Lecture 4: Bias in Linear Regressions

Jaakko Meriläinen

5304 Econometrics @ Stockholm School of Economics

Introduction

- In previous lectures, we have discussed the following:
  - Selection problems and the experimental ideal
  - Derivation of the OLS estimator (and how it is unbiased under the MLR assumptions)
  - How to conduct inference on the OLS estimators (going from the sample to population objects we care about)

- This lecture focuses on cases where the ZCM assumption does not hold

Plan for Today

1. Introduction
2. Bias in linear regression
   - Omitted variables
   - Directed acyclic graphs
   - Bad controls
   - Measurement error in OLS
   - Simultaneity and reverse causation
3. Selection on observables vs. unobservables

Biased Coefficients

- Recall that our focus is on estimating causal effects
- Our ability to interpret regression coefficients causally comes from assuming $E(u) = E(u|\mathbf{x})$
  - Otherwise, the estimates from OLS will be biased
  - Probing whether it is plausible is the core of much applied work
- In this lecture, we will focus on four types of violations:
  - Omitted variable bias
  - Bias due to measurement error
  - Bad controls
  - Simultaneity bias

Omitted Variables

- Perhaps the most common concern with modeling the relationships that we estimate is one of omitted variables:
  - The econometrician does not observe everything about the world!
  - Hence the error term u which captures things other than our covariates which affect the outcome
  - We have until now assumed that $E(u) = E(u|\mathbf{x})$, i.e, the errors do not depend on x

- But if an omitted variable z is correlated with $\mathbf{x}$ and determines y, we are in trouble...

Omitted Variable Bias: The Simple Case

- Let us first think of the two variable case - and let the true model be

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

- If we observed both variables, we could easily estimate $\widehat{\beta}_0$, $\widehat{\beta}_1$, $\widehat{\beta}_2$

- Now suppose we only have data on x_1

- What happens if we ignore x_2?

Omitted Variable Bias: The Simple Case

- Regressing y on x_1 gives us:

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$$

- It could be that $\tilde{\beta}_1 \neq \hat{\beta}_1$!

- Imagine a regression of x_2 on x_1 such that:

$$\tilde{x_2} = \tilde{\delta}_0 + \tilde{\delta}_1 x_1$$

- It can be shown that $E(\tilde{\beta}_1) = E(\hat{\beta}_1 + \hat{\beta}_2 \tilde{\delta}_1)$

Omitted Variable Bias: The Simple Case

$$E(\tilde{\beta_1}) = E(\widehat{\beta_1}) + E(\widehat{\beta_2})\tilde{\delta_1} = \beta_1 + \beta_2\tilde{\delta_1}$$

- In general, then, we see that $E(\tilde{\beta_1}) \neq \beta_1$
- Indeed $E(\tilde{\beta_1}) = \beta_1$ in only two special cases:
  - Either $\beta_2 = 0$ - x_2 does not have a direct relationship with y
  - Or $\tilde{\delta_1} = 0$ - x_1 and x_2 are uncorrelated

Omitted Variable Bias: The Simple Case

- This expression

$$E(\tilde{\beta_1}) = \beta_1 + \beta_2 \tilde{\delta_1}$$

allows us to put a sign to the bias in our estimate of β (given by $E(\tilde{\beta_1}) - E(\widehat{\beta_1})$)

- Positive (upwards) bias if $E(\tilde{\beta_1}) > E(\widehat{\beta_1})$
- Negative (downwards) bias if $E(\tilde{\beta_1}) < E(\widehat{\beta_1})$

	$corr(x_1, x_2) > 0$	$corr(x_1, x_2) < 0$
$\beta_2 > 0$	positive bias	negative bias
$\beta_2 < 0$	negative bias	positive bias

An Example

- Suppose we were interested in the effect of poverty on health and estimate the following regression:

$$health_i = \alpha + \beta_1 poor_i + u$$

- What are omitted variables which might lurk in u?
- Will they be positively or negatively correlated with being poor ($poor_i$)?
- Will it have a positive or negative direct effect on health status ($health_i$)?

