Intermediate Causal Inference

Sensitivity analysis

Andy Eggers

April 2019

Motivation

Scope

Broadly, *sensitivity analysis* checks how results depend on assumptions and choices made in analysis.

Our focus: how sensitive are your conclusions to reasonable amounts of omitted variable bias?

Scope (2)

"In an observational study, a sensitivity analysis replaces qualitative claims about whether unmeasured biases are present with an objective quantitative statement about the magnitude of bias that would need to be present to change the conclusions. In this sense, a sensitivity analysis speaks to the assertion 'it might be bias' in much the same way that a P-value speaks to the assertion 'it might be bad luck'."

— Paul Rosenbaum, "Two R Packages for Sensitivity Analysis in Observational Studies", (2015)

Why do we need sensitivity analysis?

We implicitly do sensitivity analysis all the time.

- *Effect size is really large:* unlikely that such a large effect would be caused by an omitted variable
- *p-value is really small:* unlikely that I would get such a small p-value if there wasn't really an effect

We want to be more precise.

Classic approaches

The simplest example

"If cigarette smokers have 9 times the risk of nonsmokers for developing lung cancer, and this is not because cigarette smoke is a causal agent, but only because cigarette smokers produce hormone X, then the proportion of hormone-X-producers among cigarette smokers must be at least 9 times greater than that of nonsmokers."

— Cornfield et al (1959), "Smoking and lung cancer: recent evidence and a discussion of some questions"

Key point: quantifying the type of bias that would be necessary to explain the result.

The simplest example: formally

Notation:

- y_i: outcome for unit i (e.g. lung cancer)
- $D_i \in \{0, 1\}$: i's exposure to treatment (e.g. smoking)
- $X_i \in \{0, 1\}$: i's exposure to some other factor (e.g. hormone X)
- $p_1 \equiv \Pr(X_i = 1 \mid D_i = 1)$: rate of $X_i = 1$ in treated group
- $p_0 \equiv \Pr(X_i = 1 \mid D_i = 0)$: rate of $X_i = 1$ in control group

The simplest example: formally (2)

Assume no effect of treatment and (w/o loss of generality) $p_1 > p_0$.

Avg outcome by treatment status:

$$E[y_i \mid D_i = 1] = p_1 E[y_i \mid X_i = 1] + (1 - p_1)E[y_i \mid X_i = 0]$$

$$E[y_i \mid D_i = 0] = p_0 E[y_i \mid X_i = 1] + (1 - p_0)E[y_i \mid X_i = 0]$$

Rearranging,

$$p_1 = \frac{E[y_i \mid D_i = 1] - E[y_i \mid X_i = 0]}{E[y_i \mid X_i = 1] - E[y_i \mid X_i = 0]}$$

$$p_0 = \frac{E[y_i \mid D_i = 0] - E[y_i \mid X_i = 0]}{E[y_i \mid X_i = 1] - E[y_i \mid X_i = 0]}$$

The simplest example: formally (3)

It follows that

$$\frac{p_1}{p_0} = \frac{E[y_i \mid D_i = 1] - E[y_i \mid X_i = 0]}{E[y_i \mid D_i = 0] - E[y_i \mid X_i = 0]} \geq \frac{E[y_i \mid D_i = 1]}{E[y_i \mid D_i = 0]}$$

(where last inequality depends on $E[y_i \mid X_i = 1] > E[y_i \mid X_i = 0]$).

So if rate of death 9 times higher among smokers and smoking does not cause death, then a confounder that does cause death must be at least 9 times more prevalent among smokers.

Limitations of simplest example

- Binary treatment and binary confounder
- Ratio of average outcomes (not difference)
- No statistical inference (e.g. p-values)

How can we be more general?

Review: interpretation of p-value

Consider randomized experiment in which treatment is randomly assigned within **matched pairs** (block randomization).

Observing outcomes, we measure a test statistic:

- difference in means between treatment and control group
- Wilcoxon's signed rank statistic
- etc

We compute a p-value under the *sharp null hypothesis*: Suppose the treatment has no effect. Under what proportion of randomizations would we observe a test statistic as large or larger than the one actually observed?

An example (from Rosenbaum 2002)

Hammond (1964) matched 35,975 heavy smokers with non-smokers on the basis of 18 covariates (age, race, alcohol consumption, occupational exposure to dust, …)

In 122 pairs, one person died of lung cancer; in 110 of these it was the smoker who died.

Initial question: What is the probability that it would be the smoker who died in at least 110 of the pairs if

- treatment was randomly assigned, and
- smoking does not cause lung cancer?

