Lecture 2: Causal Inference and Randomized Controlled Trials

Rigissa Megalokonomou
University of Queensland

Reading for lecture 2

- In **Wooldridge 2013**,
  - Sections 3.2, 3.3 and 3.4 in chapter 3
  - Sections 6.2 in chapter 6
  - Sections 8.1 and 8.2 in chapter 8
- Introduction to randomized controlled experiments
  - Burtless, Gary (1995), The Case for Randomized Field Trials in Economic and Policy Research, Journal of Economic Perspectives, Vol. 9, No. 2, pp. 63-84.
  - Duflo, Esther and Abhijit V. Banerjee (2009), The Experimental Approach to Development Economics, Annual Review of Economics, Vol. 1, pp. 151-178.
- **Chapters 2 and 3 in "Mostly Harmless Econometrics** - An Empiricist's Companion" by Angrist and Psischke (2009)

Our objective: Obtain causal effect of policy change

- Causal effects give answers to **'what if'** questions:
  - For instance, what would happen to smoking if cigarette taxes were raised?
  - Economic theory is to tell us that the demand for cigarettes is likely to fall when their prices rise.
  - But the government wants to know the answer to the question **by how much**:

$$\frac{\partial E(y|\mathbf{x})}{\partial x_1}$$

  - where ∂x_1 in the increase in cigarette prices, $\mathbf{x} = (x_1, x_2, ..., x_k)$.
- The interest is in the effect of a tax on the demand for cigarettes as mentioned above.

Our objective: Obtain causal effect of policy change

- Other examples:
  - What would happen to students' performance (y) if class sizes (x_1) were reduced?
    Policy Change: Programs that reduce class sizes
  - What would happen if people living in 'bad' neighborhoods were given the opportunity to move to better ones? e.g. Health or income (y) and neighborhood quality (x_1).
    Policy Change: Programs that re-shuffle people depending on socio-economic background.
  - What would happen to workers' productivity (y) if they spend a long period unemployed (x_1).
    Policy Change: Programs that train workers or unemployment insurance.

How can we get an estimate of the causal effect?

- How to estimate causal effects?
  - Want to estimate the effect of x_1 on distribution of y, other relevant things being held constant.
  - Most common to be interested in effect on mean of y, i.e.:

$$\frac{\partial E(y|x_1, others)}{\partial x_1}$$

- This will be equal to the causal effect of β_1 where *others* are held fixed.
- It turns out that a **regression** is the most obvious way to estimate causal effects.
- In practice, data come in the form of samples and rarely consist of the entire population.

- Assumptions we made last week:
- (A1) The model is linear.
- (A2) Random sample of (y, x)
- (A3) None of the Xs is constant and there is no perfect multicollinearity.
- (A4) Zero conditional mean of errors (exogeneity).
- (A5) Constant variance of errors (homoskedasticity)
- (A6) Normality of error term.

We need these assumptions to hold for our β estimate to have the desired properties.

Potential problems of using regression models

- A regression tells us about correlation (association), but 'correlation is not causation' without *proper* **conditioning variables**.
- General **problems** when we estimate causal effects using linear regression are:
  - Omitted Variables;
  - Reverse Causality;
  - Measurement Error;
  - Sample selection;
  - Problems due to **omitted variables** and **sample selection** are receiving particular attention in microeconomics since these problems are very common in applications. These problems can be framed as violations of (A4) in lecture 1 (namely, exogeneity).
- Regressions should account for these problems to give causal effects.

A violation of A4: Omitted Variable Problem

- Framing the omitted variable problem...
- A model is assumed to be linear in $(\mathbf{x}, \mathbf{w})$, and a researcher wants to estimate the parameters γ and β. The regression is:

$$y = \mathbf{x}\beta + \mathbf{w}\gamma + u_1$$

(Hint: y could be wages, x education and w ability)

- However, you estimate (e.g. since we don't observe **w** so it is missing/omitted)

$$y = \mathbf{x}b + u_2$$

so w becomes part of the error term, thus $u_2 = \mathbf{w}\gamma + u_1$

- Then:

$$E(\widehat{b}|\mathbf{x}, \mathbf{w}) = \beta + \gamma \frac{Cov(x, w)}{Var(x)} \quad (1)$$

(Hint: "short equals long plus the effect of omitted times the regression of omitted on included", Mostly Harmless Econometrics, page 60)

so (1) becomes:

$$OVB = E(\widehat{b}|\mathbf{x}, \mathbf{w}) - \beta = \gamma \frac{Cov(x, w)}{Var(x)} \quad (2)$$

