Regression and Causal Inference

Prof. Tzu-Ting Yang
楊子霆

Institute of Economics, Academia Sinica
中央研究院經濟研究所

September 24, 2024

Regression: Main Idea

Main Idea of Regression

- A multivariate regression can help us study the relationship between treatment D_i and outcome Y_i

$$\mathrm{Y}_i = \delta + \alpha D_i + X_i\beta + \epsilon_i$$

- Here, X is a vector of covariates and β is a vector of coefficients

$$X = (x_1', x_2', \ldots, x_k')$$

$$\beta = (\beta_1, \beta_2, \ldots, \beta_k)$$

$$X\beta = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

Main Idea of Regression

- We can interpret α as the causal effect of treatment by including all observed confounding factors X_i in the regression
- The inclusion of X allows for a like-with-like comparison
  - We compare units with the same values of X but different values of D
  - But the like-with like comparison is only valid if X contains all confounding factors
  - Again, this interpretation is based on Conditional Independence Assumption (CIA)

Regression: Potential Outcomes Framework

Conditional Independence Assumption (CIA)

Conditional Independence Assumption

$$(\mathrm{Y}_i^1, \mathrm{Y}_i^0) \perp\!\!\!\perp D_i | X_i$$

- Both matching and regression require CIA (selection on observable) to get causal affects
  - But regression implicitly assume a specific functional form of **potential outcomes**
    - $\mathrm{Y}^0 = \delta + X_i\beta$
    - $\mathrm{Y}^1 = \delta + \alpha + X_i\beta$
- Thus, it ensures that selection bias is eliminated:
  - $\mathrm{E}[\mathrm{Y}_i^0 | X_i, D_i = 1] = \mathrm{E}[\mathrm{Y}_i^0 | X_i, D_i = 0]$
  - $\mathrm{E}[\mathrm{Y}_i^1 | X_i, D_i = 1] = \mathrm{E}[\mathrm{Y}_i^1 | X_i, D_i = 0]$

Regression and Potential Outcome

- Under the CIA, we can estimate the following regression to get causal effect of D by including all possible observed confounding factors X

$$\mathrm{Y}_i = \delta + \alpha D_i + X_i\beta + \epsilon_i$$

- We use the regression estimates to "predict" the counterfactual outcomes
  - For treated units ($D = 1$), we can get counterfactual outcome Y^0

$$\mathrm{Y}_i^0 = \delta + X_i\beta$$

  - For untreated units ($D = 0$), we can get counterfactual outcome Y^1

$$\mathrm{Y}_i^1 = \delta + \alpha + X_i\beta$$

  - CIA implies $\mathrm{E}[\epsilon_i|X_i] = 0$

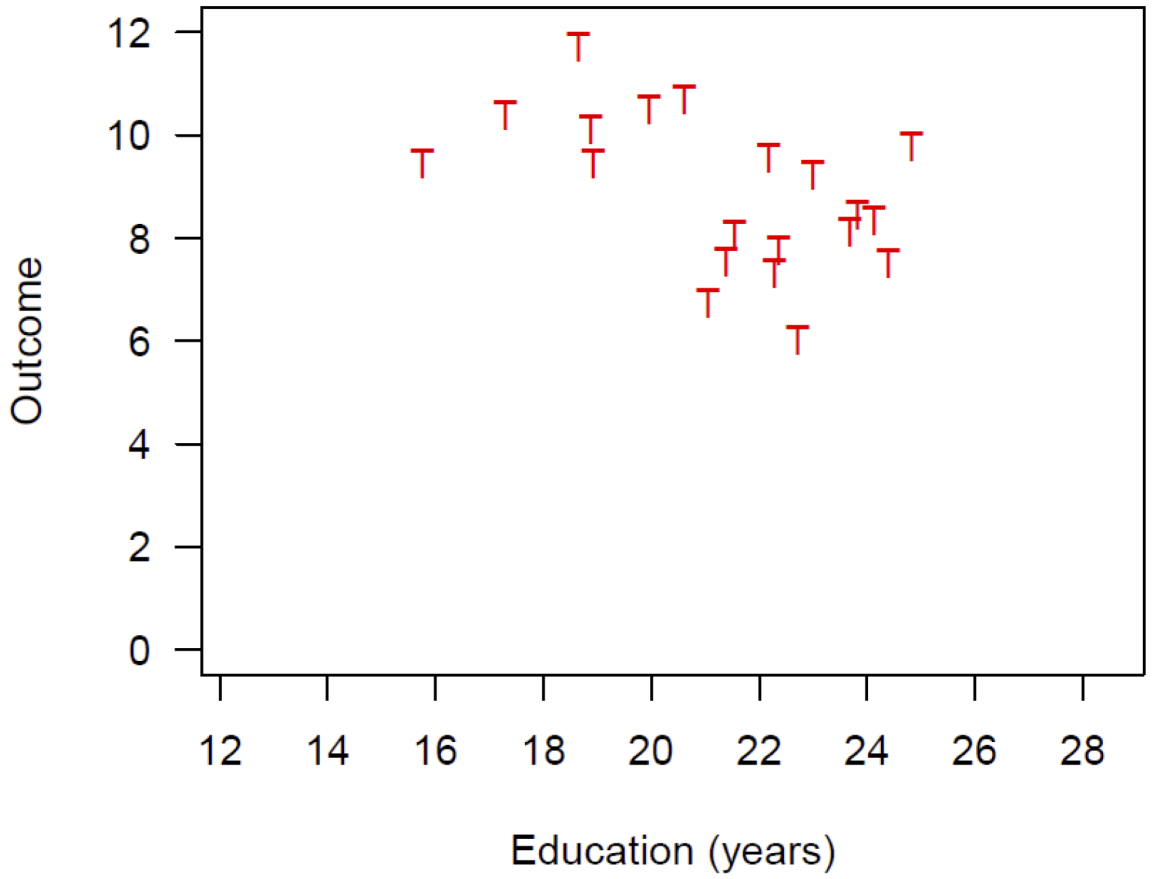
Regression and Matching

A Graphical Example

- Suppose we want to examine the effect of a treatment D on an outcome Y
  - Education is a observed confounding factor X
- Matching methods:
  - Require sufficient overlap in covariate distributions (X) between treated and control groups
  - This is known as the common support assumption
  - Ensures valid counterfactual comparisons