In Case You Think This Is Obvious... (forbes.com)

CORONAVIRUS | 130,196 views | Jun 6, 2020, 11:26am EDT

Bald Men At Higher Risk Of Severe Coronavirus Symptoms

Marla Milling Contributor ⓘ

Healthcare

I am a Forbes.com Contributor specializing in geriatric health and women's health articles.

Omitted Variable Correlated with Both the LHS and the RHS...

Updated (6/8/20) This piece has been clarified to note that the study did not control for age, which is a risk factor for hair loss and severe Covid-19.

New research is showing why a larger percentage of men—particularly bald men—are

Omitted Variable in the General Case

- In the general case, let the true model be:

$$E(y|x_1, x_2, ..., x_k, q) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \gamma q$$

where $x_1, x_2, ..., x_k$ are variables we observe while q is omitted

- Write the above equation in error form as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$$

$u \equiv v + \gamma q$, where v is the **structural error**

Omitted Variable in the General Case

- If q is correlated with any x_j, u is correlated as well $\Rightarrow$ We cannot consistently estimate any β_j for sure
- Write the **linear projection** of q on $\mathbf{x}$ as:

$$q = \delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k + r$$

- Plugging this into previous equation, we can rewrite as:

$$y = (\beta_0 + \gamma\delta_0) + (\beta_1 + \gamma\delta_1)x_1 + \cdots + (\beta_k + \gamma\delta_k)x_k + v + \gamma r$$

Omitted Variable in the General Case

$$y = (\beta_0 + \gamma\delta_0) + (\beta_1 + \gamma\delta_1)x_1 + \cdots + (\beta_k + \gamma\delta_k)x_k + v + \gamma r$$

- We can then just read off that $plim\, \widehat{\beta_k} = \beta_k + \gamma\delta_j$ as in the simple case
- The bias does not go away as $n \to \infty$
- If all other δ_j are 0 except for δ_k, then $plim\, \widehat{\beta_k} = \beta_k + \gamma[Cov(x_k, q)/Var(x_k)]$

(Cross-Sectional) OLS Solutions for OVB

- **Measure q so that you can include it directly**
  - Easier said than done
  - If we could have measured it, of course we should!
  - But it can be very hard to argue that everything determining y is controlled for
- **Randomize!**
  - Removes the correlation with x_j by design
  - Again, if possible, then obviously we should!
  - But many times you just cannot!
- **Use a proxy variable**

Proxy Variables

- Proxy variable solution: find an alternative measure (z) which "proxies" for q such that it does not otherwise determine y directly
- Conditioning on z removes partial correlation between $\mathbf{x}$ and q
- For example, imagine if you were missing innate ability in a wage equation but had IQ scores available as a proxy

Example: Washington (2008)

- Research question: How does having daughters affect their legislator parents' voting behavior in the U.S.?
- Sociologists have shown that child gender affects parental support for feminist policies—does
- Does this also show up in Congress members' voting behavior?
- Gender of a child is essentially random $\Rightarrow$ Comparing two legislators with children of different gender is as good as a real randomized experiment
- But...

Example: Washington (2008)

While there are no natural methods that will affect the probability that any one child is a boy, a couple with male preferences could follow a natural method to ensure a certain number of sons. For example, a couple could follow a stopping rule in which they continue having children until a son is born. While such a stopping rule can never alter the proportion of female children in the overall population, it will have within family consequences for the ratio of daughters to sons.

- Families following such a stopping rule could be different from others
- These differences could also show up in legislators' policy preferences and voting behavior
- Solution: conditional on total number of children, the number of female children appears to be random

Example: Washington (2008)