- Extent of omitted variables bias (OVB) depends on:
  - (i) correlation between x and w ($Cov(x, w)$),
  - (ii) relationship between y and w (γ), and
  - (iii) variance of x.
- OVB is increasing in γ, $cov(x, w)$, and $var(x)^{-1}$.
  - OVB increases with higher value of γ. This implies that w (the omitted variable) has a greater effect on y.
  - OVB increases with higher values of $cov(x, w)$. This is because the effects from w are captured by the coefficient of x. The effect that should be captured by γ is captured by b mistakenly, because of the correlation between w and x.
  - OVB decreases with higher values of $cov(x, x) = var(x)$.
- In practice, we focus on $Cov(x, w)$ to make sure that the omitted variable and the included variable are uncorrelated!!
- First two methods that deal with this are: (i) **randomized controlled experiment design** (Lecture 2, today) and (ii) instrumental variables method (Lecture 3).

Examples of omitted variable bias

$$ln(wage) = \beta_0 + \beta educ + \gamma abil + u$$

Suppose *abil* is not observed.

$$E(\widehat{b}|educ, abil) = \beta + \gamma \frac{cov(educ, abil)}{var(educ)} \quad (3)$$

	$cov(educ, abil) > 0$	$cov(educ, abil) < 0$
$\gamma > 0$	positive bias	negative bias
$\gamma < 0$	negative bias	positive bias

$$OVB = E(\widehat{b}|educ, abil) - \beta = \gamma \frac{cov(educ, abil)}{var(educ)} \quad (4)$$

- So (*abil*) and *educ* are positively correlated and b is upward biased.
- If omitted variable (*abil*) is uncorrelated with *educ*, then no bias occurs for OLS estimator of β.
- If we want an RCT on *educ*, the key is to make *educ* to be completely random so that *educ* is not correlated with omitted variables (*abil* and other inputs).

An example of omitted variable bias: the Mozart effect

$$Score = \alpha + \beta Music/Art + \mathbf{x}\gamma + u$$

where **x** is other observed variables and *Music*/*Art* taking music or art courses

- Rauscher, Shaw and Ky (1993) in Nature suggested that listening to Mozart for 10-15 minutes could temporarily raise your IQ by 8 or 9 points. (This study made big news.)
- Subsequently a review of dozens of studies by Winner and Cooper (2000) and by Hetland (2000) in Journal of Aesthetic Education found that students who take optional music or arts courses do better in English and math tests.

Examples of omitted variable bias: the Mozart effect

- A closer look at these studies suggests that these **music classes were optional classes** and the **correlation** between performing well (i.e. high score in math) and taking music or art courses could arise due the fact:
- It is not a random decision for a student to take these classes, but: (i) academically better students (i.e. higher ability students) might have more time to take optional courses or
  (ii) only better quality school (i.e. schools in rich districts) can provide a deeper music and arts curriculum.
- In the regression terminology, the estimated relationship between test scores and taking optional music/art courses appears to have *omitted variable bias problem*.
- Corresponding omitted factors are (i) students' innate **ability** or (ii) the overall **quality of the school**
- So is there a Mozart effect? One way to find out is to do an RCT!

Examples of omitted variable bias: the Mozart effect

- If we select students in a **random way** to be in the treatment of Art/Music classes in a RCT, that would eliminate the omitted variable bias.
  - Then $Cov(Music/Art, ability) = 0$, which eliminates the OVB.
  - This is because assignment of having "*Music*/*Art*" is **random** in experiments.
  - The random assignment of participants to "treatment" and "control" groups **eliminates differences in observed characteristics**, and thus the OVB.
- Taken together, many controlled experiments on the Mozart effect **fail** to show that listening to Mozart improves IQ.

Another example

- Assume that you want to find whether there is any link between class size and student learning.
- Perhaps school systems can save money by hiring fewer teachers, with no reduction in class size.
- Using the existing data would not be very informative from a causal perspective... Why?
- The reason is that weaker students are often deliberately grouped into smaller classes!
- So we need an RCT!

Another well-known example of RCT

- Krueger (1999) econometrically analyses a randomized experiment of the effect of class size on student achievement.
- The project is known as Tennessee Student/Teacher Achievement Ratio (STAR) and was run in the 1980s.

- It was a very ambitious and influential experiment, and costed around 12 million dollars.
- The average class size in regular Tenessee classes was about 22.3 kids.

