Sensitivity analysis for observational studies: Looking back and moving forward

Qingyuan Zhao

Statistical Laboratory, University of Cambridge

September 8, 2020 (Yale Biostats Seminar)

Based on ongoing work with Bo Zhang, Ting Ye, Dylan Small (U Penn) and Joe Hogan (Brown U).
Slides can be found at `http://www.statslab.cam.ac.uk/~qz280/`.

Sensitivity analysis

Sensitivity analysis is widely found in any area that uses mathematical models.

The broader concept [Saltelli et al., 2004]

- "The study of how **the uncertainty in the output** of a mathematical model or system (numerical or otherwise) can be apportioned to different sources of **uncertainty in its inputs**".
- Model inputs may be any factor that "can be changed in a model prior to its execution", including **"structural and epistemic sources of uncertainty"**.

In observational studies

- The most typical question is:

  How do the qualitative and/or quantitative conclusions of the observational study change if the **no unmeasured confounding assumption** *is violated?*

Sensitivity analysis for observational studies

State of the art

- Gazillions of methods specifically designed for different problems.
- Various forms of statistical guarantees.
- Often not straightforward to interpret

Goal of this talk: A high-level overview

1. What is the **common structure** behind?
2. What are some **good principles and ideas**?

The perspective of this talk: **global** and **frequentist**.

Prototypical setup

Observed iid copies of $\boldsymbol{O} = (\boldsymbol{X}, A, Y)$ from the underlying full data $\boldsymbol{F} = (\boldsymbol{X}, A, Y(0), Y(1))$, where A is a binary treatment, $\boldsymbol{X}$ is covariates, Y is outcome.

Outline

Motivating example

Component 1: Sensitivity model

Component 2: Statistical inference

Component 3: Interpretation

Example: Child soldiering [Blattman and Annan, 2010]

- From 1995 to 2004, about $60,000$ to $80,000$ youths were abducted in Uganda by a rebel force.
- Question: What is the impact of child soldiering (e.g. on the years of education)?
- The authors controlled for a variety of covariates $\boldsymbol{X}$ (age, household size, parental education, etc.) but were concerned about **ability to hide from the rebel** as a unmeasured confounder.
- They used the following model proposed by Imbens [2003]:

$$
\begin{aligned}
& A \perp\!\!\!\perp Y(a) \mid \boldsymbol{X}, U, \text{ for } a = 0, 1, \\
& U \mid \boldsymbol{X} \sim \text{Bernoulli}(0.5), \\
& A \mid \boldsymbol{X}, U \sim \text{Bernoulli}(\text{expit}(\boldsymbol{\kappa}^T \boldsymbol{X} + \lambda U)), \\
& Y(a) \mid \boldsymbol{X}, U \sim \text{N}(\beta a + \boldsymbol{\nu}^T \boldsymbol{X} + \delta U, \sigma^2) \text{ for } a = 0, 1,
\end{aligned}
$$

- U is an unobserved confounder. (λ, δ) are sensitivity parameters; $\lambda = \delta = 0$ corresponds to a primary analysis assuming no unmeasured confounding.

Main results of Blattman and Annan [2010]

- Their primary analysis found that the ATE is -0.76 (s.e. 0.17).
- Sensitivity analysis can be summarized with a single calibration plot:

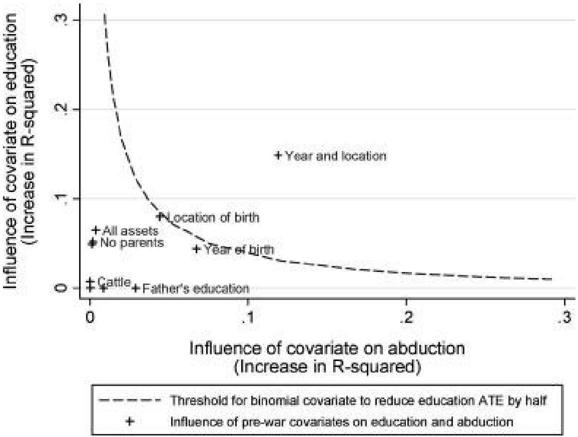


Figure 5 of Blattman and Annan [2010].