Regression and Matching

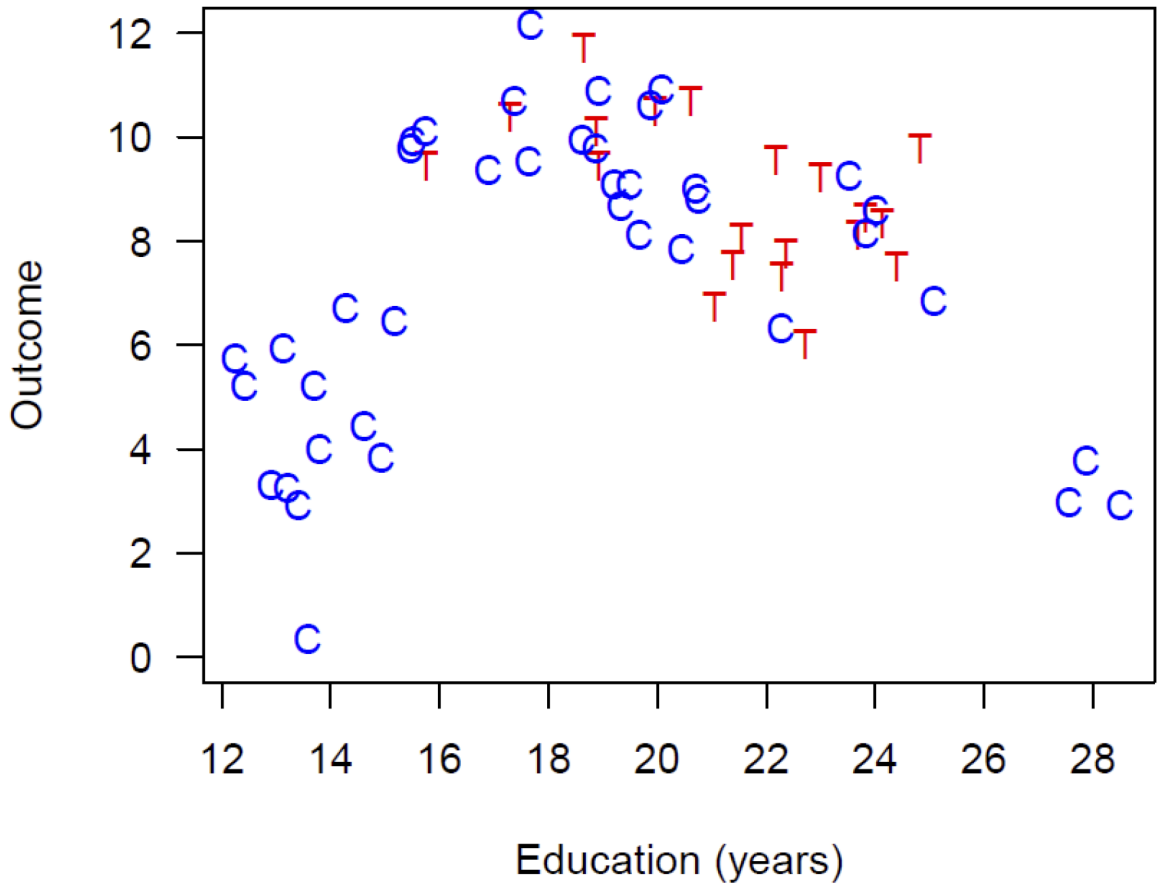
A Graphical Example



Source: Ben Elsner's slides

Regression and Matching

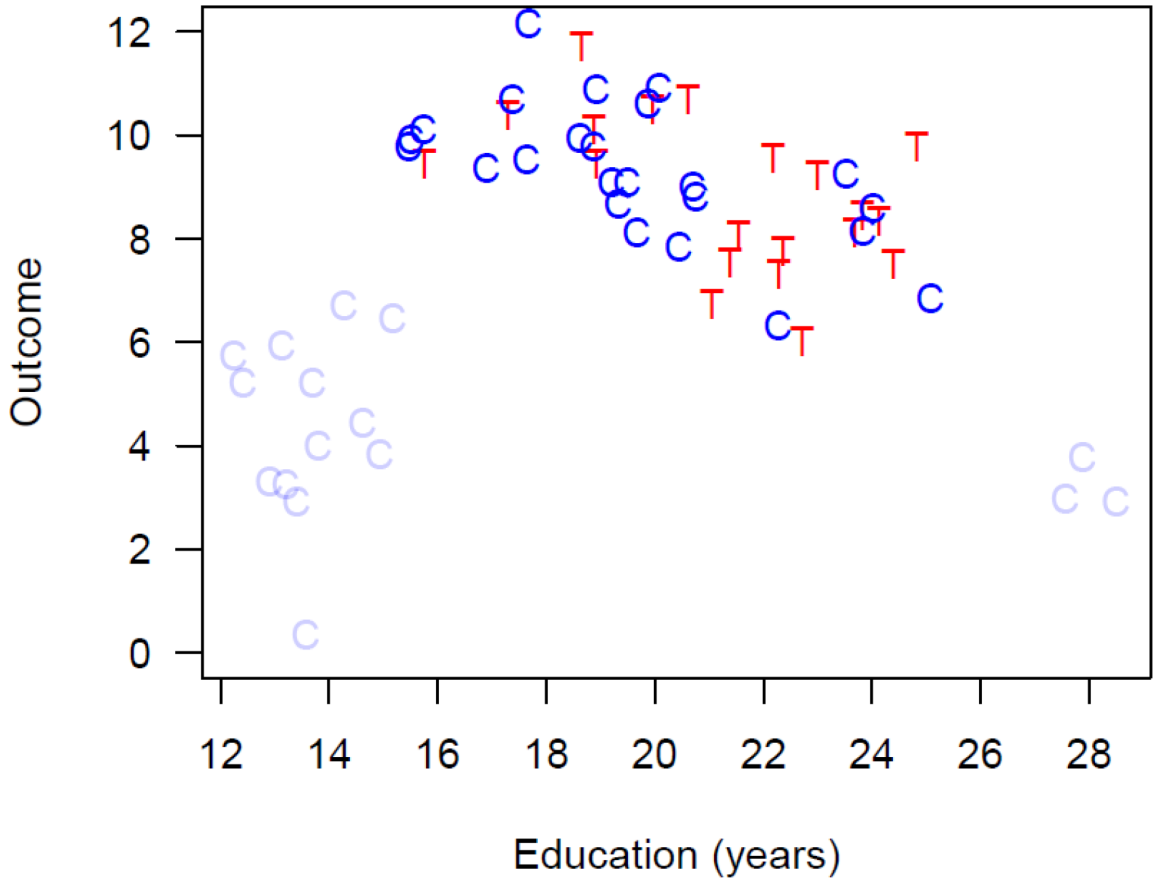
A Graphical Example



Source: Ben Elsner's slides

Regression and Matching

A Graphical Example



Source: Ben Elsner's slides

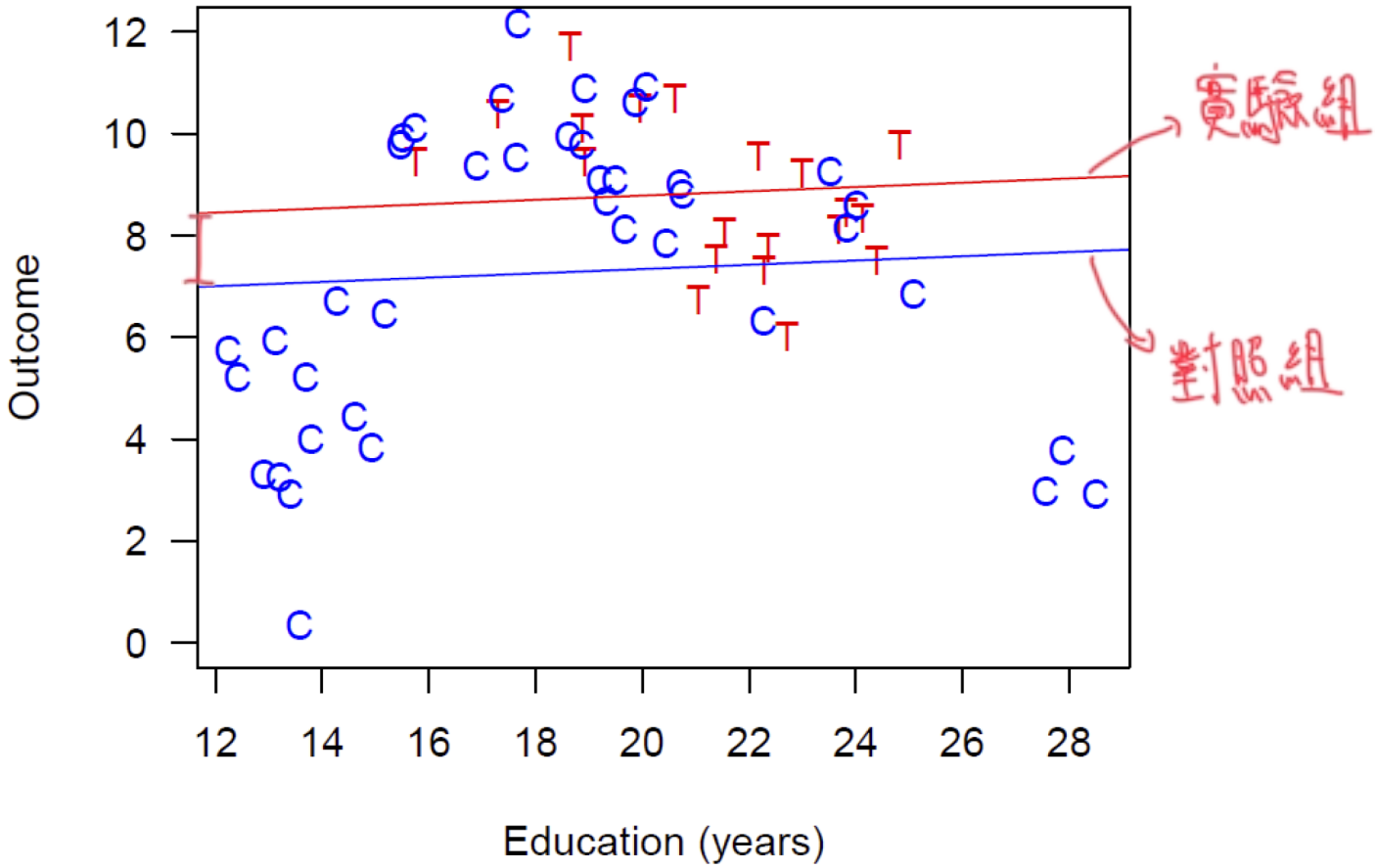
Regression and Matching

A Graphical Example

- Regression methods:
  - Can potentially extrapolate beyond the common support region
  - By relying on the specified regression model to predict counterfactual outcomes
    - Linear term for education: $Y_i = \delta + \alpha D_i + \beta_1 X_i + \epsilon_i$
    - Quadratic term for education: $Y_i = \delta + \alpha D_i + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i$
    - Estimated effect of treatment D can be different for these two models
  - The extrapolation may be unreliable if:
    - Model is misspecified
    - Extrapolation region is too far from data

