UNIVERSITY OF CALIFORNIA
RIVERSIDE

Sensitivity Analysis of Unmeasured Confounding in Causal Inference based on Exponential Tilting and Super Learner

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Applied Statistics

by

Mi Zhou

June 2021

Dissertation Committee:

Dr. Weixin Yao, Chairperson
Dr. Analisa Flores
Dr. Esra Kurum

The Dissertation of Mi Zhou is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

It is a blessing for me to have such an exciting, challenging, and fulfilling Ph.D life. I would like to express my sincere gratitude to my advisor, Prof. Weixin Yao. When I felt frustrated and almost gave up, his guidance and support pushed me through and lit up the darkness. When I was willing to accept an initial solution, he would always encourage me to pursue for a better one, which has influenced me fundamentally. I would not have come this far without him. I cannot thank him enough.

The faculty in my department, and particularly my dissertation committee members had a large impact on me. I am grateful to my dissertation committee memebers, Prof. Esra Kurum and Dr. Analisa Flores, for their support and encouragement. I am also thankful to my former advisor, Dr. Zhiwei Zhang, who had led me to this path and opened the door to the wonderful world of causal inference for me. I would also like to thank all the professors and faculty in the Statistics department. The knowledge I learned from them makes up the foundation of this dissertation and helped me to develop my passion.

I also could not have completed this dissertation without the support of my friends. I want to thank all my friends from the Statistics department, especially Arnab Chowdhury, Judy Yun, Naijin Zhang, Song Zhai, Jinhui Yang, Lin Cong, Luke Klein, Fangjie Xie and my friend from the Economics department Ran Wang for their friendship. They have inspired me in both statistics and life. I also want to thank my best friends Lika Liu, Xiaolong Wang and my cousin Xian Yin for their trust and having always been there for me.

Finally, I am extremely thankful to my dad Yuan Zhou, my mom Cuilin Zheng and my fiance Sakar Sigdel for their unconditional love and for letting me shine.

To my beloved family, without whose support I would never have been here.

ABSTRACT OF THE DISSERTATION

Sensitivity Analysis of Unmeasured Confounding in Causal Inference based on Exponential Tilting and Super Learner

by

Mi Zhou

Doctor of Philosophy, Graduate Program in Applied Statistics
University of California, Riverside, June 2021
Dr. Weixin Yao, Chairperson

Causal inference under the potential outcome framework relies on the strongly ignorable treatment assumption. This assumption is usually unverifiable,and therefore we will never know if there is unmeasured confounding. Unmeasured confounding is one of the fundamental challenges in causal inference. Sensitivity studies are often used to address this issue and evaluate how sensitive a causal estimate is to the unmeasured confounder. In this dissertation, we propose a new sensitivity analysis method to evaluate the impact of the unmeasured confounder by combining ideas of the doubly robust estimator, the exponential tilt method, and the super learner algorithm for both binary and continuous outcomes in chapters 2 and 3, respectively. Compared to other existing methods of sensitivity analysis that parameterize the unmeasured confounder as a latent variable in the working models, the exponential tilting method does not impose any restrictions on the structure or models of the unmeasured confounders. Therefore, the unmeasured confounder could be continuous, binary, or categorical, and could be univariate or multivariate. In order to reduce the modeling bias of traditional parametric methods, we propose incorporating the *super learner machine learn-*

ing algorithm to perform nonparametric model estimation and the corresponding sensitivity analysis. In addition, we employ the data-driven trimming method to handle the estimated extreme propensity scores that can hamper the performance of the proposed doubly robust estimator. Furthermore, most existing sensitivity analysis methods require multivariate sensitivity parameters, which makes it more difficult and subjective to specify a reasonable range of the sensitivity parameters in practice. However, the new method has a univariate sensitivity parameter with a nice and simple interpretation of log-odds ratios and deviation in the conditional means of the outcomes for binary and continuous outcomes respectively. This makes choosing a range for the sensitivity parameter easier for the application. The simulation studies demonstrate the effectiveness of the proposed method.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Important research questions in biomedical, epidemiological, economic, social and behavioral sciences are frequently not associational but causal in nature. For example, what is the efficacy of a new drug in a target population of patients? Does a particular microorganism cause a particular disease? What fraction of past crimes could have been avoided by a given policy? Causal inference is the process of learning about causality from data with the help of subject matter knowledge. Understanding the causal effect of a treatment or exposure is an important goal in various scientific studies.

1.1 Potential Outcomes

A common framework for causality research is based on potential outcomes, which was originally proposed by Neyman 1923 in the context of completely randomized experiments and later Rubin 1974 extended it to both observational and experimental studies. Most of the existing methods for causal inference can be cast in terms of potential outcomes.

Suppose we are interested in the causal effect of some treatment T on some outcome Y. The potential outcome for a given treatment is the outcome that would result under that treatment. For each possible value t of T, let $Y(t)$ be the outcome for treatment t, i.e., the outcome that would be observed if, possibly contrary to fact, treatment was set to $T = t$. In this proposal, we consider a binary treatment, with $t = 1$ indicating the treatment of interest and $t = 0$ indicating a control treatment. Thus, in the target population, each person has two potential outcomes: $Y(1)$ and $Y(0)$. The potential outcomes are usually not completely observed. Typically, we can only observe either $Y(1)$ or $Y(0)$, depending on which treatment is actually given.

1.2 Causal effect

1.2.1 Average treatment effect

For a population of interest, the average treatment effect is defined as the difference between the population average value of Y if everyone was treated with $T = 1$ and the population average value of Y if everyone was treated with $T = 0$: $\delta = \mu_1 - \mu_0$, where $\mu_t = E\{Y(t)\}$, $t = 0, 1$.

1.2.2 Other causal effects

In addition to the average treatment effect, there are many other causal effects that might be of interest. For example, we might be interested in the difference in a specified quantile between two potential outcome distributions. For binary outcomes, a causal effect

could be defined as the relative risk $P\{Y(1)=1\}/P\{Y(0)=1\}$, or the odds ratio:

$$\frac{P\{Y(1)=1\}}{P\{Y(1)=0\}}/\frac{P\{Y(0)=1\}}{P\{Y(0)=0\}} = P\{Y(1)=1\}P\{Y(0)=0\}/[P\{Y(1)=0\}P\{Y(0)=1\}].$$

Sometimes interest lies in the average treatment effect on the treated, $E\{Y(1)-Y(0)|T=1\}$, which measures the average effect of treatment among those who are actually treated. This may be of interest if the untreated individuals would never be interested in this particular treatment (for example, some people are just unable or unwilling to have surgery). For those people who want to be treated, we would like to understand the average effect of the treatment in that subpopulation.

The causal effect of interest depends on the scientific context – the research question, the design of a study, and possibly the data available. For later discussions in this proposal, we focus on the average treatment effect.

1.3 Confounding

Randomized experiments are considered the gold standard for causal inference as in a randomized study subjects are randomly assigned to treatment groups, which makes the treatment assignment independent of the potential outcomes. Outcomes in different treatment groups can be compared directly, because the subjects are not systematically different. Formally, randomization implies that T is independent of $Y(1)$ and $Y(0)$, both of

which are baseline variables (which exist before randomization occurs). It follows that

$$E(Y|T=t) = E\{Y(t)|T=t\} = E\{Y(t)\} \quad t = 0, 1. \tag{1.1}$$

Thus, the average outcome in each treatment group estimates the mean of the corresponding potential outcome.

In some situations, randomization may be unethical or impractical, and one has to conduct an observational study in which treatment assignment is observed but not controlled by investigators. Without randomization, treatment assignment is not known to be independent of the potential outcomes, and equation (1.1) is generally false. This issue, known as confounding, is the main challenge in understanding causality from observational data. To adjust for confounding in an observational study, we first need to identify the confounders, the variables that cause the confounding. These are usually identified as the baseline variables that are predictive of both treatment assignment and the outcome of interest.

1.4 Standard Assumptions

1.4.1 Strongly Ignorable (SI) Treatment Assignment

A standard assumption for causal inference is the assumption of strongly ignorable treatment assignment. Let X be a collection of measured baseline covariates which are considered possible confounders. The strong ignorability assumption says that the treatment

assignment T is conditionally independent of the potential outcomes $Y(t)$ given the value of X, written

$$T \perp Y(t)|X, \quad t = 0, 1. \tag{1.2}$$

In other words, within each level of X, T is assumed to be independent of the potential outcomes, so we have a quasi-randomized experiment. This assumption is not testable with observed data and must be based on external information such as subject matter knowledge. In practice, to make the strong ignorability assumption plausible, one needs to identify and measure all baseline variables that are important predictors of treatment assignment and the outcome of interest.

1.4.2 Positivity Assumption

Another standard assumption for causal inference is the positivity assumption:

$$P(T = t|X = x) > 0, \quad \forall\ t,\ x, \tag{1.3}$$

which requires that both treatments be possible at each level of X. In other words, there cannot be a subpopulation defined by X that never receives one of the treatments. Unlike assumption(3.1), assumption (1.3) is testable with observed data. Without this assumption, treatment assignment would be deterministic for some values of X, and there would be no observed outcome data in one treatment group for those values of X. This would have a negative impact on parameter identification, as we now discuss.

Together, assumptions (3.1) and (1.3) are sufficient for nonparametric identification of μ_t $(t = 0, 1)$ and hence δ; that is, the parameters can be expressed in terms of the observable (X, T, Y) under the two assumptions. To see this, note that assumption (3.1) implies

$$E\{Y(t)|X\} = E\{Y(t)|T = t, X\} = E(Y|T = t, X),$$

which further implies that

$$\mu_t = E\{E(Y|T = t, X)\},$$

where the outer expectation is with respect to the marginal distribution of X. Assumption (1.3) ensures that $\mathrm{E}(Y|T = t, X)$ is identified from the observed data. If the positivity assumption fails, so that $\mathrm{P}(T = t|X = x) = 0$ $\forall x \in \mathcal{X}_0$ for some $\mathcal{X}_0$ with $\mathrm{P}(X \in \mathcal{X}_0) > 0$, then $\mathrm{E}(Y|T = t, X = x)$ is unidentified for $x \in \mathcal{X}_0$.

1.5 Existing methods assuming SI

Suppose an observational study is conducted on a random sample of size n from the target population. The observed data will be conceptualized as independent copies of (X, T, Y), and will be denoted by (X_i, T_i, Y_i), $i = 1, \ldots, n$. The goal is to estimate $\delta = \mu_1 - \mu_0$ using the observed data. In this section, we describe several methods that assume SI and positivity.

1.5.1 Outcome Regression

The previous discussion of identification suggests an estimation approach based on the regression function $\mathrm{E}(Y|T,X)$. Estimator obtained based on this method is referred to as the outcome regression estimator, and it has been widely used. Some well-known works include Paul R Rosenbaum and Rubin 1982, Imbens 2003 and Tan 2006. If X is discrete with a small number of levels, one can estimate $\mathrm{E}(Y|T,X)$ non-parametrically with the sample mean of Y within each stratum defined by (T,X). When X has several components, some of which may be continuous, this nonparametric approach quickly becomes impractical due to the curse of dimensionality. One way to deal with the curse of dimensionality is to specify a model for $\mathrm{E}(Y|T,X)$, say $m(T,X;\beta)$, where m is a known function and β represents the unknown parameters to be estimated. Now the average treatment effect can be rewritten as:

$$\delta = \mathrm{E}\{m(1,X;\beta) - m(0,X;\beta)\}. \tag{1.4}$$

For example, the outcome regression model may be specified as a generalized linear model (GLM):

$$m(t,x,\beta) = \psi\{\beta_0 + \beta_X x + \beta_T t + \beta_{TX}(tx)\}, \tag{1.5}$$

where ψ is an inverse link function. In this model, the regression coefficients β_T and β_{TX} together describe the conditional effect of T on Y given X on the scale of the link function. Specifically, we have

$$\beta_T + \beta_{TX}X = \psi^{-1}(\mathrm{E}\{Y(1)|X\}) - \psi^{-1}(\mathrm{E}\{Y(0)|X\}).$$

The model is also related to the marginal effect through the relationship $\mu_t = \mathrm{E}\{m(t, X, \beta)\}$, where the expectation is with respect to the marginal distribution of X.

The outcome regression model can be fitted using standard techniques. Let $\hat{\beta}$ denote the resulting estimate of β, we can estimate μ_t $(t = 0, 1)$ with

$$\widehat{\mu}_t^{\mathrm{OR}} = \frac{1}{n} \sum_{i=1}^{n} m(t, X_i; \hat{\beta}),$$

and estimate δ with $\widehat{\delta}_{\mathrm{OR}} = \widehat{\mu}_1^{\mathrm{OR}} - \widehat{\mu}_0^{\mathrm{OR}}$.

1.5.2 Inverse Probability Weighting (IPW)

One way to look at confounding in observational studies is to regard each treatment group as a biased sample from the target population, with some subjects over-sampled and other under-sampled. For example, subjects with high propensity scores, which indicates a high probability of receiving the treatment, would be over-sampled in the treated group and under-sampled in the control group. Similarly, subjects with small propensity scores, which indicates a small probability of receiving the treatment, would be under-sampled in the treated group and over-samples in the control group. The idea of IPW is to create a pseudo-population where treatment is no longer dependent on confounders by weighting the available subjects in each treatment group.

The appropriate weight turns out to be the inverse of the probability of the treatment received, conditional on X:

For treated subjects: weighted by the inverse of $P(T = 1|X) = PS$.

For control subjects: weighted by the inverse of $P(T = 0|X) = 1 - PS$.

The IPW approach (P. Rosenbaum, Colton, and Armitage 1998; Lunceford and Davidian 2004; Robins, Hernan, and Brumback 2000) can be formally justified as follows:

$$\begin{aligned}
\mathrm{E}\left\{\frac{TY}{\pi(X)}\right\} &= \mathrm{E}\left\{\frac{TY(1)}{\pi(X)}\right\} \\
&= \mathrm{E}\left[\mathrm{E}\left\{\frac{TY(1)}{\pi(X)}\middle| X, Y(1)\right\}\right] \\
&= \mathrm{E}\left[\frac{\mathrm{E}\{T|X, Y(1)\}Y(1)}{\pi(X)}\right] \\
&= \mathrm{E}\left\{\frac{\mathrm{E}(T|X)Y(1)}{\pi(X)}\right\} \\
&= \mathrm{E}\left\{\frac{\pi(X)Y(1)}{\pi(X)}\right\} \\
&= \mathrm{E}\{Y(1)\} \\
&= \mu_1,
\end{aligned} \tag{1.6}$$

which follows from the SI and positivity assumptions. (The positivity assumption ensures that the denominator $\pi(X)$ is positive and can be cancelled with $\pi(X)$ in the numerator.) A similar argument shows that $\mu_0 = \mathrm{E}\left\{\frac{(1-T)Y}{1-\pi(X)}\right\}$.