Another well-known example of RCT

- The experiment assigned 11,600 students and their teachers to one of three groups:
- 1. Small classes (13-17 students).
- 2. Regular classes (22-25 students) and a part time teacher's aide (regular arrangement).
- 3. Regular classes (22-25 students) with a full time teacher's aide.
- Schools with at least three classes could choose to participate in the experiment.
- After the assignment, the design called for students to remain in the same class type for four years.
- Randomization occurred within schools.

How RCTs eliminate omitted variable bias?

$$E(\widehat{b}|x,w) = \beta + \gamma \frac{Cov(x,w)}{Var(x)} \tag{5}$$

where β is coefficient on x (included variable, in this case *Classsize*) and w is an omitted variable (could be prior score/achievement).

- A RCT (like the STAR experiment) which is randomly assigning participants to "treatment" (x=1) and "control" (x=0) groups guarantees that $Cov(x,w) = 0$.

- When $Cov(x,w) = 0$, then this eliminates the OVB which is equal to $\gamma \frac{cov(x,w)}{var(x)}$ and thus $E(\widehat{b}|x,w) = \beta$

- To check this **in practice** we usually compare pre-treatment characteristics or other covariates across treatment groups.
- For example, we would like to compare the following variables across the three treatment groups (being assigned to a small class/regular class with part time aid/regular class with full time aid): students' race, free lunch variable (proxy for family income), students' age, pre-treatment test scores etc. and find no statistical differences across the treatment groups.

2. Another violation of A4: Reverse Causality

- Idea is that correlation between y and x may be because it is y that causes x not the other way round.
- Interested in causal model:

$$y = \beta x + u \tag{6}$$

- But there is also a relationship in the other direction:

$$x = \alpha y + \epsilon \tag{7}$$

- Reduced form is (if you substitute (6) into (7)):

$$x = \frac{\epsilon + \alpha * u}{1 - \alpha\beta} = \frac{1}{1 - \alpha\beta}\epsilon + \frac{\alpha}{1 - \alpha\beta} \cdot u \tag{8}$$

- x is correlated with u in (6) because of (8). Thus $cov(x, u) \neq 0$. Thus **A4 is violated, which is the exogeneity assumption** and thus x is endogenous.
- **Reverse causality leads to bias in OLS estimator.**

Example: Do hospitals make people healthier? If we run a regression of HEALTHPROB (takes the value 1 if you currently have a health problem) on PATIENT(takes the value 1 if you have spent at least one night in hospital in the past year, 0 otherwise), we get the following output from STATA:

(hint: reg HEALTHPROB PATIENT in stata)


```
      Source |       SS       df       MS              Number of obs =   16592
-------------+------------------------------           F(  1, 16590) =  764.28
       Model |  108.427588     1  108.427588            Prob > F      =  0.0000
    Residual |  2353.60803 16590  .141880908           R-squared     =  0.0440
-------------+------------------------------           Adj R-squared =  0.0440
       Total |  2462.03562 16591  .148395854           Root MSE      =  .37666

------------------------------------------------------------------------------
  HEALTHPROB |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     PATIENT |    .229932   .0095126    27.65   0.000     .2443362    .2816277
       _cons |    .153447   .0030092    50.99   0.000     .1475386    .1595077
------------------------------------------------------------------------------
```


- Regression model used is:

$$HEALTHPROB = 0.153 + 0.262PATIENT$$

Example: Do hospitals make people healthier?

$$HEALTHPROB = 0.153 + 0.262PATIENT$$

- We can interpret that as having spent an additional night in hospital (and thus having got additional medical treatment) increases the probability of having a health problem. This is counter-intuitive.

- This could be due to **endogeneity problem** from **reverse causality**. It could be that:
  health problem (Y) causes $->$ hospitalisation (X),
  or it could be that
  because you got hospitalised (X) $->$ you have additional health problems (Y).

- The health problem could be the **cause** of getting medical treatment or the **outcome** because you got hospitalised.

- The Stata result shows a very significant positive relationship between the two variables (hospitalisation and health problem). The "causal interpretation" of this relationship is that **going to hospital makes you sick** (??), but a moment's thought should convince you that *this might not be the case.* **It is very likely that a reverse causality problem exists in this model.**
- You should be critical when you make your own models and you should always think intuitively.
- One solution is to use instrumental variable method (lecture3).