TABLE 2—IMPACT OF FEMALE CHILDREN ON LEGISLATOR VOTING ON WOMEN'S ISSUES

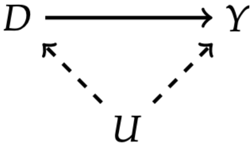
	NOW	AAUW			
	105th (1)	105th (2)	106th (3)	107th (4)	108th (5)
Number of female children	2.3**	2.38**	1.69	2.42**	2.25**
	(1.04)	(1.12)	(1.14)	(1.09)	(1.15)
Other legislator characteristics					
Female	10.83***	9.19***	10.44***	7.56***	6.91**
	(2.69)	(2.91)	(2.88)	(2.62)	(2.73)
White	1.86	0.14	2.59	−2.63	1.94
	(3.45)	(3.68)	(3.83)	(3.15)	(3.21)
Republican	−44.9***	−60.47***	−55.93***	−63.22***	−63.93***
	(2.11)	(2.28)	(2.34)	(2.12)	(2.44)
Age	0.66	0.85	2.03**	1.3	2.3***
	(0.80)	(.86)	(0.9)	(0.8)	(0.86)
Age squared	−0.01	−0.01	−0.02**	−0.01	−0.02***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Service length	0.24	−0.21	−0.73*	−0.1	−0.14
	(0.30)	(0.32)	(0.38)	(0.35)	(0.33)
Service length squared	−0.01	0.00	0.02*	−0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
No religion	7.26	5.67	5.35	7.03	−7.14
	(7.02)	(7.61)	(7.79)	(7.18)	(7.5)
Catholic	−3.97**	−4.5**	−2.28	−4.02**	−5.47***
	(1.94)	(2.09)	(2.13)	(1.99)	(2.08)
Other Christian	0.77	3.2	1.69	1.65	3.87
	(4.60)	(4.98)	(4.91)	(4.49)	(4.68)
Other religion[a]	10.87**	9.68**	11.89***	10.29***	3.16
	(3.75)	(4.05)	(4.34)	(3.79)	(3.96)
Democratic vote share in district	84.16***	62.15***	57.44***	56.21***	66.95***
(most recent presidential election)	(10.87)	(11.57)	(12.02)	(9.09)	(10.89)
N[b]	430	434	434	434	433

Note: All specifications include region and number of children fixed effects. Standard errors in parentheses.
* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.
[a] The omitted religious category is Protestant.
[b] Sample size varies due to missing child gender and voting score information.

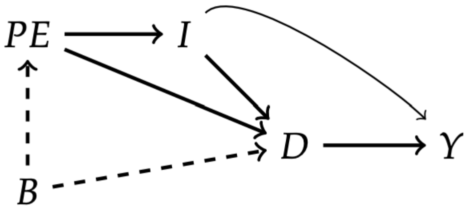
Directed Acyclic Graphs

- Some people find it useful to present the data generating process using *directed acycilc graphs* (DAGs)
- Random variables are represented with letters and causal relationships are represented with arrows
- Let us consider some examples from Cunningham (2021, Ch. 3)

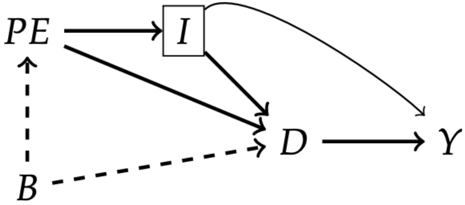
A Confounding Variable



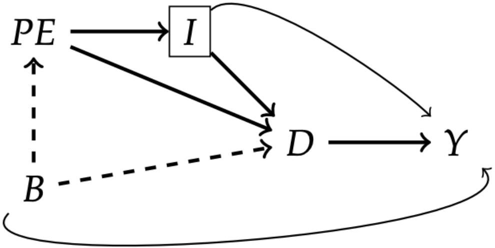
The Effect of College (D) on Income (Y)



Conditioning on Parental Income Solves the Confounding Problem...



...But Only If You Believe the DAG!



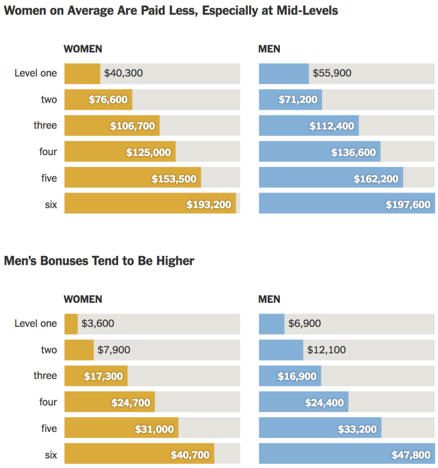
Bad Controls

- We discussed how having more controls could address potential bias
- **Important:** Not all controls are reasonable to include
- Deciding which are requires judgment!
- Angrist and Pischke (2009) discuss this as the "bad control" problem

Bad Controls: What Are They?

