Sensitivity Analysis via the Proportion of Unmeasured Confounding

Matteo Bonvini and Edward H. Kennedy
Department of Statistics & Data Science, Carnegie Mellon University, Pittsburgh, PA

ABSTRACT
In observational studies, identification of ATEs is generally achieved by assuming that the correct set of confounders has been measured and properly included in the relevant models. Because this assumption is both strong and untestable, a sensitivity analysis should be performed. Common approaches include modeling the bias directly or varying the propensity scores to probe the effects of a potential unmeasured confounder. In this article, we take a novel approach whereby the sensitivity parameter is the “proportion of unmeasured confounding”: the proportion of units for whom the treatment is not as good as randomized even after conditioning on the observed covariates. We consider different assumptions on the probability of a unit being unconfounded. In each case, we derive sharp bounds on the average treatment effect as a function of the sensitivity parameter and propose nonparametric estimators that allow flexible covariate adjustment. We also introduce a one-number summary of a study’s robustness to the number of confounded units. Finally, we explore finite-sample properties via simulation, and apply the methods to an observational database used to assess the effects of right heart catheterization. Supplementary materials for this article are available online.

1. Introduction
In an experiment, the random assignment of the treatment to the units ensures that any measured and unmeasured factors are balanced between the treatment and control groups, thereby allowing the researcher to attribute any observed effect to the treatment. In observational studies, however, achieving such balance requires the untestable assumption that all confounders, roughly variables affecting both the treatment A and the outcome Y, are collected. To gauge the consequences of departures from the no-unmeasured-confounding assumption, a sensitivity analysis generally posits the existence of an unmeasured confounder U and varies either the U-A association or the U-Y association or both. The minimal strength of these associations that would drive the observed Y-A association to zero is often reported as a measure of the study’s robustness to unmeasured confounding.

Since the seminal work of Cornfield et al. (1959) on the association between smoking and lung cancer, a plethora of sensitivity analysis frameworks have been proposed. Here, we mention a few of them and refer to Liu, Kuramoto, and Stuart (2013) and Richardson et al. (2014) for excellent reviews. In the context of matched studies, Rosenbaum’s framework (Rosenbaum 1987, 2002) is likely the most commonly used. It governs the U-A association via a parameter \( \Gamma \geq 1 \) by requiring that, within each pair, the ratio of the odds that unit 1 is treated to the odds that unit 2 is treated falls in the interval \( [\Gamma^{-1}, \Gamma] \). The U-Y association is often left unrestricted or bounded as in Gastwirth, Krieger, and Rosenbaum (1998). More recently, Zhao, Small, and Bhattacharya (2017) and Yadlowsky et al. (2018) have proposed extensions to this framework that do not require matching.

In addition, Ding and VanderWeele (2016) and VanderWeele and Ding (2017) have derived a bounding factor for certain treatment effects in terms of two sensitivity parameters governing the U-A and U-Y relationships. Other authors have proposed modeling the distribution of U and the relationships U-Y and U-A directly (Rosenbaum and Rubin 1983; Imbens 2003), which has been recently extended to the case where the distribution of U is left unspecified by Zhang and Tchetgen Tchetgen (2019). In the context of time-varying treatments, sensitivity analyses have been proposed for marginal structural models (Brumback et al. 2004) and cause-specific selection models (Rotnitzky et al. 2001).

In this article, we propose a novel approach to sensitivity analysis based on a mixture model for confounding. We conceptualize that an unknown fraction \( \epsilon \) of the units in the sample is arbitrarily confounded while the rest is not. The parameter \( \epsilon \) is unknown and not estimable but can be varied as a sensitivity parameter. As discussed below, our model generalizes some relaxations to the no-unmeasured-confounding assumption that have been previously proposed in the literature. Furthermore, our framework yields a natural one-number summary of a study’s robustness: the minimum proportion of confounded units such that bounds on the average treatment effect (ATE) contain zero. All the code can be found in the Github repository matteobonvini/experiments-sensitivity-paper.
1.1. Motivation

The most widely adopted frameworks for sensitivity analysis generally assume that each unit in the sample could be subject to unmeasured confounding and then proceed by specifying the maximal extent of such confounding. However, just like a treatment effect can be heterogeneous, confounding, too, can differ between units. We propose a complementary approach: in some instances, the researcher may have failed to measure relevant confounders but may hope that there is a subset of units, possibly unknown, for whom the treatment is as good as randomized given the measured covariates.

As a toy example, suppose it is observed that adolescent alcohol drinking (treatment $A$) is positively associated with the occurrence of liver diseases (outcome $Y$). Suppose all confounders $X$ have been recorded except for parental smoking, which could be associated with both $A$ (de Oliveira et al. 2019; Pengpid and Peltzer 2019) and $Y$ due to second-hand smoking (Lammert et al. 2013). Previously proposed sensitivity analyses would check whether a small association between parental smoking and $A$ or $Y$ can explain away the observed $A$-$Y$ association. Instead, we propose to leverage on the observation that parental smoking is a confounder only for units whose parents smoke at home. For instance, some parents may only smoke at work, in which case parental smoking would not have an effect on $Y$. The sample is thus composed of two groups: those units for which $A$ is as good as randomized given $X$ because they are not subject to second-hand smoking regardless of whether their parents smoke and those for which it is not. Depending on how prevalent the former group is, the observed $A$-$Y$ association might be at least partially attributed to the effect of $A$. This toy example generalizes to other cases. For instance, if a confounder is measured with error, the observed covariates may be sufficient to de-confound the treatment-outcome relationship only for an unknown subset of units. In such case, the sample can be thought of containing two groups: those units for whom the confounder was measured correctly, for example, if the questionnaire on motivation or drugs usage was answered truthfully, and those for whom it was not.

The possibility that a sample comes from a mixture of distributions has been studied in great detail in statistics. In robust statistics, for example, it is assumed that a small unknown fraction of the sample comes from a “corrupted” or “contaminated” distribution that is not the target of inference (see Remark 2). In causal inference, unmeasured confounding takes the role of contamination. Borrowing the contaminated model from this literature, we conceptualize that an unknown fraction of the sample suffers from unmeasured confounding.

For example, consider Figure 1. In the shaded region of the space defined by the two observed covariates, the treatment is not assigned randomly; units with covariates’ values falling in this region may have self-selected into the treatment arms and therefore estimating the effect of the treatment on their outcomes is impossible without making further, untestable assumptions. For brevity, we say these units are “confounded,” while the other units are “unconfounded.” Note that, except in special cases, some of which are discussed next, the region is not identifiable from the observed data. However, even if the region is not identifiable, its measure, termed $\epsilon$ in our model, might be specified or upper bounded using subject-matter knowledge. More generally, $\epsilon$ can be varied as a sensitivity parameter. In Figure 1, despite covering different sets of units, all three regions have the same mass, with approximately 20% of the points falling inside them. Given a value for $\epsilon$, we show how to find the region yielding the most conservative inference.

Special cases of our model have already been discussed in the literature when it is known who the confounded units are. For example, in introducing the selective ignorability framework, Joffe, Yang, and Feldman (2010) discussed estimating the effect of erythropoietin alpha (EPO) on mortality using an observational database containing information on all subjects in the United States on hemodialysis. The treatment is thought to be unconfounded only after conditioning on hematocrit, which, however, is not recorded for 10.6% of the subjects. Thus, one may view 10.6% of the sample as coming from a “confounded distribution.” In addition, the differential effects framework proposed in Rosenbaum (2006), too, can be regarded as a special case of our model. Differential effects are treatment contrasts that are immune to certain types of biases called “generic biases.” For example, suppose two treatments are under study. In certain cases, it is plausible that, while units might self select into either treatment arm, the choice of the treatment among units who take exactly one treatment is as good as random. Notice that this setup is a special case of our model: the confounded units are precisely those who are not taking any treatment or are taking a combination of both of them.

Figure 1. The shaded region represents the set of units for whom the treatment is not assigned randomly, even after conditioning on observed covariates. All three figures show approximately the same number of points falling within the “confounded region,” albeit covering different sets of units. The probability $\epsilon$ that a unit falls within the region is our model’s sensitivity parameter, here $\epsilon \approx 0.2$.
Finally, a standard instrumental variables (IVs) setting, too, can be thought of as a case where a fraction of the units is unconfounded. For example, consider an experiment with binary treatment that suffers from units’ noncompliance. The treatment assignment is randomized but the treatment received is not. For the units who complied with the experimental guidelines, the treatment received is equal to the treatment assigned, which is randomly assigned. Thus, the compliers can be considered the units for whom the treatment/outcome relationship is not confounded. In fact, in their detailed analysis of the binary IV model, Richardson and Robins (2010) proposed a sensitivity analysis for the ATE where the sensitivity parameter can be expressed as the proportion of compliers. For the observational setting considered in this article, however, the instrument is never observed, thus, contrarily to a standard IV analysis, the sample contains no information regarding who the confounded units are. In this light, our contribution can also be regarded as an attempt to infer ATEs when it is plausible that nature is acting via an unobservable IV.

