Inference on Breakdown Frontiers*

Matthew A. Masten† Alexandre Poirier‡

October 20, 2017

Abstract

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 deviations from random assignment of treatment, and one which indexes deviations from 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. This frontier provides a quantitative measure of robustness of conclusions to deviations from the 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 the conclusions we consider are fairly robust to failure of rank invariance, when random assignment holds, but these conclusions are much more sensitive to both assumptions for small deviations from random assignment.

JEL classification: C14; C18; C21; C25; C51

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

*This paper was presented at the 2017 IAAE Conference, the 2017 "Inference in Nonstandard Problems" CEME Conference, the 2017 University of Chicago Interactions Workshop, Duke, UC Irvine, the University of Pennsylvania, UC Berkeley, Yale, the University of Georgia, and Auburn University. We thank audiences at those conferences and seminars, 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.

†Department of Economics, Duke University, matt.masten@duke.edu

‡Department of Economics, University of Iowa, alexandre-poirier@uiowa.edu

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 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 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 (for example, 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 non-tight bounds critique of partial identification methods (for example, see section 7.2 of Ho and Rosen 2016). 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 (2013), who proposed doing inference on a breakdown point. Finally, our focus on multi-dimensional frontiers builds on the graphical sensitivity analysis of Imbens (2003). We discuss these papers and others in more detail at the end of this section.

To illustrate the breakdown frontier approach, we study a simple potential outcomes model with a binary treatment. 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

[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) give 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.

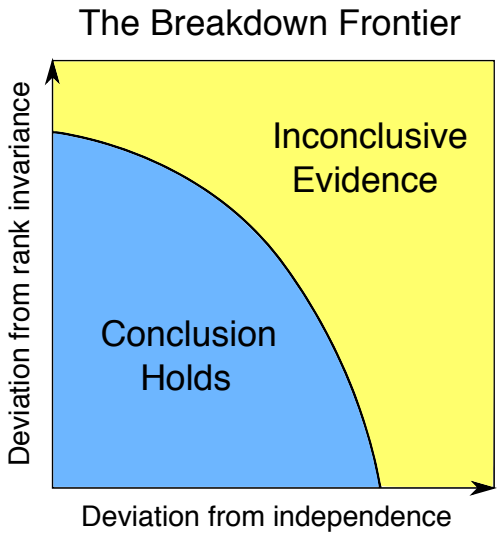


Figure 1: An example breakdown frontier, partitioning the space of assumptions into the set for which our conclusion of interest holds (the robust region) and the set for which our evidence is inconclusive.

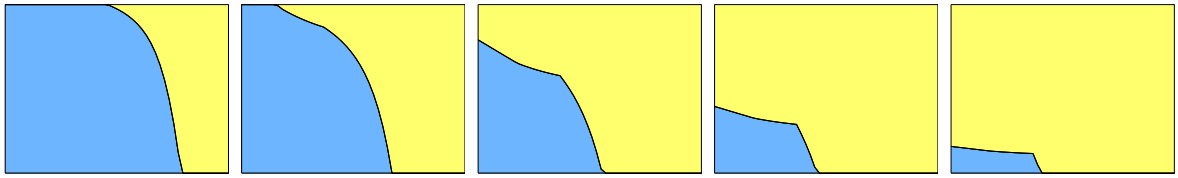
identified. One may be concerned, however, that these two assumptions are too strong. We relax rank invariance by supposing that there are two types of units in the population: one type for which rank invariance holds and another type for which it may not. The proportion t of the second type is a measure of deviation from the rank invariance assumption. We relax random assignment using a propensity score distance $c \geq 0$ as in our previous work, Masten and Poirier (2017). We give more details on both of these relaxations in section 2. We derive the sharp 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 deviation from the random assignment assumption. The vertical axis measures t, the deviation from rank invariance. The origin represents the 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 assignment. 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.

Figure 2a illustrates how the breakdown frontier changes as our conclusion of interest changes.

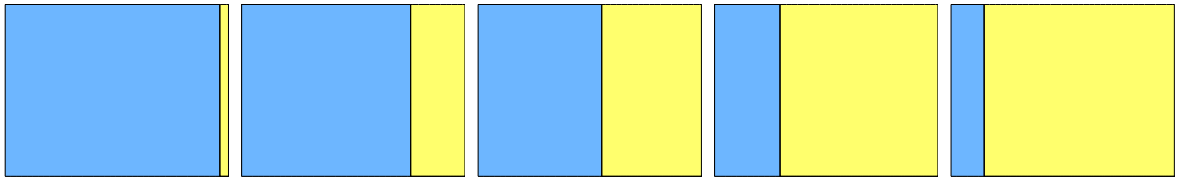
Figure 2: Example breakdown frontiers

(a) For the claim $\mathbb{P}(Y_1 > Y_0) \geq \underline{p}$, for five different values of $\underline{p}$.



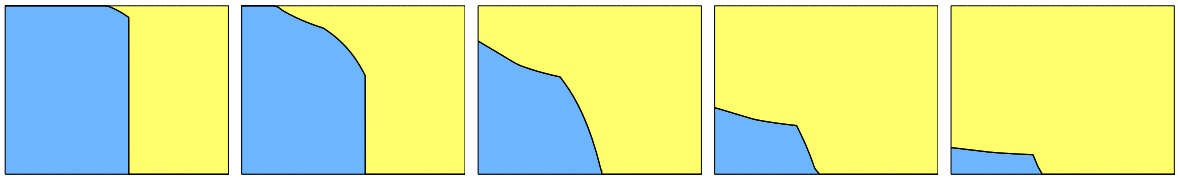
$\underline{p}$ =0.1 $\underline{p}$ =0.25 $\underline{p}$ =0.5 $\underline{p}$ =0.75 $\underline{p}$ =0.9

(b) For the claim ATE $\geq \mu$, for five different values of μ.



$\mu = -1.5$ $\mu = -0.5$ $\mu = 0$ $\mu = 0.5$ $\mu = 0.75$

(c) For the joint claim $\mathbb{P}(Y_1 > Y_0) \geq \underline{p}$ and ATE ≥ 0, for five different values of $\underline{p}$.



$\underline{p}$ =0.1, $\mu = 0$ $\underline{p}$ =0.25, $\mu = 0$ $\underline{p}$ =0.5, $\mu = 0$ $\underline{p}$ =0.75, $\mu = 0$ $\underline{p}$ =0.9, $\mu = 0$

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 towards 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 towards zero, we are making progressively weaker claims about the true parameter, and hence the set of assumptions for which the conclusion holds grows larger.

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.

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.

We also study breakdown frontiers and derive asymptotic distributional results for the average treatment effect $\text{ATE} = \mathbb{E}(Y_1 - Y_0)$ and quantile treatment effects $\text{QTE}(\tau) = Q_{Y_1}(\tau) - Q_{Y_0}(\tau)$ for $\tau \in (0, 1)$. Because the breakdown frontier analysis for these two parameters is nearly identical, we focus on ATE for brevity. Since ATE does not rely on the dependence structure between potential outcomes, assumptions regarding rank invariance do not affect the identified set. Hence, for a specific conclusion like $\text{ATE} \geq 0$, the breakdown frontiers are vertical lines. This was not the case for the breakdown frontier for claims about $\mathbb{P}(Y_1 > Y_0)$, where the assumptions on independence of treatment assignment and on rank invariance interact.

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

$$\text{ATE} \geq \mu$$

for five different values for μ. The figure shows the corresponding breakdown frontiers. As μ increases, 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. This happens when $\mathbb{E}(Y \mid X = 1) - \mathbb{E}(Y \mid X = 0) < \mu$. Conversely, as μ decreases, we are making progressively weaker claims about the true parameter, and hence the set of assumptions for which the conclusion holds grows larger.

Breakdown frontiers can be defined for many simultaneous claims. For example, figure 2c illustrates breakdown frontiers for the joint claim that $\mathbb{P}(Y_1 > Y_0) \geq \underline{p}$ and $\text{ATE} \geq 0$, for five different values of $\underline{p}$. Compare these plots to those of figures 2a and 2b. For small $\underline{p}$, many pairs (c, t) lead to the conclusion $\mathbb{P}(Y_1 > Y_0) \geq \underline{p}$. But many of these pairs do not also lead to the conclusion $\mathbb{E}(Y_1 - Y_0) \geq 0$, and hence those pairs are removed. As $\underline{p}$ increases, however, the constraint that we also want to conclude $\text{ATE} \geq 0$ eventually does not bind. Similarly, if we look at the breakdown frontier solely for our claim about ATE, many points (c, t) are included which are ruled out when we also wish to make our claim about $\mathbb{P}(Y_1 > Y_0)$. This is especially true as $\underline{p}$ gets larger.

Although we focus on one-sided claims like $\mathbb{P}(Y_1 > Y_0) \geq \underline{p}$, one may also be interested in the simultaneous claim that $\mathbb{P}(Y_1 > Y_0) \geq \underline{p}$ and $\mathbb{P}(Y_1 > Y_0) \leq \overline{p}$, where $\underline{p} < \overline{p}$. Similarly to the joint one-sided claims about $\mathbb{P}(Y_1 > Y_0)$ and ATE discussed above, the breakdown frontier for this two-sided claim on $\mathbb{P}(Y_1 > Y_0)$ can be obtained by taking minimum of the two breakdown frontier functions for each one-sided claim separately. Equivalently, the area under the breakdown frontier for the two-sided claim is just the intersection of the areas under the breakdown frontiers for the one-sided claims.

As mentioned above, although identified sets are used in its definition, the breakdown frontier is always point identified. Hence for estimation and inference we use mostly standard nonparametric techniques. We first propose a consistent sample-analog estimator of the breakdown frontier. We then do inference via two asymptotically valid bootstrap procedures. Our main approach, discussed in section 3, relies on important work by Dümbgen (1993), Fang and Santos (2015), and Hong and Li (2015). We discuss an alternative approach in appendix C. In both approaches we construct one-sided lower uniform confidence bands for the breakdown frontier. Figure 6 on page 31 illustrates such a band. A one-sided band is appropriate because our inferential goal is to determine the set of assumptions for which we can still draw our conclusion. If α denotes the nominal size, then approximately $(1-\alpha)100\%$ of the time, all elements (c,t) which are below the confidence band lead to population level identified sets for the parameters of interest which allow us to conclude that our conclusions of interest hold; we also discuss a testing interpretation in appendix A. Keep in mind that our reason for using one-sided confidence bands is unrelated to whether our conclusion of interest is a one-sided or two-sided claim. We examine the finite sample performance of our main approach in several Monte Carlo simulations.

Finally, we use our results to examine the role of assumptions in determining the effects of child soldiering on wages, as studied in Blattman and Annan (2010). We illustrate how our results can be used as a sensitivity analysis within a larger empirical study. Specifically, we begin by first estimating and doing inference on parameters under point identifying assumptions. Next, we estimate breakdown frontiers for several claims about these parameters. We present our one-sided confidence bands as well. We then use these breakdown frontier point estimates and confidence bands to judge the sensitivity of our conclusion to the point identifying assumptions: A large inner confidence set for the robust region which includes many different points (c,t) suggests robust results while a small inner confidence set close to the origin suggests non-robust results.

Related literature

In the rest of this section, we review the 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 pages 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 non-tight bounds; see section 7.2 of Ho and Rosen 2016). 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 deviations from the missing-at-random assumption in a model of missing data. Kline and Santos (2013) study a different class of deviations from 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) studies 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 non-monotonicities 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.

Neither Horowitz and Manski (1995) nor Stoye (2005, 2010) discuss estimation or inference of breakdown points. Imbens (2003) estimates his level sets in an empirical application, but does not discuss inference. Kline and Santos (2013), on the other hand, is the first and only paper we're aware of that explicitly suggests estimating and doing inference on a breakdown point. We build on their work by proposing to do inference on the multi-dimensional breakdown frontier. This allows us to study the tradeoff 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 figures 2a, 2b, and 2c. Inference on a frontier rather than a point also raises additional issues they did not discuss; see our appendix A 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.

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 (2015). We then use the numerical bootstrap of Dümbgen (1993) and Hong and Li (2015) to construct our confidence bands. For other applications of Hadamard directional differentiability, see Lee and Bhattacharya (2016), Kaido (2016), and Hansen (2017).

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 (2017), 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) provide a comprehensive discussion of this literature; also see Abbring and Heckman (2007) section 2. Here we focus on the two 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. Fan and Park (2009) provide formal identification results for the joint cdf of potential outcomes and the distribution of treatment effects under the known Kendall's τ assumption. They provide estimation and inference methods for their bounds, but do not study breakdown points. Finally, none of these papers studies the specific deviation from 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 A. In that section we also briefly discuss Bayesian inference, although we use frequentist inference in this paper.

2 Model and identification

To illustrate the breakdown frontier approach, we study 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 deviation 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. We observe the scalar outcome variable

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

For simplicity we omit covariates throughout most of this paper.

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

Assumption A1. For each $x, x' \in \{0,1\}$:

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

Via A1.1, we restrict attention to continuously distributed potential outcomes. A1.2 states that the unconditional and conditional supports of Y_x are equal, and are a possibly infinite closed interval. This assumption implies that the endpoints $\underline{y}_x$ and $\overline{y}_x$ are point identified. We maintain A1.2 for simplicity, but it can be relaxed using similar derivations as in Masten and Poirier (2016). A1.3 is an overlap assumption.

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

Identifying assumptions

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

1. Random assignment of treatment: $X \perp\!\!\!\perp Y_1$ and $X \perp\!\!\!\perp Y_0$.
2. Rank invariance: $U_1 = U_0$ almost surely.

Note that the joint independence assumption $X \perp\!\!\!\perp (Y_1, Y_0)$ provides no additional identifying power beyond the marginal independence assumption stated above. Any functional of F_{Y_1,Y_0} 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 deviations from random assignment of treatment, and one which indexes deviations from 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.

We begin with our measure of distance from independence.

Definition 1. Let c be a scalar between 0 and 1. Say X is *c-dependent* on the potential outcomes Y_x if

$$\sup_{y \in \text{supp}(Y_x)} |\mathbb{P}(X = x \mid Y_x = y) - \mathbb{P}(X = x)| \leq c \tag{2}$$

for $x \in \{0, 1\}$.

For $c = 0$, c-dependence implies $X \perp\!\!\!\perp Y_1$ and $X \perp\!\!\!\perp Y_0$. For $c > 0$, however, c-dependence allows for some deviations from independence. We discuss c-dependence in detail in our previous paper Masten and Poirier (2017).

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

$$F_{Y_1,Y_0}(y_1, y_0) = C(F_{Y_1}(y_1), F_{Y_0}(y_0))$$

where $C(\cdot, \cdot)$ is a unique 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

$$C(u_1, u_0) = \min\{u_1, u_0\} \equiv M(u_1, u_0),$$

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

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

We next define a class of assumptions which includes both 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 rank invariance* if their copula C satisfies

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

where $M(u_1, u_0) = \min\{u_1, u_0\}$ and H is an arbitrary unknown copula.

This assumption says the population is a mixture of two parts: In one part, rank invariance holds. This part contains $100 \cdot t\,\%$ of the overall population. In the second part, rank invariance fails in an arbitrary, unknown way. Hence, for this part, the dependence structure is unconstrained beyond the Fréchet-Hoeffding bounds. This part contains $100(1-t)\%$ of the overall population. Thus for $t = 0$ the usual 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 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 $\mathbb{P}(T = 1) = t$, where T is independent of (Y_1, Y_0). Suppose that individuals for whom $T_i = 1$ have an arbitrary dependence structure, while those with $T_i = 0$ have rank invariant potential outcomes. Then by the law of iterated expectations,

$$\begin{aligned} F_{Y_1,Y_0}(y_1, y_0) &= (1-t)F_{Y_1,Y_0|T}(y_1, y_0 \mid 0) + tF_{Y_1,Y_0|T}(y_1, y_0 \mid 1) \\ &= (1-t)M(F_{Y_1}(y_1), F_{Y_0}(y_0)) + tH(F_{Y_1}(y_1), F_{Y_0}(y_0)). \end{aligned}$$

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 always available and hence can be used to facilitate breakdown frontier analyses.

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

Assumption A2.

1. X is c-dependent of the potential outcomes Y_x, where $c < \min\{p_1, p_0\}$.
2. (Y_1, Y_0) satisfies $(1-t)$-percent rank invariance where $t \in [0, 1]$.

For brevity we focus on the case $c < \min\{p_1, p_0\}$ throughout this section.

Partial identification under deviations from independence and rank invariance

We next study identification under the deviations from full independence and rank invariance defined above. We begin by briefly recalling results from Masten and Poirier (2017) on identification of the quantile treatment effect $\text{QTE}(\tau) = Q_{Y_1}(\tau) - Q_{Y_0}(\tau)$, the average treatment effect $\text{ATE} = \mathbb{E}(Y_1 - Y_0)$, and the 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 (2017), we showed that A1 and A2.1 imply that the identified set for $\text{QTE}(\tau)$ is[2]

$$\left[\underline{\text{QTE}}(\tau, c), \overline{\text{QTE}}(\tau, c)\right] \equiv \left[\underline{Q}^c_{Y_1}(\tau) - \overline{Q}^c_{Y_0}(\tau),\, \overline{Q}^c_{Y_1}(\tau) - \underline{Q}^c_{Y_0}(\tau)\right] \tag{4}$$

where

$$\begin{aligned} \overline{Q}^c_{Y_x}(\tau) &= Q_{Y|X}\left(\tau + \frac{c}{p_x}\min\{\tau, 1-\tau\} \mid x\right) \\ \underline{Q}^c_{Y_x}(\tau) &= Q_{Y|X}\left(\tau - \frac{c}{p_x}\min\{\tau, 1-\tau\} \mid x\right). \end{aligned} \tag{5}$$

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

$$[\underline{\text{ATE}}(c), \overline{\text{ATE}}(c)] \equiv \left[\int_0^1 \underline{\text{QTE}}(\tau, c)\, du,\, \int_0^1 \overline{\text{QTE}}(\tau, c)\, du\right] \tag{6}$$

and that

$$\overline{F}^c_{Y_x}(y) = \min\left\{\frac{F_{Y|X}(y \mid x)p_x}{p_x - c},\, \frac{F_{Y|X}(y \mid x)p_x + c}{p_x + c}\right\} \tag{7}$$

and

$$\underline{F}^c_{Y_x}(y) = \max\left\{\frac{F_{Y|X}(y \mid x)p_x}{p_x + c},\, \frac{F_{Y|X}(y \mid x)p_x - c}{p_x - c}\right\} \tag{8}$$

are functionally sharp bounds on the cdf F_{Y_x}. We use these marginal cdf bounds in our DTE bounds. These results are unchanged if we further impose A2.2. That is, assumptions on rank invariance 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 (DTE), the cdf

$$F_{Y_1-Y_0}(z) = \mathbb{P}(Y_1 - Y_0 \leq z)$$

for a fixed z. While the ATE only depends on the marginal distributions of potential outcomes, the DTE depends on the joint distribution of (Y_1, Y_0). Consequently, as we'll see below, the identified set for the DTE depends on the value of t.

For any $z \in \mathbb{R}$ define $\mathcal{Y}_z = [\underline{y}_1, \overline{y}_1] \cap [\underline{y}_0 + z, \overline{y}_0 + z]$. Note that $\text{supp}(Y_1 - Y_0) \subseteq [\underline{y}_1 - \overline{y}_0, \overline{y}_1 - \underline{y}_0]$. Let z be an element of $[\underline{y}_1 - \overline{y}_0, \overline{y}_1 - \underline{y}_0]$ such that $\mathcal{Y}_z$ is nonempty. If z is such that $\mathcal{Y}_z$ is empty, then the DTE is either 0 or 1 depending solely on the relative location of the two supports, which is point identified by A1.2.

[2]In that paper we also extended the bounds in equations (4)–(8) to the $c \geq \min\{p_1, p_0\}$ case.

Define

$$\overline{\text{DTE}}(z,c,t) = (1-t)\mathbb{P}(\underline{Q}^c_{Y_1}(U) - \overline{Q}^c_{Y_0}(U) \leq z) + t\left(1 + \min\left\{\inf_{y\in\mathcal{Y}_z}(\overline{F}^c_{Y_1}(y) - \underline{F}^c_{Y_0}(y-z)), 0\right\}\right)$$

$$\underline{\text{DTE}}(z,c,t) = (1-t)\mathbb{P}(\overline{Q}^c_{Y_1}(U) - \underline{Q}^c_{Y_0}(U) \leq z) + t\max\left\{\sup_{y\in\mathcal{Y}_z}(\underline{F}^c_{Y_1}(y) - \overline{F}^c_{Y_0}(y-z)), 0\right\}$$

where $U \sim \text{Unif}[0,1]$. The following result shows that these are sharp bounds on the DTE.

Theorem 1 (DTE bounds)**.** Suppose the joint distribution of (Y,X) is known. Suppose A1 and A2 hold. Let $z \in [\underline{y}_1 - \overline{y}_0, \overline{y}_1 - \underline{y}_0]$ be such that $\mathcal{Y}_z$ is nonempty. Then the identified set for $\mathbb{P}(Y_1 - Y_0 \leq z)$ is

$$[\underline{\text{DTE}}(z,c,t), \overline{\text{DTE}}(z,c,t)].$$

The bound functions $\underline{\text{DTE}}(z,\cdot,\cdot)$ and $\overline{\text{DTE}}(z,\cdot,\cdot)$ are continuous and monotonic. When both random assignment ($c=0$) and rank invariance ($t=0$) hold, these bounds collapse to a single point and we obtain point identification. If we impose random assignment ($c=0$) but allow arbitrary dependence between Y_1 and Y_0 ($t=1$) then we obtain the well known Makarov (1981) 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 et al. (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

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

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 sensitivity of this parameter to deviations from full independence and rank invariance.

