Lecture 9: Research ethics of causal inference

Assessing Unconfoundedness

Xi Chen

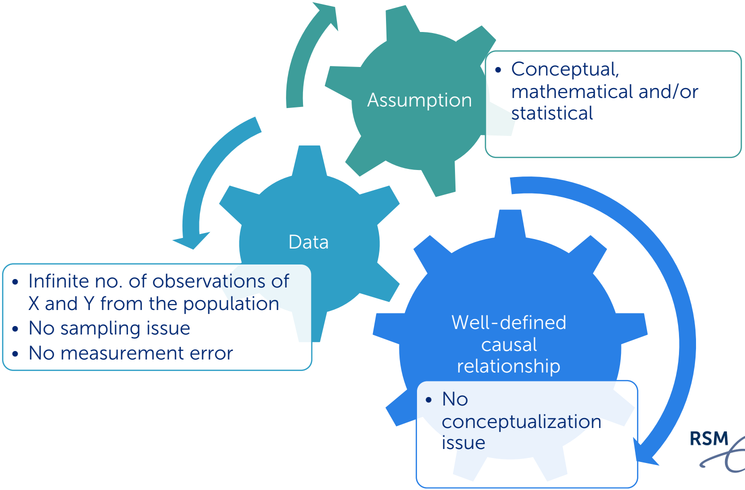
Rotterdam School of Management
Erasmus University Rotterdam

June 9, 2024

Outline

Outline

Elements of causal identification



Examples of assumptions

Stable Unit Treatment Value Assumption (SUTVA)

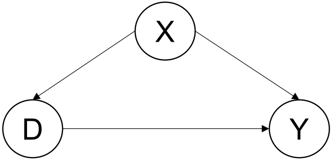
For a completely randomized experiment (CRE), we must have 1) the potential outcomes for any unit do not vary with the treatments assigned to other units, and 2) there are no different forms or versions of each treatment level.

Question: suppose we have data from a CRE, can we statistically test SUTVA with the data?

Examples of assumptions

(Conditional) Unconfoundedness

The potential outcomes are independent from the treatment status (given a set of control variables X).

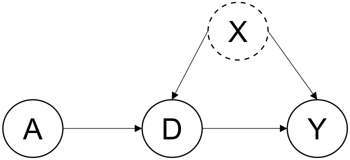


Question: With data on $\{Y, D, X\}$, can we statistically test the unconfoundedness assumption?

Examples of assumptions

Non-compliance (instrumental variables)

To identify the local average treatment effect, the treatment assignment (instrumental variables) must be 1) exogenous, 2) relevant, 3) excluded, and 4) monotonic (no defiers).

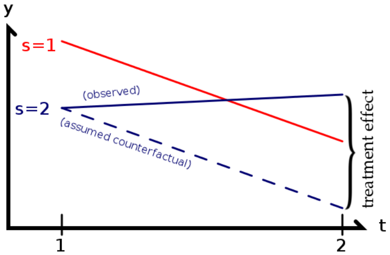


Question: With data on $\{Y, A, D\}$, can we statistically test the assumptions on instrumental variables?

Examples of assumptions

Parallel Trend Assumption in DID

In the absence of treatment, the difference between the treatment and control group is constant over time.

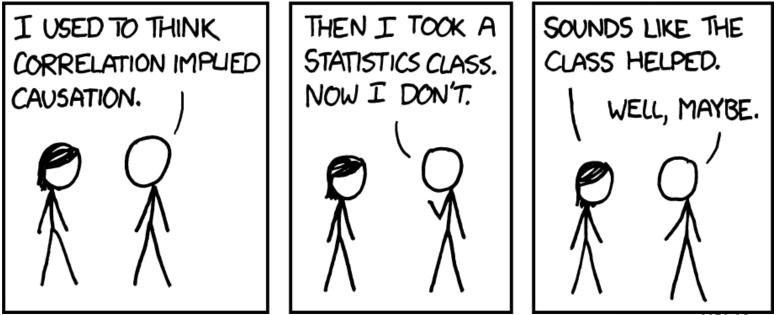


Question: With data on $\{Y_0, Y_1, D_0, D_1\}$, can we statistically test the parallel trend assumption?

Conceptual assumptions are statistically untestable!

Fact (Conceptual assumptions are untestable!)