Randomization inference

Equivalently: What is prob. of at least 110 heads in 122 flips?

```
# Manual approximation
M <- 1000000  # number of simulations
nheads_fair_coin <- rep(NA, M)   # for storage
for(i in 1:M){
  samp <- sample(x = c(0,1), size = 122, replace = T, prob = c(.5, .5))
  nheads_fair_coin[i] <- sum(samp)
}
mean(nheads_fair_coin >= 110)
```

```
## [1] 0
```

```
# Exact answer:
pbinom(110, 122, 1/2, lower.tail = F)
```

```
## [1] 2.923364e-22
```

Rosenbaum's bounds for p-values (1)

Hammond (1964) matched on 18 covariates but might have missed something.

Fisher (1958) suggested there might be a gene that makes people like smoking and gives them lung cancer.

Even if smoking does not affect lung cancer, observed results not surprising if some confounder very strongly predicts smoking w/in pairs and causes death from lung cancer with certainty.

Rosenbaum's approach: posit limit on degree of confounding (i.e. how well gene predicts smoking); report p-value for worst-case *(ex ante)* confounding given this limit.

Rosenbaum's bounds for p-values (2)

Suppose each subject i's probability of being a smoker is in interval $\left[\frac{1}{2} - \delta, \frac{1}{2} + \delta\right]$ (so in randomized experiment $\delta = 0$).

Imposes limit on possible confounding.

Worst-case ex ante confounding given δ: in each pair, the person who was destined to die from lung cancer became a smoker with probability $\frac{1}{2} + \delta$.

New question: What is the probability that the person destined to die from lung cancer became a smoker in at least 110 of the pairs?

Randomization inference (2)

Equivalently: What is prob. of at least 110/122 heads from *biased* coin?

```
delta <- .25 # degree of confounding/bias on coin
nheads_biased_coin <- rep(NA, M)   # for storage
for(i in 1:M){
  samp <- sample(x = c(0,1), size = 122, replace = T,
                 prob = 1/2 + c(-delta, delta))
  nheads_biased_coin[i] <- sum(samp)
}
mean(nheads_biased_coin >= 110) # manual approx
```

```
## [1] 2.5e-05
```

```
pbinom(110, 122, 1/2 + delta, lower.tail = F) # exact answer
```

```
## [1] 6.326946e-06
```

R task

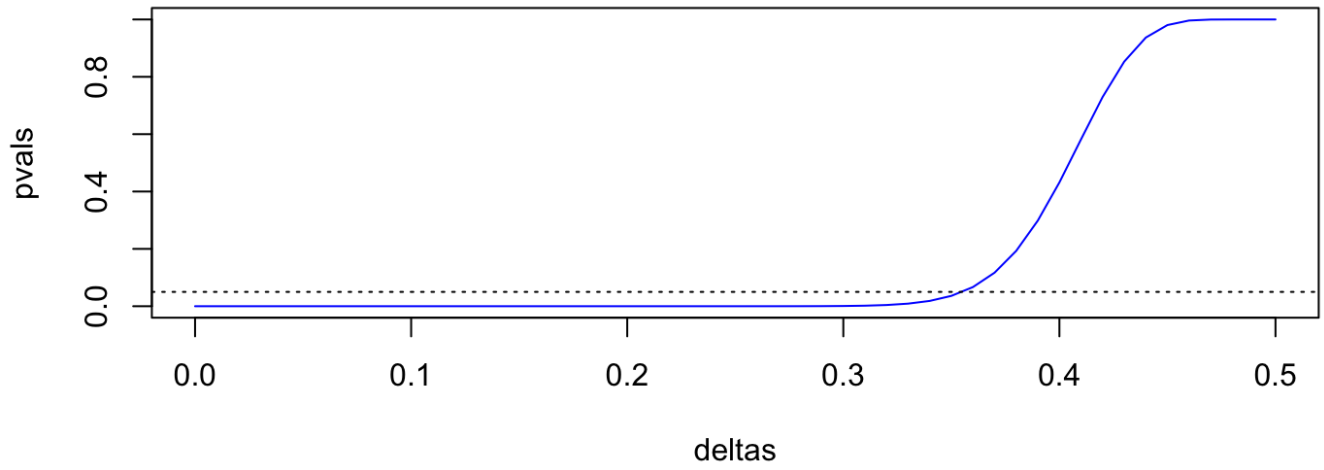
To calculate p-value given 110/122 successes and δ, we used

```
pbinom(110, 122, 1/2 + delta, lower.tail = F)
```

How large does δ have to be before the worst-case p-value rises above .05?