Finally, notice that all of the bounds and identified sets discussed in this section are analytically tractable and depend on just two functions identified from the population—the conditional cdf $F_{Y|X}$ and the probability masses p_x. 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 random assignment and 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: What are the *weakest* assumptions that allow us to draw this conclusion, given the observed distribution of (Y, X)? 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 $1 - \overline{\mathrm{DTE}}(z, c, t)$ (by theorem 1), the robust region for the conclusion that $\mathbb{P}(Y_1 - Y_0 > z) \geq \underline{p}$ is

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

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 $\overline{\mathrm{DTE}}(z, c, t)$ is increasing in c and t, the robust region will be empty if $\overline{\mathrm{DTE}}(z, c, t) > 1 - \underline{p}$, and non-empty if $\overline{\mathrm{DTE}}(z, c, t) \leq 1 - \underline{p}$. That is, if the conclusion of interest doesn't even 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 $\overline{\mathrm{DTE}}(z, 0, 0) \leq 1 - \underline{p}$ so that $\mathrm{RR}(z, \underline{p}) \neq \emptyset$.

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

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

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

$$\mathrm{bf}(z, c, \underline{p}) = \frac{1 - \underline{p} - \mathbb{P}(\underline{Q}^c_{Y_1}(U) - \overline{Q}^c_{Y_0}(U) \leq z)}{1 + \min\left\{\inf_{y \in \mathcal{Y}_z}(\overline{F}^c_{Y_1}(y) - \underline{F}^c_{Y_0}(y - z)), 0\right\} - \mathbb{P}(\underline{Q}^c_{Y_1}(U) - \overline{Q}^c_{Y_0}(U) \leq z)}. \tag{9}$$

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

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

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

section 5.

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

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

and the breakdown frontier is

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

These sets are nonempty if $\underline{\mathrm{ATE}}(0) \geq \mu$; 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

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

This point c^* is a breakdown point for the conclusion that ATE $\geq \mu$. Note that continuity of $\underline{\mathrm{ATE}}(\cdot)$ implies $\underline{\mathrm{ATE}}(c^*) = \mu$. 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 ATE $\geq \mu$ hold. Then the robust region for this joint conclusion is just the intersection of the two individual robust regions:

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

The breakdown frontier for the joint conclusion is just 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. For example, see figure 2c on page 4.

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 the probability of treatment. Hence we propose simple sample analog plug-in 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 conclude by discussing selection of the tuning parameter for this bootstrap procedure.

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 below may be used as inputs to traditional inference on partially identified parameters. See Canay and Shaikh (2017) for a survey of this literature.

To establish our main asymptotic results, we present a sequence of results. Each result focuses on a different component of the overall breakdown frontier: (1) the underlying conditional cdf of outcomes given treatment and the probability of treatment, (2) the bounds on the marginal distributions of potential outcomes, (3) the QTE bounds, (4) the ATE bounds, (5) breakdown points for ATE, (6) the DTE under rank invariance but without full independence, (7) the DTE without either rank invariance or full independence, and finally (8) the breakdown frontier itself. Results (2)–(8) are new, while (1) follows from standard arguments.

We first suppose we observe a random sample of data.

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

We begin with the underlying parameters $F_{Y|X}(y \mid x)$ and $p_x = \mathbb{P}(X = x)$. Let

$$\widehat{F}_{Y|X}(y \mid x) = \frac{\frac{1}{N}\sum_{i=1}^N \mathbb{1}(Y_i \leq y)\mathbb{1}(X_i = x)}{\frac{1}{N}\sum_{i=1}^N \mathbb{1}(X_i = x)}$$

and

$$\widehat{p}_x = \frac{1}{N}\sum_{i=1}^N \mathbb{1}(X_i = x)$$

denote the sample analog estimators of these two quantities. Let $\rightsquigarrow$ denote weak convergence. Using standard methods, the following result shows that these estimators converge uniformly in y and x to a mean-zero Gaussian process.

Lemma 1. Suppose A3 holds. Then

$$\sqrt{N}\begin{pmatrix} \widehat{F}_{Y|X}(y \mid x) - F_{Y|X}(y \mid x) \\ \widehat{p}_x - p_x \end{pmatrix} \rightsquigarrow \mathbf{Z}_1(y, x),$$

a mean-zero Gaussian process in $\ell^\infty(\mathbb{R} \times \{0,1\}, \mathbb{R}^2)$ with continuous paths and covariance kernel equal to

$$\begin{aligned}\boldsymbol{\Sigma}_1(y,x,\tilde{y},\tilde{x}) &= \mathbb{E}[\mathbf{Z}_1(y,x)\mathbf{Z}_1(\tilde{y},\tilde{x})'] \\ &= \begin{pmatrix} \frac{F_{Y|X}(\min\{y,\tilde{y}\}|x) - F_{Y|X}(y|x)F_{Y|X}(\tilde{y}|x)}{p_x}\mathbb{1}(x=\tilde{x}) & 0 \\ 0 & p_x\mathbb{1}(x=\tilde{x}) - p_x p_{\tilde{x}} \end{pmatrix}.\end{aligned}$$

Next consider the bounds (7) and (8) on the marginal distributions of potential outcomes. These population bounds are a functional ϕ_1 evaluated at $(F_{Y|X}(\cdot \mid \cdot), p_{(\cdot)})$ where $p_{(\cdot)}$ denotes the probability p_x as a function of $x \in \{0,1\}$. We estimate these bounds by the plug-in estimator $\phi_1(\widehat{F}_{Y|X}(\cdot,\cdot), \widehat{p}_{(\cdot)})$. 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 (2015).

Definition 3. Let $\mathbb{D}$ and $\mathbb{E}$ be Banach spaces with norms $\|\cdot\|_{\mathbb{D}}$ and $\|\cdot\|_{\mathbb{E}}$. Let $\mathbb{D}_\phi \subseteq \mathbb{D}$ and $\mathbb{D}_0 \subseteq \mathbb{D}$. The map $\phi : \mathbb{D}_\phi \to \mathbb{E}$ is *Hadamard directionally differentiable* at $\theta \in \mathbb{D}_\phi$ tangentially to $\mathbb{D}_0$ if there is a continuous map $\phi'_\theta : \mathbb{D}_0 \to \mathbb{E}$ such that

$$\lim_{n\to\infty} \left\| \frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} - \phi'_\theta(h) \right\|_{\mathbb{E}} = 0$$

for all sequences $\{h_n\} \subset \mathbb{D}$ and $\{t_n\} \in \mathbb{R}_+$ such that $t_n \searrow 0$, $\|h_n - h\|_{\mathbb{D}} \to 0$, $h \in \mathbb{D}_0$ as $n \to \infty$, and $\theta + t_n h_n \in \mathbb{D}_\phi$ for all n.

If we further have that ϕ'_θ is linear, then we say ϕ is *Hadamard differentiable* (proposition 2.1 of Fang and Santos 2015). Not every Hadamard directional derivative ϕ'_θ must be linear, however.

We use the functional delta method for Hadamard directionally differentiable mappings (e.g., theorem 2.1 in Fang and Santos 2015) 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 (7) and (8), we estimate these by

$$\begin{aligned}\widehat{\overline{F}}^c_{Y_x}(y) &= \min\left\{ \frac{\widehat{F}_{Y|X}(y \mid x)\widehat{p}_x}{\widehat{p}_x - c}, \frac{\widehat{F}_{Y|X}(y \mid x)\widehat{p}_x + c}{\widehat{p}_x + c} \right\} \\ \widehat{\underline{F}}^c_{Y_x}(y) &= \max\left\{ \frac{\widehat{F}_{Y|X}(y \mid x)\widehat{p}_x}{\widehat{p}_x + c}, \frac{\widehat{F}_{Y|X}(y \mid x)\widehat{p}_x - c}{\widehat{p}_x - c} \right\}.\end{aligned} \tag{10}$$

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

Assumption A4.

1. For each $x \in \{0,1\}$, $-\infty < \underline{y}_x < \overline{y}_x < +\infty$.
2. For each $x \in \{0,1\}$, $F_{Y|X}(y \mid x)$ is continuously differentiable everywhere with density $f_{Y|X}(y \mid x)$ uniformly continuous in y, uniformly bounded from above, and uniformly bounded away from zero on $\text{supp}(Y \mid X = x)$.

A4.1 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 (10), 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 $\mathbb{P}(Q_{Y_1}(U) - Q_{Y_0}(U) \leq z)$. This is a well known issue when estimating quantile processes; for example, see van der Vaart (2000) lemma 21.4(ii). A4.2 requires the density of $Y \mid X$ to be bounded away from zero uniformly. This ensures that conditional quantiles of $Y \mid X$ are uniquely defined. It also implies that the limiting distribution of the estimated quantile bounds will be well-behaved. Uniform continuity of the density implies that the derivative of the conditional quantile function with respect to τ are uniformly continuous.

For some of our main results in this section, we establish convergence uniformly over $c \in \mathcal{C}$ for some finite grid $\mathcal{C} = \{c_1, c_2, \ldots, c_J\} \subset [0, \min\{p_1, p_0\})$. We constrain this grid to be below $\min\{p_1, p_0\}$ 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 $c \geq \min\{p_1, p_0\}$ case given in Masten and Poirier (2017). Uniformity over an interval of c does not hold since some of the functionals below are not Hadamard directionally differentiable when their 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.

The next result establishes convergence in distribution of the cdf bound estimators.

Lemma 2. Suppose A1, A3, and A4 hold. Let $\mathcal{Y} \subset \mathbb{R}$ be a finite grid of points. Then

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

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

$\mathbf{Z}_2$ is not Gaussian itself, but it is a continuous transformation of Gaussian processes. A characterization of this limiting process is given in the proof in appendix B.

Next consider the quantile bounds (5), which we estimate by

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

The next result establishes uniform convergence in distribution of these quantile bounds estimators. For the following results, let $\overline{C} \in (0, \min\{p_1, p_0\})$.

Lemma 3. Suppose A1, A3, and A4 hold. Then

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

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

This result is uniform in c on an interval, in $x \in \{0,1\}$, and in $\tau \in (0,1)$. This result directly implies convergence over $c \in \mathcal{C}$ as well. Unlike the distribution of cdf bounds estimators, this process is Gaussian. This follows by Hadamard differentiability of the mapping between $\theta_0 \equiv (F_{Y|X}(\cdot \mid \cdot), p_{(\cdot)})$ and the quantile bounds. By applying the functional delta method, we can show asymptotic normality of smooth functionals of these bounds. A first set of functionals are the QTE bounds of equation (4), which are a linear combination of the quantile bounds. Let

$$\widehat{\underline{\mathrm{QTE}}}(\tau, c) = \widehat{\underline{Q}}^c_{Y_1}(\tau) - \widehat{\overline{Q}}^c_{Y_0}(\tau) \qquad \text{and} \qquad \widehat{\overline{\mathrm{QTE}}}(\tau, c) = \widehat{\overline{Q}}^c_{Y_1}(\tau) - \widehat{\underline{Q}}^c_{Y_0}(\tau).$$

Then,

$$\sqrt{N}\begin{pmatrix} \widehat{\overline{\mathrm{QTE}}}(\tau, c) - \overline{\mathrm{QTE}}(\tau, c) \\ \widehat{\underline{\mathrm{QTE}}}(\tau, c) - \underline{\mathrm{QTE}}(\tau, c) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \mathbf{Z}_3^{(1)}(\tau, 1, c) - \mathbf{Z}_3^{(2)}(\tau, 0, c) \\ \mathbf{Z}_3^{(2)}(\tau, 1, c) - \mathbf{Z}_3^{(1)}(\tau, 0, c) \end{pmatrix},$$

where the superscript $\mathbf{Z}^{(j)}$ denotes the jth component of the vector $\mathbf{Z}$.

A second set of functionals are the ATE bounds from equation (6). These bounds are smooth linear functionals of the QTE bounds. Therefore the joint asymptotic distribution of these bounds can be established by the continuous mapping theorem. Let

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

Then, by the linearity of the integral operator, these estimated ATE bounds converge to their population counterpart at a $\sqrt{N}$-rate and therefore

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

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

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, $\underline{\text{ATE}}(0) > \mu$. Let

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

be the estimated breakdown point. This is the estimated smallest deviation from 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 non-decreasing and directionally differentiable everywhere. We now present a result about the asymptotic distribution of $\widehat{c}^*$. For a function $f : \mathbb{R} \to \mathbb{R}$, let

$$\partial_u^- f(u_0) = \lim_{t \nearrow 0} \frac{f(u_0+t) - f(u_0)}{t}$$

denote the *left*-derivative of f at u_0. Similarly, let

$$\partial_u^+ f(u_0) = \lim_{t \searrow 0} \frac{f(u_0+t) - f(u_0)}{t}$$

denote the *right*-derivative of f at u_0.

Proposition 1. Suppose A1, A3, and A4 hold. Assume $c^* \in [0,\overline{C}]$. Suppose also that $\partial_c^- \underline{\text{ATE}}(c^*) \neq 0$ and $\partial_c^+ \underline{\text{ATE}}(c^*) \neq 0$. Then $\sqrt{N}(\widehat{c}^* - c^*) \rightsquigarrow \mathbf{Z}_4$, a random variable.

The assumption that $c^* \in [0,\overline{C}]$ can be relaxed to the general case where $c^* \in [0,1]$ but we maintain the stronger assumption for brevity. The derivatives of the ATE lower bound with respect to c are assumed to be different from zero in both directions to allow for the limiting distribution of the inverse of this function to exist.

Under rank invariance, we can also establish asymptotic normality of bounds for $\mathbb{P}(Q_{Y_1}(U) - Q_{Y_0}(U) \leq z)$, which are given by $(\underline{P}(c),\overline{P}(c)) \equiv (\underline{\text{DTE}}(z,c,0),\overline{\text{DTE}}(z,c,0))$. Estimates for these quantities are provided by

$$\widehat{\underline{P}}(c) = \int_0^1 \mathbb{1}(\widehat{\overline{Q}}^c_{Y_1}(u) - \widehat{\underline{Q}}^c_{Y_0}(u) \leq z)\, du$$
$$\widehat{\overline{P}}(c) = \int_0^1 \mathbb{1}(\widehat{\underline{Q}}^c_{Y_1}(u) - \widehat{\overline{Q}}^c_{Y_0}(u) \leq z)\, du.$$

Asymptotic normality can be established using the Hadamard directional differentiability of the mapping from the differences in quantile bounds to the bounds $(\underline{P}(c),\overline{P}(c))$. This mapping 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 et al. (2010) to the case where the quantile function has a point of non-differentiability by showing Hadamard directional differentiability of this mapping.

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

Assumption A5. For each $c \in \mathcal{C}$,

1. The number of elements in each of the sets

$$\mathcal{U}_1^*(c) = \{u \in (0,1) : \partial_u^-(\overline{Q}_{Y_1}^c(u) - \underline{Q}_{Y_0}^c(u)) = 0 \text{ or } \partial_u^+(\overline{Q}_{Y_1}^c(u) - \underline{Q}_{Y_0}^c(u)) = 0\}$$

$$\mathcal{U}_2^*(c) = \{u \in (0,1) : \partial_u^-(\underline{Q}_{Y_1}^c(u) - \overline{Q}_{Y_0}^c(u)) = 0 \text{ or } \partial_u^+(\underline{Q}_{Y_1}^c(u) - \overline{Q}_{Y_0}^c(u)) = 0\}$$

   is finite.

2. The following hold.

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

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

These assumptions imply that the respective function's derivatives change signs a finite number of times, and 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 that the functions are 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 5 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 C. Like in Chernozhukov et al. (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 A5 is refutable, since it is expressed as a function of identified quantities, namely the QTE bounds for all $u \in (0, 1)$.

With this additional assumption we can show $\sqrt{N}$-convergence of the bounds uniformly in $\mathcal{C}$.

Lemma 4. Suppose A1, A3, A4, and A5 hold. Then

$$\sqrt{N}\begin{pmatrix} \widehat{\overline{P}}(c) - \overline{P}(c) \\ \widehat{\underline{P}}(c) - \underline{P}(c) \end{pmatrix} \rightsquigarrow \mathbf{Z}_5(c),$$

a tight random element in $\ell^\infty(\mathcal{C}, \mathbb{R}^2)$ with continuous paths.

If random assignment holds ($c = 0$) in addition to rank invariance ($t = 0$), then the DTE is point identified and lemma 4 gives the asymptotic distribution of the sample analog DTE estimator (in this case the upper and lower bound functions are equal). This can be considered an estimator of the DTE in one of the models of Matzkin (2003).

We now establish the limiting distribution of the DTE bounds uniformly in (c, t). Let

$$\widehat{\underline{\text{DTE}}}(z,c,t) = (1-t)\widehat{\underline{P}}(c) + t\max\left\{\sup_{a\in\mathcal{Y}_z}(\widehat{\underline{F}}^c_{Y_1}(a) - \widehat{\overline{F}}^c_{Y_0}(a-z)), 0\right\}$$
$$\widehat{\overline{\text{DTE}}}(z,c,t) = (1-t)\widehat{\overline{P}}(c) + t\left(1 + \min\left\{\inf_{a\in\mathcal{Y}_z}(\widehat{\overline{F}}^c_{Y_1}(a) - \widehat{\underline{F}}^c_{Y_0}(a-z)), 0\right\}\right).$$

We have shown in lemma 4 that the terms $\underline{P}(c)$ and $\overline{P}(c)$ are estimated at a $\sqrt{N}$-rate by the Hadamard directional differentiability of the mapping linking empirical cdfs and these terms. We now show that the second components of the DTE bounds are a Hadamard directionally differentiable functional as well, leading to the $\sqrt{N}$ joint convergence of these bounds to a tight, random element uniformly in c and t.

Lemma 5. Suppose A1, A3, A4, and A5 hold. Then

$$\sqrt{N}\begin{pmatrix} \widehat{\overline{\text{DTE}}}(z,c,t) - \overline{\text{DTE}}(z,c,t) \\ \widehat{\underline{\text{DTE}}}(z,c,t) - \underline{\text{DTE}}(z,c,t) \end{pmatrix} \rightsquigarrow \mathbf{Z}_6(c,t), \tag{11}$$

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

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

$$\widehat{\text{BF}}(0,c,\underline{p}) = \min\{\max\{\widehat{\text{bf}}(0,c,\underline{p}), 0\}, 1\} \tag{12}$$

where

$$\widehat{\text{bf}}(0,c,\underline{p}) = \frac{1 - \underline{p} - \widehat{\overline{P}}(c)}{1 + \min\left\{\inf_{y\in\mathcal{Y}_0}(\widehat{\overline{F}}^c_{Y_1}(y) - \widehat{\underline{F}}^c_{Y_0}(y)), 0\right\} - \widehat{\overline{P}}(c)}. \tag{13}$$

By combining our previous lemmas, we can show that the estimated breakdown frontier converges in distribution.

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

$$\sqrt{N}(\widehat{\text{BF}}(0,c,\underline{p}) - \text{BF}(0,c,\underline{p})) \rightsquigarrow \mathbf{Z}_7(c,\underline{p}),$$

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

This convergence is uniform over the deviation from independence c on any grid $\mathcal{C}$, and over the strength of the conclusion $\underline{p}$ on any grid $\mathcal{P}$. This result essentially follows from the convergence of

the preliminary estimators established in lemma 1 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 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* 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, 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 the appendix, 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 $Z_i = (Y_i, X_i)$ and $Z^N = \{Z_1, \ldots, Z_N\}$. 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

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

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 $\widehat{\theta}$:

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

For these choices, let $\mathbf{Z}_N^* = \sqrt{N}(\widehat{\theta}^* - \widehat{\theta})$. Recall from lemma 1 that $\mathbf{Z}_1$ denotes the limiting distribution of $\sqrt{N}(\widehat{\theta} - \theta_0)$. 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 (2015). They further show that ϕ is Hadamard differentiable *if and only if*

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

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 (2015) show that a different bootstrap procedure is consistent. Specifically, let $\widehat{\phi}'_{\theta_0}$ be a consistent estimator of ϕ'_{θ_0}. Then their results imply that

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

Analytical consistent estimates of ϕ'_{θ_0} are often difficult to obtain, so Dümbgen (1993) and Hong and Li (2015) propose 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

$$\widehat{\phi}'_{\theta_0}(\sqrt{N}(\widehat{\theta}^* - \widehat{\theta})) = \frac{\phi\left(\widehat{\theta} + \varepsilon_N\sqrt{N}(\widehat{\theta}^* - \widehat{\theta})\right) - \phi(\widehat{\theta})}{\varepsilon_N} \tag{14}$$

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 (2015) show

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

where the left hand side is defined in equation (14).