Regression and Matching

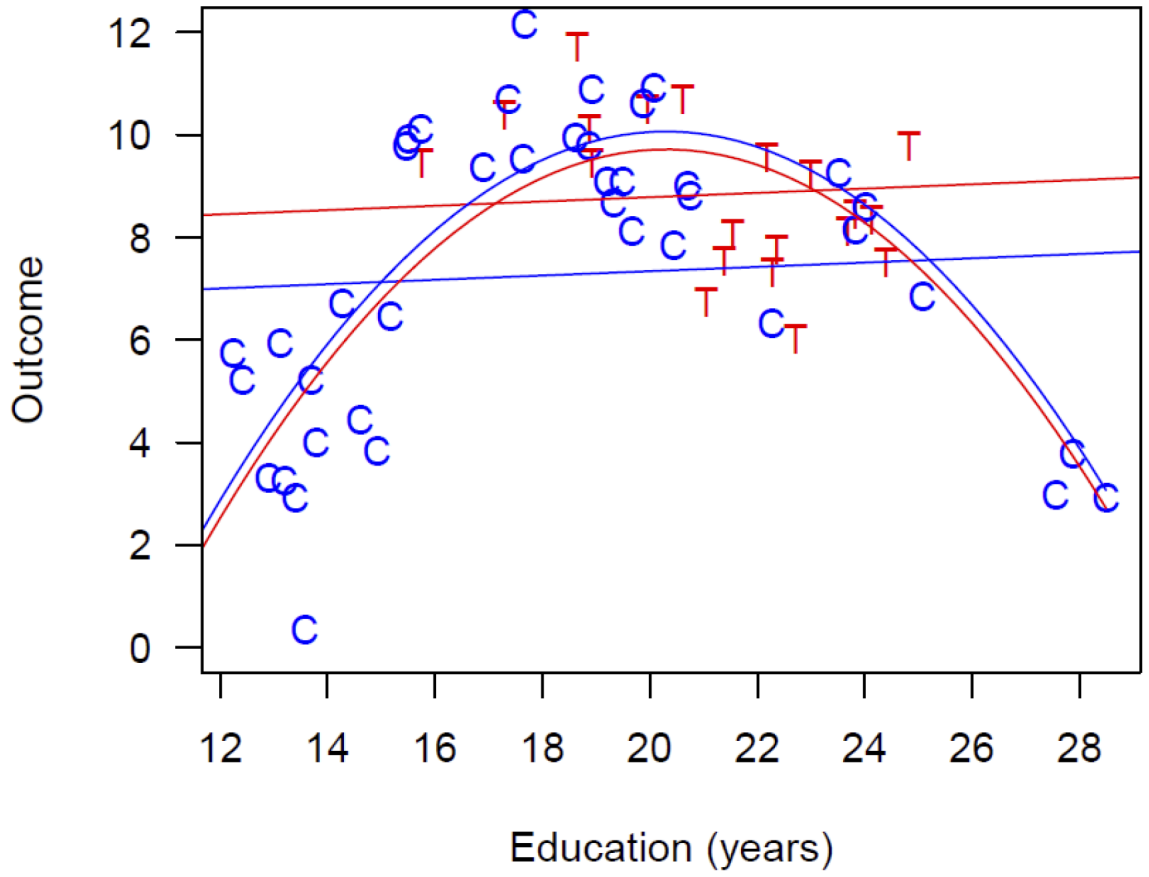
A Graphical Example



Source: Ben Elsner's slides

Regression and Matching

A Graphical Example



Source: Ben Elsner's slides

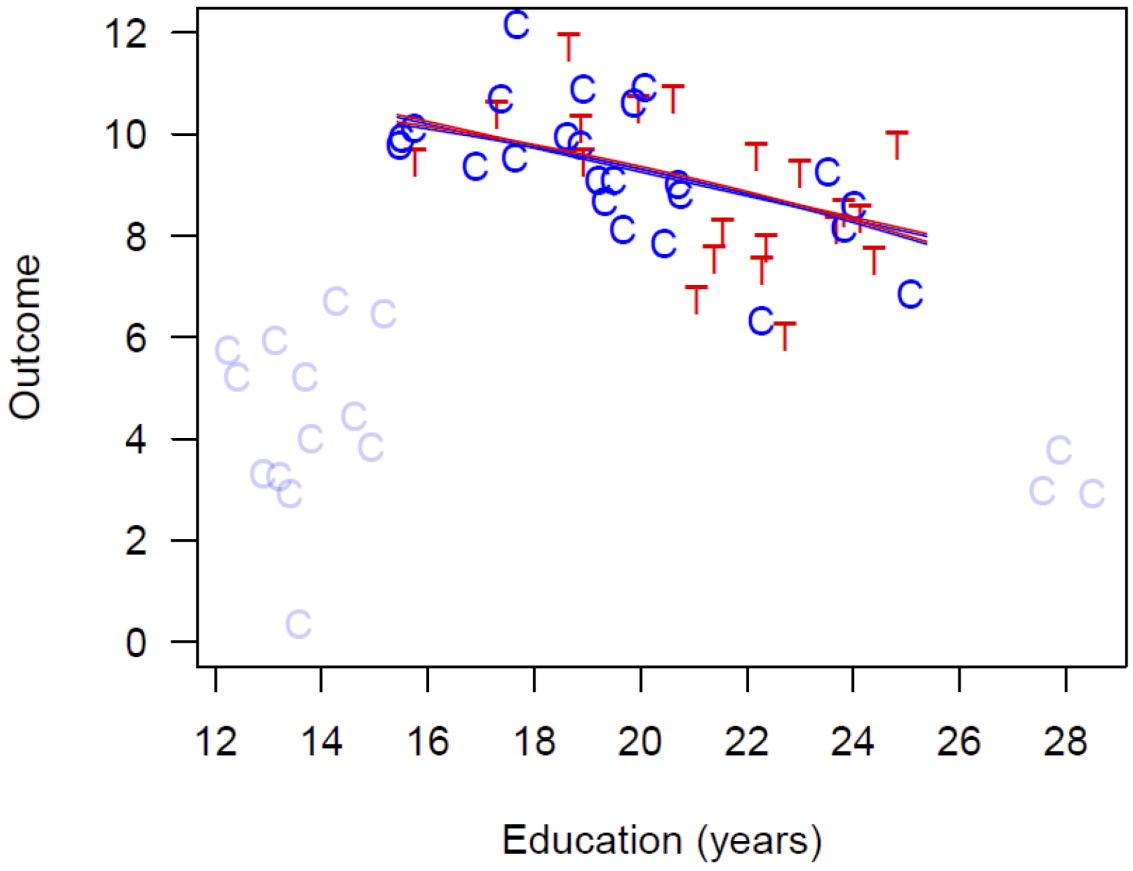
Regression and Matching

A Graphical Example

- Regression methods:
  - Among these units within common support region, there is no difference in outcomes between treatment and control groups

Regression and Matching

A Graphical Example



Source: Ben Elsner's slides

Regression and Matching

Summary

- The previous slides highlight a problem with regression
- Regression methods may face challenges due to lack of common support
  - When covariate distributions do not overlap between treated and control groups
  - Even if both groups have the same average covariate values (e.g., education)
  - The regression line can be influenced by control units in regions without treated units

Regression and Matching

Summary

- Matching methods avoid this issue by restricting comparisons to common support region
  - Only units with similar covariate values are compared
  - Ensures comparisons are made between fundamentally comparable units
  - Avoids extrapolation to regions without data support
- However, matching may discard useful data outside the common support region
  - Regression can potentially utilize this information, if model is correctly specified
  - Trade-off between bias and variance/efficiency

Identification Results for Regression

- We estimate the following regression:

$$\mathrm{Y}_i = \delta + \alpha D_i + X_i\beta + \epsilon_i$$

- The estimated coefficient of treatment D is the following:

$$\alpha(X) = \underbrace{\mathrm{E}[\mathrm{Y}_i|X_i, D_i = 1] - \mathrm{E}[\mathrm{Y}_i|X_i, D_i = 0]}_{\text{ODO at given } X_i}$$

- Based on CIA, including key observed covariates X_i into regression can help us eliminate selection bias

Identification Results for Regression

$$\alpha(X) = \underbrace{\mathrm{E}[\mathrm{Y}_i|X_i, D_i = 1] - \mathrm{E}[\mathrm{Y}_i|X_i, D_i = 0]}_{\text{ODO at given } X_i}$$

$$= \mathrm{E}[\mathrm{Y}_i^1|X_i, D_i = 1] - \mathrm{E}[\mathrm{Y}_i^0|X_i, D_i = 0]$$

$$= \mathrm{E}[\mathrm{Y}_i^1|X_i, D_i = 1] - \mathrm{E}[\mathrm{Y}_i^0|X_i, D_i = 1]$$

$$+ \mathrm{E}[\mathrm{Y}_i^0|X_i, D_i = 1] - \mathrm{E}[\mathrm{Y}_i^0|X_i, D_i = 0]$$

$$= \underbrace{\mathrm{E}[\mathrm{Y}_i^1 - \mathrm{Y}_i^0|X_i, D_i = 1]}_{\text{CATT}}$$

$\delta + \beta X$ ← $\delta + \beta X$

$$+ \underbrace{\mathrm{E}[\mathrm{Y}_i^0|X_i, D_i = 1] - \mathrm{E}[\mathrm{Y}_i^0|X_i, D_i = 0]}_{\text{Selection Bias} = 0}$$

$$= \underbrace{\mathrm{E}[\mathrm{Y}_i^1 - \mathrm{Y}_i^0|X_i, D_i = 1]}_{\text{CATT}}$$

$$+ \underbrace{\beta\mathrm{E}[\mathrm{X}_i|X_i, D_i = 1] - \beta\mathrm{E}[\mathrm{X}_i|X_i, D_i = 0]}_{\text{Selection Bias}}$$

Identification Results for Regression

$$
\begin{aligned}
\alpha(X) &= \underbrace{\mathrm{E}[\mathrm{Y}_i|X_i, D_i = 1] - \mathrm{E}[\mathrm{Y}_i|X_i, D_i = 0]}_{\text{ODO at given } X_i} \\
&= \underbrace{\mathrm{E}[\mathrm{Y}_i^1 - \mathrm{Y}_i^0|X_i, D_i = 1]}_{\text{CATT}} \\
&+ \underbrace{\beta\mathrm{E}[\mathrm{X}_i|X_i, D_i = 1] - \beta\mathrm{E}[\mathrm{X}_i|X_i, D_i = 0]}_{\text{Selection Bias}} \\
&= \underbrace{\mathrm{E}[\mathrm{Y}_i^1 - \mathrm{Y}_i^0|X_i, D_i = 1]}_{\text{CATT}} + \underbrace{0}_{\text{Selection Bias}} \\
&= \underbrace{\mathrm{E}[\mathrm{Y}_i^1 - \mathrm{Y}_i^0|X_i, D_i = 0]}_{\text{CATU}} \\
&= \underbrace{\mathrm{E}[\mathrm{Y}_i^1 - \mathrm{Y}_i^0|X_i]}_{\text{CATE}}
\end{aligned}
$$

Identification Results for Regression

- Note that there are as many causal effects (CATE or CATT) as the number of value in X_i
- People might find it useful to boil a set of estimates down to a single summary measure
  - e.g. Average treatment effect
- Again, applying the law of iterated expectations (LIE), we can identify ATT, ATU, and ATE
  - Take average of CATT, CATU, and CATE over all subgroups (all possible X-values)

Regression: Estimation

Regression: Estimation

- Again, if we have population data, we can get the above causal effect α
- However, we usually only have sample (i.e. part of population data)
- We need to use sample data to estimate α

