Supplement to "Inference on breakdown frontiers"

(*Quantitative Economics*, Vol. 11, No. 1, January 2020, 41–111)

Matthew A. Masten
Department of Economics, Duke University

Alexandre Poirier
Department of Economics, Georgetown University

This supplemental Appendix provides an extended comparison of our results with the sensitivity analysis inference literature, a discussion of higher dimensional breakdown frontiers, Monte Carlo simulations for our proposed estimation and inference procedures, an alternative inference approach based on population smoothing, and additional empirical analyses.

Appendix D: 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 compared and contrasted 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, Robins, and Scharfstein (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 p.

Matthew A. Masten: matt.masten@duke.edu
Alexandre Poirier: alexandre.poirier@georgetown.edu

 https://doi.org/10.3982/QE1288

562). They emphasize that using uniform confidence bands is important since the functions $\theta(r)$ are often nonmonotonic, as we discussed in Appendix A with respect to Imbens (2003). A similar method is proposed by Rotnitzky, Robins, and Scharfstein (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

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

See their Figure 2. Unlike Todem, Fine, and Peng (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) considered 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) proposed 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 began 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 p. 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

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

Let

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

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

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

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

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

and

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

Then

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

PROOF OF PROPOSITION S.1. We have

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

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 S.1 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 (2018)), 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 S.1 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 S.2. *Assume that the cdf of the asymptotic distribution of* $\sqrt{N}[\widehat{\Theta}_L(r^*) - \Theta_L(r^*)]$ *is continuous and strictly increasing at its* $1-\alpha$ *quantile. Then*

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

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

PROOF OF PROPOSITION S.2. We have

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

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

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

If $\text{LB}(\cdot)$ 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 $\text{LB}(r^*) \leq \text{LB}(r_L)$ since $\text{LB}(\cdot)$ is weakly decreasing. But now notice that $\text{LB}(r_L) \leq \underline{\theta}$ by definition of r_L. Hence $\text{LB}(r^*) \leq \underline{\theta}$.

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 S.2 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

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

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

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

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) considered local-to-full-independence asymptotic distributions of MLEs. Conley, Hansen, and Rossi (2012) derived 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) generalized this local-to-zero asymptotic result to the GMM estimator for a given system of moment equalities. Unlike this literature, our approach is global: Breakdown frontier analysis focuses on the largest relaxations of 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, Rotnitzky, and Scharfstein (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, Hansen, and Rossi (2012) applied 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 nonstandard Bayesian analysis. Giacomini, Kitagawa, and Volpicella (2016) studied 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, Rotnitzky, and Scharfstein (2000) and Giacomini, Kitagawa, and Volpicella (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, Kitagawa, and Volpicella (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) provided 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.

Appendix E: Higher dimensional breakdown frontiers

In this paper we focus on breakdown frontiers where the assumption space—the domain of the sensitivity parameters that index relaxations—is two dimensional. In two dimensions, the breakdown frontier can be viewed as a function on a one-dimensional domain. Hence presenting estimated breakdown frontiers and corresponding confidence bands is conceptually straightforward. As in other settings, however, summarizing nonparametric functions of two or more parameters is difficult. In this section, we briefly discuss four ways to summarize higher dimensional breakdown frontiers.

1. *Compute the size of the robust region.* We call the area under the breakdown frontier the robust region. The area of this region provides a quantitative measure of the

robustness of the conclusion of interest. This area is a scalar statistic that can be computed regardless of the dimension of the space of assumptions. Although this statistic does not capture the trade-off between assumptions' identifying power, it can be used to compare the overall robustness of conclusions across different studies.

