Home Assignment after session 03.

PART A: Reading Assignments:

Q: Which part of the slides corresponds to chapter 3?

Ans: Part II.1: Linear Model and OLS
Regression corresponds to the Chapter 3.

Look at Appendix A on Matrix Al, gebra, and browse over the sections I have indicated. (A1, A3, A4, A6, A7, A8, A10, A11, A20 – For this course it's enough to be able to look up, whenever you are uncertain).
□□Provide a sentence listing all sections you read, and explain whether you understood all, some or none of them.

A1 explains vector and scalar. Transpose of matrix square matrix , identity matrix and partitioned matrix.

A3 explains matrix addition

A4 explains matrix multiplication, orthogonal and conferrable vectors. Further matrix multiplication is not commutative but associative and distributive.

A6 explains rank of the matrix, inverse of the matrix and nonsigular matrix A, Woodbury matrix identity, Sherman–Morrison formula, Moore-Penrose generalized inverse *A*

A7 orthogonal , orthonormal matrices and unit vector.

A8 provides definition of determinant

Laplace's expansion, properties of determinant

A9 explains eigenvalue ,latent roots, and spectral decomposition

A10 Definition of positive Definite Matrices, Properties of positive semi-definite matrices

A11 provide idempotent matrix definition ,Moore-
Penrose generalized condition

A20 provide details on Matrix Calculus

And Properties of matrix derivatives (did not do proofs)

More: **I also recommend to read Appendices B1-B3 and B5.**
Provide a short sentence about each of these chapters, to document your reading
B.1 Inequalities for Real Numbers Expressions of Triangle Inequality

Jensen’s Inequality

Geometric Mean Inequality

Loève’s *cr* Inequality

NormMonotonicity

B.3 Inequalities forMatrices

Schwarz Matrix Inequality

Triangle Inequality

Trace Inequality

Quadratic Inequality.

Strong SchwarzMatrix Inequality

NormEquivalence

Eigenvalue Product Inequality

B.5 Proofs*
Proof of Triangle Inequality (B.1).

Proof of $c2$ Inequality

Proof of Schwarz Inequality

Proof of Minkowski’s Inequality

Appendix B-3 Proof of Triangle Inequality

2) Revision of slides

3) Outlook

Check out the slide deck on Experiments, and discuss the difference of the ATE and the ATET.
Check out the slide deck on IV: o Give the definition of the simple IV in the univariate case.
List the two key assumptions that have to hold for an IV?

Difference of ATE and ATET
ATE is the average of all value for unit level casual effect in the population. Average outcome under the policy - average outcome without policy $ATE = E[YD=1]-E[YD=0\ ]$

Averaged treatment effect on treated (ATET)is average causal effect is for subsets of the units

$ATET = E[YD=1\ ]-E[YD=0] \mid D=1$

The principal econometric problem in the estimation of treatment effects is selection bias, which arises from the fact that treated individuals differ from the non-treated for reasons other than treatment status per se.

Instrumental variable

An instrumental variable is a **third variable introduced into regression analysis** that is correlated with the predictor variable, but uncorrelated with the response variable. By using this variable, it becomes possible to estimate the true causal effect that some predictor variable has on a response variable.

IV Assumptions

Lets say Z is instrumental variable, *Z*, to satisfy the following assumptions:

- *Relevance*: Z can predict ΔX i.e. $cov(Z, X) \neq 0$
- *Exogeneity*: Z is uncorrelated with the error term i.e. $cov(Z, \varepsilon) = 0$

Relevance is important because it essentially states that our instrument of choice is correlated with our independent variable of our choice . *Exogeneity* is important because it states that our instrument is uncorrelated with the error term

Bonus Go over confidence intervals and hypothesis testing
□□□□Revisit the testing section in the second Stats-Primer (UEA_Week_01_005A_Fast_primer_02_Lecture_MRB).
□□□□Read the last two chapters on testing in the Extra slide deck on asymptotics (UEA_ecoR2PhD_ExtraLectB03_Asymptotics_StK_SLDs
□□□□Compare the two testing section, and provide a personal note, which treatment you prefer, and why 3-5 sentences)

The testing method 2, using asymptotic distributions' is better than testing method 1, in asymptotic distribution when sample size tends to infinity the estimator is converge to its true value / behaves normally.