This bootstrap procedures requires evaluating ϕ at two values, which is computationally simple. It also requires selecting the tuning parameter ε_N, which we discuss later. 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 $\text{BF}(z, \cdot, \underline{p})$ to characterize this frontier. We focus on $z = 0$ and a single fixed $\underline{p}$ so that our breakdown frontier is for the claim that $\mathbb{P}(Y_1 > Y_0) \geq \underline{p}$. Thus our goal is to do inference on the function $\text{BF}(0, \cdot, \underline{p})$. We specifically construct one-sided *lower* uniform confidence bands. That is, we will construct a lower

band function $\widehat{\text{LB}}(c)$ such that

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

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

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

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

$$\lim_{N\to\infty} \mathbb{P}\left(\text{RR}_L \subseteq \text{RR}\right) = 1 - \alpha.$$

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 on page 48 in appendix A.

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 A for more discussion.

We consider bands of the form

$$\widehat{\text{LB}}(c) = \widehat{\text{BF}}(0, c, \underline{p}) - \widehat{k}(c)$$

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

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

or, equivalently, if

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

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 (2017) 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 A1, A3, A4, and A5 hold. Define $\phi : \ell^\infty(\mathbb{R} \times \{0, 1\}, \mathbb{R}^2) \to \ell^\infty(\mathcal{C})$ such that

$$\widehat{\text{BF}}(0, c, \underline{p}) = [\phi(\widehat{\theta})](c).$$

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

$$\left[\widehat{\phi}'_{\theta_0}(\sqrt{N}(\widehat{\theta}^* - \widehat{\theta}))\right] \overset{P}{\rightsquigarrow} \left[\phi'_{\theta_0}(\mathbf{Z}_1)\right] \equiv \mathbf{Z}_7. \tag{15}$$

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

$$\widehat{z}_{1-\alpha} = \inf\left\{ z \in \mathbb{R} : \mathbb{P}\left(\sup_{c \in \mathcal{C}} \frac{\left[\widehat{\phi}'_{\theta_0}(\sqrt{N}(\widehat{\theta}^* - \widehat{\theta}))\right](c, \underline{p})}{\sigma(c)} \leq z \mid Z^N \right) \geq 1 - \alpha \right\}. \tag{16}$$

Finally, suppose also that the cdf of

$$\sup_{c \in \mathcal{C}} \frac{[\phi'_{\theta_0}(\mathbf{Z}_1)](c, \underline{p})}{\sigma(c)} = \sup_{c \in \mathcal{C}} \frac{\mathbf{Z}_7(c, \underline{p})}{\sigma(c)}$$

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). As a consequence of this result, the lower $1 - \alpha$ band $\widehat{\text{LB}}(c) = \widehat{\text{BF}}(0, c, \underline{p}) - \widehat{z}_{1-\alpha}\sigma(c)$ 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:

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

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, $\widetilde{\text{LB}}(c)$ is a uniform lower $1-\alpha$ band for $\text{BF}(0, c, \underline{p})$ over $c \in [0, \overline{C}]$.

Rather than fixing the number of grid points, one could let $J \to \infty$ sufficiently slowly as $N \to \infty$ and show that any interpolated band gets arbitrarily close to a band uniform over $[0, \overline{C}]$. Horowitz and Lee (2012, 2017) discuss this approach in different settings.

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 let $\widehat{k}(c) = \widehat{z}_{1-\alpha}\widehat{\sigma}(c)$. 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 $\widehat{k}(c_1), \ldots, \widehat{k}(c_J)$ solve

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

subject to

$$\mathbb{P}\left(\sup_{c\in\{c_1,\ldots,c_J\}} \sqrt{N}\Big(\widehat{\text{BF}}(0, c, \underline{p}) - \text{BF}(0, c, \underline{p}) - k(c)\Big) \leq 0\right) = 1 - \alpha,$$

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.

Bootstrap selection of ε_N

While Dümbgen (1993) and Hong and Li (2015) provide rate constraints on ε_N, they do not recommend a procedure for picking ε_N in practice. In this section, we propose a bootstrap method for picking ε_N. We use this method for our empirical illustration in section 5; 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 $\text{CP}_N(\varepsilon; F_{Y,X,W})$ 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, $F_{Y,X,W}$. Here we allow for covariates W, as discussed in Masten and Poirier (2017). The bootstrap replaces $F_{Y,X,W}$ with an estimator $\widehat{F}_{Y,X,W}$. We pick a grid $\{\varepsilon_1, \ldots, \varepsilon_K\}$ of ε's and let $\widehat{\varepsilon}_N$ solve

$$\min_{k=1,\ldots,K} |\text{CP}_N(\varepsilon_k; \widehat{F}_{Y,X,W}) - (1-\alpha)|.$$

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 of section 4. Larger grids and larger values of B can be chosen subject to computational constraints. We furthermore must choose

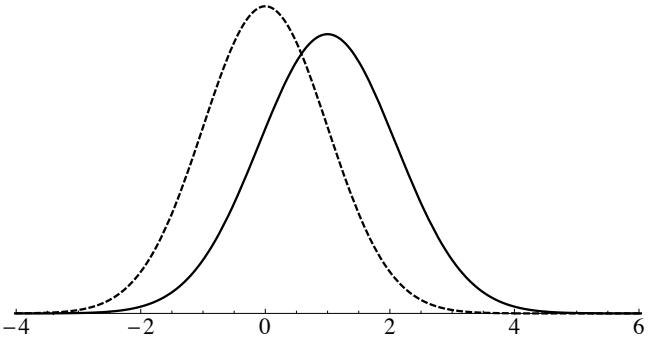


Figure 3: The two distributions of $Y \mid X = x$ for our Monte Carlo simulation. The dashed line is $Y \mid X = 0$ while the solid line is $Y \mid X = 1$.

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. 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.

4 Monte Carlo simulations

In this section we study the finite sample performance of our estimation and inference procedures. We consider the following dgp. For $x = 0, 1$, $Y \mid X = x$ has a truncated normal distribution, with density

$$f_{Y|X}(y \mid x) = \frac{1}{\gamma x + 1} \phi_{[-4,4]} \left(\frac{y - \pi x}{\gamma x + 1} \right),$$

where $\phi_{[-4,4]}$ is the truncated standard normal density. We let $\gamma = 0.1$ and $\pi = 1$. These two densities are shown in figure 3. We set $\mathbb{P}(X = 1) = 0.5$. This dgp implies a joint distribution of (Y, X), which we draw independently from. Figure 2a in the introduction shows population breakdown frontiers for this dgp.

We consider two sample sizes, $N = 500$ and $N = 2000$. For each sample size we generate $S = 500$ simulated datasets. In each dataset we compute the estimated breakdown frontier and a 95% lower bootstrap uniform confidence band, as discussed in section 3. We use $B = 1000$ bootstrap draws. We consider the same five values of $\underline{p}$ used in the introduction: 0.1, 0.25, 0.5, 0.75, and 0.9.

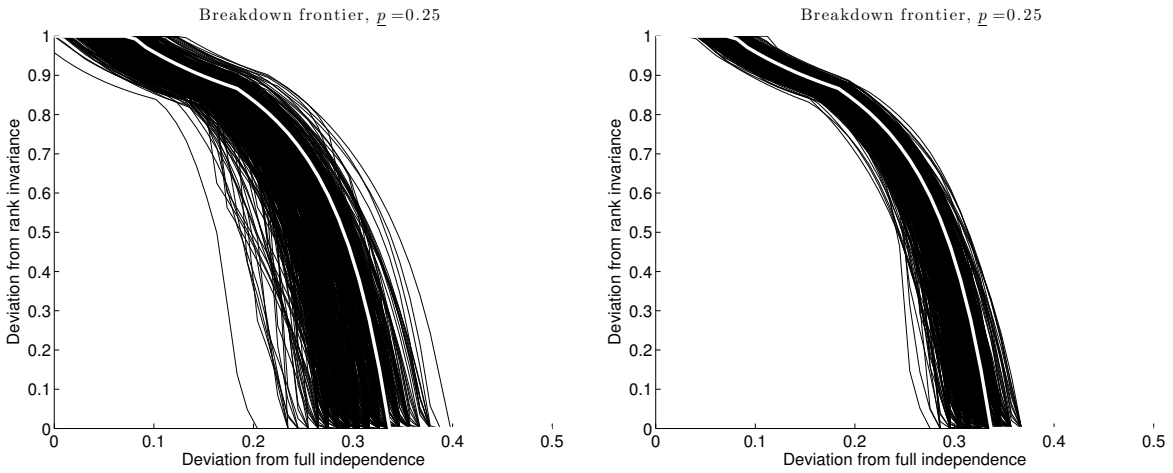


Figure 4: Left: $N = 500$. Right: $N = 2000$. These plots show the sampling distribution of our breakdown frontier estimator by gathering the point estimates of the breakdown frontier across all Monte Carlo simulations into one plot. The true breakdown frontier is shown on top in white.

First we consider the performance of our point estimator of the breakdown frontier. Figure 4 shows the sampling distribution of our breakdown frontier estimator. We show only $\underline{p} = 0.25$, but the other values of $\underline{p}$ yield similar figures. For this $\underline{p}$, we gather all point estimates of the breakdown frontier in the same plot. These plots show several features. First, as predicted by consistency, the sampling distribution becomes tighter around the truth as sample size increases. Second, the sampling distribution is not symmetric around the true frontier—it generally appears biased downwards. This is confirmed in figure 5 which plots the estimated finite sample mean function, $\widehat{\mathbb{E}}[\widehat{\text{BF}}(c)]$. This mean is estimated as the sample mean across all of our Monte Carlo datasets; that is, across all estimates shown in figure 4. The figure also shows the true breakdown frontier as a dotted line. In general the truth lies above the mean function. Again by consistency, this finite sample bias converges to zero as sample size increases, which we see when comparing the top row to the bottom row.

Next we consider the performance of our confidence bands. Figure 6 shows an example band along with the estimated frontier and the true frontier. To evaluate the performance of bands like this, we compute uniform coverage probabilities. We use 50 grid points for computing and evaluating uniform coverage of the confidence band. Table 1 shows the results. Here we present a range of choices for ε_N. Since $\varepsilon_N^{\text{naive}} = 1/\sqrt{N}$ yields the naive bootstrap, we use this choice as our baseline. We then consider seven other choices by rescaling the naive ε_N. Specifically, we consider $\varepsilon_N = K\varepsilon_N^{\text{naive}}$ for $K \in \{0.5, 1.5, 2, 4, 6, 8, 10\}$. Recall that Hong and Li (2015) impose the rate constraints that $\varepsilon_N \to 0$ and $\sqrt{N}\varepsilon_N \to \infty$. Hence asymptotically the ratio $\varepsilon_N/\varepsilon_N^{\text{naive}}$ must diverge.

First consider $N = 500$. For $\underline{p} = 0.1$, 0.25, and 0.75, the choice of ε_N which yields coverage probabilities closest to the nominal coverage of 0.95 is twice the naive choice. This is also approximately true for $\underline{p} = 0.5$. For $\underline{p} = 0.9$, the next largest ε_N has the coverage probability closest to the nominal coverage. Focusing on these choices of ε_N, the coverage probabilities are relatively close

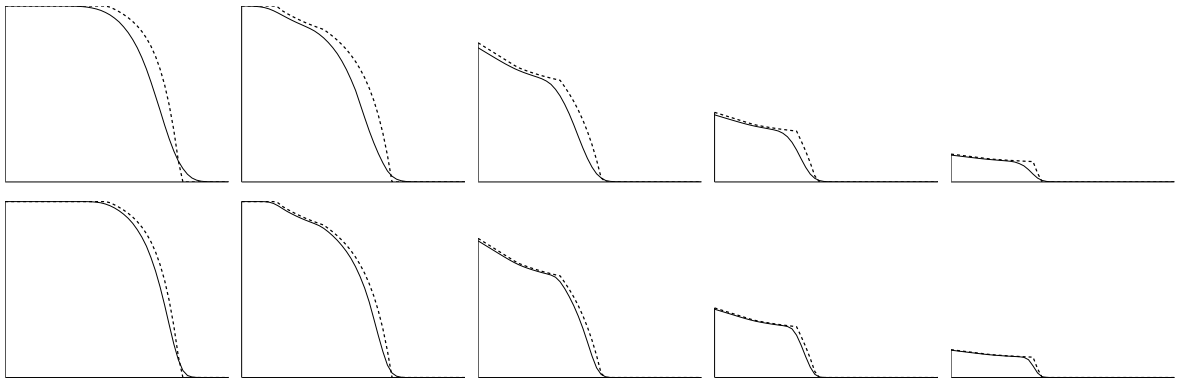


Figure 5: Rows are sample sizes (top is $N = 500$, bottom is $N = 2000$). Columns are five values of $\underline{p}$ (from left to right: $\underline{p} = 0.1$, 0.25, 0.5, 0.75, and 0.9). Dotted lines are the true breakdown frontiers. The solid lines are the Monte Carlo estimates of $\mathbb{E}[\widehat{\text{BF}}(c)]$. This plot shows the finite sample bias of our breakdown frontier estimator.

to the nominal for the 'outside' columns $\underline{p} = 0.1$, 0.25, and 0.9. For the 'inside' columns $\underline{p} = 0.25$ and $\underline{p} = 0.5$, we have substantial over-coverage. Indeed, for $\underline{p} = 0.1$, 0.25, and 0.5, all choices of ε_N's considered lead to over-coverage. For the two larger values of $\underline{p}$, some values of ε_N lead to under-coverage. Finally, with ε_N's large enough, we obtain 100% coverage for all $\underline{p}$'s.

Next consider $N = 2000$. Here we obtain similar results. For $\underline{p} = 0.1$ and 0.25, the choice of ε_N which yields coverage probabilities closest to the nominal coverage of 0.95 is four times the naive choice. This is also approximately true for $\underline{p} = 0.5$. For $\underline{p} = 0.75$, the next largest ε_N is the best (six times the naive choice). For $\underline{p} = 0.9$, an even larger ε_N is the best (eight times the naive, with the optimal scaling probably around seven). And for ε_N's large enough, we obtain essentially 100% coverage for all $\underline{p}$'s.

Before we interpret these results, we discuss one more table, table 2. While table 1 showed coverage probabilities, table 2 gives us an idea of the power of our confidence bands. For each simulation, we compute the ratio of the area under the confidence band to the area under the estimated breakdown frontier. By definition our confidence bands are all below the estimated breakdown frontier and hence this ratio can never be larger than one. Although we do not perform formal analysis of power, this ratio gives us an idea of the main trade-off in obtaining our confidence bands: We want them to be as large as possible subject to the constraint that they have correct coverage. This is how we defined our band in section 3, for a fixed ε_N. Here we compare the properties of these bands across different ε_N's. First consider $N = 500$ and $\underline{p} = 0.1$. From table 1, twice the naive choice of ε_N yields the closest to nominal coverage. All other choices gave over-coverage. We see this in table 2 since twice the naive choice gives essentially the largest area—all but one other choice have smaller area. Similarly, for $\underline{p} = 0.9$, the best choice based on table 1 is four times the naive choice, which gives an area under the confidence band of 47% that of the area under the estimated breakdown frontier. Smaller ε_N's give larger areas, but under-cover. Larger ε_N's give smaller areas, but over-cover. For large enough ε_N, the confidence bands get close to

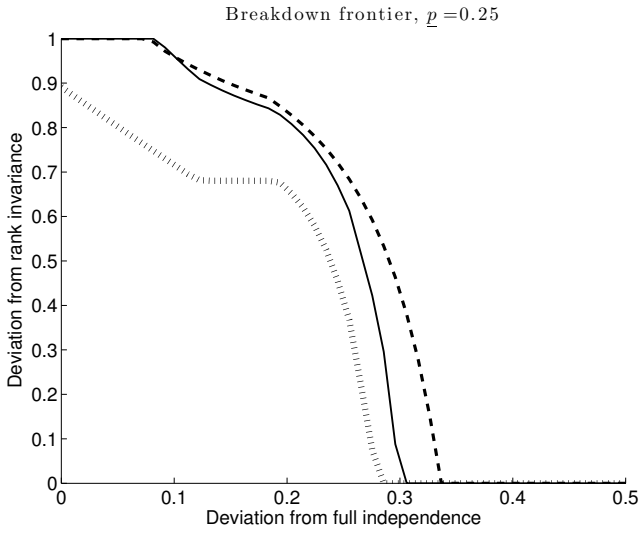


Figure 6: $N = 500$. Example 95% lower uniform confidence band (dotted line), estimated breakdown frontier (solid line), true breakdown frontier (dashed line).

zero everywhere, and hence have very small area and 100% coverage. The results for $N = 2000$ are similar.

In table 1 we saw that most combinations of $\underline{p}$ and ε_N led to over-coverage. This is caused by a downward bias in our estimated breakdown frontiers, as shown in figure 5. Since we are constructing lower confidence bands, this downward bias causes our confidence bands to over-cover. Although this finite-sample bias disappears asymptotically, one may wish to do a finite-sample bias correction to obtain higher-order refinements. Fan and Park (2009) previously studied this specific bias problem, in the case with random assignment (our $c = 0$) and no assumptions on rank invariance (our $t = 1$). They propose analytical and bootstrap bias corrected estimators of the bounds. Chernozhukov, Lee, and Rosen (2013) study a related problem. We leave such higher-order refinements to future work.

5 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 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 that paper, they use this data to examine

Table 1: Coverage Probabilities

N	ε_N	$\varepsilon_N/\varepsilon_N^{\text{naive}}$	$\underline{p}$ 0.10	0.25	0.50	0.75	0.90
500	0.0224	0.50	1.000	1.000	0.998	0.966	0.898
	0.0447	1.00	0.986	0.992	0.990	0.928	0.892
	0.0671	1.50	0.970	0.990	0.988	0.922	0.884
	0.0894	2.00	0.956	0.990	0.990	0.936	0.884
	0.1789	4.00	0.974	0.994	0.994	0.980	0.956
	0.2683	6.00	0.998	1.000	1.000	1.000	1.000
	0.3578	8.00	1.000	1.000	1.000	1.000	1.000
	0.4472	10.00	1.000	1.000	1.000	1.000	1.000
2000	0.0112	0.50	0.994	1.000	0.992	0.934	0.934
	0.0224	1.00	0.986	0.992	0.990	0.934	0.918
	0.0335	1.50	0.980	0.988	0.986	0.932	0.900
	0.0447	2.00	0.980	0.976	0.982	0.930	0.882
	0.0894	4.00	0.952	0.970	0.984	0.926	0.870
	0.1342	6.00	0.960	0.982	0.986	0.942	0.906
	0.1789	8.00	0.980	0.996	1.000	0.990	0.978
	0.2236	10.00	0.994	1.000	1.000	1.000	1.000

Nominal coverage is $1 - \alpha = 0.95$. As discussed in the body text, the choice $\varepsilon_N^{\text{naive}} = 1/\sqrt{N}$ yields the naive bootstrap. Cell values show uniform-over-c coverage probabilities of one-sided lower confidence bands, computed to maximize total area under the band.

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 note 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 pre-treatment 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

Table 2: Proportional area under the confidence bands

N	ε_N	$\varepsilon_N/\varepsilon_N^{\text{naive}}$	$\underline{p}$ 0.10	0.25	0.50	0.75	0.90
500	0.0224	0.50	0.644	0.643	0.637	0.672	0.734
	0.0447	1.00	0.759	0.716	0.705	0.740	0.774
	0.0671	1.50	0.780	0.734	0.722	0.751	0.776
	0.0894	2.00	0.779	0.730	0.722	0.746	0.763
	0.1789	4.00	0.604	0.552	0.541	0.529	0.468
	0.2683	6.00	0.252	0.174	0.117	0.069	0.024
	0.3578	8.00	0.022	0.007	0.001	0.000	0.000
	0.4472	10.00	0.000	0.000	0.000	0.000	0.000
2000	0.0112	0.50	0.869	0.832	0.808	0.834	0.884
	0.0224	1.00	0.894	0.865	0.841	0.862	0.896
	0.0335	1.50	0.901	0.876	0.853	0.873	0.901
	0.0447	2.00	0.904	0.882	0.859	0.879	0.902
	0.0894	4.00	0.906	0.879	0.862	0.877	0.890
	0.1342	6.00	0.875	0.840	0.829	0.837	0.833
	0.1789	8.00	0.814	0.755	0.732	0.717	0.665
	0.2236	10.00	0.704	0.615	0.563	0.499	0.387

Nominal coverage is $1 - \alpha = 0.95$. As discussed in the body text, the choice $\varepsilon_N^{\text{naive}} = 1/\sqrt{N}$ yields the naive bootstrap. Cell values show the average (across simulations) ratio of the area under the confidence band to the area under the estimated breakdown frontier.

> 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).

Analysis

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

Table 3: Summary statistics

Variable Name	Mean	Median	Stddev	Min	Max
Daily wage in Uganda shillings	2957.60	1400.00	6659.76	35.71	83333.34
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

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

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 3 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 1,400 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 $\overline{F}^{c}_{Y_x}(y)$ and $\underline{F}^{c}_{Y_x}(y)$ 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 wage. 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.

Table 4 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 4 are unconditional on the two covariates. Next we consider the conditional independence assumption, with age and household size in 1996 as our covariates. Under

Table 4: 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	

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.

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. To check the robustness of these point estimates to failure of conditional independence, we estimate the breakdown point c^* for the conclusion ATE ≥ 0, where we use log-wages as our outcome variable. We measure deviations from conditional independence by our conditional c-dependence distance. 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.