- One clear case of a bad control is including channels through which any treatment effect is mediated
- The classic illustration is again in the context of a wage regression:
  - Suppose we are looking at the effect of education on wages
  - We decide the condition on occupation dummies
  - But if education affects your later occupation choice directly, this is a bad measure to include!
- More generally, measures that could have been outcomes should typically not be included
  - This is true even for randomized trials
  - You can add baseline variables easily in that case, not variables after treatment was assigned!

Google Pays Men More (A Case Study That Remains Relevant...)



Source: New York Times (September 8, 2017)

Is This Discrimination?

> Three former Google employees have filed a class-action lawsuit against the company for allegedly discriminating against women. The complaint claims that Google systematically pays women less than men performing similar jobs, promotes men more often than similarly qualified women, and keeps women in lower-paying and lower-level positions.

Source: Gizmodo.com (September 14, 2017)

Gender Bias at Google? The Google Response

After the *New York Times* detailed the employee spreadsheets on Friday, Google spokesperson Gina Scigliano told Gizmodo that its own data shows, when you take "location, tenure, job role, level and performance" into account, that "women are paid 99.7% of what men are paid at Google." Scigliano described the *Times* story as "extremely flawed."

Source: Gizmodo.com (September 14, 2017)

The Labor Economist's Response

Sally Hudson
@SallyLHudson

Follow

Dear Google,

Occupation controls are literally the textbook example of how not to measure wage discrimination.

Sincerely,
Labor Economists

After the *New York Times* detailed the employee spreadsheets on Friday, Google spokesperson Gina Scigliano told Gizmodo that its own data shows, when you take "location, tenure, job role, level and performance" into account, that "women are paid 99.7% of what men are paid at Google." Scigliano described the *Times* story as "extremely flawed."

are Bad Control

aps more controls are the answer. Why not control for occupation f
e? Many data sets that report earnings also classify workers' jobs, s
her they're employed as a manager or laborer. Surely occupations are
ictors of both schooling and earnings, possibly capturing traits that
Mick and Johan from more average Joes. By the logic of OVB, therefo
ld control for occupation, a matter easily accomplished by including d
bles to indicate the types of jobs held.
spite of the fact that occupation is strongly correlated with both sch
wages, occupation dummies are *bad controls* in a wage equation. Th
Master Joshway works today as a professor and not a busboy (as h
is in part a reward for his extravagant schooling. It's a mistake to elir
benefit from our calculation by comparing only professors or busboys
npting to quantify the economic value of schooling. Even in a world wh
ssors earn a uniform one million dollars a year (may it soon come to
all busboys earn a uniform \$10,000, an experiment that randomly a
oling would show that schooling raises wages. The channel by which
ncreased in this notional experiment is the shift from lowly busboyh
ted professorness.

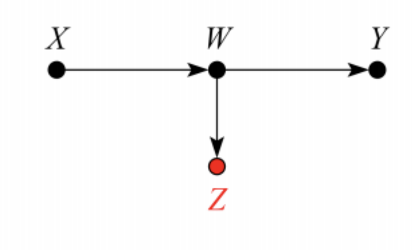
7:28 PM - 15 Sep 2017

Examples of Bad Controls (If Trying to Estimate the Effect of X on Y)

- Let us consider some examples from Cinelli, Forney and Pearl (2021)
- Suppose Z is a mediator of the causal effect of X on Y
- Controlling for Z will block the very effect we want to estimate, thus biasing our estimates ("overcontrol bias")

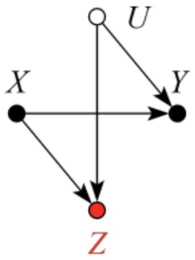
Examples of Bad Controls

- Z is not itself a mediator of the causal effect of X on Y, controlling for Z is equivalent to partially controlling for the mediator M
- This would similarly bias our estimates