Given the fundamental problem of causal inference, conceptual assumptions in causal inference are almost always statistically untestable.



$\int$

Assumptions determine conclusions

The golden formula of emprical work

Data + **Assumptions** = Conclusions

Assumptions can be subjective

In empirical analysis, researchers often make implicit assumptions based on subjectivity.

Assumptions determine conclusions

The golden formula of emprical work

Data + **Assumptions** = Conclusions

Assumptions can be subjective

In empirical analysis, researchers often make implicit assumptions based on subjectivity.

Assumptions determine conclusions

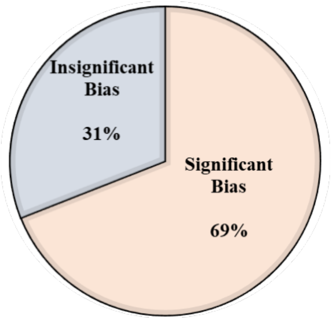
A recent study[1]:

- Asks teams of researchers to analyze the same data set - **football games**.
- The data include info. such as player clubs, birthday, height, weight, and positions...
- **Research question**: *are referees more likely to give red cards to dark- than light-skin-toned players*?

[1] Silberzahn, R. et al. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. Advances in Methods and Practices in Psychological Science, 1(3), 337-356.

Assumptions determine conclusions

Results of the study:



- Researchers make their own assumptions and get different results.
- "*What theoretical and/or statistical rationale was used for your choice of covariates?*"

Assumptions determine conclusions

Similar studies in other fields show consistent observations:

- Biology:
  - Gould, E. et al. (2023). Same data, different analysts: variation in effect sizes due to analytical decisions in ecology and evolutionary biology.
- Economics:
  - Breznau, N. et al. (2022). Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. Proceedings of the National Academy of Sciences, 119(44), e2203150119.
- Neuroscience:
  - Botvinik-Nezer, R. et al. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. Nature, 582(7810), 84-88.

A short discussion

Is a difference-in-difference design more credible than a simple regression?

A simple regression

$$Y_i = \beta D_i + \varepsilon$$

A DID regression

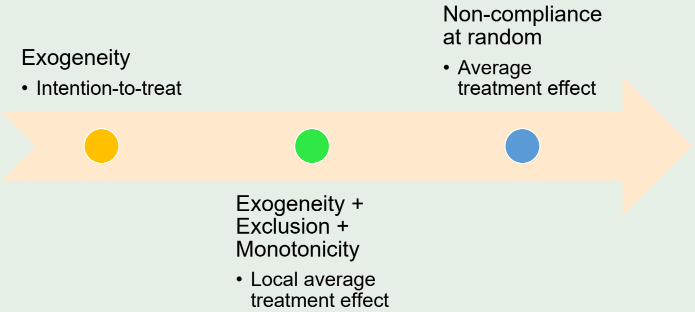
$$Y_{it} = D_i + T_t + \beta D_i T_t + \varepsilon$$

Where Y is the outcome, D is the treatment status, T is the before-after dummy.

The law of decreasing credibility

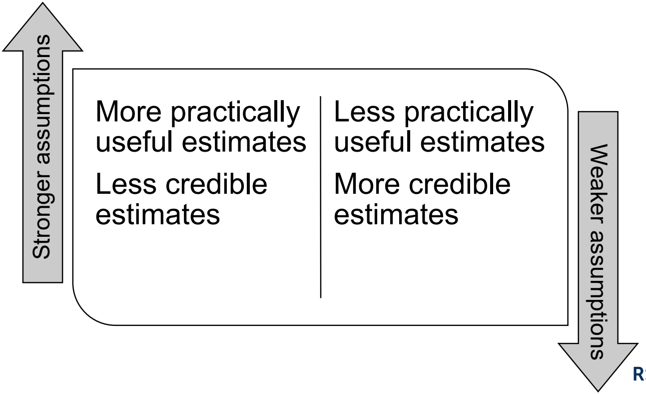
The credibility of inference decreases with the strength of the assumptions maintained.

Example (Non-compliance)



Some reflections on assumptions

- You always need assumptions for causal inference practices!
- Assumptions are the key.



Research ethics of causal inference

Communication

- Be transparent about your assumptions.

