018) to construct our confidence bands. For other applications of Hadamard directional differentiability, see Kaido (2016), Hansen (2017), and Lee and Bhattacharya (2019).

Our identification analysis builds on two strands of literature. First is the literature on relaxing statistical independence assumptions. There is a large literature on this, including important work by Rosenbaum and Rubin (1983), Robins, Rotnitzky, and Scharfstein (2000), and Rosenbaum (1995, 2002). We apply results from our paper Masten and Poirier (2018a), which discusses that literature in more detail. In that paper we did not study estimation or inference. Second is the literature on identification of the distribution of treatment effects $Y_1 - Y_0$, especially without the rank invariance assumption. In their introduction, Fan, Guerre, and Zhu (2017) provided a comprehensive discussion of this literature; also see Abbring and Heckman (2007, Section 2). Here we focus on the papers most related to our sensitivity analysis. Heckman, Smith, and Clements (1997) performed a sensitivity analysis to the rank invariance assumption by fixing the value of Kendall's τ for the joint distribution of potential outcomes, and then varying τ from -1 to 1; see Tables 5A and 5B. Their analysis is motivated by a search for breakdown points, as evident in their Section 4 title, "How far can we depart from perfect dependence and still produce plausible estimates of program impacts?" Nonetheless, they do not formally define identified sets for parameters given their assumptions on Kendall's τ, and they do not formally define a breakdown point. Moreover, they do not suggest estimating or doing inference on breakdown points. Gechter (2016) performed a sensitivity analysis to the rank invariance assumption by fixing a lower bound on the value of Spearman's ρ. Under this assumption he derives the identified set for a certain average treatment effect. He then studies estimation and inference on this set for a fixed value of the sensitivity parameter. Fan and Park (2009) provided formal identification results for the joint cdf of potential outcomes and the distribution of treatment effects under the known Kendall's τ assumption. They also discuss how to extend those results to known Spearman's ρ in their Remark 1. They provide estimation and inference methods for their bounds, but do not study breakdown points. Finally, none of these papers study the specific relaxation of rank invariance we consider (as defined in Section 2).

In this section, we have focused narrowly on the papers most closely related to ours. We situate our work more broadly in the literature on inference in sensitivity analyses in Appendix D in the Online Supplemental Material. In that section we also briefly discuss Bayesian inference, although we use frequentist inference in this paper.

Appendix B: Additional inference results and discussion

B.1 *Asymptotic results for the bound functionals*

In this section we provide asymptotic results for the various bound functionals discussed in this paper. These are preliminary results used in our breakdown analysis done in Section 3. They can also be used on their own as inputs to traditional inference on partially identified parameters.

First we establish convergence in distribution of the cdf bound estimators (14). Here and throughout the paper we use the following notation: For an arbitrary set $\mathcal{A}$ and a Banach space $\mathcal{B}$, $\ell^\infty(\mathcal{A}, \mathcal{B})$ denotes the set of all maps $z : \mathcal{A} \to \mathcal{B}$ with finite sup-norm $\|z\| = \sup_{a \in \mathcal{A}} \|z(a)\|_{\mathcal{B}}$, equipped with this norm. For example, see van der Vaart and Wellner (1996, p. 381).

Lemma 1. *Suppose Assumptions A1, A3, and A4 hold. Let $\mathcal{Y} \subset \mathbb{R}$ be a finite grid of points. Then*

$$\sqrt{N}\begin{pmatrix} \widehat{\overline{F}}^c_{Y_x|W}(y \mid w) - \overline{F}^c_{Y_x|W}(y \mid w) \\ \widehat{\underline{F}}^c_{Y_x|W}(y \mid w) - \underline{F}^c_{Y_x|W}(y \mid w) \end{pmatrix} \rightsquigarrow \mathbf{Z}_2(y, x, w, c),$$

a tight random element of $\ell^\infty(\mathcal{Y} \times \{0, 1\} \times \operatorname{supp}(W) \times \mathcal{C}, \mathbb{R}^2)$.

$\mathbf{Z}_2$ is not Gaussian itself, but it is a continuous transformation of Gaussian processes. For given (x, c, w), the limit will be Gaussian at all values of y except for

$$y \in \left\{ Q_{Y|X,W}\left(\frac{p_{x|w} - c}{2p_{x|w}} \,\Big|\, x, w \right),\ Q_{Y|X,W}\left(\frac{p_{x|w} + c}{2p_{x|w}} \,\Big|\, x, w \right) \right\}.$$

Next we consider the conditional quantile bound estimators. Recall that $\overline{C} \in (0, \min\{p_{1|w}, p_{0|w}\})$ for all $w \in \operatorname{supp}(W)$.

Lemma 2. *Suppose Assumptions A1, A3, A4, and A5 hold. Then*

$$\sqrt{N}\begin{pmatrix} \widehat{\overline{Q}}^c_{Y_x|W}(\tau \mid w) - \overline{Q}^c_{Y_x|W}(\tau \mid w) \\ \widehat{\underline{Q}}^c_{Y_x|W}(\tau \mid w) - \underline{Q}^c_{Y_x|W}(\tau \mid w) \end{pmatrix} \rightsquigarrow \mathbf{Z}_3(\tau, x, w, c),$$

a mean-zero Gaussian process in $\ell^\infty((0, 1) \times \{0, 1\} \times \operatorname{supp}(W) \times [0, \overline{C}], \mathbb{R}^2)$ with continuous paths.

This result is uniform in c on an interval, in $x \in \{0,1\}$, $w \in \operatorname{supp}(W)$, and in $\tau \in (0,1)$. This result directly implies convergence over $c \in \mathcal{C}$ as well. Unlike the distribution of the cdf bounds estimators, this process is Gaussian. This follows by Hadamard differentiability of the mapping between $\theta_0 \equiv (F_{Y|X,W}(\cdot \mid \cdot,\cdot), p_{(\cdot|\cdot)}, q_{(\cdot)})$ and the conditional quantile bounds.

Let the superscript $\mathbf{Z}^{(j)}$ denotes the jth component of the vector $\mathbf{Z}$. Since the CQTE bounds are functionals of these conditional quantile bounds, Lemma 2 implies the following convergence result:

$$\sqrt{N}\begin{pmatrix} \widehat{\overline{\text{CQTE}}}(\tau,c\mid w) - \overline{\text{CQTE}}(\tau,c\mid w) \\ \widehat{\underline{\text{CQTE}}}(\tau,c\mid w) - \underline{\text{CQTE}}(\tau,c\mid w) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \mathbf{Z}_3^{(1)}(\tau,1,w,c) - \mathbf{Z}_3^{(2)}(\tau,0,w,c) \\ \mathbf{Z}_3^{(2)}(\tau,1,w,c) - \mathbf{Z}_3^{(1)}(\tau,0,w,c) \end{pmatrix}.$$

Similarly, Lemma 2 also implies

$$\sqrt{N}\begin{pmatrix} \widehat{\overline{\text{CATE}}}(c\mid w) - \overline{\text{CATE}}(c\mid w) \\ \widehat{\underline{\text{CATE}}}(c\mid w) - \underline{\text{CATE}}(c\mid w) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \int_0^1 \big(\mathbf{Z}_3^{(1)}(u,1,w,c) - \mathbf{Z}_3^{(2)}(u,0,w,c)\big)\,du \\ \int_0^1 \big(\mathbf{Z}_3^{(2)}(u,1,w,c) - \mathbf{Z}_3^{(1)}(u,0,w,c)\big)\,du \end{pmatrix},$$

a mean-zero Gaussian process in $\ell^\infty(\operatorname{supp}(W)\times[0,\overline{C}],\mathbb{R}^2)$ with continuous paths.

Next consider the ATE bounds. The following decomposition implies that the estimated ATE upper bound converges weakly to a Gaussian element:

$$\begin{aligned} &\sqrt{N}\big(\widehat{\overline{\text{ATE}}}(c) - \overline{\text{ATE}}(c)\big) \\ &= \sqrt{N}\sum_{k=1}^K \widehat{q}_{w_k}\big(\widehat{\overline{\text{CATE}}}(c\mid w_k) - \overline{\text{CATE}}(c\mid w_k)\big) + \sum_{k=1}^K \overline{\text{CATE}}(c\mid w_k)\sqrt{N}(\widehat{q}_{w_k} - q_{w_k}) \\ &\rightsquigarrow \sum_{k=1}^K q_{w_k}\int_0^1 \big(\mathbf{Z}_3^{(1)}(u,1,w_k,c) - \mathbf{Z}_3^{(2)}(u,0,w_k,c)\big)\,du \\ &\quad + \sum_{k=1}^K \overline{\text{CATE}}(c\mid w_k)\mathbf{Z}_1^{(3)}(0,0,w_k). \end{aligned}$$

A similar result holds for the estimated ATE lower bound.

The mapping used in equation (20) is called the *pre-rearrangement operator*. Chernozhukov, Fernández-Val, and Galichon (2010) showed that this operator was Hadamard differentiable when the quantile functions are continuously differentiable for all $u \in (0,1)$. In our case, the underlying quantile functions are continuously differentiable on $(0,1/2)\cup(1/2,1)$, and continuous but not differentiable at $u = 1/2$. At this value, the left and right derivatives exist and are finite, but are generally different from one another. We extend the result of Chernozhukov, Fernández-Val, and Galichon (2010) to the case where the quantile function has a point of nondifferentiability by showing Hadamard directional differentiability of this mapping.

To do so, we make additional assumptions on the behavior of these quantile functions.

ASSUMPTION A6. *For each $c \in \mathcal{C}$ and $w \in \operatorname{supp}(W)$,*

1. *The number of elements in each of the sets*

$$\begin{aligned}\mathcal{U}_1^*(c \mid w) = \big\{u \in (0,1) : &\ \partial_u^-\big(\overline{Q}^c_{Y_1|W}(u \mid w) - \underline{Q}^c_{Y_0|W}(u \mid w)\big) = 0 \\ &\text{or } \partial_u^+\big(\overline{Q}^c_{Y_1|W}(u \mid w) - \underline{Q}^c_{Y_0|W}(u \mid w)\big) = 0\big\}, \\ \mathcal{U}_2^*(c \mid w) = \big\{u \in (0,1) : &\ \partial_u^-\big(\underline{Q}^c_{Y_1|W}(u \mid w) - \overline{Q}^c_{Y_0|W}(u \mid w)\big) = 0 \\ &\text{or } \partial_u^+\big(\underline{Q}^c_{Y_1|W}(u \mid w) - \overline{Q}^c_{Y_0|W}(u \mid w)\big) = 0\big\}\end{aligned}$$

is finite.

2. *The following hold*:

(a) *For any $u \in \mathcal{U}_1^*(c \mid w)$, $\overline{Q}^c_{Y_1|W}(u \mid w) - \underline{Q}^c_{Y_0|W}(u \mid w) \neq z$.*

(b) *For any $u \in \mathcal{U}_2^*(c \mid w)$, $\underline{Q}^c_{Y_1|W}(u \mid w) - \overline{Q}^c_{Y_0|W}(u \mid w) \neq z$.*

These assumptions imply that the respective function's derivatives change signs a finite number of times. Therefore they cross the horizontal line at z a finite number of times. These functions are continuously differentiable in u everywhere on $(0, 1/2) \cup (1/2, 1)$, and are directionally differentiable at $1/2$. The second assumption rules out the functions being flat when exactly valued at z. Failure of the second condition in this assumption implies that convergence will hold uniformly over any compact subset that excludes these values, which typically form a measure-zero set. Therefore this assumption can be satisfied by considering convergence for values of c which exclude those where the second part of Assumption A6 fails. Without knowing a priori at which values this assumption may fail, selecting grid points randomly from a continuous distribution ensures that these values are selected with probability zero.

An alternative approach to inference if the second condition fails for some values of c is to smooth the population function using methods described in Appendix G in the Online Supplemental Material. Like in Chernozhukov, Fernández-Val, and Galichon (2010, Corollary 4), we require a tuning parameters to control the level of smoothing. We show that $\sqrt{N}$-convergence holds for all parameter values when introducing any amount of fixed smoothing.

Finally, note that Assumption A6 is refutable, since it is expressed as a function of identified quantities, namely the CQTE bounds for all $u \in (0, 1)$.

With this additional assumption we can show $\sqrt{N}$-convergence of the bound estimators in equation (20) uniformly in $\operatorname{supp}(W) \times \mathcal{C}$.

LEMMA 3. *Suppose Assumptions A1, A3, A4, A5, and A6 hold. Then*

$$\sqrt{N}\begin{pmatrix}\widehat{\overline{P}}(c \mid w) - \overline{P}(c \mid w) \\ \widehat{\underline{P}}(c \mid w) - \underline{P}(c \mid w)\end{pmatrix} \rightsquigarrow \mathbf{Z}_4(w, c),$$

a tight random element in $\ell^\infty(\operatorname{supp}(W) \times \mathcal{C}, \mathbb{R}^2)$.

As discussed in Section 3, we use Lemma 3 to establish the following result.

Lemma 4. *Fix $z \in \mathbb{R}$. Suppose Assumptions A1, A3, A4, A5, and A6 hold. Then*

$$\sqrt{N}\begin{pmatrix} \widehat{\overline{\text{DTE}}}(z,c,t) - \overline{\text{DTE}}(z,c,t) \\ \widehat{\underline{\text{DTE}}}(z,c,t) - \underline{\text{DTE}}(z,c,t) \end{pmatrix} \rightsquigarrow \mathbf{Z}_5(c,t), \tag{29}$$

a tight random element of $\ell^\infty(\mathcal{C} \times [0,1], \mathbb{R}^2)$ with continuous paths.

B.2 *Estimation and inference with continuous covariates*

The estimation and inference theory in Section 3 assumes that the covariates W are discretely distributed (via Assumption A4). Those results are nonparametric in the sense that they do not impose any restrictions on the conditional distribution of $Y \mid X, W$ or on the propensity score $p_{x|w}$. But they rule out continuous covariates. In this section, we briefly discuss how to do estimation and inference with continuous covariates.

When some components of W are continuously distributed, a simple solution is to discretize W and then apply the previous estimator. Alternatively, one can smooth over different covariate values. This can be done using parametric, semiparametric, or nonparametric estimators.

For example, especially if the dimension of W is large, one could use the usual logit propensity score estimator

$$\widehat{p}_{1|w} = \widehat{\mathbb{P}}(X=1 \mid W=w) = \Lambda(\widehat{\beta}'w),$$

where $\Lambda(a) = \exp(a)/(1+\exp(a))$ is the standard logit cdf and $\widehat{\beta}$ are the maximum likelihood estimated index coefficients. The conditional quantile function $Q_{Y|X,W}(\tau \mid x, w)$ can be estimated by a linear quantile regression of Y on $(1, X, W)$, so that

$$\widehat{Q}_{Y|X,W}(\tau \mid x, w) = \widehat{\gamma}(\tau)' \begin{pmatrix} 1 \\ x \\ w \end{pmatrix},$$

where $\widehat{\gamma}(\tau)$ are estimated linear quantile regression coefficients.

