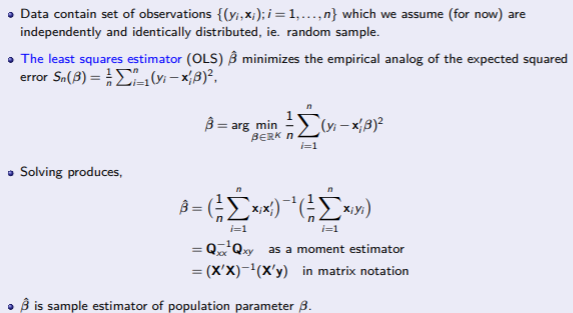
Part B1: Definitions and Etiquette. 7 Definitions:

**1.) Provide the OLS Estimator and its assumptions. Which ones will go uncaught if violated?**



Linearity (A.1): Observations are IID and satisfy

Strict Exogeneity (A.2)

Variables have finite second moments (A.3)

Invertibility (A.4): Also known as no-multicollinearity

**2.) Provide a definition of Bias. Define Selection bias (formally).**

Selection bias is a type of bias that results from non-random selection of observations.

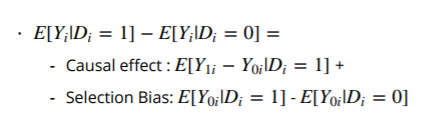
Take the content of hospital treatment in the textbook as an example. So:

Problem: we can never see both outcomes Y0i and Y1i at the same time

Solution: we must learn about the effects of hospitalization by comparing the average health of those who were and were not hospitalized

E[Yi|Di = 1] − E[Yi|Di = 0]

So selection bias will be



**3.) What happens to the OLS estimator if X is endogenous with the error term. (show formally)**

When having an internal cause or origin, implying there is correlation between the parameter or variable and the error term:

1 Omitted Variable Bias.

2 Measurement errors in regressors.

3 Simultaneous equations.

4 Program or treatment evaluation with selection into treatment.

**4.) Provide the 2 main IV assumptions?**

**a. Which of them can only be defended, which one can be tested?**

Correlation will result in weak tools that are still useful. So endogeneity needs to be tested. If there is slight endogeneity, you need to use this instrument.

**b. Show what happens to the estimator when the exclusion restriction is violated.**

Displayed when violations occur exclusion limit estimator. The estimator comprises a portion of instrument related error term.

**5.) What are the assumptions required for IV-2SLS to identify a LATE?**

Those are:

i. Random assignment

ii. Exclusion Restriction

iii. Relevance

iv. Monotonicity

**6.) Imagine a Data Generating Process follows the Linear Panel model, but you only observe a Cross Section. Show formally what happens.**

Basic linear panel data model:

Yit = α + Xitβ + ηi + λt + Uit i = 1,...,N t = 1,...,T

ηi is an individual specific effect, which captures components that are unobserved by the econometrician.

λt is some function (polynomial) in time (henceforth, for simplicity of notation, we include λt in Xit ).

Often, the individual specific effect contains omitted variables which are correlated with the regressors, E[ηi |Xi1,...,Xit] 6= 0

It has different effects in fixed effect and random, therefore, it need to follow MUNDLAK SPECIFICATION and do HAUSMAN TEST.

**7.) You are setting up a lab experiment, (about returns to a surprise increase in a piece rate called “gift”).**

**Now you wonder whether you would benefit from including a difference in differences setup in the experiment.**

**a. Provide the linear in means estimator for experiments and the regression estimator.**

**b. Write out the DID linear in means estimator (separating pre/post and treated/control), and the DID regression estimator.**

**c. How would you implement DID in an experiment, what do you think you would gain, what would the cost be?**

**DID/IV/RD Etiquette:**

**What are** **the differences in the Etiquettes for DiD and IV. Based on the key assumption in the identification strategies, try to explain why the etiquette is different for the two methods.**

DiD Etiquette:

1. Explain and Defend the Experiment

2. Present Raw Data in Terms of a Graph

3. Show treatment and control are similar pre-treatment.

4. Present Baseline Estimates.

5. Investigate pre-treatment patterns

6. Run many Robustness checks

7. Discuss the External Validity/Interest of the Treated Group

8. Apologize

IV Etiquette:

1. Explain and defend the experiment

2. Test for power and over identification

3. Do a reduced-form regression of the dependent variable (2nd stage) on the instruments

4. Conduct multiple robustness checks

5. Discuss the assumptions behind homogeneous treatment effects and/or explain why the treated

population is inherently interesting

6. Apologize for all that is still unproven and give caveats

**What are the differences in the Etiquettes for RD and IV. Based on the key assumption in the identification strategies, try to explain why the etiquette is different for the two methods.**

RD Etiquette:

1. Explain and Defend the Experiment

2. Present Raw Data in Terms of a Graph

3. Show treatment and control are similar below and above the threshold.

4. Present Baseline Estimates

5. Run many Robustness checks - Does the effect survive

6. Discuss the External Validity/Interest of the Treated Group

7. Apologize

IV Etiquette:

1. Explain and defend the experiment

2. Test for power and over identification

3. Do a reduced-form regression of the dependent variable (2nd stage) on the instruments

4. Conduct multiple robustness checks

5. Discuss the assumptions behind homogeneous treatment effects and/or explain why the treated population is inherently interesting

6. Apologize for all that is still unproven and give caveats