2. *Compute directional breakdown points.* Let the sensitivity parameter r be in $[0,1]^{\dim(r)}$ for some integer $\dim(r) \geq 2$. One can focus on values $r = m \cdot d$ where $m \in \mathbb{R}_+$ is a scalar and $d \in \mathbb{R}_+^{\dim(r)}$ is a known vector. Given the direction d, the largest possible value of m such that $m \cdot d \in [0,1]^{\dim(r)}$ is

$$\overline{m} = \sup\big\{m \in \mathbb{R}_+ : m \cdot d_\ell \leq 1 \text{ for all } \ell = 1, \ldots, \dim(r)\big\}.$$

Suppose we are interested in the conclusion that $\theta \in \mathcal{C}$ for some pre-specified set $\mathcal{C} \subseteq \Theta$. Then the breakdown point in the direction d is

$$m^* = \sup\big\{m \in [0, \overline{m}] : \Theta_I(m \cdot d) \subseteq \mathcal{C}\big\}.$$

This is the largest we can relax the assumptions in the direction d while still being able to conclude that $\theta \in \mathcal{C}$. The point $r^* = m^* \cdot d$ is on the breakdown frontier for this conclusion.

3. *Use parametric shapes to construct confidence sets.* Characterizing the breakdown frontier is equivalent to characterizing the robust region. In this paper, we study a setting where we construct inner confidence sets for the robust region. We presented these inner confidence sets graphically. It is well known, however, that summarizing higher dimensional confidence sets is difficult. A common solution is to use rectangular confidence regions. These regions can then be summarized by a set of intervals. Here we could similarly construct rectangular inner confidence sets for the robust region. One downside of using rectangles is that they may not well approximate any curvature in the shape of the breakdown frontier. To capture such curvature, while retaining the ability to summarize a higher dimensional set, we could instead construct ellipsoidal inner confidence sets.

4. *Compute average derivatives.* Suppose the space of assumptions is three dimensional. Let $r = (r_1, r_2, r_3)$ denote the sensitivity parameters. Suppose we are interested in the trade off between r_1 and r_2. Fix the value of r_3 and compute the breakdown frontier for the conclusion that $\theta \in \mathcal{C}$ as

$$\mathrm{BF}(r_2 \mid r_3) = \sup\big\{r_1 \in [0,1] : \Theta_I(r_1, r_2, r_3) \subseteq \mathcal{C}\big\}.$$

Holding r_3 fixed, this is a function from $[0,1]$ to $[0,1]$. Suppose that for each $r_3 \in [0,1]$, it is differentiable almost everywhere with respect to the Lebesgue measure. Then we can compute its average derivative

$$\int_0^1 \mathrm{BF}'(r_2 \mid r_3)\, \omega_2(r_2)\, dr_2,$$

where $\omega_2(\cdot)$ is some fixed weight function. We could present this average derivative as a function of r_3, or we could also average over r_3. This approach allows us to summarize

the average rate of substitution between the first two assumptions. If the function $\text{BF}(\cdot \mid r_3)$ is not differentiable, we could instead use secant functions to summarize the trade-offs.

Here we have just briefly discussed each method. We leave a full analysis and implementation of these to future work.

Appendix F: Monte Carlo simulations

In this section we study the finite sample performance of our estimation and inference procedures proposed in Section 3. We consider the following dgp. For $x \in \{0, 1\}$, $Y \mid X = x$ 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 $\pi = 1$. We set $\mathbb{P}(X = 1) = 0.5$. This dgp implies a joint distribution of (Y, X), from which we draw independently.

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.

First we consider the performance of our point estimator of the breakdown frontier. Figure S.1 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

