6(b). Sensitivity Analysis

ISS5096 || ECI

Jaewon ("Jay-one") Yoo
National Tsing Hua University

Where are we? Where are we going?

- Saw how to estimate the ATE with regression under selection on observables.

Where are we? Where are we going?

- Saw how to estimate the ATE with regression under selection on observables.
- What if this assumption doesn't hold?
    - **Sensitivity analysis**: try to vary the amount of unmeasured confounding to see if it changes the effect.
    - **Partial identification**: abandon point identification and try to find bounds for the ATE under different assumptions.

1/ Sensitivity Analysis

Sensitivity Analysis for Regression

- Standard regression estimator of the ATE:

$$Y_i = \widehat{\alpha} + \widehat{\tau} D_i + \mathbf{X}'_i \widehat{\beta} + \varepsilon_i$$

  - Recall that regression with covariates both in uninteracted and interacted models are consistent for SATE/PATE (W4 discussion on Freedman 2008 critique and Lin 2023 remedy).

Sensitivity Analysis for Regression

- Standard regression estimator of the ATE:

$$Y_i = \widehat{\alpha} + \widehat{\tau} D_i + \mathbf{X}'_i \widehat{\beta} + \varepsilon_i$$

  - Recall that regression with covariates both in uninteracted and interacted models are consistent for SATE/PATE (W4 discussion on Freedman 2008 critique and Lin 2023 remedy).
- What if the true regression model contained U_i which we omitted?

$$Y_i = \alpha + \tau D_i + \mathbf{X}'_i \beta + \gamma U_i + \varepsilon_i$$

Sensitivity Analysis for Regression

- Standard regression estimator of the ATE:

$$Y_i = \widehat{\alpha} + \widehat{\tau} D_i + \mathbf{X}'_i \widehat{\beta} + \varepsilon_i$$

  - Recall that regression with covariates both in uninteracted and interacted models are consistent for SATE/PATE (W4 discussion on Freedman 2008 critique and Lin 2023 remedy).

- What if the true regression model contained U_i which we omitted?

$$Y_i = \alpha + \tau D_i + \mathbf{X}'_i \beta + \gamma U_i + \varepsilon_i$$

- Standard **omitted variable bias (OVB)** formula:

$$\widehat{\tau} = \tau + \gamma \times \underbrace{\frac{\text{cov}(D_i^{\perp\!\!\!\perp \mathbf{X}}, U_i^{\perp\!\!\!\perp \mathbf{X}})}{\mathbb{V}(D_i^{\perp\!\!\!\perp \mathbf{X}})}}_{\texttt{regression of } U_i^{\perp\!\!\!\perp \mathbf{X}} \texttt{ on } D_i^{\perp\!\!\!\perp \mathbf{X}}}$$

$$= \tau + \underbrace{\gamma \times \delta}_{\text{OVB}}$$

  - Gives us $\rightsquigarrow \tau = \widehat{\tau} - \gamma\delta$

Partial R-Squared Interpretations

- Standard OVB written in terms of regression coefficients is difficult to reason about.
- Easier to reason with partial R^2 version of OVB (Cinelli and Hazlett, JRSSB 2020):

$$|\text{bias}| = \sqrt{\frac{R^2_{Y\sim U|D,\mathbf{X}} \cdot R^2_{D\sim U|\mathbf{X}}}{1 - R^2_{D\sim U|\mathbf{X}}} \cdot \frac{\mathbb{V}(Y^{\perp\mathbf{X},D})}{\mathbb{V}(D^{\perp\mathbf{X}})}}$$

Partial R-Squared Interpretations

- Standard OVB written in terms of regression coefficients is difficult to reason about.
- Easier to reason with partial R^2 version of OVB (Cinelli and Hazlett, JRSSB 2020):

$$|\text{bias}| = \sqrt{\frac{R^2_{Y\sim U|D,\mathbf{X}} \cdot R^2_{D\sim U|\mathbf{X}}}{1 - R^2_{D\sim U|\mathbf{X}}} \cdot \frac{\mathbb{V}(Y^{\perp\!\!\!\perp \mathbf{X},D})}{\mathbb{V}(D^{\perp\!\!\!\perp \mathbf{X}})}}$$

- Partial R^2 is the incremental predictive value of one variable:

$$R^2_{Y\sim U|D,\mathbf{X}} = \frac{R^2_{Y\sim D+\mathbf{X}+U} - R^2_{Y\sim D+\mathbf{X}}}{1 - R^2_{Y\sim D+\mathbf{X}}} = \frac{\text{additional variance explained by } U}{\text{variance unexplained by } D, \mathbf{X}}$$