R task (solution)

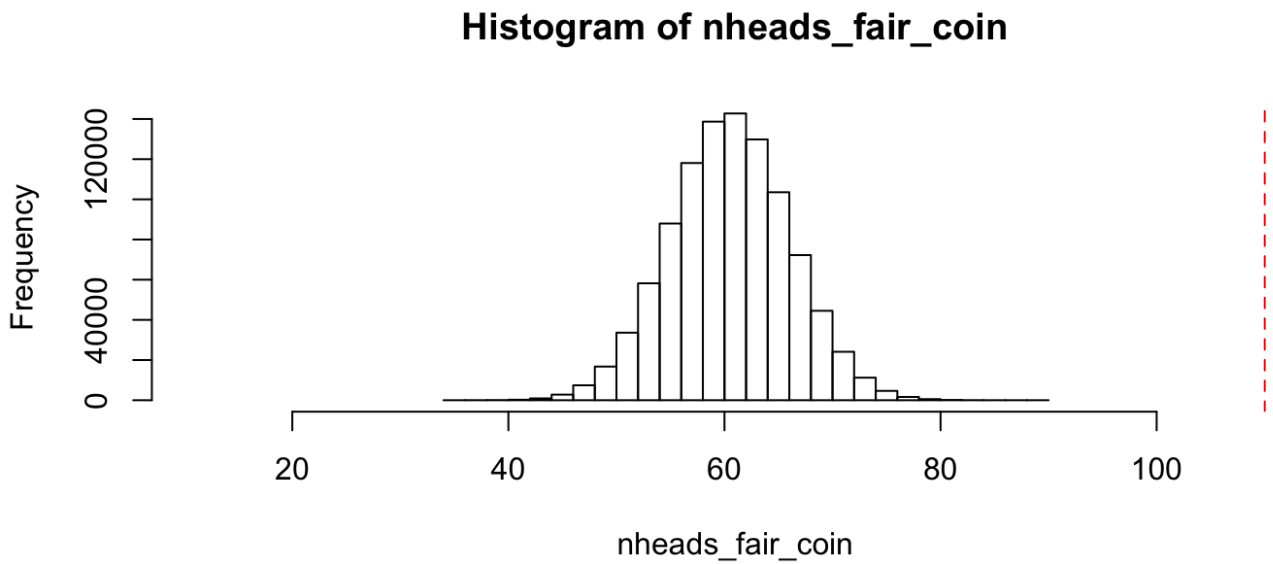
```
deltas <- seq(from = 0, to = 1/2, by = .01)
pvals <- rep(NA, length(deltas))
for(i in 1:length(deltas)){
  pvals[i] <- pbinom(110, 122, 1/2 + deltas[i], lower.tail = F)
}
plot(deltas, pvals, type = "l", col = "blue")
abline(h = .05, lty = 3)
```



Null distributions

We calculate p-values by computing sampling distribution for a test statistic under the null hypothesis (*null distribution*).

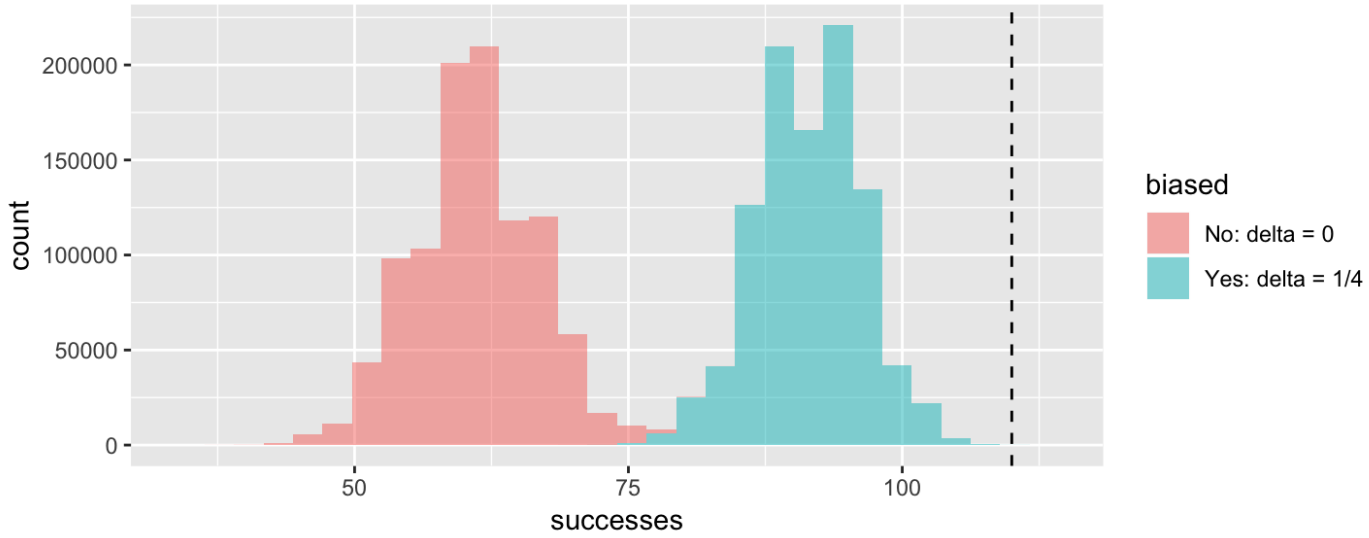
```
hist(nheads_fair_coin, xlim = c(11, 111))
abline(v = 110, col = "red", lty = 2)
```



