# Causal Analysis

Impact Evaluation and Causal Machine Learning with Applications in R

## Chapter 10: Partial Identification and Sensitivity Analysis

# Table of Contents

# Partial Identification

From point to partial identification:

- Previous approaches were based on assumptions that allow identifying a single value for the causal effect, known as point identification.
- Partial identification yields an interval or set of possible values for the causal effect.
- Arises when we impose weaker (or no) statistical assumptions.
- Preferable when stronger assumptions for point identification are not plausible.

# Trade-off in Data-based Causal Analysis

- Stronger statistical assumptions allow for more precise causal effect estimation.
- However, stronger assumptions increase the risk of bias if they fail to accurately represent real-life behavior.
- Partial identification makes these trade-offs explicit:
  - With few or no assumptions, the set of possible treatment effect values is large.
  - As more assumptions are added, the set becomes narrower.
  - Eventually, the set collapses to point identification (a single causal effect value) with constraints like selection on observables.

# Partially Observed Mean Potential Outcomes

- Consider the mean of the potential outcome under treatment:
  - $Y(1)$ is observed for treated observations, with probability $\Pr(D = 1)$.
  - $Y(1)$ is not observed for nontreated observations, with probability $\Pr(D = 0) = 1 - \Pr(D = 1)$.
- Mean of the potential outcome under treatment:

$$\begin{aligned} E[Y(1)] &= E[Y(1)|D = 1] \cdot \Pr(D = 1) + E[Y(1)|D = 0] \cdot \Pr(D = 0), \\ &= \underbrace{E[Y|D = 1] \cdot \Pr(D = 1)}_{\text{observed}} + \underbrace{E[Y(1)|D = 0]}_{\text{unobserved}} \cdot \underbrace{\Pr(D = 0)}_{\text{observed}}. \end{aligned} \tag{10.1}$$

- Analogous, mean of the potential outcome under non-treatment:

$$\begin{aligned} E[Y(0)] &= E[Y(0)|D = 1] \cdot \Pr(D = 1) + E[Y(0)|D = 0] \cdot \Pr(D = 0), \\ &= \underbrace{E[Y(0)|D = 1]}_{\text{unobserved}} \cdot \underbrace{\Pr(D = 1)}_{\text{observed}} + \underbrace{E[Y|D = 0] \cdot \Pr(D = 0)}_{\text{observed}}. \end{aligned} \tag{10.2}$$

# Bounding Mean Potential Outcomes

- Under the independence assumption in expression (3.1), we could easily identify the ATE.
- However, we refrain from this strong assumption and permit selection into treatment, which implies that:

$$E[Y(1)|D=0] \neq E[Y(1)|D=1], \text{ and}$$

$$E[Y(0)|D=1] \neq E[Y(0)|D=0].$$

- Without further assumptions, the ATE cannot be point-identified.
- By putting upper and lower bounds on the unobserved means $E[Y(1)|D=0]$ and $E[Y(0)|D=1]$, we restrict the means to lie within a range of minimum and maximum values, in order to bound the ATE.

# Upper and Lower Bounds

Assume theoretical upper and lower bounds for the outcome:

- Maximum value of the outcome: $y^{UB}$.
- Minimum value of the outcome: $y^{LB}$.

Construct bounds for the potential outcomes:

$$
\begin{aligned}
E[Y(1)]^{UB} &= E[Y|D=1] \cdot \Pr(D=1) + y^{UB} \cdot \Pr(D=0), \\
E[Y(1)]^{LB} &= E[Y|D=1] \cdot \Pr(D=1) + y^{LB} \cdot \Pr(D=0), \\
E[Y(0)]^{UB} &= y^{UB} \cdot \Pr(D=1) + E[Y|D=0] \cdot \Pr(D=0), \\
E[Y(0)]^{LB} &= y^{LB} \cdot \Pr(D=1) + E[Y|D=0] \cdot \Pr(D=0). \qquad (10.3)
\end{aligned}
$$

# Bounds on the Average Treatment Effect

As discussed in Manski (1990), upper and lower bounds on the ATE can be calculated as:

$$\Delta^{UB} = E[Y(1)]^{UB} - E[Y(0)]^{LB}, \text{ and}$$
$$\Delta^{LB} = E[Y(1)]^{LB} - E[Y(0)]^{UB}. \tag{10.4}$$

Assumptions to tighten the upper and lower bounds:

- Different upper and lower bounds for treated and nontreated observations, instead of $y^{UB}$ and $y^{LB}$.
- Monotone treatment response (Manski, 1997): For a binary treatment, the mean potential outcome under treatment cannot be lower than under nontreatment.
- Monotone treatment selection (Manski and Pepper, 2000): Subjects select into treatment such that the mean potential outcomes of the treated and nontreated groups can be ordered.