2. The Sensitivity Model

We suppose we are given an iid sample \((O_1, \ldots, O_n) \sim \mathbb{P}\) with \(O = (X, A, Y)\), for covariates \(X \in \mathcal{X} \subseteq \mathbb{R}^p\), a binary treatment \(A \in \{0, 1\}\) and an outcome \(Y \in \mathcal{Y} \subseteq \mathbb{R}\). We let \(Y^a\) denote the potential outcome that would have been observed had the treatment been set to \(A = a\) (Rubin 1974). The goal is to estimate the ATE defined as \(\psi = \mathbb{E}(Y|A = 1) - \mathbb{E}(Y|A = 0)\). To ease the notation, we let \(\pi(a|X) = \mathbb{P}(A = a|X)\),

\[
\mu_a(X) = \mathbb{E}(Y|A = a, X), \quad \text{and} \quad \eta = \{\pi(0|X), \mathbb{P}(1|X), \mu_0(X), \mu_1(X)\}.
\]

Throughout, we assume that the following two assumptions hold

**Assumption 1 (Consistency).** \(Y = AY^1 + (1 - A)Y^0\).

**Assumption 2 (Positivity).** \(\mathbb{P}\{t \leq \pi(a|X) \leq 1 - t\} = 1\) for some \(t > 0\).

Both assumptions are standard in the causal inference literature. Consistency rules out any interference between the units, whereas positivity requires that each unit has a nonzero chance of receiving either treatment arm regardless of their covariates’ values. It is well known that if, in addition to consistency and positivity, it also holds that \(Y^a \perp A|X\) (no unmeasured confounding), then \(\psi\) can be point-identified as \(\psi = \mathbb{E}(\mu_1(X) - \mu_0(X))\). In this work, we propose a sensitivity model that relaxes the no-unmeasured-confounding assumption while retaining both consistency and positivity. As a consequence of this relaxation, \(\psi\) is no longer point-identified but it can still be bounded.

Our model supposes that the observed distribution \(\mathbb{P}\) is derived from a counterfactual distribution \(\mathbb{Q}\) of \((X, A, Y^1, Y^0)\) such that

\[
\mathbb{Q} = \epsilon \mathbb{Q}_0 + (1 - \epsilon) \mathbb{Q}_1, \tag{1}
\]

where \(\mathbb{Q}_0\) is a “confounded distribution” for which \(A \perp Y^a|X\) and \(\mathbb{Q}_1\) is an “unconfounded distribution” for which \(A \perp Y^a|X\). In practice, it might be useful to think of each \(\mathbb{Q}_i\) as potentially factoring according to \(A \perp Y^a|S_i\), where \(S_i\) is a set of confounding variables such that \(S_i\) is measured but \(S_0 \setminus S_1 \neq \emptyset\).\(^1\)

The parameter \(\epsilon \in \mathcal{E} \subseteq [0, 1]\) governs the proportion of unmeasured confounding. It is unknown and not estimable but can be varied as a sensitivity parameter. Here, \(\mathcal{E}\) is an interval that the user can specify. Although \(\psi\) cannot be point-identified for \(\epsilon > 0\), it is possible to bound it as a function of \(\epsilon\). In particular, for \(\epsilon = 1\), the familiar worst-case bounds are recovered. For an outcome bounded in \([0, 1]\), these bounds have width equal to 1, which means that the sign of the treatment effect is not identified. Varying the sensitivity parameter to recover different identification regions has been proposed in other works, such as Richardson et al. (2014), Kennedy, Harris, and Keele (2019), and Diaz and van der Laan (2013), albeit for different targets of inference or sensitivity models.

An equivalent formulation of our model (1) is one where there is a latent selection indicator \(S \in \{0, 1\}\), with \(\mathbb{P}(S = 1) = 1 - \epsilon\), such that \(A \perp Y^a|X, S = 0\) but \(A \perp Y^a|X, S = 1\). The following lemma rewrites \(\psi\) in terms of \(S\).

**Lemma 1.** Let \(\lambda_0(X) = \mathbb{E}(Y^0|A = 1 - a, X, S = 0)\). Under consistency (1) and positivity (2), it holds that

\[
\psi = \mathbb{E}(1 - S)[(Y - \lambda_1(X))|A = 1)\]

\[+ S(\mathbb{E}(Y|A = 1, X, S = 1) - \mathbb{E}(Y|A = 0, X, S = 1))].
\]

All proofs can be found in the supplementary materials. As shown in Lemma 1, \(\psi\) depends on three unobservable quantities: \(\lambda_0(X), \lambda_1(X), \text{and} S\). The quantity \(\lambda_1(X)(\lambda_0(X))\) represents the average outcome for those control (treated) units subject to unmeasured confounding had they taken the treatment (control) instead. Without further assumptions, the observed distribution \(\mathbb{P}\) would not impose any restrictions on \(\lambda_0(X)\) or \(\lambda_1(X)\) even if \(S\) was known.

For any given \(\epsilon\), a sharp lower (upper) bound on \(\psi\) can be obtained by minimizing (maximizing) \(\psi\) in Lemma 1 over \(\lambda_0(X), \lambda_1(X), \text{and} S\). Without imposing some restrictions on the distribution of \(S\), the optimization step involves finding, and nonparametrically estimating, the optimal regression functions \(\mathbb{E}(Y|A = a, X, S = 1)\). Given a sample of \(n\) observations, this step would involve fitting regression functions on \(\left\{\sum_{i=0}^{n} e_i\right\}\) different subsamples of size \([n]\), which is computationally very costly even for moderate sample sizes.

Instead, we proceed by requiring that \(S \perp (Y, A)|X\); we call the resulting sensitivity model “\(X\)-mixture model.” The assumption that \(S \perp (Y, A)|X\) can be interpreted in at least three ways. First, one may hope that it holds exactly for the mechanism that

\(^1\)As pointed out by an anonymous reviewer, the mixture model (1) could be generalized to \(\mathbb{Q} = \sum_{i=1}^{k} \gamma_i \mathbb{Q}_i\), where each \(\mathbb{Q}_i\) is a distribution on the counterfactuals capturing different degrees of the confounding. While richer sensitivity analyses can yield more nuanced conclusions, the large number of parameters whose plausibility range would need to be assessed \((J - 1)\) in this case) may hinder their applications in many settings. For instance, consider the toy example above, with \(X = \{\emptyset\} \text{ and } Y, A, U \in \{0, 1\}\) for simplicity. Suppose that \(\mathbb{P}(U = 1, A) = \gamma_0 + \gamma_1 A\) and \(\mathbb{Q}_2(Y^1(A = 1, U) = 0)\). For some constants \(\gamma\) and \(\alpha\). Then, \(\mathbb{E}_{\mathbb{Q}_2}(Y^1(Y^1) = 0) = \mathbb{E}_{\mathbb{Q}_2}(Y^1(A = 1) - \mathbb{E}_{\mathbb{Q}_2}(Y^1|A = 0) = 0\), but \(\mathbb{E}_{\mathbb{Q}_2}(Y^1(A = 0) = \mathbb{E}_{\mathbb{Q}_2}(Y^1|A = 0) = \alpha_0\) \(\gamma_1\), which is generally nonzero.
generated the sample. For instance, it is trivially satisfied, for example, if \( S \) is just a possibly unknown, deterministic function of the observed covariates. An example satisfying this condition is given by the selected ignorability framework proposed in Joffe, Yang, and Feldman (2010): if the treatment is as good as randomized conditional on hematocrit (and possibly other observed covariates), then \( S \) could be an indicator of whether hematocrit is missing.

Even if it does not hold exactly, assuming \( S \perp (Y, A) \mid X \) may be a close approximation to reality that one can use to make the problem computationally tractable. This second interpretation is in the same spirit as using parametric regression models to simplify a given problem, hoping that they will be a close approximation to the true regression function. Third, even if \( S \perp (Y, A) \mid X \), the X-mixture model can help determining whether a study is not robust to unmeasured confounding. Because the bounds if no assumptions are made will be at least as wide as those under \( S \perp (Y, A) \mid X \), if a study does not appear robust in the X-mixture model, it will not appear robust in the general case either. In the following theorem, we derive closed-form expressions for sharp bounds on \( \psi \) in the X-mixture model.