Three components of sensitivity analysis

1. **Model augmentation:** Need to extend the model used by primary analysis to allow for unmeasured confounding.
2. **Statistical inference:** Vary the sensitivity parameter, estimate the causal effect, and control suitable statistical errors.
3. **Interpretation of the results:** Sensitivity analysis is often quite complicated (because we need to probe different “directions” of unmeasured confounding).

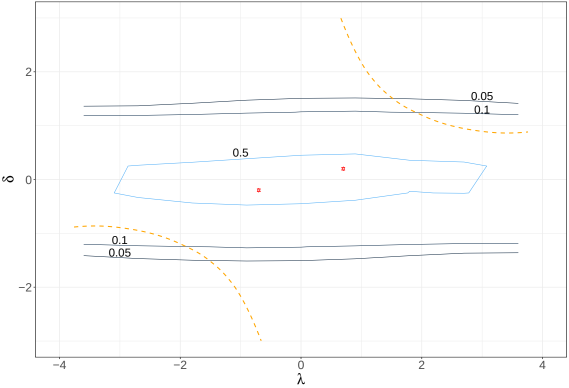
Some issues with the last analysis

Recall the model:

$$
\begin{aligned}
&A \perp\!\!\!\perp Y(a) \mid \boldsymbol{X}, U, \text{ for } a = 0, 1,\\
&U \mid \boldsymbol{X} \sim \text{Bernoulli}(0.5),\\
&A \mid \boldsymbol{X}, U \sim \text{Bernoulli}(\text{expit}(\boldsymbol{\kappa}^T \boldsymbol{X} + \lambda U)),\\
&Y(a) \mid \boldsymbol{X}, U \sim \text{N}(\beta a + \boldsymbol{\nu}^T \boldsymbol{X} + \delta U, \sigma^2) \text{ for } a = 0, 1,
\end{aligned}
$$

- Issue 1: The sensitivity parameters (λ, δ) are identifiable in this model. So it is **logically inconsistent** for us to vary the sensitivity parameter.
- Issue 2: In the calibration plot, partial R^2 for observed and unobserved confounders are not directly comparable because they use different reference models.

Visualization the the identifiability of (λ, δ)



- Red dots are the MLE;
- Solid curves are rejection regions for the likelihood ratio test;
- Dashed curves are where estimated ATE is reduced by a half.

Lesson: Parametric sensitivity models need to be carefully constructed to be useful.

What is a sensitivity model?

General setup

$$\text{Observed data } \boldsymbol{O} \overset{\textit{infer}}{\Longrightarrow} \text{Distribution of the full data } \boldsymbol{F}.$$

Recall our prototypical example: $\boldsymbol{O} = (\boldsymbol{X}, A, Y)$,
$\boldsymbol{F} = (\boldsymbol{X}, A, Y(0), Y(1))$.

An abstraction

A *sensitivity model* is a family of distributions $\mathcal{F}_{\theta,\eta}$ of $\boldsymbol{F}$ that satisfies:

1. *Augmentation:* Setting $\eta = 0$ corresponds to a primary analysis assuming no unmeasured confounders.
2. *Model identifiability:* Given η, the implied marginal distribution $\mathcal{O}_{\theta,\eta}$ of the observed data $\boldsymbol{O}$ is identifiable.

Statistical problem

Given η (or the range of η), use the observed data to make inference about some causal parameter $\beta = \beta(\theta, \eta)$.

Understanding sensitivity models

Observational equivalence

- $\mathcal{F}_{\theta,\eta}$ and $\mathcal{F}_{\theta'',\eta'}$ are said to be *observationally equivalent* if $\mathcal{O}_{\theta,\eta} = \mathcal{O}_{\theta',\eta'}$. We write this as $\mathcal{F}_{\theta,\eta} \simeq \mathcal{F}_{\theta',\eta'}$.
- Equivalence class $[\mathcal{F}_{\theta,\eta}] = \{\mathcal{F}_{\theta',\eta'} \mid \mathcal{F}_{\theta,\eta} \simeq \mathcal{F}_{\theta',\eta'}\}$.