Using these parametric estimators, define

$$\widehat{\overline{Q}}^c_{Y_x|W}(\tau \mid w) = \widehat{Q}_{Y|X,W}\left(\tau + \frac{c}{\widehat{p}_{x|w}} \min\{\tau, 1-\tau\} \mid x, w\right)$$

and

$$\widehat{\underline{Q}}^c_{Y_x|W}(\tau \mid w) = \widehat{Q}_{Y|X,W}\left(\tau - \frac{c}{\widehat{p}_{x|w}} \min\{\tau, 1-\tau\} \mid x, w\right)$$

as before. Since the asymptotic properties of $\widehat{\beta}$ and $\widehat{\gamma}(\cdot)$ are well known, it should be feasible to derive the asymptotic distribution of the functionals in Section 3. Alternatively, one could use semiparametric or nonparametric estimators of the propensity score $p_{x|w}$ and the conditional quantile function $Q_{Y|X,W}$. Again, such first step estimators can be plug-ins to obtain estimates of the various bounds we consider in Section 3. We leave a full analysis of the asymptotic properties of these estimators to future work.

B.3 *Choosing the grid points* $\mathcal{C}$

Here we suggest two simple approaches for choosing the number and location of the grid points $\mathcal{C} = \{c_1, \dots, c_J\}$. First, one can let $\mathcal{C} = \{\overline{c}_1, \dots, \overline{c}_K\}$ where K is the number of observed covariates and, for each $k \in \{1, \dots, K\}$, $\overline{c}_k$ is the maximal deviation between the observed propensity score and the "leave out variable k" propensity score, which we define and discuss in our empirical illustration on page 68. There we argue that these are natural points to consider.

Second, researchers can choose equally spaced grid points, for a fixed J. This approach can be used in combination with the first approach. In this case, researchers may want to begin with $\{\overline{c}_1, \dots, \overline{c}_K\}$ and then add a multiple m of K additional points, so that the total number of points J is $K + mK$ for some positive integer m.

In practice one could choose J to be the minimum of 100 and some small proportion of the sample size, like 5%. Values larger than 100 are not likely to affect the appearance of plots like Figure 3. This is just a rough rule of thumb, however. Unfortunately we are not aware of any clear data-driven way of picking the number of grid points. One option is to preestimate the points of nondifferentiability and then pick the grid points sufficiently far from these points of nondifferentiability. Horowitz and Lee (2012, 2017) discussed approaches like this in different settings. This approach, however, requires choosing a tuning parameter which defines what it means to be "too far" from the points of nondifferentiability, and it is not clear how to pick that parameter. Alternatively, we can avoid selecting a grid by using the smoothing approach described in Appendix G in the Online Supplemental Material. This approach requires choosing several smoothing parameters, however. Until future research provides a clear best choice, we tentatively recommend that researchers use the simple rule of thumb discussed above.

B.4 *Bootstrap selection of* ε_N

While Dümbgen (1993) and Hong and Li (2018) provide rate constraints on ε_N, they do not recommend a procedure for picking ε_N in practice. In this section, we suggest a heuristic bootstrap method for picking ε_N. We use this method for our empirical illustration in Section 4; we also present the full range of bands considered. Since the question of choosing ε_N goes beyond the purpose of the present paper, we defer a formal analysis of this method to future research. For discussions of bootstrap selection of tuning parameters in other problems, see Taylor (1989), Léger and Romano (1990), Marron (1992), and Cao, Cuevas, and Manteiga (1994).

Fix a $\underline{p}$. Let $\mathrm{CP}_N(\varepsilon; F_{Y,X,W})$ denote the finite sample coverage probability of our confidence band as described above, for a fixed ε. This statistic depends on the unknown distribution of the data, $F_{Y,X,W}$. The bootstrap replaces $F_{Y,X,W}$ with an estimator $\widehat{F}_{Y,X,W}$. We pick a grid $\{\varepsilon_1, \dots, \varepsilon_L\}$ of ε's and let $\widehat{\varepsilon}_N$ solve

$$\min_{\ell=1,\dots,L} \left|\mathrm{CP}_N(\varepsilon_\ell; \widehat{F}_{Y,X,W}) - (1-\alpha)\right|.$$

We compute CP_N by simulation. In our empirical illustration, we take $B = 500$ draws. We use the same grid of ε's as in our Monte Carlo simulations in Appendix F in the Online Supplemental Material. Larger grids and larger values of B can be chosen subject to computational constraints. Note that the rate conditions on the tuning parameter will automatically be satisfied for $\widehat{\varepsilon}_N$ if our initial grid satisfies $\varepsilon_1 \leq \cdots \leq \varepsilon_L$ with $\varepsilon_1\sqrt{N} \to \infty$ and $\varepsilon_L \to 0$ as $N \to \infty$.

We furthermore must choose an estimator $\widehat{F}_{Y,X,W}$. The nonparametric bootstrap uses the empirical distribution. We use the smoothed bootstrap (De Angelis and Young (1992), Polansky and Schucany (1997)). Specifically, we estimate the distribution of (X, W) by its empirical distribution. We then let $\widehat{F}_{Y|X,W}$ be a kernel smoothed cdf estimate of the conditional cdf of $Y|X, W$. We use the standard logistic cdf kernel and the method proposed by Hansen (2004) to choose the smoothing bandwidths. We divide these bandwidths in half since this visually appears to better capture the shape of the conditional empirical cdfs, and since smaller order bandwidths are recommended for the smoothed bootstrap (Section 4 of De Angelis and Young (1992)).

Bootstrap consistency requires sufficient smoothness of the functional of interest in the underlying cdf. It may be that the lack of smoothness that requires us to use the methods of Fang and Santos (2019) and Hong and Li (2018) in the first place also cause the naive bootstrap to be inconsistent for approximating the distribution of $\mathrm{CP}_N(\varepsilon; F_{Y,X,W})$. As mentioned earlier, formally investigating this issue is beyond the scope of this paper. Our goal here is merely to suggest a simple first-pass approach at choosing ε_N.

Appendix C: Proofs

Proofs for Section 2

Proof of Theorem 1. Let $F_1(\cdot \mid w)$ and $F_0(\cdot \mid w)$ be any strictly increasing cdfs conditional on $W = w$ for any $w \in \operatorname{supp}(W)$. Suppose $(Y_1, Y_0) \mid W$ have joint cdf

$$F_{Y_1,Y_0|W}(y_1, y_0 \mid w) = C\big(F_1(y_1 \mid w), F_0(y_0 \mid w) \mid w\big).$$

Then

$$\begin{aligned}
\mathbb{P}(Y_1 - Y_0 \leq z \mid W = w) &= \int_{\{y_1 - y_0 \leq z\}} dC\big(F_1(y_1 \mid w), F_0(y_0 \mid w) \mid w\big) \\
&= (1-t) \int_{\{y_1 - y_0 \leq z\}} dM\big(F_1(y_1 \mid w), F_0(y_0 \mid w)\big) \\
&\quad + t \int_{\{y_1 - y_0 \leq z\}} dH\big(F_1(y_1 \mid w), F_0(y_0 \mid w) \mid w\big),
\end{aligned}$$

where $M(u_1, u_0) = \min\{u_1, u_0\}$.

For fixed distributions $(F_1(\cdot \mid w), F_0(\cdot \mid w))$, the first integral is the probability that $\{Y_1 - Y_0 \leq z\}$ given $W = w$, where $(Y_1, Y_0) \mid W$ are random variables that satisfy conditional rank invariance. Hence for these random variables, the corresponding conditional ranks are equal almost surely: $U_1 = U_0$ a.s. Let $U \sim \mathrm{Unif}[0, 1]$ denote this almost sure common random variable. Using Assumption A1.1, we can thus write

$$(Y_1, Y_0) \mid W \stackrel{d}{=} \big(F_1^{-1}(U \mid W), F_0^{-1}(U \mid W)\big)$$

and therefore

$$\int_{\{y_1-y_0\leq z\}} dM\big(F_1(y_1 \mid w), F_0(y_0 \mid w)\big) = \mathbb{P}\big(F_1^{-1}(U \mid w) - F_0^{-1}(U \mid w) \leq z\big).$$

Makarov (1982) derived sharp bounds on

$$\int_{\{y_1-y_0\leq z\}} dH\big(F_1(y_1), F_0(y_0) \mid w\big).$$

Applying these bounds yields

$$\int_{\{y_1-y_0\leq z\}} dH\big(F_1(y_1 \mid w), F_0(y_0 \mid w)\big) \in \Big[\max\Big\{\sup_{y\in\mathcal{Y}_z(w)}\big(F_1(y \mid w) - F_0(y-z \mid w)\big), 0\Big\},$$
$$1 + \min\Big\{\inf_{y\in\mathcal{Y}_z(w)}\big(F_1(y \mid w) - F_0(y-z \mid w)\big), 0\Big\}\Big].$$

Therefore, for given $w \in \operatorname{supp}(W)$ and given $(F_1(\cdot \mid w), F_0(\cdot \mid w))$, sharp bounds for $\mathbb{P}(Y_1 - Y_0 \leq z \mid W = w)$ are given by

$$\big[\underline{\theta}\big(F_1(\cdot \mid w), F_0(\cdot \mid w)\big),\ \overline{\theta}\big(F_1(\cdot \mid w), F_0(\cdot \mid w)\big)\big],$$

where

$$\underline{\theta}\big(F_1(\cdot \mid w), F_0(\cdot \mid w)\big) = (1-t)\mathbb{P}\big(F_1^{-1}(U \mid w) - F_0^{-1}(U \mid w) \leq z\big)$$
$$+ t\max\Big\{\sup_{y\in\mathcal{Y}_z(w)}\big(F_1(y \mid w) - F_0(y-z \mid w)\big), 0\Big\}$$

and

$$\overline{\theta}\big(F_1(\cdot \mid w), F_0(\cdot \mid w)\big) = (1-t)\mathbb{P}\big(F_1^{-1}(U \mid w) - F_0^{-1}(U \mid w) \leq z\big)$$
$$+ t\Big(1 + \min\Big\{\inf_{y\in\mathcal{Y}_z(w)}\big(F_1(y \mid w) - F_0(y-z \mid w)\big), 0\Big\}\Big).$$

Define the first-order stochastic dominance ordering as follows: For two cdfs F and G, let $F \preceq_{\text{FSD}} G$ if $F(t) \geq G(t)$ for all $t \in \mathbb{R}$. All of the following statements refer to this ordering. For any fixed $F_1(\cdot \mid w)$,

$$\tilde{F}_0(\cdot \mid w) \preceq_{\text{FSD}} F_0(\cdot \mid w) \quad \text{implies} \quad \underline{\theta}\big(F_1(\cdot \mid w), \tilde{F}_0(\cdot \mid w)\big) \leq \underline{\theta}\big(F_1(\cdot \mid w), F_0(\cdot \mid w)\big).$$

That is, the lower bound function $\underline{\theta}(F_1(\cdot \mid w), F_0(\cdot \mid w))$ is weakly increasing in $F_0(\cdot \mid w)$. This can be shown in two steps. First, the expression

$$\mathbb{P}\big(F_1^{-1}(U \mid w) - F_0^{-1}(U \mid w) \leq z\big)$$

is weakly increasing in $F_0(\cdot \mid w)$ since, for $\tilde{F}_0(\cdot \mid w) \preceq_{\text{FSD}} F_0(\cdot \mid w)$, we have $\tilde{F}_0^{-1}(u \mid w) \leq F_0^{-1}(u \mid w)$ for $u \in (0,1)$, and therefore

$$\mathbb{P}\big(F_1^{-1}(U \mid w) - \tilde{F}_0^{-1}(U \mid w) \leq z\big) \leq \mathbb{P}\big(F_1^{-1}(U \mid w) - F_0^{-1}(U \mid w) \leq z\big).$$

Second, the expression

$$\max\Big\{\sup_{y\in\mathcal{Y}_z(w)}\big(F_1(y\mid w)-F_0(y-z\mid w)\big),0\Big\}$$

is weakly increasing in $F_0(\cdot\mid w)$ since the supremum and maximum operators are weakly increasing. Thus both components of $\underline{\theta}$ are weakly increasing in $F_0(\cdot\mid w)$. Therefore, their linear combination is also weakly increasing in $F_0(\cdot\mid w)$.

We can similarly show that $\underline{\theta}(F_1(\cdot\mid w),F_0(\cdot\mid w))$ is weakly decreasing in $F_1(\cdot\mid w)$. Thus substituting

$$\big(F_1(\cdot\mid w),F_0(\cdot\mid w)\big)=\big(\underline{F}^c_{Y_1|W}(\cdot\mid w),\overline{F}^c_{Y_0|W}(\cdot\mid w)\big)$$

yields the lower bound $\underline{\text{CDTE}}(z,c,t\mid w)$. The upper bound function $\overline{\theta}(F_1(\cdot\mid w),F_0(\cdot\mid w))$ is also weakly increasing in $F_0(\cdot\mid w)$ and weakly decreasing in $F_1(\cdot\mid w)$. Thus substituting

$$\big(F_1(\cdot\mid w),F_0(\cdot\mid w)\big)=\big(\overline{F}^c_{Y_1|W}(\cdot\mid w),\underline{F}^c_{Y_0|W}(\cdot\mid w)\big)$$

yields the upper bound $\overline{\text{CDTE}}(z,c,t\mid w)$. In making these substitutions we applied Proposition 2 from Masten and Poirier (2018a). In that paper we defined functions $F^c_{Y_x|W}(\cdot\mid w;\epsilon,\eta)$, which we now use to show sharpness of the DTE bounds.

Substitute

$$\big(F^c_{Y_1|W}(\cdot\mid w;\epsilon,0),F^c_{Y_0|W}(\cdot\mid w;1-\epsilon,0)\big)$$

into the bound functionals and continuously vary ϵ between [0, 1]. Note that we let $\eta=0$ since $c<\min\{p_{1|w},p_{0|w}\}$. By continuity of $\underline{\theta}(\cdot,\cdot)$ and $\overline{\theta}(\cdot,\cdot)$ in their arguments and continuity of $(F^c_{Y_1|W}(\cdot\mid w;\epsilon,0),F^c_{Y_0|W}(\cdot\mid w;1-\epsilon,0))$ in ϵ, the intermediate value theorem implies that every element between the bounds can be attained.

By integrating these CDTE bounds over the marginal distribution of W, we obtain the DTE bounds:

$$\begin{aligned}&\big[\underline{\text{DTE}}(z,c,t),\overline{\text{DTE}}(z,c,t)\big]\\&\quad=\bigg[\int_{\text{supp}(W)}\underline{\text{CDTE}}(z,c,t\mid w)\,dF_W(w),\int_{\text{supp}(W)}\overline{\text{CDTE}}(z,c,t\mid w)\,dF_W(w)\bigg].\end{aligned}$$

Sharpness of these bounds results from the sharpness of the CDTE bounds for every $w\in\text{supp}(W)$ and the joint attainability of

$$\big\{\big(\underline{F}^c_{Y_1|W}(\cdot\mid w),\overline{F}^c_{Y_0|W}(\cdot\mid w)\big):w\in\text{supp}(W)\big\}$$

and of

$$\big\{\big(\overline{F}^c_{Y_1|W}(\cdot\mid w),\underline{F}^c_{Y_0|W}(\cdot\mid w)\big):w\in\text{supp}(W)\big\}.\qquad\square$$

Proofs for Sections 3 and B.1

The following lemma shows that $(\widehat{F}_{Y|X,W}(\cdot\mid\cdot,\cdot),\widehat{p}_{(\cdot|\cdot)},\widehat{q}_{(\cdot)})$ converges uniformly in y, x, and w to a mean-zero Gaussian process. This result follows by applying the delta method.