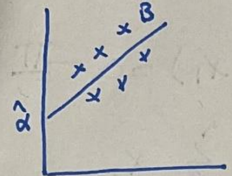
PART B

***Try to* understand how we derive the OLS estimator.**
(this is bonus: try this for 30 min, then move on and come back)
□□First, make sure you understand the minimization of the CEF. oNow enumerate all the rules from stats primer slidedeck 1 that have been used and indicate in which line of the proof.

When deriving the OLS-estimator (in the sample analogue), the optimization was skipped. o Try to do the minimization (the sum of squared residuals) for the sample analogue and derive beta_hat.

you might be asked to reproduce this.

Deriving Least Square Estimator $\hat{\alpha}$ $\hat{\beta}$



$$S = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_1 - \hat{\alpha} - \hat{\beta} x)^2$$

$$\hat{y} = \hat{\alpha} + \hat{\beta} x_1$$

minimization First order condition

$$\frac{\partial S}{\partial \hat{\alpha}} = 0 \qquad (I) \Rightarrow \frac{\partial S}{\partial \alpha} = -2 \sum_{i=1}^{n} (y_1 - \hat{\alpha} - \hat{\beta} x_1) = 0$$

$$\frac{\partial S}{\partial \hat{\beta}} = 0$$

2nd order condition

$$(II) \quad \frac{\partial S}{\partial \hat{\beta}} = -2 \sum_{i=1}^{n} x_1 (y_1 - \hat{\alpha} - \hat{\beta} x_1) = 0$$

let

$$\sum_{i=1}^{n} x_1 = N\bar{x} \qquad \sum_{i=1}^{n} (x_1 - \bar{x})(y_1 - \bar{y}) = \sum_{i=1}^{n} y(x - \bar{x})$$

$$= \sum_{i=1}^{n} x_1 (y_1 - \bar{y})$$

$$\sum_{i=1}^{n} y_1 = N\bar{y}$$

$$\boxed{\frac{1}{N} \sum_{i=1}^{N} x_1 = \bar{x}} \qquad \sum_{i=1}^{n} (x_1 - \bar{x})(y_i - \bar{y})$$

$$\boxed{\frac{1}{N} \sum_{i=1}^{n} = \bar{y}} \qquad = \sum_{i=1}^{n} (x_1 y_1 - x_1 \bar{y} - \bar{x} y_1 + \bar{x}\bar{y})$$

$$\boxed{\begin{array}{l} \sum_{i=1}^{n} x_1 = N\bar{x} \\ \sum_{i=1}^{n} y_i = N\bar{y} \end{array}} \qquad = \sum_{i=1}^{n} x_1 y' - \bar{y} \sum_{i=1}^{n} x - \bar{x} \sum_{i=1}^{n} y_1 + \bar{x}\,\bar{y} \sum_{i=1}^{n}$$

$$= \sum_{i=1}^{n} x_1 y_1 - N\bar{y}\bar{x} - \cancel{N\bar{y}\bar{x}} + \cancel{N\bar{y}\bar{x}}$$

$$= \sum_{i=1}^{n} x_1 y_i - N\bar{y}\bar{x} = \sum_{i=1}^{n} x_1 y_1 - \bar{x} \sum_{i=1}^{n} y_i$$

$$\sum_{i=1}^{n} (y_1 x_1 - y_i \bar{x}) = \sum_{i=1}^{n} y_1 (x_1 - \bar{x})$$

1st order condition $\rightarrow -2\sum_{i=1}^{n}(y_1 - \hat{\alpha} - \hat{\beta}x_1) = 0$ — I