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

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

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 (2017). Specifically, we consider the difference between the "full" propensity score and the "leave out variable k" propensity score which omits variable k: Define

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

and

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

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 latter result 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 deviations from 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' deviations in 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 from independence in various settings, see Rotnitzky, Robins, and Scharfstein (1998), Robins (1999), Imbens (2003), Altonji, Elder, and Taber (2005, 2008), and Oster (2016).

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 rank invariance assumption to obtain point identification. Under these assumptions, our point estimate is 0.93 with a 95% CI of $[0.44, 0.99]$. Given the severity of being abducted, it is not surprising that our point estimate is so high.

Is this point estimate robust to failures of full conditional independence and rank invariance? We examine this question by estimating 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, 0.75, and 0.9 as in our Monte Carlo simulations. 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 point identifying assumptions. In our application this is 0.465, which is close to 0.5 so we omit it. Imbens (2003) suggests a similar choice of cutoff in his approach.

Figure 7 shows the results. As in our earlier plots, the horizontal axis plots c, deviations from full conditional independence, while the vertical axis plots t, deviations from 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's valid to examine various (c, t) combinations since we use uniform confidence bands.

First consider the top 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 five plots. If we impose full independence, then our estimated frontier suggests that we can almost fully relax 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 if rank invariance fails for up to 80% of the population. Moreover, looking at all choices of ε_N—not just our selected one—the lowest the vertical intercept ever gets is about 58%. Next suppose we relax full independence. Recall that the maximal deviation 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 smaller than the horizontal intercept of our selected confidence band. Hence, if we impose full rank invariance, our conclusion that $\mathbb{P}(Y_1 > Y_0) \geq 0.1$ is robust to deviations from full independence. Suppose instead that we think selection on

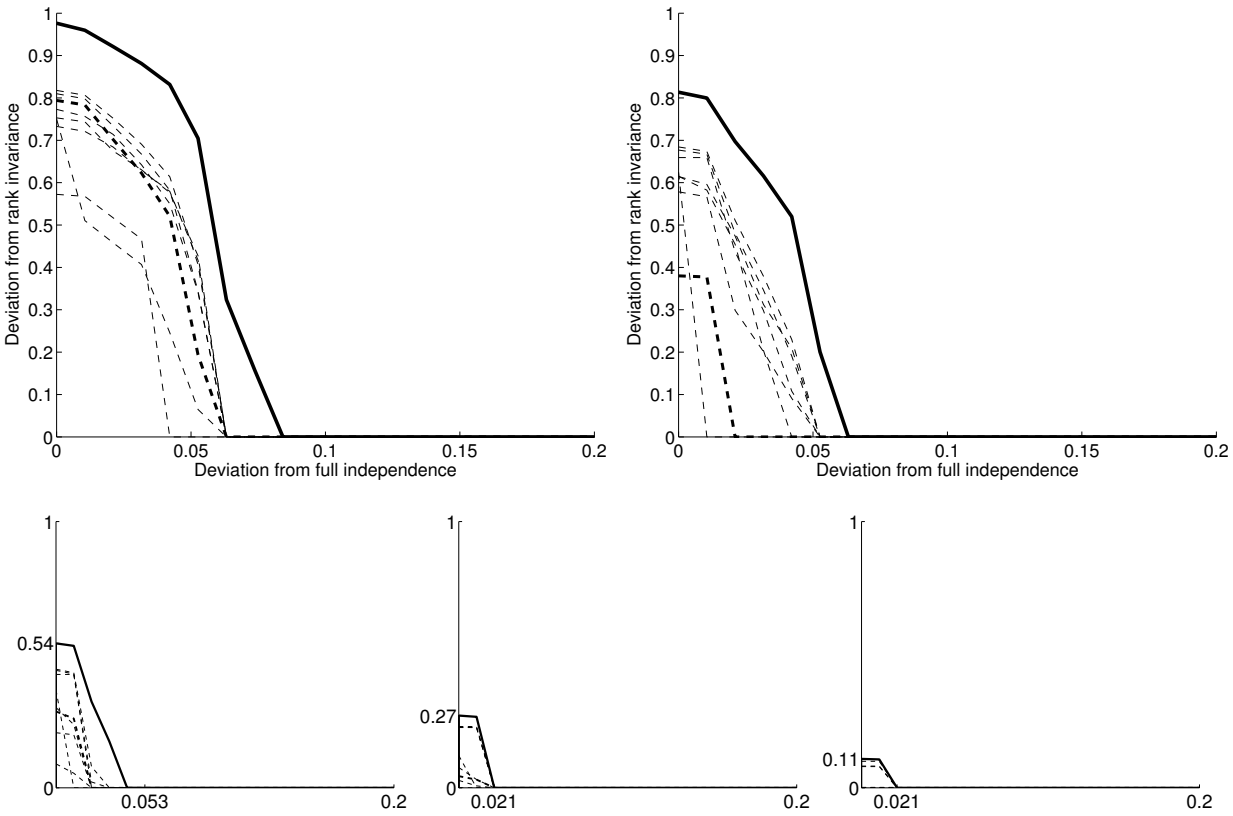


Figure 7: Estimated breakdown frontiers (solid lines) and confidence bands (dashed lines) for the claim $\mathbb{P}(Y_1 > Y_0) \geq \underline{p}$. Top, left to right: $\underline{p} = 0.1$, 0.25. Bottom, left to right: $\underline{p} = 0.5$, 0.75, 0.9. Light dashed lines are confidence bands for all eight values of ε_N considered. The darker dashed line is the band selected by our bootstrap procedure.

unobservables is at most the smaller $\bar{c}$ value, about 0.04. Then for c's in the range $[0, 0.04]$, we can still conclude $\mathbb{P}(Y_1 > Y_0) \geq 0.1$ so long as at least 20–40% 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. The rate of substitution between these two assumptions quickly changes for c's larger than 0.04, however. Specifically, if we increase it to 0.07 then we must impose rank invariance on the entire population if we still want to conclude that at least 10% of people benefit from not being forced into military service. All the breakdown frontiers we estimate in this application have this feature: we can relax full independence some while only requiring a small increase in the proportion of the population satisfying rank invariance. But after a point, further deviations from independence require much stronger assumptions on rank invariance. Understanding 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 deviations from full independence up to the size we see between the observed and leave out variable k propensity scores, depending on how much rank invariance failure we allow. For deviations from full independence up to the

smaller value of $\bar{c}$, we can allow up to 60% of the population to deviate from rank invariance.

Next consider the top right plot, $\underline{p} = 0.25$. Since this is a stronger conclusion than the previous one, all the frontiers are shifted towards the origin. Consequently, by construction, this conclusion is not as robust as the other one. If we focus on the estimated breakdown frontier, we obtain similar qualitative conclusions as for $\underline{p} = 0.1$. However, our selected confidence band for $\underline{p} = 0.25$ is substantially shifted inward relative to the $\underline{p} = 0.1$ confidence band. Thus even if we impose full independence we can only allow for about 38% failure of rank invariance. Relaxing full independence to a 2 percentage point deviation in the propensity score requires 100% rank invariance if we still want to maintain our conclusion. One caveat, however, is that our selected confidence band is substantially smaller than the bands corresponding to other choices of ε_N. Hence alternative choices of ε_N would suggest more robust results. This point underscores the need for future work on the choice of ε_N.

Next consider the bottom left 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 independence, and accounting for sampling uncertainty, then we can allow rank invariance to fail for about 30% of the population. This is quite large, but it relies on full independence holding exactly. If we also relax independence to $c = 0.02$ then we need 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. 0.02 is smaller than 0.09, the maximal deviation in observed propensity scores. Hence we might not be comfortable with such small values of c, which suggests the data do not definitively support the conclusion $\mathbb{P}(Y_1 > Y_0) \geq 0.5$. Similar results hold for $\underline{p} = 0.75$ and $\underline{p} = 0.9$. For these conclusions, even under full independence we cannot allow for too much rank invariance failure. And these conclusions are not robust to essentially any deviations from full independence. Coincidentally, the lower bound of our confidence interval for $\widehat{p}_{0,0} = 0.93$ is 0.44, which implies that even if we imposed the point identifying assumptions, the sampling uncertainty in the data prevent us from rejecting the hypothesis that $p_{0,0} < \underline{p}$ for $\underline{p} = 0.5$, 0.75, and 0.9.

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 independence and 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 propensity scores up to the same value as $\bar{c}_{\texttt{age}}$, which is also about two-thirds as large as $\bar{c}_{\texttt{hhSize}}$; 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 reasonable deviations in rank invariance and full independence for $\underline{p} = 10\%$, but not for larger percentages. In general, the results appear more sensitive to full independence than to rank invariance. This robustness to rank invariance matches the findings of Heckman et al. (1997), who imposed full independence and studied deviations form 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 deviation from rank invariance, and also because of differences in the empirical applications.

6 Conclusion

In this paper we advocated the *breakdown frontier* approach to sensitivity analysis. This approach defines the population breakdown frontier as the weakest set of assumptions such that a given 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 deviations from 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 deviations from 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 some conclusions about $\mathbb{P}(Y_1 > Y_0)$ are fairly robust to failure of rank invariance, when random assignment holds, but conclusions are much more sensitive to both assumptions for small deviations from random assignment.

As with the previous literature on breakdown points, breakdown frontier analysis can in principle be done in most models. It requires an indexed class of assumptions which deliver a nested sequence of identified sets, with non-refutable point identification obtained at one extreme and the no assumptions bounds obtained at the other. One then must characterize identified sets for the parameter of interest as a function of the sensitivity parameters. Given these identified sets, the breakdown frontier for a conclusion about that parameter can be defined, and inference procedures for this frontier can be developed.

A key conceptual step is deciding how to define the parametric classes of assumptions such that the magnitude of the deviation 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 deviation but not another. Thus one goal of future research is to explore this space of assumption deviations, 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 deviation from the random assignment assumption, including the one used here. In the present paper, we also used a general method for spanning two discrete assumptions by defining a $(1-t)$-percent deviation, as we did with rank invariance. But much work still remains to be done.

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.

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.

——— (2008): "Using selection on observed variables to assess bias from unobservables when evaluating swan-ganz catheterization," *American Economic Review P&P*, 98, 345–350.

Andrews, I., M. Gentzkow, and J. M. Shapiro (2017): "Measuring the sensitivity of parameter estimates to estimation moments," *The Quarterly Journal of Economics*, 132, 1553–1592.

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

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

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

Canay, I. A. and A. M. Shaikh (2017): "Practical and theoretical advances in inference for partially identified models," in *Advances in Economics and Econometrics: 11th World Congress of the Econometric Society (Forthcoming)*, Cambridge University Press.

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.

Chernozhukov, V., I. Fernández-Val, and A. Galichon (2010): "Quantile and probability curves without crossing," *Econometrica*, 78, 1093–1125.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013): "Intersection bounds: estimation and inference," *Econometrica*, 81, 667–737.

Conley, T. G., C. B. Hansen, and P. E. Rossi (2012): "Plausibly exogenous," *Review of Economics and Statistics*, 94, 260–272.

Copas, J. and S. Eguchi (2001): "Local sensitivity approximations for selectivity bias," *Journal of the Royal Statistical Society, Series B*, 63, 871–895.

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.

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

DiTraglia, F. and C. García-Jimeno (2016): "A framework for eliciting, incorporating, and disciplining identification beliefs in linear models," *Working paper*.

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

Dümbgen, L. (1993): "On nondifferentiable functions and the bootstrap," *Probability Theory and Related Fields*, 95, 125–140.

Escanciano, J. C. and L. Zhu (2013): "Set inferences and sensitivity analysis in semiparametric conditionally identified models," *Working paper*.

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.

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*, Emerald Group Publishing Limited, vol. 25, 3–70.

——— (2010): "Sharp bounds on the distribution of treatment effects and their statistical inference," *Econometric Theory*, 26, 931–951.

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

Fang, Z. and A. Santos (2015): "Inference on directionally differentiable functions," *arXiv preprint arXiv:1404.3763*.

Freyberger, J. and Y. Rai (2017): "Uniform confidence bands: Characterization and optimality," *Working paper*.

Giacomini, R., T. Kitagawa, and A. Volpicella (2016): "Uncertain identification," *Working paper*.

Hampel, F. R. (1968): "Contributions to the theory of robust estimation," *PhD Dissertation, University of California, Berkeley*.

——— (1971): "A general qualitative definition of robustness," *The Annals of Mathematical Statistics*, 1887–1896.

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.

Hansen, B. E. (2004): "Bandwidth selection for nonparametric distribution estimation," *Working paper*.

——— (2017): "Regression kink with an unknown threshold," *Journal of Business & Economic Statistics*, 35, 228–240.

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.

Ho, K. and A. M. Rosen (2016): "Partial identification in applied research: benefits and challenges," *Working paper*.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999): "Bayesian model averaging: A tutorial," *Statistical Science*, 382–401.

Hong, H. and J. Li (2015): "The numerical delta method and bootstrap," *Working paper*.

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

——— (2017): "Nonparametric estimation and inference under shape restrictions," *Journal of Econometrics*, 201, 108–126.

Horowitz, J. L. and C. F. Manski (1995): "Identification and robustness with contaminated and corrupted data," *Econometrica*, 281–302.

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

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

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

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

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

Kline, B. and E. Tamer (2016): "Bayesian inference in a class of partially identified models," *Quantitative Economics*, 7, 329–366.

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.

Koop, G. and D. J. Poirier (1997): "Learning about the across-regime correlation in switching regression models," *Journal of Econometrics*, 78, 217–227.

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

Leamer, E. E. (1978): *Specification Searches: Ad Hoc Inference with Nonexperimental Data*, vol. 53, John Wiley & Sons Incorporated.

Lee, Y.-Y. and D. Bhattacharya (2016): "Applied welfare analysis for discrete choice with interval-data on income," *Working paper*.

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

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.

Lindley, D. V. (1972): *Bayesian Statistics: A Review*, SIAM.

Luenberger, D. G. (1969): *Optimization by Vector Space Methods*, John Wiley & Sons.

Makarov, G. (1981): "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.

Manski, C. F. (2007): *Identification for Prediction and Decision*, Harvard University Press.

——— (2013): "Response to the review of 'Public policy in an uncertain world'," *The Economic Journal*, 123, F412–F415.

Marron, J. S. (1992): "Bootstrap bandwidth selection," in *Exploring the Limits of Bootstrap*, John Wiley, 249–262.

Masten, M. A. and A. Poirier (2016): "Partial independence in nonseparable models," *cemmap Working Paper, CWP26/16*.

——— (2017): "Identification of treatment effects under conditional partial independence," *Econometrica (forthcoming)*.

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

Moon, H. R. and F. Schorfheide (2012): "Bayesian and frequentist inference in partially identified models," *Econometrica*, 80, 755–782.

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

Oster, E. (2016): "Unobservable selection and coefficient stability: Theory and evidence," *Journal of Business & Economic Statistics (forthcoming)*.

Pitman, E. J. G. (1949): "Notes on non-parametric statistical inference," *Unpublished lecture notes, Columbia University*.

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.

Richardson, A., M. G. Hudgens, P. B. Gilbert, and J. P. Fine (2014): "Nonparametric bounds and sensitivity analysis of treatment effects," *Statistical Science*, 29, 596.

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

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*, Springer, 1–94.

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

——— (2002): *Observational Studies*, Springer, second ed.

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.

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.

Rotnitzky, A., D. O. Scharfstein, T.-L. Su, and J. M. Robins (2001): "Methods for conducting sensitivity analysis of trials with potentially nonignorable competing causes of censoring," *Biometrics*, 57, 103–113.

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

Sklar, M. (1959): *Fonctions de répartition à n dimensions et leurs marges*, Université Paris 8.

Stein, E. M. and R. Shakarchi (2011): *Functional Analysis: An Introduction to Further Topics in Analysis*, vol. 4, Princeton University Press.

Stoye, J. (2005): "Essays on partial identification and statistical decisions," Ph.D. thesis, Northwestern University.

——— (2010): "Partial identification of spread parameters," *Quantitative Economics*, 1, 323–357.

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

Todem, D., J. Fine, and L. Peng (2010): "A global sensitivity test for evaluating statistical hypotheses with nonidentifiable models," *Biometrics*, 66, 558–566.

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

van der Vaart, A. W. (2000): *Asymptotic Statistics*, Cambridge University Press.

Vansteelandt, S., E. Goetghebeur, M. G. Kenward, and G. Molenberghs (2006): "Ignorance and uncertainty regions as inferential tools in a sensitivity analysis," *Statistica Sinica*, 953–979.

A Inference in sensitivity analyses

In this section we provide additional details explaining how our results compare to several approaches in the literature. We focus on the different inference methods used in sensitivity analyses. Most methods can be grouped by whether the population level sensitivity analysis is a parametric path or nonparametric neighborhood approach. In Masten and Poirier (2016) we compare and contrast these population level approaches in more detail. The parametric path approach has two key features: (1) a specific parametric deviation r from a baseline assumption of $r = 0$ and (2) a parameter $\theta(r)$ that is point identified given that deviation. The nonparametric neighborhood approach specifies increasing nested neighborhoods around a baseline assumption of $r = 0$ such that $\Theta(r)$ is the identified set for the parameter given a specific neighborhood r. Typically $\Theta(r) = [\Theta_L(r), \Theta_U(r)]$ for point identified lower and upper bound functions Θ_L and Θ_U.

Parametric paths

The most common approach for a parametric path analysis is to report the estimated function $\widehat{\theta}(r)$ along with pointwise confidence bands. For example, see figure 1 of Rotnitzky et al. (1998), figure 1 of Robins (1999), and figure 1 of Vansteelandt, Goetghebeur, Kenward, and Molenberghs (2006). Uniform confidence bands can be used instead, as in figure 3 of Todem, Fine, and Peng (2010). Those authors use their uniform confidence bands to test hypotheses about $\theta(r)$ uniformly over r. They also suggest projecting these bands onto their domain to obtain confidence sets for the set $\{r : |\theta(r)| > 0\}$, although they do not discuss this in detail (see the last few sentences of page 562). They emphasize that using uniform confidence bands is important since the functions $\theta(r)$ are often not monotonic, as we discussed earlier with respect to Imbens (2003). A similar method is proposed by Rotnitzky et al. (1998). They study a model with two scalar sensitivity parameters $r = (r_1, r_2)$ and two parameters $\theta_1(r)$ and $\theta_2(r)$. They construct a standard test statistic $T(r)$ for testing the null that $\theta_1(r) = \theta_2(r)$. They then plot the contours

$$\{r \in \mathbb{R}^2 : T(r) = -1.96\} \qquad \text{and} \qquad \{r \in \mathbb{R}^2 : T(r) = 1.96\}.$$

See their figure 2. Unlike Todem et al. (2010), they do not account for multiple testing concerns. Also see figures 2–4 of Rotnitzky, Scharfstein, Su, and Robins (2001). Several papers also suggest picking a set $\mathcal{R}$ to form an identified set $\{\theta(r) : r \in \mathcal{R}\}$ and then doing inference on this identified set. For example, see Vansteelandt et al. (2006). Escanciano and Zhu (2013) consider the diameter of such identified sets,

$$d = \sup_{r,r' \in \mathcal{R}} \|\theta(r) - \theta(r')\|,$$

and study estimation and inference on d.

Finally, Rosenbaum (1995, 2002) proposes a sensitivity analysis within the context of finite sample randomization inference for testing the sharp null hypotheses of no unit level treatment effects for the units in our dataset. This is a very different approach to the approaches discussed above and what we do in the present paper.

Nonparametric neighborhoods

Our population level sensitivity analysis uses nonparametric neighborhoods, not parametric paths. Thus for each r we obtain an identified set $\Theta(r)$. There is a large literature on how to do inference on a single identified set; see Canay and Shaikh (2017) for an overview. Few papers discuss inference on a continuous sequence of identified sets $\Theta(r)$, however. The simplest approach arises when the identified set is characterized by point identified upper and lower bounds: $\Theta(r) = [\Theta_L(r), \Theta_U(r)]$. In this case one can plot estimated bound functions $\widehat{\Theta}_L$ and $\widehat{\Theta}_U$ along with outer confidence bands for these functions. For example, see figure 2 of Richardson, Hudgens, Gilbert, and Fine (2014). They informally discuss how to use these bands to check robustness of the claim that the true parameter is nonzero, but they do not formally discuss breakdown points or inference on them.

Kline and Santos (2013) similarly begin by constructing pointwise confidence bands for the bound functions. They then use level sets of these bands to construct their confidence intervals for a breakdown point (see equation 41 on page 249). In their remark 4.4 on page 250 they mention the approach we take—doing inference based directly on the asymptotic distribution of breakdown point estimators. In order to compare these two approaches, we discuss the approach of projecting confidence bands for lower bound functions in more detail here.

Let the sensitivity parameter r be in $[0,1]^{d_r}$ for some integer $d_r \geq 1$. Let $\Theta_L(r)$ denote the lower bound function for a scalar parameter θ. By construction, $\Theta_L(\cdot)$ is weakly decreasing in

its components. Suppose it is also continuous. Suppose we are interested in the conclusion that $\theta_{\text{true}} \geq \underline{\theta}$. Suppose for simplicity that it is known that $\Theta_L(0) \geq \underline{\theta}$. This allows us to ignore the upper bound function and its confidence band. Define the breakdown frontier for the claim that $\theta_{\text{true}} \geq \underline{\theta}$ by

$$\text{BF} = \{r \in [0,1]^{d_r} : \Theta_L(r) = \underline{\theta}\}.$$

Let

$$\text{RR} = \{r \in [0,1]^{d_r} : \Theta_L(r) \geq \underline{\theta}\}.$$

denote the robust region, the set of sensitivity parameters that lie on or below the breakdown frontier. The following proposition shows that, in general, projections of uniform lower confidence bands for Θ_L produce valid uniform lower confidence bands for the breakdown frontier.