# Monotone Treatment Response and Treatment Selection

Monotone Treatment Response (MTR):

- MTR assumption tightens bounds by assuming $E[Y(1)] \geq E[Y(0)]$.
- ATE is assumed to be nonnegative: $\Delta = E[Y(1)] - E[Y(0)] \geq 0$.
- Lower bound of the ATE is adjusted to $\Delta^{LB} = \max(E[Y(1)]^{LB} - E[Y(0)]^{UB}, 0)$.

Monotone Treatment Selection (MTS):

- (Positive) MTS implies that treated individuals have weakly higher mean potential outcomes than nontreated.
- $E[Y(1)|D = 1] \geq E[Y(1)|D = 0]$, $E[Y(0)|D = 1] \geq E[Y(0)|D = 0]$.
- Upper and lower bounds simplify to $E[Y(0)]^{LB} = E[Y|D = 0]$ and $E[Y(1)]^{UB} = E[Y|D = 1]$.

# Instrumental Variable Assumptions

Partial identification with instrumental variables:

- Bounds can also be tightened using instrumental variables, see e.g. Robins (1989) and Balke and Pearl (1997).
- Assuming mean independence for a binary instrument $Z$ implies that $E[Y(1)|Z = 1] = E[Y(1)|Z = 0]$ and $E[Y(0)|Z = 1] = E[Y(0)|Z = 0]$.
- The instrument must not affect the mean potential outcome except through the treatment, which implies an exclusion restriction.

# Bounds with Instrumental Variables

- The mean independence assumption entails the following bounds on the mean potential outcomes:

$$
\begin{aligned}
E[Y(1)]^{UB} = \min \big( & E[Y|D=1,Z=1] \cdot \Pr(D=1|Z=1) + y^{UB} \cdot \Pr(D=0|Z=1), \\
& E[Y|D=1,Z=0] \cdot \Pr(D=1|Z=0) + y^{UB} \cdot \Pr(D=0|Z=0) \big), \\
E[Y(1)]^{LB} = \max \big( & E[Y|D=1,Z=1] \cdot \Pr(D=1|Z=1) + y^{LB} \cdot \Pr(D=0|Z=1), \\
& E[Y|D=1,Z=0] \cdot \Pr(D=1|Z=0) + y^{LB} \cdot \Pr(D=0|Z=0) \big), \\
E[Y(0)]^{UB} = \min \big( & y^{UB} \cdot \Pr(D=1|Z=1) + E[Y|D=0,Z=1] \cdot \Pr(D=0|Z=1), \\
& y^{UB} \cdot \Pr(D=1|Z=0) + E[Y|D=0,Z=0] \cdot \Pr(D=0|Z=0) \big), \\
E[Y(0)]^{LB} = \max \big( & y^{LB} \cdot \Pr(D=1|Z=1) + E[Y|D=0,Z=1] \cdot \Pr(D=0|Z=1), \\
& y^{LB} \cdot \Pr(D=1|Z=0) + E[Y|D=0,Z=0] \cdot \Pr(D=0|Z=0) \big). \quad (10.5)
\end{aligned}
$$

# Intuition of Instrument-Based Bounds

Intuition:

- Since $Z$ does not affect potential outcomes on average, we can compute bounds conditional on $Z$ and intersect these bounds across different values of $Z$ to tighten them.
- After computing bounds conditional on $Z$, take the minimum of the upper bounds and the maximum of the lower bounds on the mean potential outcomes across $Z = 1$ and $Z = 0$.

Note that:

- Imposes no assumptions about the relationship between the treatment and the instrument such as the existence of first-stage effects or monotonicity.
- Implies that the LATE on compliers is not point identified.
- In analogy to the ATE, LATE can be bounded under IV independence assumptions.

