5. Observational Studies

ISS5096 || ECI
Jaewon (“Jay-one”) Yoo
National Tsing Hua University

Roadmap

1. Identification in Observational Studies

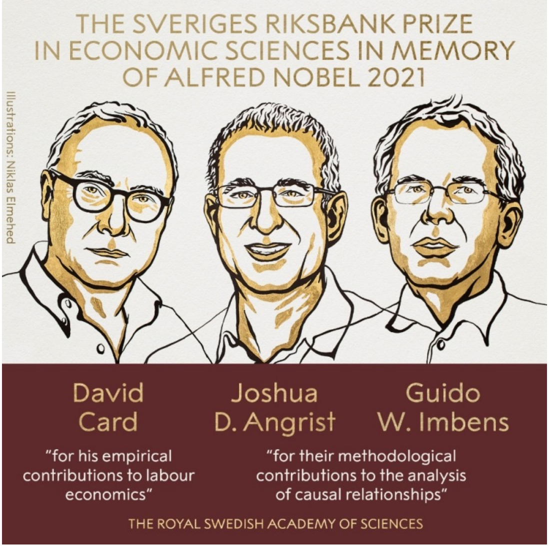
2. Selection on Observables

3. Sensitivity Analysis

Where are we? Where are we going?

- So far: experiments where design makes things easier.
- Today: what happens when we have observational studies to work with?
  - **Identification** and **selection on observables**: assumptions and estimation.
  - What if the assumptions are wrong? $\rightsquigarrow$ **sensitivity analysis** or **partial identification**.
  - Rest of the course will cover different designs for observational studies.
- Q: Why are observational studies in causal inference important? (What are the limitations of RCTs?)

Where are we? Where are we going?



Source: Twitter *@NobelPrize*

Where are we? Where are we going?

Scientific Background on the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2021

ANSWERING CAUSAL QUESTIONS USING OBSERVATIONAL DATA

The Committee for the Prize in Economic Sciences in Memory of Alfred Nobel

"Taken together, therefore, the Laureates' contributions have played a central role in establishing the so-called design-based approach in economics. This approach – aimed at emulating a randomized experiment to answer a causal question ***using observational data*** *– has transformed applied work and improved researchers' ability to answer causal questions of great importance for economic and social policy using observational data." (p.2)*

1/ Identification in Observational Studies

Randomized Experiment Review

- **Experiment**: when the researcher controls the treatment assignment.
  - $p_i = \mathbb{P}[D_i = 1]$ is the probability of treatment assignment.
  - p_i is controlled by & known to researcher in an experiment.
- **Randomized experiment** is an experiment with two properties:
  1. **Positivity**: assignment is probabilistic (and not deterministic): $0 < \mathbb{P}[D_i = 1] < 1$
  2. **Unconfoundedness**: $\mathbb{P}[D_i = 1 | \mathbf{Y}(1), \mathbf{Y}(0)] = \mathbb{P}[D_i = 1]$
     - Treatment assignment does not depend on any potential outcomes.
     - Sometimes written as $D_i \perp\!\!\!\perp (\mathbf{Y}(1), \mathbf{Y}(0))$.

What is the Selection Problem?

- What if we **observe** a non-randomized treatment?
  - Maybe treatment assignment is **confounded** so D_i is related to POs.
- What can we learn about the ATE here? Look at the diff-in-means:

$$
\begin{aligned}
&\mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0] \\
&= \mathbb{E}[Y_i(1)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0] \\
&= \mathbb{E}[Y_i(1)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 1] + \mathbb{E}[Y_i(0)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0] \\
&= \underbrace{\mathbb{E}[Y_i(1) - Y_i(0)|D_i = 1]}_{\text{ATT}} + \underbrace{\mathbb{E}[Y_i(0)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0]}_{\text{selection bias}}
\end{aligned}
$$

- Without unconfoundedness: naive diff-in-means = PATT + selection bias
- **Selection bias**: how different the treated and control groups are in terms of their potential outcome under control.

Selection Bias = Unidentified ATT

$$\mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0] = \underbrace{\tau_{\text{treated}}}_{\text{ATT}} + \underbrace{\mathbb{E}[Y_i(0)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0]}_{\text{selection bias}}$$

- Difference in means: a combination of two unknown quantities.
  - Can't distinguish if a diff-in-means is the ATT or selection bias.
- Example: effect of comment sections on support for online influencers.
  - Naive estimate: influencers do worse without comment sections than with them.
  - $\rightsquigarrow$ negative ATT **OR** positive ATT with large negative selection bias.
  - SB = influencers that disable user comments are worse than those that keep them, even if they posted the same content.
- With an unbounded Y_i, we cannot even bound the ATT because, in principle, SB could be anywhere from $-\infty$ to ∞.
- We say ATT (as well as ATE) are **unidentified** w/o further assumptions.