In an observational study, the propensity score is typically unknown and must be estimated. Here again, $\pi(X)$ can be estimated nonparametrically if X is discrete with a small number of levels. In most realistic situations, to deal with the curse of dimensionality, we can estimate $\pi(X)$ under a regression model for T conditional on X, say $\mathrm{P}(T = 1|X) = \pi(X, \gamma)$. Because T is binary, the propensity score model is a binary regression model (e.g., logistic regression), which can be fitted using standard techniques. Let $\widehat{\gamma}$ denote the

resulting estimate of γ. Then we can estimate δ with $\widehat{\delta}_{\text{IPW}} = \widehat{\mu}_1^{\text{IPW}} - \widehat{\mu}_1^{\text{IPW}}$, where

$$\widehat{\mu}_1^{\text{IPW}} = \frac{1}{n}\sum_{i=1}^{n} \frac{T_i Y_i}{\pi(X_i, \widehat{\gamma})},$$
$$\widehat{\mu}_0^{\text{IPW}} = \frac{1}{n}\sum_{i=1}^{n} \frac{(1-T_i) Y_i}{1-\pi(X_i, \widehat{\gamma})}.$$

Each $\widehat{\mu}_t^{\text{IPW}}$ can be regarded as a weighted average of the observed outcomes in the tth treatment group. The total weight converges to 1 but is not necessarily equal to 1 in finite samples. For improved performance, we could divide the weights by the total weight in each treatment group. The resulting normalized IPW estimator generally performs better.

1.5.3 Double Robust (DR) Estimators

The OR and IPW estimators involve different models, and their consistency depends on correct specification of the OR model and the PS model, respectively. Misspecification of the working model generally results in a bias in the estimator. Some protection is provided by a DR estimator, which is consistent under correct specification of either or both of the OR and PS models. This DR property is a significant advantage because in practice it is often difficult to specify a regression model correctly.

Motivated by semiparametric theory, DR estimators (Robins, Rotnitzky, and Zhao 1995; Rotnitzky, Robins, and D. O. Scharfstein 1998; D. O. Scharfstein, Rotnitzky, and Robins 1999; Lunceford and Davidian 2004; Rotnitzky, Lei, et al. 2012) can be constructed in different ways. A common approach to DR estimation is to augment an IPW estimator or estimating equation with a term derived from semiparametric theory. In the present context, an augmented IPW (AIPW) estimator of δ is given by $\widehat{\delta}_{\text{AIPW}} = \widehat{\mu}_1^{\text{AIPW}} - \widehat{\mu}_0^{\text{AIPW}}$,

where

$$\widehat{\mu}_1^{\text{AIPW}} = \widehat{\mu}_1^{\text{IPW}} - \frac{1}{n}\sum_{i=1}^{n} \frac{T_i - \pi(X_i, \widehat{\gamma})}{\pi(X_i, \widehat{\gamma})} m(1, X_i, \widehat{\beta}),$$

$$\widehat{\mu}_0^{\text{AIPW}} = \widehat{\mu}_0^{\text{IPW}} - \frac{1}{n}\sum_{i=1}^{n} \frac{T_i - \pi(X_i, \widehat{\gamma})}{1 - \pi(X_i, \widehat{\gamma})} m(0, X_i, \widehat{\beta}).$$

The DR property of the AIPW estimator can be seen as follows. For ease of argument, let us focus on estimation of μ_1. Let β^* and γ^* denote the probability limits of $\widehat{\beta}$ and $\widehat{\gamma}$, respectively. If the PS model is correct, then $\gamma*$ equals the true value of γ, the IPW estimator is consistent, and the difference $\widehat{\mu}_1^{\text{AIPW}} - \widehat{\mu}_1^{\text{IPW}}$ converges to $\text{E}\left\{ \frac{T-\pi(X)}{\pi(X)} m(1, X, \beta^*) \right\}$, which is easily seen to be zero by using a conditioning argument. Next, suppose the OR model is correct, so that β^* equals the true value of β and the OR estimator is consistent. In this case, the difference

$$\widehat{\mu}_1^{\text{AIPW}} - \widehat{\mu}_1^{\text{OR}} = \frac{1}{n}\sum_{i=1}^{n} \frac{T_i\{Y_i - m(1, X_i, \beta)\}}{\pi(X_i, \gamma^*)}$$

converges to

$$\text{E}\left[\frac{T\{Y - m(1, X, \beta)\}}{\pi(X, \gamma^*)}\right],$$

which is again zero. Thus, the AIPW estimator is indeed DR—consistent if either model is correct. If both models are correct, the AIPW estimator attains the nonparametric information bound; in that sense, the AIPW estimator is said to be locally efficient.

There are other approaches, such as targeted maximum likelihood estimation, to obtaining DR estimators with similar (or better) properties.

1.5.4 Propensity Score Stratification

Paul R Rosenbaum and Rubin 1983 point out that

$$T \perp Y(t)|\pi(X), \qquad t = 0, 1.$$

Thus, upon conditioning on the PS, T becomes independent of the potential outcomes so we have a quasi-randomized experiment within each subpopulation defined by $\pi(X)$. This observation motivates a PS stratification approach in which subjects are grouped into a few strata based on their estimated PS values. Let S_k, $k = 1, \ldots, K$, be a partition of the unit interval based on quantiles of the estimated PS $\pi(X_i, \widehat{\gamma})$. If each stratum is approximately homogeneous, we can treat T as approximately independent of the potential outcomes with the stratum, and estimate $\mu_t^{(k)} = \mathrm{E}\{Y(t)|\pi(X) \in S_k\}$ with $\widehat{\mu}_t^{(k)}$, the average of the observed Y-values for treatment t in stratum k. The stratified estimator of δ is $\widehat{\delta}_{\mathrm{ST}} = \widehat{\mu}_1^{\mathrm{ST}} - \widehat{\mu}_0^{\mathrm{ST}}$ with

$$\widehat{\mu}_t^{\mathrm{ST}} = \frac{1}{n} \sum_{k=1}^{K} |S_k| \widehat{\mu}_t^{(k)},$$

where $|S_k|$ is the size of S_k. This PS stratification approach is simple and easy to implement. On the other hand, because the strata are not exactly homogeneous, there may be a bias in $\widehat{\delta}_{\mathrm{ST}}$ due to residual confounding.

1.6 Existing methods dealing with possible violation of SI

The methods discussed above are valid in cases where there are no unmeasured confounders, which, practically, is a strong assumption. Researchers may hope that sufficiently

rich baseline information is collected in order to justify this assumption. In reality, the assumption is questionable in many observational studies. If unmeasured confounders exist, the strong ignorability assumption may be violated, which may result in a bias in treatment effect estimation and undermine the validity and credibility of the conclusions drawn. Unfortunately, the strong ignorability assumption cannot be validated with observed data, and there is frequently insufficient background knowledge to justify this assumption. Therefore, it is important to consider possible violations of the strong ignorability assumption.

With the strong ignorability assumption removed, the parameters μ_t and δ become un-identified, which has motivated alternative assumptions of various forms that aim to recover identifiability. These alternative assumptions are, like the strong ignorability assumption, untestable with observed data. A single set of alternative identifying assumptions may be as questionable as, or more questionable than, the strong ignorability assumption. To address such uncertainty, it is common to conduct a sensitivity analysis that considers a variety of identifying assumptions and compares results obtained under different assumptions. In fact, most existing methods that deal with violations of the SI assumption can be regarded as sensitivity analysis methods.

In the rest of this section, we describe two existing sensitivity analysis methods of Paul R Rosenbaum and Rubin 1982 and Lin, Psaty, and Kronmal 1998, as well as a E-value approach of Ding and VanderWeele 2016, which represents a different perspective.

1.6.1 Sensitivity Analysis Method of Rosenbaum and Rubin (1982)

Paul R Rosenbaum and Rubin 1982 proposed a sensitivity analysis method that explicitly adjusts for an unmeasured confounder as a latent variable. With U denoting the

unmeasured confounder, their method assumes that X and U together satisfy the SI and positivity assumptions:

$$T \perp Y(t)|X,U, \qquad t=0,1; \tag{1.7}$$

$$0 < P(T=0|X,U) < 1, \qquad \text{with probability 1.} \tag{1.8}$$

In addition, their method assumes that X is finitely discrete and that Y and U are both binary.

Under assumption (1.7), the joint distribution of $(Y(t),T,U,X)$ for $t=0,1$ can be factorized as:

$$P(Y(t)|U,X)P(T|U,X)P(U|X)P(X),$$

where the SI assumption for (X,U) allows us to write $P(Y(t)|T,U,X)$ as $P(Y(t)|U,X)$. Write $\phi_j = P(X=j)$ and $\pi_j = P(U=0|X=j), \quad j=1,...J$. Further, let $P(T|U,X)$ and $P(Y(t)|U,X)$ be parameterized as logistic regression models:

$$P(T=0|U=u,X=j) = [1+exp(\gamma_j + u\alpha_j)]^{-1},$$

and

$$P(Y(t)=0|U=u,X=j) = [1+exp(\beta_{jt} + u\eta_{jt})]^{-1}.$$

Note that both models are saturated and impose no real assumptions on the conditional probabilities on the left side. The above expressions are really just to parameterize the two

conditional probabilities. When $u = 0$, γ_j is the log odds of treatment 0 in subclass j and when $u = 1$, $\gamma_j + \alpha_j$ is the log odds of treatment 0 in subclass j. Likewise, β_{jt} is the log odds of $Y(t)$ in subclass j when $u = 0$, and $\beta_{jt}+\eta_{jt}$ is the corresponding log odds when $u = 1$.

Under this parameterization, μ_t can be expressed as

$$\mu_t = \sum_{j=1}^{J} \phi_j [(1 - \pi_j) \frac{exp(\beta_{jt} + \eta_{jt})}{1 + exp(\beta_{jt} + \eta_{jt})} + \pi_j \frac{exp(\beta_{jt})}{1 + exp(\beta_{jt})}]. \tag{1.9}$$

If all parameters on the right side can be estimated, we can then substitute their estimates into the above expression to estimate μ_t and δ. However, the parameters are not fully identified because U is unobserved. the proposal of Paul R Rosenbaum and Rubin 1982 is to specify the values of the parameters associated with the unmeasured confounding variable u (i.e., $\pi_j, \alpha_j, \eta_{jt}$). Once these values are specified, we can then estimate the other parameters by maximizing the likelihood for the observed data (X_i, T_i, Y_i), $i = 1, \ldots, n$. Varying the specified values of $(\pi_j, \alpha_j, \eta_{jt})$ leads to a sensitivity analysis.

1.6.2 Sensitivity Analysis Method of Lin et al. (1998)

Lin, Psaty, and Kronmal 1998 proposed an approach to assess the sensitivity of the regression coefficients in an outcome regression model adjusting for both the measured and unmeasured confounders. Because the regression coefficients describe the conditional effect of T on Y given confounders, the approach of Lin, Psaty, and Kronmal 1998 is targeted at the conditional effect and not the marginal effect, which is the focus of this proposal.

Again, let U denote an unmeasured confounder. The main idea of Lin, Psaty, and

Kronmal 1998 is to relate a regression model for $\mathrm{E}(Y|T,X,U)$ to a regression model for $\mathrm{E}(Y|T,X)$ of a similar form. A key observation here is that

$$\mathrm{E}(Y|T,X) = \mathrm{E}\{\mathrm{E}(Y|T,X,U)|T,X\}.$$

To deduce $\mathrm{E}(Y|T,X)$ from a given model for $\mathrm{E}(Y|T,X,U)$ requires integration with respect to the conditional distribution of U given (T,X). To this end, the authors assume that

$$U \perp X|T, \tag{1.10}$$

that is, unmeasured confounders are independent of measured confounders within each treatment group. Under this assumption, the conditional distribution of U given (T,X) is just the conditional distribution of U given T, which is easier to specify because T is binary.

As a concrete example, suppose Y is binary and consider the following log-linear model:

$$P(Y(t)=1|T=t,X,U) = P(Y=1|T=t,X,U) = e^{\alpha+\beta T+\gamma_t U+\theta' X}, \tag{1.11}$$

where $\alpha,\ \beta,\ \gamma_t(t=0,1)$ and θ are unknown regression parameters. The notation γ_t implies that T is allowed to interact with U.

Under assumption (1.10), model (1.11) implies that

$$P(Y=1|X,T) = e^{\alpha^*+\beta^* T+\theta^{*'} X}, \tag{1.12}$$

where θ^* is the same as θ but (α^*, β^*) may differ from (α, β) in model (1.11). β represents the true conditional effect of the treatment given $(X, U = 0)$, while β^* is just a regression parameter without a causal interpretation. If we assume $\gamma_1 = \gamma_0$ so there is no interaction between T and U in model (1.11), then β represents the conditional effect of treatment given (X, U) (without restricting the value of U). Let $R = e^{\beta}$ and $R^* = e^{\beta^*}$, then R is the true relative risk of disease due to treatment whereas R^* is the relative risk estimable from the reduced model, i.e., the apparent relative risk. It is of interest to ascertain the relationship between β and β^* and between R and R^*. Suppose U is a binary unmeasured confounder:

$$U|T, X \sim Bernoulli(P_T).$$

Then, after some algebra, it can be shown that

$$\beta = \beta^* - log\frac{e^{\gamma_1}P_1 + (1 - P_1)}{e^{\gamma_0}P_0 + (1 - P_0)}, \tag{1.13}$$

where $P_1 = P_{T=1} = P(U = 1|T = 1)$ and $P_0 = P_{T=0} = P(U = 1|T = 0)$;

or

$$R = R^*/A; \tag{1.14}$$

$$A = \frac{e^{\gamma_1}P_1 + (1 - P_1)}{e^{\gamma_0}P_0 + (1 - P_0)}.$$

A is called the adjustment factor and it involves the unmeasured confounder U. In applications, A can be evaluated for some specified values of the sensitivity parameters that

are associated with U (i.e., $P_0, P_1, e^{\gamma_1}, e^{\gamma_0}$). Lin, Psaty, and Kronmal 1998 also considered several other models and distributions of U, many of which admit a simple characterization of the relationship between β and β^*.

This is an elegant and interesting approach. However, as pointed out by VanderWeele 2008 , assumption (1.10) is implausible if X and U are both related to T, as is typically the case when X and U are both important confounders.

1.6.3 E-value

The two sensitivity analysis methods described earlier require specifying the values of certain parameters related to the unmeasured confounder, which may be difficult and arbitrary. There are other sensitivity analysis methods that require different specifications, which also tend to be arbitrary.

VanderWeele and Ding 2017 takes a different perspective. Instead of exploring alternative assumptions for identification, they ask how strong an unmeasured confounder has to be in order to qualitatively change the conclusion of a causal analysis. They assume a binary outcome and use the causal relative risk (or risk ratio) defined earlier as the effect measure. They introduce an E-value, defined as the minimum strength of association, on the risk ratio scale, that an unmeasured confounder would need to have with both the treatment and outcome, conditional on the measured covariates, to explain away a treatment-outcome

association. More precisely,

$$E-value = \begin{cases} R + \sqrt{R \times (R-1)}, & if R > 1, \\ R^* + \sqrt{R^* \times (R^*-1)}, & if\ R^* = \frac{1}{R} > 1 \end{cases}$$

where R is the risk ratio of an observational study after adjusting for several observed confounding covariates. The E-value is a continuous measure of the robustness of an observed association to unobserved confounders. The smallest possible E-value is 1, which corresponds to $R = 1$ and indicates no evidence for a causal relationship. Higher E-values indicate stronger evidence for a causal relationship.