Proposition 3. Let $\text{LB}(\cdot)$ be an asymptotically exact uniform lower $(1-\alpha)$-confidence band for $\Theta_L(\cdot)$. That is,

$$\lim_{N\to\infty} \mathbb{P}\left(\text{LB}(r) \leq \Theta_L(r) \text{ for all } r \in [0,1]^{d_r}\right) = 1 - \alpha.$$

(We call $\text{LB}(\cdot)$ a 'band' even though it's really a hypersurface.) Define the projections

$$\text{BF}_L = \{r \in [0,1]^{d_r} : \text{LB}(r) = \underline{\theta}\}$$

and

$$\text{RR}_L = \{r \in [0,1]^{d_r} : \text{LB}(r) \geq \underline{\theta}\}.$$

Then

$$\lim_{N\to\infty} \mathbb{P}(\text{RR}_L \subseteq \text{RR}) \geq 1 - \alpha.$$

Proof of proposition 3. We have

$$\begin{aligned}\mathbb{P}(\text{RR}_L \subseteq \text{RR}) &= \mathbb{P}(\text{For all } r \in [0,1]^{d_r} \text{ s.t. } \text{LB}(r) \geq \underline{\theta}, \text{ we have } \Theta_L(r) \geq \underline{\theta}) \\ &\geq \mathbb{P}(\text{LB}(r) \leq \Theta_L(r) \text{ for all } r \in [0,1]^{d_r}).\end{aligned}$$

Now take limits on both sides as $N \to \infty$. The inequality arises essentially because the functional inequality $\text{LB}(\cdot) \leq \Theta_L(\cdot)$ is a sufficient, but not necessary, condition for $\text{RR}_L \subseteq \text{RR}$. □

Proposition 3 shows that projecting a uniform band always yields a confidence band for the breakdown frontier which has size at least $1-\alpha$. Notice that although we did not use monotonicity of $\Theta_L(\cdot)$ here, this monotonicity implies that we can always take $\text{LB}(\cdot)$ to be weakly decreasing without loss of generality. This follows since monotonicity of $\Theta_L(\cdot)$ allows us to convert any non-monotonic lower confidence band into a monotonic one without any loss of coverage.

There are two downsides to this projection approach, compared to our direct approach:

1. In general, this projection approach may be conservative.
2. Relatedly, one must choose the lower confidence band $\Theta_L(\cdot)$. There are many such choices. The standard ones, such as equal width or inversion of a sup t-statistic (e.g., see Freyberger and Rai 2017), will likely yield conservative projection bands, since they are not chosen with the goal of doing inference on the breakdown frontier in mind.

Kline and Santos (2013) do not propose projections of *uniform* confidence bands. They propose projections of pointwise confidence bands. As we discuss next, projection of pointwise bands produces valid confidence intervals for breakdown points. But it does not generally produce valid

confidence bands for breakdown frontiers. Hence in the multidimensional r case one either must use our direct approach, or appeal to proposition 3 above.

To see that pointwise band projections are valid in the scalar r case, we expand on Kline and Santos' (2013) analysis. Define the population breakdown point by

$$r^* = \inf\{r \in [0,1] : \Theta_L(r) \leq \underline{\theta}\}.$$

Let $c_{1-\alpha}(r)$ be the $1-\alpha$ quantile of the asymptotic distribution of

$$\sqrt{N}\big[\widehat{\Theta}_L(r) - \Theta_L(r)\big].$$

Define the pointwise one-sided lower confidence band for $\Theta_L(\cdot)$ by

$$\mathrm{LB}(r) = \widehat{\Theta}_L(r) - \frac{c_{1-\alpha}(r)}{\sqrt{N}}.$$

Let

$$r_L = \inf\{r \in [0,1] : \mathrm{LB}(r) \leq \underline{\theta}\}$$

be the projection of this confidence band. The following result is a minor generalization of example 2.1 in Kline and Santos (2013).

Proposition 4. Assume that the cdf of the asymptotic distribution of $\sqrt{N}\big[\widehat{\Theta}_L(r^*) - \Theta_L(r^*)\big]$ is continuous and strictly increasing at its $1-\alpha$ quantile. Then

$$\lim_{N\to\infty} \mathbb{P}(r_L \leq r^*) \geq 1-\alpha.$$

If $\mathrm{LB}(\cdot)$ is weakly decreasing with probability one then this inequality holds with equality.

Proof of proposition 4. We have

$$\begin{aligned}
\mathbb{P}(r_L \leq r^*) &= \mathbb{P}(r^* \geq \inf\{r \in [0,1] : \mathrm{LB}(r) \leq \underline{\theta}\}) \\
&\geq \mathbb{P}(\mathrm{LB}(r^*) \leq \underline{\theta}) \\
&= \mathbb{P}\left(\widehat{\Theta}_L(r^*) - \frac{c_{1-\alpha}(r^*)}{\sqrt{N}} \leq \underline{\theta}\right) \\
&= \mathbb{P}\left(\sqrt{N}\big(\widehat{\Theta}_L(r^*) - \underline{\theta}\big) \leq c_{1-\alpha}(r^*)\right) \\
&= \mathbb{P}\left(\sqrt{N}\big(\widehat{\Theta}_L(r^*) - \Theta_L(r^*)\big) \leq c_{1-\alpha}(r^*)\right).
\end{aligned}$$

The first line follows by definition of r_L. For the second line, notice that $\mathrm{LB}(r^*) \leq \underline{\theta}$ implies that $r^* \in \{r \in [0,1] : \mathrm{LB}(r) \leq \underline{\theta}\}$ and hence $r^* \geq \inf\{r \in [0,1] : \mathrm{LB}(r) \leq \underline{\theta}\}$ by the definition of the infimum. This gives us

$$\mathbb{P}(\mathrm{LB}(r^*) \leq \underline{\theta}) \leq \mathbb{P}(r^* \geq \inf\{r \in [0,1] : \mathrm{LB}(r) \leq \underline{\theta}\}).$$

If $\mathrm{LB}(\cdot)$ is weakly decreasing with probability one, then the reverse inequality holds, and hence we have an equality in the second line. To see this, suppose $r^* \geq r_L$ holds. Then $\mathrm{LB}(r^*) \leq \mathrm{LB}(r_L)$ since $\mathrm{LB}(\cdot)$ is weakly decreasing. But now notice that $\mathrm{LB}(r_L) \leq \underline{\theta}$ by definition of r_L. Hence $\mathrm{LB}(r^*) \leq \underline{\theta}$.

The third line follows by definition of LB. The fifth line by definition of the population breakdown point, as the solution to $\Theta_L(r) = \underline{\theta}$. The result now follows by taking limits as $N \to \infty$

on both sides, and by definition of $c_{1-\alpha}(r^*)$ and the invertibility of the limiting cdf at its $1-\alpha$ quantile. □

Proposition 4 shows that, for doing inference on scalar breakdown points, projections of monotonic lower pointwise confidence bands for the lower bound function yields a one-sided confidence interval $[r_L, 1]$ for the breakdown point r^* which has asymptotically exact size. If the lower band function is not always monotonic, however, this projection can be conservative. Moreover, since we're constructing one-sided pointwise confidence bands, we do not have any flexibility to choose the shape of this confidence band. Hence whether it is monotonic or not will be determined by the distribution of the data. Furthermore, it does not appear that this proof strategy extends to multidimensional r. Hence projections of pointwise bands are unlikely to yield uniform confidence bands for the breakdown frontier.

Overall, our analysis above shows that the projection of confidence bands approach to doing inference on breakdown points and frontiers will likely yield conservative inference. This is not surprising since, unlike our approach, these bands are not designed specifically for doing inference on the breakdown frontier. Finally, we note that if one nonetheless wants to use a projection approach, our asymptotic results in section 3 can be used to do so.

A testing interpretation of lower confidence bands for breakdown frontiers

Consider the scalar r case, as above. Suppose we want to test

$$H_0 : \Theta_L(r) \leq \underline{\theta} \qquad \text{versus} \qquad H_1 : \Theta_L(r) > \underline{\theta}$$

for a fixed $r \in [0,1]$. By definition of the breakdown point, H_0 is true if and only if $r \geq r^*$. Let $[r_L, 1]$ denote a one-sided lower confidence interval for the breakdown point r^*; that is, $\mathbb{P}([r_L, 1] \ni r^*) = 1 - \alpha$. Define the test

$$\phi = \begin{cases} \text{Choose } H_0 & \text{if } r_L < r \\ \text{Choose } H_1 & \text{if } r \leq r_L. \end{cases}$$

Then

$$\begin{aligned} \mathbb{P}(\text{Choose } H_1 \mid H_0 \text{ true}) &= \mathbb{P}(r \leq r_L) \\ &\leq \mathbb{P}(r^* \leq r_L) \\ &= \alpha. \end{aligned}$$

The second line follows since $r \geq r^*$. The last line follows by construction of r_L. Hence ϕ has size at most α. This result holds for any $r \in [0,1]$. Thus we can interpret the robust region inner confidence set $[0, r_L]$ as the set of sensitivity parameters r such that we reject the null that the true parameter might be below $\underline{\theta}$. That is, for $r \in [0, r_L]$, our test concludes that $\theta > \underline{\theta}$. For r outside the robust region inner confidence set, we do not reject the null that θ might be at or below $\underline{\theta}$.

Here we considered the scalar r case for simplicity, but this argument extends to the general case of interpreting lower confidence bands for an arbitrary dimensional breakdown frontier.

Local analyses

Since Pitman (1949), local asymptotics are sometimes used to study the behavior of a given estimator under small deviations from model assumptions. Several papers use this approach to study deviations from exogeneity-type assumptions. In a missing data model, Copas and Eguchi (2001)

consider local-to-full-independence asymptotic distributions of MLEs. Conley, Hansen, and Rossi (2012) derive the asymptotic bias of the 2SLS estimator in an IV model along sequences where violations of the exclusion restriction converge to zero. The asymptotic bias depends on a local parameter. By placing a prior on this local parameter, they do Bayesian inference on the coefficient of interest. Andrews, Gentzkow, and Shapiro (2017) generalize this local-to-zero asymptotic result to the GMM estimator for a given system of moment equalities. Unlike this literature, breakdown frontier analysis focuses on the largest deviations from an assumption under which one's conclusions still hold.

Bayesian inference and breakdown frontiers

Although we focus on frequentist inference, here we briefly discuss Bayesian approaches. In section 11 of Robins et al. (2000), Robins studied Bayesian inference in a parametric path approach to sensitivity analysis. Let r denote the sensitivity parameter and $\theta(r)$ the parameter of interest, which is point identified given r. Holding r fixed, one can do standard Bayesian inference on $\theta(r)$. Thus Robins simply suggests placing a prior on r and averaging posteriors conditional on r over this prior. Indeed, this approach is essentially just Bayesian model averaging, where r indexes the class of models under consideration. See Hoeting, Madigan, Raftery, and Volinsky (1999) for a survey of Bayesian model averaging, and Leamer (1978) for important early work. Among other approaches, Conley et al. (2012) apply these ideas to do a sensitivity analysis in an IV model. See DiTraglia and García-Jimeno (2016) for a generalization and a detailed analysis of priors in that IV setting.

Next consider the nonparametric neighborhood approach. Here the parameter of interest is only partially identified for a fixed r, and thus even holding r fixed leads to non-standard Bayesian analysis. Giacomini, Kitagawa, and Volpicella (2016) study Bayesian model averaging where one of the models is partially identified. They study averaging of a finite number of models. If their results can be extended to a continuum of models, then this method could be applied to the model and assumptions we consider in this paper.

A subtlety arises in both Robins et al. (2000) and Giacomini et al. (2016): Depending on how one specifies the joint prior for the sensitivity parameters and the remaining parameters, it may be possible to obtain some updating of the prior for the sensitivity parameters (a point mentioned more generally by Lindley 1972 in his footnote 34 on page 46; also see Koop and Poirier 1997). As Giacomini et al. (2016) discuss, however, the model posterior will not converge to the truth unless the model is refutable. None of the assumptions (c, t) in the model we study are refutable. Hence the prior over (c, t) generally matters even asymptotically. That said, the breakdown frontier determines exactly how much the model priors matter for a specific claim. For instance, suppose the model prior places all of its mass below the breakdown frontier for a specific claim. Then we conjecture that the Bayesian model averaged posterior probability that the claim is true will converge to one as $N \to \infty$, regardless of the specific choice of prior. Kline and Tamer (2016) provide results like this in the single model case. More generally, we conjecture that the proportion of model prior mass that falls below the breakdown frontier partially determines the tightness of the corresponding asymptotic posterior probability of the conclusion being true: The more mass outside the breakdown frontier, the more the model priors matter. Consequently, a sample analog estimate and perhaps even frequentist inference on the breakdown frontier can be useful even in a Bayesian analysis, to help determine the importance of one's model priors. This is similar to Moon and Schorfheide's (2012) recommendation that one report estimated identified sets along with Bayesian posteriors. Here we have just sketched the relationship between Bayesian analysis and breakdown frontiers. We leave a complete analysis of these issues to future work.

B Proofs

Proofs for section 2

Proof of theorem 1. Let F_1 and F_0 be any strictly increasing cdfs. Suppose (Y_1, Y_0) have joint cdf $F_{Y_1,Y_0}(y_1, y_0) = C(F_1(y_1), F_0(y_0))$. Then

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

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

$$\int_{\{y_1 - y_0 \leq z\}} dM(F_1(y_1), F_0(y_0)) = \mathbb{P}(F_{Y_1}^{-1}(U) - F_{Y_0}^{-1}(U) \leq z).$$

Makarov (1981) derived sharp bounds on $\int_{\{y_1 - y_0 \leq z\}} dH(F_1(y_1), F_0(y_0))$ as follows:

$$\begin{aligned}&\int_{\{y_1 - y_0 \leq z\}} dH(F_1(y_1), F_0(y_0)) \\ &\qquad \in \left[\max\left\{\sup_{y \in \mathcal{Y}_z}(F_1(y) - F_0(y-z)), 0\right\}, 1 + \min\left\{\inf_{y \in \mathcal{Y}_z}(F_1(y) - F_0(y-z)), 0\right\}\right].\end{aligned}$$

Therefore, for given (F_1, F_0), sharp bounds for $\mathbb{P}(Y_1 - Y_0 \leq z)$ are given by $[\underline{\theta}(F_1, F_0), \overline{\theta}(F_1, F_0)]$, where

$$\begin{aligned}\underline{\theta}(F_1, F_0) &= (1-t)\mathbb{P}(F_1^{-1}(U) - F_0^{-1}(U) \leq z) + t\max\left\{\sup_{y \in \mathcal{Y}_z}(F_1(y) - F_0(y-z)), 0\right\} \\ \overline{\theta}(F_1, F_0) &= (1-t)\mathbb{P}(F_1^{-1}(U) - F_0^{-1}(U) \leq z) + t\left(1 + \min\left\{\inf_{y \in \mathcal{Y}_z}(F_1(y) - F_0(y-z)), 0\right\}\right).\end{aligned}$$

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,

$$\tilde{F}_0 \preceq_{\text{FSD}} F_0 \qquad \text{implies} \qquad \underline{\theta}(F_1, \tilde{F}_0) \leq \underline{\theta}(F_1, F_0).$$

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

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

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

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

Second, the expression

$$\max\left\{\sup_{y\in\mathcal{Y}_z}(F_1(y)-F_0(y-z)),0\right\}$$

is weakly increasing in F_0 since the supremum and maximum operators are weakly increasing. Thus both components of $\underline{\theta}$ are weakly increasing in F_0. Therefore the linear combination of them is also weakly increasing in F_0.

We can similarly show that $\underline{\theta}(F_1,F_0)$ is weakly decreasing in F_1. Thus substituting $(F_1,F_0)=(\underline{F}^c_{Y_1},\overline{F}^c_{Y_0})$ yields the lower bound $\underline{\text{DTE}}(z,c,t)$. The upper bound function $\overline{\theta}(F_1,F_0)$ is also weakly increasing in F_0 and weakly decreasing in F_0. Thus substituting $(F_1,F_0)=(\overline{F}^c_{Y_1},\underline{F}^c_{Y_0})$ yields the upper bound $\overline{\text{DTE}}(z,c,t)$. In making these substitutions we applied proposition 2 from Masten and Poirier (2017). In that paper we defined functions $F^c_{Y_x}(\cdot\mid\epsilon,\eta)$, which we now use to sharpness of the DTE bounds.

Substitute $(F^c_{Y_1}(\cdot\mid\epsilon,0),F^c_{Y_0}(\cdot\mid1-\epsilon,0))$ into the bound functionals and continuously vary ϵ between $[0,1]$. Note that we let $\eta=0$ since $c<\min\{p_1,p_0\}$. By continuity of $\underline{\theta}(\cdot,\cdot)$ and $\overline{\theta}(\cdot,\cdot)$ in their arguments and continuity of $(F^c_{Y_1}(\cdot\mid\epsilon,0),F^c_{Y_0}(\cdot\mid1-\epsilon,0))$ in ϵ, the intermediate value theorem implies that every element between the bounds can be attained. □

Proofs for section 3

Proof of lemma 1. By a second order Taylor expansion,

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

By standard bracketing entropy results (e.g., example 19.6 on page 271 of van der Vaart 2000) the function classes $\{\mathbb{1}(Y\le y)\mathbb{1}(X=x):y\in\mathbb{R},x\in\{0,1\}\}$ and $\{\mathbb{1}(X=x):x\in\{0,1\}\}$ are both P-Donsker. Hence the residual is of order $O_p(N^{-1})$ uniformly in $(y,x)\in\mathbb{R}\times\{0,1\}$. Combining this with Slutsky's theorem we get the uniform over y and x asymptotic representation

$$\widehat{F}_{Y|X}(y\mid x)-F_{Y|X}(y\mid x)=\frac{1}{N}\sum_{i=1}^N\frac{\mathbb{1}(X_i=x)(\mathbb{1}(Y_i\le y)-F_{Y|X}(y\mid x))}{\mathbb{P}(X=x)}+o_p(N^{-1/2}).$$

By the same bracketing entropy arguments, the class

$$\left\{ \frac{\mathbb{1}(X = x)(\mathbb{1}(Y \leq y) - F_{Y|X}(y \mid x))}{\mathbb{P}(X = x)} : y \in \mathbb{R}, x \in \{0, 1\} \right\}$$

is P-Donsker and hence $\sqrt{N}(\widehat{F}_{Y|X}(\cdot \mid \cdot) - F_{Y|X}(\cdot \mid \cdot))$ converges in distribution to a mean-zero Gaussian process with continuous paths. The covariance kernel $\mathbf{\Sigma}_1$ can be calculated as follows:

$$\begin{aligned}
&\mathbf{\Sigma}_1(y, x, \tilde{y}, \tilde{x}) \\
&= \begin{pmatrix} \mathbb{E}\left[\frac{\mathbb{1}(X_i=x)\mathbb{1}(X_i=\tilde{x})(\mathbb{1}(Y_i\leq y)-F_{Y|X}(y|x))(\mathbb{1}(Y_i\leq \tilde{y})-F_{Y|X}(\tilde{y}|\tilde{x}))}{\mathbb{P}(X=x)\mathbb{P}(X=\tilde{x})}\right] & \mathbb{E}\left[\frac{\mathbb{1}(X_i=x)(\mathbb{1}(X_i=\tilde{x})-p_{\tilde{x}})(\mathbb{1}(Y_i\leq y)-F_{Y|X}(y|x))}{\mathbb{P}(X=x)}\right] \\ \mathbb{E}\left[\frac{\mathbb{1}(X_i=\tilde{x})(\mathbb{1}(X_i=x)-p_x)(\mathbb{1}(Y_i\leq \tilde{y})-F_{Y|X}(\tilde{y}|\tilde{x}))}{\mathbb{P}(X=\tilde{x})}\right] & \mathbb{E}[(\mathbb{1}(X_i = x) - p_x)(\mathbb{1}(X_i = \tilde{x}) - p_{\tilde{x}})] \end{pmatrix} \\
&= \begin{pmatrix} \frac{F_{Y|X}(\min\{y,\tilde{y}\}|x)-F_{Y|X}(y|x)F_{Y|X}(\tilde{y}|x)}{p_x}\mathbb{1}(x = \tilde{x}) & 0 \\ 0 & p_x\mathbb{1}(x = \tilde{x}) - p_x p_{\tilde{x}} \end{pmatrix}.
\end{aligned}$$

□

Lemma 6 (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 6. 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 $\mathbb{D}_0$

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

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

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

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. □

Proof of lemma 2. Let $\theta_0 = (F_{Y|X}(\cdot \mid \cdot), p_{(\cdot)})$ and $\widehat{\theta} = (\widehat{F}_{Y|X}(\cdot \mid \cdot), \widehat{p}_{(\cdot)})$. For fixed y and c, define the mapping

$$\phi_1 : \quad \ell^\infty(\mathbb{R} \times \{0,1\}) \times \ell^\infty(\{0,1\}) \to \ell^\infty(\{0,1\}, \mathbb{R}^2)$$

by

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

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

$$\begin{pmatrix} \overline{F}^c_{Y_x}(y) \\ \underline{F}^c_{Y_x}(y) \end{pmatrix} = [\phi_1(\theta_0)](x) \qquad \text{and} \qquad \begin{pmatrix} \widehat{\overline{F}}^c_{Y_x}(y) \\ \widehat{\underline{F}}^c_{Y_x}(y) \end{pmatrix} = [\phi_1(\widehat{\theta})](x).$$