Null distribution (2)

Assuming worst-case confounding, null distribution moves closer to the test statistic:

```
df <- data.frame(successes = c(nheads_fair_coin, nheads_biased_coin),
      biased = c(rep("No: delta = 0", M), rep("Yes: delta = 1/4", M)))
ggplot(df, aes(x = successes, fill = biased)) +
  geom_histogram(alpha = .5) + geom_vline(xintercept = 110, linetype = 2)
```



Rosenbaum bounds more generally

Rosenbaum bounds for p-values can be calculated **for matched datasets** with many test statistics (not just number of "successes" with binary outcome):

- signed rank test (continuous treatment)
- sign-score statistics
- sum statistics

Also,

- not just p-values (confidence intervals too)
- not just "prospective" studies (case-control studies too)

See R packages *sensitivitymv* and *sensitivitymw*, and Rosenbaum's book *Observational Studies* (and note use of Γ vs δ).

Regression approaches

Limitations of Rosenbaum approach

Framework is built around matched pairs/groups.

But what about when analysis is based on regression, not matching (e.g. because treatment is continuous)?

We will study approaches based on omitted-variable bias in regression.

Omitted variable bias

Suppose you estimate the effect of treatment D on outcome Y using a regression controlling for covariates $X_1, X_2, \ldots, X_k$:

$$E[Y_i] = \beta_0 + \tau D_i + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

Estimated treatment effect is $\hat{\tau}$.

How does omitting X_k from the regression affect the estimated treatment effect?