What is identification?

- **Identification** connects the counterfactual to the observed.
  - **Counterfactual distribution** $\mathbb{P}^*$ of $\{Y_i(1), Y_i(0), D_i, \mathbf{X}_i\}$.
  - **Observational distribution** $\mathbb{P}$ of $\{Y_i, D_i, \mathbf{X}_i\}$.
  - Causal quantities are functions of $\mathbb{P}^*$, but we get samples from $\mathbb{P}$.
    - $\rightsquigarrow$ We can only learn about $\mathbb{P}^*$ through $\mathbb{P}$!
- Quantity ψ (/(p)saI/) is **identified** if we can write it as function of $\mathbb{P}$.
  - Would we know this quantity if we had access to unlimited data?
  - $\rightsquigarrow$ no worrying about estimation uncertainty here.
- Connecting counterfactuals to the observational requires **assumptions**.
  - **"What is your identification strategy?"** = what are the assumptions that allow you to claim that you've estimated a causal effect?
  - Research design can help justify assumptions (experiments, RDD, etc).
  - Or you will need to justify them through argument.

Identification vs. Estimation

- Identification tells us **what** to estimate, not **how**.
  - If identified, we know our causal parameter is some function of $\mathbb{P}$.
  - For example, let's consider the **population** diff-in-means:

$$\mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0]$$

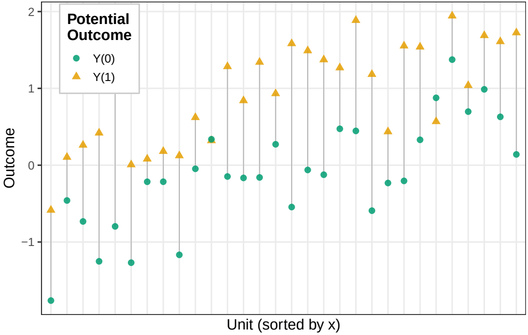
  - But, $\mathbb{P}$ is not directly observable since it's a population distribution!
- Once identified, we need to actually **estimate** the function of $\mathbb{P}$.
  - $\widehat{\tau}_{\text{diff}}$ is an estimator for population diff-in-means
  - Now just estimating conditional expectations, etc.
  - $\rightsquigarrow$ **after identification, causal inference part done**
  - Purely a statistical question from here on out.
- Identification comes first, then comes estimation.
  - Without identification, properties of the estimator are unimportant.
  - keep them separate: estimator shouldn't drive identification.

What is Confounding?

- **Confounding**: treatment and potential outcomes are not independent!
    - Due to "common causes" of Y_i and D_i.
    - Main concern in observational studies.
- Pervasive in management/social sciences:
    - Effect of job training program on employment (confounder: motivation)
    - Effect of college GPAs on salary (confounder: intelligence)
    - Effect of income on voting (confounder: age)
    - Effect of corporate giants on economic development (confounder: previous economic development)
- Confounding $\rightsquigarrow$ incomplete identification of ATE $\rightsquigarrow$ biased estimators.

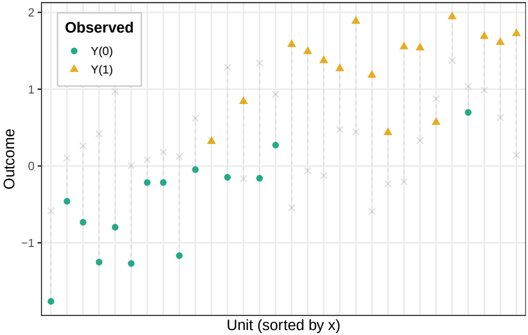
Confounding: The Science Table

- Simulated data (true ATE = 1). Ideally, we observe **both** $Y_i(1)$ and $Y_i(0)$:
  - Each line segment = unit-level treatment effect $Y_i(1) - Y_i(0)$.



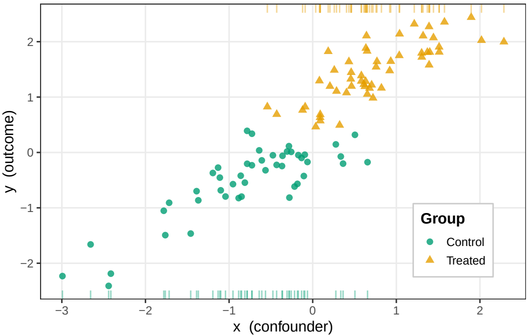
Confounding: The Fundamental Problem

- In reality, we only observe **one** potential outcome per unit.
  - Grey × = missing counterfactual. We cannot compute individual effects!