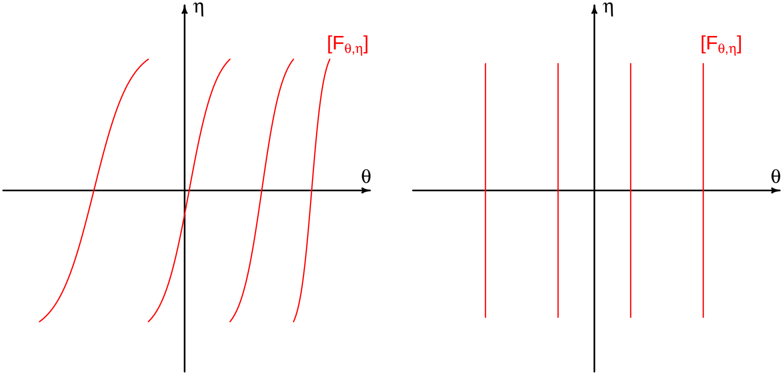
Types of sensitivity models

Testable models: When $\mathcal{F}_{\theta,\eta}$ is not rich enough, $[\mathcal{F}_{\theta,\eta}]$ is a singleton and η can be identified from the observed data (should be avoided in practice).

Global models: For any (θ, η) and η', there exists θ' s.t. $\mathcal{F}_{\theta',\eta'} \simeq \mathcal{F}_{\theta,\eta}$.

Separable models: For any (θ, η), $\mathcal{F}_{\theta,\eta} \simeq \mathcal{F}_{\theta,0}$.

A visualization



Left: **Global** sensitivity models; Right: **Separable** sensitivity models.

Model augmentation

In general, there are 3 ways to build a sensitivity model (underlined are nonidentifiable distributions):

1. Simultaneous model:

$$\begin{aligned}&f_{\boldsymbol{X},U,A,Y(a)}(\boldsymbol{x},u,a',y)\\=&f_{\boldsymbol{X}}(\boldsymbol{x})\cdot\underline{f_{U\mid\boldsymbol{X}}(u\mid\boldsymbol{x})}\cdot\underline{f_{A\mid\boldsymbol{X},U}(a'\mid\boldsymbol{x},u)}\cdot\underline{f_{Y(a)\mid\boldsymbol{X},U}(y\mid\boldsymbol{x},u)}.\end{aligned}$$

2. Treatment model (also called selection model, primal model, Tukey's factorization):

$$f_{\boldsymbol{X},A,Y(a)}(\boldsymbol{x},a',y)=f_{\boldsymbol{X}}(\boldsymbol{x})\cdot\underline{f_{A\mid Y(a),\boldsymbol{X}}(a'\mid y,\boldsymbol{x})}\cdot\underline{f_{Y(a)\mid\boldsymbol{X}}(y\mid\boldsymbol{x})}.$$

3. Outcome model (also called pattern mixture model, dual model):

$$f_{\boldsymbol{X},A,Y(a)}(\boldsymbol{x},a',y)=f_{\boldsymbol{X}}(\boldsymbol{x})\cdot f_{A\mid\boldsymbol{X}}(a'\mid\boldsymbol{x})\cdot\underline{f_{Y(a)\mid A,\boldsymbol{X}}(y\mid a',\boldsymbol{x})}.$$

Different sensitivity models amount to different ways of specifying the nonidentifiable distributions [National Research Council, 2010]. Our paper gives a comprehensive review.

Statistical inference

Modes of inference

1. **Point identified** sensitivity analysis is performed **at a fixed η**.
2. **Partially identified** sensitivity analysis is performed **simultaneously** over $\eta \in H$ **for a given range H**.

Statistical guarantees of interval estimators

1. **Confidence interval** $[C_L(\boldsymbol{O}_{1:n}; \eta), C_U(\boldsymbol{O}_{1:n}; \eta)]$ satisfies

$$\inf_{\theta_0, \eta_0} \mathbb{P}_{\theta_0, \eta_0}\big\{\beta(\theta_0, \eta_0) \in [C_L(\eta_0), C_U(\eta_0)]\big\} \geq 1 - \alpha.$$

2. **Sensitivity interval** (also called uncertainty interval, confidence interval) $[C_L(\boldsymbol{O}_{1:n}; H), C_U(\boldsymbol{O}_{1:n}; H)]$ satisfies

$$\inf_{\theta_0, \eta_0} \mathbb{P}_{\theta_0, \eta_0}\big\{\beta(\theta_0, \eta_0) \in [C_L(H), C_U(H)]\big\} \geq 1 - \alpha. \qquad (1)$$

They look almost the same, but (1) is actually equivalent to

$$\inf_{\theta_0, \eta_0} \inf_{\mathcal{F}_{\theta, \eta} \simeq \mathcal{F}_{\theta_0, \eta_0}} \mathbb{P}_{\theta_0, \eta_0}\big\{\beta(\theta, \eta) \in [C_L(H), C_U(H)]\big\} \geq 1 - \alpha.$$

Methods for sensitivity analysis

- **Point identified** sensitivity analysis is basically the same as primary analysis with known "offset" η.
- **Partially identified** sensitivity analysis is much harder.

