

SENSITIVITY ANALYSIS

Annie Chen

February 5, 2020

UNMEASURED CONFOUNDING

- With matching, we aim to achieve balance on *observed* covariates.
- But... there's no guarantee that there is balance on *unobserved* variables that we did not match on.

SENSITIVITY ANALYSIS

- How sensitive are estimates of an average causal effect to the potential effects of unobservable treatment selection patterns?
- An unobserved covariate, C , will induce a material degree of bias only if it is sufficiently associated with both treatment assignment, D , and the outcome, Y .¹
- What does the DAG look like?

¹OVB iff $C_i \not\perp\!\!\!\perp D_i$ and $C_i \not\perp\!\!\!\perp Y_i$

REVIEW OF OMITTED VARIABLE BIAS

- Suppose the true model can be represented by the “long” regression formula:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 C_i + \epsilon_i$$

where C_i denotes an unobserved (confounding) variable, and D_i is the treatment.

- In the case of OVB,

$$Y_i = \beta_0^* + \beta_1^* D_i + \epsilon_i^*$$

if

$$C_i = \gamma_0 + \gamma_1 D_i + \nu_i$$

- Then,

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 (\gamma_0 + \gamma_1 D_i + \nu_i) + \epsilon_i$$

$$Y_i = \beta_0 + \beta_2 \gamma_0 + (\beta_1 + \beta_2 \gamma_1) D_i + \beta_2 \nu_i + \epsilon_i$$

- and

$$\beta_1^* = \beta_1 + \beta_2 \gamma_1$$

IMBENS-STYLE SENSITIVITY ANALYSIS

- Where U (unobserved confounder) and D (treatment) are binary (0 or 1), the bias of the treatment effect (τ) is:

$$\begin{aligned}\mathbb{E}[\hat{\tau}] - \tau &= \mathbb{P}(U_i = 1|D_i = 1) - \mathbb{P}(U_i = 1|D_i = 0) \\ &\quad \times \mathbb{E}[Y_i|U_i = 1] - \mathbb{E}[Y_i|U_i = 0]\end{aligned}$$

- Where $\delta \equiv \mathbb{P}(U_i = 1|D_i = 1) - \mathbb{P}(U_i = 1|D_i = 0)$ is the difference in average U_i between treatment conditions
 $\gamma \equiv \mathbb{E}[Y_i|U_i = 1] - \mathbb{E}[Y_i|U_i = 0]$ represents the effect of U_i on Y_i .
- The bias is $\delta\gamma$ (similar to the OVB formula we just saw in the regression context: $\mathbb{E}[\hat{\tau}] = \tau + \delta\gamma$).
- What assumption did we make for this to be true?

RETURN TO LALONDE EXAMPLE

```
data(lalonde, package = "Matching")
```

```
model1 <- lm(re78 ~ treat + age + educ + married + black + re75, data = lalonde)
model1_HC <- list(coeftest(model1,
                           vcov = vcovHC(model1, type = "HC2"))[, 2])
```

TABLE 1: OLS Results

	re78
treat	1,651.331 (646.178)
age	53.498 (39.719)
educ	409.823 (157.929)
married	-172.412 (851.525)
black	-2,197.024 (735.325)
re75	0.146 (0.103)
Constant	738.590 (2,058.213)

- What is half the magnitude of the treatment coefficient?

```
library(lm.beta)
# using standardized coeffs
beta_half <- lm.beta(model1)$standardized.coef[2]/2
c(lm.beta(model1)$standardized.coef[2], beta_half)
```

```
##      treat      treat
## 0.1228638 0.0614319
```

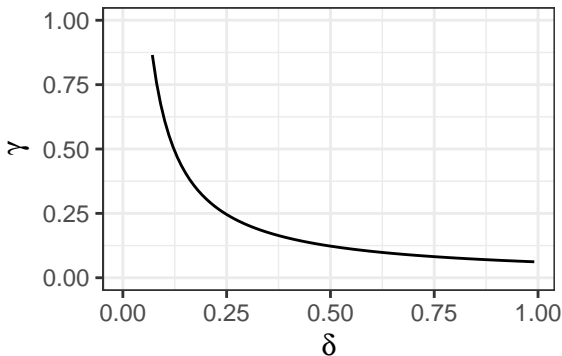
$$\delta = \mathbb{P}(U_i = 1 | D_i = 1) - \mathbb{P}(U_i = 1 | D_i = 0)$$

```
# The range of delta is 0,1
delta <- seq(0.001, 1, by = 0.01)
```

- Exercise: Using δ and beta_half , solve for γ (remember: $\mathbb{E}[\hat{\tau}] = \tau + \delta\gamma$). Then plot δ (x-axis) and γ (y-axis).

```
# Solve for gamma = E[Y_i | U_i = 1] - E[Y_i | U_i = 0]
g <- beta_half/delta
```

```
ggplot(data.frame(delta = delta, g = g), aes(x = delta, y = g)) +  
  geom_path() +  
  xlab(expression(delta)) + ylab(expression(gamma)) +  
  ylim(0, 1) + theme_bw()
```



- How are the covariates related to the treatment?

```
# Estimate the relationship between covariates (U) and treatment (D)  
mod_delta <- lm(treat ~ age + educ + married + black + re75,  
               data = lalonde)
```

TABLE 2: OLS delta

	treat
age	0.003 (0.003)
educ	0.018 (0.014)
married	0.030 (0.068)
black	0.021 (0.063)
re75	0.00001 (0.00001)
Constant	0.123 (0.164)
N	445
R ²	0.010