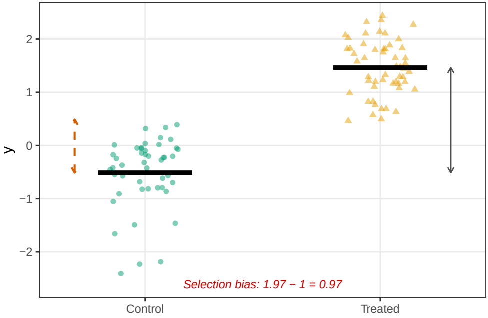
Confounding: What Does It Look Like?

- X_i affects **both** treatment assignment and outcome (confounder).
  - Higher $X_i \rightsquigarrow$ more likely to be treated. Notice the rug plots on top/bottom.
  - (New simulated data with $n = 100$ to visualize the pattern more clearly.)



Confounding: Why It Biases Naive Comparisons

- Simple mean comparison $\overline{Y}_1 - \overline{Y}_0$ is **not** the ATE!
  - Treated group has higher X_i on average $\rightsquigarrow$ higher Y_i even without treatment.



What to do?

2/ Selection on Observables

Observational Studies

- Many different types of identification assumptions we'll cover.
- Begin with most common observational assumptions:
  1. **No unmeasured confounding**: $\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp D_i | \mathbf{X}_i$
     - Also called: conditional unconfoundedness, weak ignorability, selection on observables, no omitted variables, exogenous, conditional exchangeability, etc.
     - $\rightsquigarrow$ Conditional on some covariates, D_i is (effectively) randomly assigned.
  2. **Positivity** or **Overlap**: $0 < \mathbb{P}[D_i = 1 | \mathbf{X}_i] < 1$
     - Treatment and control are both possible at every value of $\mathbf{X}_i$
     - $\rightsquigarrow$ There are both treated and untreated units for each level of $\mathbf{X}_i$ (i.e., "common support").
- We'll take $\mathbf{X}_i$ as a 'given' for now and see later how we might choose it.
- These are assumptions that **can be wrong!**

Identification of the ATE

- Positivity and no unmeasured confounders will identify the PATE:

$$
\begin{aligned}
\tau &= \mathbb{E}[Y_i(1) - Y_i(0)] \\
&= \mathbb{E}_{\mathbf{X}}[\mathbb{E}[Y_i(1) - Y_i(0) | \mathbf{X}_i]] \\
&= \mathbb{E}_{\mathbf{X}}[\mathbb{E}[Y_i(1) | \mathbf{X}_i] - \mathbb{E}[Y_i(0) | \mathbf{X}_i]] \\
&= \mathbb{E}_{\mathbf{X}}[\mathbb{E}[Y_i(1) | D_i = 1, \mathbf{X}_i] - \mathbb{E}[Y_i(0) | D_i = 0, \mathbf{X}_i]] \\
&= \mathbb{E}_{\mathbf{X}}[\mathbb{E}[Y_i | D_i = 1, \mathbf{X}_i] - \mathbb{E}[Y_i | D_i = 0, \mathbf{X}_i]]
\end{aligned}
$$

- Useful to write the treated and control CEFs:

$$
\mu_1(\mathbf{x}) = \mathbb{E}[Y_i(1) | \mathbf{X}_i = \mathbf{x}], \qquad \mu_0(\mathbf{x}) = \mathbb{E}[Y_i(0) | \mathbf{X}_i = \mathbf{x}]
$$

  - How the mean of the potential outcomes vary with the covariates.

- Key part of the identification above:

$$
\underbrace{\mu_1(\mathbf{x})}_{\text{counterfactual}} = \underbrace{\mathbb{E}[{\color{red}Y_i} | D_i = 1, \mathbf{X}_i = \mathbf{x}]}_{\text{observational}}, \qquad \mu_0(\mathbf{x}) = \mathbb{E}[{\color{red}Y_i} | D_i = 0, \mathbf{X}_i = \mathbf{x}]
$$

Regression Estimation of the ATE

- Identification done, now turn to estimation!
- Regression estimators $\widehat{\mu}_1(\mathbf{x})$ and $\widehat{\mu}_0(\mathbf{x})$.
    - Might be linear or nonlinear models.
    - Safest practice: estimate separate regressions in each treatment group.
    - Sometimes called an **imputation** or **plug-in** estimator.
- Regression/imputation estimator of the ATE:

$$\widehat{\tau}_{\text{reg}} = \frac{1}{n} \sum_{i=1}^{n} \widehat{\mu}_1(\mathbf{X}_i) - \widehat{\mu}_0(\mathbf{X}_i)$$

- Procedure:
    1. Obtain predicted values for all units when $D_i = 1$.
    2. Obtain predicted values for all units when $D_i = 0$.
    3. Take the average difference between these predicted values.