Partially identified inference

Let $\mathcal{F}_{\theta_0,\eta_0}$ be the truth. There are essentially two approaches:

Method 1 Directly make inference about the two ends:

$$\beta_L = \inf_{\eta \in H} \{\beta(\theta,\eta) \mid \mathcal{F}_{\theta,\eta} \simeq \mathcal{F}_{\theta_0,\eta_0}\},$$
$$\beta_U = \sup_{\eta \in H} \{\beta(\theta,\eta) \mid \mathcal{F}_{\theta,\eta} \simeq \mathcal{F}_{\theta_0,\eta_0}\}.$$

Method 2 Take the union of point identified interval estimators.

Method 1: Bound estimation

Suppose $H = H_\Gamma$ is indexed by a hyperparameter Γ. Consider

$$\beta_L(\Gamma) = \inf_{\eta \in H_\Gamma} \{\beta(\theta, \eta) \mid \mathcal{F}_{\theta,\eta} \simeq \mathcal{F}_{\theta_0,\eta_0}\}$$

Method 1.1: Separable bounds

- Suppose $\mathcal{F}_{\theta^*,0} \simeq \mathcal{F}_{\theta_0,\eta_0}$ (existence from global sensitivity model).
- For some models we can solve the optimization analytically and obtain
$$\beta_L(\Gamma) = g_L(\beta^*, \Gamma)$$
for known function g_L.
- "Separable" because the primary analysis (for β^*) is separated from the sensitivity analysis. Inference is thus a trivial extension of the primary analysis.
- Examples: Cornfield's bound [Cornfield et al., 1959]; E-value [Ding and VanderWeele, 2016].

Method 1: Bound estimation

Suppose $H = H_\Gamma$ is indexed by a hyperparameter Γ. Consider

$$\beta_L(\Gamma) = \inf_{\eta \in H_\Gamma} \{\beta(\theta, \eta) \mid \mathcal{F}_{\theta,\eta} \simeq \mathcal{F}_{\theta_0,\eta_0}\}$$

Method 1.2: Tractable bounds

- In other cases we may derive

$$\beta_L(\Gamma) = g_L(\theta^*, \Gamma)$$

  for some tractable functions g_L.
- Can then estimate $\beta_L(\Gamma)$ by replacing θ^* with its empirical estimate.
- Inference typically relies on establishing asymptotic normality:

$$\sqrt{n}(\hat{\beta}_L - \beta_L) \stackrel{d}{\to} \mathrm{N}(0, \sigma_L^2).$$

- Example: Vansteelandt et al. [2006]; Yadlowsky et al. [2018].
- Note: With large-sample theory, things get a bit tricky because confidence/sensitivity intervals can be pointwise or uniform. See Imbens and Manski [2004]; Stoye [2009].

Method 1: Bound estimation

Suppose $H = H_\Gamma$ is indexed by a hyperparameter Γ. Consider

$$\beta_L(\Gamma) = \inf_{\eta \in H_\Gamma} \{\beta(\theta, \eta) \mid \mathcal{F}_{\theta,\eta} \simeq \mathcal{F}_{\theta_0,\eta_0}\}$$

Method 1.3: Stochastic programming

- Suppose the model is separable and we may write $\beta(\theta, \eta) = \mathbb{E}_{\theta,\eta}[\beta(\boldsymbol{O}; \eta)] = \mathbb{E}_{\theta,0}[\beta(\boldsymbol{O}; \eta)]$.
- $\beta_L(\Gamma)$ is then the optimal value for the optimization problem

$$\begin{aligned} \text{minimize} \quad & \mathbb{E}_{\theta_0,0}[\beta(\boldsymbol{O}; \eta)] \\ \text{subject to} \quad & \eta \in H_\Gamma. \end{aligned}$$

- This is known as **stochastic programming** in the optimization literature. Solving the empirical version of the optimization problem is known as **sample average approximation**.
- In nice problems with compact H_Γ, the sample optimal value has a central limit theorem [Shapiro et al., 2014].
- Example: Tudball et al. [2019].