**Theorem 1 (Bounds in X-mixture model).** Suppose that Assumptions 1 and 2 hold. Further suppose that

\[
S \perp (A, Y) \mid X
\]

and that \( \mathbb{P}(Y \in [y_{\min}, y_{\max}]) = 1 \), for \( y_{\min}, y_{\max} \) finite. Choose \( \delta \in [0, 1] \) such that

\[
L_\delta \equiv \delta |y_{\min} - \mu_a(X)| \leq \lambda_a(X) - \mu_a(X) \leq \delta |y_{\max} - \mu_a(X)| \equiv U_\delta \text{ with prob. 1}
\]

for \( a \in \{0, 1\} \). Then, as a function of \( \epsilon \), sharp bounds on \( \psi \) are

\[
\psi_f(\epsilon) = \mathbb{E}[\mu_1(X) - \mu_0(X) + \mathbb{1}\{g(\eta) \leq q_\epsilon\} g(\eta)] - \epsilon, \\
\psi_u(\epsilon) = \mathbb{E}[\mu_1(X) - \mu_0(X) + \mathbb{1}\{g(\eta) > q_{1-\epsilon}\} g(\eta)],
\]

where \( g(\eta) = \pi(0 \mid X)U_1 - \pi(1 \mid X)L_0 \) and \( q_\epsilon \) is its \( \tau \)-quantile.

Theorem 1 yields the identification of sharp lower and upper bounds on \( \psi \) when it is suspected that 100% of the units in the sample are confounded and it is assumed that predicting whether a unit is confounded or not cannot be improved by conditioning on \( (Y, A) \). Relaxing condition (A1) to \( S \perp Y \mid (A, X) \) poses additional challenges and it is discussed in Appendix 3. We refer to this relaxed version of the X-mixture model as the “XA-mixture model.” Notably, it covers the differential effects framework of Rosenbaum (2006), as one could specify \( \mathcal{S} = 1(A_1 + A_2 = 1) \) for some binary treatment \( A_1 \) and \( A_2 \).

The bounds are in terms of the parameters \( \epsilon \) and \( \delta \), as well as the regression functions \( \pi(a \mid X) \) and \( \mu_a(X) \), and they involve nonsmooth transformations of unknown functions of \( \mathbb{P} \). The parameter \( \epsilon \) is our main sensitivity parameter and controls the proportion of unmeasured confounding in the sample. Parallely, \( \delta \) controls the extent of unmeasured confounding among the \( S = 0 \) units, as it bounds the difference between the unobservable regression \( \lambda_a(X) \) and the estimable regression \( \mu_a(X) \). Notice that (2) always holds for \( \delta = 1 \). Setting \( \delta < 1 \) imposes an untestable assumption on the severity of the unmeasured confounding, which might be sensible if some knowledge on the confounding mechanism is available. Specifically, our parameterization is such that \( \lambda_a(X) \) can be bounded by linear combinations of \( y_{\min}, y_{\max} \) and \( \mu_a(X) \):

\[
\delta y_{\min} + (1 - \delta)\mu_a(X) \leq \lambda_a(X) \leq \delta y_{\max} + (1 - \delta)\mu_a(X).
\]

Unless otherwise specified, in what follows we consider \( y_{\min} = 0, y_{\max} = 1 \), and set \( \delta = 1 \), thus yielding

\[
\psi_f(\epsilon) = \mathbb{E}[\mu_1(X) - \mu_0(X) + \mathbb{1}\{g(\eta) \leq q_\epsilon\} g(\eta)] - \epsilon, \\
\psi_u(\epsilon) = \mathbb{E}[\mu_1(X) - \mu_0(X) + \mathbb{1}\{g(\eta) > q_{1-\epsilon}\} g(\eta)],
\]

for \( g(\eta) = \pi(0 \mid X)[1 - \mu_1(X)] + \pi(1 \mid X)\mu_0(X) \). If \( Y \) is bounded, this choice does not impose any assumption since \( Y \) can be rescaled to be in \([0, 1]\). If \( Y \) is unbounded, Theorem 1 is not directly applicable, but a similar result can be derived if one is willing to assume that \( |\lambda_a(X) - \mu_a(X)| \leq \delta \) for \( a \in \{0, 1\} \) and \( \delta < \infty \). We leave further investigation of the unbounded case as future work. We conclude this section with four remarks aiming to shed some more light on the bounds derived in Theorem 1.

**Remark 1.** Suppose \( Y \) is bounded in \([0, 1]\) and take \( \delta = 1 \). The length of the bound is then

\[
\Delta(\epsilon) = [\mathbb{E}[g(\eta)] g(\eta) > q_{1-\epsilon}] - \mathbb{E}[g(\eta)] g(\eta) \leq q_\epsilon + 1] \epsilon.
\]

If \( S \) was known, the length of the bound would reduce to \( \Delta(\epsilon) = \epsilon \). Thus, we can view the term \( [\mathbb{E}[g(\eta)] g(\eta) > q_{1-\epsilon}] - \mathbb{E}[g(\eta)] g(\eta) \leq q_\epsilon] \epsilon \) as the “cost” of not knowing who the confounded units are.

**Remark 2.** The conditional independence of \( S \) and \( Y \) considerably simplifies the optimization step. To see this, notice that \( \mathbb{E}[Y \mid A = a, S = 1] = \mu_a(X) \) if \( S \perp Y \mid A, X \). In turn, this implies that \( \psi \) can be written as

\[
\psi = \mathbb{E}[\Gamma(Y, A, X) + S(\mu_1(X) - \mu_0(X) - \Gamma(Y, A, X))],
\]

where \( \Gamma(Y, A, X) = \{Y - \lambda_{1-\Lambda}(X)\}(2A - 1) \). Therefore, bounds on \( \psi \) can be derived from bounds on \( \mathbb{E}[\mu_1(X) - \mu_0(X) - \Gamma(Y, A, X)S = 1] \), which fits the framework studied by Horowitz and Manski (1995). In their work, the goal is to do inference about a distribution \( Q_\tau \) using data \( Y \) such that \( Y = ZY_1 + (1 - Z)Y_0 \), with \( Z \in [0, 1] \) and \( Y_1 \sim Q_\tau \). They discuss two models: the “contaminated sampling model,” which assumes \( Z \) to be independent of \( Y_1 \), and the “corrupted sampling model,” which does not make this assumption. If it is known that \( \mathbb{P}(Z = 0) \leq \lambda \), they derive sharp bounds on the conditional expectation of \( Y_1 \) given some covariates \( X \) when contamination or corruption does not occur in \( X \). Our setup does not immediately fit this framework because corruption applies to all observed variables \( Y, A, X \). However, if \( S \perp Y \mid A, X \), the optimal solution for \( S \) can be found by considering only the marginal distribution of the one-dimensional random variable \( \mu_1(X) - \mu_0(X) - \Gamma(Y, A, X) \). Following the terminology in Horowitz and Manski (1995), we may view the assumption that \( S \perp Y \mid A, X \) as a compromise between contamination \( S \perp (Y, A, X) \) and corruption (no assumption on \( S \)).
Remark 3. As pointed out by Robins (2002), many interesting sensitivity analyses make use of parameters that depend on the covariates collected. In turn, this might hinder the direct comparison of studies’ robustness. For example, a study where many confounders have been properly taken into account might appear more sensitive to departures from the no-unmeasured-confounding assumption than a study that failed to control for any confounder. This could happen, for instance, if the effect estimate in the former study is closer to the null value than the estimate from the latter. This apparent paradox might arise because a sensitivity analysis measures departures from a weak or strong assumption depending on whether many or few observed confounders are collected. Our proposed sensitivity analysis hinges on \( \epsilon \), the proportion of unmeasured confounding, which depends on the covariates collected. As such, it might be subject to this paradox.

Remark 4. Section 4 of Rosenbaum (1987) contains a modification to the sensitivity analysis proposed in that article, and briefly summarized in our introduction, that allows an unknown fraction \( \beta \) of the sample to suffer from unmeasured confounding. While conceptually similar to the approach presented in this article, their method relies on exact matching. In fact, if units are exactly matched on observed covariates, our sensitivity model recovers Rosenbaum’s with \( \beta = \epsilon \) and \( \Gamma = 0 \). However, exact matching is often infeasible due to the presence of continuous or high-dimensional covariates. Therefore, our work can be viewed as an extension to Rosenbaum’s Section 4 model to the case where units are not matched on observed covariates.