Investigation

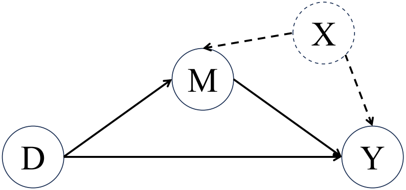
- Examine the plausibility of your assumptions.

Relaxation

- Remove implausible assumptions even if it means no positive findings.

An example of mediation analysis

The core problem of the traditional mediation analysis:

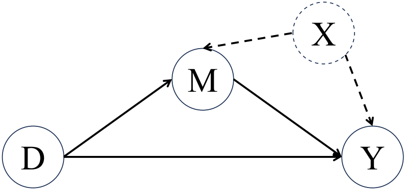


$$
\begin{cases}
Y &= \alpha_1 + \beta_1 D + e_1 \\
M &= \alpha_2 + \beta_2 D + e_2 \\
Y &= \alpha_3 + \beta_3 D + \beta_4 M + e_3
\end{cases}
$$

Traditional Mediation	Constructed IV
$\rho(e_2, e_3) = 0$	$\rho(e_2, e_3) \neq 0$
$-11.7 \ (-20.2, -4.8)$	$-5.2 \ (-12.1, 0.4)$

An example of mediation analysis

The core problem of the traditional mediation analysis:



$$\begin{cases} Y & = \alpha_1 + \beta_1 D + e_1 \\ M & = \alpha_2 + \beta_2 D + e_2 \\ Y & = \alpha_3 + \beta_3 D + \beta_4 M + e_3 \end{cases}$$

Traditional Mediation	Constructed IV
$\rho(e_2, e_3) = 0$	$\rho(e_2, e_3) \neq 0$
$-11.7 \ (-20.2, -4.8)$	$-5.2 \ (-12.1, 0.4)$

Outline

Unconfoundedness is untestable

Fact (Unconfoundedness is untestable)

Given the fundamental problem of causal inference, unconfoundedness is untestable, as only one potential outcome for an individual is observed.

- However, we can assess the plausibility of the unconfoundedness assumption by testing the implications.
- To this end, we formulate **consistency tests** and **sensitivity analysis**.

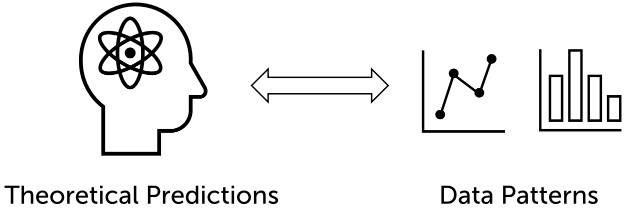
Outline

1. Assumptions in causal inference
2. Assessing unconfoundedness
   - Consistency tests
   - Sensitivity analysis

Consistency tests

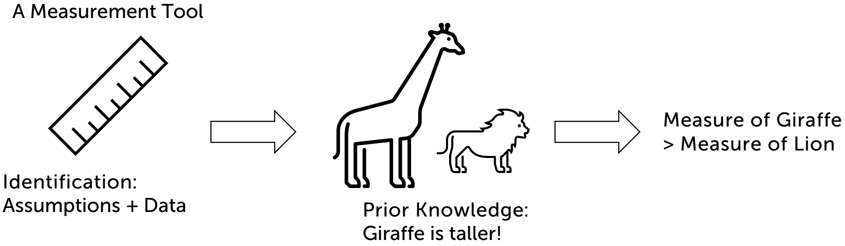
Definition (Consistency tests)

If unconfoundedness is a valid assumption, we should observe data patterns that are consistent with it.



Consistency tests

Identification assumptions and data as a measurement tool of causal effects.



Three types of consistency tests

Definition (Consistency tests with pseudo-outcomes)

If the identification strategy is valid, we will find an estimated effect of the original treatment on a pseudo-outcome to be consistent with our *a priori* belief on the pseudo-outcome.

Definition (Consistency tests with pseudo-treatments)

If the identification strategy is valid, we will find an estimated effect of the pseudo-treatment on the original outcome to be consistent with our *a priori* belief on the pseudo-treatment.

Three types of consistency tests

Definition (Consistency tests with pseudo-outcomes)

If the identification strategy is valid, we will find an estimated effect of the original treatment on a pseudo-outcome to be consistent with our *a priori* belief on the pseudo-outcome.

Definition (Consistency tests with pseudo-treatments)

If the identification strategy is valid, we will find an estimated effect of the pseudo-treatment on the original outcome to be consistent with our *a priori* belief on the pseudo-treatment.