Method 2: Combining point identified inference

Method 2.1: Union confidence interval

- Suppose $[C_L(\eta), C_U(\eta)]$ are confidence intervals that satisfy

$$\inf_{\theta_0,\eta_0} \mathbb{P}_{\theta_0,\eta_0}\big\{\beta(\theta_0,\eta_0) \in [C_L(\eta_0), C_U(\eta_0)]\big\} \geq 1-\alpha.$$

- Then $[C_L(H), C_U(H)] = \cup_{\eta\in H}[C_L(\eta), C_U(\eta)]$ is a sensitivity interval:

$$\inf_{\theta_0,\eta_0} \mathbb{P}_{\theta_0,\eta_0}\big\{\beta(\theta_0,\eta_0) \in [C_L(H), C_U(H)]\big\} \geq 1-\alpha.$$

- Proof is a simple application of the union bound.
- Note: Can be improved to cover the partially identified region if the intervals have the same tail probabilities [Zhao et al., 2019].
- Using asymptotic theory, we often have

$$[C_L(\eta), C_U(\eta)] = \hat{\beta}(\eta) \mp z_{1-\frac{\alpha}{2}} \cdot \frac{\hat{\sigma}(\eta)}{\sqrt{n}}$$

- Computationally challenging because $\hat{\sigma}(\eta)$ is usually complicated.

Method 2: Combining point identified inference

Method 2.2: Percentile bootstrap [Zhao et al., 2019]

1. For fixed η, use percentile bootstrap (b indexes data resample):

$$[C_L(\eta), C_U(\eta)] = \left[Q_{\frac{\alpha}{2}}\left(\hat{\hat{\beta}}_b(\eta)\right), Q_{1-\frac{\alpha}{2}}\left(\hat{\hat{\beta}}_b(\eta)\right)\right].$$

2. Use the generalized minimax inequality to interchange quantile and infimum/supremum:

Percentile bootstrap sensitivity interval

$$Q_{\frac{\alpha}{2}}\left(\inf_{\eta} \hat{\hat{\beta}}_b(\eta)\right) \le \inf_{\eta} Q_{\frac{\alpha}{2}}\left(\hat{\hat{\beta}}_b(\eta)\right) \le \sup_{\eta} Q_{1-\frac{\alpha}{2}}\left(\hat{\hat{\beta}}_b(\eta)\right) \le Q_{1-\frac{\alpha}{2}}\left(\sup_{\eta} \hat{\hat{\beta}}_b(\eta)\right).$$

Union sensitivity interval

Advantages

- Computation is reduced to repeating Method 1.3 over resamples.
- Only need coverage guarantee for $[C_L(\eta), C_U(\eta)]$ for **fixed** η.

An analogue

Point-identified parameter: Efron's bootstrap

Point estimator $\xrightarrow{\textit{Bootstrap}}$ Confidence interval

Partially identified parameter: Three ideas

Optimization *Percentile Bootstrap* *Minimax inequality*

Extrema estimator $\Longrightarrow$ Sensitivity interval

Method 2: Combining point identified inference

Method 2.3: Supreme of p-value

- **Rosenbaum's sensitivity analysis** is the hypothesis testing analogue of Method 2.1 (Union CI).
- Suppose we have valid p-values (for fixed η) that satisfies

$$\inf_{\theta_0,\eta_0} \mathbb{P}_{\theta_0,\eta_0}\{p(\boldsymbol{O}_{1:n};\eta_0) \le \alpha\} \le \alpha.$$

- Then their supremum can be used for partially identified inference:

$$\inf_{\theta_0,\eta_0} \mathbb{P}_{\theta_0,\eta_0}\left\{\sup_{\eta\in H} p(\boldsymbol{O}_{1:n};\eta) \le \alpha\right\} \le \alpha$$

- Rosenbaum [1987, 2002] used randomization tests to construct the p-value (for matched observational studies).
- He then used Holley's inequality in probabilistic combinatorics to efficiently compute $\sup_{\eta\in H} p(\boldsymbol{O}_{1:n};\eta)$.