Example: Lalonde dataset

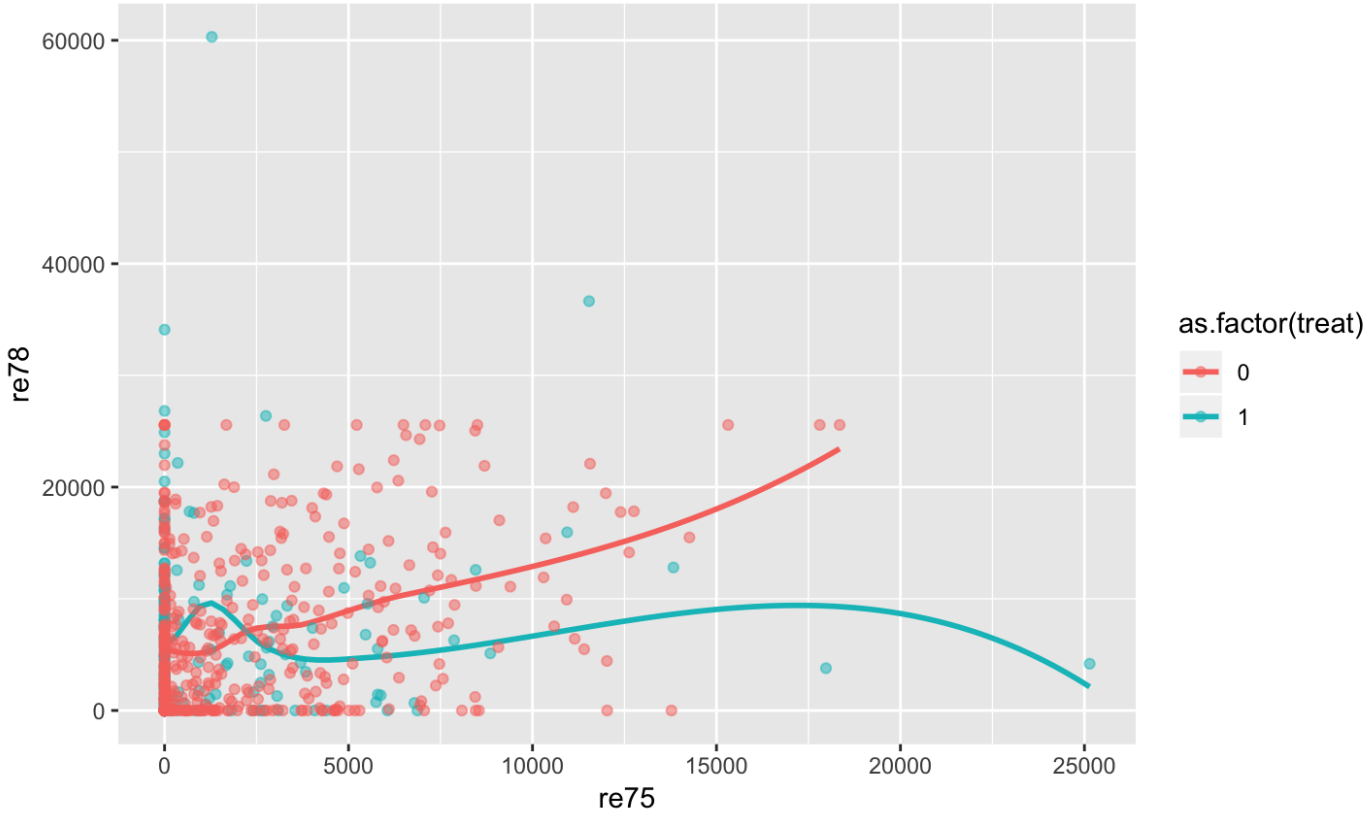
```
# load from cobalt package
library(cobalt)
data("lalonde", package = "cobalt")
# or, download from my website: http://andy.egge.rs/data/lalonde.RData

head(lalonde, 5)
```

```
##   treat age educ   race married nodegree re74 re75       re78
## 1     1  37   11  black       1        1    0    0  9930.0460
## 2     1  22    9 hispan       0        1    0    0  3595.8940
## 3     1  30   12  black       0        0    0    0 24909.4500
## 4     1  27   11  black       0        1    0    0  7506.1460
## 5     1  33    8  black       0        1    0    0   289.7899
```

- **re7x**: real earnings from 197x.
- **treat**: participated in job training program in 1976-1977

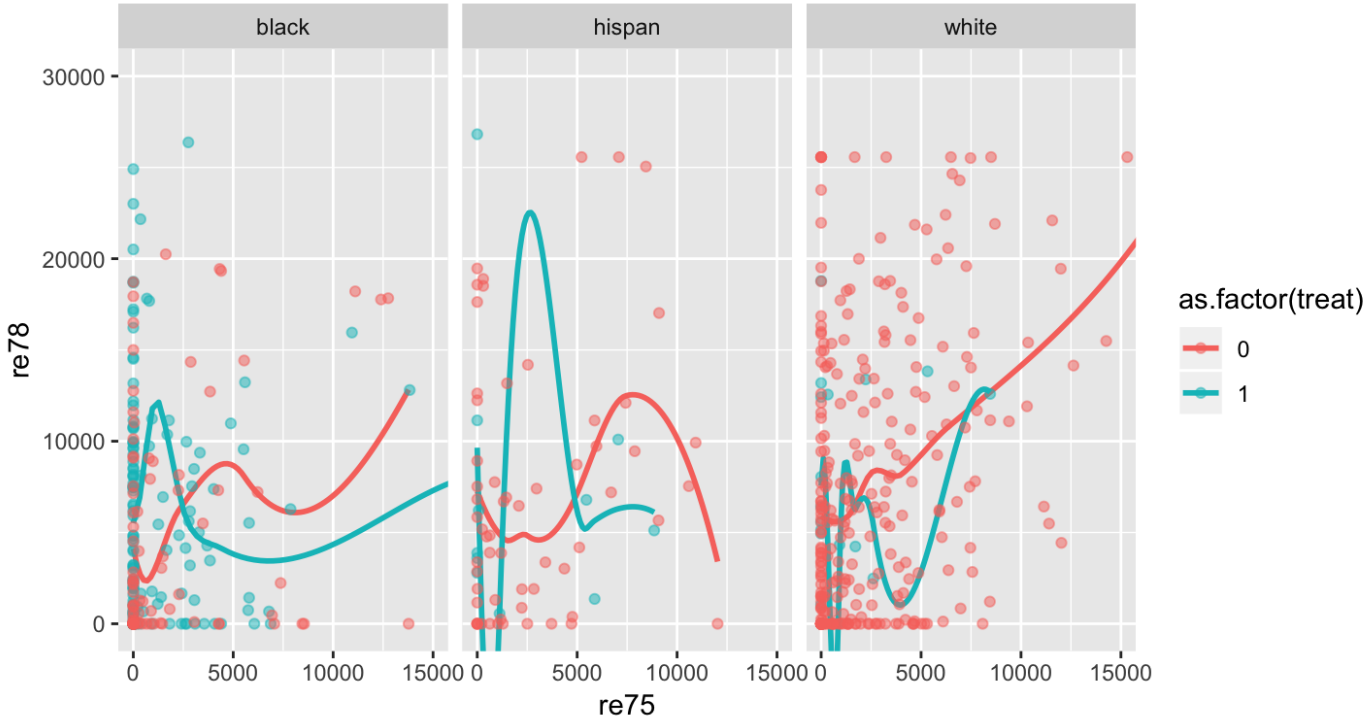
Some exploration

```
ggplot(lalonde, aes(x = re75, y = re78, color = as.factor(treat))) +
  geom_smooth(se = F) +
  geom_point(alpha = .5)
```