Three types of consistency tests

Definition (Alternative populations)

If the identification strategy is valid, we will find an estimated effect (of the original treatment and outcome) on a treatment / control group from an alternative population to be consistent with our a priori belief on the treatment effect of the population.

Consistency tests: examples

Pseudo-outcomes (Dube et al. 2013)

Objective: Spillover effects of the US gun laws on gun-related crimes in Mexico.

Population: Mexican municipalities close to U.S. border.

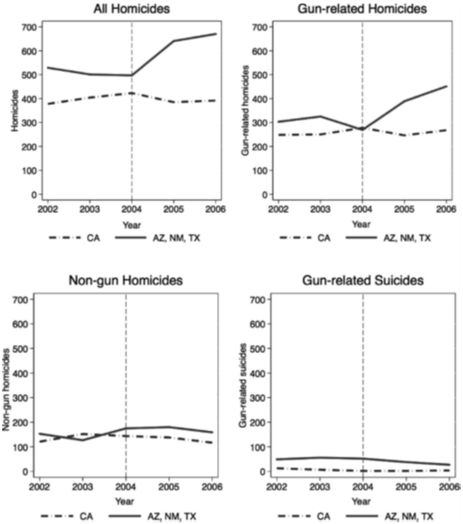
Treatment: Assault weapons from neighboring US state.

Outcome: Gun-related homicides.

Pseudo-outcomes: Accidents, non-gun homicides, and suicides.

Consistency tests: examples

Pseudo-outcomes (Dube et al. 2013)



Consistency tests: examples

Pseudo-treatment

- In DID, falsification tests that assume the treatment happens in the pre-treatment periods.
- Expecting no effect of these pseudo-treatments.

- In RDD, falsification tests that assume different cutoffs of the forcing variables.
- Expecting no effect of the pseudo-treatments from these cutoffs.

Consistency tests: examples

Alternative population (Chen et al. 2023)

Objective: How environmental uncertainty influences the diversity of alliances.

Population: Mutual fund firms in China.

Treatment: The jump of uncertainty during the Great Recession.

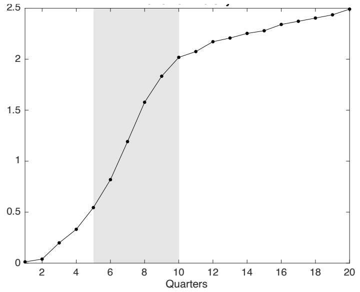
Outcome: Diversity of alliances.

Alternative population: Banks in China.

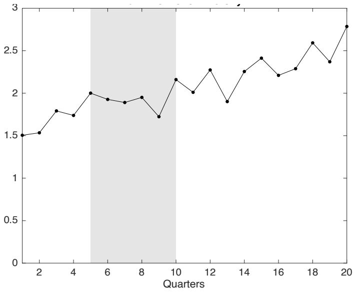
Consistency tests: examples

Alternative population (Chen et al. 2023)

- Banks are not affected by the Great Recession because of the massive stimulus package by the Chinese government.



(a) Mutual Fund Firms' Diversity



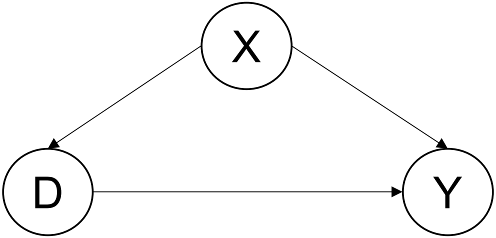
(b) Banks' Diversity

RSM Erasmus

Outline

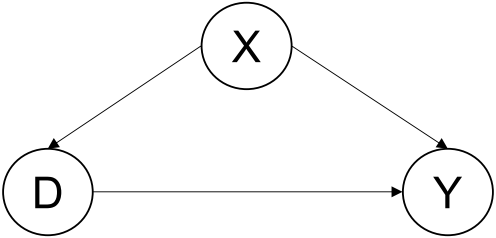
The basic idea of sensitivity analysis

- In practice, we often apply conditioning strategy and assume conditional unconfoundedness.
- This is a strong assumption, as there is always possible that confounders exist.
- **Question: if confounders exist, how they change the ATE estimation?**