Interpretation of sensitivity analysis

Two good ideas

1. Sensitivity value.
2. Calibration using measured confounders.

Idea 1: Sensitivity value

- Sensitivity value (or sensitivity frontier) is the value of the sensitivity parameter η (or hyperparameter Γ) where **some qualitative conclusions change**.
- Example: In Blattman and Annan [2010], this is where the estimated ATE is halved.
- Example: In Rosenbaum's sensitivity analysis, this is where we can no longer reject the causal null hypothesis.
- Analogue to the p-value for the primary analysis.
- Often exists a **phase transition** for partially identified inference: if Γ is too large (compared to the treatment effect), can never reject the causal null even with enormous n [Rosenbaum, 2004; Zhao, 2019].

Interpretation of sensitivity analysis

Calibration using measured confounders

- A practical solution to quantifying the sensitivity.
- Some good heuristics [e.g. Imbens, 2003; Hsu and Small, 2013] but often with subtle issues. Easier in carefully parameterized models [Cinelli and Hazlett, 2020].
- No unifying framework, lots of work needed.
- Perhaps what we need is to build calibration into the sensitivity model (e.g. let H_Γ be defined by calibration).

Take-home messages

- Three components of a sensitivity analysis: **model augmentation**, **statistical inference**, **interpretation**.
- **Sensitivity model** = Parametrizing the full data distribution = Overparameterizing the observed data distribution.
  Understand them by **observational equivalence** classes.
- Different ways of model augmentation by **different factorizations** of the full data distribution.
- **Point identified** inference versus **partially identified** inference.
- Two general approaches for partially identified inference:
  1. Bound estimation;
  2. Combining point identified inference.
- Two good ideas for interpretation:
  1. Sensitivity value;
  2. Calibration using measured confounders.
- **Lots of future work needed!**

References I

C. Blattman and J. Annan. The consequences of child soldiering. *The Review of Economics and Statistics*, 92(4):882–898, 2010. doi: 10.1162/REST_a_00036.

C. Cinelli and C. Hazlett. Making sense of sensitivity: extending omitted variable bias. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(1):39–67, 2020. doi: 10.1111/rssb.12348.

J. Cornfield, W. Haenszel, E. Hammond, A. Lilienfeld, M. Shimkin, and E. Wynder. Smoking and lung cancer. *Journal of the National Cancer Institute*, 22:173–203, 1959.

P. Ding and T. J. VanderWeele. Sensitivity analysis without assumptions. *Epidemiology*, 27:368–377, 2016.

J. Y. Hsu and D. S. Small. Calibrating sensitivity analyses to observed covariates in observational studies. *Biometrics*, 69:803–811, 2013.

G. W. Imbens. Sensitivity to exogeneity assumptions in program evaluation. *American Economic Review*, 93:126–132, 2003.

G. W. Imbens and C. F. Manski. Confidence intervals for partially identified parameters. *Econometrica*, 72(6):1845–1857, 2004.

National Research Council. *The prevention and treatment of missing data in clinical trials*. National Academies Press, 2010.

References II

P. R. Rosenbaum. Sensitivity analysis for certain permutation inferences in matched observational studies. *Biometrika*, 74:13–26, 1987.

P. R. Rosenbaum. *Observational Studies*. Springer., 2002.

P. R. Rosenbaum. Design sensitivity in observational studies. *Biometrika*, 91 (1):153–164, 2004.

A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto. *Sensitivity analysis in practice: A guide to assessing scientific models*. John Wiley & Sons, Ltd, 2004.

A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on stochastic programming: modeling and theory*. SIAM, 2014.

J. Stoye. More on confidence intervals for partially identified parameters. *Econometrica*, 77(4):1299–1315, 2009.

M. Tudball, Q. Zhao, R. Hughes, K. Tilling, and J. Bowden. An interval estimation approach to sample selection bias, 2019.

S. Vansteelandt, E. Goetghebeur, M. G. Kenward, and G. Molenberghs. Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Statistica Sinica*, 16(3):953–979, 2006.

References III

S. Yadlowsky, H. Namkoong, S. Basu, J. Duchi, and L. Tian. Bounds on the conditional and average treatment effect with unobserved confounding factors, 2018.

Q. Zhao. On sensitivity value of pair-matched observational studies. *Journal of the American Statistical Association*, 114(526):713–722, 2019.

Q. Zhao, D. S. Small, and B. B. Bhattacharya. Sensitivity analysis for inverse probability weighting estimators via the percentile bootstrap. *Journal of the Royal Statistical Society (Series B)*, 81(4):735–761, 2019.