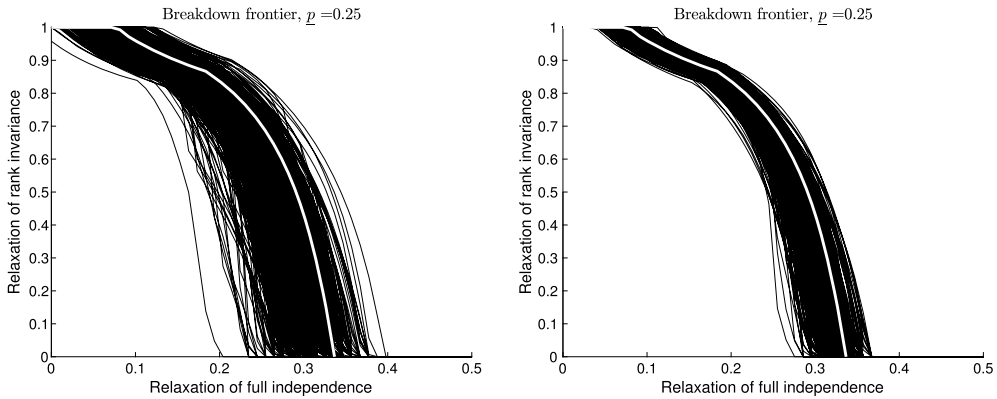


Figure S.1. Left: $N = 500$. Right: $N = 2000$. These plots show the sampling distribution of our breakdown frontier estimator by gathering the point estimates of the breakdown frontier across all Monte Carlo simulations into one plot. The true breakdown frontier is shown on top in white.

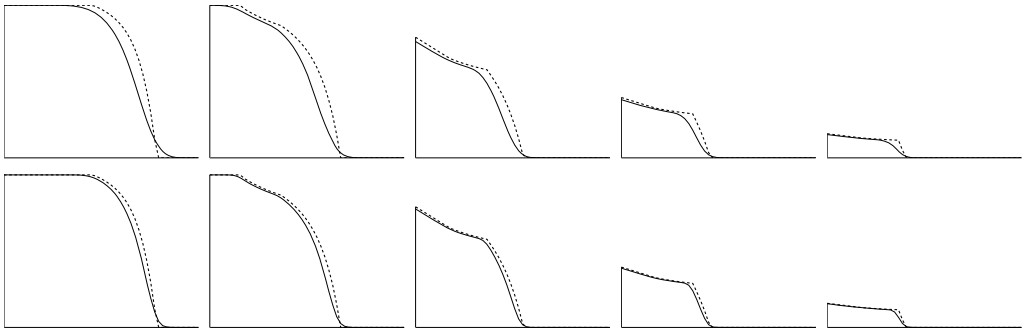


FIGURE S.2. Rows are sample sizes (top is $N = 500$, bottom is $N = 2000$). Columns are five values of $\underline{p}$ (from left to right: $\underline{p} = 0.1$, 0.25, 0.5, 0.75, and 0.9). Dotted lines are the true breakdown frontiers. The solid lines are the Monte Carlo estimates of $\mathbb{E}[\widehat{\text{BF}}(c)]$. This plot shows the finite sample bias of our breakdown frontier estimator.

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 S.2 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 S.1. 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 S.3 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 S.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

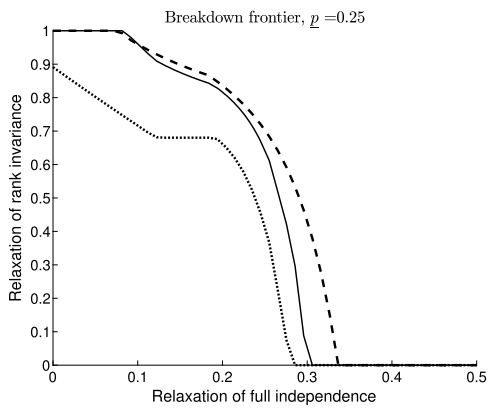


FIGURE S.3. $N = 500$. Example 95% lower uniform confidence band (dotted line), estimated breakdown frontier (solid line), true breakdown frontier (dashed line).

TABLE S.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

Note: 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.

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 (2018) imposed 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 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 overcoverage. Indeed, for $\underline{p} = 0.1$, 0.25, and 0.5, all choices of ε_N's considered lead to overcoverage. For the two larger values of $\underline{p}$, some values of ε_N lead to undercoverage. 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 S.2. While Table S.1 showed coverage probabilities, Table S.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

TABLE S.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

Note: 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.

confidence bands are all below the estimated breakdown frontier. Hence this ratio can never be larger than one. Although we do not perform a 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 S.1, twice the naive choice of ε_N yields the closest to nominal coverage. All other choices gave overcoverage. We see this in Table S.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 S.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 undercover. Larger ε_N's give smaller areas, but overcover. For large enough ε_N, the confidence bands get close to zero everywhere, and hence have very small area and 100% coverage. The results for $N=2000$ are similar.