- How do they relate to the outcome?

```
# Estimate relationship between covariates (U) and outcome (Y)  
mod_gamma <- lm(re78 ~ age + educ + married + black + re75,  
               data = lalonde)
```

TABLE 3: OLS gamma

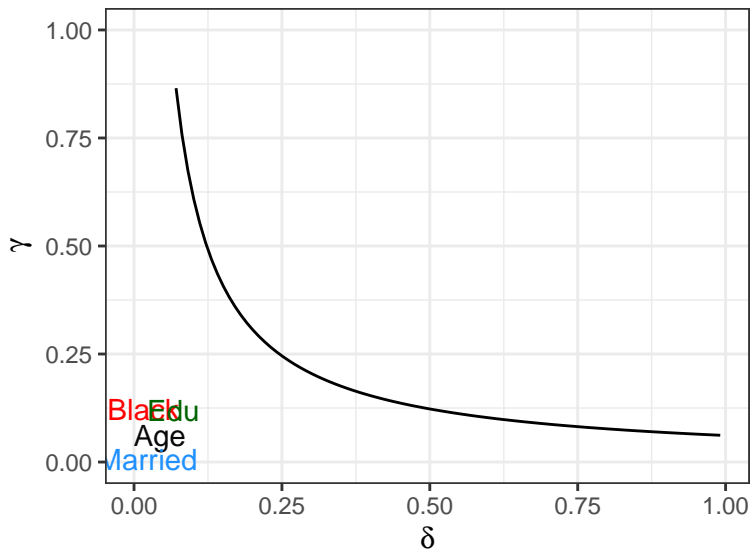
	re78
age	58.494 (39.976)
educ	440.036 (165.095)
married	-122.549 (864.001)
black	-2,161.909 (734.904)
re75	0.154 (0.103)
Constant	941.146 (2,058.587)
N	445
R ²	0.037

- To compare the hypothetical degree of confounding based on the sensitivity parameters with the actual degree of confounding created by some observed covariates, we plot them against the curve.
- We use the standardized coefficients here, but you can also calculate the raw δ and γ values, or compute the partial R-squared.²

```
library(lm.beta)
# Extract standardized coefficients for age
delta_age <- abs(lm.beta(mod_delta)$standardized.coefficients[2])
gamma_age <- abs(lm.beta(mod_gamma)$standardized.coefficients[2])
```

²Both standardized coeffs and partial r measure the unique contribution of a covariate in a model.

- Do this for other covariates in the model...



RETURN TO RANDOMIZATION INFERENCE (ROSENBAUM APPROACH)

- Components of RI:
 - Null hypothesis (i.e. no treatment effect)
 - Test-statistic (i.e. Wilcoxon Signed Rank Test)
 - Number of permutations: nCr
 - p-value is ratio of number of times we observe the test-statistic or greater to number of permutations
- Overview of Rosenbaumian SA:
 - 1 Set Γ (the sensitivity parameter).³
 - 2 Compute probability of treatment ($\pi(X_i)$) for Γ bounds (see lecture slides).
 - 3 Conduct RI under the null of $\min\{\pi(X_i)\} = \max\{\pi(X_i)\} = 0.5$.
 - 4 Repeat for different values of Γ .

³ $\Gamma = 1$ is the case where there is no bias from confounding. $\Gamma = 2$?

WILCOXON SIGNED RANK STATISTIC

- Wilcoxon Signed Rank Statistic

Match A	7	5	3	4	1
Match B	2	4	3	1	2
Abs Diff	5	1	0	3	1
Sign	+	+	NA	+	-
Rank	4.0	1.5	NA	3.0	1.5

- $W = \sum_i^N \{S_i \times R_i\}$ where N is the number of pairs where the difference $\neq 0$, S is the sign of the matched pairs $(X_i - X_j)$, and R denotes the rank.

- Why is it necessary to match first?

```
covars <- c("age", "educ", "black", "re75")
matched_dat <- Match(Y = lalonde$re78, Tr = lalonde$treat,
                     X = lalonde[,covars], Weight = 2)

summary(matched_dat)
```

```
##
## Estimate...    1913.8
## AI SE.....   821.46
## T-stat.....   2.3297
## p.val.....    0.019819
##
## Original number of observations..... 445
## Original number of treated obs..... 185
## Matched number of observations..... 185
## Matched number of observations (unweighted). 301
```

RBOUNDS PACKAGE

- `psens` gives Rosenbaum's bounds for the p-values from a Wilcoxon signed rank test.

```
library(rbounds)
rosenbaum <- psens(matched_dat, Gamma = 1.5, GammaInc = 0.1)
```

	Gamma	Lower bound	Upper bound
1	1.000	0.002	0.002
2	1.100	0.000	0.014
3	1.200	0.000	0.057
4	1.300	0.000	0.154
5	1.400	0.000	0.306
6	1.500	0.000	0.488

TABLE 4: Rosenbaum p-values

- Still need to benchmark these hypothetical values against your data by calculating the odds ratios!

ADDITIONAL NOTES

PLOT TWIST!

- Γ can be decomposed into the strength of the relationship between confounder and outcome (Δ) and strength of the relationship between the confounder and treatment (Λ) as

$$\Gamma = \frac{\Delta\Lambda + 1}{\Delta + \Lambda}$$

- Details in Rosenbaum and Silber (2009).