Chapter 2

Sensitivity Analysis of Unmeasured Confounding in Causal Inference for Binary Outcomes

2.1 Introduction

In order to explore causal inference, Neyman 1923 proposed the potential outcome framework in the context of completely randomized experiments and later Rubin 1974 extended it to both observational and experimental studies. Consider a binary treatment T with $T = 1$ indicating the treatment group and $T = 0$ indicating the control group. The potential outcome $Y(t)$ is defined as the outcome we would observe if a subject had been assigned the treatment $T = t,\ t = 0, 1$. The observed outcome will then be $Y = TY(1)+(1-T)Y(0)$. The causal effect, or average treatment effect (ATE), of the treatment is defined as

$\tau = \mu_1 - \mu_0$, where $\mu_t = E\{Y(t)\}, t = 0, 1$. In observational studies, a standard assumption for causal inference is the assumption of strongly ignorable treatment assignment. Let X be a collection of measured baseline covariates which are considered possible confounders. The strong ignorability assumption states that the treatment assignment T is conditionally independent of the potential outcomes $Y(0)$ and $Y(1)$ given X, i.e.,

$$Y(t) \perp T|X, t = 0, 1. \tag{2.1}$$

In other words, within each level of X, T is assumed to be independent from the potential outcomes, so we have a quasi-randomized experiment.

Researchers usually hope that sufficiently rich baseline information is collected so that the ignorability assumption (3.1) is reasonable. However, this assumption is usually questionable in observational studies, and the *unmeasured confounding* is one of the fundamental challenges in causal inference. If unmeasured confounders exist, the strong ignorability assumption is violated, which may result in a biased treatment effect estimation and undermine the validity and credibility of the corresponding causal inference. Unfortunately, the strong ignorability assumption cannot be validated with observed data for the estimation of causal effect, and there is frequently insufficient background knowledge to justify this assumption.

Without the strong ignorability assumption (3.1), the causal effect becomes unidentified, which has motivated alternative assumptions of various forms that aim to recover the identifiability. However, these alternative assumptions are also untestable with observed data like the strong ignorability assumption, and may be as questionable as, or even

more questionable than, the strong ignorability assumption. To address such uncertainty, it is important to conduct a sensitivity analysis that considers a variety of identifiability assumptions and compare results obtained under different assumptions.

The history of sensitivity analysis can be dated back to the work of Cornfield et al. 1959 which explored a causal link between smoking and lung cancer. Paul R Rosenbaum and Rubin 1982 proposed a sensitivity analysis framework that explicitly included a binary unmeasured confounding variable U as a latent variable in regression models for the binary treatment T and the binary outcome $Y(t)$. Their method assumed that X and U together satisfy the strong ignorability assumption, i.e., $Y(t) \perp T|X, U$. More specifically, the observed data consists of (Y, T, X) and a binary unmeasured confounder U is assumed for each individual. Suppose that the observed data and the unmeasured confounder were generated according to:

$$
\begin{aligned}
U_i &\sim \text{Ber}(P_u) \\
T_i|X_i, U_i &\sim \text{Ber}\left(\text{expit}\{\gamma_x X + \gamma_u U\}\right) \\
Y_i|T_i, X_i, U_i &\sim \text{Ber}\left(\text{expit}\{\beta_t T + \beta_x X + \beta_u U\}\right),
\end{aligned}
$$

where $\text{Ber}(\pi)$ is a Bernoulli distribution with a success probability π, $\text{expit}(s) = (1 + e^{-s})^{-1} = e^s/(1 + e^s)$. If we could observe U, the parameters $(P_u, \gamma_x, \gamma_u, \beta_t, \beta_x, \beta_u)$ can be estimated via logistic regression models and report τ as the causal effect. The two coefficients γ_u and β_u measure the strength of the association between the unmeasured confounder and the treatment and the association between the unmeasured confounder and the outcome, respectively. However, U is not observed, thus P_u, γ_u and β_u are not

identifiable using the observed data. What analysts can do is to specify plausible values of (P_u, γ_u, β_u) based on their subjective judgments about these parameters, then the other parameters can be estimated based on (P_u, γ_u, β_u). The final estimate of the causal effect can be expressed as $\hat{\tau}(P_u, \gamma_u, \beta_u)$. Veitch and Zaveri 2020 pointed out that this approach has a major drawback: it relies on a parametric model for the full data generating process. Using the assumed model is equivalent to assuming that if U had been observed, it would have been appropriate to use logistic regression to model the treatment assignment and outcome. Also, the sensitivity analysis result depends on the distribution of U. Furthermore, the choice of the sensitive parameters (P_u, γ_u, β_u) is subjective and challenging, and could be even more difficult if U is not univariate. Lin, Psaty, and Kronmal 1998 proposed an approach to parameterize the unmeasured confounder as the bias of regression coefficients in an outcome regression model adjusting for both the measured and unmeasured confounders. Imbens 2003 extended Paul R Rosenbaum and Rubin 1982 to allow for non-binary outcomes. There is also a line of work that formulated the sensitivity parameter as the bias of the outcome regression model caused by the unmeasured confounder, see, for example, Hogan, M. Daniels, and Hu 2014; Jung et al. 2018; Roy, Lum, and Michael J Daniels 2017, etc. Ding and VanderWeele 2016 and VanderWeele and Ding 2017 investigated how strong an unmeasured confounder has to be in order to qualitatively change the conclusion of a causal analysis. In recent years, another extension of the traditional sensitivity analysis was from a Bayesian perspective which used the average over the distribution of sensitivity parameters rather than varying the sensitivity parameters McCandless, Gustafson, and Levy 2007; Dorie et al. 2016; McCandless and Gustafson 2017. Most of the above sensitivity analysis

methods require specifying some parametric models for the confounding variable, which may be difficult to hold in practice. In addition, they usually contain multivariate sensitivity parameters, which makes their choice difficult and subjective when performing sensitivity analysis in practice.

To this end, we propose a new sensitivity analysis method to evaluate the impact of the unmeasured confounder by leveraging ideas of doubly robust estimators (Robins, Rotnitzky, and Zhao 1995), the exponential tilt method D. Scharfstein et al. 2014, and the super learner machine learning method Van der Laan, Polley, and Hubbard 2007. Note that if there is any unmeasured confounder, $f(Y(t)|X, T = 1)$ and $f(Y(t)|X, T = 0)$ will be different, where f is a probability distribution function (either a density function for a continuous variable or a probability mass function for a discrete variable). Inspired by this, we propose to assess the sensitivity of the difference between the conditional distribution of the observed potential outcome given covariates and that of the counterfactual potential outcome via the exponential tilting method proposed by D. Scharfstein et al. 2014. Compared to most existing sensitivity analysis methods, the exponential tilting method does not directly impose any assumptions on the distribution of the unmeasured confounders. Therefore, the new method has the flexibility to allow the unmeasured confounder to be continuous, binary, or categorical, and be univariate or multivariate. In order to reduce the modeling bias of traditional parametric methods, we propose incorporating *super learner machine learning algorithms* to perform the nonparametric model estimation and the corresponding sensitivity analysis. Super learner algorithms aim to optimally combine many machine learning algorithms together to provide a better estimation than any individual

candidate machine learning algorithm. Unlike most of existing sensitivity analysis methods which usually contain multiple sensitivity parameters, the new method has a univariate sensitivity parameter, which directly measures the bias of the strong ignorability assumption (3.1) caused by the unmeasured confounding, and has a nice and simple interpretation of log-odds ratios for binary outcomes. The used univariate sensitivity parameter also makes its choice and the application of the new sensitivity analysis method very easy for practitioners.

The rest of this paper is structured as follows. In section 2, we introduce our proposed estimation method in detail. In sections 3 and 4, we present numerical examples based on both a simulation study and a real data application. A conclusion is given in section 5.

2.2 Methodology

In this article, we mainly focus on the binary response variable due to its wide variety of applications in causal inference Paul R Rosenbaum and Rubin 1982; McCandless, Gustafson, and Levy 2007; Groenwold et al. 2010. The proposed method can be easily extended to a continuous response variable. Note that for a binary response variable, the potential outcome mean $\mu_t = E\{Y(t)\}$ can be interpreted as the success rate among the target population if everyone had been assigned the treatment t. For simplicity of explanation, we introduce our new method by focusing on estimating $\mu_1 = E\{Y(1)\}$, the mean of potential outcome if everyone in the target population had been treated. The $\mu_0 = E\{Y(0)\}$ and the corresponding ATE $\tau = \mu_1 - \mu_0$ can be similarly estimated.

When the subjects are in the treatment group, $Y(1)$ is the observed actual outcome, whereas when the subjects are in the control group, $Y(1)$ becomes the counterfactual outcome which cannot be observed. So the estimation of μ_1 essentially boils down to the imputation of counterfactual outcomes. Let

$$m(x) = E(Y(1)|X = x),\ m_1(x) = E(Y(1)|T = 1, X = x)$$

and

$$P(x, y) = P(T = 1|X = x, Y(1) = y),\ q(x) = P(T = 1|X = x).$$

Then our target parameter can be also written as $\mu_1 = \{Y(1)\} = \{m(X)\}$. Notice that when there is no unmeasured confounder,

$$m(x) = m_1(x) \text{ and } P(x, y) = q(x), \tag{2.2}$$

which can be estimated from the observed data.

2.2.1 Doubly robust estimator

Let $(X_i, Y_i, T_i)_{i=1}^n$ be an independent and identically distributed random sample from $X \in \mathbb{R}^d$, $Y \in \{0, 1\}$ and $T \in \{0, 1\}$, where X is a set of baseline covariates, and T is a binary treatment indicator. Under the strong ignorability assumption (i.e., there is no unmeasured confounder), three of the most commonly used estimators for $\mu_1 = E\{Y(1)\}$ are

$$\hat{\mu}^{OR} = \frac{1}{n}\sum_{i=1}^{n} \hat{m}_1(x_i), \tag{2.3}$$

$$\hat{\mu}^{IPW} = \frac{1}{n}\sum_{i=1}^{n}\frac{T_i Y_i}{\hat{q}(x_i)}, \tag{2.4}$$

and

$$\hat{\mu}^{DR} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{T_i Y_i}{\hat{q}(x_i)} - \frac{T_i - \hat{q}(x_i)}{\hat{q}(x_i)}\hat{m}_1(x_i)\right\}, \tag{2.5}$$

where $\hat{m}_1(x)$ and $\hat{q}(x)$ are the estimates of $m_1(x)$ and $q(x)$, respectively. Estimator (2.3) only involves an outcome regression (OR) model, $m_1(x)$, so it is referred to as an OR estimator Tan 2006. Estimator (2.4) is called an IPW estimator (P. Rosenbaum, Colton, and Armitage 1998; Lunceford and Davidian 2004; Robins, Hernan, and Brumback 2000), since it includes a model for the propensity score (PS, Paul R Rosenbaum and Rubin 1983), $q(x) = P(T = 1|X = x)$, and uses the inverse probability weight (IPW). Estimator (3.3) is a doubly robust estimator of the potential outcome mean (DR, Robins, Rotnitzky, and Zhao 1995; Rotnitzky, Robins, and D. O. Scharfstein 1998; D. O. Scharfstein, Rotnitzky, and Robins 1999; Lunceford and Davidian 2004; Rotnitzky, Lei, et al. 2012). It involves both OR and PS models. Notice that the DR estimator is an augmented form of the IPW estimator:

$$\hat{\mu}^{DR} = \hat{\mu}^{IPW} - \frac{1}{n}\sum_{i=1}^{n}\frac{T_i - \hat{q}(x_i)}{\hat{q}(x_i)}\hat{m}_1(x_i).$$

The consistency of the resulting estimators OR, IPW, and DR depends on the correct specification of the relevant models for $q(x)$ and $m_1(x)$. However, unless the parametric models for $q(x)$ and $m_1(x)$ are correctly specified, we cannot expect OR or IPW estimator to be consistent. It is worth noting that the validity of the DR estimator entails weaker conditions, since it only requires either m_1 or q to be correctly specified. Because of this,

we will adopt the DR method for the proposed sensitivity analysis.

If there exists any unmeasured confounder, then $m(x) \neq m_1(x)$ and $P(x, y) \neq q(x)$, and the traditional estimators of μ_1 in (2.3)—(3.3) will be biased. To incorporate the unmeasured confounder to correct the estimation bias, we propose the following modified doubly robust estimator for μ_1 using the estimators of $m(x)$ and $P(x, y)$ instead of $m_1(x)$ and $q(x)$, respectively,

$$\hat{\mu}^{DR} = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{T_i Y_i}{\hat{P}(x_i, y_i)} - \frac{T_i - \hat{P}(x_i, y_i)}{\hat{P}(x_i, y_i)} \hat{m}(x_i) \right\}. \tag{2.6}$$

The difficulty of the estimation for $m(x)$ and $P(x, y)$ lies in the fact that the outcome $Y(1)$ is only partially observed.

2.2.2 Exponential tilt method for the unmeasured confounder

When there are unmeasured confounders, the conditional distribution of the observed $Y(1)$ is no longer the same as the conditional distribution of the unobserved $Y(1)$ given the covariates, and thus the relationship in equation (3.2) breaks down. Our goal is to restore the relationship between the conditional distributions of the observed and the unobserved outcome so that we can further estimate $m(x)$ and $P(x, y)$ in (3.4).

Let $g(y|x)$ be the conditional distribution of $Y(1)|X$ and $g_t(y|x)$ be the conditional distribution of $Y(1)|X, T = t, t = 0, 1$. Notice that when $T = 1$, $Y(1)$ is observed, and when $T = 0$, $Y(1)$ is missing. We propose leveraging the exponential tilt method from D.

Scharfstein et al. 2014 to build the connection between g_0 and g_1,

$$g_0(y|x) = \frac{g_1(y|x)e^{\alpha y}}{E\{e^{\alpha Y}|x, T=1\}}, \tag{2.7}$$

which links the distribution of unobserved outcomes to the the distribution of observed outcomes. The denominator in (3.5) is a normalization constant to make g_0 a legitimate density. Since the exponential tilt method (3.5) does not require specifying parametric models for latent confounding variables, the unmeasured confounder could be continuous, binary, or categorical, and could be univariate or multivariate. The parameter α is a univariate sensitivity parameter and non-identifiable with the data, and governs the departure of the truth from the strong ignorability assumption that "no unmeasured confounder exists". When $\alpha = 0$, $g_0(y|x) = g_1(y|x)$, which indicates there is no unmeasured confounding. When $\alpha \neq 0$, $g_0(y|x) \neq g_1(y|x)$ and hence some unmeasured confounders exist. It can be derived from (3.5) that for a binary outcome, if

$$\begin{aligned} &[Y(1)|T=1, X=x] \sim \text{Ber}\left(\pi(X)\right), \\ &\text{then } [Y(1)|T=0, X=x] \sim \text{Ber}\left(\frac{\pi(X)e^{\alpha}}{\pi(X)e^{\alpha} + 1 - \pi(X)}\right). \end{aligned} \tag{2.8}$$

Equation (2.8) also indicates that

$$\text{logit}\{P(Y(1)=1|T=0, X=x) - \text{logit}\{P(Y(1)=1|T=1, X=x)\} = \alpha,$$

where $\text{logit}(s) = \log\frac{s}{1-s}$. Therefore, for binary response variables, e^{α} is the conditional odds ratio (and hence α is the log odds ratio) of the unobserved potential outcome and

the observed potential outcome being 1 after adjusting for the measured confounding variables X; the sensitivity parameter α can directly measure the bias/violation of the strong ignorability assumption (3.1) caused by the unmeasured confounding.