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

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

and

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

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:

$$
\begin{aligned}
[\delta_1(\theta)](x) &= \frac{\theta^{(1)}(y,x)\theta^{(2)}(x)}{\theta^{(2)}(x)+c} \text{ has Hadamard derivative equal to} \\
[\delta'_{1,\theta}(h)](x) &= \frac{\theta^{(1)}(y,x)h^{(2)}(x)+h^{(1)}(y,x)\theta^{(2)}(x)}{\theta^{(2)}(x)+c} - \frac{\theta^{(1)}(y,x)\theta^{(2)}(x)h^{(2)}(x)}{(\theta^{(2)}(x)+c)^2}, \\
[\delta_2(\theta)](x) &= \frac{\theta^{(1)}(y,x)\theta^{(2)}(x)-c}{\theta^{(2)}(x)-c} \text{ has Hadamard derivative equal to} \\
[\delta'_{2,\theta}(h)](x) &= \frac{\theta^{(1)}(y,x)h^{(2)}(x)+h^{(1)}(y,x)\theta^{(2)}(x)}{\theta^{(2)}(x)-c} - \frac{(\theta^{(1)}(y,x)\theta^{(2)}(x)-c)h^{(2)}(x)}{(\theta^{(2)}(x)-c)^2}, \\
[\delta_3(\theta)](x) &= \frac{\theta^{(1)}(y,x)\theta^{(2)}(x)}{\theta^{(2)}(x)-c} \text{ has Hadamard derivative equal to} \\
[\delta'_{3,\theta}(h)](x) &= \frac{\theta^{(1)}(y,x)h^{(2)}(x)+h^{(1)}(y,x)\theta^{(2)}(x)}{\theta^{(2)}(x)-c} - \frac{\theta^{(1)}(y,x)\theta^{(2)}(x)h^{(2)}(x)}{(\theta^{(2)}(x)-c)^2}, \\
[\delta_4(\theta)](x) &= \frac{\theta^{(1)}(y,x)\theta^{(2)}(x)+c}{\theta^{(2)}(x)+c} \text{ has Hadamard derivative equal to} \\
[\delta'_{4,\theta}(h)](x) &= \frac{\theta^{(1)}(y,x)h^{(2)}(x)+h^{(1)}(y,x)\theta^{(2)}(x)}{\theta^{(2)}(x)+c} - \frac{(\theta^{(1)}(y,x)\theta^{(2)}(x)+c)h^{(2)}(x)}{(\theta^{(2)}(x)+c)^2}.
\end{aligned}
$$

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

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

By the chain rule (lemma 6), the map ϕ_1 is Hadamard directionally differentiable at θ_0 with Hadamard directional derivative evaluated at θ_0 equal to

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

By lemma 1, $\sqrt{N}(\widehat{\theta}(y,x) - \theta_0(y,x)) \rightsquigarrow \mathbf{Z}_1(y,x)$. Hence we can use the delta method for Hadamard directionally differentiable functions (see theorem 2.1 in Fang and Santos 2015) to find that

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

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 $\mathbf{Z}_2(y, x, c)$. □

Proof of lemma 3. Let $\mathcal{S} = \{(y, x) \in \mathbb{R}^2 : x \in \{0, 1\}, y \in [\underline{y}_x, \overline{y}_x]\}$. Let $\mathscr{D}(\mathcal{S}) \subset \ell^\infty(\mathcal{S})$ denote the set of functions that are càdlàg in the first argument for each $x \in \{0, 1\}$. Define the mapping

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

by

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

By A1, A3, A4, and lemma 21.4(ii) in van der Vaart (2000) this mapping is Hadamard differentiable at θ_0 tangentially to $\mathscr{C}(\mathcal{S}) \subset \ell^\infty(\mathcal{S})$, the set functions that are continuous in the first argument for each $x \in \{0, 1\}$. Its Hadamard derivative at $\theta_0 = (F_{Y|X}(\cdot \mid \cdot), p_{(\cdot)})$ is

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

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

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

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

Now define the mapping

$$\phi_2 : \quad \ell^\infty((0, 1) \times \{0, 1\}) \times \ell^\infty(\{0, 1\}) \to \ell^\infty((0, 1) \times \{0, 1\} \times [0, \overline{C}], \mathbb{R}^2)$$

by

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

Then

$$\begin{pmatrix} \overline{Q}^c_{Y_x}(\tau) \\ \underline{Q}^c_{Y_x}(\tau) \end{pmatrix} = [\phi_2(\tilde{\phi}_2(\theta_0))](\tau, x, c) \qquad \text{and} \qquad \begin{pmatrix} \widehat{\overline{Q}}^c_{Y_x}(\tau) \\ \widehat{\underline{Q}}^c_{Y_x}(\tau) \end{pmatrix} = [\phi_2(\tilde{\phi}_2(\widehat{\theta}))](\tau, x, c).$$

We will show that ϕ_2 is Hadamard differentiable tangentially to the space of uniformly continuous functions on $(0, 1) \times \{0, 1\}$. The Hadamard derivative of the first component of ϕ_2 evaluated at $\psi_0 \equiv \tilde{\phi}_2(\theta_0)$ is

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

To see this, a Taylor expansion gives

$$\frac{\phi_2^{(1)}(\psi_0 + t_n h_n) - \phi_2^{(1)}(\psi_0)}{t_n} = h_n^{(1)}\left(\tau + \frac{c}{\psi_0^{(2)}(x) + t_n h_n^{(2)}(x)} \min\{\tau, 1-\tau\}, x\right) - \psi_0^{(1)\prime}\left(\tau + \frac{c}{\psi_0^{(2)}(x) + a_n(x)} \min\{\tau, 1-\tau\}, x\right) \frac{c \min\{\tau, 1-\tau\}}{(\psi_0^{(2)}(x) + a_n(x))^2} h_n^{(2)}(x)$$

using the fact that $\psi_0^{(1)}(\tau, x) = Q_{Y|X}(\tau \mid x)$ is continuously differentiable in τ by assumption A4.2, and noting that term $a_n(x)$ satisfies $|a_n(x)| \leq |t_n h_n^{(2)}(x)| = O(t_n)$. Next,

$$\begin{aligned}
&\sup_{\tau \in (0,1), x \in \{0,1\}, c \in [0, \overline{C}]} \left| h_n^{(1)}\left(\tau + \frac{c \min\{\tau, 1-\tau\}}{\psi_0^{(2)}(x) + t_n h_n^{(2)}(x)}, x\right) - h^{(1)}\left(\tau + \frac{c \min\{\tau, 1-\tau\}}{\psi_0^{(2)}(x)}, x\right) \right| \\
&\leq \sup_{\tau \in (0,1), x \in \{0,1\}, c \in [0, \overline{C}]} \left| h_n^{(1)}\left(\tau + \frac{c \min\{\tau, 1-\tau\}}{\psi_0^{(2)}(x) + t_n h_n^{(2)}(x)}, x\right) - h^{(1)}\left(\tau + \frac{c \min\{\tau, 1-\tau\}}{\psi_0^{(2)}(x) + t_n h_n^{(2)}(x)}, x\right) \right| \\
&\quad + \sup_{\tau \in (0,1), x \in \{0,1\}, c \in [0, \overline{C}]} \left| h^{(1)}\left(\tau + \frac{c \min\{\tau, 1-\tau\}}{\psi_0^{(2)}(x) + t_n h_n^{(2)}(x)}, x\right) - h^{(1)}\left(\tau + \frac{c \min\{\tau, 1-\tau\}}{\psi_0^{(2)}(x)}, x\right) \right| \\
&\leq \|h_n^{(1)} - h^{(1)}\|_\infty + o(1) \\
&= o(1).
\end{aligned}$$

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

$$\sup_{\tau \in (0,1), x \in \{0,1\}, c \in [0, \overline{C}]} \left| \psi_0^{(1)\prime}\left(\tau + \frac{c}{\psi_0^{(2)}(x) + a_n(x)} \min\{\tau, 1-\tau\}, x\right) \frac{c \min\{\tau, 1-\tau\}}{(\psi_0^{(2)}(x) + a_n(x))^2} h_n^{(2)}(x) - \psi_0^{(1)\prime}\left(\tau + \frac{c}{\psi_0^{(2)}(x)} \min\{\tau, 1-\tau\}, x\right) \frac{c \min\{\tau, 1-\tau\}}{(\psi_0^{(2)}(x))^2} h^{(2)}(x) \right| = o(1)$$

by uniform continuity of $\psi_0^{(1)\prime}$ (implied by A4.2) and by $a_n(x) = o(1)$. 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 $\mathbf{Z}_1(y, x)$ has uniformly continuous paths, we have that

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

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

The following lemma is a variation of lemma 21.3 on page 306 of van der Vaart (2000), allowing for a directionally differentiable function F rather than a fully differentiable F. We use this lemma in our analysis of ATE breakdown points in the proof of proposition 1 below.

Lemma 7. Let $F : [0,1] \to \mathbb{R}$ be a continuous, non-decreasing function. Let $p \in \mathbb{R}$ be such that $F(\xi_p) = p$, F is directionally differentiable at $\xi_p \in [0,1]$ with $\partial^+ F(\xi_p) > 0$ and $\partial^- F(\xi_p) > 0$. Define the mapping ϕ by $F \mapsto \inf\{y \in \mathbb{R} : F(y) \geq p\}$. Then ϕ is Hadamard directionally differentiable at F tangentially to the set of càdlàg functions on $[0,1]$, $\mathscr{D}([0,1])$, with Hadamard directional derivative equal to

$$\phi'_F(h) = -h(\xi_p) \left(\frac{\mathbb{1}(h(\xi_p) > 0)}{\partial^- F(\xi_p)} + \frac{\mathbb{1}(h(\xi_p) < 0)}{\partial^+ F(\xi_p)} \right).$$

Proof of lemma 7. This proof follows that of lemma 21.3 in van der Vaart (2000). Let $h_n \to h$ in the sup-norm, $h \in \mathscr{D}([0,1])$, $t_n \searrow 0$, and let $\xi_{pn} = \phi(F + t_n h_n)$. By the proof of lemma 21.3 in van der Vaart (2000), we can write

$$[F + t_n h_n](\xi_{pn} - \epsilon_n) \leq p \leq [F + t_n h_n](\xi_{pn})$$

where $\xi_{pn} \to \xi_p$, $\epsilon_n > 0$, and $\epsilon_n = o(t_n)$. He also shows that $h_n(\xi_{pn} - \epsilon_n) = h(\xi_p) + o(1)$. Since $p = F(\xi_p)$,

$$[F + t_n h_n](\xi_{pn} - \epsilon_n) \leq F(\xi_p) \leq [F + t_n h_n](\xi_{pn}).$$

Rearranging the second inequality yields

$$-h_n(\xi_{pn}) \leq \frac{F(\xi_{pn}) - F(\xi_p)}{t_n}$$

while rearranging the first inequality yields

$$\frac{F(\xi_{pn} - \epsilon_n) - F(\xi_p)}{t_n} \leq -h_n(\xi_{pn} - \epsilon_n).$$

Consider the left hand side:

$$\begin{aligned}
\frac{F(\xi_{pn} - \epsilon_n) - F(\xi_p)}{t_n} &= \frac{F(\xi_{pn}) - F(\xi_p)}{t_n} + \frac{F(\xi_{pn} - \epsilon_n) - F(\xi_{pn})}{t_n} \\
&= \frac{F(\xi_{pn}) - F(\xi_p)}{t_n} + O\left(\frac{\epsilon_n}{t_n}\right) \\
&= \frac{F(\xi_{pn}) - F(\xi_p)}{t_n} + o(1).
\end{aligned}$$

The second equality follows from a Taylor expansion. The last equality follows from $\epsilon_n / t_n = o(1)$. Therefore,

$$\frac{F(\xi_{pn}) - F(\xi_p)}{t_n} \leq -h_n(\xi_{pn} - \epsilon_n) + o(1).$$

Combining these two inequalities gives

$$-h_n(\xi_{pn}) \leq \frac{F(\xi_{pn}) - F(\xi_p)}{t_n} \leq -h_n(\xi_{pn} - \epsilon_n) + o(1).$$

Thus

$$\frac{F(\xi_{pn}) - F(\xi_p)}{t_n} = -h(\xi_p) + o(1)$$

since $h_n(\xi_{pn} - \epsilon_n) = h(\xi_p) + o(1)$, by uniform convergence of h_n to h, $\epsilon_n \to 0$, and by $\xi_{pn} \to \xi_p$.

Case 1: Let $h(\xi_p) > 0$. Then there is an N such that $h_n(\xi_{pn}) > 0$ for all $n \geq N$ by uniform

convergence of h_n to h, and by $\xi_{pn} = \xi_p + o(1)$. By definition of ξ_{pn}, $[F + t_n h_n](\xi_{pn}) = p$. Hence

$$\begin{aligned} F(\xi_{pn}) + t_n h_n(\xi_{pn}) &= p \\ &= F(\xi_p). \end{aligned}$$

Thus

$$\begin{aligned} \frac{F(\xi_{pn}) - F(\xi_p)}{t_n} &= -h_n(\xi_{pn}) \\ &< 0 \end{aligned}$$

for $n \geq N$. Hence $F(\xi_{pn}) < F(\xi_p)$ and therefore $\xi_{pn} < \xi_p$ for $n \geq N$. Therefore, we have that

$$\begin{aligned} \frac{F(\xi_{pn}) - F(\xi_p)}{t_n} &= \frac{F(\xi_{pn}) - F(\xi_p)}{\xi_{pn} - \xi_p} \frac{\xi_{pn} - \xi_p}{t_n} \\ &= \partial^- F(\xi_p) \times \phi'_F(h) + o(1). \end{aligned}$$

For the second line, recall that $\xi_{pn} = \phi(F + t_n h_n)$ and $\xi_p = \phi(F)$ by definition. Combining this result with our derivation above of the limit of the left hand side shows that

$$\phi'_F(h) = \frac{-h(\xi_p)}{\partial^- F(\xi_p)}$$

Case 2: Let $h(\xi_p) < 0$. Here we similarly conclude

$$\frac{F(\xi_{pn}) - F(\xi_p)}{t_n} = \partial^+ F(\xi_p) \times \phi'_F(h) + o(1).$$

and hence

$$\phi'_F(h) = \frac{-h(\xi_p)}{\partial^+ F(\xi_p)}.$$

Case 3: Finally, let $h(\xi_p) = 0$. Then

$$\frac{F(\xi_{pn}) - F(\xi_p)}{t_n} = o(1)$$

by our earlier derivations. For a given n, if $\xi_{pn} \geq \xi_p$, we have that

$$|F(\xi_{pn}) - F(\xi_p)| = \partial^+ F(\tilde{\xi}_p)|\xi_{pn} - \xi_p|$$

by a one-sided Taylor expansion for some $\tilde{\xi}_p$ between ξ_{pn} and ξ_p. If $\xi_{pn} < \xi_p$, we have

$$|F(\xi_{pn}) - F(\xi_p)| = \partial^- F(\xi_p^*)|\xi_{pn} - \xi_p|$$

for some ξ_p^* between ξ_{pn} and ξ_p. Combining these two cases we have

$$\begin{aligned} o(1) &= \left| \frac{F(\xi_{pn}) - F(\xi_p)}{t_n} \right| \\ &= \Big(\partial^- F(\xi_p^*) \mathbb{1}(\xi_{pn} < \xi_p) + \partial^+ F(\tilde{\xi}_p) \mathbb{1}(\xi_{pn} \geq \xi_p) \Big) \frac{|\xi_{pn} - \xi_p|}{t_n}. \end{aligned}$$

Since $\partial^+ F(\xi_p)$ and $\partial^- F(\xi_p)$ are bounded above, we have $\xi_{pn} - \xi_p = o(t_n)$. Hence $\phi'_F(h) = 0$.

Combining all three cases yields

$$\phi'_F(h) = -h(\xi_p) \left(\frac{\mathbb{1}(h(\xi_p) > 0)}{\partial^- F(\xi_p)} + \frac{\mathbb{1}(h(\xi_p) < 0)}{\partial^+ F(\xi_p)} \right)$$

as $n \to \infty$. Finally, since $\phi'_F(h)$ is continuous in h, ϕ is Hadamard directionally differentiable. □

Proof of proposition 1. Consider the QTE bound of equation (4) as a function of c, for a fixed τ. It satisfies the assumptions of lemma 7. The ATE lower bound also satisfies these assumptions since it is the integral of this QTE bound over $\tau \in (0, 1)$. By the discussion following lemma 3, $\sqrt{N}(\widehat{\underline{\text{ATE}}}(c) - \underline{\text{ATE}}(c))$ converges in distribution to a random element of $\ell^\infty([0, \overline{C}])$ with continuous paths.

Let

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

By $c^* \in [0, \overline{C}]$ and lemma 7 applied to the domain $[0, \overline{C}]$, we can apply the functional delta method for Hadamard directionally differentiable functions to see that $\sqrt{N}(\tilde{c}^* - c^*)$ converges in distribution to a random variable we denote by $\mathbf{Z}_4$.

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,

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

Therefore, the set $\{c \in [0, \overline{C}] : \widehat{\underline{\text{ATE}}}(c) \leq \mu\}$ is non-empty 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

$$\begin{aligned} \sqrt{N}(\widehat{c}^* - c^*) &= \sqrt{N}(\widehat{c}^* - \tilde{c}^*) + \sqrt{N}(\tilde{c}^* - c^*) \\ &= o_p(1) + \sqrt{N}(\tilde{c}^* - c^*) \\ &\rightsquigarrow \mathbf{Z}_4. \end{aligned}$$

□

The following result extends proposition 2(i) of Chernozhukov et al. (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 8. Let $\theta_0(u, c) = (\theta_0^{(1)}(u, c), \theta_0^{(2)}(u, c))$ where for $j \in \{1, 2\}$ we have that $\theta_0^{(j)}(u, c)$ is strictly increasing in $u \in [0, 1]$, bounded above and below, and differentiable everywhere except at $u = u^*$,

where it is directionally differentiable. Further, assume that the two components satisfy A5. Then, for fixed $z \in \mathbb{R}$, the mapping $\phi_3 : \ell^\infty((0,1) \times \mathcal{C}, \mathbb{R}^2) \to \ell^\infty(\mathcal{C}, \mathbb{R}^2)$ defined by

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

is Hadamard directionally differentiable tangentially to $\mathscr{C}((0,1)\times\mathcal{C}, \mathbb{R}^2)$ with Hadamard directional derivative given by equations (17) and (18) below.

