Home Assignment Week 1 :

Amir Jafarzadeh

PART A: Reading Assignments:

1. Please get hold of the main book for part 3: Hansen, B. Econometrics", [HaL]: • Familiarize with the structure. Read the intro and browse through chapter 3 (until 3.15). **Chapter 3 "The Algebra of Least Squares"**

3.1: introduction

3.2: samples: introduces the dataset or sample, the assumptions is that all variables are identically distributed.

3.3: Moment Estimators: calculation of mean and variance in sample and the universe population.

3.4: Least Squares Estimator: calculation of the concept of sum of squared errors function and why we should minimize it. The key is that "we define the estimator β^ as the minimizer of S^ (β)"

Definition 3.1 The **least squares estimator** is $\widehat{\beta} = \underset{\beta \in \mathbb{R}^k}{\operatorname{argmin}} \widehat{S}(\beta)$ where $\widehat{S}(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - X_i'\beta\right)^2$.

3.5: Solving for Least Squares with One Regressor:

Given β we calculate the "error" Yi − Xiβ by taking the vertical distance between Yi and Xiβ.

These vertical lines are the errors Yi − Xiβ.

$$\mathrm{SSE}(\beta) = \sum_{i=1}^{n}\left(Y_i - X_i\beta\right)^2 = \left(\sum_{i=1}^{n} Y_i^2\right) - 2\beta\left(\sum_{i=1}^{n} X_i Y_i\right) + \beta^2\left(\sum_{i=1}^{n} X_i^2\right).$$

This β^ can minimize the above sentence:

$$\widehat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}.$$

3.6: Solving for Least Squares with Multiple Regressors:

$$0 = \frac{\partial}{\partial\beta}\mathrm{SSE}(\widehat{\beta}) = -2\sum_{i=1}^{n} X_i Y_i + 2\sum_{i=1}^{n} X_i X_i'\widehat{\beta}.$$

As in the single regressor case this is a quadratic function in β. The difference is that in the multiple regressor case this is a vector-valued quadratic function.

$$\widehat{\beta} = \left(\sum_{i=1}^{n} X_i X_i'\right)^{-1} \left(\sum_{i=1}^{n} X_i Y_i\right).$$

3.7: Illustration: the author calculates the B in the regression of the sample (the March 2009 Current Population Survey) and interpretation of it.

3.8: Least Squares Residuals:

fitted value $\widehat{Y}_i = X_i'\widehat{\beta}$

the residual: $\widehat{e}_i = Y_i - \widehat{Y}_i = Y_i - X_i'\widehat{\beta}.$

The error ei is unobservable while the residual e^i is an estimator.

If there is no endogeneity problem, then we have: $$\sum_{i=1}^{n} X_i \widehat{e}_i = 0.$$

In which X is constant

3.9: Demeaned Regressors:

$Y_i = X_i'\beta + \alpha + e_i$ when X is not constant

Solving for the B^:

$$\begin{aligned}\widehat{\beta} &= \left(\sum_{i=1}^{n} X_i \left(X_i - \overline{X}\right)'\right)^{-1} \left(\sum_{i=1}^{n} X_i \left(Y_i - \overline{Y}\right)\right) \\ &= \left(\sum_{i=1}^{n} \left(X_i - \overline{X}\right)\left(X_i - \overline{X}\right)'\right)^{-1} \left(\sum_{i=1}^{n} \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)\right).\end{aligned}$$

the demeaned formula for the least squares estimator.

3.10: Model in Matrix Notation:

Written formula in the matrix notion:

$$\hat{\beta} = (X'X)^{-1}(X'Y)$$
$$\hat{e} = Y - X\hat{\beta}$$
$$X'\hat{e} = 0.$$

3.11: Projection Matrix:

Define the matrix p:

$$P = X(X'X)^{-1}X'.$$

The properties of P:

1. P is symmetric (P 0 = P).

2. P is idempotent (PP = P).

3. trP = k.

4. The eigenvalues of P are 1 and 0.

5. P has k eigenvalues equalling 1 and n −k equalling 0.

6. rank(P) = k.

3.12: Annihilator Matrix:

Define $M = I_n - P = I_n - X(X'X)^{-1}X'$

MX=0

M is annihilator matrix due to the property that for any matrix Z in the range space of X then

M Z = Z −P Z = 0.

The annihilator matrix M has similar properties with P.

3.13: Estimation of Error Variance:

σ 2 is a moment so a natural estimator:

$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2.$$

But since ei^2 is not observable, so we take a two-step approach to estimation.

The residuals e^i are calculated in the first step, and then we substitute e^i for ei to obtain the feasible estimator.