Regression and the Imputation Estimator

- What if we estimate $\widehat{\mu}_1(\mathbf{x})$ and $\widehat{\mu}_0(\mathbf{x})$ jointly in a single linear model?
- Uninteracted OLS: $Y_i = \alpha + \tau D_i + \beta' \mathbf{X}_i + \varepsilon_i$

$$\mathbb{E}[Y_i | D_i = d, \mathbf{X}_i] = \begin{cases} \underbrace{(\alpha + \tau) + \beta' \mathbf{X}_i}_{\mu_1(\mathbf{X}_i)} & \text{if } d = 1 \\ \underbrace{\alpha + \beta' \mathbf{X}_i}_{\mu_0(\mathbf{X}_i)} & \text{if } d = 0 \end{cases}$$

  - $\mu_1(\mathbf{X}_i) - \mu_0(\mathbf{X}_i) = \tau$ for all i (same slope β).
  - $\rightsquigarrow$ Imputation estimator = coefficient on $D_i = \widehat{\tau}$.

Regression and the Imputation Estimator

- Fully interacted OLS (centered):

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

$$\mathbb{E}[Y_i | D_i = d, \widetilde{\mathbf{X}}_i] = \begin{cases} \underbrace{(\alpha + \tau) + (\beta + \gamma)' \widetilde{\mathbf{X}}_i}_{\mu_1(\widetilde{\mathbf{X}}_i)} & \text{if } d = 1 \\ \underbrace{\alpha + \beta' \widetilde{\mathbf{X}}_i}_{\mu_0(\widetilde{\mathbf{X}}_i)} & \text{if } d = 0 \end{cases}$$

  - $\mu_1(\widetilde{\mathbf{X}}_i) - \mu_0(\widetilde{\mathbf{X}}_i) = \tau + \gamma' \widetilde{\mathbf{X}}_i$ — varies across units!
  - But on average $= \tau$, since $\overline{\widetilde{\mathbf{X}}} = 0$.
  - $\rightsquigarrow$ Imputation estimator = coefficient on $D_i = \widehat{\tau}$.

- Different slopes for treated and control $\rightsquigarrow$ equivalent to separate regressions.

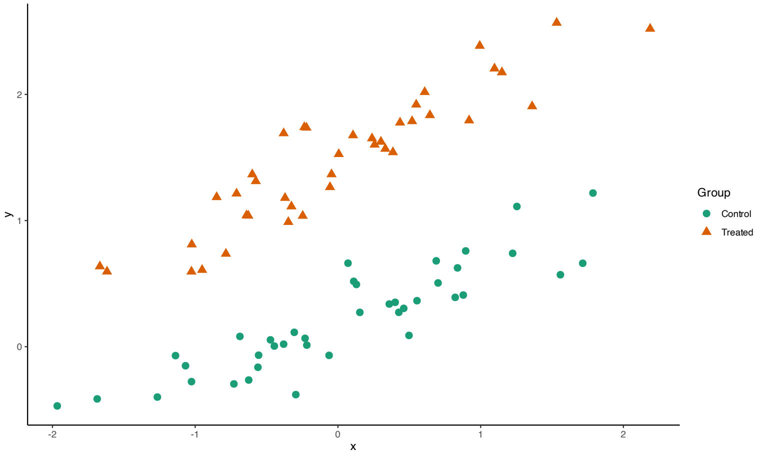
- Key difference: uninteracted assumes **constant** treatment effect; fully interacted allows **heterogeneous** effects.

Variance Estimation

- How do we get estimates of the variance of $\widehat{\tau}_{\text{reg}}$?
- If an OLS coefficient $\rightsquigarrow$ use heteroskedasticity-consistent (robust) variance estimator (e.g., HC2).
- For the general imputation estimator (separate regressions), analytic variance expressions exist but involve estimation uncertainty from both $\widehat{\mu}_1$ and $\widehat{\mu}_0$ — complicated!
- Computational alternative: **(nonparametric) bootstrap**
    - Randomly resample n rows of the data with replacement.
    - Refit the regressions on the bootstrapped data.
    - Calculate $\widehat{\tau}_{\text{reg}}$ in each bootstrap.
    - Repeat several times and use empirical variance of the bootstraps.