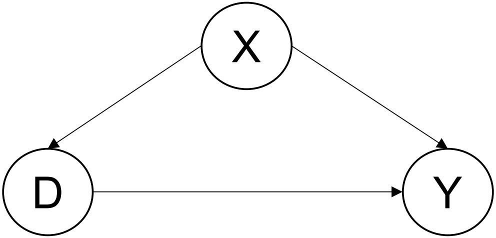
The basic idea of sensitivity analysis

- In practice, we often apply conditioning strategy and assume conditional unconfoundedness.
- This is a strong assumption, as there is always possible that confounders exist.
- **Question: if confounders exist, how they change the ATE estimation?**



The basic idea of sensitivity analysis

- In practice, we often apply conditioning strategy and assume conditional unconfoundedness.
- This is a strong assumption, as there is always possible that confounders exist.
- **Question: if confounders exist, how they change the ATE estimation?**



The logic of sensitivity analysis

Assume a confounder exists as U

Simulate the values of $U \sim f(U \mid D, Y, X)$

Use both U and X and re-estimate $P(Y \mid D)$

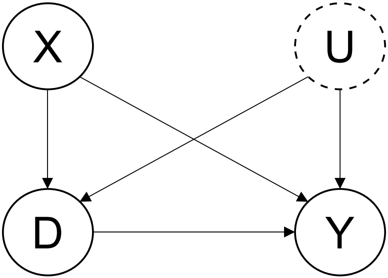
Examine the change in $P(Y \mid D)$ with vs. without U

A small change in $P(Y \mid D)$ lends credibility to the conditional unconfoundedness.

Sensitivity analysis: A toy model

A simple model with treatment D, outcome Y, observable X and a single unobservable $U \perp X$.

$$
\begin{aligned}
D &:= a_x X + a_u U \\
Y &:= b_x X + b_u U + \underbrace{\tau}_{\text{Objective}} D
\end{aligned}
$$



Sensitivity analysis: A toy model

$$\begin{cases} D & := \alpha_x X + \alpha_u U \\ Y & := \beta_x X + \beta_u U + \tau D \end{cases}$$

If we observe both X and U, we can recover τ:

$$E\left(Y_1 - Y_0\right) = E_{X,U}\left[E\left(Y \mid D = 1, X, U\right) - E\left(Y \mid D = 0, X, U\right)\right] = \tau$$

Yet, we do not observe U. If condition on X only, we get:

$$E_X\left[E\left(Y \mid D = 1, X\right) - E\left(Y \mid D = 0, X\right)\right] = \tau + \underbrace{\frac{\beta_u}{\alpha_u}}_{\text{Bias}}$$

Sensitivity analysis: a toy model

Technical proof:

$$
\begin{aligned}
E_X\left[E\left(Y \mid D=d, X\right)\right] &= E_X\left[E\left(\beta_x X + \beta_u U + \tau D \mid D=d, X\right)\right] \\
&= E_X\left[\beta_x X + \beta_u E\left(U \mid D=d, X\right) + \tau d\right] \\
&= E_X\left[\beta_x X + \beta_u \underbrace{\left(\frac{d-\alpha_x X}{\alpha_u}\right)}_{\text{From the treatment equation}} + \tau d\right] \\
&= \left(\tau + \frac{\beta_u}{\alpha_u}\right) d + \left(\beta_x - \frac{\beta_u \alpha_x}{\alpha_u}\right) E\left(X\right)
\end{aligned}
$$

Therefore, we have the following,

$$
E_X\left[E\left(Y \mid D=1, X\right) - E\left(Y \mid D=0, X\right)\right] = \left(\tau + \frac{\beta_u}{\alpha_u}\right)(1-0) = \tau + \frac{\beta_u}{\alpha_u}
$$

Sensitivity analysis: a toy model

With a closed-form solution for the bias, how do we use it?