In Table S.1, we saw that most combinations of $\underline{p}$ and ε_N led to overcoverage. This is caused by a downward bias in our estimated breakdown frontiers, as shown in Figure S.2. 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) studied a related problem. Constructing such bias corrected estimators of nondifferentiable functionals, however, is delicate due to the results of Doss and Sethuraman (1989, Theorem 1) and Hirano and Porter (2012, Theorem 2(a)). They show that achieving asymptotic unbiasedness for estimators of nondifferentiable functionals generally requires the variance to diverge. This result motivates consideration of alternative criteria, like the half-median unbiasedness property used by Chernozhukov, Lee, and Rosen (2013). We leave a full analysis of such corrections to future work.

Appendix G: Inference via population smoothing

In this section we develop an alternative approach to constructing lower uniform confidence bands for the breakdown frontier. For simplicity, we omit covariates. As discussed in Section 3, the population breakdown frontier $\text{BF}(\cdot, \underline{p})$ 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), Hong and Li (2018), and Fang and Santos (2019) to do inference. In this section, we instead replace $\text{BF}(\cdot, \underline{p})$ 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. For simplicity, we omit covariates throughout this section.

Specifically, recall from Section 2 that

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

where

$$\text{bf}(c, \underline{p}) = \frac{1 - \underline{p} - \mathbb{P}\big(\underline{Q}^c_{Y_1}(U) - \overline{Q}^c_{Y_0}(U) \leq 0\big)}{1 + \min\Big\{\inf\limits_{y \in \mathcal{Y}_0}\big(\overline{F}^c_{Y_1}(y) - \underline{F}^c_{Y_0}(y-0)\big), 0\Big\} - \mathbb{P}\big(\underline{Q}^c_{Y_1}(U) - \overline{Q}^c_{Y_0}(U) \leq 0\big)}. \tag{10}$$

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

Definition S.1. 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 $\mathrm{SBF}_\kappa(\cdot, \underline{p})$ such that this functional is a smooth lower approximation of the functional mapping θ_0 to $\mathrm{BF}(\cdot, \underline{p})$. Given such a functional, we estimate the smoothed breakdown frontier by sample analog:

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

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

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

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 S.3. *Suppose Assumptions* A1, A3, *and* A5 *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*

$$\sqrt{N}\big(\psi(\widehat{\theta}^*) - \psi(\widehat{\theta})\big) \overset{P}{\rightsquigarrow} \psi'_{\theta_0}(\mathbf{Z}_1), \tag{S.1}$$

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

$$\widehat{z}_{1-\alpha} = \inf\left\{ z \in \mathbb{R} : \mathbb{P}\left(\sup_{c \in [0,\overline{C}]} \frac{\sqrt{N}\big(\big[\psi(\widehat{\theta}^*)\big](c) - \big[\psi(\widehat{\theta})\big](c)\big)}{\sigma(c)} \leq z \mid Z^N \right) \geq 1 - \alpha \right\}. \tag{S.2}$$

Finally, suppose also that the cdf of

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

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 1. *Suppose the assumptions of Proposition S.3 hold. Let $\widehat{k}(c) = \widehat{z}_{1-\alpha}\sigma(c)$, where $\widehat{z}_{1-\alpha}$ is defined in equation (S.2). Then*

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

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, Fernández-Val, and Galichon (2010) in their Corollary 4. They suggested replacing the nonsmooth 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 $\text{SBF}_\kappa(c, \underline{p})$. We consider each piece composing the function $\text{BF}(c, \underline{p})$ in turn. First consider $\overline{F}^c_{Y_1}(y)$. This bound is a minimum of two terms (see equation (4)). 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 in using the functional $\psi_{1,\kappa} : \mathbb{D}_1 \to \ell^\infty(\mathcal{D})$ defined by

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

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

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

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

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

LEMMA S.1. *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 (10), 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 (10). 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 (10).