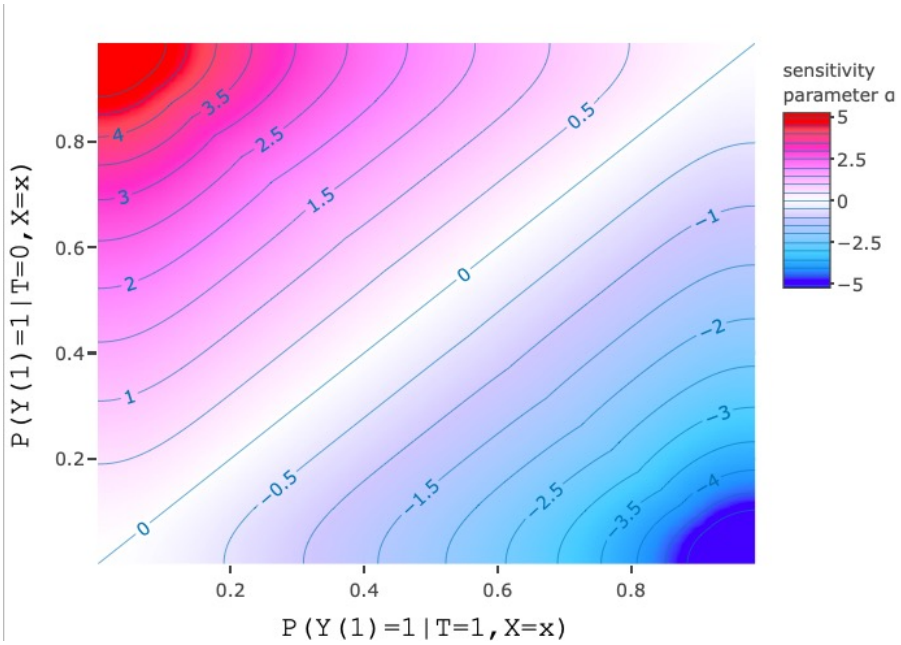


Figure 2.1: The contour plot of log-odds ratio

Figure 2.1 demonstrates how the conditional log-odds ratio relates to the underlying probabilities of the unobserved and observed potential outcome being 1 as the sensitivity parameter α varying from -5 to 5. When there is no unmeasured confounder, $P(Y(1) = 1|T = 0, X = x) = P(Y(1) = 1|T = 1, X = x)$ and thus the log odds ratio is $\alpha = 0$. If $\alpha > 0$, it implies that the probability of the unobserved potential outcome $Y(1)$ (given $T = 0$) being 1 is larger than that of the observed potential outcome (given $T = 1$) being 1 after adjusting for the measured confounding variables X. When $\alpha < 0$, the relationship is reversed, i.e., the probability of the unobserved potential outcome $Y(1)$ (given $T = 0$) being 1 is smaller than that of the observed potential outcome (given $T = 1$)

being 1 conditional on any X.

For example, if $Y = 1$ indicates a certain disease being cured, then e^{α} is the conditional odds ratio of being cured between the control patients had they been treated and the treated patient after adjusting for the measured covariates. A value of $\alpha = 1$ implies an odds ratio of $e^1 = 2.72$, meaning that the odds of being cured for the control patients had they been treated is almost 2.72 times as that of the treated patients after adjusting for measured covariates. The choice of α in practice relies on some subject-matter guidance, such as experts' experience and prior knowledge that is experiment-specific. In practice, usually, the α value from -2 to 2 (with the corresponding odds ratio from $e^{-2} = 0.14$ to $e^2 = 7.39$) or even -1 to 1 (with the odds ratio from 0.37 to 2.72) is a reasonable choice for a sensitivity analysis.

As mentioned earlier, the exponential tilt method of (3.5) can be also applied to the continuous outcomes easily. For example, based on the assumption of (3.5), if $[Y(1)|T = 1, X = x] \sim N(\eta, \sigma^2)$, then $[Y(1)|T = 0, X = x] \sim N(\eta + \alpha\sigma^2, \sigma^2)$ for the unobserved outcome variable. Thus, for continuous response variables, the sensitivity parameter α determines the mean shift/difference, which is $\alpha\sigma^2$, between the observed outcome $Y(1)$ from the treated arm and the unobserved outcome from the control arm due to the unmeasured confounder. Note that σ can be estimated based on the observed outcome. If the exponential tilt method is applied to the standardized data (i.e., $\sigma = 1$), then the sensitivity parameter α is exactly equal to the mean difference/shift between the unobserved potential outcome from the control arm and the observed potential outcome from the treatment arm.

To apply the proposed method (3.4), we need to estimate $m(x)$ and $P(x, y)$. Note that $g_1(y|x)$ and $q(x)$ can be estimated from observed data. We introduce a super learner machine learning estimation method for estimating $g_1(y|x)$ and $q(x)$ in Section 3.2.3. The $m(x)$ can then be estimated by $\hat{m}(x) = \hat{g}(1|x)$, where

$$\hat{g}(y|x) = \hat{q}(x)\hat{g}_1(y|x) + [1 - \hat{q}(x)]\hat{g}_0(y|x),$$

and $g_0(y|x)$ can be estimated based on $g_1(y|x)$ according to equation (3.5). To estimate $P(x, y)$, note that

$$\begin{aligned} P(x,y) =& \frac{g_1(y|x)q(x)}{g_1(y|x)q(x) + g_0(y|x)\{1 - q(x)\}} \\ =& \frac{g_1(y|x)q(x)}{g_1(y|x)q(x) + \frac{g_1(y|x)e^{\alpha y}}{E(e^{\alpha Y}|x,T=1)}\{1 - q(x)\}} \\ =& \text{expit}\left[\text{logit}\{q(x)\} - \alpha y + \log\left\{e^{\alpha} g_1(1|x) + g_1(0|x)\right\}\right], \end{aligned} \tag{2.9}$$

Then, $P(x, y)$ can be estimated by plugging the estimates to equation (3.7) as

$$\hat{P}(x,y) = \text{expit}\left[\text{logit}\{\hat{q}(x)\} - \alpha y + \log\left\{e^{\alpha}\hat{g}_1(1|x) + \hat{g}_1(0|x)\right\}\right]. \tag{2.10}$$

2.2.3 Super learner machine learning estimators

One way to estimate $g_1(x)$ and $q(x)$ is to specify some parametric models for them, for example, a linear or a logistic regression model for $g_1(x)$ for a continuous or binary response, respectively, and a logistic regression model for $q(x)$. However, if there are some violations of the parametric assumptions, which is usually the case in practice, it would

aggravate the bias in the causal inference even for a doubly robust estimator. To this end, we propose employing a super learner machine learning algorithm to nonparametrically estimate $g_1(x)$ and $q(x)$. Super learner is a general loss-based learning algorithm that was proposed and analyzed theoretically by Van der Laan, Polley, and Hubbard 2007. The algorithm optimally combines a library of machine learners by minimizing the cross-validation error and is aimed to estimate the regression function flexibly without over-fitting the data (Van der Laan and Rose 2011).

To illustrate this learning process, let the observed data be $(Y_i, X_i)_{i=1}^n$, where Y is the outcome and X is a set of covariates with dimension d. The super learner algorithm aims to estimate $m(x) = E(Y|X = x)$ using a library of machine learners $m_1, ...m_K$ weighted by $\lambda_1, \ldots, \lambda_K$, such that $\hat{m}(x) = \sum_{k=1}^K \lambda_k \hat{m}_k(X)$, where $\hat{m}_k$ is the estimator of $m(x)$ based on k^{th} machine learner. The selection of a library of machine learners will be discussed in Section 3.3. The weight vector $\lambda = (\lambda_1, \ldots, \lambda_K)$ can be chosen by the following cross validation procedure:

Randomly split the sample $(Y_i, X_i)_{i=1}^n$ into J equally sized subsets. For each $j \in (1, ..., J)$, the j^{th} subset, denoted by S_j, is used as a validation set and the other subsets are used as the training sets. Let $\hat{m}_k^{(-j)}$ be the estimator of $m(x)$ using the k^{th} machine learner based on the training data without the j^{th} subset S_j. Then we can find the weight vector by

$$(\hat{\lambda}_1, ..., \hat{\lambda}_K) = \underset{(\lambda_1,...,\lambda_K)}{\arg\min} \sum_{j=1}^{J} \sum_{(X_i,Y_i)\in S_j} \left\{ Y_i - \sum_{k=1}^{K} \lambda_k \hat{m}_k^{(-j)}(X_i) \right\}^2 .$$

Polley, Rose, and Van der Laan 2011 suggested bounding λ_k and using the constraints $\sum_{k=1}^K \lambda_k = 1,\ \lambda_k \geq 0,\ \forall k$. A non-negative binomial likelihood maximization, maximizing

AUC (Area Under The Curve) of ROC (Receiver Operating Characteristics) curve or minimizing the misclassification error rate can also be used as a cross validation criteria for binary outcomes.

2.2.4 New sensitivity analysis method

Below we summarize our proposed sensitivity analysis by combining the ideas of the doubly robust estimator, the exponential tilt method, and the super learner machine learning method introduced in Sections 2.1 to 2.3, respectively.

Step 1: Train the super learner method for $q(x)$ based on the data (T_i, X_i) using T as the response variable and X as the independent variable to obtain the estimate $\hat{q}(x)$.

Step 2: Train the super learner method for $g_1(x)$ based on the subset of the data $\{(X_i, Y_i),\ T_i = 1, i = 1, \ldots, n\}$ using Y as the response variable and X as the independent variable to obtain the estimate $\hat{g}_1(x)$.

Step 3: For a given value of the sensitivity parameter α, calculate:

$$\hat{g_0}(y|x) = \frac{e^{\alpha y}\hat{g}_1(y|x)}{e^{\alpha}\hat{g}_1(1|x) + \hat{g}_1(0|x)},$$

$$\hat{g}(y|x) = \hat{q}(x)\hat{g}_1(y|x) + \{1 - \hat{q}(x)\}\hat{g}_0(y|x),$$

$$\hat{m}(x) = \hat{g}(1|x),$$

and

$$\hat{P}(x, y) = \text{expit}\left[\text{logit}\{\hat{q}(x)\} - \alpha y + \log\left\{e^{\alpha}\hat{g}_1(1|x) + \hat{g}_1(0|x)\right\}\right].$$

Step 4: Repeat steps 1-3 for a set of the sensitivity parameter α, and compare results across different α values, say from -2 to 2.

Note that the new sensitivity analysis method only has one sensitive parameter α, which also enjoys a nice interpretation based on the odds ratio. Therefore, unlike most existing sensitivity analysis methods containing multiple sensitivity parameters, the new method is much easier to implement and choose the sensitivity parameter.

2.3 Examples

In this section, we conduct two sets of simulation studies. Similar to Lin, Psaty, and Kronmal 1998, we use the first simulation study to demonstrate the effectiveness of the proposed super learner based doubly robust estimation method (3.4) for adjusting unmeasured confounders in the estimation of μ_1 for any given sensitivity parameter α. The second simulation study is conducted to illustrate the performance of the proposed sensitivity analysis method in Section 3.2.5 by varying the sensitivity parameter. We compare our proposed nonparametric doubly robust estimator using the super learner algorithm (DR_np) with two parametric estimators: OR and IPW as shown in equations (2.3) and (2.4), respectively. We use parametric logistic regression to estimate the outcome and the propensity score models for OR and IPW estimators. For the proposed DR-np method, the super learner is based on the library of learners of generalized linear models (GLM), generalized additive models (GAM, Hastie and Tibshirani 1990) and recursive partitioning and regression trees (rpart, Breiman et al. 1984).

2.3.1 Estimating μ_1

Given three independent baseline covariates $X = (X_1, X_2, X_3)$ generated from a standard normal distribution, we generate the treatment assignment T and the observable binary outcome $Y(1)$ in the treatment group ($T = 1$) from the following two logistic model:

$$\text{logit}\{P(T = 1|X)\} = -2X_1 + X_2 + X_3 + 2\left\{X_1^2 + \sin(X_2) + (X_1 + 0.5)^3 e^{X_3}\right\},$$

$$\text{logit}\{P(Y(1) = 1|X, T = 1)\} = 3X_1 - 2X_2 - X_3 + 2\left\{X_1^2 X_2 - \sin(X_2) + (X_1 + 1)^3 e^{X_3}\right\}.$$

Note that the models used to generate both the treatment and the outcome include non-linear terms in X. The purpose is to make it unlikely for the analysts to formulate correct parametric models for the treatment and the outcome. In practice, it seems more natural for them to specify models that are linear in the observed covariates. When unmeasured confounding exists, the strong ignorability assumption (3.1) is violated and hence $\text{logit}\{P(Y(1) = 1|X, T = 1)\}$ will be different from $\text{logit}\{P(Y(1) = 1|X, T = 0)\}$ with the difference modeled by the exponential tilt relationship (2.8). The sensitivity parameter α in the exponential model can measure the departure of the truth from the assumption (3.1) regardless of the distribution and dimension of the unmeasured confounder. We examine $\alpha = (\pm 2, \pm 1.5, \pm 1, \pm 0.5)$ for estimating μ_1 (the corresponding odds ratio is from 0.14 to 7.39). The simulation was performed for the sample size $n = 1000$ with 1000 replicates.

Figure 2.2 displays squared errors of estimates for μ_1 based on the exponential tilt model (2.8) for each $\alpha = (\pm 2, \pm 1.5, \pm 1, \pm 0.5)$. It can be seen that our proposed method DR_np can estimate μ_1 well after adjusting the confounding effect using the exponential

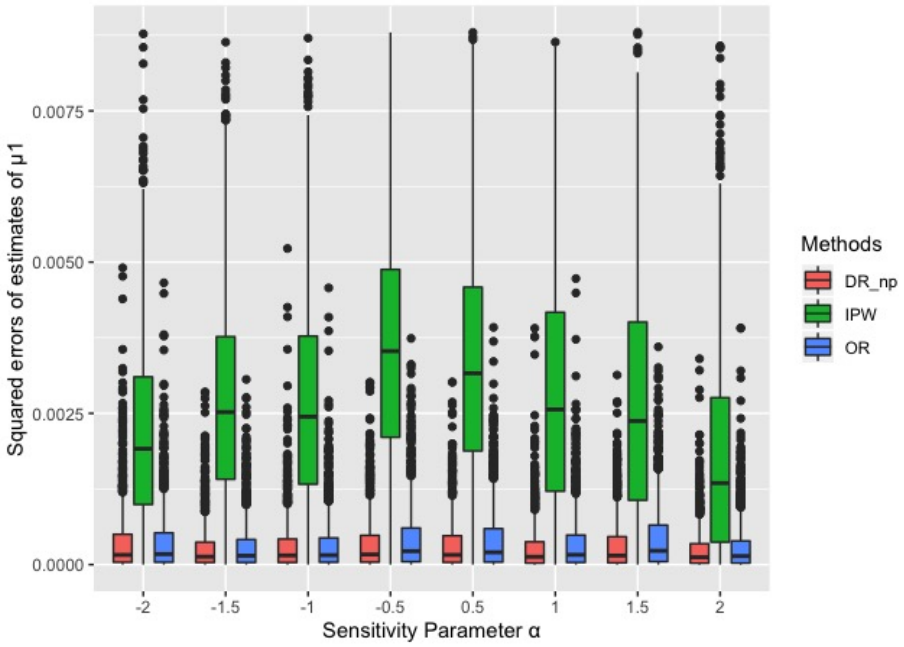


Figure 2.2: Squared errors of estimates for μ_1 in Simulation 1.

tilt model and result in estimates with the smallest median squared error and the smallest variation, with the OR method a close second. In addition, both DR_np and OR have much better performance than IPW.