Lemma C1. *Suppose Assumption A3 and Assumption A4 hold. Then*

$$\sqrt{N}\begin{pmatrix}\widehat{F}_{Y|X,W}(y \mid x,w)-F_{Y|X,W}(y \mid x,w)\\ \widehat{p}_{x|w}-p_{x|w}\\ \widehat{q}_w-q_w\end{pmatrix}\rightsquigarrow \mathbf{Z}_1(y,x,w),$$

a mean-zero Gaussian process in $\ell^\infty(\mathbb{R}\times\{0,1\}\times\operatorname{supp}(W),\mathbb{R}^3)$ *with continuous paths and covariance kernel equal to*

$$\begin{aligned}&\boldsymbol{\Sigma}_1(y,x,w,\tilde{y},\tilde{x},\tilde{w})\\ &\quad=\mathbb{E}\big[\mathbf{Z}_1(y,x,w)\mathbf{Z}_1(\tilde{y},\tilde{x},\tilde{w})'\big]\\ &\quad=\operatorname{diag}\begin{pmatrix}\dfrac{F_{Y|X,W}\big(\min\{y,\tilde{y}\}\mid x,w\big)-F_{Y|X,W}(y\mid x,w)F_{Y|X,W}(\tilde{y}\mid x,w)}{p_{x|w}q_w}\mathbb{1}(x=\tilde{x},w=\tilde{w})\\ \dfrac{p_{x|w}}{q_w}\mathbb{1}(x=\tilde{x},w=\tilde{w})-\dfrac{p_{x|w}p_{\tilde{x}|w}}{q_w}\mathbb{1}(w=\tilde{w})\\ q_w\mathbb{1}(w=\tilde{w})-q_wq_{\tilde{w}}\end{pmatrix}.\end{aligned}$$

Proof of Lemma C1. By a second-order Taylor expansion,

$$\begin{aligned}&\widehat{F}_{Y|X,W}(y\mid x,w)-F_{Y|X,W}(y\mid x,w)\\ &\quad=\frac{\dfrac{1}{N}\displaystyle\sum_{i=1}^N\mathbb{1}(Y_i\le y)\mathbb{1}(X_i=x,W_i=w)}{\dfrac{1}{N}\displaystyle\sum_{i=1}^N\mathbb{1}(X_i=x,W_i=w)}-\frac{\mathbb{P}(Y\le y,X=x,W=w)}{\mathbb{P}(X=x,W=w)}\\ &\quad=\frac{\dfrac{1}{N}\displaystyle\sum_{i=1}^N\mathbb{1}(Y_i\le y)\mathbb{1}(X_i=x,W_i=w)-\mathbb{P}(Y\le y,X=x,W=w)}{\mathbb{P}(X=x,W=w)}\\ &\qquad-\frac{F_{Y|X,W}(y\mid x,w)}{\mathbb{P}(X=x,W=w)}\left(\frac{1}{N}\sum_{i=1}^N\mathbb{1}(X_i=x,W_i=w)-\mathbb{P}(X=x,W=w)\right)\\ &\qquad+O_p\left[\left(\frac{1}{N}\sum_{i=1}^N\mathbb{1}(Y_i\le y)\mathbb{1}(X_i=x,W_i=w)-F_{Y|X,W}(y\mid x,w)\mathbb{P}(X=x,W=w)\right)\right.\\ &\qquad\qquad\left.\cdot\left(\frac{1}{N}\sum_{i=1}^N\mathbb{1}(X_i=x,W_i=w)-\mathbb{P}(X=x,W=w)\right)\right]\\ &\qquad+O_p\left[\left(\frac{1}{N}\sum_{i=1}^N\mathbb{1}(X_i=x,W_i=w)-\mathbb{P}(X=x,W=w)\right)^2\right].\end{aligned}$$

By standard bracketing entropy results (e.g., example 19.6 on p. 271 of van der Vaart (2000)) the function classes $\{\mathbb{1}(Y\le y)\mathbb{1}(X=x)\mathbb{1}(W=w): y\in\mathbb{R}, x\in\{0,1\}, w\in\operatorname{supp}(W)\}$

and $\{\mathbb{1}(X=x)\mathbb{1}(W=w) : x \in \{0,1\}, w \in \operatorname{supp}(W)\}$ are both P-Donsker. Hence the residual is of order $O_p(N^{-1})$ uniformly in $(y,x,w) \in \mathbb{R} \times \{0,1\} \times \operatorname{supp}(W)$. Combining this with Slutsky's theorem, we get the uniform over y, x, and w asymptotically linear representation

$$\begin{aligned}
&\widehat{F}_{Y|X,W}(y \mid x, w) - F_{Y|X,W}(y \mid x, w) \\
&\quad = \frac{1}{N}\sum_{i=1}^{N} \frac{\mathbb{1}(X_i = x, W_i = w)\big(\mathbb{1}(Y_i \le y) - F_{Y|X,W}(y \mid x, w)\big)}{\mathbb{P}(X = x, W = w)} + o_p\big(N^{-1/2}\big).
\end{aligned}$$

By the same bracketing entropy arguments, the class

$$\left\{\frac{\mathbb{1}(X = x, W = w)\big(\mathbb{1}(Y \le y) - F_{Y|X,W}(y \mid x, w)\big)}{\mathbb{P}(X = x, W = w)} : y \in \mathbb{R}, x \in \{0,1\}, w \in \operatorname{supp}(W)\right\}$$

is P-Donsker, and hence $\sqrt{N}(\widehat{F}_{Y|X,W}(\cdot \mid \cdot, \cdot) - F_{Y|X,W}(\cdot \mid \cdot, \cdot))$ converges in distribution to a mean-zero Gaussian process with continuous paths.

A similar argument yields the asymptotically linear representations

$$\widehat{p}_{x|w} - p_{x|w} = \frac{1}{N}\sum_{i=1}^{N} \frac{\mathbb{1}(W_i = w)\big(\mathbb{1}(X_i = x) - p_{x|w}\big)}{q_w} + o_p\big(N^{-1/2}\big)$$

and

$$\widehat{q}_w - q_w = \frac{1}{N}\sum_{i=1}^{N}\big(\mathbb{1}(W_i = w) - q_w\big).$$

The covariance kernel $\boldsymbol{\Sigma}_1$ can be calculated as follows:

$$\begin{aligned}
&\big[\boldsymbol{\Sigma}_1(y, x, w, \tilde{y}, \tilde{x}, \tilde{w})\big]_{1,1} \\
&\quad = \mathbb{E}\left[\frac{\mathbb{1}(X_i = x, W_i = w)\mathbb{1}(X_i = \tilde{x}, W_i = \tilde{w})\big(\mathbb{1}(Y_i \le y) - F_{Y|X,W}(y \mid x, w)\big)\big(\mathbb{1}(Y_i \le \tilde{y}) - F_{Y|X,W}(\tilde{y} \mid \tilde{x}, \tilde{w})\big)}{\mathbb{P}(X = x, W = w)\mathbb{P}(X = \tilde{x}, W = \tilde{w})}\right] \\
&\quad = \frac{F_{Y|X,W}\big(\min\{y, \tilde{y}\} \mid x, w\big) - F_{Y|X,W}(y \mid x, w)F_{Y|X,W}(\tilde{y} \mid x, w)}{p_{x|w}q_w}\mathbb{1}(x = \tilde{x}, w = \tilde{w}), \\
&\big[\boldsymbol{\Sigma}_1(y, x, w, \tilde{y}, \tilde{x}, \tilde{w})\big]_{1,2} \\
&\quad = \mathbb{E}\left[\frac{\mathbb{1}(X_i = x, W_i = \tilde{w})\big(\mathbb{1}(X_i = \tilde{x}) - p_{\tilde{x}|\tilde{w}}\big)\big(\mathbb{1}(Y_i \le y) - F_{Y|X,W}(y \mid x, w)\big)}{p_{x|w}q_w q_{\tilde{w}}}\right] = 0, \\
&\big[\boldsymbol{\Sigma}_1(y, x, w, \tilde{y}, \tilde{x}, \tilde{w})\big]_{1,3} \\
&\quad = \mathbb{E}\left[\frac{\big(\mathbb{1}(W_i = \tilde{w}) - q_{\tilde{w}}\big)\mathbb{1}(X_i = x, W_i = w)\big(\mathbb{1}(Y_i \le y) - F_{Y|X,W}(y \mid x, w)\big)}{\mathbb{P}(X = x, W = w)}\right] = 0, \\
&\big[\boldsymbol{\Sigma}_1(y, x, w, \tilde{y}, \tilde{x}, \tilde{w})\big]_{2,1} = \big[\boldsymbol{\Sigma}_1(\tilde{y}, \tilde{x}, \tilde{w}, y, x, w)\big]_{1,2} = 0,
\end{aligned}$$

$$
\begin{aligned}
&\big[\boldsymbol{\Sigma}_1(y,x,w,\tilde{y},\tilde{x},\tilde{w})\big]_{2,2} \\
&\quad = \mathbb{E}\bigg[\frac{\mathbb{1}(W_i = w)\mathbb{1}(W_i = \tilde{w})\big(\mathbb{1}(X_i = x) - p_{x|w}\big)\big(\mathbb{1}(X_i = \tilde{x}) - p_{\tilde{x}|\tilde{w}}\big)}{q_w q_{\tilde{w}}}\bigg] \\
&\quad = \frac{p_{x|w}}{q_w}\mathbb{1}(x = \tilde{x}, w = \tilde{w}) - \frac{p_{x|w}p_{\tilde{x}|w}}{q_w}\mathbb{1}(w = \tilde{w}), \\
&\big[\boldsymbol{\Sigma}_1(y,x,w,\tilde{y},\tilde{x},\tilde{w})\big]_{2,3} = \mathbb{E}\bigg[\frac{\big(\mathbb{1}(W_i = \tilde{w}) - q_{\tilde{w}}\big)\big(\mathbb{1}(X_i = x) - p_{x|w}\big)}{q_w}\bigg] = 0, \\
&\big[\boldsymbol{\Sigma}_1(y,x,w,\tilde{y},\tilde{x},\tilde{w})\big]_{3,1} = \big[\boldsymbol{\Sigma}_1(\tilde{y},\tilde{x},\tilde{w},y,x,w)\big]_{1,3} = 0, \\
&\big[\boldsymbol{\Sigma}_1(y,x,w,\tilde{y},\tilde{x},\tilde{w})\big]_{3,2} = \big[\boldsymbol{\Sigma}_1(\tilde{y},\tilde{x},\tilde{w},y,x,w)\big]_{2,3} = 0, \\
&\big[\boldsymbol{\Sigma}_1(y,x,w,\tilde{y},\tilde{x},\tilde{w})\big]_{3,3} = \mathbb{E}\big[\big(\mathbb{1}(W_i = w) - q_w\big)\big(\mathbb{1}(W_i = \tilde{w}) - q_{\tilde{w}}\big)\big] \\
&\qquad\qquad\qquad\qquad\qquad\quad = q_w \mathbb{1}(w = \tilde{w}) - q_w q_{\tilde{w}}. \qquad \square
\end{aligned}
$$

Lemma C2 (Chain rule for Hadamard directionally differentiable functions). *Let $\mathbb{D}$, $\mathbb{E}$, and $\mathbb{F}$ be Banach spaces with norms $\|\cdot\|_{\mathbb{D}}$, $\|\cdot\|_{\mathbb{E}}$, and $\|\cdot\|_{\mathbb{F}}$. Let $\mathbb{D}_\phi \subseteq \mathbb{D}$ and $\mathbb{E}_\psi \subseteq \mathbb{E}$. Let $\phi : \mathbb{D}_\phi \to \mathbb{E}_\psi$ and $\psi : \mathbb{E}_\psi \to \mathbb{F}$ be functions. Let $\theta \in \mathbb{D}_\phi$ and ϕ be Hadamard directionally differentiable at θ tangentially to $\mathbb{D}_0 \subseteq \mathbb{D}$. Let ψ be Hadamard directionally differentiable at $\phi(\theta)$ tangentially to the range $\phi'_\theta(\mathbb{D}_0) \subseteq \mathbb{E}_\psi$. Then $\psi \circ \phi : \mathbb{D}_\phi \to \mathbb{F}$ is Hadamard directionally differentiable at θ tangentially to $\mathbb{D}_0$ with Hadamard directional derivative equal to $\psi'_{\phi(\theta)} \circ \phi'_\theta$.*

This result is a version of proposition 3.6 in Shapiro (1990), who omits the proof. We give the proof here because this result is key to our paper.

Proof of Lemma C2. Let $\{h_n\}_{n\geq 1}$ be in $\mathbb{D}$ and $h_n \to h \in \mathbb{D}_0$. By Hadamard directional differentiability of ϕ tangentially to $\mathbb{D}_0$,

$$
\bigg\| \frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} - \phi'_\theta(h) \bigg\|_{\mathbb{E}} = o(1)
$$

as $n \to \infty$ for any $t_n \searrow 0$. That is,

$$
g_n \equiv \frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} \xrightarrow{\mathbb{E}} \phi'_\theta(h) = g,
$$

where $\phi'_\theta \in \phi'_\theta(\mathbb{D}_0)$. Therefore, by Hadamard directional differentiability of ψ, we have

$$
\frac{\psi\big(\phi(\theta + t_n h_n)\big) - \psi\big(\phi(\theta)\big)}{t_n} = \frac{\psi\big(\phi(\theta) + t_n g_n\big) - \psi\big(\phi(\theta)\big)}{t_n} \xrightarrow{\mathbb{F}} \psi'_{\phi(\theta)}(g) = \psi'_{\phi(\theta)}\big(\phi'_\theta(h)\big).
$$

By Hadamard directional differentiability of ϕ at θ and ψ at $\phi(\theta)$, ϕ'_θ and $\psi'_{\phi(\theta)}$ are continuous mappings. Hence their composition $\psi'_{\phi(\theta)} \circ \phi'_\theta$ is continuous. This combined with our derivations above imply that $\psi \circ \phi$ is Hadamard directionally differentiable tangentially to $\mathbb{D}_0$ at θ. $\square$

Proof of Lemma 1. Let $\theta_0 = (F_{Y|X,W}(\cdot \mid \cdot, \cdot), p_{(\cdot|\cdot)}, q_{(\cdot)})$ and $\widehat{\theta} = (\widehat{F}_{Y|X,W}(\cdot \mid \cdot, \cdot), \widehat{p}_{(\cdot|\cdot)}, \widehat{q}_{(\cdot)})$. For fixed y and c, define the mapping

$$\begin{aligned}\phi_1 : \ell^\infty\big(\mathbb{R} \times \{0,1\} \times \operatorname{supp}(W)\big) \times \ell^\infty\big(\{0,1\} \times \operatorname{supp}(W)\big) \times \ell^\infty\big(\operatorname{supp}(W)\big) \\ \to \ell^\infty\big(\{0,1\} \times \operatorname{supp}(W), \mathbb{R}^2\big)\end{aligned}$$

by

$$\big[\phi_1(\theta)\big](x,w) = \begin{pmatrix} \min\left\{ \dfrac{\theta^{(1)}(y,x,w)\theta^{(2)}(x,w)}{\theta^{(2)}(x,w) - c}, \dfrac{\theta^{(1)}(y,x,w)\theta^{(2)}(x,w) + c}{\theta^{(2)}(x,w) + c} \right\} \\ \max\left\{ \dfrac{\theta^{(1)}(y,x,w)\theta^{(2)}(x,w)}{\theta^{(2)}(x,w) + c}, \dfrac{\theta^{(1)}(y,x,w)\theta^{(2)}(x,w) - c}{\theta^{(2)}(x,w) - c} \right\} \end{pmatrix},$$

where $\theta^{(j)}$ is the jth component of θ. Note that

$$\begin{aligned}\begin{pmatrix} \overline{F}^c_{Y_x|W}(y \mid w) \\ \underline{F}^c_{Y_x|W}(y \mid w) \end{pmatrix} &= \big[\phi_1(\theta_0)\big](x,w), \\ \begin{pmatrix} \widehat{\overline{F}}^c_{Y_x|W}(y \mid w) \\ \widehat{\underline{F}}^c_{Y_x|W}(y \mid w) \end{pmatrix} &= \big[\phi_1(\widehat{\theta})\big](x,w).\end{aligned}$$