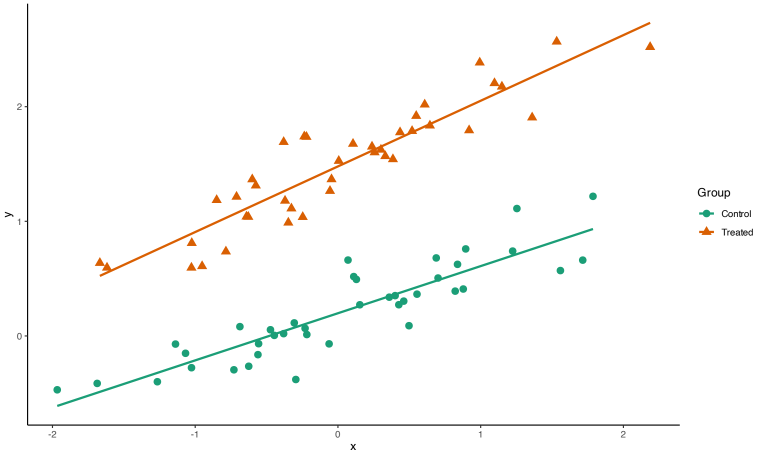
Imputation Estimator Visualization

```
> toy_data <- read_csv("https://bit.ly/3vOy2Ao")
```



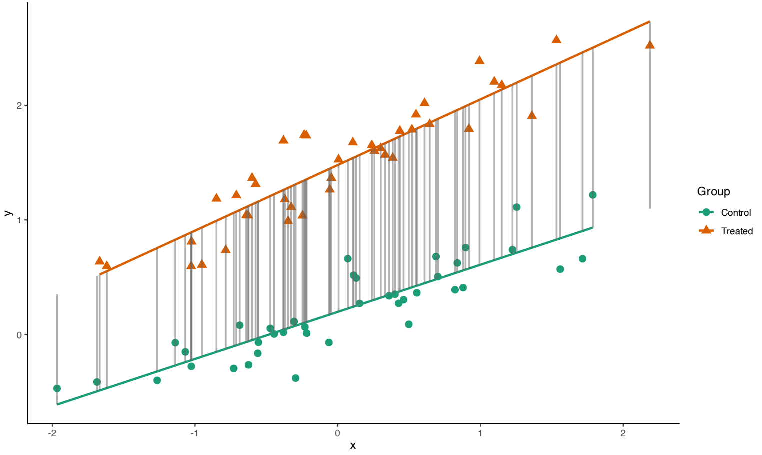
Imputation Estimator Visualization

```
> lm0 <- lm(y~x, data = toy_data, subset = d==0); lm1 <- lm(y~x, data = toy_data, subset = d==1)
```



Imputation Estimator Visualization

```
> mu0.imps = predict(lm0, toy_data); mu1.imps = predict(lm1, toy_data)
> cat("Estimate of ATE:", mean(mu1.imps - mu0.imps))
Estimate of ATE: 1.285176
```



Fully Interacted OLS & Imputation Estimator

- Recall: fully interacted OLS with centered covariates = imputation estimator.
- Let's verify numerically:

```
> toy_data$x_tilde <- toy_data$x - mean(toy_data$x)
> mod_full <- lm(y ~ d + x_tilde + d * x_tilde, data = toy_data)

> cat("\nEstimate of ATE (Imputation):", mean(mu1.imps - mu0.imps),
    "\nEstimated coefficient on Di from full int.", mod_full$coefficients["d"])

Estimate of ATE (Imputation): 1.285176
Estimated coefficient on Di from full int. 1.285176
```

- $\rightsquigarrow$ With centered covariates, $\widehat{\tau}_{\text{reg}} \equiv$ estimated coefficient on D_i.
  - Would be the same for uninteracted model, except the variance will be larger (less precision).