Some exploration (2)

```
lalonde %>% ggplot(aes(x = re75, y = re78, color = as.factor(treat))) +
  geom_smooth(se = F) + geom_point(alpha = .5) +
  coord_cartesian(xlim = c(0, 15000), y = c(0, 30000)) +
  facet_wrap( ~ race)
```



R assignment: known OVB

Using linear regression,

- Regress earnings in 1978 on treatment and all covariates, store as **reg_long**
- Regress earnings in 1978 on treatment and all covariates **except re74**, store as **reg_short**
- Compute OVB from omitting **re74**, store as **bias_direct**

R assignment: known OVB

```
# Regress earnings in 1978 on treatment and all covariates
reg_long <- lm(re78 ~ treat + age + educ + race + married +
                 nodegree + re74 + re75, data = lalonde)
## Regress earnings in 1978 on treatment and all covariates except re74
reg_short <- lm(re78 ~ treat + age + educ + race + married +
                  nodegree + re75, data = lalonde)
## Bias in our estimate of treatment effect
(bias_direct <- coef(reg_short)["treat"] - coef(reg_long)["treat"])


##    treat
## -327.028
```

R assignment: known OVB (fancier code)

```
# Regress earnings in 1978 on treatment and all covariates
reg_long <- lm(re78 ~ ., data = lalonde)
## Regress earnings in 1978 on treatment and all covariates except re74
reg_short <- update(reg_long, . ~ . - re74)
## Bias in our estimate of treatment effect
(bias_direct <- coef(reg_short)["treat"] - coef(reg_long)["treat"])
```

```
##    treat
## -327.028
```

OVB from included covariates as basis for sensitivity analysis

What does omitted variable bias (OVB) due to *observed* covariates tell us about OVB due to *unobserved* covariates?

Possibly nothing (e.g. if observed covariates are noise and the key covariate is unobserved).

Still, might be useful to know this ratio:

$$\frac{\text{estimated effect size in full model}}{\text{maximum bias due to omitting an observed covariate from full model}}$$

i.e. $\frac{\text{OVB necessary to eliminate estimated effect}}{\text{maximum observed OVB}}$

OVB ratio: Implementation

What we need to do:

- Estimate OVB from omitting each variable in full model
- Get maximum
- Take ratio

```
source("http://andy.egge.rs/code/ovb_decomp.R")
ovb_decomp <- OVB_from_included_covariates(re78 ~ treat + age + educ
                + race + married + nodegree + re74 + re75,
                data = lalonde, treatment_varname = "treat")
ovb_decomp$bias_ratio
```

```
## |treatment effect| / maximum absolute OVB from model covariates
##                                                       2.040728
```

What does this mean?

Another approach: decomposing OVB

We just did sensitivity analysis by comparing

- estimated treatment effect
- OVB from dropping observed covariates

What if we want to consider OVB from a *hypothetical* omitted variable, e.g. one that is as related to the treatment as X_1 **and** as related to the outcome as X_2?

We follow approach of Hosman, Hansen, and Holland ("HHH", 2010).