2.1. One-Number Summary of a Study’s Robustness

In practice, one might want to report a one-number summary of how robust the estimated effect is to the number of confounded units. An example of such summary is the minimum proportion of confounded units \( \epsilon_0 \) such that the bounds on \( \psi \) are no longer informative about the sign of the effect, that is, that they contain zero. Larger values of \( \epsilon_0 \) indicate that the estimated effect is more robust to potential unmeasured confounding. Mathematically,

\[
\epsilon_0 = \arg \min_{\epsilon \in \mathcal{E}} \mathbb{E} \left[\left| \text{sgn}(\psi(\epsilon)) - \text{sgn}(\psi_u(\epsilon)) \right| \right],
\]

where \( \text{sgn}(x) \) measures the sign of \( x \), \( \text{sgn}(x) = -1(x < 0) + 1(x > 0) \). Because \( \psi_u(\epsilon = 1) - \psi(\epsilon = 1) = 1 \), the minimum is guaranteed to be attained in \( \mathcal{E} = [0, 1] \). Furthermore, under certain mild conditions, the bounds are continuous and strictly monotone in \( \epsilon \), hence \( \epsilon_0 \) is generally the unique value such that \( \psi(\epsilon_0) = 0 \) or \( \psi_u(\epsilon_0) = 0 \). This motivates the moment condition \( \psi(\epsilon_0)\psi_u(\epsilon_0) = 0 \), which we use to construct a Z-estimator of \( \epsilon_0 \).

Other authors have proposed one-number summaries of a study’s robustness to unmeasured confounding. For example, the minimum value for \( \Gamma \) in Rosenbaum’s framework and its extensions (Rosenbaum 1987; Gastwirth, Krieger, and Rosenbaum 1998; Zhao, Small, and Bhattacharya 2017; Yadlowsky et al. 2018) such that the observed effect ceases to be statistically significant can be used as a summary of study’s robustness to unmeasured confounding. Recently, Ding and VanderWeele (2016) and VanderWeele and Ding (2017) have introduced the E-value, which measures the minimum strength of association, on the risk ratio scale, that an unmeasured confounder would need to have with both the outcome and the treatment to “explain away” the observed effect of the treatment on the outcome. To derive the elegant formula for the E-value, the unobserved confounder is assumed to be associated with the treatment and with the outcome in equal magnitude. Furthermore, the derivation makes use of a bounding factor that needs to be computed for each stratum of the covariates. Computing such bounding factor when the observed covariates are continuous or high-dimensional can be problematic. Moreover, their method requires additional approximations if the outcome is not binary. On the other hand, the one-number summary proposed here does not require any further assumption other than the restriction on \( S \) described above. Hence, we view these summary measures as complementary and the specific context would generally dictate which one is more appropriate.

3. Estimation and Inference

3.1. Proposed Estimators

There are at least two types of bias that can arise when estimating a causal effect using observational data: the bias arising from incorrectly assuming that all confounders have been collected and the statistical bias of the chosen estimator (Luedtke, Diaz, and van der Laan 2015). In Section 2, we constructed a model to probe the effects of the former bias. In this section, we propose estimators that aim to minimize the latter. Our estimators of the bounds are built using the efficient influence functions (IFs) and cross-fitting. IFs play a crucial role in nonparametric efficiency theory, as the variance of the efficient IF can be considered the nonparametric counterpart of the Cramer–Rao lower bound in parametric models. Furthermore, estimators constructed using the efficient IF have favorable properties, such as doubly-robustness or second-order bias. Here, we note that \( \psi(\epsilon) \) and \( \psi_u(\epsilon) \) do not possess an IF, as they are not pathwise differentiable. However, certain terms appearing in their expressions, such as \( E(\mu_1(X)) \), are pathwise differentiable; as such, they can be estimated using IFs. For terms that are not pathwise differentiable we resort to plug-in estimators. We refer to Bickel et al. (1993), van der Vaart (2002), van der Laan, Laan, and Robins (2003), Tsiatis (2007), Chernozhukov et al. (2016), and others for detailed accounts on IFs and their use.

To ease the notation in this section, let

\[
v(O; \eta) = \frac{(2A - 1) \{Y - \mu_A(X)\}}{\pi(A|X)} + \mu_1(X) - \mu_0(X)
\]

denote the uncentered IF for the parameter \( E(\mu_1(X) - \mu_0(X)) \). Furthermore, let \( \tau(O; \eta) \) denote the uncentered IF for \( E(g(\eta)) \):

\[
\tau(O; \eta) = \frac{(1 - 2A) \{Y - \mu_A(X)\}}{\pi(A|X) \pi(1 - A|X)} + A\mu_0(X)
\]

and let

\[
\psi(\epsilon_0)\psi_u(\epsilon_0) = 0,
\]

where

\[
\psi(\epsilon_0) = \epsilon_0 - \frac{1}{\pi(A|X)} - \frac{\mu_1(X) - \mu_0(X)}{(1 - A)(1 - \mu_1(X))},
\]

and let

\[
\psi_u(\epsilon_0; q_1) = \psi(\epsilon_0) + \mathbb{I}[g(\eta) > q_{1-\epsilon}]\tau(O; \eta) - \epsilon,
\]

where

\[
\psi_u(\epsilon_0; q_{1-\epsilon}) = \psi(\epsilon_0) + \mathbb{I}[g(\eta) > q_{1-\epsilon}]\tau(O; \eta).
\]
Then, it holds that $\psi_t(\epsilon) = \mathbb{E}[\psi_t(O; \eta; q_t)]$ and $\psi_u(\epsilon) = \mathbb{E}[\psi_u(O; \eta; q_{1-\epsilon})]$. Following Robins et al. (2008), Zheng and van der Laan (2010), and Chernozhukov et al. (2016), among others, we use cross-fitting to allow for arbitrarily complex estimators of the nuisance functions $\eta$ and $q_t$ to avoid empirical process conditions. Specifically, we split the data into $B$ disjoint groups of size $n/B$ and we let $K_i = k$ indicate that subject $i$ is split into group $k$, for $k \in \{1, \ldots, B\}$. Notice that it is not required that the groups have equal size, for example each $K_i$ could be drawn uniformly from $\{1, \ldots, B\}$. For simplicity, we proceed with having equal-size groups. We let $P_n$ denote the empirical measure as $P_n[f(O)] = \frac{1}{n} \sum_{i=1}^{n} f(O_i)$ and $P^n_k$ denote the sub-empirical measure as $P^n_k[f(O)] = \sum_{i=1}^{n} f(O_i) 1(K_i = k)/\sum_{i=1}^{n} 1(K_i = k)$. In addition, we let $\hat{\eta}_{-k}$ denote the estimator of $\eta$ computed without observations from fold $K = k$ and $\hat{q}_{1-k}$ denote the estimator of $q_t$ equal to the empirical quantile of $g(\hat{\eta}_{-k})$ solving $P^n_k[1\{g(\hat{\eta}_{-k}) \leq q_{1-k}\}] = \tau + o_p(n^{-1/2})$. Then, we estimate the bounds as

\[
\hat{\psi}_t(\epsilon) = \frac{1}{B} \sum_{k=1}^{B} P^n_k[1\{\psi(O; \hat{\eta}_{-k}) + 1\{g(\hat{\eta}_{-k}) \leq \hat{q}_{1-k}\}\} > \epsilon] \\
\hat{\psi}_u(\epsilon) = \frac{1}{B} \sum_{k=1}^{B} P^n_k[1\{\psi(O; \hat{\eta}_{-k}) + 1\{g(\hat{\eta}_{-k}) > \hat{q}_{1-k}\}\} > \epsilon]
\]

The computation of the estimators above is straightforward as it amounts to fitting regression functions on $B - 1$ subsets of the data and evaluate the estimated functions at the values of the covariates on the corresponding test set. The use of cross-fitting lends itself naturally to the use of parallel computing as one can estimate the regression functions on different subsets of the data simultaneously. We incorporate this possibility in our implementation of the methods in R. Moreover, it is worth noting that cross-fitting does not discard any data point in the estimation step, since each observation is used twice without overfitting: once for estimating the regression functions and once for estimating the expectation operator. In addition, because we are working under a fully nonparametric model, there exists only one $P$; therefore, our estimators of the pathwise differentiable terms are efficient in the sense that they asymptotically achieve the semiparametric efficiency bound.