Variance Estimation w/ Bootstrap

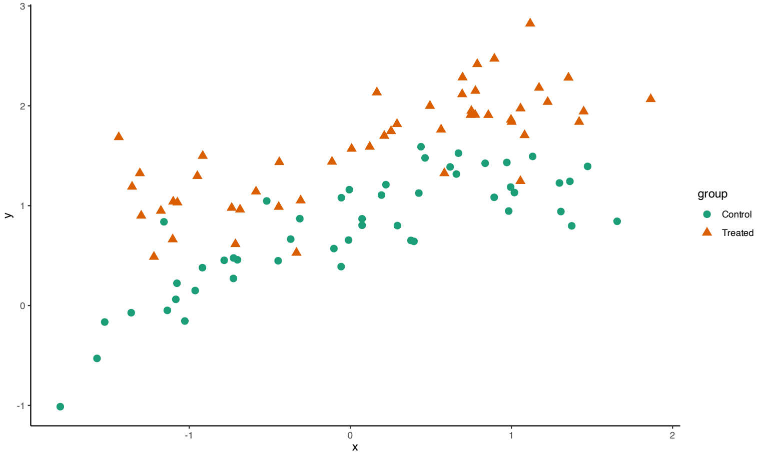
- Again, how do we get estimates of the variance of $\widehat{\tau}_{\text{reg}}$?
  - (Nonparametric) bootstrap: a simulation-based alternative to analytic standard errors.
  - Idea: resample from the data to approximate the distribution of $\widehat{\tau}_{\text{reg}}$.

```r
> set.seed(02138); sims<-500; tau_hat_draws<-rep(NA, sims)
> for (i in 1:sims) { # Repeat the following several times
    # 1. Randomly resample n rows of the data with replacement
    sample_boot <- dplyr::slice_sample(toy_data, n = nrow(toy_data), replace = TRUE)

    # 2. Refit the regressions on the bootstrapped data
    model <- lm(y ~ d + x_tilde + d*x_tilde, data = toy_data)
    dat1 <- sample_boot; dat1$d <- 1
    dat0 <- sample_boot; dat0$d <- 0
    mu1_hat <- predict(model, newdata = dat1)
    mu0_hat <- predict(model, newdata = dat0)

    # 3. Calculate tau_hat in each bootstrap
    tau_hat_draws[i] <- mean(mu1_hat - mu0_hat)
  }

> # 4. Use empirical variance of the bootstraps
> var(tau_hat_draws)
[1] 0.000254049
```

Nonlinear Relationships

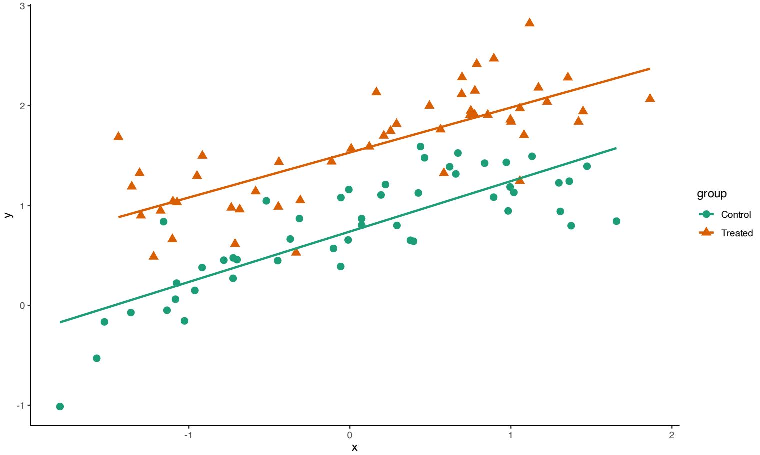
- Same idea but with nonlinear relationship between Y_i and X_i:

```
> toy_data_02 <- read_csv("https://bit.ly/4cOg0Pl") # sim data with nonlinear relationship
```



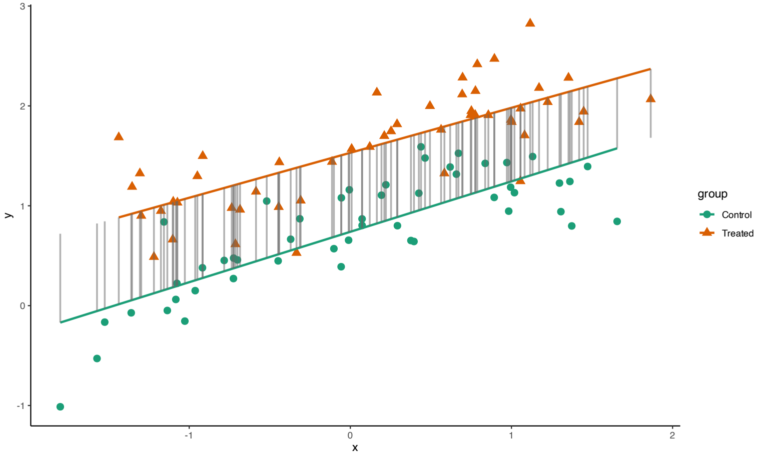
Nonlinear Relationships

- Same idea but with nonlinear relationship between Y_i and X_i:



Nonlinear Relationships

- Same idea but with nonlinear relationship between Y_i and X_i:

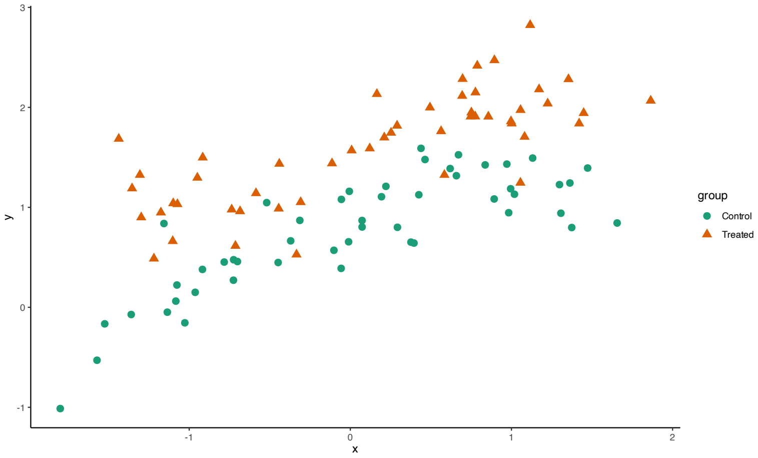


Using Semiparametric Regression

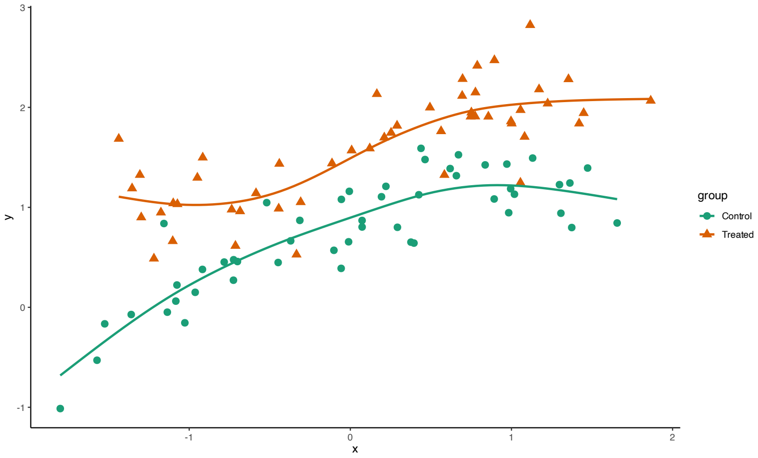
- Here, CEFs are nonlinear, but we don't know their form.
- We can use generalized additive models (GAMs) from the mgcv package for 'flexible' estimation:

```
> library(mgcv)
> gam1 <- gam(y~s(x), data = toy_data_02, subset = group=="Treated")
> gam0 <- gam(y~s(x), data = toy_data_02, subset = d==0); summary(gam1)

Family: gaussian
Link function: identity

Formula:
y ~ s(x)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.59527    0.04571    34.9   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
      edf Ref.df     F p-value
s(x) 3.73  4.642 19.85  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.651   Deviance explained = 67.7%
GCV = 0.11539  Scale est. = 0.10447  n = 50
```

Using GAMS



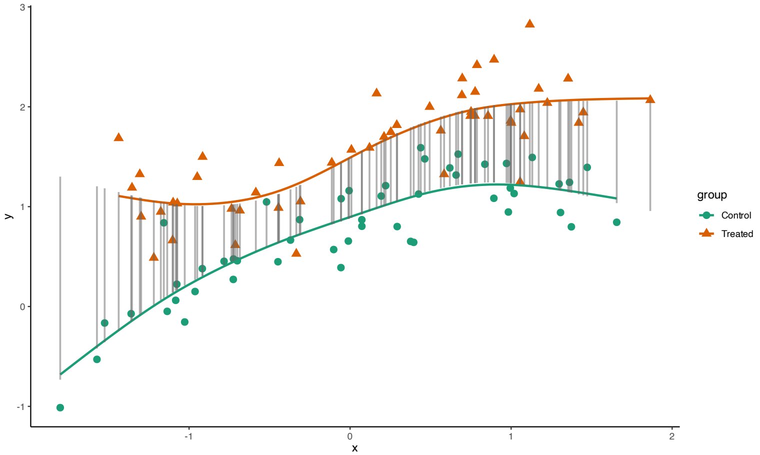
Using GAMS



Using GAMS

- We can estimate $\widehat{\tau}_{\text{reg}}$ using the imputation estimator.

```
> cat("Estimate of ATE (GAM):",mean(predict(gam1) - predict(gam0)))

Estimate of ATE (GAM): 0.8379884
```



3/ Sensitivity Analysis

Cf. Cinelli & Hazlett (2020), *Making Sense of Sensitivity*, JRSS-B

Sensitivity Analysis for Regression

- Standard regression estimator of the ATE:

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

  - Note: applies equally to fully interacted models (cf. W4).