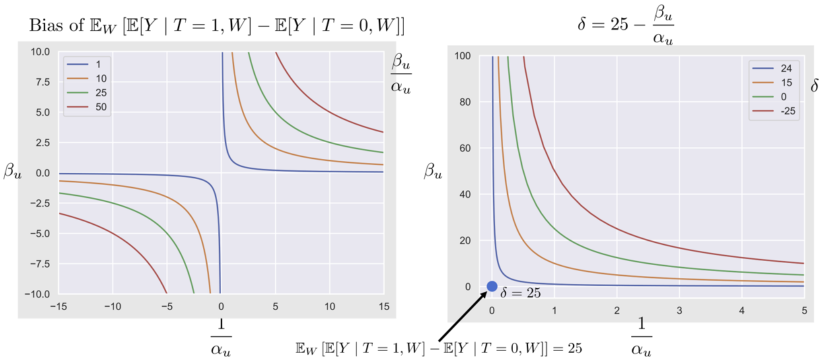
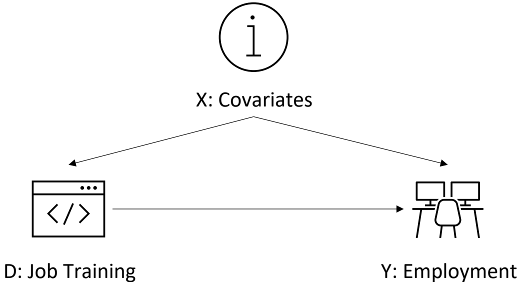


Figure: Contour Plots of Bias (left) and "True" Treatment Effect (right)

Sensitivity analysis: Rosenbaum's approach

Rosenbaum developed a sensitivity analysis procedure for matching estimators (Rosenbaum 1987).

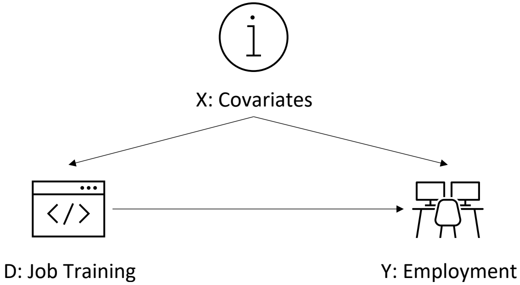
- Suppose we want to know how **a job training program** (D) can help the unemployed **to find a job** (Y).
- The participation of the program is self-selected, but a set of covariates X are observed.
- Define the propensity score $\pi_i = P(D_i = 1 \mid X_i)$.



Sensitivity analysis: Rosenbaum's approach

Rosenbaum developed a sensitivity analysis procedure for matching estimators (Rosenbaum 1987).

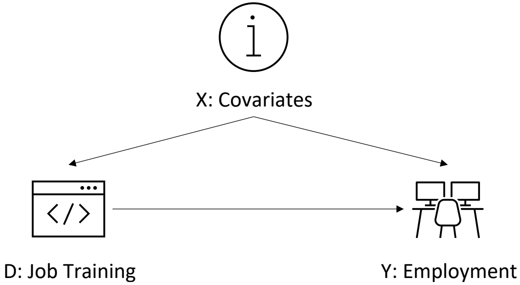
- Suppose we want to know how **a job training program** (D) can help the unemployed **to find a job** (Y).
- The participation of the program is self-selected, but a set of covariates X are observed.
- Define the propensity score $\pi_i = P(D_i = 1 \mid X_i)$.



Sensitivity analysis: Rosenbaum's approach

Rosenbaum developed a sensitivity analysis procedure for matching estimators (Rosenbaum 1987).

- Suppose we want to know how **a job training program** (D) can help the unemployed **to find a job** (Y).
- The participation of the program is self-selected, but a set of covariates X are observed.
- Define the propensity score $\pi_i = P(D_i = 1 \mid X_i)$.

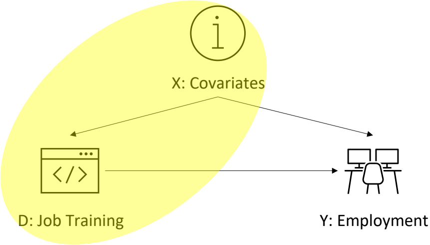


Sensitivity analysis: Rosenbaum's approach

Observation

For **a matched pair** with the same covariates X_i (e.g., age, gender...), the propensity score is equal, if there is no confounders.

$$\begin{cases} \text{Odds}_{\text{Bob}} & = \dfrac{\pi_{\text{Bob}}}{1-\pi_{\text{Bob}}} \\ \text{Odds}_{\text{Jim}} & = \dfrac{\pi_{\text{Jim}}}{1-\pi_{\text{Jim}}} \end{cases} \text{ and } \Gamma = \dfrac{\dfrac{\pi_{\text{Bob}}}{1-\pi_{\text{Bob}}}}{\dfrac{\pi_{\text{Jim}}}{1-\pi_{\text{Jim}}}} = 1$$