Finally, while the estimators of the bounds discussed in this section have several attractive properties in terms of computational tractability and convergence rates, they might not be monotone in $\epsilon$ in finite samples. To remedy this, the estimators can be “rearranged” using the procedure described in Chernozhukov, Fernandez-Val, and Galichon (2009). We apply this procedure in Section 4, although we find that the original, non-rearranged estimators achieve low bias and nominal uniform coverage as well.

### 3.2. Establishing Weak Convergence

To state asymptotic guarantees for the proposed estimators, we first make the following technical assumption:

**Assumption 3 (Margin condition).** The random variable $g(\eta)$ has absolutely continuous CDF and there exists $\alpha > 0$ such that for all $t > 0$ and $\tau \in \mathbb{E}$, it holds that $\mathbb{P}\left(\{g(\eta) - q_{1-\epsilon} \leq t\}\right) \lesssim \epsilon^{\tau}$ and $\mathbb{P}\left(\{g(\eta) - q_{1-\epsilon} \geq t\}\right) \lesssim \epsilon^{\tau}$. 

Assumption 3 requires that there is not too much mass around any $\epsilon$-quantile and $(1-\epsilon)$-quantile of $g(\eta)$, for $\epsilon \in \mathbb{E}$. It is essentially equivalent to the margin condition used in classification problems (Audibert and Tsybakov 2007), optimal treatment regime settings (van der Laan and Luedtke 2014; Luedtke and van der Laan 2016), and other problems involving estimation of nonsmooth functionals (Kennedy, Balakrishnan, and G’Sell 2018; Kennedy, Harris, and Kellee 2019). Notably it is satisfied for $\alpha = 1$ if, for instance, the density of $g(\eta)$ is bounded on $\mathbb{E}$. We give the main convergence theorem for $\hat{\psi}_t(\epsilon)$. A similar statement holds for $\hat{\psi}_u(\epsilon)$.

**Theorem 2.** Let

\[
\hat{\sigma}^2_u(\epsilon) = \mathbb{E}\left[\left((\psi_u(O; \hat{\eta}_{-k}, \hat{q}_{1-k}) - \psi_u(\epsilon)) - (\hat{q}_{1-k} - \hat{q}_{1-\epsilon})(1\{\hat{g}(\hat{\eta}_{-k}) > \hat{q}_{1-\epsilon}\} - \hat{q}_{1-\epsilon})\right)^2\right]
\]

be the estimator of the variance function

\[
\sigma^2_u(\epsilon) = \mathbb{E}\left[\left((\psi_u(O; \eta, q_{1-\epsilon}) - \psi_u(\epsilon)) - q_{1-\epsilon}(1\{g(\eta) > q_{1-\epsilon}\} - 1)\right)^2\right]
\]

If Assumptions 1–3 hold, and the following conditions also hold:

1. $\mathbb{P}\{t \leq \sqrt{n} |X| \leq 1 - t\} = 1$ for $a = 0, 1$ and some $t > 0$.
2. $\sup_{\epsilon \in \mathbb{E}} |\hat{\sigma}_u(\epsilon) - \sigma_u(\epsilon)| = o_p(1)$.
3. $\sup_{\epsilon \in \mathbb{E}} \mathbb{E}[|\psi_u(O; \eta, q_{1-\epsilon}) - \psi_u(\epsilon) - q_{1-\epsilon}(1\{g(\eta) > q_{1-\epsilon}\} - 1)|] = o_p(1)$.
4. $\mathbb{E} \left( |g(\eta) - g(\eta)| + \mathbb{E}[\hat{q}_{1-\epsilon} - q_{1-\epsilon}] \right)^{1+\alpha} = o_p(n^{-1/2})$, for $\alpha$ satisfying Assumption 3.
5. $\|\sqrt{n}(\hat{\psi}_u(\epsilon) - \psi_u(\epsilon)) \sqrt{\mathbb{E}}(1 |X|) \|_{\infty} \leq \sup_{a \in \mathbb{E}} \|\hat{\sigma}_u(X) - \mu_u(X)\| = o_p(n^{-1/2})$.

Then $\sqrt{n}(\hat{\psi}_u(\epsilon) - \psi_u(\epsilon)) / \hat{\sigma}_u(\epsilon) \rightarrow G(\epsilon)$ in $\mathcal{L}^\infty(\mathbb{E})$, with $\mathcal{E} \subseteq [0, 1]$, where $G(\cdot)$ is a mean-zero Gaussian process with covariance $\mathbb{E}[G(\epsilon)G(\eta)] = \mathbb{E}\left[\phi_u(O; \eta, q_{1-\epsilon})\phi_u(O; \eta, q_{1-\epsilon})\right]$ and $\phi_u(O; \eta, q_{1-\epsilon})$ and $\phi_u(O; \eta, q_{1-\epsilon}) - \psi_u(\epsilon) - q_{1-\epsilon}(1\{g(\eta) > q_{1-\epsilon}\} - 1) \rightarrow \sigma_u(\epsilon)$.

Theorem 2 gives sufficient conditions so that the estimated curves tracing the lower and upper bounds as a function of $\epsilon$ converge to a Gaussian process. In turn, this enables the computation of confidence bands trapping the ATE with any desired confidence level uniformly over $\epsilon$. The first three conditions of the theorem are quite mild. Condition 1 is a positivity condition requiring that the estimator of the propensity score is bounded away from 0 and 1. Condition 2 requires uniform consistency of the variance estimator at any rate. Condition 3 holds if, in addition to satisfying the margin Assumption 3, $g(\eta)$ and $\hat{q}_i$ converge uniformly, in $\mathbb{E}$ and $\epsilon$, respectively, to the truth at any rate.

The key assumptions are conditions 4 and 5. While more restrictive than the first three, these conditions can be satisfied even if flexible machine learning tools are used. In fact,
condition 5 only requires that the product of the $L_2$ errors in estimating $\pi(a|X)$ and $\mu_Y(X)$ is of order $n^{-1/2}$, which means that, for example, each regression function can be estimated at the slower rate $n^{-1/4}$. A rate of convergence in $L_{\infty}$ norm of order $n^{-1/4}$ is also sufficient to satisfy condition 4 if the density of $g(\eta)$ is bounded because the marg In our settings, a natural way to define the bandwidth $h$ is to use the slower rate $n^{-1/4}$, and $h$ is achieved if nonparametric smoothness, sparsity or other structural assumptions are imposed on the true regression functions. Theorem 3. Suppose that the CDF of $g(\eta)$ is strictly increasing in neighborhoods of $q_{\epsilon_0}$ and $q_{1-\epsilon_0}$. The asymptotic normality of $\epsilon_0$ relies on the existence (and nonsingularity) of the derivative of the map $e \mapsto \psi(e)\psi_{\epsilon_0}(e)$ at $e = \epsilon_0$. Calculating such derivative requires computing the derivative of the quantile function, which is why we require the CDF of $g(\eta)$ to be strictly increasing in the relevant neighborhoods. We expect all these conditions to be satisfied in practice in the presence of continuous covariates and enough smoothness or sparsity for the regression functions.

Asymptotic normality allows the straightforward calculation of a Wald-type confidence interval for $\epsilon_0$ using a consistent estimate for the variance. We thus propose reporting both a point-estimate for $\epsilon_0$ and $1-\alpha$ confidence interval as a summary of the study's robustness to unmeasured confounding.

4. Illustrations

4.1. Simulation Study

In this section, we report the results of the simulations we performed to investigate the finite-sample performance of our proposed estimators. We consider the following data generating mechanism:

$$X_i \sim \text{TruncNorm}(\mu = 0, \sigma = 1, lb = -2, ub = 2)$$

for $i \in \{1, 2\}$, $U \sim \text{Bern}(0.5)$,

$$S|X_1, X_2 \sim \text{Bern}(\Phi(X_1)),$$

$$A|X_1, X_2, U, S \sim \text{Bern}(0.5(\Phi(X_1) + 0.5S + (1 - S)U)),$$

$$Y^d|X_1, X_2, U, S, A \sim \text{Bern}(0.25 + 0.5\Phi(X_1 + X_2) + (a - 0.5)r - 0.1U),$$

$$Y = AY^1 + (1 - A)Y^0,$$

where $\Phi(\cdot)$ denotes the CDF of a standard normal random variable. Notice that

$$\Pr(A = 1|X_1, X_2, S = 0) = \Pr(A = 1|X_1, X_2, S = 1) = 0.5\Phi(X_1) + 0.25,$$

thus this model satisfies the assumptions of Theorem 2 and it implies that $E(Y^1 - Y^0) = r$. The random variable $U$ acts as a binary unmeasured confounder; given the observed covariates $X$, units with $S = 0$ and $U = 1$ are more likely to be treated and exhibit $Y = 0$ than those with $S = 0$ and $U = 0$. Therefore, under this setup, one would expect the treatment effect to be underestimated if the no-unmeasured-confounding assumption is (incorrectly) assumed to be true.\(^2\)

\(^2\)In principle, one could construct the empirical moment condition after performing the rearrangement procedure of Chernozhukov, Fernandez-Val, and Galichon (2009). Whether or not the rearrangement is done, we expect the inference about $\epsilon_0$ to be equivalent asymptotically and vary minimally in finite samples.