- 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:

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

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

  - But γ and δ are coefficients of an **unobserved** variable: we don't know what U is, let alone its scale.

Why Not Use Regression Coefficients?

- Recall: bias $= \gamma \times \delta$. Can't we just benchmark against observed coefficients?

- Problem: regression coefficients are **scale-dependent**.
  - `female` is binary (0/1), age is continuous (18–80), U is…unknown.
  - If you rescale a variable by 10, its coefficient shrinks by 10.
  - ⇝ Comparing $\gamma = 0.5$ to a coefficient of 0.3 is meaningless if the scales differ.

- **Partial** R^2: proportion of residual variance explained by one variable (scale-free).

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

OVB in Terms of Partial R^2

- Cinelli and Hazlett (JRSS-B, 2020) rewrite OVB using partial R^2:

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

- Same structure as $\gamma \times \delta$:
  - $R^2_{Y\sim U|D,\mathbf{X}}$: how much U predicts Y (partialing out $D, \mathbf{X}$) $\rightsquigarrow$ the $U \longrightarrow Y$ path.
  - $R^2_{D\sim U|\mathbf{X}}$: how much U predicts D (partialing out $\mathbf{X}$) $\rightsquigarrow$ the $U \longrightarrow D$ path.
  - $1 - R^2_{D\sim U|\mathbf{X}}$: "clean" treatment variation not driven by U. As U dominates D, this $\longrightarrow 0$ and bias explodes.
  - $\mathbb{V}(\cdot)/\mathbb{V}(\cdot)$: scales the bias to the same units as $\widehat{\tau}$ (computed from data).

Why This Helps: Benchmarking

- Both partial R^2's are unknown (since U is unobserved).
- But partial R^2 of **observed** covariates *can* be computed from data!
  - E.g., $R^2_{Y\sim\texttt{female}|D,\mathbf{X}}$ and $R^2_{D\sim\texttt{female}|\mathbf{X}}$ are known.
- $\rightsquigarrow$ We can ask: "How strong would U need to be, *relative to observed confounders*, to overturn our finding?"
- **Sensitivity analysis**: vary the two unknown R^2's $\in [0, 1]$ and trace out how $\widehat{\tau}$ changes.
  - This is exactly what the contour plot in the next slides will show.

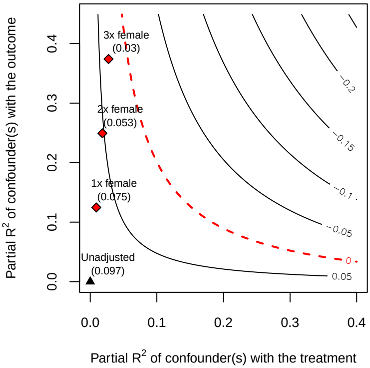
R Package: sensemakr

```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)
```

- Darfur conflict (Sudan, 2003) survey: D = directly harmed by violence, Y = support for peacemaking.
- `female` is the strongest known confounder ⇝ used as benchmark.
  - kd (knock-on distance): how many times stronger would the unobserved confounder need to be (relative to `female`) to change our conclusion?

Sensitivity Analysis with Partial R^2

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



- Contour lines = adjusted estimate $\widehat{\tau}$.
- Red diamonds = benchmarks at 1–3× strength of `female`.

- Here: even a confounder 3× as strong as `female` would not eliminate the effect. ⇝ Our estimate is robust to plausible levels of unobserved confounding.

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
- Sensitivity analysis asks: how strong would an unobserved confounder need to be to overturn the finding? The partial R^2 framework (Cinelli and Hazlett, 2020) provides a principled way to answer this question.
- See also: **5b. Partial Identification** for what we can learn under weaker assumptions (bounds instead of point estimates).
- Up next: DAGs, i.e., a graphical framework for reasoning about which covariates to condition on.

Onto the presentations & discussions!

Contact Information:
jaewon.yoo@iss.nthu.edu.tw
https://j1yoo.github.io/