Sensitivity analysis: Rosenbaum's approach

Existence of confounders

If no confounders exist, $\Gamma = 1$. The existence of confounders $\rightarrow$ $\Gamma \neq 1$.

$$\Gamma = \frac{\frac{\pi_{\text{Bob}}}{1-\pi_{\text{Bob}}}}{\frac{\pi_{\text{Jim}}}{1-\pi_{\text{Jim}}}} \neq 1$$

- Γ represent the severity of the confoundedness.
- Increasing the value of Γ to see the changes in the treatment effects.

Sensitivity analysis: Rosenbaum's approach

		Job Training	
		$D = 0$	$D = 1$
Employment	$Y = 0$	10	c $= 50$
	$Y = 1$	b $= 20$	140

- Suppose use McNemar's exact p-value to test the difference in the employment rate with vs. without the job training ($n = b + c$).

$$\text{p-value} = 2 \sum_{k=b}^{n} \binom{n}{k} \left(\frac{\Gamma}{1+\Gamma} \right)^{k} \left(\frac{1}{1+\Gamma} \right)^{n-k}$$

- Under unconfoundedness, $\Gamma = 1$. With a larger Γ, more serious confounding.

Sensitivity analysis: Rosenbaum's approach

p-value	prob $\left(\frac{\Gamma}{1+\Gamma}\right)$	Γ
0.004	0.5000	1.00
0.0018	0.5238	1.10
0.0057	0.5455	1.20
0.0148	0.5652	1.30
0.0328	0.5833	1.40
0.0635	0.6000	1.50

- When $\Gamma = 1.5$, the p-value is larger than 0.05.
- To nullify the effect: a confounder that makes Bob 1.5 times more likely to join the program than Jim.
- Is $1.5\times$ a good indicator of validity? You be the judge!

Sensitivity analysis: Rosenbaum's approach

Technical proof: The probability in McNemar's exact p-value is equal to $\Gamma/(1+\Gamma)$.

The probability in the McNemar's test is θ_i, indicating $P\left(d_i = 1 \mid d_i + d_j = 1\right)$, or the probability of a subject i is treated, conditional on one of the matched pairs $\langle i, j\rangle$ is treated.

$$
\begin{aligned}
\theta_i &= \frac{P\left(d_i = 1, d_j = 0\right)}{P\left(d_i = 1, d_j = 0\right) + P\left(d_i = 0, d_j = 1\right)} \\
&= \frac{\pi_i\left(1 - \pi_j\right)}{\pi_i\left(1 - \pi_j\right) + \pi_j\left(1 - \pi_i\right)}
\end{aligned}
$$

Then we must have

$$
\frac{\theta_i}{1 - \theta_i} = \frac{\pi_i\left(1 - \pi_j\right)}{\pi_j\left(1 - \pi_i\right)} \overset{\text{Definition}}{=} \Gamma
$$

Therefore, $\theta_i = \Gamma/(1+\Gamma)$.

Sensitivity analysis: selection bias

- Rosenbaum's approach focuses on primarily on matching.
- In practice, we often use a linear regression:
  $Y = \beta X + \tau D + \varepsilon$.
- Oster 2019 developed a more general method with insights from Altonji et al. 2005.

Variables	**Model 1**	**Model 2**	**...**	**Full Model**
Treatment D	1.00	1.20	...	1.50
X_1	Excl.	Incl.	...	Incl.
...	...	...	...	...
X_K	...	...	...	Incl.

Table: Stability of Treatment Effects across Models

Sensitivity analysis: selection bias

Is the stability of the treatment effect sufficient?

Variables	**Null Model**	**Full Model**
Treatment D	1.49	1.50
X_1	Excl.	Incl.
...	...	...
X_K	Excl.	Incl.

Two alternative explanations for stability

1. The bias from confounders are minimal.
2. Control variables are not "controlling anything."

Sensitivity analysis: selection bias

To examine the “efficacy” of the control variables, we need to check the model fit.