Review: Ordinary Least Squares Estimation

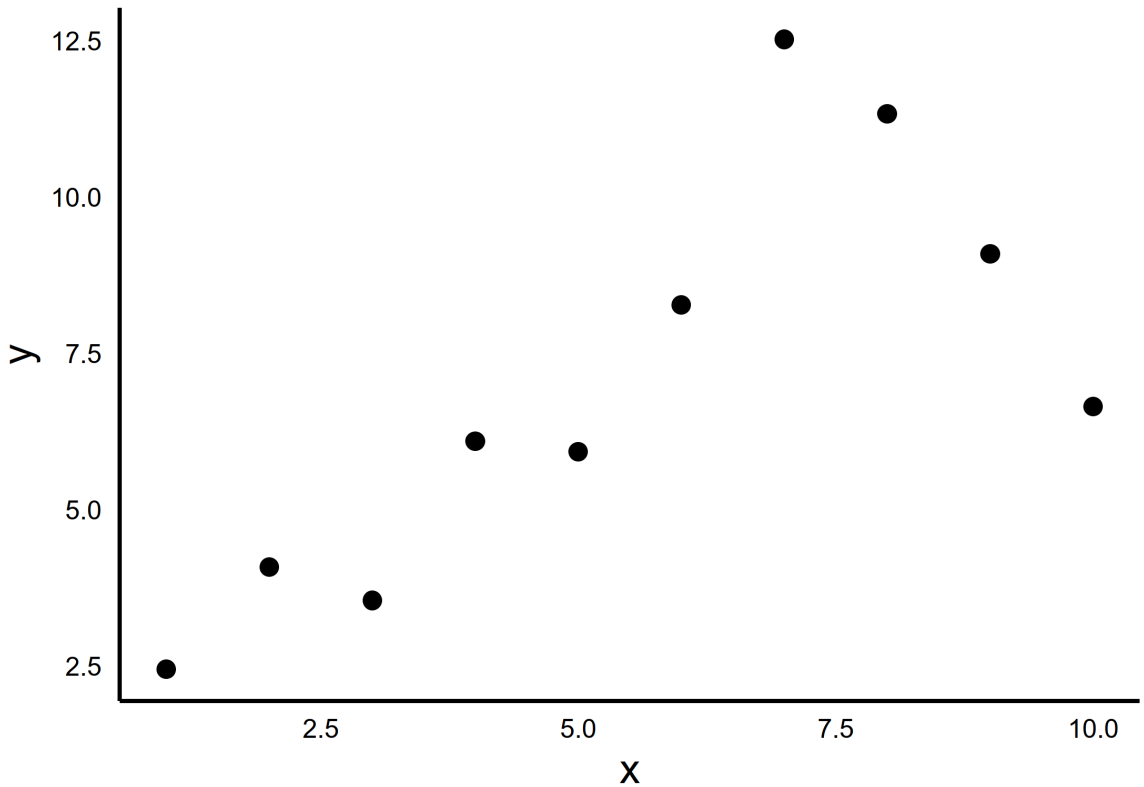
- Regression analysis assigns values to model parameters (δ and α) so as to make $\hat{\mathrm{Y}}_i$ as close as possible to Y_i
- Ordinary Least Squares (OLS) estimation accomplish it by choosing values that **minimize the sum of squared error (SSE)**

$$(\hat{\delta}, \hat{\alpha}) = \min_{\delta,\alpha} \frac{1}{N} \sum_{i=1}^{N} (\mathrm{Y}_i - \delta - \alpha D_i)^2$$

- The result is the best-fitting line that describes the relationship between D and Y in the sample