SUMMARY: Reverse causality $->$ endogeneity
$(Cov(x, u) \neq 0) -> \beta$ is biased

Another example or reverse causality: Smoking and Depression

- Another example: A study may find that smoking and depression are linked. It could be hat:
  smoking (Y) $->$ depression(X)
  or it could be that
  depression(X) $->$ smoking (Y).

- The likelihood is that they both cause each other (smoking causes depression and depression causes smoking) because smokers may feel societal pressure to quit, but may not have the willpower to do so.

3. Another violation of A4: Measurement Error (ME)

- When we use the imprecise measure of an economic variable in a regression model, then our model contains measurement error.
- Our recorded measures of the variable may contain errors (for example: inaccuracies in measuring family savings)
  - **Measurement Error in the Dependent Variable (Y)**
  - Let Y^* could be annual family savings, but we report Y because families are not perfect in their reporting of annual family savings; it is easy to leave out categories or to overestimate the amount contributed to a fund
  - Consider a savings function:

$$Y = \beta_0 + \beta_1 size + \beta_2 educ + \beta_3 age + \beta_4 income + u$$

where the measurement error is

$$ME = Y - Y^*$$

3. Another violation of A4: Measurement Error (ME)

- It might be reasonable to assume that the measurement error is not correlated with size, and age. On the other hand, we might think that families with higher incomes, or more education, report their savings more (or less) accurately.
- **We can never know, unless we can collect data on actual savings.**
- If you can prove that the measurement error is not related to the explanatory variables, then this measurement error is considered to be random.
- **If the measurement error is just a random reporting error that is independent of the explanatory variables, then OLS is perfectly appropriate.**

Measurement Error in the Independent Variable (X)

- The measurement error could also refer to the independent variable.
- Consider the model:

$$CollegeGPA = \beta + \gamma Income^* + \vartheta highschoolGPA + \theta Familybackground + u$$

- You have collected precise data on collegeGPA, highschoolGPA, and family background.
- But **family income**, *especially as reported by students*, could be easily mismeasured.
- But still we observe $Income^*$ (what students reported), instead of the actual Income.
- If Income=$Income^*$+e, then using this reported family income in place of actual family income ($Income^*$) will bias the OLS estimator.

- You need to make assumptions about the measurement error. The most common one is that cov(x_i^*,ME)=0. But in some cases it is not very easy for your audience to believe these assumptions.

Another violation of A4: Sample selection

Sample selection or Selection Bias: The bias introduced by the selection of individuals or groups in such a way that proper randomization is not achieved.

When there is selection bias the sample studied differs systematically from the population of interest intended to be analyzed, leading to systematic error in an association or outcome.

Another violation of A4: Sample selection

- Miguel and Kremer (2004) study the impact of a treatment against intestinal worms in primary school in rural Kenya.
- Intestinal worms affect one in four people worldwide and are particularly prevalent among school-age children in developing countries.
- Using an RCT, they evaluated the effect of the de-worming treatment on health, school absenteeism, and test scores.
- Treatment schools received half yearly (or yearly for different worms) treatment and medical education of how to avoid worm infection.
- Overall 75 schools were treated, while others were not. Schools were randomly assigned to treated and control groups.

Another example of sample selection

- **Sample selection bias** could arise because:
  - Individuals assigned to comparison/control group could attempt to move into the treatment group
    - Deworming program: parents could attempt to move their children from comparison/control schools to treatment schools.
- Alternatively, individuals allocated to treatment group may not receive the treatment.
    - Deworming program: some students assigned to treatment in treated schools schools did not receive medical treatment. In treated schools not all children received the treatment, mostly because of school absence on the treatment day.

Common features of problems

- All problems – omitted variables, reverse causality, measurement error, sample selection etc – have an econometric form: **the same one**: A4 is violated, $E(u|x) \neq 0$, $cov(X, u) \neq 0$
- **How to overcome this problem if we cannot design an RCT?**
  - **One approach** is to find econometric procedures that provide consistent estimates of the causal effects even in the presence of these problems.
    - One possible way is to use **instrumental variables**.
    - Or other more sophisticated econometric methods than OLS including **fixed effects, propensity score matching, regression discontinuity design** etc.
  - **Another approach** is to find better data.
    - Griliches: "Since it is the 'badness' of the data that provides us with our living, perhaps it is not at all surprising that we have shown little interest in improving it."