Partial R-Squared Interpretations

- Standard OVB written in terms of regression coefficients is difficult to reason about.
- Easier to reason with partial R^2 version of OVB (Cinelli and Hazlett, JRSSB 2020):

$$|\text{bias}| = \sqrt{\frac{R^2_{Y\sim U|D,\mathbf{X}} \cdot R^2_{D\sim U|\mathbf{X}}}{1 - R^2_{D\sim U|\mathbf{X}}} \cdot \frac{\mathbb{V}(Y^{\perp\mathbf{X},D})}{\mathbb{V}(D^{\perp\mathbf{X}})}}$$

- Partial R^2 is the incremental predictive value of one variable:

$$R^2_{Y\sim U|D,\mathbf{X}} = \frac{R^2_{Y\sim D+\mathbf{X}+U} - R^2_{Y\sim D+\mathbf{X}}}{1 - R^2_{Y\sim D+\mathbf{X}}} = \frac{\text{additional variance explained by } U}{\text{variance unexplained by } D, \mathbf{X}}$$

- **Sensitivity analysis** can then vary two unknown parameters:
  - $R^2_{Y\sim U|D,\mathbf{X}} \in [0,1]$: incremental predictive value of U for the outcome.
  - $R^2_{D\sim U|\mathbf{X}} \in [0,1]$: incremental predictive value of U for treatment.
  - From these we can determine the bias and thus the true value of τ.

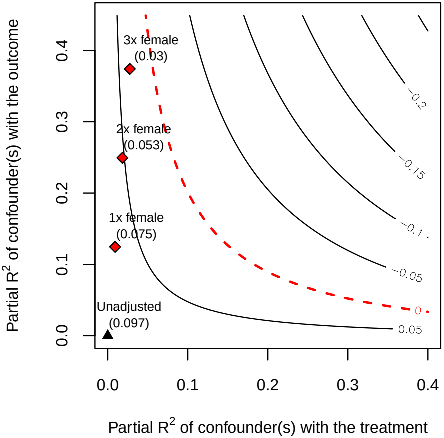
R Package: sensemakr

```r
# load package
library(sensemakr)

# import dataset
data("darfur")

# run the regression model
model <- lm(peacefactor ~ directlyharmed + age + farmer_dar + herder_dar +
              pastvoted + hhsize_darfur + female + village, data = darfur)

# conduct the sensitivity analysis
sensitivity <- sensemakr(model = model,
                         treatment = "directlyharmed",
                         benchmark_covariates = "female",
                         kd = 1:3)
```

- `female` in our data is the strongest known confounder we are using as benchmark.
    - kd (knock-on distance): how many times stronger the confounder is related to the treatment in comparison to the observed benchmark covariate.

Sensitivity Analysis with Partial R^2

```r
# plot bias contour of point estimate
plot(sensitivity)
```



Summing Up

- Conditional unconfoundedness is an assumption on unmeasured data, and thus inherently untestable!
  - The data are uninformative about the distribution of $Y(0)$ for treated units and $Y(1)$ for control units.
  - The assumption, however, can be ‘indirectly’ assessed. $\rightsquigarrow$ sensitivity analysis
- Up next: instrumental variables and the noncompliance issue

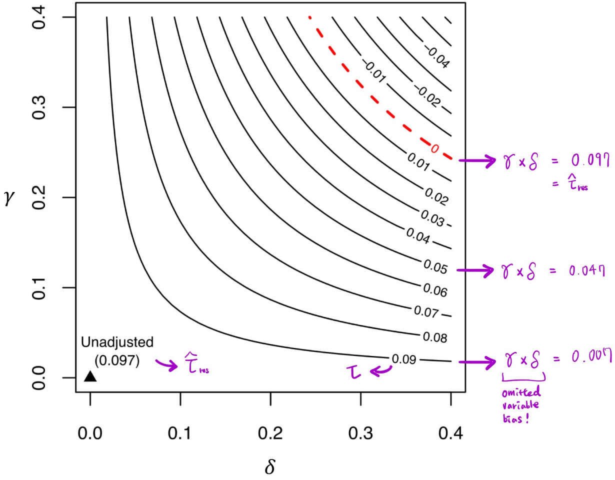
Have a great weekend! :)

Contact Information:
jaewon.yoo@iss.nthu.edu.tw
https://j1yoo4.github.io/

Appendix

Sensitivity Contour Plots

- Two-dimensional plot of bias contours parameterized by γ and δ:



Source: Cinelli, C., & Hazlett, C. (2020). Making sense of sensitivity: Extending omitted variable bias. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 82(1), 39-67.