2.3.2 Sensitivity analysis

In the previous section, we have demonstrated that the proposed method can successfully adjust the unmeasured confounder for any given sensitivity parameter α. In practice, however, the α is unknown. Next we demonstrate the proposed sensitivity analysis method by checking how the estimate changes when varying the sensitivity parameter α for the exponential tilt model.

To illustrate how our new sensitivity analysis method can be applied to unmeasured confounding settings used by existing sensitivity analysis methods, we incorporate

unmeasured confounding by explicitly including a latent variable $U \sim N(1,1)$ as one of the covariates in the models to generate the treatment T and the outcome $Y(t)$ as the following:

$$\begin{aligned}\text{logit}\{P(T=1|X,U)\} = &-2X_1 + X_2 + X_3 + \beta_u^t U \\ &+ 2\left\{X_1^2 + \sin(X_2) + (X_1+0.5)^3 e^{X_3}\right\},\end{aligned}$$

$$\begin{aligned}\text{logit}\{P(Y(t)=1|X,U)\} = &\, 3X_1 - 2X_2 - X_3 + 2t + \beta_u^y U \\ &+ 2\left\{X_1^2 X_2 - \sin(X_2) + (X_1+1)^3 e^{X_3}\right\},\ t=0,1.\end{aligned}$$

The observed outcome is $Y = TY(1) + (1-T)Y(0)$, and the data set used for analysis is $\{Y, T, X\}$. We assume $\beta_u^t = \beta_u^y = \beta_u$ and consider four cases of $\beta_u = (0,1,2,3)$ to represent different strengths of unmeasured confounding in the simulation. For each fixed $\beta_u = (0,1,2,3)$, we perform a sensitivity analysis by checking how the estimate of μ_1 changes when α varies from -4 to 4 (the corresponding odds ratio is from 0.2 to 54.6). The simulation is performed for the sample size $n = 1000$ with 1000 replicates.

Notice that case 1 with $\beta_u = 0$ indicates "there is no unmeasured confounder" which corresponds to $\alpha = 0$. However, for other nonzero β_u, it is hard to derive the explicit relationship between β_u and α. The sensitivity parameter α used in our new method can measure the difference between the conditional log-odds of the unobserved outcome being 1 and the log-odds of the observed outcome being 1. In other words, α directly governs the difference between $\text{logit}\{P(Y(t)=1|X,T=0)\}$ and $\text{logit}\{P(Y(t)=1|X,T=1)\}$ caused by the unmeasured confounding U with sensitivity parameters (β_u^t, β_u^y) and the distribution for U. Therefore, compared to the sensitivity parameters (β_u^t, β_u^y) and the distribution for U used by existing sensitivity analysis methods, the univariate sensitivity parameter α used

by our new method has better and more direct interpretation, and is also much easier to choose in practice .

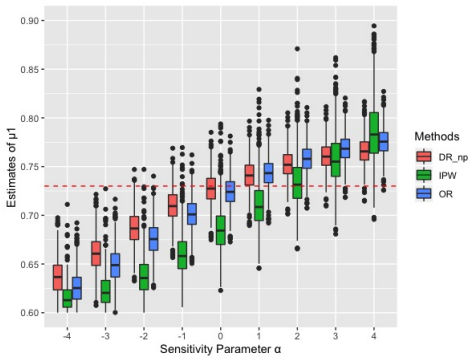


Figure 2.3: Simulation results of case 1: $\beta_u = 0$

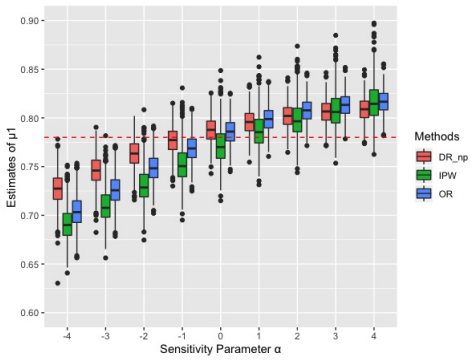


Figure 2.4: Simulation results of case 2: $\beta_u = 1$

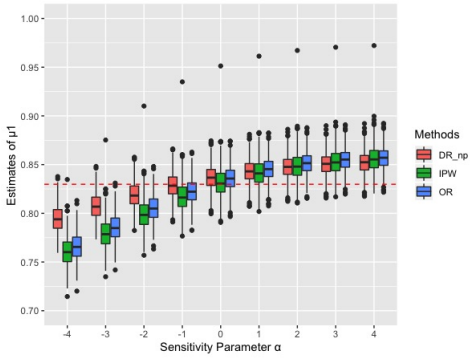


Figure 2.5: Simulation results of case 3: $\beta_u = 2$

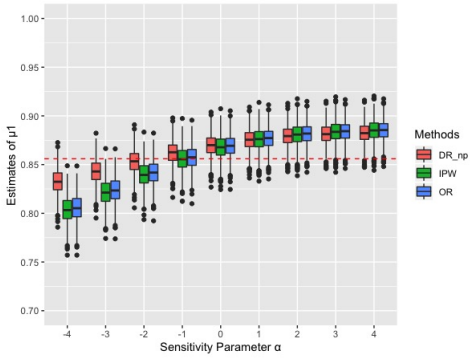


Figure 2.6: Simulation results of case 4: $\beta_u = 3$

Figures 3.2-2.6 display box plots of the estimates for μ_1 obtained by OR, IPW and the proposed method DR_np after adjusting the unmeasured confounding using the exponential tilt method with varying sensitivity parameter α for $\beta_u = 0, 1, 2$, and 3, respectively. The dashed line indicates the true μ_1. It can be seen that, at most specified α values, the proposed DR_np provides best estimates for μ_1.

2.4 Application

In this section, we discuss an application of the proposed sensitivity analysis to evaluate the causal relationship between heart failure death rate and low ejection fraction (EF$\leq$ 30). The ejection fraction (EF) measures how much blood the left ventricle pumps out with each contraction, which is usually represented as a percentage with a normal range between 50%-75%. As stated by the World Health Organization Cardiovascular diseases (CVDs) fact sheet, CVD causes 31% of all global deaths and is the number one cause of death. We use the data set which was originally analyzed by Ahmad et al. 2017. The main objective of their study was to estimate death rates due to heart failure and to investigate its link with some major risk factors in the city of Faisalabad (the third most populous city of Pakistan). The data set contains medical records of 299 heart failure patients in Faisalabad from April to December 2015. 10 confounding variables are addressed in this study and summarized in Table 2.1. Ahmad et al. 2017 used traditional survival models such as Cox regression and Kaplan-Meier plots to predict the death rate and identify the risk factors. According to their conclusion, low EF was a significant risk factor associated with death caused by heart failure. They also stated that there was a major difference among the Kaplan-Meier curves for patients with EF$\leq$30 and patients with EF$>$30. Thus, 30 is used as a cutoff to divide the EF into two groups: $T = 1,$ if EF ≤ 30 and $T = 0,$ if EF > 30 in our application. Our goal is to estimate the causal effect of low EF on heart failure deaths and to assess the sensitivity of the result to some unmeasured confounder. The outcome is defined based on whether heart failure death had occurred: $Y = 1,$ if the patient died and $Y = 0,$ otherwise. A summary of the death event variable Y is also provided in Table 2.1.

Variable	$EF \leq 30$ (n=93)	$EF > 30$ (n=207)	measurement
Age	59.63(10.80)	61.38(12.34)	Years
Anaemia	0.46(0.50)	0.42(0.49)	Binary
log(creatinine_phosphokinase)	5.68(1.09)	5.65(1.16)	mcg/L (log)
Diabetes	0.41(0.49)	0.42(0.50)	Binary
High blood pressure	0.37(0.48)	0.34(0.48)	Binary
Platelets	257836.45(94691.67)	265850.78(99303.30)	platelets/mL
Serum creatinine	1.45(0.77)	1.47(1.14)	mg/dL
Serum sodium	135.61(4.62)	137.08(4.25)	mEa/L
Sex	0.70(0.46)	0.63(0.48)	Binary
Smoking	0.34(0.48)	0.31(0.46)	Binary
Death event	0.55(0.50)	0.22(0.41)	Binary

Table 2.1: Mean (and standard deviation) of Ahmad et al. 2017 dataset.

Then, the causal effect of low EF is $\tau = \mu_1 - \mu_0 = P[Y(1) = 1] - P[Y(0) = 1]$. We use our proposed modified doubly robust estimator to estimate μ_1 and μ_0, and include GLM, GAM, rpart and random forest (Ho 1995) algorithms in the super leaner library.

With the adjustment of these 10 measured confounding variables, the estimated effect of low EF on heart failure death τ is 0.275 with a 95% confidence interval of (0.149, 0.402). Since the factors that lead to heart failure deaths remain unclear, it is informative and worthwhile to evaluate the sensitivity of the result of this study. For the potential outcome if everyone in this study had $EF \leq 30$, we assumed $\text{logit}[P\{Y(1)|T = 0, X\}] - \text{logit}[P\{Y(1)|T = 1, X\}] = \alpha_1$, where α_1 is the difference between the conditional log-odds of heart failure death for patients with normal EF($EF > 30$) if they had become $EF \leq 30$ and the conditional log-odds of heart failure death for patients with $EF \leq 30$. For the potential outcome if everyone in this study had $EF > 30$, we assume $\text{logit}[P\{Y(0)|T = 1, X\}] - \text{logit}[P\{Y(0)|T = 0, X\}] = \alpha_2$. The interpretation of α_2 is similar to α_1. For simplicity,

let $\alpha_1 = \alpha_2 = \alpha$, with the corresponding odds ratio being e^{α}. Table 2.2 displays the point estimates and 95% confidence intervals (obtained by a bootstrap with 100 replications) for the low EF effect on heart failure death rate τ with the adjustment of the unmeasured confounder represented by α.

α:	-4	-3	-2	-1
Odds ratio:	0.018	0.05	0.14	0.37
τ:	0.080	0.098	0.135	0.196
$C.I$:	(-0.003,0.162)	(0.009,0.188)	(0.035,0.235)	(0.097,0.295)
α:	1	2	3	4
Odds ratio:	2.72	7.39	20.09	54.60
τ :	0.343	0.372	0.369	0.360
$C.I$:	(0.222,0.464)	(0.276,0.468)	(0.260,0.479)	(0.248,0.471)

Table 2.2: Point estimates and 95% confidence intervals of the low EF effect on heart failure death rate.

The point estimates of τ change from approximately 0.08 to 0.36, while α being varied from -4 to 4. None of the α values makes the point estimate of τ null value zero (which would indicate that low EF would have no effect on heart failure death) or negative (which would indicate that low EF would decrease the heart failure death compared to higher EF). It should be pointed out that if the investigators have made efforts to collect the measured confounders, it might not be very likely to have unmeasured confounders that result in an $|\alpha| > 2$ or even $|\alpha| > 1$, although those values were examined in this sensitivity analysis. Notice that when $\alpha = -4$ with an odds ratio of 0.018, which is very unlikely to occur in practice, the null value $\tau = 0$ is included in the 95% confidence interval, but the lower limit of the confidence interval -0.003 is quite close to 0. The results of this sensitivity study demonstrate that the potential unmeasured confounder will not likely change the sign of τ (except for extreme α values). Therefore, we can safely conclude that

the low EF increases the heart failure death rate when compared to higher EF and such conclusion is robust against the possible unmeasured confounders.

Chapter 3

Sensitivity Analysis of Unmeasured Confounding in Causal Inference for Continuous Outcomes

3.1 Introduction

In observational and experimental studies, the identification of the causal effect, or the average treatment effect (ATE), of a treatment T on an outcome Y can be achieved under the potential outcome framework with the strong ignorability assumption proposed by Rubin 1974. For a binary treatment T with $T = 1$ indicating the treatment group and $T = 0$ indicating the control group, the potential outcome $Y(t)$ is the outcome that we would observe, had the treatment been assigned to $T = t$. Then, ATE can be defined as $\tau = \mu_1 - \mu_0 = E\{Y(1)\} - E\{Y(0)\}$, where $\mu_t = E\{Y(t)\}, (t = 0, 1)$, is the mean of the

potential outcome if everyone received $T = t$. The strong ignorability assumption assumes

$$Y(t) \perp T|X,\ t = 0, 1, \tag{3.1}$$

where X is a vector of measured confounders which are usually identified as some baseline covariates. Assumption (3.1) implies "there exists no unmeasured confounder" and requires all confounding variables to be measured and included in X. The above ignorability assumption (3.1) is usually questionable and very difficult to satisfy in practice. In addition, it cannot be tested using the observed data. Therefore, we can never confidently rule out the existence of unmeasured confounding in the estimation for ATE.

The existence of unmeasured confounding makes ATE unidentifiable using the observed data and would induce serious bias in causal estimates if it is not adjusted. To address this issue, sensitivity analysis is often used. Since the work of Cornfield et al. 1959 which explored a causal relationship between smoking and lung cancer, a number of sensitivity study approaches have been proposed. Paul R Rosenbaum and Rubin 1982 proposed a sensitivity analysis that included a binary unmeasured confounding variable U as a latent variable in the regression models for the binary treatment T and the binary outcome $Y(t)$ with the assumption that X and U together satisfy the strong ignorability assumption, i.e., $Y(t) \perp T|X, U$. The regression coefficients that are associated with U are not identifiable with the data and were used as sensitivity parameters. In a sensitivity analysis, different values of such sensitivity parameters are assumed and other parameters are estimated based on these specified values of sensitivity parameters. Some extension of this parameterization of unmeasured confounding include Imbens 2003 that extended Paul

R Rosenbaum and Rubin 1982 to allow for non-binary outcomes. Sensitivity studies that use different parameterization of unmeasured confounding were discussed in Lin, Psaty, and Kronmal 1998, Hogan, M. Daniels, and Hu 2014, Jung et al. 2018, Roy, Lum, and Michael J Daniels 2017 and Ding and VanderWeele 2016. Liu, Kuramoto, and Stuart 2013 and Richardson et al. 2014 provided excellent reviews about sensitivity analysis methods. Note that the above methods require a specification of whether the latent variable is continuous or discrete, and require a parameterization of unmeasured latent variable(s). In addition, the above methods contain multiple sensitivity parameters, which makes it difficult to specify a reasonable range of values for them.