The HHH decomposition of OVB

Consider estimating a "long" regression

$$E[Y_i] = \beta_{0,l} + \tau_l D_i + \beta_{1,l} X_1 + \beta_{2,l} X_2 + \ldots + \beta_{k,l} X_k,$$

a "short" regression omitting X_k

$$E[Y_i] = \beta_{0,s} + \tau_s D_i + \beta_{1,s} X_1 + \beta_{2,s} X_2 + \ldots + \beta^s_{k-1,s} X_{k-1},$$

and a "treatment assignment" regression

$$E[D_i] = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_k X_k.$$

Hosman, Hansen & Holland (2010) show that

$$\text{OVB} \equiv \tau_s - \tau_l = \text{SE}(\tau_s) \frac{\alpha_k}{\text{SE}(\alpha_k)} \sqrt{\frac{(1 - R^2_s) - (1 - R^2_l)}{1 - R^2_s}}$$

HHH decomposition of OVB (2)

$$\text{OVB} \equiv \tau_s - \tau_l = \text{SE}(\tau_s) \times \frac{\alpha_k}{\text{SE}(\alpha_k)} \times \sqrt{\frac{(1 - R_s^2) - (1 - R_l^2)}{1 - R_s^2}}$$

Bias from omitting X_k is product of three things:

- standard error of treatment effect in short regression
- t-statistic from coefficient on X_k in treatment assignment regression
- degree to which X_k helps predict Y_i (reduction in unexplained variance from including X_k)

HHH bias decomposition: R task

- [Already done] Regress earnings in 1978 on treatment and all covariates, store as **reg_long**
- [Already done] Regress earnings in 1978 on treatment and all covariates **except re74**, store as **reg_short**
- [Already done] Compute OVB from omitting **re74**, store as **bias_direct**
- Regress treatment on all covariates, store as **treatment_mod**
- Compute OVB HHH-decomposition: product of
  - standard error on **treat** in **reg_short**
  - t-stat on **re74** in **treatment_mod**
  - square root of reduction in unexplained variance from **reg_short** to **reg_long**
- Compare to **bias_direct**

HHH bias decomposition: R solutions

See above for **reg_long**, **reg_short**, and **bias_direct**.

```
treatment_mod <- lm(treat ~ . - re78, data = lalonde)
seb <- sqrt(diag(vcov(reg_short)))["treat"]
tW <- summary(treatment_mod)$coefficients["re74", "t value"]
uv_long <- 1 - summary(reg_long)$r.squared # uv: unexplained variance
uv_short <- 1 - summary(reg_short)$r.squared
delta_uv <- (uv_short - uv_long)/uv_short
c(seb, tW, sqrt(delta_uv))
```

```
##       treat
## 794.4701222  -2.0311951   0.2026543
```

```
c(seb*tW*sqrt(delta_uv), bias_direct)
```

```
##    treat    treat
## -327.028 -327.028
```

HHH variance computations

Omitted variables also affect **standard errors**.

HHH derive SE's as function of the same 3 parameters:

$$\text{SE}(\hat{\tau}_l) = \text{SE}(\hat{\tau}_s)\sqrt{1 + \frac{1 + \frac{\alpha_k}{\text{SE}(\alpha_k)}}{\text{df} - 1}}\sqrt{1 - \frac{(1 - R_s^2) - (1 - R_l^2)}{1 - R_s^2}}$$

Thank you HHH!

Sensitivity intervals

For a given vector of parameters and a given direction of bias, you get a single confidence interval.

For a range of parameters (e.g. t-stat no larger than z) and unknown direction of bias, you get a set of intervals, the union of which HHH call the **sensitivity interval**.

Like Rosenbaum's bounds, sensitivity interval is informative about *ex ante* worst case.

HHH show how to compute edges of **sensitivity interval** from their three parameters.

Choosing parameters

Remember, HHH's **sensitivity interval** is based on:

1. standard error of treatment effect
2. t-stat for OV in treatment assignment model
3. reduction in unexplained variance (RUV) from adding OV to outcome model

Item (1) can come directly from regression model.

What should we choose for items (2) and (3)?