\(^3\)To incorporate finite sampling uncertainty in sensitivity analyses, one-number summaries of a study's robustness are generally computed as the values of the sensitivity parameter(s) such that a $\alpha$-level confidence interval for the ATE under no unmeasured confounding includes the null value. Choosing different $\alpha$s to estimate the ATE with no residual confounding may then yield different conclusions regarding the study's robustness to unmeasured confounding, despite the latter being a separate inferential task. Constructing a confidence interval for $\epsilon_0$ directly bypasses this issue.

\(^4\)In the context of the toy example of Section 1.1, $U$ and $S$ indicate whether the parents are smokers and whether they would smoke at home, respectively.
We estimate the lower bound \( \psi_l(\epsilon) \), the upper bound \( \psi_u(\epsilon) \) and \( \epsilon_0 \) using the methods outlined in Section 3.1. In particular, we use 5-fold cross-fitting to estimate the nuisance functions, fitting both generalized linear and additive models via the SuperLearner method (van der Laan, Polley, and Hubbard 2007). The performance of the proposed estimators is evaluated via integrated bias, root-mean-squared-error (RMSE), and coverage. These evaluation metrics offer insight into what sample size is required to achieve a good performance of the multiplier bootstrap, which relies on the convergence of the bounds' estimators to a Gaussian process.

\[
\begin{align*}
\hat{\text{bias}} &= \frac{1}{I} \sum_{i=1}^{I} \left[ \sum_{j=1}^{J} (\hat{\psi}_l(\epsilon_j) - \psi_l(\epsilon_j)) \right], \\
\text{RMSE} &= \sqrt{\frac{1}{I} \sum_{i=1}^{I} \left[ \sum_{j=1}^{J} (\hat{\psi}_l(\epsilon_j) - \psi_l(\epsilon_j))^2 \right]}^{1/2}
\end{align*}
\]

and suitably modified formulas for \( \psi_u(\epsilon) \) and \( \epsilon_0 \). We run \( I = 500 \) simulations across \( I = 21 \) values of \( \epsilon \) equally spaced in \( \epsilon = \{0, 0.2\} \). To better estimate \( \epsilon_0 \) we make the grid finer and consider 201 values of \( \epsilon \) equally spaced in \( \epsilon \). To evaluate 95% uniform coverage, we say that the uniform band covers if it contains the true region \( [\psi_l(\epsilon), \psi_u(\epsilon)] \) for all \( \epsilon \in \epsilon \). Finally, we assess bias and 95% coverage for \( \epsilon_0 \). Table 1 shows the results of our simulation for \( r = 0.05 \). This set up is such that \( \epsilon_0 = 0.041 \). In addition, if no-unmeasured-confounding is erroneously thought to hold (\( \epsilon = 0 \)), \( \psi \) is, on average, underestimated since \( E \{\mu_1(\mathbf{X}) - \mu_0(\mathbf{X})\} \approx 0.023 < r \). This simple simulation setup exemplifies what our theory predicts. Even for moderate sample sizes, we achieve approximately correct nominal uniform coverage for the identification region and \( \epsilon_0 \). Furthermore, the \( \sqrt{n} \times \text{RMSE} \) remains roughly constant as the sample size increases. Finally, in Section 7 of the Appendix, we extend this simulation study to investigate how conservative our model would be if the true \( \epsilon_0 \) is actually zero, that is, there is no unmeasured confounding.

### 4.2. Application

In this section, we illustrate the proposed sensitivity analysis by reanalyzing the data from the study on Right Heart Catheterization (RHC) conducted by Connors et al. (1996).\(^5\) The data consist of 5735 records from critically ill adult patients receiving care in an ICU for certain disease categories in one out of five US teaching hospitals between 1989 and 1994. For each patient, demographic variables, comorbidities and diagnosis variables as well as several laboratory values were recorded. A total of 2184 patients underwent RHC within the first 24 hours in the ICU. Within 30 days of admission, 1918 patients died, approximately 38.00% and 30.64% of the treated and control groups, respectively. After conditioning on the measured confounders, the authors concluded that patients treated with RHC had, on average, lower probability of surviving (30-day mortality: OR = 1.24, 95% CI = [1.03, 1.49]). Notably, sensitivity analyses targeting potential violations of the propensity score model suggested robustness of the study’s conclusions to unmeasured confounding.

We investigate the effects of varying the proportion of confounded units while avoiding any parametric assumptions on the nuisance regression functions. One reason to believe that a fraction of the sample might be effectively unconfounded is the following. Suppose there are two types of surgeons: those who prefer performing RHC (R-surgeon) and those who do not (NR-surgeon). One might believe that the surgeon’s preference for RHC is a valid instrument. Roughly, an instrument is a variable that is unconfounded, associated with the treatment receipt, and that affects the outcome only through the treatment. It appears plausible that a surgeon’s preference for RHC would satisfy these conditions if, for instance, the efficacy of RHC was not well understood at the time the study was conducted. In fact, physicians’ preferences for a treatment have been used as IV’s before (see, e.g., Hernán and Robins 2006; Baiocchi, Cheng, and Small 2014 for reviews and discussions). Then, the patients who would undergo RHC if assigned to an R-surgeon but would not undergo RHC if assigned to a NR-surgeon represent the unconfounded unknown fraction of the sample.

Consider the group of patients who underwent RHC. A unit in this group can be either a “complier” or a “noncomplier.” She’s a complier if she would not have undergone RHC if assigned to an NR-surgeon, whereas she’s a noncomplier if she would have undergone RHC regardless of the type of surgeon or only if assigned to a NR-surgeon. In many instances, these two types will differ in terms of observed covariates \( \mathbf{X} \). However, for certain values \( \mathbf{x} \) of \( \mathbf{X} \), a unit might be either a complier or a noncomplier with non-zero probability. In this scenario, our relaxed \( \mathbf{X} \)-model posits that the probability of survival conditional on receiving RHC is the same for a complier and a noncomplier sharing the same \( \mathbf{x} = \mathbf{x} \). Notice that this is not imposing any assumption on what would have happened to the noncomplier had she not been treated. In fact, we derived the lower (upper) bound on the average effect of RHC by assuming that she would have certainly survived (died) had she not undergone RHC. This maximal conservativeness in deriving the bounds likely protects our conclusions from mild violations of our \( \mathbf{X} \)- and \( \mathbf{X} \)-models.

To construct the curves tracing the bounds using the data, we estimate the nuisance regression functions via the cross-validation-based SuperLearner ensemble (van der Laan, Polley, and Hubbard 2007), combining generalized additive models, random forests, splines, support vector machines as well as generalized linear models. We perform 5-fold cross-fitting. We also construct pointwise and uniform confidence bands. Results are reported in Figure 2.