2nd order $\rightarrow$ $-2\sum_{i=1}^{n} x_1(y_1 - \hat{\alpha} - \hat{\beta}x_1) = 0$ — II

$$\sum_{i=1}^{n} y_1 = \hat{\alpha}\sum_{i=1}^{n} 1 + \hat{\beta}\sum_{i=1}^{n} x_1$$

$$N\bar{Y} = \hat{\alpha}N + \hat{\beta}N\bar{X}$$

$$\bar{Y} = \hat{\alpha} + \hat{\beta}\bar{X} \longrightarrow \text{III}$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} \text{ — IV}$$

using 2nd order condition

$$\sum_{i=1}^{n} x_1 y_1 = \hat{\alpha}\sum_{i=1}^{n} X + \hat{\beta}\sum_{i=1}^{n} x_1^2$$

$$\sum_{i=1}^{n} x_1 y_1 = \hat{\alpha}N\bar{X} + \hat{\beta}\sum_{i=1}^{n} x^2$$

$$\sum_{i=1}^{n} x_1 y_1 = (\bar{Y} - \hat{\beta}\bar{X})N\bar{X} + \hat{\beta}\sum_{i=1}^{n} x^2$$

$$\sum_{i=1}^{n} x_1 y_1 = N\bar{X}\bar{Y} - \hat{\beta}N\bar{X}^2 + \hat{\beta}\sum_{i=1}^{n} x^2 \text{ — V}$$

$$\therefore \sum_{i=1}^{n} x_1 y_1 - N\bar{X}\bar{Y} = \hat{\beta}\left(\sum_{i=1}^{n} x_1^2 - \bar{N}x^2\right)$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_1 y_1 - N\bar{X}\bar{Y}}{\sum_{i=1}^{n} x^2 - N\bar{X}^2}$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_1 y_1 - \bar{X}y_i)}{\sum_{i=1}^{n} x^2 - \bar{X}x_1}$$

$$\boxed{\hat{\beta} = \frac{Cov(X_1 Y_1)}{Var(X_1)}}$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\cdot\bar{X}$$

$$\boxed{\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(x - \bar{X})^2}}$$

□□***make sure that*** **you understand the proof of unbiasedness.**
□□**As before, indicate all the rules from the first stats slidedeck 0 and indicate where they are used (linearity etc - revisit the first slidedeck from the stats-primer as needed). o If you are not yet entirely sure about matrix inversion and multiplication, I recommend the Hansen Appendix or looking for a good cheat sheet on the web.**

Rules

i. linearity in parameters
ii. The expected value of error is zero for al observations E (e) = 0
iii. The conditional variance of error terms is constant in all x overtime
iv. Error term is independently distributed and not correlated

2.) Assume you are a senior economist and your intern/research assistant is showing you an OLS-model. However, you quickly realize that you have no reason to believe that E(u|X)=0. In econometrics, a violation of E(u|X)=0 is called "Endogeneity." As a thought experiment, you consider E(u|X) = 2 to be a much better assumption and would like to quantify for yourself whether it is a problem to use OLS, if, indeed E(u|X) = 2.

1. ⇨ Try to derive a formal expression for E(beta_hat) under E(u|X)= 2
2. ⇨ Can you pin down the bias? Is it zero, positive or negative?
3. ⇨ Be prepared to do this with pen and paper at a later point.

$$E(\hat{\beta} | u) = \beta + [X'X]^{-1} + E(u|X)$$

$$\text{Bias} = [X'X]^{-1} \times E[u|X) = 2$$

$$2\,[X'X]^{-1} X \neq 0 \quad \text{un.} = 2$$

OLS $\overset{\text{Bias}}{\text{estimate}}$ of β

$$\hat{\beta} = (X_1' X_1)^{-1} X_1 y$$

$$\beta + (X_1' X_1)^{-1} X_1 x_{2\gamma} + (X_1' X_1)^{-1} X_1' u$$

$$\beta_1 + \delta\gamma$$

$$\delta = (E \, \cancel{E}(x_{i1} \; x_{i1})^{-1} E(x_{i1} x_{i2})$$

$$\delta \neq 0 \rightarrow E x_{i1} x_{i2} \neq 0 \Longrightarrow \text{Bias.}$$