Next consider the $\mathbb{P}(\underline{Q}^c_{Y_1}(U) - \overline{Q}^c_{Y_0}(U) \leq 0)$ 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

$$\big[\mathrm{pr}(f)\big](c) = \int_0^1 \mathbb{1}\big[f(u,c) \leq 0\big]\, du$$

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

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

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

Lemma S.2. *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 $\mathbb{1}(x \geq 0)$. Here we mention just one:

$$\mathrm{ss}^+_\kappa(x) = S_1(\kappa x - 1) \quad \text{and} \quad \mathrm{ss}^-_\kappa(x) = S_1(\kappa x), \tag{3}$$

where

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

Lemma S.3. *Let* $\kappa \in \mathbb{R}_+$*. Consider* ss^+_κ *and* ss^-_κ *defined in equation* (3).

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

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 (10) by $\mathrm{spr}_\kappa(f)$ where the step function is approximated by ss^-_κ. Likewise we replace the pre-rearrangement operator in the denominator by $\mathrm{spr}_\kappa(f)$ where the step function is approximated by ss^+_κ. Both of these changes decrease the value of equation (10).

Next consider the infimum piece in the denominator of equation (10). First notice that

$$\inf_{y \in \mathcal{Y}_0} \big(\overline{F}^c_{Y_1}(y) - \underline{F}^c_{Y_0}(y-0)\big) = 1 + \inf_{y \in \mathcal{Y}_0} \big(-1 + \overline{F}^c_{Y_1}(y) - \underline{F}^c_{Y_0}(y-0)\big).$$

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

$$\|f\|_p = \left(\int_{\mathcal{Y}_0} |f(y)|^p dy \right)^{1/p}$$

denote the $L_p(\mathcal{Y}_0)$-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 $\underline{y} = \inf \mathcal{Y}_0$ and $\overline{y} = \sup \mathcal{Y}_0$, which are both finite since $\mathcal{Y}_0$ 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

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

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 S.4. *$\psi_{2,p}$ is a smooth upper approximation to the infimum function, as $p \to \infty$.*

Using this result, we replace

$$1 + \inf_{y \in \mathcal{Y}_0} \big(-1 + \overline{F}^c_{Y_1}(y) - \underline{F}^c_{Y_0}(y - 0) \big)$$

by its smooth upper approximation

$$1 + \big[\psi_{2,p}\big(-1 + \overline{F}^{(\cdot)}_{Y_1}(\cdot) - \underline{F}^{(\cdot)}_{Y_0}(\cdot - 0) \big)\big](c)$$

in equation (10), 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 $\text{BF}(c, \underline{p})$ 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 1 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 A6 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 G

PROOF OF PROPOSITION S.3. Equation (S.1) follows by the functional delta method (e.g., Theorem 3.1 of Fang and Santos (2019)), since ψ is Hadamard differentiable. The rest of the proof follows as in the proof of Proposition 2. □

PROOF OF COROLLARY 1. This result follows immediately by the lower envelope property of the smoothed breakdown frontier. □

PROOF OF LEMMA S.1. 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

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

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 S.5 below. Finally, note that Fréchet differentiability implies Hadamard differentiability. □

LEMMA S.5. *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*

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

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

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

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

PROOF OF LEMMA S.5. We show the $n = 1$ case, where

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

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

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

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

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

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 $\tilde{f} \in \mathbb{R}$,

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

Therefore,

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

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

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

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

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

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

PROOF OF LEMMA S.2. 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:

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

as $\kappa \to \infty$.

3. This follows immediately from Lemma S.6 below. □

LEMMA S.6. *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*

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

Then ϕ is Fréchet differentiable, where

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

is the Fréchet derivative of ϕ at f.

PROOF OF LEMMA S.6. This result is a modification of example 5 on page 174 of Luenberger (1969). We have

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

By the usual mean value theorem,

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

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 $\tilde{f} \in \mathbb{R}$,

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

Therefore,

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

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

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

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

Proof of Lemma S.3. We give the proof for part 1. Part 2 is analogous.