HHH's approach

HHH (2010) report sensitivity intervals for

- t-stats from six deliberately excluded covariates, and
- three values for RUV: .01, .1, 1

TABLE 3

95% sensitivity intervals for the treatment coefficient, with the putative unobserved variable's treatment confounding ($|t_W|$) hypothesized to be no greater than the treatment confounding of 6 deliberately omitted variables. The decrease it would bring to the variance of response residuals is hypothesized to be no greater than either of 2 index values, 1% and 10%, or is not restricted

Variable	Treatment confounding benchmark		% decrease in unexplained variation ($100\rho^2_{y\cdot w\mid zx}$)		
			1%	10%	Unrestricted
Insurance class	12.2	most	(0.03, 0.20)	(−0.04, 0.26)	(−0.21, 0.43)
Respiratory eval.	8.9	some	(0.04, 0.19)	(−0.01, 0.23)	(−0.12, 0.35)
Mean blood press.	8.6	some	(0.04, 0.19)	(−0.01, 0.23)	(−0.12, 0.34)
Cardiovascular eval.	8.5	some	(0.04, 0.19)	(−0.01, 0.23)	(−0.11, 0.34)
Weight (kg)	6.1	some	(0.04, 0.18)	(0.01, 0.21)	(−0.05, 0.28)
Immunosuppression	0.4	least	(0.06, 0.16)	(0.06, 0.16)	(0.06, 0.16)

Black et al's approach

Black et al (2018) report sensitivity intervals using t-stat and RUV from

1. variable with highest RUV
2. variable with highest t-stat
3. RUV from 1, t-stat from 2
4. RUV and t-stat from all covariates jointly

Implementation in R (1)

```
source("http://andy.egge.rs/code/HHH_sensitivity_interval.R")
```

For perspective, look at t-stats and pct reduction in unexplained variance from Lalonde covariates:

```
tab <- make_specpars_table(formula = re78 ~ ., data = lalonde,
                           treatment_varname = "treat")
round(tab, 3)
```

```
##              |tW|  PRUV
## age         1.298 0.026
## educ        2.465 1.059
## race       22.697 0.618
## married     2.584 0.057
## nodegree    2.239 0.016
## re74        2.031 4.107
## re75        0.689 0.804
```

Implementation in R (2)

Get sensitivity interval assuming an omitted variable like **re74**:

```
re74_si <- make_sensitivity_interval(formula = re78 ~ ., data = lalonde,
          treatment_varname = "treat", tw = 2.03, r.par = .041)
round(re74_si, 3)
```

```
## [1] -278.848 3375.335
```

Get sensitivity interval assuming an omitted variable as related to the outcome as **re74** and as related to the treatment as **race**:

```
hybrid_si <- make_sensitivity_interval(formula = re78 ~ ., data = lalonde,
          treatment_varname = "treat", tw = 22.7, r.par = .041)
round(hybrid_si, 3)
```

```
## [1] -4086.686  7183.173
```

Sensitivity analysis: extra material

The standard OVB decomposition

We estimate a "long" regression

$$E[Y_i] = \beta_0 + \tau D_i + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

and a "short" regression omitting X_k

$$E[Y_i] = \beta_{0,s} + \tau_s D_i + \beta_{1,s} X_1 + \beta_{2,s} X_2 + \ldots + \beta_{k-1,s} X_{k-1}.$$

Consider this "auxiliary" regression:

$$E[X_k] = \alpha_0 + \tau_a D_i + \alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_{k-1} X_{k-1}.$$

Brace yourself, because:

$$\tau_s - \tau = \tau_a \beta_k$$

Standard OVB decomposition in R

We said $\tau_s - \tau = \tau_a \beta_k$ where

- τ_s is treatment effect in short regression (omitting X_k)
- τ is treatment effect in long regression (including X_k)
- β_k is coefficient on X_k in long regression

Confirm this in the Lalonde data where **re74** is the omitted variable.

OVB decomposition in R (sol'n)

```
reg_long <- lm(re78 ~ treat + age + educ + race + married +
                 nodegree + re74 + re75, data = lalonde)
reg_short <- lm(re78 ~ treat + age + educ + race + married +
                  nodegree + re75, data = lalonde)
(bias_direct <- coef(reg_short)["treat"] - coef(reg_long)["treat"])
```

```
##    treat
## -327.028
```

```
# auxiliary regression
reg_aux <- lm(re74 ~ treat + age + educ + race + married +
                nodegree + re75, data = lalonde)
(bias_decomp <- coef(reg_aux)["treat"]*coef(reg_long)["re74"])
```

```
##    treat
## -327.028
```

Applying to sensitivity analysis

We could base our sensitivity analysis on the standard decomposition, i.e.

$$\overbrace{\tau_s - \tau}^{\text{OVB}} = \overbrace{\widetilde{\tau_a}}^{\text{OV-treat link}} \times \overbrace{\widetilde{\beta_k}}^{\text{Outcome-OV link}}$$

Two downsides relative to HHH:

- Both components here are tied to units; in HHH only $\text{SE}(\tau)$ is
- Regression of omitted variable on treatment and other covariates (here) less interpretable than regression of treatment on omitted variable and other covariates (HHH)