Recent trends

- A lot of emphasis on **good quality data** and **research design** than 'statistical fixes'. This leads to the explosion of **randomized controlled experiments.**
- Started in labor economics and development economics but now arriving in **most** fields.
- Suppose that we want to estimate the causal effect of x on y. If you could *choose* the source of variation in x, how would you do it?
- The answer is that you would want to allocate x at random (i.e. to give treatment to all sample members with **equal probability**). This is what is known as a randomized controlled trial/experiment (RCT).

Randomized controlled trials/experiments

- Evidence from randomized controlled experiments are referred as the '**gold standard**' - since **random assignment allows us to talk about causal effects**.
- By design, a policy variable, x_1, (i.e. variable of interest) will be independent of **any other influences** on y, whether they are observed or unobserved.

$$y = x\beta + u$$

- Start with example where x_1 is binary (though simple to generalize):
  - $x_1 = 0$ is control group
  - $x_1 = 1$ is treatment group
- **Randomization implies everyone has the same probability of getting the treatment!**
- Why is Randomization Good?
  - It solves the problems of omitted variables, sample selection etc

An Example: Racial Discrimination

- Black men earn less than white men in US and we want to know whether the difference in earnings is due to discrimination.
- It could be due to discrimination or other unobserved by the researcher factors but observed by the employer

$$Employability = \alpha + \beta_1 black + \beta_2 educ + \beta_3 abil + \mathbf{x}\gamma + \epsilon$$

- Hard to fully resolve non-experimental data problems due to 'omitted variables'.

An Example: Racial Discrimination

- Bertrand/Mullainathan "Are Emily and Greg More Employable Than Lakisha and Jamal" American Economic Review, 2004
  - Create **fake resumes** and send them to job adverts.
  - Allocate **names at random to resumes** - some given black-sounding names (treatment, $x_1 = 1$), others white-sounding names (control, $x_1 = 0$);
  - Outcome variable is call-back rates (y).
  - Interpretation of this study: these are probably not direct measures of racial discrimination, just the effect of having a black-sounding name on outcome.
  - But the name is uncorrelated *by construction* with **other factors** on resume (education, ability etc.).

Notation for RCT

- Interested in treatment effect:

$$E(y|x_1 = 1) - E(y|x_1 = 0) \tag{9}$$

We will learn about discrimination effects by comparing the average outcomes of those who have black-sounding names and those who have white-sounding names.

- Estimating Treatment Effects. The Statistical Approach:
  - Take mean of outcome variable in the treatment group
  - Take mean of outcome variable in the control group
  - Take the difference between the two
- The coefficient β_1 in the following regression is equivalent to (9).

$$y = \alpha + \beta_1 \cdot x_1 + u \tag{10}$$

- No problems in (9). It is fine if x_1 is a binary variable. But:
  - Does not directly compute standard errors
- **Suggestion**: Use a regression approach of (10) to estimate treatment effects.

Regression for RCT

Run the regression:

$$Employability = \alpha + \beta_1 x_1 + u$$

However, employability depends on other factors beside race.

$$y = \beta_1 x_1 + \mathbf{x}\gamma + u$$

where y is employability, x_1 is a black-sounding name dummy, $\mathbf{x} = (1, x_2, x_3, ..., x_k)$ are other factors

- **Proposition: The OLS estimator of β_1 is an unbiased estimator of the causal effect of x_1 on y**

- By the definition of randomization of x_1: x_1 should not be correlated to any other variables.
- Hence, the explanatory variables should be uncorrelated to the treatment dummy: $cov(x_1, \mathbf{x}) = 0$

Computing standard errors

- Stata gives us estimates of the standard errors for β_1.
- Unless told otherwise the regression package will compute standard errors assuming errors are homoskedastic.
- Heteroskedasticity-robust standard errors can be easily implemented
  - simple to use this in Stata

    reg y x, robust

Problems with RCTs:

- Expensive
  - Randomized Controlled trials are often very **expensive**. Project STAR costed 12m - whereas non-experimental data are often available at little or no additional cost.
- Ethical Issues
  - Ethical issues related to some people receiving treatment and others not. (e.g. Who will get the de-worming treatment? Can you randomly assign race or sex?)