The maps $(a_1, a_2) \mapsto \min\{a_1, a_2\}$ and $(a_1, a_2) \mapsto \max\{a_1, a_2\}$ are Hadamard directionally differentiable with Hadamard directional derivatives at (a_1, a_2) equal to

$$h \mapsto \begin{cases} h^{(1)} & \text{if } a_1 < a_2, \\ \min\big\{h^{(1)}, h^{(2)}\big\} & \text{if } a_1 = a_2, \\ h^{(2)} & \text{if } a_1 > a_2 \end{cases}$$

and

$$h \mapsto \begin{cases} h^{(2)} & \text{if } a_1 < a_2, \\ \max\big\{h^{(1)}, h^{(2)}\big\} & \text{if } a_1 = a_2, \\ h^{(1)} & \text{if } a_1 > a_2 \end{cases}$$

respectively, where $h \in \mathbb{R}^2$; for example, see equation (18) in Fang and Santos (2015). The mapping ϕ_1 is comprised of compositions of these min and max operators, along with four other functions. We can show that these four mappings are ordinary Hadamard differentiable. Here, we compute these Hadamard derivatives with respect to θ:

$$\big[\delta_1(\theta)\big](x,w) = \frac{\theta^{(1)}(y,x,w)\theta^{(2)}(x,w)}{\theta^{(2)}(x,w) + c} \quad \text{has Hadamard derivative equal to}$$

$$\begin{aligned}\big[\delta'_{1,\theta}(h)\big](x,w) = {} & \frac{\theta^{(1)}(y,x,w)h^{(2)}(x,w) + h^{(1)}(y,x,w)\theta^{(2)}(x,w)}{\theta^{(2)}(x,w) + c} \\ & - \frac{\theta^{(1)}(y,x,w)\theta^{(2)}(x,w)h^{(2)}(x,w)}{\big(\theta^{(2)}(x,w) + c\big)^2},\end{aligned}$$

$$\big[\delta_2(\theta)\big](x,w)=\frac{\theta^{(1)}(y,x,w)\theta^{(2)}(x,w)-c}{\theta^{(2)}(x,w)-c} \quad \text{has Hadamard derivative equal to}$$

$$\begin{aligned}\big[\delta'_{2,\theta}(h)\big](x,w)=&\ \frac{\theta^{(1)}(y,x,w)h^{(2)}(x,w)+h^{(1)}(y,x,w)\theta^{(2)}(x,w)}{\theta^{(2)}(x,w)-c}\\&-\frac{\big(\theta^{(1)}(y,x,w)\theta^{(2)}(x,w)-c\big)h^{(2)}(x,w)}{\big(\theta^{(2)}(x,w)-c\big)^2},\end{aligned}$$

$$\big[\delta_3(\theta)\big](x,w)=\frac{\theta^{(1)}(y,x,w)\theta^{(2)}(x,w)}{\theta^{(2)}(x,w)-c} \quad \text{has Hadamard derivative equal to}$$

$$\begin{aligned}\big[\delta'_{3,\theta}(h)\big](x,w)=&\ \frac{\theta^{(1)}(y,x,w)h^{(2)}(x,w)+h^{(1)}(y,x,w)\theta^{(2)}(x,w)}{\theta^{(2)}(x,w)-c}\\&-\frac{\theta^{(1)}(y,x,w)\theta^{(2)}(x,w)h^{(2)}(x,w)}{\big(\theta^{(2)}(x,w)-c\big)^2},\end{aligned}$$

$$\big[\delta_4(\theta)\big](x,w)=\frac{\theta^{(1)}(y,x,w)\theta^{(2)}(x,w)+c}{\theta^{(2)}(x,w)+c} \quad \text{has Hadamard derivative equal to}$$

$$\begin{aligned}\big[\delta'_{4,\theta}(h)\big](x,w)=&\ \frac{\theta^{(1)}(y,x,w)h^{(2)}(x,w)+h^{(1)}(y,x,w)\theta^{(2)}(x,w)}{\theta^{(2)}(x,w)+c}\\&-\frac{\big(\theta^{(1)}(y,x,w)\theta^{(2)}(x,w)+c\big)h^{(2)}(x,w)}{\big(\theta^{(2)}(x,w)+c\big)^2}.\end{aligned}$$

All these derivatives are well defined at θ_0 because $\theta_0^{(2)}(x,w)=p_{x|w}>\overline{C}\geq c$. With this notation, we can write the functional ϕ_1 as

$$\phi_1(\theta)=\begin{pmatrix}\min\big\{\delta_3(\theta),\delta_4(\theta)\big\}\\ \max\big\{\delta_1(\theta),\delta_2(\theta)\big\}\end{pmatrix}.$$

By the chain rule (Lemma C2), the map ϕ_1 is Hadamard directionally differentiable at θ_0 with Hadamard directional derivative evaluated at θ_0 equal to

$$\phi'_{1,\theta_0}(h)=\begin{pmatrix}\mathbb{1}\big(\delta_3(\theta_0)<\delta_4(\theta_0)\big)\cdot\delta'_{4,\theta_0}(h)\\ \quad+\mathbb{1}\big(\delta_3(\theta_0)=\delta_4(\theta_0)\big)\cdot\min\big\{\delta'_{3,\theta_0}(h),\delta'_{4,\theta_0}(h)\big\}\\ \quad+\mathbb{1}\big(\delta_3(\theta_0)>\delta_4(\theta_0)\big)\cdot\delta'_{3,\theta_0}(h)\\ \mathbb{1}\big(\delta_1(\theta_0)<\delta_2(\theta_0)\big)\cdot\delta'_{1,\theta_0}(h)\\ \quad+\mathbb{1}\big(\delta_1(\theta_0)=\delta_2(\theta_0)\big)\cdot\max\big\{\delta'_{1,\theta_0}(h),\delta'_{2,\theta_0}(h)\big\}\\ \quad+\mathbb{1}\big(\delta_1(\theta_0)>\delta_2(\theta_0)\big)\cdot\delta'_{2,\theta_0}(h)\end{pmatrix}.$$

By Lemma C1, $\sqrt{N}(\widehat{\theta}(y,x,w)-\theta_0(y,x,w))\rightsquigarrow\mathbf{Z}_1(y,x,w)$. Hence we can use the delta method for Hadamard directionally differentiable functions (see Theorem 2.1 in Fang and Santos (2019)) to find that

$$\begin{aligned}\left[\sqrt{N}\big(\phi_1(\widehat{\theta}) - \phi_1(\theta_0)\big)\right](x, w) &\rightsquigarrow \left[\phi'_{1,\theta_0}(\mathbf{Z}_1)\right](x, w) \\ &\equiv \tilde{\mathbf{Z}}_2(x, w).\end{aligned}$$

This result holds uniformly over any finite grid of values for $y \in \mathbb{R}$ and $c \in \mathcal{C}$ by considering the Hadamard directional differentiability of a vector of these mappings indexed at different values of y and c, which yields the process $\mathbf{Z}_2(y, x, w, c)$. □

Proof of Lemma 2. Let $\mathcal{S} = \{(y, x, w) \in \mathbb{R}^2 : y \in [\underline{y}_x(w), \overline{y}_x(w)], x \in \{0, 1\}, w \in \operatorname{supp}(W)\}$. Let $\mathscr{D}(\mathcal{S}) \subset \ell^\infty(\mathcal{S})$ denote the set of functions that are càdlàg in the first argument for each $x \in \{0, 1\}$ and $w \in \operatorname{supp}(W)$. Define the mapping

$$\tilde{\phi}_2 : \mathscr{D}(\mathcal{S}) \times \ell^\infty\big(\{0, 1\} \times \operatorname{supp}(W)\big) \times \ell^\infty\big(\operatorname{supp}(W)\big) \to \ell^\infty\big((0, 1) \times \{0, 1\} \times \operatorname{supp}(W), \mathbb{R}^2\big)$$

by

$$\left[\tilde{\phi}_2(\theta)\right](\tau, x, w) = \begin{pmatrix} \left(\theta^{(1)}\right)^{-1}(\tau, x, w) \\ \theta^{(2)}(x, w) \end{pmatrix}.$$

By Assumptions A1, A3, A5, and Lemma 21.4(ii) in van der Vaart (2000) this mapping is Hadamard differentiable at θ_0 tangentially to $\mathscr{C}(\mathcal{S}) \times \ell^\infty(\{0, 1\} \times \operatorname{supp}(W)) \times \ell^\infty(\operatorname{supp}(W))$, where $\mathscr{C}(\mathcal{S}) \subset \ell^\infty(\mathcal{S})$ is the set functions that are continuous in the first argument for each $x \in \{0, 1\}$ and $w \in \operatorname{supp}(W)$. Its Hadamard derivative at $\theta_0 = (F_{Y|X,W}(\cdot \mid \cdot, \cdot), p_{(\cdot|\cdot)}, q_{(\cdot)})$ is

$$\left[\tilde{\phi}'_{2,\theta_0}(h)\right](\tau, x, w) \mapsto \left(-\frac{h^{(1)}\big(Q_{Y|X,W}(\tau \mid x, w), x, w\big)}{f_{Y|X,W}\big(Q_{Y|X,W}(\tau \mid x, w) \mid x, w\big)}, h^{(2)}(x, w)\right).$$

By the functional delta method and Theorem 7.3.3 part (iii) of Bickel and Doksum (2015),

$$\left[\sqrt{N}\big(\tilde{\phi}_2(\widehat{\theta}) - \tilde{\phi}_2(\theta_0)\big)\right](\tau, x, w) \rightsquigarrow \tilde{\mathbf{Z}}_3(\tau, x, w),$$

where $\tilde{\mathbf{Z}}_3$ is a mean-zero Gaussian process in $\ell^\infty((0, 1) \times \{0, 1\} \times \operatorname{supp}(W), \mathbb{R}^2)$ with uniformly continuous paths.

Now define the mapping

$$\begin{aligned}\phi_2 : \ell^\infty\big((0, 1) \times \{0, 1\} \times \operatorname{supp}(W)\big) &\times \ell^\infty\big(\{0, 1\} \times \operatorname{supp}(W)\big) \\ &\to \ell^\infty\big((0, 1) \times \{0, 1\} \times \operatorname{supp}(W) \times [0, \overline{C}], \mathbb{R}^2\big)\end{aligned}$$

by

$$\left[\phi_2(\psi)\right](\tau, x, w, c) = \begin{pmatrix} \psi^{(1)}\left(\tau + \dfrac{c}{\psi^{(2)}(x, w)} \min\{\tau, 1 - \tau\}, x, w\right) \\ \psi^{(1)}\left(\tau - \dfrac{c}{\psi^{(2)}(x, w)} \min\{\tau, 1 - \tau\}, x, w\right) \end{pmatrix}.$$

Then

$$\begin{pmatrix} \overline{Q}^c_{Y_x|W}(\tau \mid w) \\ \underline{Q}^c_{Y_x|W}(\tau \mid w) \end{pmatrix} = \big[\phi_2\big(\tilde{\phi}_2(\theta_0)\big)\big](\tau, x, w, c),$$
$$\begin{pmatrix} \widehat{\overline{Q}}^c_{Y_x|W}(\tau \mid w) \\ \widehat{\underline{Q}}^c_{Y_x|W}(\tau \mid w) \end{pmatrix} = \big[\phi_2\big(\tilde{\phi}_2(\widehat{\theta})\big)\big](\tau, x, w, c).$$

We will show that ϕ_2 is Hadamard differentiable tangentially to the space $\mathscr{C}_U((0,1) \times \{0,1\} \times \operatorname{supp}(W)) \times \ell^\infty(\{0,1\} \times \operatorname{supp}(W))$, where $\mathscr{C}_U(\mathcal{A})$ denotes the set of uniformly continuous functions on $\mathcal{A}$. The Hadamard derivative of the first component of ϕ_2 evaluated at $\psi_0 \equiv \tilde{\phi}_2(\theta_0)$ is

$$\begin{aligned}
&\big[\phi^{(1)\prime}_{2,\psi_0}(h)\big](\tau, x, w, c) \\
&\quad = h^{(1)}\left(\tau + \frac{c}{\psi_0^{(2)}(x,w)} \min\{\tau, 1-\tau\}, x, w\right) \\
&\qquad - \psi_0^{(1)\prime}\left(\tau + \frac{c}{\psi_0^{(2)}(x,w)} \min\{\tau, 1-\tau\}, x, w\right) \frac{c \min\{\tau, 1-\tau\}}{\big(\psi_0^{(2)}(x,w)\big)^2} h^{(2)}(x,w).
\end{aligned}$$

To see this, a Taylor expansion gives

$$\begin{aligned}
&\left[\frac{\phi_2^{(1)}(\psi_0 + t_n h_n) - \phi_2^{(1)}(\psi_0)}{t_n}\right](\tau, x, w, c) \\
&\quad = h_n^{(1)}\left(\tau + \frac{c}{\psi_0^{(2)}(x,w) + t_n h_n^{(2)}(x,w)} \min\{\tau, 1-\tau\}, x, w\right) \\
&\qquad - \psi_0^{(1)\prime}\left(\tau + \frac{c}{\psi_0^{(2)}(x,w) + a_n(x,w)} \min\{\tau, 1-\tau\}, x, w\right) \\
&\qquad \times \frac{c \min\{\tau, 1-\tau\}}{\big(\psi_0^{(2)}(x,w) + a_n(x,w)\big)^2} h_n^{(2)}(x,w)
\end{aligned}$$

using the fact that $\psi_0^{(1)}(\tau, x, w) = Q_{Y|X}(\tau \mid x, w)$ is continuously differentiable in τ by Assumption A5.2, and noting that term $a_n(x,w)$ satisfies $|a_n(x,w)| \leq |t_n h_n^{(2)}(x,w)| = O(t_n)$. Next,

$$\begin{aligned}
&\sup_{\tau,x,w,c}\left| h_n^{(1)}\left(\tau + \frac{c \min\{\tau, 1-\tau\}}{\psi_0^{(2)}(x,w) + t_n h_n^{(2)}(x,w)}, x, w\right) \right. \\
&\qquad \left. - h^{(1)}\left(\tau + \frac{c \min\{\tau, 1-\tau\}}{\psi_0^{(2)}(x,w)}, x, w\right)\right| \\
&\quad \leq \sup_{\tau,x,w,c}\left| h_n^{(1)}\left(\tau + \frac{c \min\{\tau, 1-\tau\}}{\psi_0^{(2)}(x,w) + t_n h_n^{(2)}(x,w)}, x, w\right)\right.
\end{aligned}$$

$$
\begin{aligned}
&\qquad - h^{(1)}\left(\tau + \frac{c\min\{\tau, 1-\tau\}}{\psi_0^{(2)}(x,w) + t_n h_n^{(2)}(x,w)}, x, w\right)\Bigg| \\
&+ \sup_{\tau,x,w,c}\left| h^{(1)}\left(\tau + \frac{c\min\{\tau, 1-\tau\}}{\psi_0^{(2)}(x,w) + t_n h_n^{(2)}(x,w)}, x, w\right) \right. \\
&\qquad \left. - h^{(1)}\left(\tau + \frac{c\min\{\tau, 1-\tau\}}{\psi_0^{(2)}(x,w)}, x, w\right)\right| \\
&\leq \left\| h_n^{(1)} - h^{(1)} \right\|_\infty + o(1) \\
&= o(1),
\end{aligned}
$$

where all three suprema are taken over $\tau \in (0,1)$, $x \in \{0,1\}$, $w \in \operatorname{supp}(W)$, $c \in [0, \overline{C}]$. The last inequality follows from uniform continuity of $h^{(1)}$. The last line follows from uniform convergence of h_n to h.