Examples of Bad Controls

- Adjusting for Z leads to "collider stratification bias"



Measurement Error

- We typically estimate regressions on data that are **measured with error**
  - Errors in surveys
  - Data coding errors
  - The imperfect relationship between the proxy collected in the survey and the “true” value of the variable we want

- We will now examine what happens when data are noisy measures of true values

Measurement Error in Regressors

- Let us start in the simplest one-regressor case:

$$y = \beta_0 + \beta_1 x_1^* + u$$

- Assume this satisfies the first 4 Gauss-Markov assumptions.
- We do not observe x_1^* but observe a noisy measure $x_1 = x_1^* + e_1$
- For simplicity, $E(e_1) = 0$
- Assume $E(y|x_1^*, x_1) = E(y|x_1^*)$, i.e., e is just noise/not correlated with u

Classical Error-in-Variables Problem

- Assume that the error is uncorrelated with the true value x^*:

$$Cov(x_1^*, e_1) = 0$$

- This necessarily implies x_1 and e_1 are correlated

$$Cov(x_1, e_1) = E(x_1 e_1) = E(x_1^* e_1) + E(e_1^2) = \sigma_{e1}^2$$

- This correlation between x_1 and e_1 is problematic

- The DGP can be written as:

$$y = \beta_0 + \beta_1 x_1 + (\beta_1 e_1 + u)$$

- It is clear that the error term of the regression and x_1 are correlated

Classical Error-in-Variables Problem

- Under the assumptions above, we can show that:

$$
\begin{aligned}
plim(\widehat{\beta_1}) &= \beta_1 + \frac{Cov(x_1, u - \beta_1 e_1)}{Var(x_1)} \\
&= \beta_1 - \frac{\beta_1 \sigma^2_{e_1}}{\sigma^2_{x_1^*} + \sigma^2_{e_1}} \\
&= \beta_1 \left(\frac{\sigma^2_{x_1^*}}{\sigma^2_{x_1^*} + \sigma^2_{e_1}} \right)
\end{aligned}
$$

- Note that the multiplicative term, $(\frac{\sigma^2_{x_1^*}}{\sigma^2_{x_1^*} + \sigma^2_{e_1}})$, is always less than 1
- This implies that $\widehat{\beta_1}$ is always closer to 0 than the true β_1
- This is called **attenuation bias**

Potential Solutions to Measurement Error

- We are thinking of classical measurement error here—not strategic mis-reporting etc.

- The first-best solution is to collect better data
  - If σ_{e_1} is very small in comparison to $\sigma^2_{x_1^*}$ i.e. the noise is small compared to the signal, the bias may be negligible

- Sometimes this comes from better data sources—for instance:
  - Consumption: might be more accurate from supermarket scanner data than by asking survey respondents to recall
  - Earnings data: might be more reliable from tax records than survey reports

- More generally, as we will see in our lectures on instrumental variables (IV), having multiple measures for the same variable can be very advantageous

Measurement Error in the Dependent Variable

- In contrast, **classical** measurement error in the dependent variable is not a problem
- Specifically, let true DGP be

$$y^* = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$$

- We observe $y = y^* + e$

Measurement Error in the Dependent Variable

- If $E(e) = E(e|\mathbf{x}) = 0$, then OLS regression of y on x will give consistent estimates of the parameters
- To see this, note that DGP can be written as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + (u + e)$$

- The error term $(u + e)$ is still uncorrelated with $\mathbf{x}$
- All OLS results on unbiasedness go through!