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

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

$$\int_{\mathbb{R}}\big|\mathrm{ss}^+_\kappa(x)-\mathbb{1}(x\ge 0)\big|\,dx\le\frac{1}{\kappa}$$

which converges to zero as $\kappa\to\infty$. We use this property in Lemma S.2.

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 S.4. 1. This follows from

$$\left(\int_{\mathcal{Y}_0} f(y,c)^p\,dx\right)^{1/p}\le\sup_{y\in\mathcal{Y}_0}f(y,c)\left(\int_{\mathcal{Y}_0}1\,dx\right)^{1/p}\le\big\|f(\cdot,c)\big\|_\infty(\overline{y}-\underline{y})^{1/p}.$$

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

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

The last equality follows since $-f$ is nonnegative. By Lemma S.6, ψ is Fréchet differentiable with Fréchet derivative

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

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 S.5, the Fréchet derivative of θ is

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

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

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

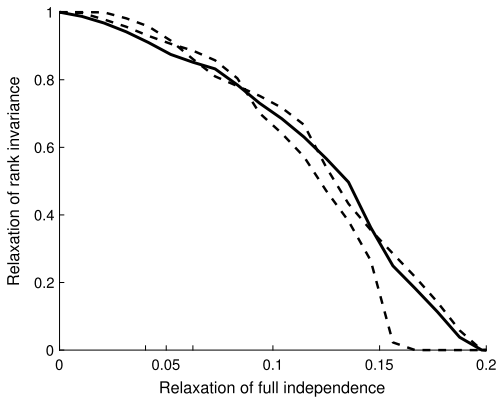
This derivative is not defined when f has zero L_p-norm. □

Appendix H: Additional empirical analyses

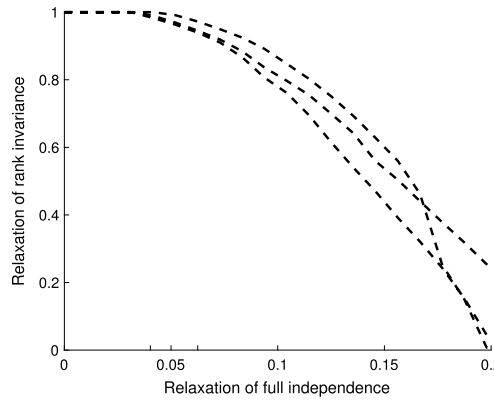
In our empirical analysis of Section 4 we collapse our two discrete covariates into two binary indicators of whether one is above or below the median value of the covariate. In this section we explore the impact of alternative coarsenings of these covariates on our empirical results. Specifically, we consider six different coarsenings total: three binary coarsenings and three trinary coarsenings.

For the binary coarsenings, we collapse each covariate W_k into a binary variable based on whether the value of the covariate is above or below the τth quantile $Q_{W_k}(\tau)$, for $\tau \in \{0.25, 0.5, 0.75\}$. This includes our baseline case, $\tau = 0.5$. For the trinary coarsenings, we collapse each covariate W_k into three bins: below $Q_{W_k}(\tau_1)$, between $Q_{W_k}(\tau_1)$ and $Q_{W_k}(\tau_2)$, and above $Q_{W_k}(\tau_2)$, where $Q_{W_k}(\tau)$ is the τth quantile of the covariate W_k. We use three choices of (τ_1, τ_2): $(0.35, 0.65)$, $(0.30, 0.70)$, and $(0.25, 0.75)$. As discussed in Section 4, given our overall sample size of 448 observations, using finer coarsenings (that is, more cells) or more asymmetric choices of τ yields cells with too few observations.

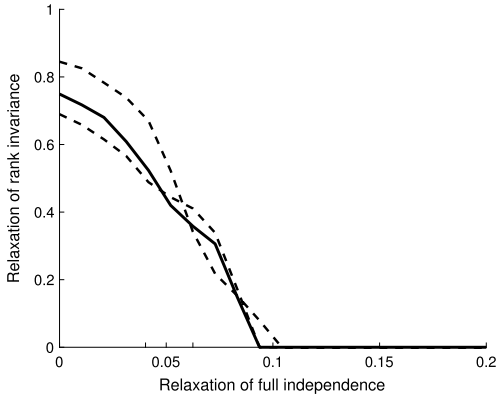
Figure S.4 shows the estimated breakdown frontiers for each of the six different coarsenings, and for $\underline{p} \in \{0.1, 0.25, 0.5\}$. The solid line is our baseline estimate—the binary coarsening split at the median. The dotted lines show our alternative coarsenings.