Similarly, we have that

$$
\begin{aligned}
\sup_{\tau,x,w,c}&\left| \psi_0^{(1)\prime}\left(\tau + \frac{c}{\psi_0^{(2)}(x,w) + a_n(x,w)}\min\{\tau, 1-\tau\}, x, w\right) \right. \\
&\qquad \times \frac{c\min\{\tau, 1-\tau\}}{\left(\psi_0^{(2)}(x,w) + a_n(x,w)\right)^2} h_n^{(2)}(x,w) \\
&- \psi_0^{(1)\prime}\left(\tau + \frac{c}{\psi_0^{(2)}(x,w)}\min\{\tau, 1-\tau\}, x, w\right) \\
&\qquad \left. \times \frac{c\min\{\tau, 1-\tau\}}{\left(\psi_0^{(2)}(x,w)\right)^2} h^{(2)}(x,w) \right| = o(1)
\end{aligned}
$$

by uniform continuity of $\psi_0^{(1)\prime}$ (implied by Assumption A5.2) and by $a_n(x,w) = o(1)$. Again, the sup is over $\tau \in (0,1)$, $x \in \{0,1\}$, $w \in \operatorname{supp}(W)$, $c \in [0, \overline{C}]$. Therefore $\phi_2^{(1)}$ is Hadamard differentiable tangentially to the space of uniformly continuous functions. A similar argument can be made for $\phi_2^{(2)}$. By composition, $\phi_2 \circ \tilde{\phi}_2$ is Hadamard differentiable tangentially to $\mathscr{C}(\mathcal{S})$.

By the functional delta method and the fact that $\tilde{\mathbf{Z}}_3(y,x,w)$ has uniformly continuous paths, we have that

$$
\begin{aligned}
\left[\sqrt{N}\left(\phi_2\left(\tilde{\phi}_2(\widehat{\theta})\right) - \phi_2\left(\tilde{\phi}_2(\theta_0)\right)\right)\right](\tau,x,w,c) &\rightsquigarrow \left[\phi_{2,\psi_0}' \circ \tilde{\phi}_{2,\theta_0}'(\mathbf{Z}_1)\right](\tau,x,w,c) \\
&\equiv \mathbf{Z}_3(\tau,x,w,c),
\end{aligned}
$$

a mean-zero Gaussian process with continuous paths in $\tau \in (0,1)$ and $c \in [0, \overline{C}]$. □

Proof of Proposition 1. Consider the lower CQTE bound of equation (8) as a function of c. Its first component is the lower bound of the conditional quantile of $Y_1 \mid W = w$.

By Assumption A5.2, the derivative of that conditional quantile with respect to c equals

$$\begin{aligned}&\frac{\partial}{\partial c}Q_{Y|X,W}\left(\tau-\frac{c}{p_{1|w}}\min\{\tau,1-\tau\}\,\Big|\,1,w\right)\\&\quad=\frac{-\min\{\tau,1-\tau\}}{p_{1|w}f_{Y|X,W}\left(Q_{Y|X,W}\left(\tau-\frac{c}{p_{1|w}}\min\{\tau,1-\tau\}\,\Big|\,1,w\right)\,\Big|\,1,w\right)}.\end{aligned}$$

The second component of the lower CQTE bound is the upper bound of the conditional quantile of $Y_0 \mid W=w$. The derivative of that conditional quantile with respect to c equals

$$\begin{aligned}&\frac{\partial}{\partial c}Q_{Y|X,W}\left(\tau+\frac{c}{p_{0|w}}\min\{\tau,1-\tau\}\,\Big|\,0,w\right)\\&\quad=\frac{\min\{\tau,1-\tau\}}{p_{0|w}f_{Y|X,W}\left(Q_{Y|X,W}\left(\tau+\frac{c}{p_{0|w}}\min\{\tau,1-\tau\}\,\Big|\,0,w\right)\,\Big|\,0,w\right)}.\end{aligned}$$

Moreover, these derivatives are bounded away from zero and infinity uniformly over $c \in (0,\overline{C}]$. This implies that the derivative of the CQTE is negative and uniformly bounded away from zero.

Next recall that

$$\underline{\text{CATE}}(c\mid w)=\int_0^1\underline{\text{CQTE}}(\tau,c\mid w)\,d\tau.$$

Its derivative with respect to c exists by the dominated convergence theorem (by Assumptions A1 and A5). Moreover, it is bounded away from zero for all $c\in(0,\overline{C}]$. By taking another expectation over the marginal distribution of W, $\partial\underline{\text{ATE}}(c)/\partial c$ exists (by Assumption A4), is negative, and is bounded away from zero for all $c\in(0,\overline{C}]$.

c^* is defined implicitly by $\underline{\text{ATE}}(c^*)=\mu$. We have shown that the function $\underline{\text{ATE}}(c)$ satisfies the assumptions of Lemma 21.3 on page 306 of van der Vaart (2000). Thus the mapping $\underline{\text{ATE}}(\cdot)\mapsto c^*$ is Hadamard differentiable tangentially to the set of càdlàg functions on $(0,\overline{C}]$ with derivative

$$\frac{-h(c^*)}{\frac{\partial}{\partial c}\underline{\text{ATE}}(c^*)}.$$

By the discussion following Lemma 2, $\sqrt{N}(\widehat{\underline{\text{ATE}}}(c)-\underline{\text{ATE}}(c))$ converges in distribution to a random element of $\ell^\infty([0,\overline{C}])$ with continuous paths.

Let

$$\tilde{c}^*=\inf\bigl\{c\in[0,\overline{C}]:\widehat{\underline{\text{ATE}}}(c)\le\mu\bigr\}.$$

We can then apply the functional delta method to see that $\sqrt{N}(\tilde{c}^*-c^*)$ converges in distribution to a Gaussian variable we denote by $\mathbf{Z}_{\text{BP}}$.

Since $c^* \in (0, \overline{C}]$ and by monotonicity of $\underline{\text{ATE}}(\cdot)$, we have $\underline{\text{ATE}}(\overline{C}) \leq \mu$. By $\sqrt{N}$-convergence of the ATE bounds,

$$\mathbb{P}\big(\widehat{\underline{\text{ATE}}}(\overline{C}) > \mu\big) = \mathbb{P}\big(-\sqrt{N}\big(\underline{\text{ATE}}(\overline{C}) - \mu\big) < \sqrt{N}\big(\widehat{\underline{\text{ATE}}}(\overline{C}) - \underline{\text{ATE}}(\overline{C})\big)\big)$$
$$\to 0.$$

Therefore, the set $\{c \in [0, \overline{C}] : \widehat{\underline{\text{ATE}}}(c) \leq \mu\}$ is nonempty with probability approaching one. This implies that $\tilde{c}^* \in [0, \overline{C}]$ with probability approaching one and therefore $\mathbb{P}(\tilde{c}^* = \widehat{c}^*)$ also approaches one as $N \to \infty$. Using these results, we obtain

$$\begin{aligned}\sqrt{N}\big(\widehat{c}^* - c^*\big) &= \sqrt{N}\big(\widehat{c}^* - \tilde{c}^*\big) + \sqrt{N}\big(\tilde{c}^* - c^*\big) \\ &= o_p(1) + \sqrt{N}\big(\tilde{c}^* - c^*\big) \\ &\rightsquigarrow \mathbf{Z}_{\text{BP}}. \qquad \square\end{aligned}$$

The following result extends Proposition 2(i) of Chernozhukov, Fernández-Val, and Galichon (2010) to allow for input functions which are directionally differentiable, but not fully differentiable, at one point. It can be extended to allow for multiple points of directional differentiability, but we omit this since we do not need it for our application.

Lemma C3. *Let $\theta_0(u, c, w) = (\theta_0^{(1)}(u, c, w), \theta_0^{(2)}(u, c, w))$ where for $j \in \{1, 2\}$ we have that $\theta_0^{(j)}(u, c, w)$ is bounded above and below, and differentiable everywhere except at $u = u^*$, where it is directionally differentiable. Further, assume that the two components satisfy Assumption A6. Then, for fixed $z \in \mathbb{R}$, the mapping $\phi_3 : \ell^\infty((0, 1) \times \text{supp}(W) \times \mathcal{C}, \mathbb{R}^2) \to \ell^\infty(\text{supp}(W) \times \mathcal{C}, \mathbb{R}^2)$ defined by*

$$\big[\phi_3(\theta)\big](w, c) = \begin{pmatrix} \displaystyle\int_0^1 \mathbb{1}\big(\theta^{(2)}(u, c, w) \leq z\big)\, du \\ \displaystyle\int_0^1 \mathbb{1}\big(\theta^{(1)}(u, c, w) \leq z\big)\, du \end{pmatrix}$$

is Hadamard directionally differentiable tangentially to $\mathscr{C}((0, 1) \times \text{supp}(W) \times \mathcal{C}, \mathbb{R}^2)$ with Hadamard directional derivative given by equations (30) and (31) below.

Proof of Lemma C3. For clarity we suppress the dependence on w in the expressions below. Uniformity of convergence over $w \in \text{supp}(W)$ follows from the discreteness of $\text{supp}(W)$ (Assumption A4). Our proof follows that of Proposition 2(i) in Chernozhukov, Fernández-Val, and Galichon (2010). Let

$$\mathcal{U}_1(c) = \big\{u \in (0, 1) : \theta_0^{(1)}(u, c) = z\big\}$$

denote the set of roots to the equation $\theta_0^{(1)}(u, c) = z$ for fixed z and c. By Assumption A6.1, this set contains a finite number of elements. We denote these by

$$\mathcal{U}_1(c) = \big\{u_k^{(1)}(c),\ \text{for } k = 1, 2, \ldots, K^{(1)}(c) < \infty\big\}.$$

Assumption A6.1 also implies that $\mathcal{U}_1(c) \cap \mathcal{U}_1^*(c) = \emptyset$ for any $c \in \mathcal{C}$.

We will show the first component of the Hadamard directional derivative is given by

$$\big[\phi_{3,\theta_0}^{(1)\prime}(h)\big](c) = -\sum_{k=1}^{K^{(1)}(c)} h\big(u_k^{(1)}(c), c\big)\left(\frac{\mathbb{1}\big(h\big(u_k^{(1)}(c), c\big) > 0\big)}{\big|\partial_u^- \theta_0^{(1)}\big(u_k^{(1)}(c), c\big)\big|} + \frac{\mathbb{1}\big(h\big(u_k^{(1)}(c), c\big) < 0\big)}{\big|\partial_u^+ \theta_0^{(1)}\big(u_k^{(1)}(c), c\big)\big|}\right), \quad (30)$$

where $h \in \mathscr{C}((0,1) \times \mathcal{C})$.

First, suppose $u^* \notin \mathcal{U}_1(c)$ for any $c \in \mathcal{C}$. In this case we can apply Proposition 2(i) of Chernozhukov, Fernández-Val, and Galichon (2010) directly to obtain

$$\left|\frac{\big[\phi_3^{(1)}(\theta_0 + t_n h_n)\big](c) - \big[\phi_3^{(1)}(\theta_0)\big](c)}{t_n} - \left(-\sum_{k=1}^{K^{(1)}(c)} \frac{h\big(u_k^{(1)}(c), c\big)}{\big|\partial_u \theta_0^{(1)}\big(u_k^{(1)}(c), c\big)\big|}\right)\right| = o(1)$$

for any $c \in \mathcal{C}$, where $t_n \searrow 0$, $h_n \in \ell^\infty((0,1) \times \mathcal{C})$, and

$$\sup_{(u,c) \in (0,1) \times \mathcal{C}} \big|h_n(u,c) - h(u,c)\big| = o(1)$$

as $n \to \infty$. Hence

$$\big[\phi_{3,\theta_0}^{(1)\prime}(h)\big](c) = -\sum_{k=1}^{K^{(1)}(c)} \frac{h\big(u_k^{(1)}(c), c\big)}{\big|\partial_u \theta_0^{(1)}\big(u_k^{(1)}(c), c\big)\big|},$$

a linear map in h.

Now suppose $u^* \in \mathcal{U}_1(c)$ for some $c \in \mathcal{C}$. Without loss of generality, let $u_1^{(1)}(c) = u^*$. Let $B_\epsilon(u)$ denote a ball of radius ϵ centered at u. By equation (A.1) in Chernozhukov, Fernández-Val, and Galichon (2010), for any $\delta > 0$ there exists an $\epsilon > 0$ and a large enough n such that

$$\begin{aligned}&\frac{\big[\phi_3^{(1)}(\theta_0 + t_n h_n)\big](c) - \big[\phi_3^{(1)}(\theta_0)\big](c)}{t_n}\\ &\quad \leq \sum_{k=1}^{K^{(1)}(c)} \int_{B_\epsilon(u_k^{(1)}(c))} \frac{\mathbb{1}\big(\theta_0(u,c) + t_n\big(h\big(u_k^{(1)}(c), c\big) - \delta\big) \leq z\big) - \mathbb{1}\big(\theta_0(u,c) \leq z\big)}{t_n}\, du.\end{aligned}$$

Likewise, for any $\delta > 0$ there exists $\epsilon > 0$ and large enough n such that

$$\begin{aligned}&\frac{\big[\phi_3^{(1)}(\theta_0 + t_n h_n)\big](c) - \big[\phi_3^{(1)}(\theta_0)\big](c)}{t_n}\\ &\quad \geq \sum_{k=1}^{K^{(1)}(c)} \int_{B_\epsilon(u_k^{(1)}(c))} \frac{\mathbb{1}\big(\theta_0(u,c) + t_n\big(h\big(u_k^{(1)}(c), c\big) + \delta\big) \leq z\big) - \mathbb{1}\big(\theta_0(u,c) \leq z\big)}{t_n}\, du.\end{aligned}$$

The $k = 1$ element in the first sum is

$$\int_{B_\epsilon(u^*)} \frac{\mathbb{1}\big(\theta_0(u,c) + t_n\big(h(u^*, c) - \delta\big) \leq z\big) - \mathbb{1}\big(\theta_0(u,c) \leq z\big)}{t_n}\, du.$$

$\theta_0(u, c)$ is absolutely continuous in u and, by the change of variables formula for absolutely continuous functions, the transformation $z' = \theta_0(u, c)$ implies that this $k = 1$ term is

$$\frac{1}{t_n} \int_{J_1 \cap [z, z - t_n(h(u^*, c) - \delta)]} \frac{1}{\left|\partial_u \theta_0\left(\theta_0^{-1}\left(z', c\right), c\right)\right|} \, dz',$$

where J_1 is the image of $B_\epsilon(u^*)$ under $\theta_0(\cdot, c)$ and the change of variables follows from the monotonicity of θ_0 in $B_\epsilon(u^*)$ for small enough ϵ (this monotonicity follows from Assumption A6.1, which implies that the derivative of θ_0 changes sign a finite number of times). The closed interval $[z, z - t_n(h(u^*, c) - \delta)]$ should be interpreted as $[z - t_n(h(u^*, c) - \delta), z]$ when $z - t_n(h(u^*, c) - \delta) < z$. Next consider three cases:

1. When $h(u^*, c) > 0$, the interval $[z, z - t_n(h(u^*, c) - \delta)]$ has the form $[z - \psi_n, z]$ for an arbitrarily small $\psi_n > 0$. Therefore the denominator $|\partial_u \theta_0(\theta_0^{-1}(z', c), c)|$ converges to $|\partial_u^- \theta_0(u^*, c)|$ as $n \to \infty$, by continuous differentiability on $(0, u^*)$, directional differentiability at $u = u^*$, and by $\theta_0^{-1}(z', c) = u^* + o(1)$. This holds by $z' \in [z - t_n(h(u^*, c) - \delta), z]$, an interval shrinking to $\{z\}$. Therefore,

$$\begin{aligned}
&\frac{1}{t_n} \int_{J_1 \cap [z, z - t_n(h(u^*, c) - \delta)]} \frac{1}{\left|\partial_u \theta_0\left(\theta_0^{-1}\left(z', c\right), c\right)\right|} \, dz' \\
&\quad = \frac{1}{t_n} \int_{z - t_n(h(u^*, c) - \delta)}^{z} \frac{1}{\left|\partial_u^- \theta_0\left(u^*, c\right)\right| + o(1)} \, dz' \\
&\quad = \frac{-h\left(u^*, c\right) + \delta}{\left|\partial_u^- \theta_0\left(u^*, c\right)\right|} + o(1).
\end{aligned}$$

By a similar argument,

$$\int_{B_\epsilon(u^*)} \frac{\mathbb{1}\left(\theta_0(u, c) + t_n\left(h\left(u^*, c\right) + \delta\right) \leq z\right) - \mathbb{1}\left(\theta_0(u, c) \leq z\right)}{t_n} \, du = \frac{-h\left(u^*, c\right) - \delta}{\left|\partial_u^- \theta_0\left(u^*, c\right)\right|} + o(1).$$

Letting $\delta > 0$ be arbitrarily small and by the squeeze theorem, we obtain

$$\frac{\left[\phi_3^{(1)}(\theta_0 + t_n h_n)\right](c) - \left[\phi_3^{(1)}(\theta_0)\right](c)}{t_n} = -\sum_{k=1}^{K^{(1)}(c)} \frac{h\left(u_k^{(1)}(c), c\right)}{\left|\partial_u^- \theta_0^{(1)}\left(u_k^{(1)}(c), c\right)\right|} + o(1).$$

2. When $h(u^*, c) < 0$, the interval $[z, z - t_n(h(u^*, c) - \delta)]$ is of the form $[z, z + \psi_n]$ for arbitrarily small $\psi_n > 0$. Using the same argument as in case 1, $|\partial_u \theta_0(\theta_0^{-1}(z', c), c)|$ converges to $|\partial_u^+ \theta_0(u^*, c)|$ as $n \to \infty$. Therefore, proceeding as in the previous case, we obtain that

$$\frac{\left[\phi_3^{(1)}(\theta_0 + t_n h_n)\right](c) - \left[\phi_3^{(1)}(\theta_0)\right](c)}{t_n} = -\sum_{k=1}^{K^{(1)}(c)} \frac{h\left(u_k^{(1)}(c), c\right)}{\left|\partial_u^+ \theta_0^{(1)}\left(u_k^{(1)}(c), c\right)\right|} + o(1).$$

3. When $h(u^*, c) = 0$, this $k = 1$ term converges to zero.

Combining these three cases into a single expression, we find that

$$\begin{aligned}
&\frac{1}{t_n}\int_{J_1\cap[z,z-t_n(h(u^*,c)-\delta)]}\frac{1}{\left|\partial_u\theta_0\left(\theta_0^{-1}(z',c),c\right)\right|}\,dz' \\
&\quad = -h\left(u^*,c\right)\left(\frac{\mathbb{1}\left(h\left(u^*,c\right)>0\right)}{\left|\partial_u^-\theta_0\left(u^*,c\right)\right|}+\frac{\mathbb{1}\left(h\left(u^*,c\right)<0\right)}{\left|\partial_u^+\theta_0\left(u^*,c\right)\right|}\right)+o(1).
\end{aligned}$$

This expression coincides with the Hadamard derivative under continuous differentiability at $u = u^*$, since that implies $\partial_u^-\theta_0(u^*,c) = \partial_u^+\theta_0(u^*,c)$. It follows from the remainder of the proof in Chernozhukov, Fernández-Val, and Galichon (2010) that

$$\sup_{c\in\mathcal{C}}\left|\frac{\left[\phi_3^{(1)}(\theta_0+t_nh_n)\right](c)-\left[\phi_3^{(1)}(\theta_0)\right](c)}{t_n}-\left[\phi_{3,\theta_0}^{(1)\prime}(h)\right](c)\right| = o(1),$$

where $\|\cdot\|_e$ is the Euclidean norm, and where $\phi_{3,\theta_0}^{(1)\prime}$ is defined in equation (30). Note that $\phi_{3,\theta_0}^{(1)\prime}$ is continuous in h, and therefore it is a Hadamard directional derivative.

That completes our analysis of the first component of the Hadamard directional derivative of ϕ_3 with respect to θ at θ_0. By similar arguments, the second component is

$$\left[\phi_{3,\theta_0}^{(2)\prime}(h)\right](c) = -\sum_{k=1}^{K^{(2)}(c)} h\left(u_k^{(2)}(c),c\right)\left(\frac{\mathbb{1}\left(h\left(u_k^{(2)}(c),c\right)>0\right)}{\left|\partial_u^-\theta_0^{(2)}\left(u_k^{(2)}(c),c\right)\right|}+\frac{\mathbb{1}\left(h\left(u_k^{(2)}(c),c\right)<0\right)}{\left|\partial_u^+\theta_0^{(2)}\left(u_k^{(2)}(c),c\right)\right|}\right). \quad (31)$$

$\square$

Proof of Lemma 3. Let

$$\theta_0(\tau,w,c) = \begin{pmatrix}\underline{Q}_{Y_1|W}^c(\tau\mid w)-\overline{Q}_{Y_0|W}^c(\tau\mid w)\\ \overline{Q}_{Y_1|W}^c(\tau\mid w)-\underline{Q}_{Y_0|W}^c(\tau\mid w)\end{pmatrix},$$

$$\widehat{\theta}(\tau,w,c) = \begin{pmatrix}\widehat{\underline{Q}}_{Y_1|W}^c(\tau\mid w)-\widehat{\overline{Q}}_{Y_0|W}^c(\tau\mid w)\\ \widehat{\overline{Q}}_{Y_1|W}^c(\tau\mid w)-\widehat{\underline{Q}}_{Y_0|W}^c(\tau\mid w)\end{pmatrix}.$$

Therefore

$$\begin{pmatrix}\overline{P}(c\mid w)\\ \underline{P}(c\mid w)\end{pmatrix} = \left[\phi_3(\theta_0)\right](w,c) \quad \text{and} \quad \begin{pmatrix}\widehat{\overline{P}}(c\mid w)\\ \widehat{\underline{P}}(c\mid w)\end{pmatrix} = \left[\phi_3(\widehat{\theta})\right](w,c).$$

By Lemma 2,

$$\begin{aligned}
&\sqrt{N}\begin{pmatrix}\widehat{\underline{Q}}_{Y_1|W}^c(\tau\mid w)-\widehat{\overline{Q}}_{Y_0|W}^c(\tau\mid w)-\left(\underline{Q}_{Y_1|W}^c(\tau\mid w)-\overline{Q}_{Y_0|W}^c(\tau\mid w)\right)\\ \widehat{\overline{Q}}_{Y_1|W}^c(\tau\mid w)-\widehat{\underline{Q}}_{Y_0|W}^c(\tau\mid w)-\left(\overline{Q}_{Y_1|W}^c(\tau\mid w)-\underline{Q}_{Y_0|W}^c(\tau\mid w)\right)\end{pmatrix} \\
&\quad \rightsquigarrow \begin{pmatrix}\mathbf{Z}_3^{(2)}(\tau,1,w,c)-\mathbf{Z}_3^{(1)}(\tau,0,w,c)\\ \mathbf{Z}_3^{(1)}(\tau,1,w,c)-\mathbf{Z}_3^{(2)}(\tau,0,w,c)\end{pmatrix},
\end{aligned}$$

a mean-zero Gaussian processes in $\ell^\infty((0,1)\times \operatorname{supp}(W)\times \mathcal{C}, \mathbb{R}^2)$ with continuous paths.

By Lemma C3 with $u^* = 1/2$, the mapping ϕ_3 is Hadamard directionally differentiable tangentially to $\mathscr{C}((0,1)\times \operatorname{supp}(W)\times \mathcal{C}, \mathbb{R}^2)$. By the functional delta method for Hadamard directionally differentiable functions (e.g., Theorem 2.1 in Fang and Santos (2019)), we obtain

$$\begin{aligned}
&\sqrt{N}\begin{pmatrix} \widehat{\overline{P}}(c \mid w) - \overline{P}(c \mid w) \\ \widehat{\underline{P}}(c \mid w) - \underline{P}(c \mid w) \end{pmatrix} \\
&\rightsquigarrow \begin{pmatrix} \big[\phi'^{(2)}_{3,\underline{Q}^{(\cdot)}_{Y_1|W}(\cdot|\cdot)-\overline{Q}^{(\cdot)}_{Y_0|W}(\cdot|\cdot)}\big(\mathbf{Z}_3^{(2)}(\cdot,1,\cdot,\cdot)-\mathbf{Z}_3^{(1)}(\cdot,0,\cdot,\cdot)\big)\big](w,c) \\ \big[\phi'^{(1)}_{3,\overline{Q}^{(\cdot)}_{Y_1|W}(\cdot|\cdot)-\underline{Q}^{(\cdot)}_{Y_0|W}(\cdot|\cdot)}\big(\mathbf{Z}_3^{(1)}(\cdot,1,\cdot,\cdot)-\mathbf{Z}_3^{(2)}(\cdot,0,\cdot,\cdot)\big)\big](w,c) \end{pmatrix} \\
&\equiv \mathbf{Z}_4(w,c),
\end{aligned}$$

a tight random element of $\ell^\infty(\operatorname{supp}(W)\times \mathcal{C}, \mathbb{R}^2)$. □

The following lemma shows that the sup operator is Hadamard directionally differentiable. It is a very minor extension of Lemma B.1 in Fang and Santos (2015), where we take the supremum over just one of two arguments.

Lemma C4. *Let $\mathcal{A}$ and $\mathcal{C}$ be compact subsets of $\mathbb{R}$. Define the map $\phi : \ell^\infty(\mathcal{A}\times\mathcal{C}) \to \ell^\infty(\mathcal{C})$ by*

$$\big[\phi(\theta)\big](c) = \sup_{a\in\mathcal{A}} \theta(a,c).$$

Let

$$\Psi_{\mathcal{A}}(\theta,c) = \operatorname*{arg\,max}_{a\in\mathcal{A}} \theta(a,c)$$

be a set-valued function. Then ϕ is Hadamard directionally differentiable tangentially to $\mathscr{C}(\mathcal{A}\times\mathcal{C})$ at any $\theta\in\mathscr{C}(\mathcal{A}\times\mathcal{C})$, and $\phi'_\theta : \mathscr{C}(\mathcal{A}\times\mathcal{C}) \to \mathscr{C}(\mathcal{C})$ is given by

$$\big[\phi'_\theta(h)\big](c) = \sup_{a\in\Psi_{\mathcal{A}}(\theta,c)} h(a,c)$$

for any $h\in\mathscr{C}(\mathcal{A}\times\mathcal{C})$.

Proof of Lemma C4. This proof follows that of Lemma B.1 in Fang and Santos (2015). Let $t_n \searrow 0$, and $h_n \in \ell^\infty(\mathcal{A}\times\mathcal{C})$ such that

$$\sup_{(a,c)\in\mathcal{A}\times\mathcal{C}} \big|h_n(a,c) - h(a,c)\big| \equiv \|h_n - h\|_\infty = o(1)$$

for $h\in\mathscr{C}(\mathcal{A}\times\mathcal{C})$. Since $\mathcal{A}$ is a closed and bounded subset of $\mathbb{R}$, their lemma shows that tangential Hadamard directional differentiability holds for any fixed $c\in\mathcal{C}$. We show that this holds uniformly in $c\in\mathcal{C}$ as well. First, by their equation (B.1), we note that for some

$t_n \searrow 0$,

$$\begin{aligned}\sup_{c\in\mathcal{C}}\Bigl|\sup_{a\in\mathcal{A}}\bigl(\theta(a,c)+t_n h_n(a,c)\bigr)-\sup_{a\in\mathcal{A}}\bigl(\theta(a,c)+t_n h(a,c)\bigr)\Bigr| &\le \sup_{c\in\mathcal{C}} t_n \sup_{a\in\mathcal{A}}\bigl|h_n(a,c)-h(a,c)\bigr| \\ &= t_n\|h_n-h\|_\infty \\ &= o(t_n). \end{aligned} \tag{32}$$

Second, by their equations leading to (B.3)

$$\begin{aligned}&\sup_{c\in\mathcal{C}}\Bigl|\sup_{a\in\mathcal{A}}\bigl(\theta(a,c)+t_n h(a,c)\bigr)-\sup_{a\in\Psi_{\mathcal{A}}(\theta,c)}\bigl(\theta(a,c)+t_n h(a,c)\bigr)\Bigr| \\ &\quad\le t_n\sup_{c\in\mathcal{C}}\sup_{a_0,a_1\in\mathcal{A}:|a_0-a_1|\le\delta_n}\bigl|h(a_0,c)-h(a_1,c)\bigr| \\ &\quad= o(t_n)\end{aligned} \tag{33}$$

by uniform continuity of $h(a,c)$ in a and c, which follows from the continuity of h on its compact support $\mathcal{A}\times\mathcal{C}$. Finally, combining equations (32) and (33) as in equation (B.4) from Fang and Santos (2019), it follows that

$$\begin{aligned}&\sup_{c\in\mathcal{C}}\Bigl|\sup_{a\in\mathcal{A}}\bigl(\theta(a,c)+t_n h_n(a,c)\bigr)-\sup_{a\in\mathcal{A}}\theta(a,c)-t_n\sup_{a\in\Psi_{\mathcal{A}}(\theta,c)}h(a,c)\Bigr| \\ &\le\sup_{c\in\mathcal{C}}\Bigl|\sup_{a\in\Psi_{\mathcal{A}}(\theta,c)}\bigl(\theta(a,c)+t_n h(a,c)\bigr)-\sup_{a\in\Psi_{\mathcal{A}}(\theta,c)}\theta(a,c)-t_n\sup_{a\in\Psi_{\mathcal{A}}(\theta,c)}h(a,c)\Bigr|+o(t_n) \\ &=0+o(t_n),\end{aligned}$$

which completes the proof. □

Proof of Lemma 4. We begin by showing that the first component in equation (29) converges to a tight random element of $\ell^\infty(\mathcal{C}\times[0,1])$. Fix c and w and define

$$\phi_4:\ell^\infty(\mathbb{R})\to\mathbb{R}$$

by

$$\phi_4(\theta)=\max\Bigl\{\sup_{a\in\mathcal{Y}_z(w)}\theta(a,w,c),0\Bigr\}.$$

As in the proof of Lemma 1, the four mappings $(\delta_1,\delta_2,\delta_3,\delta_4)$ when considered from $\ell^\infty(\mathbb{R}\times\{0,1\}\times\operatorname{supp}(W))\times\ell^\infty(\{0,1\}\times\operatorname{supp}(W))\times\ell^\infty(\operatorname{supp}(W))$ to $\ell^\infty(\mathbb{R}\times\{0,1\}\times\operatorname{supp}(W))$ are all Hadamard differentiable when evaluated at θ_0.