In our earlier study, we proposed a modified doubly robust estimator together with a sensitivity analysis to address the unmeasured confounder based on exponential tilting (D. Scharfstein et al. 2014) and Super learner algorithm (Van der Laan, Polley, and Hubbard 2007). Similar to the standard doubly robust estimator (DR,Robins, Rotnitzky, and Zhao 1995; Rotnitzky, Robins, and D. O. Scharfstein 1998; D. O. Scharfstein, Rotnitzky, and Robins 1999; Lunceford and Davidian 2004; Rotnitzky, Lei, et al. 2012, this modified doubly robust estimator also involves an outcome regression(OR) model and an inverse probability weight (IPW) with the adjustment of unmeasured confounding. Although the super learner algorithm is used to nonparametrically estimate the OR and IPW models, severe misspecifications of both working models along with strong unmeasured confounders can still unduly hamper the performance of this method, since it can possibly result in extreme values of inverse probability weights. Trimming is likely the most frequently used solution to this problem. Observations with probability weights below a certain cutoff point are discarded

from subsequent analysis. However, most trimming methods are sensitive to the choice of the trimming threshold and the proportion of the sample excluded from the analysis. Therefore trimming is often criticized as ad-hoc. We extend our sensitivity analysis to cases where extreme probability weights exist with the implementation of trimming that uses a data-driven threshold proposed by Ma and Wang 2020.

In this article, we propose a new sensitivity analysis for continuous outcomes by leveraging the exponential tilting method to parameterize the unmeasured confounding as the departure from the ignorable treatment assumption Rubin 1974 regardless of the distribution of the unmeasured confounder. Compared to existing sensitivity analysis methods, which usually parameterize the unmeasured confounder as a latent variable in the working models with multivariate sensitivity parameters, the new method only has a univariate sensitivity parameter and does not impose any restrictions on the structure of the unmeasured confounders. In addition, we employ the data-driven trimming method proposed by Ma and Wang 2020 to handle the estimated extreme propensity scores that can hamper the performance of the modified doubly robust (DR) estimator. Furthermore, we use simulation studies and a real data example to demonstrate how we can adjust for the bias in the estimated average treatment effect (ATE) caused by the unmeasured confounder using the proposed method.

The rest of this paper is structured as follows. In section 2, we introduce our proposed estimation method in detail. In sections 3 and 4, we present simulated numerical examples and an empirical application to illustrate how our method works. A conclusion is given in section 5.

3.2 Methodology

3.2.1 Doubly robust estimator

For explanation purposes, we discuss our method by focusing on the estimation of $\mu_1 = E\{Y(1)\}$, the mean of potential outcome if everyone in the target population had been treated. Then $\mu_0 = E\{Y(0)\}$ and the corresponding ATE $\tau = \mu_1 - \mu_0$ can be estimated similarly.

For treated patients, their $Y(1)$'s are observed, so they are the actual outcomes. Whereas for control patients, $Y(1)$'s cannot be observed and become the counterfactual outcomes–the outcomes if the control patients had been treated. In order to estimate μ_1, the counterfactual outcomes need to be imputed first. Let $(X_i, Y_i, T_i)_{i=1}^n$ be an independent and identically distributed random sample from $X \in \mathbb{R}^d$, $Y \in (-\infty, \infty)$ and $T \in \{0, 1\}$, where X is a set of baseline covariates, and T is a binary treatment indicator. Define

$$\mu(x) = E(Y(1)|X = x),\ \mu_t(x) = E(Y(1)|T = t, X = x),\ t = 0, 1$$

and

$$p(x, y) = P(T = 1|X = x, Y(1) = y),\ p(x) = P(T = 1|X = x).$$

Our goal is to estimate $\mu_1 = \{Y(1)\} = \{\mu(X)\}$.

Note that

$$\mu(x) = p(x)\mu_1(x) + \{1 - p(x)\}\mu_0(x).$$

When no unmeasured confounder exists,

$$\mu_1(x) = \mu_0(x) = \mu(x) \text{ and } p(x, y) = p(x). \tag{3.2}$$

The outcome regression (OR) model for the treatment group $\mu_1(x)$ and the propensity score (PS) model for $p(x)$ can be estimated based on the observed data, The standard doubly robust estimator of μ_1 (DR, Robins, Rotnitzky, and Zhao 1995; Rotnitzky, Robins, and D. O. Scharfstein 1998; D. O. Scharfstein, Rotnitzky, and Robins 1999; Lunceford and Davidian 2004; Rotnitzky, Lei, et al. 2012) is defined as

$$\hat{\mu}^{DR} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{T_i Y_i}{\hat{p}(x_i)} - \frac{T_i - \hat{p}(x_i)}{\hat{p}(x_i)}\hat{\mu}_1(x_i)\right\}. \tag{3.3}$$

This DR estimator (3.3) involves both an OR model and a PS model. It has the doubly robust property since it can reduce the likelihood of bias by requiring that only one of the OR and PS models be correctly specified.

3.2.2 Addressing the unmeasured confounding

When the strong ignorability assumption (3.1) is violated, (i.e., there exists unmeasured confounding), the equations in (3.2) no longer hold. In this case, the traditional DR estimator (3.3) will be biased. To correct the bias, we can modify the DR estimator for μ_1 by replacing $\mu_1(x)$ and $p(x)$ by $\mu(x)$ and $p(x, y)$, respectively:

$$\hat{\mu}_m^{DR} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{T_i Y_i}{\hat{p}(x_i, y_i)} - \frac{T_i - \hat{p}(x_i, y_i)}{\hat{p}(x_i, y_i)}\hat{\mu}(x_i)\right\}. \tag{3.4}$$

We refer to estimator (3.4) as the modified DR estimator and $p(x, y)$ as the adjusted propensity score (PS). When the equations in (3.2) do not hold, the difficulty of estimating $\mu(x)$ and $p(x, y)$ (and hence μ_1) lies in that the conditional distribution of the observed $Y(1)$ differs from conditional distribution of the unobserved $Y(1)$ given the covariates. Thus we can no longer use the model fitted with the observed data to impute the unobserved potential outcome. Our goal is to restore the relationship between the conditional distributions of the observed and the unobserved outcome, so that we can estimate $\mu(x)$ and $p(x, y)$ in (3.4).

Let $f(y|x)$ be the conditional density of $Y(1)|X$ and $f_t(y|x)$ be the conditional density of $Y(1)|X, T = t, t = 0, 1$. When $T = 1$, $Y(1)$ is observed, and when $T = 0$, $Y(1)$ is missing. We use the exponential tilt method of D. Scharfstein et al. 2014 to build the connection between f_0 and f_1,

$$f_0(y|x) = f_1(y|x)e^{\gamma_x + \alpha y} \tag{3.5}$$

where $\gamma_x = -\log\{E(e^{\alpha Y}|x, T = 1)\} = -\log\{\int e^{\alpha y} f_1(y|x)dy\}$ is a normalization constant. Equation (3.5) connects the conditional distribution of the observed and unobserved potential outcomes together. The parameter α is a sensitivity parameter that cannot be identified using the observed data. It reflects a deviation from the benchmark assumption (i.e.,the strong ignorability assumption) that "no unmeasured confounder exists". When $\alpha = 0$, $f_0(y|x) = f_1(y|x)$, which corresponds to no unmeasured confounding. When $\alpha \neq 0$, $f_0(y|x) \neq f_1(y|x)$, hence some unmeasured confounders exist.

Let $M_x(t)$ be the moment generating function corresponding to $f_1(y|x)$ and $R_x(t) =$

$\log M_x(t)$ be the corresponding cumulant-generating function. Note that the normalization constant γ_x can be also expressed as $\gamma_x = -R_x(\alpha)$.

Proposition 1. *Supposing f_0 and f_1 have the exponential tilt relationship* (3.5)*, we have*

$$\mu_1(x) = R'_x(0) \text{ and } \mu_0(x) = R'_x(\alpha), \tag{3.6}$$

where $R'_x(t)$ is the first derivative of $R_x(t)$ with respect to t.

The proof of Proposition 1 is trivial and thus omitted here. Based on the above result, we know that the departure of $\mu_0(x)$ from $\mu_1(x)$ due to the violation of the strong ignorability assumption can be characterised by the shape of $R'_x(\cdot)$. Note that for a small α around 0,

$$\mu_0(x) - \mu_1(x) = R'_x(\alpha) - R'_x(0) \approx R''_x(0)\alpha = (Y(1)|T = 1, x)\alpha.$$

Hence, the departure of $\mu_0(x)$ from $\mu_1(x)$ when the strong ignorability assumption is violated depends on both α and the variability of the observed outcome $Y(1)$ given the covariates. More specifically, if $f_1(y|x)$ has a normal distribution with mean $\mu_1(x)$ and variance σ^2, then $M_x(t) = \exp\{\mu_1(x)t + \frac{1}{2}\sigma^2t^2\}$ and $R_x(t) = \mu_1(x)t + \frac{1}{2}\sigma^2t^2$. Based on Proposition 1, we have the following result when the outcome has a conditional normal distribution. Suppose f_0 and f_1 have the exponential tilt relationship (3.5). If $[Y(1)|T = 1, X = x] \sim N\left(\mu_1(x), \sigma^2\right)$, then $[Y(1)|T = 0, X = x] \sim N\left(\mu_0(x), \sigma^2\right)$ with $\mu_0(x) = \mu_1(x) + \alpha\sigma^2$. Based on Lemma 3.2.2, when the outcome has a conditional normal distribution, the sensitivity parameter α determines the difference, through $\alpha\sigma^2$, between the conditional mean of the observed

potential outcomes and the unobserved potential outcomes after adjusting for the observed covariates. If equation (3.5) is applied to standardized data (σ=1), α itself represents this mean shift/difference. A positive $\alpha(> 0)$ indicates that the control group, if they had been treated, would have had a higher mean value of $Y(1)$ than the treatment group. Conversely, a negative $\alpha(< 0)$ indicates that the control group, if they had been treated, would have had a lower mean value of $Y(1)$ than the treatment group. The parameter σ can be estimated using the observed data, whereas the sensitivity parameter α is non-identifiable.

Our proposed sensitivity analysis is to compute the causal estimates based on a set of numerical values of α which needs to be specified with some subject-matter guidance, such as experts' experience and other prior information about the experiment. Compared to the parameterizations that explicitly include the unmeasured confounder as a latent variable in the PS and OR models, the proposed formulation of the sensitivity parameter α does not impose any parametric form of the unmeasured confounder. Therefore, the unmeasured confounder could be continuous or discrete, and could be univariate or multivariate. Regardless of the distribution and dimension of the unmeasured confounder, this one-dimensional sensitivity parameter α can reflect the bias in the conditional mean of potential outcomes caused by the unmeasured confounder.

To apply the modified DR estimator (3.4), we need to estimate $\mu(x)$ and $p(x, y)$. Noting that $\mu_1(x)$ and $p(x)$ can be estimated based on the observed data, we can estimate $m_0(x)$ based on the exponential tilt equation (3.5) and Proposition 1. When the outcome has a conditional normal distribution, $m_0(x)$ can be easily estimated based on $\mu_1(x)$ through

the relationship in Lemma 3.2.2. The $p(x, y)$ can be also easily estimated by noting that

$$\begin{aligned} p(x,y) =& \frac{f_1(y|x)p(x)}{f_1(y|x)p(x) + f_0(y|x)\{1-p(x)\}} \\ =& \frac{f_1(y|x)p(x)}{f_1(y|x)p(x) + f_1(y|x)e^{\gamma_x+\alpha y}\{1-p(x)\}} \\ =& \frac{p(x)}{p(x) + e^{\gamma_x+\alpha y}\{1-p(x)\}}. \end{aligned} \tag{3.7}$$

Note that if $\alpha = 0$, then $\gamma_x = 0, e^{\gamma_x+\alpha y} = 1$, and $p(x, y) = p(x)$. If $f_1(y|x)$ is a normal density with mean $\mu_1(x)$ and variance σ^2, $\gamma_x = -\mu_1(x)\alpha - \frac{1}{2}\sigma^2\alpha^2$.

3.2.3 Super learner machine learning estimations

A standard way to estimate $\mu_1(x)$ and $p(x)$ based on the observed data is to specify some parametric models. However, the parametric assumptions are sensitive to model misspecification. A biased and even misleading doubly robust estimator would arise if incorrect parametric structures are specified on the outcome regression and the propensity score model. For this reason, we propose employing the super learner machine learning algorithm proposed and analyzed theoretically by Van der Laan, Polley, and Hubbard 2007 to nonparametrically estimate $\mu_1(x)$ and $p(x)$ for the modified DR estimator.

The super learner includes multiple learners and generates a set of optimal weights to combine those learners via the cross-validation. Given the observed data $(Y_i, X_i)_{i=1}^n$, the super learner estimates $\mu(x) = E(Y|X = x)$ with a library of L machine learners $m_1, \ldots, m_L$ and a set of weights $\omega_1, \ldots, \omega_L$. The resulting super learner estimate of $\mu(x)$ is a weighted average of all the learners used: $\hat{\mu}(x) = \sum_{l=1}^{L} \omega_l \hat{m}_l(x)$, where $\hat{m}_l$ is the estimator of $\mu(x)$ based on the l^{th} machine learner. The weight vector $\omega = (\omega_1, \ldots, \omega_L)$ can be chosen

by the following external cross validation procedure: We first randomly split the sample $(Y_i, X_i)_{i=1}^n$ into J equally sized disjoint subsets: $D_1, \ldots, D_J$ and then find the weight vector by minimizing the squared error loss function

$$(\hat{\omega}_1, ..., \hat{\omega}_L) = \underset{(\omega_1,...,\omega_L)}{\arg\min} \sum_{j=1}^{J} \sum_{(X_i,Y_i)\in D_j} \left\{ Y_i - \sum_{l=1}^{L} \omega_l \hat{m}_l^{(-j)}(X_i) \right\}^2 .$$

where $\hat{m}_l^{(-j)}$ is the estimate of $\mu(x)$ using the l^{th} machine learner based on the training data without the j^{th} subset D_j. Polley, Rose, and Van der Laan 2011 suggested bounding ω_l and using the constraints $\sum_{l=1}^{L} \omega_l = 1,\ \omega_l \geq 0,\ \forall l$. The selection of a library of machine learners will be discussed in Section 3.3.

3.2.4 Data-driven trimming approach

Despite the fact that the super learner algorithm allows for flexibility to capture the real pattern of the data, the performance of the modified DR estimator can be severely undermined if some of the estimated propensity scores are close to zero. Close-to-zero estimated propensity scores will unduly induce large bias and standard error in the final causal estimates. One popular solution is to reduce the impact of extreme weights through a trimming method. Fixed trimming is probably the most frequently used trimming scheme. It excludes observations whose estimated propensity score is less than a predetermined threshold $b_n = b > 0$. Note that if $b_n = 0$, there is no trimming. A rule of thumb is to set $b = 0.1$ recommended by Crump et al. 2009. Another popular choice of the trimming threshold is to use the q^{th} quantile of the estimated propensity scores and remove observations with PS that are below the q^{th} quantile. Stürmer et al. 2010 proposed a

trimming scheme that requires multiple steps. Firstly, subjects with propensity scores outside of a certain range are discarded. Then, among the treated units, subjects whose propensity score are below the q^{th} quantile are further excluded. The estimated propensity scores enter the final estimator through the inverse weighting and the trimming function. While these trimming approaches serve to reduce the bias of causal estimates, they have also been criticized for being too ad-hoc.