Review: Ordinary Least Squares Estimation

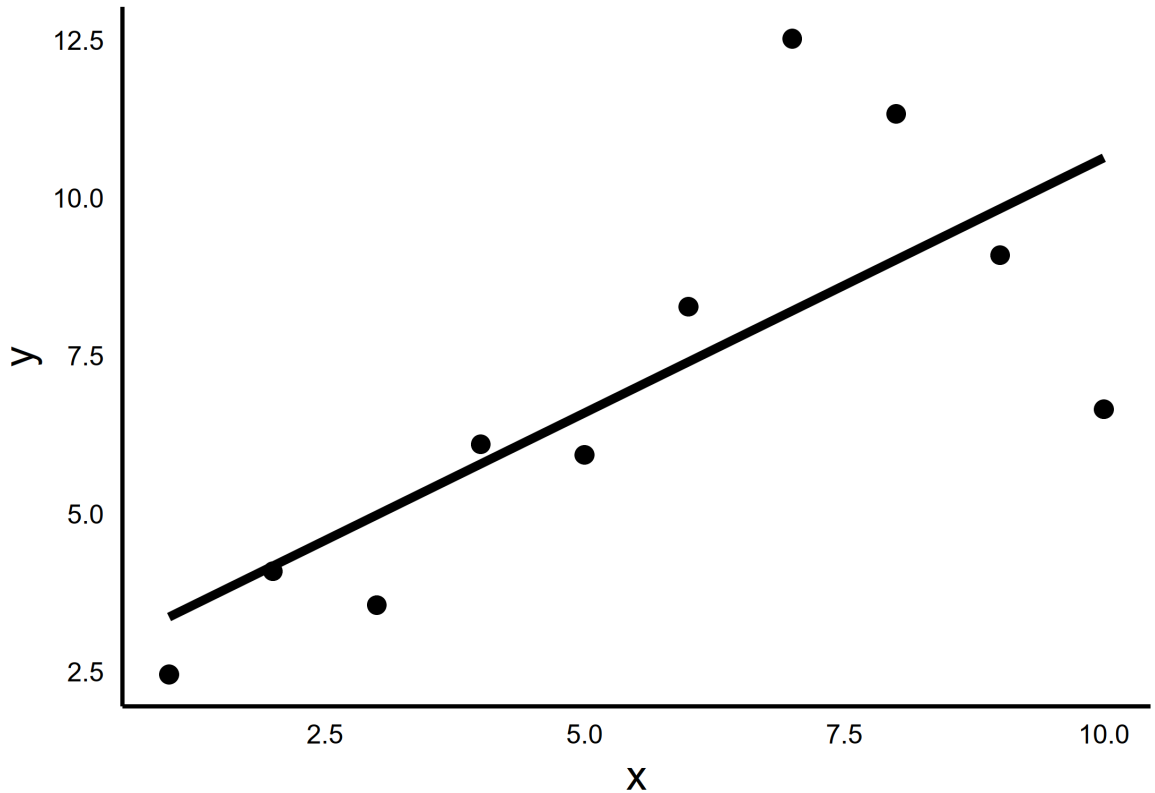
A Graphical Example



Source: Ben Elsner's slides

Review: Ordinary Least Squares Estimation

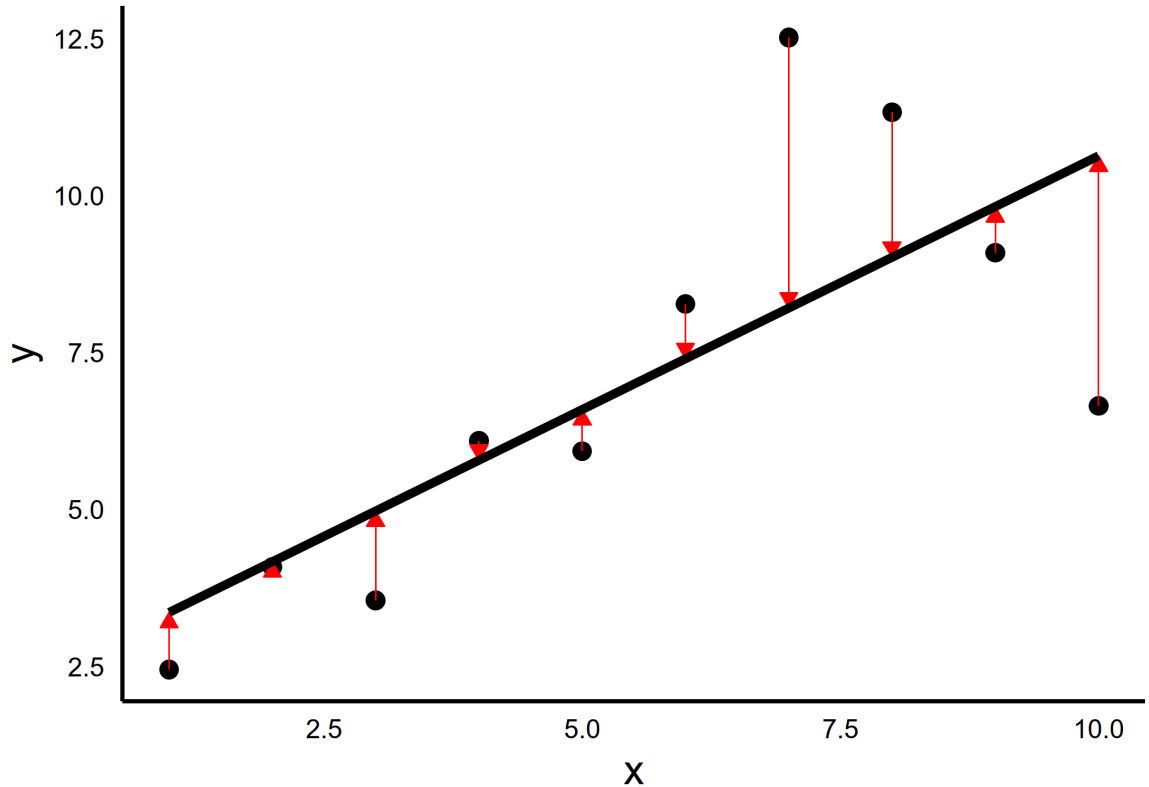
A Graphical Example



Source: Ben Elsner's slides

Review: Ordinary Least Squares Estimation

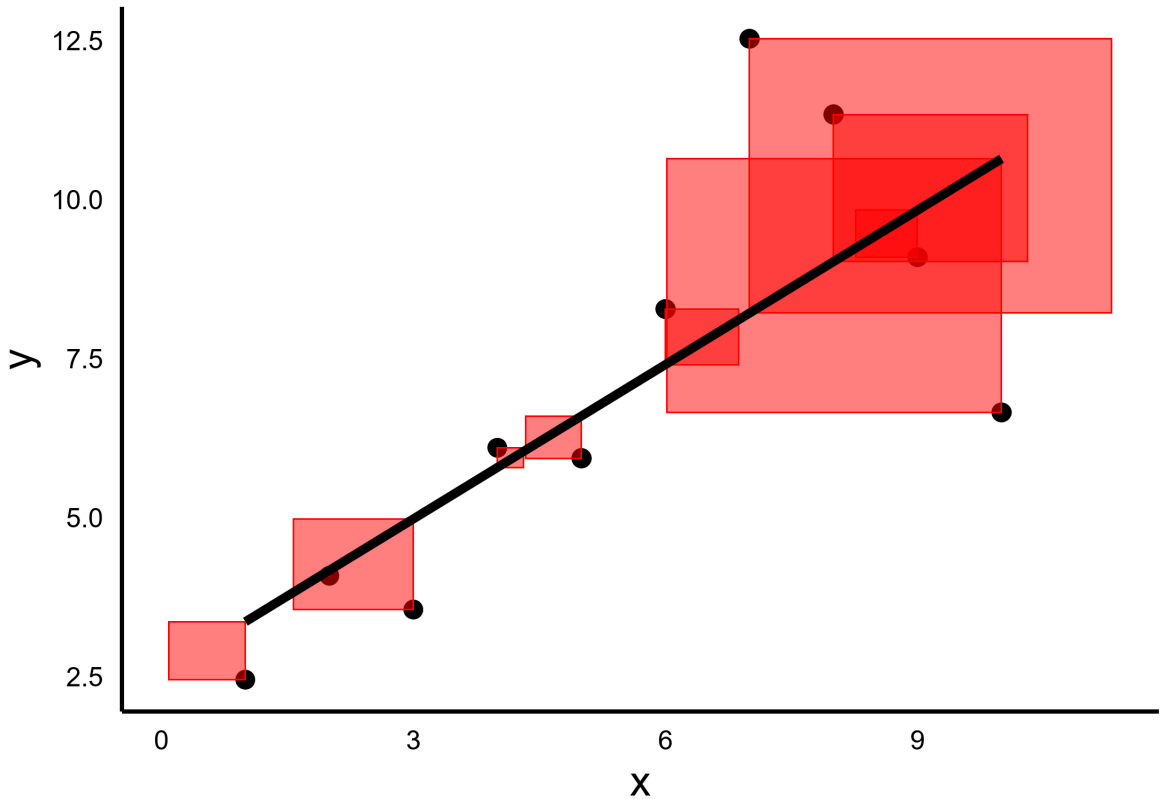
A Graphical Example



Source: Ben Elsner's slides

Review: Ordinary Least Squares Estimation