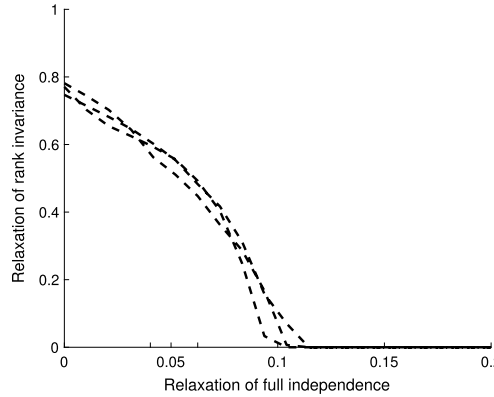
(a) $\underline{p} = 0.1$. Binary coarsenings.



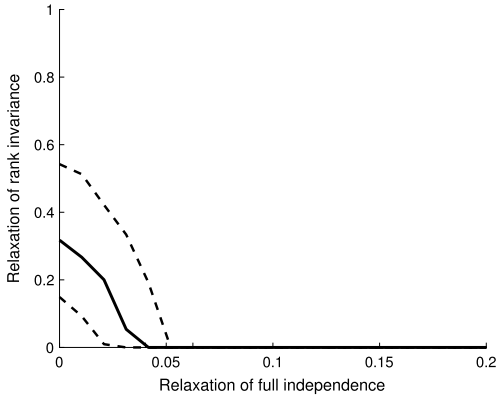
(b) $\underline{p} = 0.1$. Trinary coarsenings.



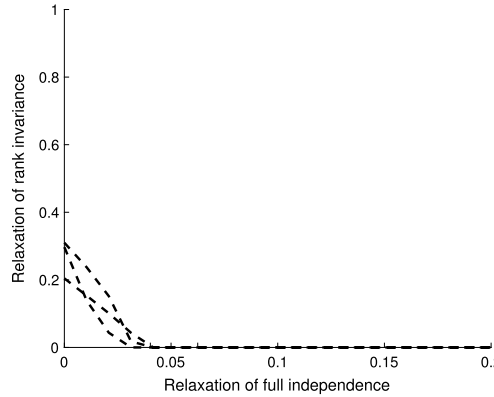
(c) $\underline{p} = 0.25$. Binary coarsenings.



(d) $\underline{p} = 0.25$. Trinary coarsenings.



(e) $\underline{p} = 0.5$. Binary coarsenings.



(f) $\underline{p} = 0.5$. Trinary coarsenings.

FIGURE S.4. Estimated breakdown frontiers using various coarsenings of the conditioning variables. The solid bold line is our baseline estimate. The dotted lines represent the five alternative coarsenings we consider.

Comparing the dotted lines with the solid line, we see that the estimated breakdown frontier generally does not change much as we vary how we coarsen the covariates. This is especially true for $\underline{p} = 0.1$ and $\underline{p} = 0.25$. Thus our overall conclusions regarding the robustness of claims about $\mathbb{P}(Y_1 > Y_0)$ do not appear to depend strongly on the specific choice of coarsening.