To this end, we extend a newly developed data-driven trimming threshold selection by Ma and Wang 2020 to the modified DR estimator (3.4). More specifically, we choose the trimming threshold data adaptively by minimizing an empirical analogue of the asymptotic mean squared error. The threshold b_n for PS can be obtained by solving the following equation:

$$P\left[p(x,y) \leq b_n\right] = \frac{1}{2nb_n}\frac{\eta_2(0)}{\eta_1(0)^2}, \tag{3.8}$$

where $\eta_s(p) = E\left[Y(1)^s | p(x,y) = p, T = 1\right]$ is the s^{th} order central moment of the response variable in the treatment group. We discard observations with estimated $p(x,y)$ below b_n, and then use the remaining subsample to calculate the modified DR estimator by the equation (3.9), i.e.,

$$\hat{\mu}_{tm}^{DR}(\alpha) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{T_i Y_i}{\hat{p}(x_i,y_i)} - \frac{T_i - \hat{p}(x_i,y_i)}{\hat{p}(x_i,y_i)}\hat{\mu}(x_i)\right\} I\{\hat{p}(x_i,y_i) > b_n\}, \tag{3.9}$$

where $I\{A\} = 1$, if A is true, and 0, otherwise. The above estimator $\hat{\mu}_{tm}^{DR}(\alpha)$ is the final proposed modified DR estimator. The asymptotic properties of the trimmed estimator are discussed in detail in Ma and Wang 2020.

3.2.5 A new sensitivity analysis

Note that the proposed modified DR estimator $\hat{\mu}_{tm}^{DR}(\alpha)$ in (3.9) depends on α. Under the strong ignorability assumption (3.1), $\alpha = 0$ and hence $\hat{\mu}_{tm}^{DR}(0)$ is the resulting estimate of μ_1.

In order to perform a sensitivity analysis of the causal estimate, we propose evaluating the $\hat{\mu}_{tm}^{DR}(\alpha)$ for a set of α values. Next, we summarize our proposed sensitivity analysis by combining the ideas of the doubly robust estimator, the exponential tilt method, the super learner machine learning method, and the data-driven trimming scheme introduced in Sections 2.1 to 2.4, respectively.

Step 1: Train the super learner method for $p(x)$ based on the data (T_i, X_i) using T as the response variable and X as the independent variable to obtain the estimate $\hat{p}(x)$.

Step 2: Train the super learner method for $\mu_1(x)$ based on the subset of the data $\{(X_i, Y_i),\ T_i = 1, i = 1, \ldots, n\}$ using Y as the response variable and X as the independent variable to obtain the estimate $\hat{\mu}_1(x)$.

Step 3: For a given sensitivity parameter α, estimate $\mu(x) = p(x)\mu_1(x) + \{1 - p(x)\}\mu_0(x)$ based on Proposition 1, and estimate $p(x, y)$ based on (3.7).

Step 4: Calculate the data-driven threshold $\hat{b}_n$ by solving the following equation:

$$\frac{1}{n}\sum_{i=1}^{n} I\{\hat{p}(x_i, y_i) \leq \hat{b}_n\} = \frac{1}{2n\hat{b}_n} \frac{\frac{1}{n}\sum_{i=1}^{n}\{y_i^2 | T = 1\}}{\left[\frac{1}{n}\sum_{i=1}^{n}\{y_i | T = 1\}\right]^2}.$$

Then, we can estimate μ_1 by $\hat{\mu}_{tm}^{DR}(\alpha)$ in (3.9) with the threshold $\hat{b}_n$.

Step 5: The sensitivity analysis is completed by repeating Steps 1-4 for a set of different values of the sensitivity parameter α, and comparing results across different α values.

If the observed outcome follows a normal distribution (i.e., $f_1(y|x)$ is a normal density), the estimate for $\mu(x)$ can be easily implemented based on Lemma 3.2.2.

3.3 Simulation studies

In this section, we conduct two sets of simulation studies. The first simulation study is conducted to illustrate the effectiveness of the proposed modified DR estimator $\hat{\mu}_{tm}^{DR}(\alpha)$ of (3.9) in the estimation of μ_1 for any given sensitivity parameter α. The second simulation study is conducted to illustrate the performance of the proposed sensitivity analysis method in Section 3.2.5 to address unmeasured confounding by varying the sensitivity parameter values.

We compare the proposed super learner based modified DR estimator (DR_np) with the corresponding parametric version (DR_par) that uses a logistic regression model and linear regression model to estimate $p(x)$ and $\mu_1(x)$, respectively. For the proposed DR-np method, the super learner used to estimate $p(x)$ is based on the library of learners of generalized linear model (GLM), generalized additive models (GAM, Hastie and Tibshirani 1990) and recursive partitioning and regression trees (rpart, Breiman et al. 1984), the super learner for estimating $\mu_1(x)$ is based on the library of linear regression, GAM, neural networks (nnet,Hopfield 1982), locally estimated scatterplot smoothing (LOESS, Cleveland 1979) and random forest (Hastie and Tibshirani 1990).

3.3.1 Estimation of mean outcome μ_1

We include three independent baseline covariates X_1, X_2, and X_3 generated from a standard normal distribution. Given the covariates, we generate the treatment assignment T from the following regression:

$$\text{logit}\{P(T=1|X)\} = -0.5X_1 + X_2 + X_3 + 2\left\{X_1^2 + \sin(X_2) + (X_1+0.5)^3 e^{X_3}\right\}.$$

The continuous outcome $Y(1)$ in the treatment group ($T=1$) is generated from the normal distributions $N(\mu_1(X), 1)$, where

$$\mu_1(X) = X_1 - 2X_2 - X_3 - 0.5\left\{\sin(X_1) - X_2^2 + e^{X_3}\right\}.$$

We include non-linear terms in X in both the treatment and the outcome models. Therefore, there are some misspecifications for the traditional parametric logistic regression model and the linear regression model. When unmeasured confounding exists, $\mu_0(x) = \{Y(1)|T = 0, X\}$ is no longer the same as $\mu_1(x) = \{Y(1)|T=1, X\}$ with α representing this difference (note that $\sigma = 1$). A sequence of equal-spaced values from -2 to 2 with increment of 0.25 are checked for the sensitivity parameter α in our simulation. All methods described above are applied to 1000 samples of size $n = 1000$.

Figure 3.1 displays the squared error of trimmed DR_par and DR_np estimates of μ_1 as the sensitivity parameter α varying from -2 to 2 with an increment of 0.25. For most α values, the median squared error and the variation of the proposed method DR_np are

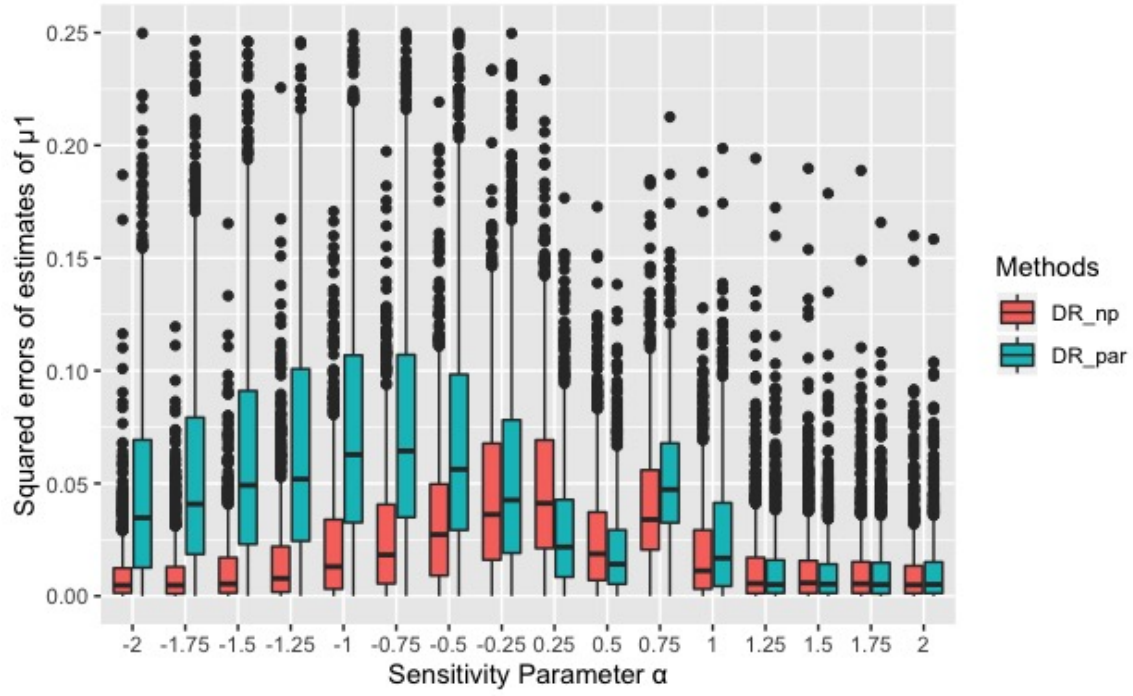


Figure 3.1: The box plot of squared errors of trimmed DR_par and trimmed DR_np estimates for μ_1.

smaller than those of the DR_par estimates.

3.3.2 Sensitivity analysis

From the previous section, we know that the proposed method can successfully adjust the unmeasured confounder for any given sensitivity parameter α. However, in practice, the α is unknown. Next, to demonstrate how our new sensitivity analysis method can be applied to unmeasured confounding settings used by existing sensitivity analysis methods, we introduce unmeasured confounding by explicitly including a latent variable $U \sim N(1, 1)$ as one of the covariates in the models to generate the treatment T and the

outcome $Y(t)$ as the following:

$$\begin{aligned}\text{logit}\{P(T=1|X,U)\} = &-0.5X_1 + X_2 + X_3 + \beta_u^t U \\ &+ 0.5\left(X_1^3 + \sin(X_2) + X_3^2\right),\end{aligned}$$

and then the outcome $Y(t)$ is generated from the following outcome regression:

$$Y(t) = X_1 - 2X_2 - X_3 - 0.5\left(e^{X_1} - X_2^2 + X_1X_3\right) + \beta_u^y U + 2t + \epsilon,$$

where $t = 0, 1$ and $\epsilon \sim N(0,1)$.

The data we use for our analysis is $\{X, T, Y\}$, where $Y = TY(1)+(1-T)Y(0)$ is the observed outcome. For simplicity, we let $\beta_u^t = \beta_u^y = \beta_u$ and examine four different values of $\beta_u = \{0, 0.5, 1, 2\}$ representing cases of no unmeasured confounding, weak, mild and strong unmeasured confounding, respectively. For each value of β_u, the proposed sensitivity study is conducted by varying α, the sensitivity parameter, from -1 to 1 with an increment of 0.25. The analysis is performed to 1000 samples of size $n = 1000$. In our sensitivity study, the univariate sensitivity parameter α measures the bias due to unmeasured confounding and $\beta_u = 0$ implies "no unmeasured confounder" which corresponds to $\alpha = 0$. However, there is no clear/explicit relationship between nonzero β_u and α.

Figures 3.2-3.5 display box plots of the estimates for μ_1 obtained by DR_par and DR_np addressing unmeasured confounding using the exponential tilt method. True μ_1 is represented by the dashed line. Each figure shows how the estimates change with different values of the sensitivity parameter α. It can be seen that, at most specified α values, the

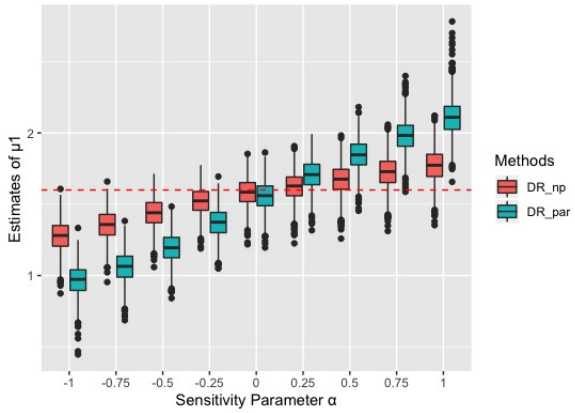


Figure 3.2: Simulation results of case 1: $\beta_u = 0$

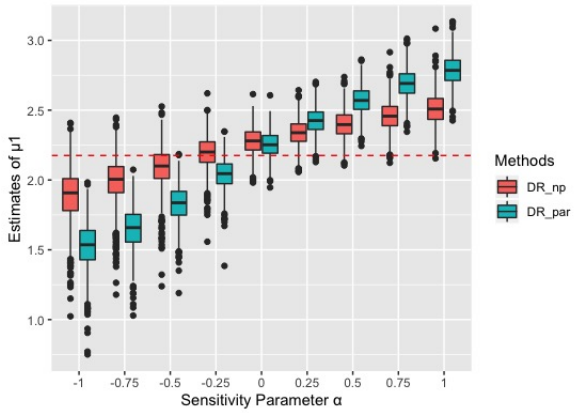


Figure 3.3: Simulation results of case 2: $\beta_u = 0.5$

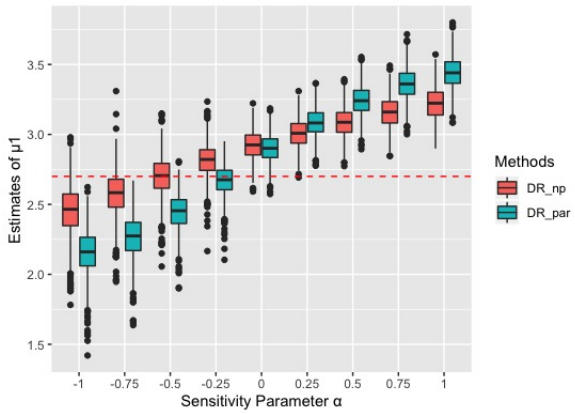


Figure 3.4: Simulation results of case 1: $\beta_u = 1$

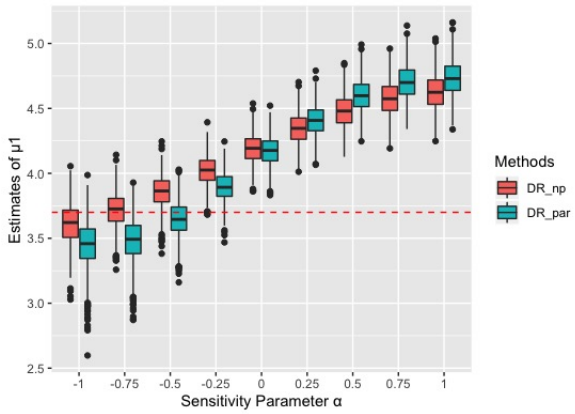


Figure 3.5: Simulation results of case 2: $\beta_u = 2$

proposed DR_np provides better estimates for μ_1 in all four cases corresponding to different strengths of the unmeasured confounder.

3.4 Empirical application

In this section, we discuss an application of the proposed sensitivity study to evaluate how the unmeasured confounder affects the causal effect (ATE) of the National Supported Work (NSW) Demonstration. NSW was a labor training program conducted in

the 1970's and provided work experience to selected participants. The study measured the baseline covariates: age, education, black, Hispanic, married, earning in 1974, earning in 1975, unemployed in 1974, and unemployed in 1975. The effect of the NSW program on post-intervention annual income levels, 1978 earnings, was originally studied by LaLonde 1986 and has been analyzed by various studies ever since, such as Imbens 2003; Dehejia and Wahba 1999; Carnegie, Harada, and Hill 2016. The data analyzed here is the same as in Dehejia and Wahba 1999. We use this data set to demonstrate how our proposed sensitivity procedure can address the impact of unmeasured confounding on the effect of NSW on 1978 earnings based on our proposed modified DR estimator with data-driven trimming. Let $T = 1$ indicate enrollment in NSW and $T = 0$ otherwise. Let X be the vector of all measured confounders described in Table 3.1. The outcome variable Y is the 1978 earnings.