Simultaneity

- Our discussion so far has always focused on single-equation models
- But sometimes, the true model we have in mind is a system of equations
- The classic example relates to the demand and supply

$$
\begin{aligned}
Q_s &= \alpha_1 p_s + \beta_1 z_s + u_1 \\
Q_d &= \alpha_2 p_d + \beta_2 z_d + u_2
\end{aligned}
$$

- Assuming markets clear, we only observe equilibrium objects (p^*, Q^*)
- So, without stronger assumptions, we cannot identify the demand and supply curves separately

Reverse Causation

- A particular form of simultaneity is **reverse causation**
- We are modelling x as determining y, but what if the true relationship is the other way around?
- Suppose we want to estimate the effect of policing on crime using the following model

$$Murders\ per\ capita = \alpha + \beta_1 Police\ per\ capita + \beta_2 X + u_1$$

- But policing is itself determined by the rate of crime in a neighborhood!

$$Police\ per\ capita = \gamma_0 + \gamma_1 Murders\ per\ capita + \gamma_2 Z + u_2$$

- In general, we cannot identify the effect of policing on crime consistently, but as we will see in Lecture 6, if X and Z are different, we may have options

Conditional Independence Assumption in Potential Outcomes Framework

- What does all of this mean in the potential outcomes framework?
- In binary treatment case, to get a causal treatment effects, we need potential outcomes (Y_{0i}, Y_{1i}) to be independent of the treatment variable (D_i), conditional on other variables X
- This is stated as $\{Y_{0i}, Y_{1i}\} \perp D_i | X$ in the binary treatment case
- Effectively means we are assuming "selection-on-observables"
  - Sometimes referred to as (conditional) ignorability of treatment
  - Conditional on covariates, we are comparing apples-to-apples

Example: Assessing the Effect of Selective Colleges

- Does going to a selective college raise wages?
- Labor economists care about returns to different types of education
- One particular question is the returns to attending a more vs. less selective college
- For instance, what would happen if you went to Harvard vs. Boston University?
- In the US, these questions are particularly prominent since college education is expensive and college choices are perceived to be very important for future outcomes
- Dale and Krueger (2002) address this question in a matching plus regression framework

The Selection Problem

- Students who go to Harvard may not be comparable to those going to, e.g., Michigan State
- Suppose that college j admits student i if and only if

$$\underbrace{Z_{ij}}_{\textit{application quality}} = \gamma_1 \underbrace{X_{1i}}_{\substack{\textit{observed} \\ (e.g.\ SAT_i)}} + \gamma_2 \underbrace{X_{2i}}_{\substack{\textit{unobserved} \\ (e.g.\ motivation)}} + \underbrace{e_{ij}}_{\textit{luck}} > C_j$$

  Here X_{2i} is unobserved only for researchers—the admissions committee may see it based on personal statements, essays, recommendation letters etc.

The Selection Problem

- Suppose wages are determined as follows

$$lnW_i = \beta_0 + \beta_1 \underbrace{SAT_{j*}}_{avg\ SAT\ score} + \beta_2 X_{1i} + \beta_3 X_{2i} + \epsilon_i$$

- But because researchers do not observe X_{2i}, they can only estimate

$$lnW_i = \beta_0^{'} + \beta_1^{'} SAT_{j*} + \beta_2^{'} X_{1i} + u_i$$

- Since SAT_{j*} is positively correlated with X_{2i}, the estimated coefficient $\beta_1^{'}$ will be biased upwards

Identification Strategy

- Dale and Kruger observe the full set of schools students applied to, the ones they were accepted to, and the ones they attended
- One identification strategy: controlling for admissions selectivity controls
  - Divide students into groups based on the schools that admitted them
  - Control for dummy variables for each of these groups
  - Conditional on acceptance, some students should have gone to less-selective colleges
  - Unbiased if the factors affecting which school you attend do not enter the wage equation directly