Proof of lemma 8. Our proof follows that of proposition 2(i) in Chernozhukov et al. (2010). Let

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

denote the set of roots to the equation $\theta_0^{(1)}(u,c) = z$ for fixed z and c. By A5.1 this set contains a finite number of elements. We denote these by

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

A5.1 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

$$[\phi_{3,\theta_0}^{(1)\prime}(h)](c) = -\sum_{k=1}^{K^{(1)}(c)} h(u_k^{(1)}(c), c) \left(\frac{\mathbb{1}(h(u_k^{(1)}(c), c) > 0)}{|\partial_u^- \theta_0^{(1)}(u_k^{(1)}(c), c)|} + \frac{\mathbb{1}(h(u_k^{(1)}(c), c) < 0)}{|\partial_u^+ \theta_0^{(1)}(u_k^{(1)}(c), c)|} \right), \tag{17}$$

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 et al. (2010) directly to obtain

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

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

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

as $n \to \infty$. Hence

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

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 et al. (2010), for any

$\delta > 0$ there exists an $\epsilon > 0$ and a large enough n such that

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

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

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

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

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

$\theta_0(u, c)$ 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

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

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 A5.1, 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 $(0, u^*)$ and directional differentiability as $u = u^*$ 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,

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

   By a similar argument,

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

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

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

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

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

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

Combining these three cases into a single expression we find that

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

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 et al. (2010) that

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

where $\| \cdot \|_e$ is the Euclidean norm, and where $\phi_{3,\theta_0}^{(1)\prime}$ is defined in equation (17). 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

$$[\phi_{3,\theta_0}^{(2)\prime}(h)](c) = -\sum_{k=1}^{K^{(2)}(c)} h(u_k^{(2)}(c), c) \left(\frac{\mathbb{1}(h(u_k^{(2)}(c), c) > 0)}{|\partial_u^- \theta_0^{(2)}(u_k^{(2)}(c), c)|} + \frac{\mathbb{1}(h(u_k^{(2)}(c), c) < 0)}{|\partial_u^+ \theta_0^{(2)}(u_k^{(2)}(c), c)|} \right). \tag{18}$$

□

Proof of lemma 4. Let

$$\theta_0(\tau, c) = \begin{pmatrix} \overline{Q}_{Y_1}^c(\tau) - \overline{Q}_{Y_0}^c(\tau) \\ \overline{Q}_{Y_1}^c(\tau) - \underline{Q}_{Y_0}^c(\tau) \end{pmatrix} \qquad \text{and} \qquad \widehat{\theta}(\tau, c) = \begin{pmatrix} \widehat{\overline{Q}}_{Y_1}^c(\tau) - \widehat{\overline{Q}}_{Y_0}^c(\tau) \\ \widehat{\overline{Q}}_{Y_1}^c(\tau) - \widehat{\underline{Q}}_{Y_0}^c(\tau) \end{pmatrix}.$$

Therefore

$$\begin{pmatrix} \overline{P}(c) \\ \underline{P}(c) \end{pmatrix} = [\phi_3(\theta_0)](c) \qquad \text{and} \qquad \begin{pmatrix} \widehat{\overline{P}}(c) \\ \widehat{\underline{P}}(c) \end{pmatrix} = [\phi_3(\widehat{\theta})](c).$$

By lemma 3,

$$\sqrt{N}\begin{pmatrix} \widehat{\underline{Q}}^c_{Y_1}(\tau) - \widehat{\overline{Q}}^c_{Y_0}(\tau) - (\underline{Q}^c_{Y_1}(\tau) - \overline{Q}^c_{Y_0}(\tau)) \\ \widehat{\overline{Q}}^c_{Y_1}(\tau) - \widehat{\underline{Q}}^c_{Y_0}(\tau) - (\overline{Q}^c_{Y_1}(\tau) - \underline{Q}^c_{Y_0}(\tau)) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \mathbf{Z}^{(2)}_3(\tau,1,c) - \mathbf{Z}^{(1)}_3(\tau,0,c) \\ \mathbf{Z}^{(1)}_3(\tau,1,c) - \mathbf{Z}^{(2)}_3(\tau,0,c) \end{pmatrix},$$

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

By lemma 8 with $u^* = 1/2$, the mapping ϕ_3 is Hadamard directionally differentiable tangentially to $\mathscr{C}((0,1)\times\mathcal{C},\mathbb{R}^2)$. By the functional delta method for Hadamard directionally differentiable functions (e.g., theorem 2.1 in Fang and Santos 2015), we obtain

$$\begin{aligned}\sqrt{N}\begin{pmatrix} \widehat{\overline{P}}(c) - \overline{P}(c) \\ \widehat{\underline{P}}(c) - \underline{P}(c) \end{pmatrix} &\rightsquigarrow \begin{pmatrix} \left[\phi^{(2)\prime}_{3,\underline{Q}^{(\cdot)}_{Y_1}(\cdot)-\overline{Q}^{(\cdot)}_{Y_0}(\cdot)}(\mathbf{Z}^{(2)}_3(\cdot,1,\cdot) - \mathbf{Z}^{(1)}_3(\cdot,0,\cdot))\right](c) \\ \left[\phi^{(1)\prime}_{3,\overline{Q}^{(\cdot)}_{Y_1}(\cdot)-\underline{Q}^{(\cdot)}_{Y_0}(\cdot)}(\mathbf{Z}^{(1)}_3(\cdot,1,\cdot) - \mathbf{Z}^{(2)}_3(\cdot,0,\cdot))\right](c) \end{pmatrix} \\ &\equiv \mathbf{Z}_5(c),\end{aligned}$$

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

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 9. 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

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

Let

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

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

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

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

Proof of lemma 9. 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

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

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 $t_n \searrow 0$,

$$\begin{aligned}\sup_{c\in\mathcal{C}}\left|\sup_{a\in\mathcal{A}}\big(\theta(a,c)+t_n h_n(a,c)\big)-\sup_{a\in\mathcal{A}}\big(\theta(a,c)+t_n h(a,c)\big)\right| &\le \sup_{c\in\mathcal{C}} t_n \sup_{a\in\mathcal{A}}|h_n(a,c)-h(a,c)| \\ &= t_n\|h_n-h\|_\infty \\ &= o(t_n). \end{aligned} \tag{19}$$

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

$$\begin{aligned}&\sup_{c\in\mathcal{C}}\left|\sup_{a\in\mathcal{A}}\big(\theta(a,c)+t_n h(a,c)\big)-\sup_{a\in\Psi_{\mathcal{A}}(\theta,c)}\big(\theta(a,c)+t_n h(a,c)\big)\right| \\ &\le t_n \sup_{c\in\mathcal{C}}\sup_{a_0,a_1\in\mathcal{A}:|a_0-a_1|\le\delta_n}|h(a_0,c)-h(a_1,c)| \\ &= o(t_n)\end{aligned} \tag{20}$$

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 (19) and (20) as in equation (B.4) from Fang and Santos (2015), it follows that

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

which completes the proof. □

Proof of lemma 5. We begin by showing that the first component in equation (11) converges to a tight random element of $\ell^\infty(\mathcal{C}\times[0,1])$. Fix c and define

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

by

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

As in the proof of lemma 2, the four mappings $(\delta_1,\delta_2,\delta_3,\delta_4)$ from $\ell^\infty(\mathbb{R}\times\{0,1\})\times\ell^\infty(\{0,1\})$ to $\ell^\infty(\mathbb{R}\times\{0,1\})$ are all Hadamard differentiable when evaluated at θ_0.

We can write

$$
\begin{aligned}
&\phi_4(\theta_0) \\
&= \max\left\{\sup_{a\in\mathcal{Y}_z}(\underline{F}^c_{Y_1}(a) - \overline{F}^c_{Y_0}(a-z)), 0\right\} \\
&= \max\left\{\sup_{a\in\mathcal{Y}_z}\Big(\max\{[\delta_1(\theta_0)](a,1), [\delta_2(\theta_0)](a,1)\} - \min\{[\delta_3(\theta_0)](a-z,0), [\delta_4(\theta_0)](a-z,0)\}\Big), 0\right\} \\
&= \max\left\{\sup_{a\in\mathcal{Y}_z}\Big([\delta_1(\theta_0)](a,1) - [\delta_3(\theta_0)](a-z,0)\Big), \sup_{a\in\mathcal{Y}_z}\Big([\delta_1(\theta_0)](a,1) - [\delta_4(\theta_0)](a-z,0)\Big),\right. \\
&\qquad\left.\sup_{a\in\mathcal{Y}_z}\Big([\delta_2(\theta_0)](a,1) - [\delta_3(\theta_0)](a-z,0)\Big), \sup_{a\in\mathcal{Y}_z}\Big([\delta_2(\theta_0)](a,1) - [\delta_4(\theta_0)](a-z,0)\Big), 0\right\}.
\end{aligned}
$$

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 6) and lemma 9, the mappings

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

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. Uniformity over $c \in \mathcal{C}$ is obtained from considering the vector of Hadamard directional derivatives for all $c \in \mathcal{C}$.

By lemma 4, the mapping $(F_{Y|X}(\cdot \mid \cdot), p_{(\cdot)}) \mapsto \underline{P}(\cdot)$ is Hadamard directionally differentiable. Linearity of the Hadamard directional derivative operator yields that the mapping $(F_{Y|X}(\cdot \mid \cdot), p_{(\cdot)}) \mapsto \underline{\text{DTE}}(z, \cdot, \cdot)$ is Hadamard directionally differentiable.

Since

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

the infimum operator is Hadamard directionally differentiable. As in the proof of lemma 2, the minimum operator is Hadamard directionally differentiable. Following lemma 4, the mapping $(F_{Y|X}(\cdot \mid \cdot), p_{(\cdot)}) \mapsto \overline{P}(\cdot)$ is Hadamard directionally differentiable. A similar argument as above implies the mapping $(F_{Y|X}(\cdot \mid \cdot), p_{(\cdot)}) \mapsto \overline{\text{DTE}}(z, \cdot, \cdot)$ is Hadamard directionally differentiable.

Combining these results with lemma 1 allows us to conclude that equation (11) holds. □

Proof of theorem 2. By lemma 4, the numerator of equation (13) converges in process over $c \in \mathcal{C}$. By lemmas 4 and 5, the denominator also converges uniformly over $c \in \mathcal{C}$. By the delta method,

$$
\begin{aligned}
\sqrt{N}\left(\widehat{\text{bf}}(0,c,\underline{p}) - \text{bf}(0,c,\underline{p})\right) &= \sqrt{N}\left(\frac{1-\underline{p}-\widehat{\overline{P}}(c)}{1+\min\left\{\inf_{y\in\mathcal{Y}_0}(\widehat{\overline{F}}^c_{Y_1}(y) - \widehat{\underline{F}}^c_{Y_0}(y)), 0\right\} - \widehat{\overline{P}}(c)}\right. \\
&\qquad\left. - \frac{1-\underline{p}-\overline{P}(c)}{1+\min\left\{\inf_{y\in\mathcal{Y}_0}(\overline{F}^c_{Y_1}(y) - \underline{F}^c_{Y_0}(y)), 0\right\} - \overline{P}(c)}\right) \\
&\rightsquigarrow \frac{-\mathbf{Z}^{(1)}_5(c)}{1+\min\left\{\inf_{y\in\mathcal{Y}_0}(\overline{F}^c_{Y_1}(y) - \underline{F}^c_{Y_0}(y)), 0\right\} - \overline{P}(c)} \\
&\qquad - \frac{1-\underline{p}-\overline{P}(c)}{\left(1+\min\left\{\inf_{y\in\mathcal{Y}_0}(\overline{F}^c_{Y_1}(y) - \underline{F}^c_{Y_0}(y)), 0\right\} - \overline{P}(c)\right)^2}\tilde{\mathbf{Z}}(c)
\end{aligned}
$$

where

$$\sqrt{N}\left(1-\underline{p}-\widehat{\overline{P}}(c)-(1-\underline{p}-\overline{P}(c))\right) \rightsquigarrow -\mathbf{Z}_5^{(1)}(c)$$

and

$$\sqrt{N}\left(\min\left\{\inf_{y\in\mathcal{Y}_0}(\widehat{\overline{F}}^c_{Y_1}(y)-\widehat{\underline{F}}^c_{Y_0}(y)),0\right\}-\widehat{\overline{P}}(c)-\min\left\{\inf_{y\in\mathcal{Y}_0}(\overline{F}^c_{Y_1}(y)-\underline{F}^c_{Y_0}(y)),0\right\}+\overline{P}(c)\right)\rightsquigarrow \tilde{\mathbf{Z}}(c).$$

Here $\tilde{\mathbf{Z}}(c)$ is a random element of $\ell^\infty(\mathcal{C})$ by lemmas 4 and 5. Therefore,

$$\sqrt{N}\left(\widehat{\mathrm{bf}}(0,c,\underline{p})-\mathrm{bf}(0,c,\underline{p})\right)$$

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

As discussed in the proof of lemma 2, the maximum and minimum operators in equation (12) are Hadamard directionally differentiable. By lemma 6, their composition is Hadamard directionally differentiable. Therefore, by the delta method for Hadamard directionally differentiable functions, $\sqrt{N}(\widehat{\mathrm{BF}}(0,c,\underline{p})-\mathrm{BF}(0,c,\underline{p}))$ converges in process as in the statement of the theorem. □

Lemma 10. 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 10. For functions h and h',

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

By a symmetric argument,

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

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

Proof of proposition 2. Hadamard directional differentiability of ϕ follows from the chain rule (lemma 6) and from the proof of theorem 2, since the breakdown frontier is a Hadamard directionally differentiable mapping of $\overline{P}(\cdot)$ and $\overline{\mathrm{DTE}}(z,\cdot,\cdot)$, which are themselves Hadamard directionally differentiable mappings of θ_0.

Lemma 1 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 (2015), equation (15) holds.

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

$$\frac{\left[\widehat{\phi}'_{\theta_0}(\sqrt{N}(\widehat{\theta}^*-\widehat{\theta}))\right](c,\underline{p})}{\sigma(c)}\overset{P}{\rightsquigarrow}\frac{\mathbf{Z}_7(c,\underline{p})}{\sigma(c)}.$$

By lemma 10, 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

$$\sup_{c\in\mathcal{C}} \frac{\left[\widehat{\phi}'_{\theta_0}(\sqrt{N}(\widehat{\theta}^* - \widehat{\theta}))\right](c,\underline{p})}{\sigma(c)} \overset{P}{\rightsquigarrow} \sup_{c\in\mathcal{C}} \frac{\mathbf{Z}_7(c,\underline{p})}{\sigma(c)}.$$

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

Proof of corollary 1. This follows immediately from proposition 2 and lemma 11 below. □

Lemma 11. Let $\overline{C} > 0$ and $\mathcal{C} = \{c_1, \ldots, c_J\} \subseteq [0, \overline{C}]$ be a finite grid of points. Let $f : [0, \overline{C}] \to \mathbb{R}_+$ be a nonincreasing function. Let $\widehat{\text{LB}}(\cdot)$ be an asymptotically exact uniform lower $1-\alpha$ confidence band for f on the grid points:

$$\lim_{N\to\infty} \mathbb{P}\left(\widehat{\text{LB}}(c_j) \leq f(c_j) \text{ for } j = 1, \ldots, J\right) = 1 - \alpha.$$

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

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

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

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

Proof of lemma 11. Define the events

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

and

$$B = \{\widehat{\text{LB}}(c_j) \leq f(c_j) \text{ for } j = 1, \ldots, J\}.$$

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

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

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

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

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

C Inference via population smoothing

In this section we develop an alternative approach to constructing lower uniform confidence bands for the breakdown frontier. As discussed in section 3, the population breakdown frontier $\text{BF}(0, \cdot, \underline{p})$ evaluated on a finite grid of c's is a Hadamard directionally differentiable functional of the underlying parameters $\theta_0 = (F_{Y|X}(\cdot \mid \cdot), p_{(\cdot)})$, but it is not necessarily ordinary Hadamard differentiable. We therefore applied the work of Dümbgen (1993), Fang and Santos (2015), and Hong and Li (2015) to do inference. In this section, we instead replace $\text{BF}(0, \cdot, \underline{p})$ by a smoother lower envelope function. We then construct uniform lower confidence bands for this smoothed breakdown frontier, which are asymptotically valid—but potentially conservative—for the original breakdown frontier. We compare and contrast these two approaches to inference at the end of this section.

Specifically, recall from section 2 that

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

where

$$\text{bf}(z, c, \underline{p}) = \frac{1 - \underline{p} - \mathbb{P}(\underline{Q}^c_{Y_1}(U) - \overline{Q}^c_{Y_0}(U) \leq z)}{1 + \min\left\{\inf_{y \in \mathcal{Y}_z}(\overline{F}^c_{Y_1}(y) - \underline{F}^c_{Y_0}(y - z)), 0\right\} - \mathbb{P}(\underline{Q}^c_{Y_1}(U) - \overline{Q}^c_{Y_0}(U) \leq z)}. \tag{9}$$

For simplicity, we fix $z \in \mathbb{R}$ and $\underline{p} \in [0, 1]$ throughout this section. We use κ throughout to denote a scalar or vector of smoothing parameters. We replace $\text{BF}(z, \cdot, \underline{p})$ by a *smooth lower approximation* $\text{SBF}_\kappa(z, \cdot, \underline{p})$, defined as follows.

Definition 4. Let $(\Theta, \|\cdot\|_\Theta)$ and $(\mathscr{G}, \|\cdot\|_\mathscr{G})$ be Banach spaces. Let $\leq$ be a partial order on $\mathscr{G}$. Let $f : \Theta \to \mathscr{G}$ be a function. Consider a function $f_\kappa : \Theta \to \mathscr{G}$, where $\kappa \in \mathbb{R}_+^{\dim(\kappa)}$ is a vector of bandwidths. We say f_κ is a *smooth lower approximation* of f if it satisfies the following:

1. (Lower envelope) $f_\kappa(\theta) \leq f(\theta)$ for all $\theta \in \Theta$ and $\kappa \in \mathbb{R}_+^{\dim(\kappa)}$.
2. (Approximation) For each $\theta \in \Theta$, $f_\kappa(\theta) \to f(\theta)$ as all components of κ converge to infinity. Here we take $\to$ to mean pointwise convergence.
3. (Smoothness) f_κ is Hadamard differentiable, possibly only tangentially to a specified set.

Define smooth *upper* approximations analogously.

Throughout this section we let $\leq$ denote the component-wise order when applied to functions with Euclidean codomain. Recall our notation $\theta_0 = (F_{Y|X}(\cdot \mid \cdot), p_{(\cdot)})$, $\widehat{\theta} = (\widehat{F}_{Y|X}(\cdot \mid \cdot), \widehat{p}_{(\cdot)})$, and $\mathbf{Z}_1$ as the limiting distribution of $\sqrt{N}(\widehat{\theta} - \theta_0)$. Below we show how to construct a functional $\psi : \ell^\infty(\mathbb{R} \times \{0, 1\}) \times \ell^\infty(\{0, 1\}) \to \ell^\infty([0, \overline{C}])$ which maps θ_0 into $\text{SBF}_\kappa(0, \cdot, \underline{p})$ such that this functional is a smooth lower approximation of the functional mapping θ_0 to $\text{BF}(0, \cdot, \underline{p})$. Given such a functional, we estimate the smoothed breakdown frontier by sample analog:

$$\widehat{\text{SBF}}_\kappa(0, c, \underline{p}) = [\psi(\widehat{\theta})](c).$$

We then construct uniform confidence bands for the breakdown frontier as follows. As in section 3, consider bands of the form

$$\widehat{\text{LB}}(c) = \widehat{\text{SBF}}_\kappa(0, c, \underline{p}) - \widehat{k}(c)$$

for some function $\widehat{k}(\cdot) \geq 0$. We specifically focus on $\widehat{k}(c) = \widehat{z}_{1-\alpha}\sigma(c)$ for a scalar $\widehat{z}_{1-\alpha}$ and a known function σ, for simplicity. We now immediately obtain the following result.

Proposition 5. Suppose A1, A3, and A4 hold. Let ψ denote the functional described above, a smooth lower approximation to the breakdown frontier functional. Let $\widehat{\theta}^*$ denote a draw from the nonparametric bootstrap distribution of $\widehat{\theta}$. Then

$$\sqrt{N}(\psi(\widehat{\theta}^*) - \psi(\widehat{\theta})) \overset{P}{\rightsquigarrow} \psi'_{\theta_0}(\mathbf{Z}_1) \tag{21}$$

where ψ'_{θ_0} denotes the Hadamard derivative of ψ at θ_0. For a given function $\sigma(\cdot)$ such that $\inf_{c\in[0,\overline{C}]} \sigma(c) > 0$, define

$$\widehat{z}_{1-\alpha} = \inf\left\{ z \in \mathbb{R} : \mathbb{P}\left(\sup_{c\in[0,\overline{C}]} \frac{\sqrt{N}([\psi(\widehat{\theta}^*)](c) - [\psi(\widehat{\theta})](c))}{\sigma(c)} \leq z \mid Z^N \right) \geq 1-\alpha \right\}. \tag{22}$$

Finally, suppose also that the cdf of

$$\sup_{c\in[0,\overline{C}]} \frac{[\psi'_{\theta_0}(\mathbf{Z}_1)](c)}{\sigma(c)}$$

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

Corollary 2. Suppose the assumptions of proposition 5 hold. Let $\widehat{k}(c) = \widehat{z}_{1-\alpha}\sigma(c)$, where $\widehat{z}_{1-\alpha}$ is defined in equation (22). Then

$$\lim_{N\to\infty} \mathbb{P}\left(\widehat{\text{SBF}}_\kappa(0, c, \underline{p}) - \widehat{k}(c) \leq \text{BF}(0, c, \underline{p}) \text{ for all } c \in [0, \overline{C}] \right) \geq 1-\alpha.$$

Importantly, the level of smoothing κ is *fixed* asymptotically. This is analogous to the requirement that the grid of c's must be fixed asymptotically in the approach discussed in section 3. It is also similar to a proposal by Chernozhukov et al. (2010) in their corollary 4. They suggested replacing the non-smooth function with a smoothed version and doing inference on the smoothed version. Their approach delivers valid inference on the smoothed function, but not the original function. This follows since their smoothed function does not satisfy an envelope property. Our modification of their suggestion, however, delivers valid inference on both the smoothed and original functions.

All that remains is to construct such a function $\text{SBF}_\kappa(z, c, \underline{p})$. We consider each piece composing the function $\text{BF}(z, c, \underline{p})$ in turn. First consider $\overline{F}^c_{Y_1}(y)$. This bound is a minimum of two terms (see equation 7). In general, consider the minimum of a finite number of terms $x_1, \ldots, x_n$. There are many smooth approximations of this function. Here we just consider one:

$$\text{sm}_\kappa\{x_1, \ldots, x_n\} = \sum_{i=1}^n x_i \frac{\exp(\kappa x_i)}{\sum_{j=1}^n \exp(\kappa x_j)}$$

for $\kappa < 0$. This same function approximates $\max\{x_1, \ldots, x_n\}$ for $\kappa > 0$. Let $\mathcal{D}$ be a subset of a Euclidean space. Let $\mathbb{D}_1$ denote the set of functions in $\ell^\infty(\mathcal{D})^n$ with range contained in some compact set $\mathcal{Y} \subseteq \mathbb{R}^n$. In our application, we are interested using the functional $\psi_{1,\kappa} : \mathbb{D}_1 \to \ell^\infty(\mathcal{D})$ defined by

$$[\psi_{1,\kappa}(f_1, \ldots, f_n)](y) = \text{sm}_\kappa\{f_1(y), \ldots, f_n(y)\}$$

to approximate the functionals $\psi_{1,\max} : \mathbb{D}_1 \to \ell^\infty(\mathcal{D})$ defined by

$$[\psi_{1,\max}(f_1, \ldots, f_n)](y) = \max\{f_1(y), \ldots, f_n(y)\}$$

and $\psi_{1,\min} : \mathbb{D}_1 \to \ell^\infty(\mathcal{D})$ defined by

$$[\psi_{1,\min}(f_1, \ldots, f_n)](y) = \min\{f_1(y), \ldots, f_n(y)\}.$$

Lemma 12. Let $\kappa \in \mathbb{R}_+$.