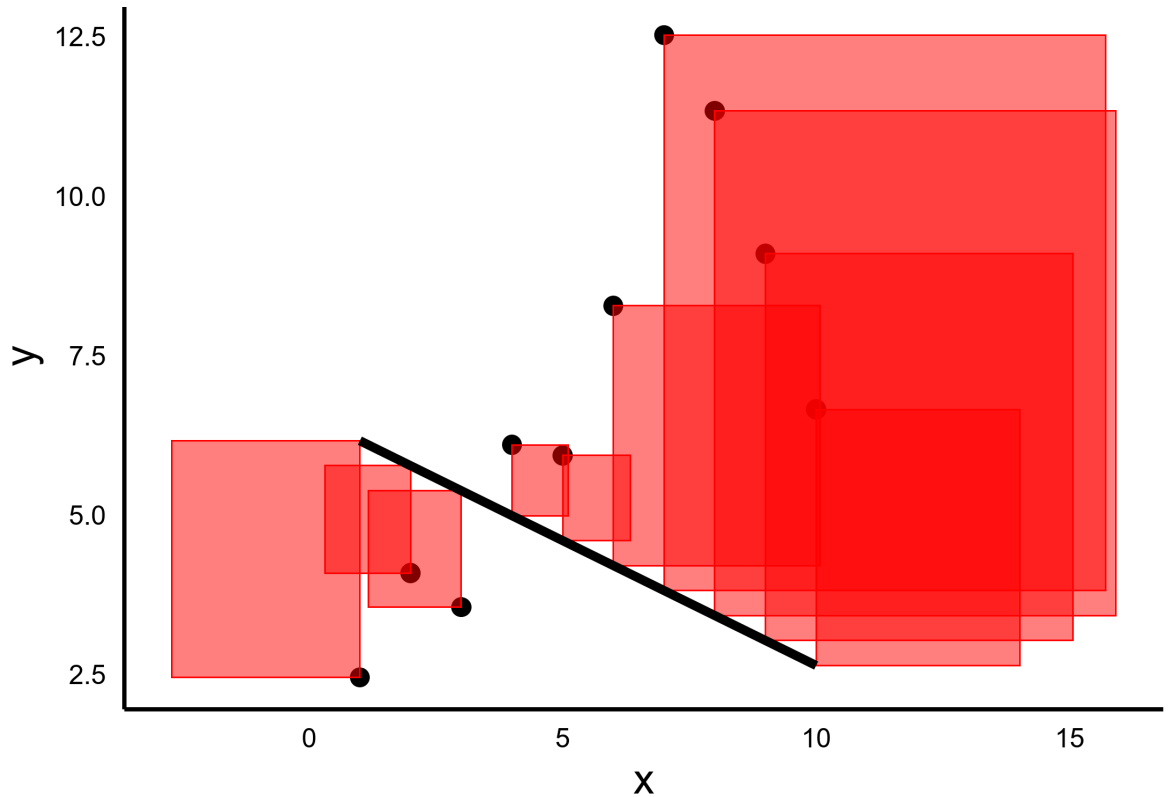
A Graphical Example



Source: Ben Elsner's slides

Review: Ordinary Least Squares Estimation

A Graphical Example



Source: Ben Elsner's slides

Review: Omitted Variable Bias

- OLS estimator for treatment effect α:

$$\hat{\alpha} = \frac{Cov(Y_i, D_i)}{V(D_i)}$$

- Failure to include enough (right) control variables in the regression would result in selection bias