### Table 1. Simulation results across 500 simulations.

| \( n \) | \( \psi_l(\epsilon) \times 100 \) | \( \psi_u(\epsilon) \times 100 \) | \( \sqrt{n} \times \text{RMSE} \times 100 \) | Coverage (\( \times 100 \)) |
|-----|-----------------|-----------------|-----------------|-----------------|
| 500 | 0.38 | 0.12 | 2.47 | 0.95 | 0.96 | 1.32 | 95.4 | 97.0 |
| 1000 | 0.51 | 0.14 | 1.59 | 0.95 | 0.95 | 1.45 | 93.2 | 95.6 |
| 5000 | 0.04 | 0.10 | 0.16 | 0.99 | 0.98 | 1.72 | 92.4 | 95.4 |
| 10,000 | 0.05 | 0.09 | 0.07 | 0.95 | 0.96 | 1.75 | 93.6 | 94.8 |

\(^5\)Available at [http://biostat.mc.vanderbilt.edu/wiki/Main/DataSets](http://biostat.mc.vanderbilt.edu/wiki/Main/DataSets).
In line with the results in Connors et al. (1996), if no-unmeasured-confounding holds, patients treated with RHC show a statistically significant decrease in 30-day survival rates. The risk difference equals \(-3.74\%\) (95% CI = \([-6.00\%, -1.49\%]\)). Under the X-mixture model, the bounds on the difference in survival rate would include zero if more than 4.89% (95% CI = \([1.50\%, 8.28\%]\)) of the patients were confounded. The value reduces to 4.02% (95% CI = \([1.59\%, 6.45\%]\)) under the relaxed XA-mixture model. Whether robustness to 5% of potentially confounded units is enough to attach a causal interpretation to the study's result largely depends on subject-matter knowledge. Earlier we have described \(\epsilon_0\) as the proportion of "noncompliers," but other interpretations are also possible. For instance, suppose it is known that, before deciding whether a patient undergoes RHC, most surgeons look at lab value \(v_1\), but some may check lab value \(v_2\) as well. Both values are correlated with survival, but only \(v_1\) is measured. If reviewers of the study have an idea of how common it is for surgeons to check \(v_2\) in addition to \(v_1\), then they would be able to decide whether \(\epsilon = 5\%\) is large or small. In the supplementary materials, we consider varying \(\delta\), the parameter governing the severity of the unmeasured confounding. For instance, if \(\delta = 0.5\) is thought to be reasonable, robustness would increase to 11.00% (95% CI = \([3.84\%, 18.16\%]\)) under the X-mixture model.

Finally, we refer the readers to Lin, Psaty, and Kronmal (1998) and Altonji, Elder, and Taber (2008), among others, for additional sensitivity analyses applied to this dataset. In particular, in the context of Cox proportional hazard regression, and under certain simplifying assumptions, Lin, Psaty, and Kronmal (1998) derived that a confidence interval for the relative hazard of death would include 1 as long as the prevalence of a binary unmeasured confounder is at least 10% greater in the group that underwent RHC than in the control group. Using a probit model of mortality at day 90, Altonji, Elder, and Taber (2008) showed that the observed positive association between mortality and RHC usage could be "explained away" if the correlation between the unmeasured factors determining RHC usage and mortality is approximately 0.15. In addition, in Section 6.1 of the supplementary materials, we apply the sensitivity analysis designed for linear models proposed in Cinelli and Hazlett (2020). We find that an unmeasured confounder that explains 4.2+% of the variance in mortality not captured by RHC usage and the measured covariates and 4.2+% of the variance in RHC usage not captured by the measured covariates would be sufficient to drive the observed effect (\(\approx 0.04\)) to zero. Notice that these approaches are designed for specific models used in the primary analysis, whereas our framework is agnostic regarding modeling choices. Further, they assume that the treatment-outcome association may be confounded for every unit, while our sensitivity model captures departures from such homogeneity by allowing the treatment-outcome association to be unconfounded for an unknown subgroup of units.

5. Discussion

In this article, we propose a novel approach to sensitivity analysis in observational studies where the sensitivity parameter is the proportion of unmeasured confounding. A strength of our model is that it captures a rich form of unmeasured confounding heterogeneity. While even richer models may allow for a more flexible characterization of confounding heterogeneity, we believe our approach strikes a nice balance between complexity and transparency. In fact, it captures heterogeneity with just one, intuitive sensitivity parameter: an unknown fraction \(\epsilon\) of the units can be arbitrarily confounded while the rest are not. The model is general enough to cover some relaxations to the no-unmeasured-confounding assumption already proposed in the literature. As \(\epsilon\) is varied, lower and upper bounds on the ATE are derived under certain assumptions on the distribution of the confounded units. The parameter \(\epsilon\) is interpretable and yields a natural one-number summary of a study's robustness to unmeasured confounding, namely the minimal proportion of confounding such that the bounds on the ATE contain zero. We provide sufficient conditions to construct both pointwise and uniform confidence bands around the curves tracing the lower and upper bounds on the ATE as a function of \(\epsilon\). We
also describe the asymptotic normality of a Z-estimator of $\epsilon_0$; we propose reporting an estimate of $\epsilon_0$ together with a Wald-type confidence interval when discussing results from an observational study.

Several questions remain unanswered and could be the subject of future research. First, bounding the ATE under no restrictions on the distribution of the confounded units is currently computationally intractable. Therefore, the discovery of a clever way to compute the bounds in this setting would generalize the current version of our model. Second, generalizing the approach of Imbens and Manski (2004) to construct uniform confidence bands trapping the true ATE $\psi$, rather than the identification region $[\psi_l(\epsilon), \psi_u(\epsilon)]$, would allow far more precise inference. Lastly, extensions to our model other than the one considered in Appendix 4 would likely lead to a richer set of sensitivity models, ultimately allowing the user to gauge the effects of departures from the no-unmeasured-confounding assumption in more nuanced ways. For example, it would be interesting to extend our sensitivity model to accommodate time-varying or continuous exposures, as well as to explore the possibility of tighter bounds by employing specific sensitivity analysis models to the confounded fraction of the sample.

Supplementary Materials

The supplementary materials contain proofs of all results along with discussions of possible extensions to the sensitivity analysis model proposed and some additional data analysis.

Acknowledgments

The authors thank Sivaraman Balakrishnan, Colin Fogarty, Marshall Joffe, Alan Mishler, Pratik Patil, and members of the Causal Group at Carnegie Mellon University for helpful discussions.

Funding

Edward Kennedy gratefully acknowledges financial support from NSF grant DMS1810979.

References

Altonji, J. G., Elder, T. E., and Taber, C. R. (2008), “Using Selection on Observed Variables to Assess Bias From Unobservables When Evaluating Swan–Ganz Catheterization,” *American Economic Review*, 98, 345–350. [1548]

Audibert, J.-Y., and Tsybakov, A. B. (2007), “Fast Learning Rates for Plug-In Classifiers,” *The Annals of Statistics*, 35, 608–633. [1545]

Baiocchi, M., Cheng, J., and Small, D. S. (2014), “Instrumental Variable Methods for Causal Inference,” *Statistics in Medicine*, 33, 2297–2340. [1547]

Benkeser, D., and van der Laan, M. (2016), “The Highly Adaptive Lasso Estimator,” in *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, IEEE, pp. 689–696. [1546]

Bickel, P., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993), *Efficient and Adaptive Estimation for Semiparametric Models* (Vol. 4), New York: Springer-Verlag. [1544]

Brumback, B. A., Hernán, M. A., Haneuse, S. J. P. A., and Robins, J. M. (2004), “Sensitivity Analyses for Unmeasured Confounding Assuming a Marginal Structural Model for Repeated Measures,” *Statistics in Medicine*, 23, 749–767. [1540]

Chernozhukov, V., Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., and Newey, W. K. (2016), “Double Machine Learning for Treatment and Causal Parameters,” Technical Report, Cemmap Working Paper. [1544,1545]

Chernozhukov, V., Fernandez-Val, I., and Galichon, A. (2009), “Improving Point and Interval Estimators of Monotone Functions by Rearrangement,” *Biometrika*, 96, 559–575. [1545,1546]

Cinelli, C., and Hazlett, C. (2020), “Making Sense of Sensitivity: Extending Omitted Variable Bias,” *Journal of the Royal Statistical Society, Series B*, 82, 39–67. [1548]

Connors, A. F., Speroff, T., Dawson, N. V., Thomas, C., Harrell, F. E., Wagner, D., Desbiens, N., Goldman, L., Wu, A. W., Califf, R. M., and Fulkerson, W. J. (1996), “The Effectiveness of Right Heart Catheterization in the Initial Care of Critically Ill Patients,” *Jama*, 276, 889–897. [1547,1548]

Cornfield, J., Haenszel, W., Hammond, E. C., Lilienfeld, A. M., Shimkin, M. B., and Wynder, E. L. (1959), “Smoking and Lung Cancer: Recent Evidence and a Discussion of Some Questions,” *Journal of the National Cancer Institute*, 22, 173–203. [1540]

de Oliveira, L. M. F. T., dos Santos, A. R. M., Farah, B. Q., Ritti-Dias, R. M., de Freitas, C. M. S. M., and Diniz, P. R. B. (2019), “Influence of Parental Smoking on the Use of Alcohol and Illicit Drugs Among Adolescents,” *Einstein (São Paulo)*, 17, EAO4377. [1541]

Diaz, I., and van der Laan, M. J. (2013), “Sensitivity Analysis for Causal Inference Under Unmeasured Confounding and Measurement Error Problems,” *The International Journal of Biostatistics*, 9, 149–160. [1542]