We can write

$$\begin{aligned}\phi_4(\theta_0)&=\max\Bigl\{\sup_{a\in\mathcal{Y}_z(w)}\bigl(\underline{F}^c_{Y_1|W}(a\mid w)-\overline{F}^c_{Y_0|W}(a-z\mid w)\bigr),0\Bigr\} \\ &=\max\Bigl\{\sup_{a\in\mathcal{Y}_z(w)}\bigl(\max\bigl\{\bigl[\delta_1(\theta_0)\bigr](a,1,w),\bigl[\delta_2(\theta_0)\bigr](a,1,w)\bigr\}\end{aligned}$$

$$
\begin{aligned}
& - \min\big\{\big[\delta_3(\theta_0)\big](a-z,0,w), \big[\delta_4(\theta_0)\big](a-z,0,w)\big\}\big), 0\Big\} \\
= \max\Big\{ & \sup_{a \in \mathcal{Y}_z(w)} \big(\big[\delta_1(\theta_0)\big](a,1,w) - \big[\delta_3(\theta_0)\big](a-z,0,w)\big), \\
& \sup_{a \in \mathcal{Y}_z(w)} \big(\big[\delta_1(\theta_0)\big](a,1,w) - \big[\delta_4(\theta_0)\big](a-z,0,w)\big), \\
& \sup_{a \in \mathcal{Y}_z(w)} \big(\big[\delta_2(\theta_0)\big](a,1,w) - \big[\delta_3(\theta_0)\big](a-z,0,w)\big), \\
& \sup_{a \in \mathcal{Y}_z(w)} \big(\big[\delta_2(\theta_0)\big](a,1,w) - \big[\delta_4(\theta_0)\big](a-z,0,w)\big),\ 0\Big\}.
\end{aligned}
$$

By linearity $(\delta_j - \delta_k)(\theta)$ is Hadamard differentiable at θ_0 for $j = 1, 2$ and $k = 3, 4$. By the chain rule (Lemma C2) and Lemma C4, the mappings

$$
\theta \mapsto \sup_{a \in \mathcal{Y}_z(w)} \big(\big[\delta_j(\theta)\big](a,1,w) - \big[\delta_k(\theta)\big](a-z,0,w)\big)
$$

are Hadamard directionally differentiable at θ_0 for $j = 1, 2$ and $k = 3, 4$. Finally, the maximum operator over five arguments is Hadamard directionally differentiable, and by another application of the chain rule, ϕ_4 is Hadamard directionally differentiable for fixed c and w. Uniformity over $c \in \mathcal{C}$ and $w \in \operatorname{supp}(W)$ is obtained from considering the vector of Hadamard directional derivatives for all $c \in \mathcal{C}$ and $w \in \operatorname{supp}(W)$.

By Lemma 3, the mapping $(F_{Y|X,W}(\cdot \mid \cdot, \cdot), p_{(\cdot|\cdot)}, q_{(\cdot)}) \mapsto \underline{P}(\cdot \mid \cdot)$ is Hadamard directionally differentiable. Linearity of the Hadamard directional derivative operator yields that the mapping $(F_{Y|X,W}(\cdot \mid \cdot, \cdot), p_{(\cdot|\cdot)}, q_{(\cdot)}) \mapsto \underline{\mathrm{CDTE}}(z, \cdot, \cdot \mid \cdot)$ is Hadamard directionally differentiable.

Since

$$
\inf_{a \in \mathcal{A}} \theta(a, c, w) = -\sup_{a \in \mathcal{A}} \big(-\theta(a, c, w)\big),
$$

the infimum operator is Hadamard directionally differentiable. As in the proof of Lemma 1, the minimum operator is Hadamard directionally differentiable. Following Lemma 3, the mapping $(F_{Y|X,W}(\cdot \mid \cdot, \cdot), p_{(\cdot|\cdot)}, q_{(\cdot)}) \mapsto \overline{P}(\cdot \mid \cdot)$ is Hadamard directionally differentiable. A similar argument as above implies the mapping $(F_{Y|X,W}(\cdot \mid \cdot, \cdot), p_{(\cdot|\cdot)}, q_{(\cdot)}) \mapsto \overline{\mathrm{CDTE}}(z, \cdot, \cdot \mid \cdot)$ is Hadamard directionally differentiable.

Combining these results with Lemma C1 allows us to conclude that

$$
\sqrt{N} \begin{pmatrix} \widehat{\overline{\mathrm{CDTE}}}(z, c, t \mid w) - \overline{\mathrm{CDTE}}(z, c, t \mid w) \\ \widehat{\underline{\mathrm{CDTE}}}(z, c, t \mid w) - \underline{\mathrm{CDTE}}(z, c, t \mid w) \end{pmatrix} \rightsquigarrow \tilde{\mathbf{Z}}_5(c, t, w),
$$

a tight random element of $\ell^\infty(\mathcal{C} \times [0,1] \times \operatorname{supp}(W), \mathbb{R}^2)$ with continuous paths.

Finally, to see that equation (29) holds, consider the lower bound estimator. We have

$$
\begin{aligned}
& \sqrt{N}\big(\widehat{\underline{\mathrm{DTE}}}(z, c, t) - \underline{\mathrm{DTE}}(z, c, t)\big) \\
& \quad = \sum_{k=1}^{K} \sqrt{N}\big(\widehat{\underline{\mathrm{CDTE}}}(z, c, t \mid w_k) - \underline{\mathrm{CDTE}}(z, c, t \mid w_k)\big) q_{w_k}
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{k=1}^{K}\underline{\mathrm{CDTE}}(z,c,t\mid w_k)\sqrt{N}(\widehat{q}_{w_k}-q_{w_k}) \\
&\rightsquigarrow \sum_{k=1}^{K}\tilde{\mathbf{Z}}_5(c,t,w_k)q_{w_k}+\sum_{k=1}^{K}\underline{\mathrm{CDTE}}(z,c,t\mid w_k)\mathbf{Z}_1^{(3)}(0,0,w_k).
\end{aligned}
$$

A similar derivation holds for the upper bound estimator. □

Proof of Theorem 2. By Lemmas 3 and C1, the numerator of equation (22) converges uniformly over $c\in\mathcal{C}$. By Lemmas 3 and 4, the denominator also converges uniformly over $c\in\mathcal{C}$. By the delta method,

$$
\begin{aligned}
&\sqrt{N}\big(\widehat{\mathrm{bf}}(c,\underline{p})-\mathrm{bf}(c,\underline{p})\big) \\
&=\sqrt{N}\left(\frac{1-\underline{p}-\sum_{k=1}^{K}\widehat{\overline{P}}(c\mid w_k)\widehat{q}_{w_k}}{1+\sum_{k=1}^{K}\Big[\min\Big\{\inf_{y\in\mathcal{Y}_0(w_k)}\big(\widehat{\overline{F}}^c_{Y_1|W}(y\mid w_k)-\widehat{\underline{F}}^c_{Y_0|W}(y\mid w_k)\big),0\Big\}-\widehat{\overline{P}}(c\mid w_k)\Big]\widehat{q}_{w_k}}\right. \\
&\qquad\left.-\frac{1-\underline{p}-\sum_{k=1}^{K}\overline{P}(c\mid w_k)q_{w_k}}{1+\sum_{k=1}^{K}\Big[\min\Big\{\inf_{y\in\mathcal{Y}_0(w_k)}\big(\overline{F}^c_{Y_1|W}(y\mid w_k)-\underline{F}^c_{Y_0|W}(y\mid w_k)\big),0\Big\}-\overline{P}(c\mid w_k)\Big]q_{w_k}}\right) \\
&\rightsquigarrow\frac{-\sum_{k=1}^{K}\mathbf{Z}_4^{(1)}(w_k,c)q_{w_k}-\sum_{k=1}^{K}\overline{P}(c\mid w_k)\mathbf{Z}_1^{(3)}(0,0,w_k)}{1+\sum_{k=1}^{K}\Big[\min\Big\{\inf_{y\in\mathcal{Y}_0(w_k)}\big(\overline{F}^c_{Y_1|W}(y\mid w_k)-\underline{F}^c_{Y_0|W}(y\mid w_k)\big),0\Big\}-\overline{P}(c\mid w_k)\Big]q_{w_k}} \\
&\quad-\frac{1-\underline{p}-\sum_{k=1}^{K}\overline{P}(c\mid w_k)q_{w_k}}{\left(1+\sum_{k=1}^{K}\Big[\min\Big\{\inf_{y\in\mathcal{Y}_0(w_k)}\big(\overline{F}^c_{Y_1|W}(y\mid w_k)-\underline{F}^c_{Y_0|W}(y\mid w_k)\big),0\Big\}-\overline{P}(c\mid w_k)\Big]q_{w_k}\right)^2}\tilde{\mathbf{Z}}(c),
\end{aligned}
$$

where

$$
\begin{aligned}
&\sqrt{N}\left(1-\underline{p}-\sum_{k=1}^{K}\widehat{\overline{P}}(c\mid w_k)\widehat{q}_{w_k}-\left[1-\underline{p}-\sum_{k=1}^{K}\overline{P}(c\mid w_k)q_{w_k}\right]\right) \\
&\rightsquigarrow-\sum_{k=1}^{K}\mathbf{Z}_4^{(1)}(w_k,c)q_{w_k}-\sum_{k=1}^{K}\overline{P}(c\mid w_k)\mathbf{Z}_1^{(3)}(0,0,w_k)
\end{aligned}
$$

and

$$\sqrt{N}\Bigg(\sum_{k=1}^{K}\Big[\min\Big\{\inf_{y\in\mathcal{Y}_0(w_k)}\big(\widehat{\overline{F}}^c_{Y_1|W}(y \mid w_k)-\widehat{\underline{F}}^c_{Y_0|W}(y \mid w_k)\big),0\Big\}-\widehat{\overline{P}}(c \mid w_k)\Big]\widehat{q}_{w_k}$$
$$-\sum_{k=1}^{K}\Big[\min\Big\{\inf_{y\in\mathcal{Y}_0(w_k)}\big(\overline{F}^c_{Y_1|W}(y \mid w_k)-\underline{F}^c_{Y_0|W}(y \mid w_k)\big),0\Big\}-\overline{P}(c \mid w_k)\Big]q_{w_k}\Bigg) \rightsquigarrow \tilde{\mathbf{Z}}(c).$$

Here $\tilde{\mathbf{Z}}(c)$ is a random element of $\ell^\infty(\mathcal{C})$ by Lemmas 3 and 4. Therefore,

$$\sqrt{N}\big(\widehat{\mathrm{bf}}(c,\underline{p})-\mathrm{bf}(c,\underline{p})\big)$$

converges to a random element in $\ell^\infty(\mathcal{C}\times\mathcal{P})$.

As discussed in the proof of Lemma 1, the maximum and minimum operators in equation (21) are Hadamard directionally differentiable. By Lemma C2, their composition is Hadamard directionally differentiable. Therefore, by the delta method for Hadamard directionally differentiable functions, $\sqrt{N}(\widehat{\mathrm{BF}}(c,\underline{p})-\mathrm{BF}(c,\underline{p}))$ converges in process as in the statement of the theorem. □

Lemma C5. *Let $h: A\to\mathbb{R}$ where $A\subseteq\mathbb{R}$. Let $F(h)=\sup_{x\in A}h(x)$. Let $\|\cdot\|_\infty$ denote the sup-norm $\|h\|_\infty=\sup_{x\in A}|h(x)|$. Then F is Lipschitz continuous with respect to the sup-norm $\|\cdot\|_\infty$ and has Lipschitz constant equal to one.*

Proof of Lemma C5. For functions h and h',

$$\begin{aligned}\sup_{x\in A}h(x)-\sup_{\tilde{x}\in A}h'(\tilde{x}) &= \sup_{x\in A}\Big(h(x)-\sup_{\tilde{x}\in A}h'(\tilde{x})\Big)\\ &\le \sup_{x\in A}\big(h(x)-h'(x)\big)\\ &\le \sup_{x\in A}\big|h(x)-h'(x)\big|.\end{aligned}$$

By a symmetric argument,

$$\begin{aligned}\sup_{x\in A}h'(x)-\sup_{\tilde{x}\in A}h(\tilde{x}) &\le \sup_{x\in A}\big|h'(x)-h(x)\big|\\ &= \sup_{x\in A}\big|h(x)-h'(x)\big|.\end{aligned}$$

Therefore $|F(h)-F(h')|\le\|h-h'\|_\infty$. □

Proof of Proposition 2. Hadamard directional differentiability of ϕ follows from the chain rule (Lemma C2) and from the proof of Theorem 2, since the breakdown frontier is a Hadamard directionally differentiable mapping of $\overline{P}(\cdot)=\mathbb{E}[\overline{P}(\cdot \mid W)]$ and $\overline{\mathrm{DTE}}(z,\cdot,\cdot)$, which are themselves Hadamard directionally differentiable mappings of θ_0.

Lemma C1 combined with Theorem 3.6.1 of van der Vaart and Wellner (1996) implies consistency of the nonparametric bootstrap for our underlying parameters: $\mathbf{Z}^*_N =$

$\sqrt{N}(\widehat{\theta}^* - \widehat{\theta}) \overset{P}{\rightsquigarrow} \mathbf{Z}_1$. By this result, $\varepsilon_N \to 0$, $\sqrt{N}\varepsilon_N \to \infty$, and Theorem 3.1 in Hong and Li (2018), equation (25) holds.