References

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. [6]

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

Chernozhukov, V., I. Fernández-Val, and A. Galichon (2010), "Quantile and probability curves without crossing." *Econometrica*, 78, 1093–1125. [15]

Chernozhukov, V., S. Lee, and A. M. Rosen (2013), "Intersection bounds: Estimation and inference." *Econometrica*, 81, 667–737. [13]

Conley, T. G., C. B. Hansen, and P. E. Rossi (2012), "Plausibly exogenous." *Review of Economics and Statistics*, 94, 260–272. [6]

Copas, J. and S. Eguchi (2001), "Local sensitivity approximations for selectivity bias." *Journal of the Royal Statistical Society, Series B*, 63, 871–895. [6]

DiTraglia, F. and C. García-Jimeno (2016), "A framework for eliciting, incorporating, and disciplining identification beliefs in linear models." Working paper. [7]

Doss, H. and J. Sethuraman (1989), "The price of bias reduction when there is no unbiased estimate." *The Annals of Statistics*, 17, 440–442. [13]

Dümbgen, L. (1993), "On nondifferentiable functions and the bootstrap." *Probability Theory and Related Fields*, 95, 125–140. [13]

Escanciano, J. C. and L. Zhu (2013), "Set inferences and sensitivity analysis in semiparametric conditionally identified models." Working paper. [2]

Fan, Y. and S. S. Park (2009), "Partial identification of the distribution of treatment effects and its confidence sets." In *Nonparametric Econometric Methods: Advances in Econometrics*, Vol. 25, 3–70, Emerald Group Publishing Limited. [12]

Fang, Z. and A. Santos (2015), "Inference on directionally differentiable functions." Working paper. [17]

Fang, Z. and A. Santos (2019), "Inference on directionally differentiable functions." *The Review of Economic Studies*, 86, 377–412. [13, 18]

Freyberger, J. and Y. Rai (2018), "Uniform confidence bands: Characterization and optimality." *Journal of Econometrics*, 204, 119–130. [4]

Giacomini, R., T. Kitagawa, and A. Volpicella (2016), "Uncertain identification." Working paper. [7]

Hirano, K. and J. R. Porter (2012), "Impossibility results for nondifferentiable functionals." *Econometrica*, 80, 1769–1790. [13]

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999), "Bayesian model averaging: A tutorial." *Statistical Science*, 14, 382–401. [6]

Hong, H. and J. Li (2018), "The numerical delta method." *Journal of Econometrics*, 206, 379–394. [11, 13]

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

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

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. [2, 4]

Koop, G. and D. J. Poirier (1997), "Learning about the across-regime correlation in switching regression models." *Journal of Econometrics*, 78, 217–227. [7]

Leamer, E. E. (1978), *Specification Searches: Ad hoc Inference With Nonexperimental Data*. John Wiley & Sons Incorporated. [6]

Lindley, D. V. (1972), *Bayesian Statistics: A Review*. SIAM. [7]

Luenberger, D. G. (1969), *Optimization by Vector Space Methods*. John Wiley & Sons. [20]

Masten, M. A. and A. Poirier (2016), "Partial independence in nonseparable models." cemmap Working Paper, CWP26/16. [1]

Moon, H. R. and F. Schorfheide (2012), "Bayesian and frequentist inference in partially identified models." *Econometrica*, 80, 755–782. [7]

Pitman, E. J. G. (1949), "Notes on non-parametric statistical inference." Unpublished lecture notes, Columbia University. [6]

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. [2]

Robins, J. M. (1999), "Association, causation, and marginal structural models." *Synthese*, 121, 151–179. [1]

Robins, J. M., A. Rotnitzky, and D. O. Scharfstein (2000), "Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models." In *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, 1–94, Springer. [6, 7]

Rosenbaum, P. R. (1995), *Observational Studies*. Springer. [2]

Rosenbaum, P. R. (2002), *Observational Studies*, second edition. Springer. [2]

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. [1, 2]

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. [2]

Stein, E. M. and R. Shakarchi (2011), *Functional Analysis: An Introduction to Further Topics in Analysis*, Vol. 4. Princeton University Press. [21]

Todem, D., J. Fine, and L. Peng (2010), "A global sensitivity test for evaluating statistical hypotheses with nonidentifiable models." *Biometrics*, 66, 558–566. [1, 2]

Vansteelandt, S., E. Goetghebeur, M. G. Kenward, and G. Molenberghs (2006), "Ignorance and uncertainty regions as inferential tools in a sensitivity analysis." *Statistica Sinica*, 16, 953–979. [1, 2]

Co-editor Christopher Taber handled this manuscript.

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