Problems with RCTs

- Threats to **Internal Validity**
  - Failure to follow experiment: (i) non-compliance (ii) attrition/missing data (Examples of non-compliance: -The individual is assigned to treatment but she does not take the treatment. -The individual is not assigned to treatment but she does take the treatment)
  - Experimental effects (Hawthorne effects): People behave differently because they are part of an experiment. If they operate differently on treatment and control groups they may introduce biases.
- Threats to **External Validity**
  - Non-representative programme
    - All social programmes are different but external validity requires a particular programme to have the same impact in other places at other times, when the context might be very different. (e.g. a programme to help the unemployed find work might have very different effects in a boom and recession)
  - Non-representative sample (e.g. whether students in Tennessee represent the whole population)
  - Treatment vs. Eligibility Effects
    - Participation in many social programmes is often voluntary. Often we give people an opportunity and we do not force them to do it.

Summary and Conclusions on RCTs

- Well-implemented RCTs do represent the 'gold standard' of research
  - No bias due to omitted variable and sample selection bias which are prevalent in microeconomics data.
- Although not necessary, we want to include other relevant variables in the regression of RCTs in order to:
  - Enhance efficiency, check/improve randomization, conditional randomization
- Not many RCTs to keep us busy and, for many important questions, we lack evidence from social experiments and will continue to do so for a long time.
- So we will keep also working with non-experimental data.

Reasons to include other regressors

- Is data on other variables of any use at all for RCTs?: Not necessary but useful, so there are reasons to include other regressors even with a randomized controlled experiment
  - Improved efficiency
  - Check for randomization
  - Improve randomization
  - Control for conditional randomization

1. Improved Efficiency

- Don't just want consistent estimate of causal effect - also want low standard error (or high precision or efficiency).
- Standard formula for standard error of OLS estimate of β_j is:

$$Var(\hat{\beta}_j|\mathbf{x}) = \frac{\sigma^2}{SST_j(1-R_j^2)}$$

- σ^2 comes from variance of residual. In regression: $(1-R^2) * Var(y)$.

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} = 1 - \frac{\sigma^2}{Var(y)}$$

$$\sigma^2 = (1-R^2) \cdot Var(y)$$

  - Include more variables raises R^2 so that reduces σ^2.
  - $plim Var(\widehat{\beta})$ falls as σ^2 falls.
  - x_j (random assignment) is independent of other $\mathbf{x}$, R_j=0. Thus, including more variables reduces $Var(\hat{\beta}_j|\mathbf{x})$.

2. Check for Randomization

- Randomization can go wrong
  - Poor implementation of research design
  - Non-compliance and attrition
- If randomization on x_1 is done well, then $\mathbf{x}$ (other regressors) should be independent of x_1 - this is testable:

$$x_1 = \mathbf{x}\delta + u_2$$

  - Hypothesis test using $H_0 : \delta = 0$
  - (Example) Test for differences in $\mathbf{x}$ between treatment ($x_1 = 1$) and control groups($x_1 = 0$) (e.g. The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR, The Economic Journal 2001 by Krueger and Whitmore.)
  - We can apply probit/logit model for x_1 on $\mathbf{x}$.

3. Improve upon randomization: conditional randomization

- Conditional randomization is where probability of treatment is different for people with different values of $\mathbf{x}$, but random conditional on $\mathbf{x}$ (e.g. Project STAR, random assignment of class types within school. x_1 is known to be correlated with school characteristics but within schools, it is not correlated with teachers, student, and other within school characteristics.)
- This is a case where we must include $\mathbf{x}$ to get consistent estimates of treatment effects (i.e. schools in wealthier regions with more resources have more small classes so students in schools of wealthier regions are more likely to be assigned in a small class).

- consider following two models:

$$y = x_1\beta + \mathbf{x}\gamma + u_4$$

$$y = x_1\beta + u_5$$

- But, conditional on $\mathbf{x}$, x_1 is independent of other factors (e.g. x_1 is small class, $\mathbf{x}$=(school, teacher's experience, others)). In other words, x_1 is correlated with u_5 but not correlated with u_4.

- Randomized controlled experiment on x_1:

$$Score = \beta \cdot small + u_5$$

- It requires $cov(u_5, small) = 0$ when $u_5 \equiv \mathbf{x} \cdot \gamma + u_4$
- However, in actual implementation, you may want to do (or cannot avoid doing) **conditional** randomization in which treatment is randomized conditional on some observable variables but is different for people with different values of those conditioning variables.
- In regression model, it is equivalent to including other relevant variables other than randomized variables.

$$log(wage) = \beta \cdot small + \mathbf{x} \cdot \gamma + u_4$$

- It allows (i) $cov(small, \mathbf{x}) \neq 0$ (and only requires $cov(u_4, x_1) = 0$) and, moreover, (ii) adding relevant variables lead to more **precise** estimation of β in general.