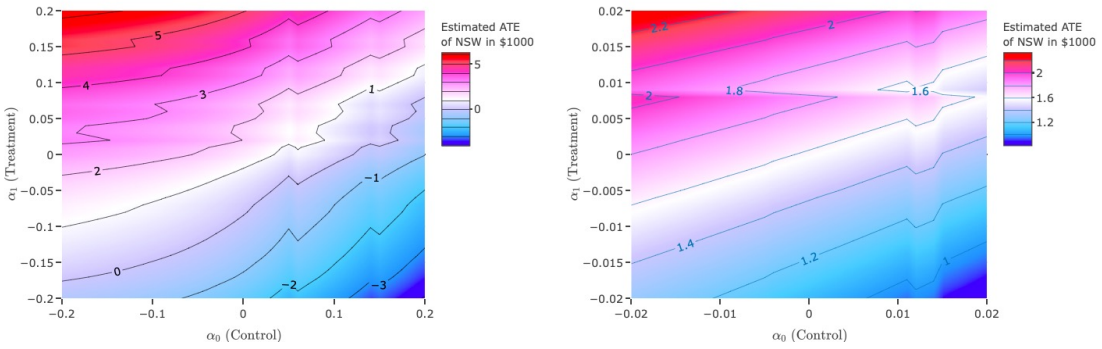
Variable	Treatment (n=185)	Control (n=260)	measurement
Age	25.82(7.16)	25.05(7.06)	Years
Education	10.35(2.01)	10.09(1.61)	Years of schooling
Black	0.84(0.36)	0.83(0.38)	Binary
Hispanic	0.06(0.23)	0.11(0.31)	Binary
Married	0.19(0.39)	0.15(0.36)	Binary
High school diploma	0.71(0.46)	0.83(0.37)	Binary
1974 Earnings	2.10(4.89)	2.11(5.69)	1000 US dollar
1975 Earnings	1.53(3.22)	1.27(3.10)	1000 US dollar
1974 Unemployment	0.71(0.46)	0.75(0.43)	Binary
1975 Unemployment	0.60(0.49)	0.68(0.47)	Binary
1978 Earnings	6.35(7.87)	4.55(5.48)	1000 US dollar

Table 3.1: Mean (and standard deviation) of LaLonde 1986 dataset.

If no unmeasured confounding exists ($\alpha = 0$), the estimated causal effect (ATE) of the NSW program on 1978 earning is 1.65 thousand dollars using the proposed modified

DR estimator $\hat{\mu}_{tm}^{DR}(0)$ of (3.9). However, as mentioned in Imbens 2003, strong motivation to enroll in a job-training program may lead to more favorable outcomes. Thus motivation to join the program can definitely be considered unmeasured confounding and would be worthwhile to address using the sensitivity analysis. For the potential outcomes if everyone had enrolled in NSW, we employ the exponential tilt method of (3.5) to assume $E\{Y(1)|T = 0, X\} - E\{Y(1)|T = 1, X\} = \alpha_1\sigma_1^2$, where $\sigma_1^2 = \{Y(1)|T = 1, X\}$ and $\alpha_1\sigma_1^2$ represents the difference between the conditional mean of 1978 outcome for nonparticipants of NSW if they had been enrolled in NSW and that of those who were actually enrolled in the NSW program. Similarly, for the potential outcome if everyone were not enrolled, we assume $E\{Y(0)|T = 1, x\} - E\{Y(0)|T = 0, x\} = \alpha_0\sigma_0^2$, where $\sigma_0^2 = \{Y(0)|T = 0, X\}$ and $\alpha_0\sigma_0^2$ has similar interpretation to $\alpha_1\sigma_1^2$. The sensitivity parameters α_0 and α_1 are taken from -0.2 to 0.2 with an increment of 0.01 (all combinations are considered). The standard deviations (σ_1 and σ_0) of 1978 earnings for the treatment group and the control group are 7.9 and 5.5, respectively, which are relatively large compared to the means. Therefore, a value of $\alpha_1 = 0.2$ (the interpretation of α_0 can be done similarly) indicates that the conditional mean of 1978 earning of nonparticipants of NSW if they had joined the program would be $\alpha_1\sigma_1^2 = 0.2 \times 7.9^2 = 12.48$ thousand US dollars *more* than the participants who were in NSW. It also means that if the nonparticipants joined NSW, their 1978 earning would be approximately 3 times that of the participants, which can be considered a big impact on the outcome resulting from the unmeasured confounder. In practice, an α value between -0.02 and 0.02, representing a difference of approximately 1.25 thousand dollars between the conditional means of the observed and unobserved potential outcomes for NSW, is more

likely.



Left: the range of sensitivity parameters is $(-0.2, 0.2)$;

Right: the range of sensitivity parameters is $(-0.02, 0.02)$.

Figure 3.6: Sensitivity analysis results of the ATE of NSW on 1978 earnings.

Figure 3.6 displays the contour plot of the ATE estimates for different combinations of α_1 and α_0. We notice that, when α_0 and α_1 are very different, the estimates tend to be very different from the benchmark value 1.65. For example, when $\alpha_1 = 0.2$ and $\alpha_1 = -0.1$, the estimated ATE is approximately 5.6 thousand US dollars, and the unmeasured confounder would significantly change our causal results. Also notice that, when the range of both α_0 and α_1 are between -0.02 and 0.02, none of the combinations of sensitivity parameter values would result in a significant departure more than 1 thousand US dollars from the bench mark value or change the direction of our conclusion. Thus, if α_0 and α_1 are both bounded between -0.02 and 0.02, this unmeasured confounding would only have a mild impact on our causal inference.

Chapter 4

Conclusions

In Chapters 2 and 3, We have proposed a new sensitivity analysis method for causal inference to adjust for unmeasured confounding in the estimation of the mean outcome by combining the ideas of the doubly robust estimator, the exponential tilting method, and the super learner algorithm. In causal inference, when unmeasured confounders exist, the conditional distribution of the observed outcome is different from that of the unobserved outcome given the covariates. As a result, the model estimated from the observed data can no longer be used to model the unobserved potential outcomes. This relationship between these distributions is restored by the exponential tilting method with a one-dimensional sensitivity parameter, which adjusts for the bias due to the unmeasured confounding, and addresses how sensitive a causal inference is when the strong ignorability assumption does not hold. Compared to most of the existing sensitivity analysis in the literature, our method does not require modeling assumptions for the unmeasured confounders as latent variables and hence the unmeasured confounder could be continuous, binary, or categorical, and could

be univariate or multivariate. In addition, the sensitivity parameter can be interpreted as a log-odds ratio for a binary outcome and deviation of the conditional outcome means for a continuous outcome, which makes the choice of its range relatively easy for practitioners. To increase the accuracy of traditional parametric methods, We propose a nonparametric modified doubly robust estimator with the super learner algorithm to reduce the bias caused by the possible misspecification of the parametric working models used by the traditional doubly robust estimator. To avoid observations with an extremely small estimated propensity score, we incorporate a newly developed data-driven trimming method to provide more stable/reliable causal estimates.

Bibliography

Neyman, Jerzy S: *On the application of probability theory to agricultural experiments. essay on principles. section 9.(tlanslated and edited by dm dabrowska and tp speed, statistical science (1990), 5, 465-480).* In: *Annals of Agricultural Sciences* 10 (1923), pp. 1–51.

Rubin, Donald B: *Estimating causal effects of treatments in randomized and nonrandomized studies.* In: *Journal of educational Psychology* 66.5 (1974), p. 688.

Rosenbaum, Paul R and Donald B Rubin: *Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome.* Wisconsin Clinical Cancer Center, Biostatistics, 1982.

Imbens, Guido W: *Sensitivity to exogeneity assumptions in program evaluation.* In: *American Economic Review* 93.2 (2003), pp. 126–132.

Tan, Zhiqiang: *Regression and weighting methods for causal inference using instrumental variables.* In: *Journal of the American Statistical Association* 101.476 (2006), pp. 1607–1618.

Rosenbaum, PR, T Colton, and P Armitage: *Encyclopedia of biostatistics.* In: (1998).

Lunceford, Jared K and Marie Davidian: *Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study.* In: *Statistics in medicine* 23.19 (2004), pp. 2937–2960.

Robins, James M, Miguel Angel Hernan, and Babette Brumback: *Marginal structural models and causal inference in epidemiology.* 2000.

Robins, James M, Andrea Rotnitzky, and Lue Ping Zhao: *Analysis of semiparametric regression models for repeated outcomes in the presence of missing data.* In: *Journal of the american statistical association* 90.429 (1995), pp. 106–121.

Rotnitzky, Andrea, James M Robins, and Daniel O Scharfstein: *Semiparametric regression for repeated outcomes with nonignorable nonresponse.* In: *Journal of the american statistical association* 93.444 (1998), pp. 1321–1339.

Scharfstein, Daniel O, Andrea Rotnitzky, and James M Robins: *Adjusting for nonignorable drop-out using semiparametric nonresponse models.* In: *Journal of the American Statistical Association* 94.448 (1999), pp. 1096–1120.

Rotnitzky, Andrea, Quanhong Lei, et al.: *Improved double-robust estimation in missing data and causal inference models.* In: *Biometrika* 99.2 (2012), pp. 439–456.

Rosenbaum, Paul R and Donald B Rubin: *The central role of the propensity score in observational studies for causal effects.* In: *Biometrika* 70.1 (1983), pp. 41–55.

Lin, Danyu Y, Bruce M Psaty, and Richard A Kronmal: *Assessing the sensitivity of regression results to unmeasured confounders in observational studies.* In: *Biometrics* (1998), pp. 948–963.

Ding, Peng and Tyler J VanderWeele: *Sensitivity analysis without assumptions.* In: *Epidemiology (Cambridge, Mass.)* 27.3 (2016), p. 368.

VanderWeele, Tyler J: *Sensitivity analysis: distributional assumptions and confounding assumptions.* In: *Biometrics* 64.2 (2008), pp. 645–649.

VanderWeele, Tyler J and Peng Ding: *Sensitivity analysis in observational research: introducing the E-value.* In: *Annals of internal medicine* 167.4 (2017), pp. 268–274.

Cornfield, Jerome et al.: *Smoking and lung cancer: recent evidence and a discussion of some questions.* In: *Journal of the National Cancer institute* 22.1 (1959), pp. 173–203.

Veitch, Victor and Anisha Zaveri: *Sense and Sensitivity Analysis: Simple Post-Hoc Analysis of Bias Due to Unobserved Confounding.* In: *arXiv preprint arXiv:2003.01747* (2020).

Hogan, JW, MJ Daniels, and L Hu: *A Bayesian perspective on assessing sensitivity to assumptions about unobserved data.* In: *Handbook of Missing Data Methodology. Boca Raton, FL, USA: Chapman & Hall/CRC Press.[Google Scholar]* (2014).

Jung, Jongbin et al.: *Algorithmic decision making in the presence of unmeasured confounding.* In: *arXiv preprint arXiv:1805.01868* (2018).

Roy, Jason, Kirsten J Lum, and Michael J Daniels: *A Bayesian nonparametric approach to marginal structural models for point treatments and a continuous or survival outcome.* In: *Biostatistics* 18.1 (2017), pp. 32–47.

McCandless, Lawrence C, Paul Gustafson, and Adrian Levy: *Bayesian sensitivity analysis for unmeasured confounding in observational studies.* In: *Statistics in medicine* 26.11 (2007), pp. 2331–2347.

Dorie, Vincent et al.: *A flexible, interpretable framework for assessing sensitivity to unmeasured confounding.* In: *Statistics in medicine* 35.20 (2016), pp. 3453–3470.

McCandless, Lawrence C and Paul Gustafson: *A comparison of Bayesian and Monte Carlo sensitivity analysis for unmeasured confounding.* In: *Statistics in Medicine* 36.18 (2017), pp. 2887–2901.

Scharfstein, Daniel et al.: *Global sensitivity analysis for repeated measures studies with informative dropout: A fully parametric approach.* In: *Statistics in Biopharmaceutical Research* 6.4 (2014), pp. 338–348.

Van der Laan, Mark J, Eric C Polley, and Alan E Hubbard: *Super learner.* In: *Statistical applications in genetics and molecular biology* 6.1 (2007).

Groenwold, Rolf HH et al.: *Sensitivity analyses to estimate the potential impact of unmeasured confounding in causal research.* In: *International journal of epidemiology* 39.1 (2010), pp. 107–117.

Van der Laan, Mark J and Sherri Rose: *Targeted learning: causal inference for observational and experimental data.* Springer Science & Business Media, 2011.

Polley, Eric C, Sherri Rose, and Mark J Van der Laan: *Super learning.* In: *Targeted learning.* 2011, pp. 43–66.

Hastie, Trevor J and Robert J Tibshirani: *Generalized additive models.* Vol. 43. CRC press, 1990.

Breiman, Leo et al.: *Classification and regression trees.* CRC press, 1984.

Ahmad, Tanvir et al.: *Survival analysis of heart failure patients: A case study.* In: *PloS one* 12.7 (2017), e0181001.

Ho, Tin Kam: "Random decision forests". In: *Proceedings of 3rd international conference on document analysis and recognition.* Vol. 1. IEEE. 1995, pp. 278–282.

Liu, Weiwei, S Janet Kuramoto, and Elizabeth A Stuart: *An introduction to sensitivity analysis for unobserved confounding in nonexperimental prevention research.* In: *Prevention science* 14.6 (2013), pp. 570–580.

Richardson, Amy et al.: *Nonparametric bounds and sensitivity analysis of treatment effects.* In: *Statistical science: a review journal of the Institute of Mathematical Statistics* 29.4 (2014), p. 596.

Ma, Xinwei and Jingshen Wang: *Robust inference using inverse probability weighting*. In: *Journal of the American Statistical Association* 115.532 (2020), pp. 1851–1860.

Crump, Richard K et al.: *Dealing with limited overlap in estimation of average treatment effects*. In: *Biometrika* 96.1 (2009), pp. 187–199.

Stürmer, Til et al.: *Treatment effects in the presence of unmeasured confounding: dealing with observations in the tails of the propensity score distribution—a simulation study*. In: *American journal of epidemiology* 172.7 (2010), pp. 843–854.

Hopfield, John J: *Neural networks and physical systems with emergent collective computational abilities*. In: *Proceedings of the national academy of sciences* 79.8 (1982), pp. 2554–2558.

Cleveland, William S: *Robust locally weighted regression and smoothing scatterplots*. In: *Journal of the American statistical association* 74.368 (1979), pp. 829–836.

LaLonde, Robert J: *Evaluating the econometric evaluations of training programs with experimental data*. In: *The American economic review* (1986), pp. 604–620.

Dehejia, Rajeev H and Sadek Wahba: *Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs*. In: *Journal of the American statistical Association* 94.448 (1999), pp. 1053–1062.

Carnegie, Nicole Bohme, Masataka Harada, and Jennifer L Hill: *Assessing sensitivity to unmeasured confounding using a simulated potential confounder*. In: *Journal of Research on Educational Effectiveness* 9.3 (2016), pp. 395–420.