# Monotone Instrumental Variable Assumption

- Monotone instrumental variable (MIV) assumption: mean potential outcomes are monotonic in the instrument: $E[Y(1)|Z=1] \geq E[Y(1)|Z=0]$, $E[Y(0)|Z=1] \geq E[Y(0)|Z=0]$.
- MIV (Manski and Pepper, 2000) typically entails wider ATE bounds than (stronger) mean independence.

Further possible assumptions:

- Assume specific ordering of potential outcomes across treatment compliance types (introduced in figure 6.1. of chapter 6), as in Flores and Flores-Lagunes (2013) to bound the LATE.
- For instance, assume that always takers ($D(1)=1, D(0)=1$) have weakly higher mean potential outcomes under treatment than the compliers ($D(1)=1, D(0)=0$): $E[Y(1)|D(1)=1, D(0)=1] \geq E[Y(1)|D(1)=1, D(0)=0]$.
- Combine multiple assumptions such as MTR and MIV.

# Sample Selection and Outcome Attrition

- Sample selection occurs if outcomes are observed only for a selective subpopulation, creating potential endogeneity issues.
- Define selection compliance types using a binary selection indicator $O$ for whether $Y$ is observed as a function of $D$:

$$O(1) = 1,\ O(0) = 1 \text{ (always selected)}$$
$$O(1) = 0,\ O(0) = 0 \text{ (never selected)}$$
$$O(1) = 1,\ O(0) = 0 \text{ (selection compliers)}$$
$$O(1) = 0,\ O(0) = 1 \text{ (selection defiers)}$$

**Example**

- Evaluation of a training program, wage outcome is only observed for employed individuals.
- Selected individuals with $O = 1$ are those who are employed.

# Bounds under Sample Selection

- Under sample selection, one may focus on the effect on always selected (with outcomes observed in either treatment state).
- Assuming dominance implies that treated observations have weakly higher potential outcomes than nontreated observations.
- Assuming monotonicity of selection in the treatment rules out selection defiers, see Lee (2009).
- These assumptions yield the following bounds on the ATE for the always selected group, see Zhang and Rubin (2003):

$$\Delta^{UB}_{O(1)=1,O(0)=1} = E[Y|D=1,O=1,Y\geq y^*] - E[Y|D=0,O=1],$$
$$\Delta^{LB}_{O(1)=1,O(0)=1} = E[Y|D=1,O=1] - E[Y|D=0,O=1], \qquad (10.6)$$

  where $y^*$ is chosen such that the lowest outcomes in the group with $D=1$ and $O=1$ that correspond to the share of compliers in that group are below this value.

# Further Partial Identification Approaches

- Semenova (2020) considers partial identification under sample selection and monotonicity of selection in the treatment and uses machine learning to control for key covariates $X$ jointly affecting selection $O$ and outcome $Y$ in a data-driven way.
- Covariates may be useful to make assumptions (like monotonicity) more plausible and to tighten bounds.
- Chen and Flores (2015) extend partial identification under sample selection (monotonicity, dominance) to the IV context (e.g. noncompliance) to bound the LATE on compliers.
- Bounding strategies have also been applied in mediation analysis with random treatment and endogenous mediator (e.g. Sjölander, 2009).

# Inference without Min/Max Operators

- To estimate the bounds and their variances, one needs to pay attention to whether the partial identification results contain minimum or maximum operators.
- Bounds on mean potential outcomes in equations (10.3) do not include such operators.
- These bounds can be estimated $\sqrt{n}$-consistently with an asymptotically normal distribution, similar to linear regression.
- Imbens and Manski (2004) suggest computing the 95% confidence interval for the partially identified ATE as follows, under the condition that the difference between the upper and lower bounds of the ATE is nonnegligible:

$$\left(\hat{\Delta}^{LB} - 1.645 \cdot \hat{\sigma}^{LB}, \hat{\Delta}^{UB} + 1.645 \cdot \hat{\sigma}^{UB}\right), \tag{10.7}$$

where $\hat{\Delta}^{LB}, \hat{\Delta}^{UB}$ are the lower and upper bounds, and $\hat{\sigma}^{LB}$ and $\hat{\sigma}^{UB}$ denote the respective standard errors.