By $1/\sigma(c)$ being uniformly bounded, we have that

$$\frac{\left[\widehat{\phi}'_{\theta_0}\left(\sqrt{N}(\widehat{\theta}^* - \widehat{\theta})\right)\right](c, \underline{p})}{\sigma(c)} \overset{P}{\rightsquigarrow} \frac{\mathbf{Z}_{\mathrm{BF}}(c, \underline{p})}{\sigma(c)}.$$

By Lemma C5, the sup operator is Lipschitz with Lipschitz constant equal to 1. Therefore, by Proposition 10.7 on page 189 of Kosorok (2008), we can apply a continuous mapping theorem to get

$$\sup_{c \in \mathcal{C}} \frac{\left[\widehat{\phi}'_{\theta_0}\left(\sqrt{N}(\widehat{\theta}^* - \widehat{\theta})\right)\right](c, \underline{p})}{\sigma(c)} \overset{P}{\rightsquigarrow} \sup_{c \in \mathcal{C}} \frac{\mathbf{Z}_{\mathrm{BF}}(c, \underline{p})}{\sigma(c)}.$$

The rest of the proof follows from Corollary 3.2 of Fang and Santos (2015). □

Lemma C6. *Let $\overline{C} > 0$. Let $\mathcal{C} = \{c_1, \dots, c_J\} \subseteq [0, \overline{C}]$ be a finite grid of points. Let $f : [0, \overline{C}] \to \mathbb{R}_+$ be a nonincreasing function. Let $\widehat{LB}(\cdot)$ be an asymptotically exact uniform lower $1-\alpha$ confidence band for f on the grid points*:

$$\lim_{N \to \infty} \mathbb{P}\big(\widehat{LB}(c_j) \le f(c_j) \textit{ for } j = 1, \dots, J\big) = 1 - \alpha.$$

Define $\widetilde{LB} : [0, \overline{C}] \to \mathbb{R}_+$ by

$$\widetilde{LB}(c) = \begin{cases} \widehat{LB}(c_1) & \textit{if } c \in [0, c_1], \\ \vdots & \\ \widehat{LB}(c_j) & \textit{if } c \in (c_{j-1}, c_j], \textit{for } j = 2, \dots, J, \\ \vdots & \\ 0 & \textit{if } c \in (c_J, \overline{C}]. \end{cases}$$

Then $\widetilde{LB}(\cdot)$ is an asymptotically exact uniform lower $1-\alpha$ confidence band on $[0, \overline{C}]$:

$$\lim_{N \to \infty} \mathbb{P}\big(\widetilde{LB}(c) \le f(c) \textit{ for all } c \in [0, \overline{C}]\big) = 1 - \alpha.$$

Proof of Lemma C6. Define the events

$$A = \big\{\widehat{\mathrm{LB}}(c_j) \le f(c) \text{ for all } c \in (c_{j-1}, c_j], \text{ for } j = 1, \dots, J\big\}$$

and

$$B = \big\{\widehat{\mathrm{LB}}(c_j) \le f(c_j) \text{ for } j = 1, \dots, J\big\}.$$

A immediately implies B. Since f is nonincreasing, B implies A. Thus

$$\begin{aligned} \mathbb{P}\big(\widetilde{\mathrm{LB}}(c) \le f(c) \text{ for all } c \in [0, \overline{C}]\big) &= \mathbb{P}\big(\widehat{\mathrm{LB}}(c_j) \le f(c) \text{ for all } c \in (c_{j-1}, c_j], \text{ for } j = 1, \dots, J\big) \\ &= \mathbb{P}\big(\widehat{\mathrm{LB}}(c_j) \le f(c_j) \text{ for } j = 1, \dots, J\big). \end{aligned}$$

The first line follows by definition of $\widetilde{\mathrm{LB}}$ and since f is nonnegative. Taking limits as $N \to \infty$ gives

$$\begin{aligned}\lim_{N\to\infty} \mathbb{P}\big(\widetilde{\mathrm{LB}}(c) \le f(c) \text{ for all } c \in [0, \overline{C}]\big) &= \lim_{N\to\infty} \mathbb{P}\big(\widehat{\mathrm{LB}}(c_j) \le f(c_j) \text{ for } j = 1, \dots, J\big) \\ &= 1 - \alpha,\end{aligned}$$

where the last equality follows from the validity of the band $\widehat{\mathrm{LB}}(\cdot)$ on $\mathcal{C}$. □

Proof of Corollary 1. This follows immediately from Proposition 2 and Lemma C6. □

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. [78]

Altonji, J. G., T. E. Elder, and C. R. Taber (2005), "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools." *Journal of Political Economy*, 113, 151–184. [69]

Altonji, J. G., T. E. Elder, and C. R. Taber (2008), "Using selection on observed variables to assess bias from unobservables when evaluating Swan–Ganz catheterization." *American Economic Review P&P*, 98, 345–350. [69]

Angrist, J. D. (1990), "Lifetime earnings and the Vietnam era draft lottery: Evidence from social security administrative records." *The American Economic Review*, 313–336. [65]

Annan, J., C. Blattman, and R. Horton (2006), "The State of Youth and Youth Protection in Northern Uganda." Final Report for UNICEF Uganda. [64]

Bedoya, G., L. Bittarello, J. Davis, and N. Mittag (2017), "Distributional impact analysis: Toolkit and illustrations of impacts beyond the average treatment effect." Policy Research Working Paper No. 8139, World Bank. [65]

Bickel, P. J. and K. A. Doksum (2015), *Mathematical Statistics: Basic Ideas and Selected Topics*, Vol. 2. CRC Press. [92]

Blattman, C. and J. Annan (2010), "The consequences of child soldiering." *The Review of Economics and Statistics*, 92, 882–898. [64, 65, 66, 67]

Brent, R. P. (1973), *Algorithms for Minimization Without Derivatives*. Prentice-Hall. [68]

Canay, I. A. and A. M. Shaikh (2017), "Practical and theoretical advances in inference for partially identified models." In *Advances in Economics and Econometrics: Eleventh World Congress*, Vol. 2 (B. Honoré, A. Pakes, M. Piazzesi, and L. Samuelson, eds.), 271–306, Cambridge University Press. [54]

Cao, R., A. Cuevas, and W. G. Manteiga (1994), "A comparative study of several smoothing methods in density estimation." *Computational Statistics & Data Analysis*, 17, 153–176. [83]

Card, D. and A. R. Cardoso (2012), "Can compulsory military service raise civilian wages? Evidence from the peacetime draft in Portugal." *American Economic Journal: Applied Economics*, 4, 57–93. [65]

Chernozhukov, V., I. Fernández-Val, and A. Galichon (2010), "Quantile and probability curves without crossing." *Econometrica*, 78, 1093–1125. [70, 80, 81, 96, 97, 99]

Chesher, A. and A. Rosen (2017), "Generalized instrumental variable models." *Econometrica*, 85, 959–989. [43, 75]

Cornfield, J., W. Haenszel, E. C. Hammond, A. M. Lilienfeld, M. B. Shimkin, and E. L. Wynder (1959), "Smoking and lung cancer: Recent evidence and a discussion of some questions." *Journal of the National Cancer Institute*, 22, 173–203. [76]

De Angelis, D. and G. A. Young (1992), "Smoothing the bootstrap." *International Statistical Review*, 45–56. [84]

Donoho, D. L. and P. J. Huber (1983), "The notion of breakdown point." *A Festschrift for Erich L. Lehmann*, 157–184. [42]

Dümbgen, L. (1993), "On nondifferentiable functions and the bootstrap." *Probability Theory and Related Fields*, 95, 125–140. [55, 61, 78, 83]

Fan, Y., E. Guerre, and D. Zhu (2017), "Partial identification of functionals of the joint distribution of "potential outcomes"." *Journal of Econometrics*, 197, 42–59. [51, 78]

Fan, Y. and S. S. Park (2009), "Partial identification of the distribution of treatment effects and its confidence sets." In *Nonparametric Econometric Methods: Advances in Econometrics*, Vol. 25, 3–70, Emerald Group Publishing Limited. [78]

Fan, Y. and S. S. Park (2010), "Sharp bounds on the distribution of treatment effects and their statistical inference." *Econometric Theory*, 26, 931–951. [51]

Fan, Y. and A. J. Patton (2014), "Copulas in econometrics." *Annual Review of Economics*, 6, 179–200. [47]

Fang, Z. and A. Santos (2015), "Inference on directionally differentiable functions." Working paper. [63, 90, 100, 105]

Fang, Z. and A. Santos (2019), "Inference on directionally differentiable functions." *The Review of Economic Studies*, 86, 377–412. [55, 60, 78, 84, 91, 92, 100, 101]

Freyberger, J. and Y. Rai (2018), "Uniform confidence bands: Characterization and optimality." *Journal of Econometrics*, 204, 119–130. [62, 63]

Gechter, M. (2016), "Generalizing the results from social experiments: Theory and evidence from Mexico and India." Working paper. [78]

Gundersen, C., B. Kreider, and J. Pepper (2012), "The impact of the National School Lunch Program on child health: A nonparametric bounds analysis." *Journal of Econometrics*, 166, 79–91. [78]

Hampel, F. R. (1968), *Contributions to the Theory of Robust Estimation*. PhD Dissertation, University of California, Berkeley. [42]

Hampel, F. R. (1971), "A general qualitative definition of robustness." *The Annals of Mathematical Statistics*, 1887–1896. [42]

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel (1986), *Robust Statistics: The Approach Based on Influence Functions*. John Wiley & Sons. [76]

Hansen, B. E. (2004), "Bandwidth selection for nonparametric distribution estimation." Working paper. [67, 71, 84]

Hansen, B. E. (2017), "Regression kink with an unknown threshold." *Journal of Business & Economic Statistics*, 35, 228–240. [78]

Heckman, J. J., J. Smith, and N. Clements (1997), "Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts." *The Review of Economic Studies*, 64, 487–535. [65, 75, 78]

Ho, K. and A. M. Rosen (2017), "Partial identification in applied research: Benefits and challenges." In *Advances in Economics and Econometrics: Eleventh World Congress* (B. Honoré, A. Pakes, M. Piazzesi, and L. Samuelson, eds.), Econometric Society Monographs, Vol. 2, 307–359, Cambridge University Press. [42, 77]

Hong, H. and J. Li (2018), "The numerical delta method." *Journal of Econometrics*, 206, 379–394. [61, 78, 83, 84, 105]

Horowitz, J. L. and S. Lee (2012), "Uniform confidence bands for functions estimated nonparametrically with instrumental variables." *Journal of Econometrics*, 168, 175–188. [83]

Horowitz, J. L. and S. Lee (2017), "Nonparametric estimation and inference under shape restrictions." *Journal of Econometrics*, 201, 108–126. [83]

Horowitz, J. L. and C. F. Manski (1995), "Identification and robustness with contaminated and corrupted data." *Econometrica*, 63, 281–302. [42, 48, 76, 77, 78]

Huber, P. J. (1964), "Robust estimation of a location parameter." *The Annals of Mathematical Statistics*, 35, 73–101. [42]

Huber, P. J. and E. M. Ronchetti (2009), *Robust Statistics*. John Wiley & Sons. [76]

Imbens, G. W. (2003), "Sensitivity to exogeneity assumptions in program evaluation." *American Economic Review P&P*, 93, 126–132. [43, 65, 69, 77]

Imbens, G. W. and D. B. Rubin (2015), *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press. [76]

Kaido, H. (2016), "A dual approach to inference for partially identified econometric models." *Journal of Econometrics*, 192, 269–290. [78]

Kline, P. and A. Santos (2013), "Sensitivity to missing data assumptions: Theory and an evaluation of the U.S. wage structure." *Quantitative Economics*, 4, 231–267. [42, 43, 62, 77]

Kosorok, M. R. (2008), *Introduction to Empirical Processes and Semiparametric Inference*. Springer Science & Business Media. [105]

Kreider, B., J. V. Pepper, C. Gundersen, and D. Jolliffe (2012), "Identifying the effects of SNAP (food stamps) on child health outcomes when participation is endogenous and misreported." *Journal of the American Statistical Association*, 107, 958–975. [78]

Kreider, B., J. V. Pepper, and M. Roy (2016), "Identifying the effects of WIC on food insecurity among infants and children." *Southern Economic Journal*, 82, 1106–1122. [78]

Lee, Y.-Y. and D. Bhattacharya (2019), "Applied welfare analysis for discrete choice with interval-data on income." *Journal of Econometrics*, 211, 361–387. [78]

Léger, C. and J. P. Romano (1990), "Bootstrap choice of tuning parameters." *Annals of the Institute of Statistical Mathematics*, 42, 709–735. [83]

Li, Q. and J. S. Racine (2008), "Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data." *Journal of Business & Economic Statistics*, 26, 423–434. [66]

Makarov, G. (1982), "Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed." *Theory of Probability & Its Applications*, 26, 803–806. [51, 85]

Manski, C. F. (2007), *Identification for Prediction and Decision*. Harvard University Press. [42, 77]

Manski, C. F. (2013), "Response to the review of 'Public policy in an uncertain world'." *The Economic Journal*, 123, F412–F415. [42]

Manski, C. F. and J. V. Pepper (2018), "How do right-to-carry laws affect crime rates? Coping with ambiguity using bounded-variation assumptions." *The Review of Economics and Statistics*, 100, 232–244. [43, 77]

Marron, J. S. (1992), "Bootstrap bandwidth selection." In *Exploring the Limits of Bootstrap*, 249–262, John Wiley. [83]

Masten, M. A, and A. Poirier (2020), "Supplement to 'Inference on breakdown frontiers'." *Quantitative Economics Supplemental Material*, 11, https://doi.org/10.3982/QE1288. [45]

Masten, M. A. and A. Poirier (2016), "Partial independence in nonseparable models." cemmap Working Paper, CWP26/16. [46, 76]

Masten, M. A. and A. Poirier (2018a), "Identification of treatment effects under conditional partial independence." *Econometrica*, 86, 317–351. [43, 47, 49, 56, 67, 68, 78, 86]

Masten, M. A. and A. Poirier (2018b), "Interpreting quantile independence." Working paper. [76]

Matzkin, R. L. (2003), "Nonparametric estimation of nonadditive random functions." *Econometrica*, 71, 1339–1375. [58]

Mullahy, J. (2018), "Individual results may vary: Inequality-probability bounds for some health-outcome treatment effects." *Journal of Health Economics*, 61, 151–162. [65]

Nelsen, R. B. (2006), *An Introduction to Copulas*, second edition. Springer. [47, 48]

Oster, E. (2019), "Unobservable selection and coefficient stability: Theory and evidence." *Journal of Business & Economic Statistics*, 37, 187–204. [69]

Polansky, A. M. and W. Schucany (1997), "Kernel smoothing to improve bootstrap confidence intervals." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59, 821–838. [84]

Robins, J. M. (1999), "Association, causation, and marginal structural models." *Synthese*, 121, 151–179. [69]

Robins, J. M., A. Rotnitzky, and D. O. Scharfstein (2000), "Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models." In *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, 1–94, Springer. [78]

Rosenbaum, P. R. (1995), *Observational Studies*. Springer. [78]

Rosenbaum, P. R. (2002), *Observational Studies*, second edition. Springer. [78]

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*, 212–218. [76, 78]

Rotnitzky, A., J. M. Robins, and D. O. Scharfstein (1998), "Semiparametric regression for repeated outcomes with nonignorable nonresponse." *Journal of the American Statistical Association*, 93, 1321–1339. [69]

Shapiro, A. (1990), "On concepts of directional differentiability." *Journal of Optimization Theory and Applications*, 66, 477–487. [55, 89]

Sklar, M. (1959), "Fonctions de répartition à n dimensions et leurs marges." *Publications de l'Institut de statistique de l'Université de Paris*, 8, 229–231. [47]

Stoye, J. (2005), *Essays on Partial Identification and Statistical Decisions.* PhD Dissertation, Northwestern University. [42, 52, 77]

Stoye, J. (2010), "Partial identification of spread parameters." *Quantitative Economics*, 1, 323–357. [42, 77]

Taylor, C. C. (1989), "Bootstrap choice of the smoothing parameter in kernel density estimation." *Biometrika*, 76, 705–712. [83]

Torgovitsky, A. (2019), "Partial identification by extending subdistributions." *Quantitative Economics*, 10, 105–144. [43, 75]

van der Vaart, A. and J. Wellner (1996), *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Science & Business Media. [60, 79, 104]

van der Vaart, A. W. (2000), *Asymptotic Statistics*. Cambridge University Press. [56, 87, 92, 95]

Co-editor Christopher Taber handled this manuscript.

Manuscript received 2 February, 2019; final version accepted 8 August, 2019; available online 28 August, 2019.