- The **OLS version** of the **selection bias** generated by inadequate controls is called **Omitted Variable Bias (OVB)**

Review: Omitted Variable Bias

- Suppose the true model is:

$$Y_i = \delta + \alpha D_i + \beta X_i + \epsilon_i$$

- X_i is the observed characteristics (e.g. family wealth)
- But we estimate this model:

$$Y_i = \delta + \alpha D_i + u_i$$

  - where $u_i = \beta X_i + \epsilon_i$
  - Assume $\mathrm{E}[\epsilon_i | X_i] = 0$

Review: Omitted Variable Bias

- OVB formula:

$$\hat{\alpha} \xrightarrow{p} \alpha + \frac{Cov(u_i, D_i)}{V(D_i)}$$

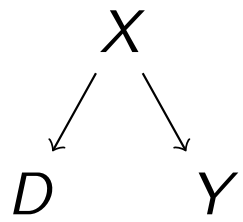
$$= \alpha + \boxed{\beta \frac{Cov(X_i, D_i)}{V(D_i)}} \quad \text{bias}$$

  - Covariance between X_i and D_i: $\text{Cov}(X_i, D_i) = \frac{1}{N} \sum_{i=1}^{N} (X_i - \bar{X})(D_i - \bar{D})$
  - Variance of D_i: $V(D_i) = \frac{1}{N} \sum_{i=1}^{N} (D_i - \bar{D})^2$

- The difference between estimated treatment effect $\hat{\alpha}$ and true effect α depends on two components:

1. β: The effect of omitted variable X_i on outcome variable Y_i
2. $\frac{Cov(X_i, D_i)}{V(D_i)}$: The relationship between omitted variable X_i and treatment variable D_i

Review: Omitted Variable Bias



- The confounding factor X can result in the co-movement between treatment D and outcome Y
- Even if treatment D has no causal effect on outcome Y

Review: Omitted Variable Bias

Example

- OVB formula:

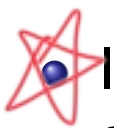
$$\hat{\alpha} \xrightarrow{p} \alpha + \frac{Cov(u_i, D_i)}{V(D_i)}$$
$$= \alpha + \beta \frac{Cov(X_i, D_i)}{V(D_i)}$$

- The difference between estimated effect of attending graduate school $\hat{\alpha}$ and true effect of attending graduate school α depends on two components:

1. β: The effect of family wealth (omitted) X_i on earnings Y_i
2. $\frac{Cov(X_i, D_i)}{V(D_i)}$: The relationship between family wealth X_i and attending graduate school D_i

Review: Omitted Variable Bias

- In RCT, we can eliminate OVB since treatment assignment D_i is unrelated to other confounding factors X_i
  - $\frac{Cov(X_i, D_i)}{V(D_i)} = 0$



- In the regression, we can eliminate OVB by including other **observed** confounding factors X_i into regression
  - $\frac{Cov(u_i, D_i)}{V(D_i)} = 0$
  - When we include X_i in regression model, $u_i = \epsilon_i$ which is unrelated to treatment status D_i

Review: Omitted Variable Bias

- OVB formula is a tool that allows us to consider the impact of controlling for variables we wish we had
- We cannot use data to check the consequences of omitted variables that we do not observe
- But we can use the OVB formula to make a educated guess as to the likely consequences of their omission

$$\hat{\alpha} \xrightarrow{p} \alpha + \beta \frac{Cov(X_i, D_i)}{V(D_i)}$$

Anatomy of Regression and Omitted Variable Bias

The Frisch-Waugh-Lovell Theorem

- The Frisch-Waugh-Lovell (FWL) Theorem states that three OLS estimators for α are equivalent:
  1. Regressing y on D and X
  2. Regressing y on $\tilde{D}$, the residuals from a regression of D on X
  3. Regressing $\tilde{Y}$ on $\tilde{D}$, with $\tilde{Y}$ being the residuals from a regression of Y on X

Anatomy of Regression and Omitted Variable Bias

The Frisch-Waugh-Lovell Theorem

- The main insight comes from point 2...
  - OVB arises when we exclude relevant variables that are correlated with both X and Y
  - By regressing D on X, we remove the part of D that is correlated with X
  - Leaving only the part of D that is uncorrelated with X in the residuals $\tilde{D}$
  - This allows us to estimate the true effect of D on Y without bias from omitted variables that are correlated with X

Anatomy of Regression and Omitted Variable Bias

The Frisch-Waugh-Lovell Theorem

- Multivariate regression does two steps simultaneously:
  1. Purges the correlation between D and X from D; the residuals $\tilde{D}$ are "what's left" and uncorrelated with X
  2. Estimates the effect of $\tilde{D}$ on Y, i.e., after purging the correlation between D and X
- Why is this important?
  - If X is a confounder, we can purge its influence by including it in the regression
  - The same holds for more than one control variable

Regression: Hypothesis Testing

Summary of Hypothesis Testing for Regression

- We estimate the following regression and want to test whether there is treatment effect:

$$Y_i = \delta + \alpha D_i + X_i\beta + \epsilon_i$$

1. Choose a null hypothesis:
    - We usually test whether there is **no average effect** of treatment
    - $H_0 : \alpha = 0$

Summary of Hypothesis Testing for Regression

2. Choose a test statistic
    - We use a t-statistic to measure whether our sample estimates support/against this null hypothesis
    - $t = \dfrac{(\hat{\alpha} - \alpha)}{\hat{\mathrm{SE}}(\hat{\alpha})}$

Summary of Hypothesis Testing for Regression

3. Estimate standard error of the estimator

- $\hat{\mathrm{SE}}(\hat{\alpha}) = \sqrt{\frac{\sum_{i=1}^{N} \hat{\epsilon}_i^2 \widetilde{D}_i^2}{\left(\sum_{i=1}^{N} \widetilde{D}_i^2\right)^2}}$ 考量異質變異數的標準差
  - $\hat{\epsilon}_i$ are the residuals from the main regression
  - $\widetilde{D}_i$ are the residuals obtained from regressing D_i on X_i
- The addition of covariates X has two opposing effects on $\hat{\mathrm{SE}}(\hat{\alpha})$.
  1. $\hat{\epsilon}_i$ might decrease since addition covariates explain some of the variation in Y_i
  2. $\widetilde{D}_i$ falls when covariates that predict D_i are added to the regressions
- This is known as **heteroskedasticity-robust standard errors**
  - Provide valid standard errors of estimator α even in the presence of heteroskedasticity (i.e., non-constant variance)

Summary of Hypothesis Testing for Regression

4. Evaluate whether the sample estimator is against null hypothesis or not
    - **Goal:** Calculate **p-value**
        - **p-value:** Given null hypothesis is true, the probability of obtaining the sample estimates or more extreme ones
        - If this probability is high, it means the sample estimate might support for null hypothesis
        - If this probability is low, it means the sample estimate might be against null hypothesis

Summary of Hypothesis Testing for Regression

4. Evaluate whether the sample estimator is against null hypothesis or not
    - In order to calculate this probability (p-value), we need to know the distribution of the t-statistic under the null hypothesis
        - If sample size is sufficiently large, using **Central Limit Theorem (CLT)**, t-statistic will have standard normal distribution