Variables	Null Model	Full Model
D	1.49	1.50
X_1	Excl.	Incl.
...	...	...
X_K	Excl.	Incl.
R^2	10.00%	10.01%

Bad Controls

Variables	Null Model	Full Model
D	1.49	1.50
X_1	Excl.	Incl.
...	...	...
X_K	Excl.	Incl.
R^2	10.00%	90.01%

Good Controls

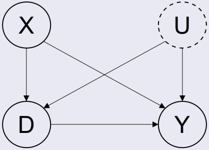
RSM Erasmus

Sensitivity analysis: selection bias

- Observed variables could inform the selection on unobservables.
- We need to consider the **changes in model fit** (e.g., R^2), along with the **changes in coefficients**.

The setup of Oster's sensitivity analysis

$$Y = \alpha + \beta D + \varphi X + \omega U + \varepsilon$$



Both X and U can include multiple variables.

Sensitivity analysis: selection bias

Assumptions of Osters' sensitivity analysis

The coefficients of X of a regression $D \sim X$ is proportional to the coefficients of X of a regression $Y \sim D + X$.

Under this assumption, the bias-adjusted treatment effect is:

$$\beta^* = \beta_1 - \delta\left(\beta_0 - \beta_1\right)\frac{R^2_{max} - R^2_1}{R^2_1 - R^2_0}$$

- δ is the sensitivity parameter, indicating the relative importance of unobservables over the observables.
- β_0 and R^2_0 are the treatment effect and R-squared of the uncontrolled model $Y \sim D$.
- β_1 and R^2_1 are the treatment effect and R-squared of the controlled model $Y \sim D + X$.

Sensitivity analysis: selection bias

To understand the result, we do a little bit algebra and assume $\delta = 1$.

$$\beta^* = \beta_1 - (\beta_0 - \beta_1)\,\frac{R^2_{max} - R^2_1}{R^2_1 - R^2_0}$$

$$\Longrightarrow \frac{\beta_1 - \beta^*}{\beta_0 - \beta_1} = \frac{R^2_{max} - R^2_1}{R^2_1 - R^2_0}$$

- Under the equal selection assumption, the ratio of the movement in coefficients is equal to the ratio of the movement in R-squared.
- The changes in model fit are informative for the changes in coefficients!

Sensitivity analysis: selection bias

To apply the result, we can follow these steps:

1. Run a regression of $Y \sim D$ and record β_0 and R_0^2;
2. Run a regression of $Y \sim D + X$ and record β_1 and R_1^2;
3. Examine the value of δ that nullify the effects or $\beta^* = 0$.
4. Discuss whether the value of δ is reasonable.

Sensitivity analysis: selection bias

To apply the result, we can follow these steps:

1. Run a regression of $Y \sim D$ and record β_0 and R_0^2;
2. Run a regression of $Y \sim D + X$ and record β_1 and R_1^2;
3. Examine the value of δ that nullify the effects or $\beta^* = 0$.
4. Discuss whether the value of δ is reasonable.

Sensitivity analysis: selection bias

To apply the result, we can follow these steps:

1. Run a regression of $Y \sim D$ and record β_0 and R_0^2;
2. Run a regression of $Y \sim D + X$ and record β_1 and R_1^2;
3. Examine the value of δ that nullify the effects or $\beta^* = 0$.
4. Discuss whether the value of δ is reasonable.

Sensitivity analysis: selection bias

To apply the result, we can follow these steps:

1. Run a regression of $Y \sim D$ and record β_0 and R_0^2;
2. Run a regression of $Y \sim D + X$ and record β_1 and R_1^2;
3. Examine the value of δ that nullify the effects or $\beta^* = 0$.
4. Discuss whether the value of δ is reasonable.

Summary

- Assumptions are the central piece of causal inference.
- Be transparent about your assumptions.
- Examine the plausibility of your assumptions with consistency tests and sensitivity analysis.
- Relax implausible assumptions even if it nullifies your positive findings.

References I

- Dube, A., Dube, O., & García-Ponce, O. (2013). Cross-border spillover: US gun laws and violence in Mexico. American Political Science Review, 107(3), 397-417.
- Rosenbaum, P. R. (1987). Model-based direct adjustment. Journal of the American statistical Association, 82(398), 387-394.
- Chen, X., Rajagopalan, N., & Yang, S. (2023). The effects of environmental uncertainty on alliance strategies of firms: Evidence from the Great Recession. Working paper.
- Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. Journal of Business & Economic Statistics, 37(2), 187-204.

References II

Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. Journal of political economy, 113(1), 151-184.