1. $\psi_{1,\kappa}$ is a smooth lower approximation of $\psi_{1,\max}$.
2. $\psi_{1,-\kappa}$ is a smooth upper approximation of $\psi_{1,\min}$.

Since $\overline{F}^c_{Y_1}(y)$ enters the denominator of equation (9), and since sm_κ for $\kappa < 0$ is an upper envelope for the minimum, replacing the minimum in the definition of $\overline{F}^c_{Y_1}(y)$ with sm_κ for $\kappa < 0$ decreases the value of equation (9). Similarly, replacing the maximum in the definition of $\underline{F}^c_{Y_0}(y)$ by sm_κ for some $\kappa > 0$ decreases the value of equation (9).

Next consider the $\mathbb{P}(\underline{Q}^c_{Y_1}(U) - \overline{Q}^c_{Y_0}(U) \le z)$ term. As discussed in section 3, this term is the pre-rearrangement operator $\mathrm{pr} : \ell^\infty((0,1) \times \mathcal{C}) \to \ell^\infty(\mathcal{C})$ defined by

$$[\mathrm{pr}(f)](c) = \int_0^1 \mathbb{1}[f(u,c) \le z]\, du$$

evaluated at the difference of the quantile bounds, where $\mathcal{C} \subseteq [0,1]$. Define the smoothed pre-rearrangement operator $\mathrm{spr}_\kappa : \ell^\infty((0,1) \times \mathcal{C}) \to \ell^\infty(\mathcal{C})$ by

$$[\mathrm{spr}_\kappa(f)](c) = 1 - \int_0^1 \mathrm{ss}_\kappa(f(u,c))\, du$$

where ss_κ is a smooth (upper or lower) approximation to the step function $\mathbb{1}(x \ge 0)$.

Lemma 13. Let $\mathrm{ss}_\kappa : \mathbb{R} \to \mathbb{R}$ be a smooth upper (lower) approximation to the step function. Suppose further that ss_κ approximates the step function in the L_1-norm and ss_κ's derivative is uniformly continuous on its domain. Then spr_κ is a smooth lower (upper) approximation to pr.

As with the maximum and minimum, there are many ways to construct smooth approximations to the step function $\mathbb{1}(x \ge 0)$. Here we mention just one:

$$\mathrm{ss}^+_\kappa(x) = S_1(\kappa x - 1) \qquad \text{and} \qquad \mathrm{ss}^-_\kappa(x) = S_1(\kappa x) \tag{23}$$

where

$$S_1(x) = \begin{cases} 0 & \text{if } x \le 0 \\ 3x^2 - 2x^3 & \text{if } x \in (0,1) \\ 1 & \text{if } x \ge 1. \end{cases}$$

Lemma 14. Let $\kappa \in \mathbb{R}_+$. Consider ss^+_κ and ss^-_κ defined in equation (23).

1. $\mathrm{ss}^+_\kappa : \mathbb{R} \to \mathbb{R}$ is a smooth upper approximation of $\mathbb{1}(x \ge 0)$.
2. $\mathrm{ss}^-_\kappa : \mathbb{R} \to \mathbb{R}$ is a smooth lower approximation of $\mathbb{1}(x \ge 0)$.

Moreover, both ss^-_κ and ss^+_κ approximate the step function in the L_1-norm and have uniformly continuous derivative on $\mathbb{R}$.

We can now replace the pre-rearrangement operator in the numerator of equation (9) by $\text{spr}_\kappa(f)$ where the step function is approximated by ss_κ^-. Likewise we replace the pre-rearrangement operator in the denominator by $\text{spr}_\kappa(f)$ where the step function is approximated by ss_κ^+. Both of these changes decrease the value of equation (9).

Next consider the infimum piece in the denominator of equation (9). First notice that

$$\inf_{y\in\mathcal{Y}_z}(\overline{F}^c_{Y_1}(y)-\underline{F}^c_{Y_0}(y-z))=1+\inf_{y\in\mathcal{Y}_z}(-1+\overline{F}^c_{Y_1}(y)-\underline{F}^c_{Y_0}(y-z)).$$

This ensures the argument of the infimum is nonpositive, a property we use below. As Fang and Santos (2015) note in their example 2.3 on page 10, if this infimum always has a unique optimizer, then the infimum operator is actually ordinary Hadamard differentiable. To avoid assuming that such a unique optimizer always exists, however, we will also replace the infimum by a smooth approximation. Specifically, let

$$\|f\|_p=\left(\int_{\mathcal{Y}_z}|f(y)|^p\,dy\right)^{1/p}$$

denote the $L_p(\mathcal{Y}_z)$-norm. As $p\to\infty$, the L_p-norm converges to the sup-norm. We use this result to construct our smooth approximation to the infimum. Let $\underline{y}=\inf\mathcal{Y}_z$ and $\overline{y}=\sup\mathcal{Y}_z$, which are both finite since $\mathcal{Y}_z$ is compact. Let $\mathbb{D}_2$ denote the set of all nonpositive functions in $\ell^\infty(\mathbb{R}\times\mathcal{C})$ with L_p-norm bounded away from zero and range contained in $[-2,0]$. Let $p\geq 1$. Define $\psi_{2,p}:\mathbb{D}_2\to\ell^\infty(\mathcal{C})$ by

$$[\psi_{2,p}(f)](c)=-\frac{1}{(\overline{y}-\underline{y})^{1/p}}\|-f(\cdot,c)\|_p.$$

We scale the L_p-norm to ensure that we obtain a lower approximation to the supremum. The two minus signs then switch this to an upper approximation to the infimum.

Lemma 15. $\psi_{2,p}$ is a smooth upper approximation to the infimum function, as $p\to\infty$.

Using this result, we replace

$$1+\inf_{y\in\mathcal{Y}_z}(-1+\overline{F}^c_{Y_1}(y)-\underline{F}^c_{Y_0}(y-z))$$

by its smooth upper approximation

$$1+\left[\psi_{2,p}\left(-1+\overline{F}^{(\cdot)}_{Y_1}(\cdot)-\underline{F}^{(\cdot)}_{Y_0}(\cdot-z)\right)\right](c)$$

in equation (9), which decreases the value of the breakdown frontier. In this step we require the argument of $\psi_{2,p}$ to have nonzero L_p-norm. This assumption rules out extreme cases, such as when both $0\in\mathcal{C}$ and $F_{Y|X}(\cdot\mid 0)=F_{Y|X}(\cdot\mid 1)$.

Finally, the maximum in the definition of $\text{BF}(z,c,\underline{p})$ can be replaced with sm_κ. For the minimum in this definition we want a smooth *lower* approximation. Since $\min\{0,x\}=x[1-\mathbb{1}(x\geq 0)]$, one such approximation is $x[1-\text{ss}_\kappa^+(x)]$.

In section 3 we showed that the population breakdown frontier is a composition of Hadamard directionally differentiable functionals. In this section, we showed how to replace each functional in this composition by an ordinary Hadamard differentiable functional in such a way that the overall function is weakly smaller than the original breakdown frontier. Moreover, the difference between the original and smoothed frontiers can be made arbitrarily small by choosing appropriate values of

the tuning parameters. Corollary 2 above shows how to use this construction to do valid inference on the original breakdown frontier.

We conclude by comparing our two approaches to inference: The first based on Hadamard directional differentiability and the second based on smoothing the population frontier. For any fixed finite grid of c values, the first approach provides asymptotically exact inference while the second approach will always be possibly conservative. Visually, the first approach uses step functions to obtain a uniform band while the second approach produces a smoother appearing frontier. The first approach required assumption A5 while the second approach does not. The first approach requires choosing the ε_N tuning parameter for the numerical delta method bootstrap, while the second approach does not since the ordinary bootstrap is valid. The second approach, however, requires choosing a large number of smoothing functions and bandwidths, unlike the first approach. Too much smoothing will lead to conservative inference, while too little smoothing will likely lead to poor finite sample performance. Overall, neither approach appears to strictly dominate.

Proofs for appendix C

Proof of proposition 5. Equation (21) follows by the functional delta method (e.g., theorem 3.1 of Fang and Santos 2015), since ψ is Hadamard differentiable. The rest of the proof follows as in the proof of proposition 2. □

Proof of corollary 2. This result follows immediately by the lower envelope property of the smoothed breakdown frontier. □

Proof of lemma 12. We give the proof for part 1. Part 2 is analogous.

1. Let $x_1, \ldots, x_n \in \mathbb{R}$. $\text{sm}_\kappa\{x_1, \ldots, x_n\}$ is a weighted average of $x_1, \ldots, x_n$ where the weights are in $(0, 1)$. Hence it must always be weakly smaller than the maximum of $x_1, \ldots, x_n$. Thus $\text{sm}_\kappa\{f_1(y), \ldots, f_n(y)\} \leq \max\{f_1(y), \ldots, f_n(y)\}$ for any functions $f_1, \ldots, f_n$ and any $y \in \mathcal{D}$.

2. Let $x_1, \ldots, x_n \in \mathbb{R}$. Suppose $x_k = \max\{x_1, \ldots, x_n\}$. Without loss of generality, suppose this maximum is unique. Multiplying and dividing by $\exp(-\kappa x_k)$ yields

$$\text{sm}_\kappa\{x_1, \ldots, x_n\} = \sum_{i=1}^{n} x_i \frac{\exp(\kappa[x_i - x_k])}{\sum_{j=1}^{n} \exp(\kappa[x_j - x_k])}.$$

   For all $i \neq k$, $x_i - x_k < 0$ hence $\exp(\kappa[x_i - x_k]) \to 0$ as $\kappa \to \infty$. Thus the weights on all $i \neq k$ converge to zero while the weight on x_k converges to one. Hence for any fixed $f_1, \ldots, f_n$ and $y \in \mathcal{D}$, $[\psi_{1,\kappa}(f_1, \ldots, f_n)](y)$ converges to $\psi_{1,\max}(f_1, \ldots, f_n)](y)$.

3. For any fixed κ, the derivatives of the weights with respect to each x_i are uniformly bounded. This follows by the functional form of the weights and compactness of $\mathcal{Y}$. Therefore $\psi_{1,\kappa}$ is Fréchet differentiable by lemma 16 below. Finally, note that Fréchet differentiability implies Hadamard differentiability.

□

Lemma 16. Let $g : \mathbb{R}^n \to \mathbb{R}$ be an everywhere differentiable function with uniformly continuous derivative on $\mathcal{Y} \subseteq \mathbb{R}^n$. Let $\mathcal{D}$ be a subset of a Euclidean space. Define the functional $\phi : \ell^\infty(\mathcal{D})^n \to \ell^\infty(\mathcal{D})$ by

$$[\phi(f_1, \ldots, f_n)](y) = g(f_1(y), \ldots, f_n(y)).$$

Let $\mathbb{D}$ denote the set of functions in $\ell^\infty(\mathcal{D})^n$ with range contained in $\mathcal{Y}$. Let $f \in \mathbb{D}$. Then ϕ is Fréchet differentiable at f with derivative

$$[\phi'_f(h)](y) = \sum_{i=1}^{n} [\nabla_i g](f_1(y), \ldots, f_n(y)) \cdot h_i(y)$$

where $\nabla_i g$ denotes the ith partial derivative of g.

Proof of lemma 16. We show the $n = 1$ case, where

$$[\phi'_f(h)](y) = g'(f(y))h(y).$$

The $n \geq 2$ case is similar. We also suppose $\mathcal{Y} = \mathbb{R}$ for simplicity. We have

$$\begin{aligned}
\big|[\phi(f+h)](y) - \big([\phi(f)](y) + [\phi'_f(h)](y)\big)\big| &= \big|\big(g(f(y)+h(y)) - g(f(y))\big) - g'(f(y))h(y)\big| \\
&= |g'(\bar{f}(y))h(y) - g'(f(y))h(y)| \\
&= |g'(\bar{f}(y)) - g'(f(y))| \cdot |h(y)|.
\end{aligned}$$

The second line follows by the mean value theorem, which says that there exists a $\bar{f}(y)$ such that

$$g(f(y)+h(y)) - g(f(y)) = g'(\bar{f}(y))h(y)$$

and $|\bar{f}(y) - f(y)| \leq |h(y)|$. We apply this argument for each $y \in \mathcal{D}$.

Next, fix $\varepsilon > 0$. By uniform continuity of $g'(\cdot)$, there is a $\delta > 0$ such that for all $\tilde{h} \in \mathbb{R}$ with $|\tilde{h}| < \delta$ and all $\tilde{f} \in \mathbb{R}$,

$$|g'(\tilde{f} + \tilde{h}) - g'(\tilde{f})| < \varepsilon.$$

Therefore,

$$|g'(\bar{f}(y)) - g'(f(y))| \cdot |h(y)| \leq \varepsilon |h(y)|$$

for all $\|h\|_\infty < \delta$. Hence

$$\sup_{y \in \mathcal{D}} \big|[\phi(f+h)](y) - \big([\phi(f)](y) + [\phi'_f(h)](y)\big)\big| \leq \varepsilon \sup_{y \in \mathcal{D}} |h(y)|$$

for all $\|h\|_\infty < \delta$. That is,

$$\frac{\|\phi(f+h) - \big(\phi(f) + \phi'_f(h)\big)\|_\infty}{\|h\|_\infty} \leq \varepsilon$$

for $\|h\|_\infty < \delta$. Since ε was arbitrary, this shows that the left hand side is $o(1)$. □

Proof of lemma 13. We show that spr_κ is a smooth lower approximation to pr when ss_κ is a smooth upper approximation to the step function. The second part is analogous.

1. This follows immediately since ss_κ is an upper approximation to the step function, which is then multiplied by a negative sign in the definition of spr_κ.

2. This follows by our assumption that ss_κ approximates the step function in the L_1-norm:

$$\int_{\mathbb{R}} |\text{ss}_\kappa(u) - \mathbb{1}(u \geq 0)| \, du \to 0$$

as $\kappa \to \infty$.

3. This follows immediately from lemma 17 below.

□

Lemma 17. Let $\Lambda : \mathbb{R} \to \mathbb{R}$ be an everywhere differentiable function with uniformly continuous derivative on its domain. Let $\mathcal{C} \subseteq \mathbb{R}$. Define the functional $\phi : \ell^\infty((0,1) \times \mathcal{C}) \to \ell^\infty(\mathcal{C})$ by

$$[\phi(f)](c) = \int_0^1 \Lambda[f(u,c)]\, du.$$

Then ϕ is Fréchet differentiable, where

$$[\phi'_f(h)](c) = \int_0^1 \Lambda'[f(u,c)]h(u,c)\, du$$

is the Fréchet derivative of ϕ at f.

Proof of lemma 17. This result is a modification of example 5 on page 174 of Luenberger (1969). We have

$$\begin{aligned}\big|[\phi(f+h)](c) - \big([\phi(f)](c) + [\phi'_f(h)](c)\big)\big| \\ = \left|\int_0^1 \big(\Lambda[f(u,c)+h(u,c)] - \Lambda[f(u,c)] - \Lambda'[f(u,c)]h(u,c)\big)\, du\right|\end{aligned}$$

By the usual mean value theorem,

$$\Lambda[f(u,c)+h(u,c)] - \Lambda[f(u,c)] = \Lambda'[\bar{f}(u,c)]h(u,c)$$

where $|f(u,c) - \bar{f}(u,c)| \leq |h(u,c)|$. We apply this argument for each $u \in (0,1)$ and $c \in \mathcal{C}$.

Next, fix $\varepsilon > 0$. $\Lambda'(\cdot)$ is uniformly continuous by assumption. Hence there is a $\delta > 0$ such that for all $\tilde{h} \in \mathbb{R}$ with $|\tilde{h}| < \delta$ and all $\tilde{f} \in \mathbb{R}$,

$$|\Lambda'(\tilde{f} + \tilde{h}) - \Lambda'(\tilde{f})| < \varepsilon.$$

Therefore,

$$\begin{aligned}&\sup_{c\in\mathcal{C}} \left|\int_0^1 \big(\Lambda[f(u,c)+h(u,c)] - \Lambda[f(u,c)] + \Lambda'[f(u,c)]h(u,c)\big)\, du\right| \\ =\ &\sup_{c\in\mathcal{C}} \left|\int_0^1 \Big(\Lambda'[\bar{f}(u,c)] - \Lambda'[f(u,c)]\Big)h(u,c)\, du\right| \\ \leq\ &\sup_{c\in\mathcal{C}} \left|\int_0^1 \varepsilon h(u,c)\, du\right| \\ \leq\ &\varepsilon \sup_{c\in\mathcal{C}} \sup_{u\in(0,1)} |h(u,c)| \\ =\ &\varepsilon \|h\|_\infty.\end{aligned}$$

The first inequality holds for $\|h\|_\infty < \delta$. Thus

$$\frac{\|\phi(f+h) - \big(\phi(f) + \phi'_f(h)\big)\|_\infty}{\|h\|_\infty} \leq \varepsilon$$

for $\|h\|_\infty < \delta$. Since ε was arbitrary, this shows that the left hand side is $o(1)$. □

Proof of lemma 14. We give the proof for part 1. Part 2 is analogous.

1. By construction, $0 \leq S_1(x) \leq 1$ on $(0, 1)$.

2. By construction, $\text{ss}_\kappa^+(x)$ equals $\mathbb{1}(x \geq 0)$ everywhere except on $(0, 1/\kappa)$. Thus we immediately obtain the desired pointwise convergence. Furthermore, note that

$$\int_{\mathbb{R}} |\text{ss}_\kappa^+(x) - \mathbb{1}(x \geq 0)| \, dx \leq \frac{1}{\kappa}$$

   which converges to zero as $\kappa \to \infty$. We use this property in lemma 13.

3. This follows since $S_1 : \mathbb{R} \to \mathbb{R}$ is differentiable.

$S_1'(x) = 6x(1-x)$ is bounded by 3 on $x \in [0, 1]$. For $x \notin [0, 1]$, $S_1'(x) = 0$. Hence S_1 is Lipschitz. Therefore it is uniformly continuous. Thus both ss_κ^+ and ss_κ^- are also uniformly continuous. □

Proof of lemma 15.

1. This follows from

$$\begin{aligned}\left(\int_{\mathcal{Y}_z} f(y,c)^p \, dx\right)^{1/p} &\leq \sup_{y \in \mathcal{Y}_z} f(y,c) \left(\int_{\mathcal{Y}_z} 1 \, dx\right)^{1/p} \\ &\leq \|f(\cdot,c)\|_\infty (\overline{y} - \underline{y})^{1/p}.\end{aligned}$$

2. This follows, for example, from proposition 2.2 on page 8 of Stein and Shakarchi (2011).

3. Define $\phi : \mathbb{D}_2 \to \ell^\infty(\mathcal{C})$ by $[\phi(f)](c) = \| - f(\cdot, c)\|_p$. $\psi_{2,p}$ is just a scaled version of this functional. Define $\psi : \mathbb{D}_2 \to \ell^\infty(\mathcal{C})$ by

$$[\psi(f)](c) = [\phi(f)](c)^p = \| - f(\cdot, c)\|_p^p = \int_0^1 [-f(u,c)]^p \, du.$$

   The last equality follows since $-f$ is nonnegative. By lemma 17, ψ is Fréchet differentiable with Fréchet derivative

$$[\psi_f'(h)](c) = -\int_0^1 p[-f(u,c)]^{p-1} h(u,c) \, du.$$

   Here we use uniform continuity of x^p on the compact set $[-2, 0]$.

   Let $\mathbb{D}_3$ denote the set of nonnegative functions in $\ell^\infty(\mathcal{C})$ with L_p-norm bounded away from zero. Note that the range of ψ is contained in $\mathbb{D}_3$. Consider the functional $\theta : \mathbb{D}_3 \to \ell^\infty(\mathcal{C})$ defined by $[\theta(g)](c) = g(c)^{1/p}$. By arguments as in lemma 16, the Fréchet derivative of θ is

$$[\theta_g'(h)](c) = \frac{1}{p} g(c)^{1/p-1} h(c).$$

   Here we will use both the bounded range of our input functions f and their L_p-norm bounded away from zero to ensure uniform continuity of $(1/p)x^{1/p-1}$. Note that $[\phi(f)](c) = [\theta(\psi(f))](c)$.

The result now follows by the chain rule, which further states that $[\phi'_f(h)](c) = [\theta'_{\psi(f)}(\psi'_f(h))](c)$. Hence

$$\begin{aligned}[\phi'_f(h)](c) &= \frac{1}{p}\left(\int_0^1 [-f(u,c)]^p \, du\right)^{1/p-1} (-1)\left(\int_0^1 p[-f(u,c)]^{p-1}h(u,c)\, du\right) \\ &= -\frac{\| - f(\cdot,c)\|_p}{\int_0^1 [-f(u,c)]^p \, du}\int_0^1 [-f(u,c)]^{p-1}h(u,c)\, du.\end{aligned}$$

This derivative is not defined when f has zero L_p-norm.

□