# Inference with Min/Max Operators

- Bounds involving nondifferentiable operators, such as minimum or maximum functions, are challenging to estimate due to their non-smoothness.
- As shown in Hirano and Porter (2012), these bounds cannot be estimated without bias, even in large samples.
- Conventional methods often fail to provide confidence intervals with accurate coverage probabilities.
- Alternative methods:
  - Half-median-unbiased confidence intervals, see Chernozhukov et al. (2013).
  - Bootstrap procedures for bias correction, see Kreider and Pepper (2007), see next slide for details.
  - Repeated sampling for directly constructing confidence intervals, see Chernozhukov, Hong, and Tamer (2007) and Romano and Shaikh (2008).

# Bootstrap Bias Correction for Inference with Min/Max Operators

- Bootstrap bias correction is based on drawing repeated random samples from the original data, estimating bounds in each sample, and averaging the estimates to approximate the bias.
- The bias of the lower bound is approximated as follows:

$$\frac{1}{B}\sum_{b=1}^{B}\hat{\Delta}_b^{LB} - \hat{\Delta}^{LB},$$

where $B$ is the number of bootstrap samples, $\hat{\Delta}_b^{LB}$ is the lower bound in the $b$-th bootstrap sample, and $\hat{\Delta}^{LB}$ is the lower bound in the original sample.

# Table of Contents

# Sensitivity Analysis

From partial identification to sensitivity analysis:

- Partial identification and sensitivity analysis approach uncertainty about identifying assumptions from different directions.
- Partial identification drops point identifying assumptions altogether (or replaces them by weaker restrictions).
- Sensitivity analysis investigates the sensitivity of the causal effect to deviations from assumptions that yield point identification.

# Considering the Selection-on-Observables Assumption

- Selection-on-observables assumption in expression (4.1):

$$\{Y(1), Y(0)\} \perp D|X.$$

- Potential outcomes are independent of the treatment $D$, conditional on observed covariates $X$.
- Violated if confounder $U$ jointly affects treatment $D$ and outcome $Y$ conditional on $X$.
- In this case, we should also control for $U$:

$$\{Y(1), Y(0)\} \perp D|X, U. \tag{10.8}$$

- However, we cannot control for $U$ because it is not observed.
- Sensitivity analysis consists of makings assumptions about how strongly $U$ is associated with $D$ and $Y$.

# Parametric and Nonparametric Approaches

- Several approaches are based on (1) parametrically modeling $\Pr(D = 1|X, U)$ or the outcome $Y$ as a function of $D$, $X$, and $U$ and (2) varying the values of $U$ over a presumably plausible range.
- See, for instance, Rosenbaum and Rubin (1983a), Imbens (2003), and Altonji, Elder, and Taber (2008).
- Ichino, Mealli, and Nannicini (2008) provide a nonparametric method (without parametric assumptions on the treatment or outcome models), but restrict $U$ to be discrete.
- We subsequently consider two approaches to sensitivity analysis that neither rely on parametric treatment/outcome models nor restrict the distribution of unobservables.

# Sensitivity Analysis by Rosenbaum

Rosenbaum (1995) suggests the sensitivity parameter $\Gamma$:

- Based on the odds ratios of the observed propensity score $\Pr(D = d|X)$ and the unknown propensity score $\Pr(D = d|X, Y(d))$.
- Formally, $\Gamma$ is assumed to satisfy:

$$\frac{1}{\Gamma} \leq \frac{\Pr(D = d|X = x)/(1 - \Pr(D = d|X = x))}{\Pr(D = d|X = x, Y(d) = y)/(1 - \Pr(D = d|X = x, Y(d) = y))} \leq \Gamma \quad (10.9)$$

  for any feasible covariate and outcome values $x$ and $y$.

- $\Gamma = 1$: No confounding, i.e., selection on observables holds, implying that $\Pr(D = d|X = x) = \Pr(D = d|X = x, Y(d) = y)$ such that both odds ratios are the same.
- $\Gamma > 1$: Confounding, i.e., deviations from the selection-on-observables assumption.

# Bounds on Mean Potential Outcomes

- Mean potential outcomes given $X$ can be computed as follows, as discussed by Kallus, Mao, and Zhou (2019):

$$E[Y(d)|X=x] = \frac{\int y \frac{f(D=d,Y=y|X=x)}{\Pr(D=d|X=x,Y(d)=y)} dy}{\int \frac{f(D=d,Y=y|X=x)}{\Pr(D=d|X=x,Y(d)=y)} dy}. \tag{10.10}$$

- Observed conditional density of $D$ and $Y$ given covariates $X$, denoted by $f(D=d, Y=y|X=x)$.
- Unobserved propensity score $\Pr(D=d|X=x, Y(d)=y)$ of expression (10.9)
- Compute upper bounds $E[Y(d)|X=x]^{UB}$ and lower bounds $E[Y(d)|X=x]^{LB}$ from equation (10.10), considering all $\Pr(D=d \mid X=x, Y(d)=y)$ satisfying constraint (10.9) under worst-case confounding $\Gamma$.