How the Matched Groups Were Created

TABLE I
ILLUSTRATION OF HOW MATCHED-APPLICANT GROUPS WERE CONSTRUCTED

		Student applications to college							
		Application 1		Application 2		Application 3		Application 4	
Student	Matched-applicant group	School average SAT	School admissions decision	School average SAT	School admissions decision	School average SAT	School admissions decision	School average SAT	School admissions decision
Student A	1	1280	Reject	1226	Accept*	1215	Accept	na	na
Student B	1	1280	Reject	1226	Accept	1215	Accept*	na	na
Student C	2	1360	Accept	1310	Reject	1270	Accept*	1155	Accept
Student D	2	1355	Accept	1316	Reject	1270	Accept*	1160	Accept
Student E	2	1370	Accept*	1316	Reject	1260	Accept	1150	Accept
Student F	Excluded	1180	Accept*	na	na	na	na	na	na
Student G	Excluded	1180	Accept*	na	na	na	na	na	na
Student H	3	1360	Accept	1308	Accept*	1260	Accept	1160	Accept
Student I	3	1370	Accept*	1311	Accept	1255	Accept	1155	Accept
Student J	3	1350	Accept	1316	Accept*	1265	Accept	1155	Accept
Student K	4	1245	Reject	1217	Reject	1180	Accept*	na	na
Student L	4	1235	Reject	1209	Reject	1180	Accept*	na	na
Student M	5	1140	Accept	1055	Accept*	na	na	na	na
Student N	5	1145	Accept*	1060	Accept	na	na	na	na
Student O	No match	1370	Reject	1038	Accept*	na	na	na	na

An Alternative Strategy: Self-Revelation

- Assume that students know their latent quality (Z) best and tailor their applications accordingly
- Idea: if I know myself to be "high quality", I apply to more selective schools
- Suppose the following:
  - The error term u_i can be modeled as $u_i = \tau_0 + \tau_1 AVG_i + v_i$
  - AVG is the average SAT score of the schools I have applied to and
  - v_i is uncorrelated with the SAT score of the school the student attended
- Then including AVG_i will solve the selection problem

Results

TABLE III

LOG EARNINGS REGRESSIONS USING COLLEGE AND BEYOND SURVEY, SAMPLE OF MALE AND FEMALE FULL-TIME WORKERS

	Model					
	Basic model: no selection controls		Matched-applicant model	Alternative matched-applicant models		Self-revelation model
	Full sample	Restricted sample	Similar school-SAT matches*	Exact school-SAT matches**	*Barron's* matches***	
Variable	1	2	3	4	5	6
School-average SAT score/100	0.076	0.082	−0.016	−0.106	0.004	−0.001
	(0.016)	(0.014)	(0.022)	(0.036)	(0.016)	(0.018)
Predicted log(parental income)	0.187	0.190	0.163	0.232	0.154	0.161
	(0.024)	(0.033)	(0.033)	(0.079)	(0.028)	(0.025)
Own SAT score/100	0.018	0.006	−0.011	0.003	−0.005	0.009
	(0.006)	(0.007)	(0.007)	(0.014)	(0.005)	(0.006)
Female	−0.403	−0.410	−0.395	−0.476	−0.400	−0.396
	(0.015)	(0.018)	(0.024)	(0.049)	(0.017)	(0.014)
Black	−0.023	−0.026	−0.057	−0.028	−0.057	−0.034
	(0.035)	(0.053)	(0.053)	(0.049)	(0.039)	(0.035)
Hispanic	0.015	0.070	0.020	−0.248	0.036	0.007
	(0.052)	(0.076)	(0.099)	(0.206)	(0.066)	(0.053)
Asian	0.173	0.245	0.241	0.368	0.163	0.155
	(0.036)	(0.054)	(0.064)	(0.141)	(0.049)	(0.037)

Recap

1. Introduction
2. Bias in linear regression
   - Omitted variables
   - Directed acyclic graphs
   - Bad controls
   - Measurement error in OLS
   - Simultaneity and reverse causation
3. Selection on observables vs. unobservables

Readings

- Causal Inference: The Mixtape - the chapter on DAGs
- Video: Angrist (2017) AEA Continuing Education Lecture 2 (click here for link)
- Video: Angrist (2017) AEA Continuing Education Lecture 3 (roughly the first 37 minutes) (click here for link)
- Angrist and Pischke (2009) - Section 3.2
- Wooldridge (2013) - Chapter 9

References

- Cinelli, C., Forney, A., & Pearl, J. (2021). A crash course in good and bad controls. Forthcoming in Sociological Methods and Research.
- Dale, S. B., & Krueger, A. B. (2002). Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. Quarterly Journal of Economics, 117(4), 1491-1527.
- Washington, E. (2008). Female Socialization: How Daughters Affect Their Legislator Fathers. American Economic Review 98(1), 311-332.