Ding, P., and VanderWeele, T. J. (2016), “Sensitivity Analysis Without Assumptions,” *Epidemiology*, 27, 368. [1540,1544]

Farrell, M. H. (2015), “Robust Inference on Average Treatment Effects With Possibly More Covariates Than Observations,” *Journal of Econometrics*, 189, 1–23. [1546]

Gastwirth, J. L., Krieger, A. M., and Rosenbaum, P. R. (1998), “Dual and Simultaneous Sensitivity Analysis for Matched Pairs,” *Biometrika*, 85, 907–920. [1540,1544]

Györfi, L., Kohler, M., Krzyzak, A., and Walk, H. (2006), *A Distribution-Free Theory of Nonparametric Regression*, New York: Springer. [1546]

Hernán, M. A., and Robins, J. M. (2006), “Instruments for Causal Inference: An Epidemiologist’s Dream?” *Epidemiology*, 17, 360–372. [1547]

Horowitz, J. L. (2009), *Semiparametric and Nonparametric Methods in Econometrics* (Vol. 12), New York: Springer. [1546]

Horowitz, J. L., and Manski, C. F. (1995), “Identification and Robustness With Contaminated and Corrupted Data,” *Econometrica: Journal of the Econometric Society*, 63, 281–302. [1543]

Imbens, G. W. (2003), “Sensitivity to Exogeneity Assumptions in Program Evaluation,” *American Economic Review*, 93, 126–132. [1540]

Imbens, G. W., and Manski, C. F. (2004), “Confidence Intervals for Partially Identified Parameters,” *Econometrica*, 72, 1845–1857. [1546,1548,1549]

Joffe, M. M., Yang, W. P., and Feldman, H. I. (2010), “Selective Ignorability Assumptions in Causal Inference,” *The International Journal of Biostatistics*, 6, 11. [1541,1543]

Kandasamy, K., and Yu, Y. (2016), “Additive Approximations in High Dimensional Nonparametric Regression via the SALSA,” in *International Conference on Machine Learning*, pp. 69–78. [1546]

Kennedy, E. H. (2019), “Nonparametric Causal Effects Based on Incremental Propensity Score Interventions,” *Journal of the American Statistical Association*, 114, 645–656. [1546]

Kennedy, E. H., Balakrishnan, S., and G’Sell, M. (2018), “Sharp Instruments for Classifying Compliers and Generalizing Causal Effects,” arXiv no. 1801.03635. [1545]

Kennedy, E. H., Harris, S., and Keele, L. J. (2019), “Survivor-Complier Effects in the Presence of Selection on Treatment, With Application to a Study of Prompt ICU Admission,” *Journal of the American Statistical Association*, 114, 93–104. [1542,1545]

Lammert, C., Nguyen, D. L., Juran, B. D., Schlicht, E., Larson, J. J., Atkinson, E. J., and Lazaridis, K. N. (2013), “Questionnaire Based Assessment of Evidence and a Discussion of Some Questions,” *Journal of the Royal Statistical Society, Series B*, 85, 179–195. [1548]
Rosenbaum, P. R. (1987), “Sensitivity Analysis for Unobserved Confounding in Nonexperimental Prevention Research,” *Prevention Science*, 14, 570–580. [1540]

Luedtke, A. R., Diaz, I., and van der Laan, M. J. (2015), “The Statistics of Sensitivity Analyses,” U.C. Berkeley Division of Biostatistics Working Paper Series 341. [1544]

Luedtke, A. R. and van der Laan, M. J. (2016), “Statistical Inference for the Mean Outcome Under a Possibly Non-Unique Optimal Treatment Strategy,” *The Annals of Statistics*, 44, 713. [1545]

Pengpid, S., and Peltzer, K. (2019), “Alcohol Use and Misuse Among School-Going Adolescents in Thailand: Results of a National Survey in 2015,” *International Journal of Environmental Research and Public Health*, 16, 1898. [1541]

Raskutti, G., Wainwright, M. J., and Yu, B. (2012), “Minimax-Optimal Rates for Sparse Additive Models Over Kernel Classes via Convex Programming,” *Journal of Machine Learning Research*, 13, 389–427. [1546]

Richardson, A., Hudgens, M. G., Gilbert, P. B., and Fine, J. P. (2014), “Nonparametric Bounds and Sensitivity Analysis of Treatment Effects,” *Statistical Science: A Review Journal of the Institute of Mathematical Statistics*, 29, 596. [1540,1542]

Richardson, T. S. and J. M. Robins (2010), “Analysis of the Binary Instrumental Variable Model,” in R. Dechter, H. Geffner, and J. Halpern (Eds.), *Heuristics, Probability and Causality: A Tribute to Judea Pearl*, Chapter 25, pp. 415–444. London: College Publications. [1542]

Robins, J. M. (2002), “Comment on ‘Covariance adjustment in randomized experiments and observational studies’ by Paul Rosenbaum,” *Stat Sci.*, 17, 286–327. [1545]

Robins, J. M. (2002), “[Covariance Adjustment in Randomized Experiments and Observational Studies]: Comment,” *Statistical Science*, 17, 309–321. [1544]

Rosenbaum, P. R. (1987), “Sensitivity Analysis for Certain Permutation Inferences in Matched Observational Studies,” *Biometrika*, 74, 13–26. [1540,1544]

——— (2002), “Covariance Adjustment in Randomized Experiments and Observational Studies,” *Statistical Science*, 17, 286–327. [1540]

——— (2006), “Differential Effects and Generic Biases in Observational Studies,” *Biometrika*, 93, 573–586. [1541,1543]

Rosenbaum, P. R., and Rubin, D. B. (1983), “Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study With Binary Outcome,” *Journal of the Royal Statistical Society*, Series B, 45, 212–218. [1540]

Rotnitzky, A., Scharfstein, D., Su, T.-L., and Robins, J. (2001), “Methods for Conducting Sensitivity Analysis of Trials With Potentially Nonignorable Competing Causes of Censoring,” *Biometrics*, 57, 103–113. [1540]

Rubin, D. B. (1974), “Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies,” *Journal of Educational Psychology*, 66, 688. [1542]

Tsitatis, A. (2007), *Semiparametric Theory and Missing Data*, New York: Springer. [1544]

van der Laan, M. J. (2017), “A Generally Efficient Targeted Minimum Loss Based Estimator Based on the Highly Adaptive Lasso,” *The International Journal of Biostatistics*, 13, 297. [1546]

van der Laan, M. J., Laan, M. J., and Robins, J. M. (2003), *Unified Methods for Censored Longitudinal Data and Causality*, New York: Springer. [1544]

van der Laan, M. J., and Luedtke, A. R. (2014), “Targeted Learning of an Optimal Dynamic Treatment, and Statistical Inference for Its Mean Outcome,” U.C. Berkeley Division of Biostatistics Working Paper Series 329. [1545]

van der Laan, M. J., Polley, E. C., and Hubbard, A. E. (2007), “Super Learner,” *Statistical Applications in Genetics and Molecular Biology*, 6, 25. [1547]

van der Vaart, A. W. (2002), “Semiparametric Statistics,” in *Lectures on Probability Theory and Statistics*, eds. S. Tavaré and J. Picard, Springer, pp. 331–457. [1544]

van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes With Application to Statistics*, New York: Springer-Verlag. [1546]

VanderWeele, T. J., and Ding, P. (2017), “Sensitivity Analysis in Observational Research: Introducing the E-Value,” *Annals of Internal Medicine*, 167, 268–274. [1540,1544]

Yadlowsky, S., Namkoong, H., Basu, S., Duchi, J., and Tian, L. (2018), “Bounds on the Conditional and Average Treatment Effect in the Presence of Unobserved Confounders,” arXiv no. 1808.09521. [1540,1544]

Yang, Y., and Tokdar, S. T. (2015), “Minimax-Optimal Nonparametric Regression in High Dimensions,” *The Annals of Statistics*, 43, 652–674. [1546]

Zhang, B., and Tchetgen Tchetgen, E. J. (2019), “A Semiparametric Approach to Model-Based Sensitivity Analysis in Observational Studies,” arXiv no. 1910.14130. [1540]

Zhao, Q., Small, D. S., and Bhattacharya, B. B. (2017), “Sensitivity Analysis for Inverse Probability Weighting Estimators via the Percentile Bootstrap,” arXiv no. 1711.11286. [1540,1544]

Zheng, W., and van der Laan, M. J. (2010), “Asymptotic Theory for Cross-Validated Targeted Maximum Likelihood Estimation,” U.C. Berkeley Division of Biostatistics Working Paper Series 273. [1545]