# Bounds on Treatment Effects

- Using the bounds on the potential outcomes, bounds on the conditional average treatment effect (CATE) are obtained:

$$\Delta_x^{UB} = E[Y(1)|X = x]^{UB} - E[Y(0)|X = x]^{LB},$$
$$\Delta_x^{LB} = E[Y(1)|X = x]^{LB} - E[Y(0)|X = x]^{UB}. \quad (10.11)$$

- $\Delta_x^{UB}$: Upper bound on the CATE for covariate composition $X = x$.
- $\Delta_x^{LB}$: Lower bound on the CATE for covariate composition $X = x$.
- Averaging over these bounds also yields upper and lower bounds on the ATE, $\Delta^{UB} = E[\Delta_x^{UB}]$ and $\Delta^{LB} = E[\Delta_x^{LB}]$.

# Sensitivity Analysis by Masten and Poirier

Masten and Poirier (2018) suggest the sensitivity parameter $\mathcal{C}$:

- Based on maximum absolute difference in the propensity scores when controlling for $X$ alone versus both $X$ and $Y(d)$:

$$|\Pr(D = 1|X = x) - \Pr(D = 1|X = x, Y(d) = y)| \leq \mathcal{C}, \tag{10.12}$$

  for $d \in \{1, 0\}$ and any values $x, y$ occurring in the population.
- $\mathcal{C}$ bounds the absolute difference in treatment probabilities due to confounding.
- $\mathcal{C}$ can take values from 0 (no confounding) to 1 (maximum confounding).
- To assess the sensitivity of causal effects, consider bounds on conditional quantiles of potential outcomes.

# Quantile-based Bounds

- Let $F^{-1}_{Y(d)|X=x}(\tau)$ be the $\tau$-quantile of the potential outcome $Y(d)$ given $X = x$.
- Let $F^{-1}_{Y|D=d,X=x}(\tau)$ be the $\tau$-quantile of the observed outcome given $D = d$ and $X = x$.
- Derive bounds using sensitivity parameter $\mathcal{C}$:

$$F^{-1}_{Y(d)|X=x}(\tau)^{UB} = F^{-1}_{Y|D=d,X=x}(\tau'),$$

$$\text{with } \tau' = \min\left(\tau + \frac{\mathcal{C}}{\Pr(D=d|X=x)} \cdot \min(\tau, 1-\tau), \frac{\tau}{\Pr(D=d|X=x)}, 1\right),$$

$$F^{-1}_{Y(d)|X=x}(\tau)^{LB} = F^{-1}_{Y|D=d,X=x}(\tau''),$$

$$\text{with } \tau'' = \max\left(\tau - \frac{\mathcal{C}}{\Pr(D=d|X=x)} \cdot \min(\tau, 1-\tau), \frac{\tau-1}{\Pr(D=d|X=x)} + 1, 0\right). \quad (10.13)$$

- Averaging the bounds in equation (10.13) across all ranks $\tau$ from 0 to 1 yields upper and lower bounds on the CATE, $\Delta^{UB}_X$ and $\Delta^{LB}_X$.
- Averaging the CATEs across all values of the covariates yields the upper and lower bounds on the ATE.

# Further Sensitivity Analyses

- Sensitivity analyses have also been suggested in the context of causal machine learning (Chernozhukov et al., 2021; Dorn, Guo, and Kallus, 2021) and beyond selection on observables.
- Example using IV framework: Assuming a violation of monotonicity of $D$ in $Z$ (such that defiers exist), the sensitivity of the LATE can be assessed by making assumptions about the shares of defier and outcome differences across compliance types (Huber, 2014b; Noack, 2021).
- Mediation analysis: Sensitivity analyses for robustness of direct and indirect effects to endogeneity of mediator (and treatment if not randomized), see for instance Tchetgen Tchetgen and Shpitser (2012), Vansteelandt and VanderWeele (2012), VanderWeele and Chiba (2014), and Hong, Qin and Yang (2018).