Summary of Hypothesis Testing for Regression

4. Evaluate whether the sample estimator is against null hypothesis or not
    - Based on standard normal distribution and sample estimator, we can get p-value
    - We reject the null hypothesis $H_0 : \alpha = 0$ when p-value is sufficiently low
        - We usually select an arbitrarily pre-defined threshold value θ, which is referred to as the **level of significance**
        - By convention, θ is commonly set to 0.1 or 0.05
    - If p-value is smaller than θ, we would say the sample estimate is **significantly different from the null hypothesis**

Interpretation of Regression Results

- We are only interested in α, the causal effect of treatment D on Y
  - The other coefficients $\beta_1, \beta_2, \ldots, \beta_k$ are NOT of interest
  - We include the covariates X to control for observed confounding factors
- Interpretation of α when controlling X
  - Holding all other variables X constant, a one unit increase in D leads to a α unit increase in Y

Interpretation of Regression Results

- Suppose the estimated regression is the following:

$$\hat{\mathrm{Y}}_{\mathrm{i}} = 35000 + 5000 D_i + 0.5 X_i$$

- Suppose the estimated standard error is:

$$\hat{\mathrm{SE}}(\hat{\alpha}) = 1000$$

- So the t-statistic for testing $H_0 : \alpha = 0$:

$$t = \frac{(\hat{\alpha} - \alpha)}{\hat{\mathrm{SE}}(\hat{\alpha})} = \frac{5000 - 0}{1000} = 5$$

Interpretation of Regression Results

- Using t-statistic, we can compute the p-value $= 0.00001$, which is much lower than 0.05 or 0.01
  - Given null hypothesis $H_0 : \alpha = 0$ is true, our estimate is unlikely to happen (but it happens!!)
  - It suggests our estimate is against the null hypothesis
  - Thus, we should reject the null hypothesis

Interpretation of Regression Results

- Based on sample estimates and its standard deviation, we can construct a confidence interval for α
- Note that the t-statistic for 5% two-sided significance level is 1.96

$$\hat{\alpha} \pm 1.96\hat{\mathrm{SE}}(\hat{\alpha}) = 5000 \pm 1.96 \times 1000$$

  - The 95% confidence interval does not include zero
  - Null hypothesis $H_0 : \alpha = 0$ is rejected at the 5% level

Regression – STATA Example

STATA Example

- See **reg.do**
- Use cps_2014_16.dta

Prepare Data for Estimation

```
1 gen college = educ99>= 15
2 replace incwage=. if incwage==9999999
3 drop if incwage==.
```

- **generate:** Create a binary variable `college` indicating if education level is college or above
- **replace:** Replace missing values in `incwage` with "." if `incwage` equals 9999999
- **drop:** Drop observations with missing values in `incwage`

Prepare Data for Estimation

```
forv i=1(1)5{
gen health_`i' = health==`i'
}
```

- **forvalues:** Loop through values 1 to 5 and create binary variables `health_1`, `health_2`, ..., `health_5` indicating if `health` equals the corresponding value
- **generate:** Create a new binary variable based on the condition `health==`i'`
- The loop generates 5 binary variables capturing different values of the `health` variable

STATA Command: reg

- **reg**: Linear regression
- Syntax:

```
reg depvar [indepvars] [if] [in] [weight] [,
    options]
```

Reducing OVB by including covariates

```
1 reg incwage college , vce(robust)
2 reg incwage college health_1 - health_4, vce(
      robust)
3 reg incwage college health_1 - health_4 age i.
      race, vce(robust)
```

- Regress `incwage` on `college` using robust standard errors
  - Add health indicator variables (`health_1` to `health_4`) to the regression
  - Further control for `age` and `race` (using indicator variables) in the regression
- Option **vce(robust)**: use robust standard errors

Reducing OVB by including covariates

Output

```
. reg incwage college health_1 - health_4 age i.race, vce(robust)

Linear regression                               Number of obs     =     46,299
                                                F(20, 46276)      =          .
                                                Prob > F          =          .
                                                R-squared         =     0.1106
                                                Root MSE          =      48789
```

incwage	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
college	32661.38	659.6271	49.51	0.000	31368.5	33954.26
health_1	25663.82	771.8364	33.25	0.000	24151.01	27176.63
health_2	24268.25	639.2598	37.96	0.000	23015.29	25521.21
health_3	18432.33	610.8693	30.17	0.000	17235.02	19629.65
health_4	7670.004	667.9725	11.48	0.000	6360.768	8979.241
age	89.56983	10.82239	8.28	0.000	68.35778	110.7819

Understanding the Frisch-Waugh-Lovell Theorem

```
1 reg college health_1 - health_4 age i.race, vce(
      robust)
2 predict college_rid, residuals
3
4 reg incwage college health_1 - health_4 age i.
      race, vce(robust)
5 reg incwage college_rid, vce(robust)
```

- Regress `college` on all other covariates to obtain residuals `college_rid`
  - `college_rid` represents the part of `college` that is unrelated to other covariates
- Regress `incwage` on `college` and other covariates
- Regress `incwage` on `college_rid` gives same coefficient as previous regression

Understanding the Frisch-Waugh-Lovell Theorem

Output

```
. reg incwage college health_1 - health_4 age i.race, vce(robust)

.inear regression                               Number of obs     =     46,299
                                                F(20, 46276)      =          .
                                                Prob > F          =          .
                                                R-squared         =     0.1106
                                                Root MSE          =      48789
```

incwage	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
college	32661.38	659.6271	49.51	0.000	31368.5	33954.26
health_1	25663.82	771.8364	33.25	0.000	24151.01	27176.63
health_2	24268.25	639.2598	37.96	0.000	23015.29	25521.21
health_3	18432.33	610.8693	30.17	0.000	17235.02	19629.65
health_4	7670.004	667.9725	11.48	0.000	6360.768	8979.241
age	89.56983	10.82239	8.28	0.000	68.35778	110.7819

Understanding the Frisch-Waugh-Lovell Theorem

Output

```
. reg incwage college_rid, vce(robust)

.inear regression                               Number of obs     =     46,299
                                                F(1, 46297)       =    2381.93
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0772
                                                Root MSE          =      49684
```

incwage	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
college_rid	32661.38	669.2215	48.81	0.000	31349.69	33973.06
_cons	29208.86	230.9059	126.50	0.000	28756.28	29661.44

Subgroup Analysis

限定子群體

```
1 reg incwage college i.health age year i.race if
      sex==1, vce(robust)
2 predict incwage_hat_m if e(sample)
```

- Option **if**: restrict sample to specific subgroup
- Option **if e(sample)**: obtain linear prediction for male (if $sex == 1$)

Subgroup Analysis

Output

```
. reg incwage college health_1 - health_4 age i.race if sex==1, vce(robust)

Linear regression                               Number of obs     =     22,173
                                                F(17, 22150)      =          .
                                                Prob > F          =          .
                                                R-squared         =     0.1303
                                                Root MSE          =      58414
```

incwage	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
college	43138.43	1208.246	35.70	0.000	40770.18	45506.68
health_1	34501.94	1382.755	24.95	0.000	31791.64	37212.24
health_2	31922.23	1145.355	27.87	0.000	29677.25	34167.21
health_3	24665.21	1109.318	22.23	0.000	22490.86	26839.55
health_4	11095.74	1286.766	8.62	0.000	8573.588	13617.89
age	170.9962	19.49211	8.77	0.000	132.7903	209.2021

Suggested Readings

- Chapter 2, Mastering Metrics: The Path from Cause to Effect
- Chapter 3, Mostly Harmless Econometrics
- Chapter 2, Causal Inference: The Mixtape