Then we obtain this inequality:

$$\tilde{\sigma}^2 - \hat{\sigma}^2 = n^{-1}\boldsymbol{e}'\boldsymbol{e} - n^{-1}\boldsymbol{e}'\boldsymbol{M}\boldsymbol{e} = n^{-1}\boldsymbol{e}'\boldsymbol{P}\boldsymbol{e} \geq 0.$$

his shows that the feasible estimator σ^ 2 is numerically smaller than the idealized estimator

3.14: Analysis of Variance:

Calculating coefficient of determination or R-squared with analysis of variance

$$R^2 = \frac{\sum_{i=1}^{n}\left(\hat{Y}_i - \overline{Y}\right)^2}{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2} = 1 - \frac{\sum_{i=1}^{n}\hat{e}_i^2}{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}.$$

It is often described as “the fraction of the sample variance of Y which is explained by the least squares fit

3.15: Projections:

visualizing least squares fitting as a projection operation

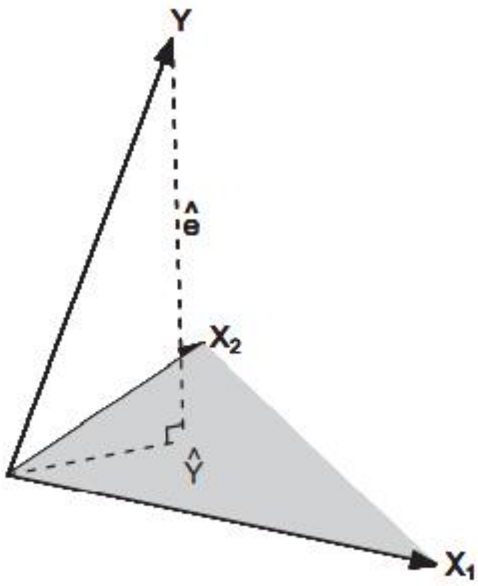


Figure 3.4: Projection of $\boldsymbol{Y}$ onto $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$

- **Which part of the slides corresponds to chapter 3?**

  UEA_Week_01_005A_Fast_primer_02_Lecture_MRB(1):

  - ✓ Section2: population, param.s & sampling: slides 10-14
  - ✓ Section3: finite sample properties of estimators: slides 16-26

- **Look at Appendix A on Matrix Al, gebra, and browse over the sections I have indicated**

- **A.1 Notation:** some basic matrix notion such as scalar, vector, matrix, transpose, square, identity matrix and partitioned matrix
- **A.3 Matrix Addition:** how to add a matrix with other matrices.
- **A.4 Matrix Multiplication:** how to multiple a matrix to other matrices. **if** A is k ×r and B is r ×s, so that the number of columns of A equals the number of rows of B, we say that A and B are conformable. A and B are orthogonal if A'B = 0
- **A.6 Rank and Inverse: explain rank of matrix and how to invers a matrix.**
- **A.7 Orthogonal and Orthonormal Matrices:** definition of orthogonal and unit vector
- **A.8 Determinant:** definition of determine and how to calculate it and properties of the determine.
- **A.10 Positive Definite Matrices:** definition of positive definite matrices and and semi-definite matrices and 7 properties of these matrices.
- **A.11 Idempotent Matrices:** some definition. The following is new for me. a k ×k square matrix A is idempotent if AA = A. if A is idempotent and symmetric with rank r then it has r eigenvalues which equal 1 and k − r eigenvalues which equal 0.
- **A.20 Matrix Calculus: definition of** The vector derivative and 6 properties of it and prove for some of them.

PART A: 2. Revise the slides: done all of them.

PART A: 3. Outlook:

- **Check out the slide deck on Experiments, and discuss the difference of the ATE and the ATET:**
  - ATE is Average Treatment Effect and means the differences between outcomes of subjects who are given intervention and the outcomes of those are not given (Simply take the difference in means outcomes between treatment and control groups).
  - . Both are factual factors, so observable.

$$ATE = \mathrm{E}[\mathrm{Y}_i^{\mathrm{D}=1} - \mathrm{Y}_i^{\mathrm{D}=0}] = \mathrm{E}[\mathrm{Y}_i^{\mathrm{D}=1}] - \mathrm{E}[\mathrm{Y}_i^{\mathrm{D}=0}]$$

  - ATET is Average Treatment Effect on the Treated and means the differences between outcomes of subjects who are given intervention and the outcomes of the similar individuals if they were not given. The first factor is factual but the second one is not, so the second one is not observable and we should estimate it.

$$ATET = \mathrm{E}[\mathrm{Y}_i^{\mathrm{D}=1} - \mathrm{Y}_i^{\mathrm{D}=0} \mid \mathrm{D}_i = 1] = \mathrm{E}[\mathrm{Y}_i^{\mathrm{D}=1} \mid \mathrm{D}_i = 1] - \mathrm{E}[\mathrm{Y}_i^{\mathrm{D}=0} \mid \mathrm{D}_i = 1]$$

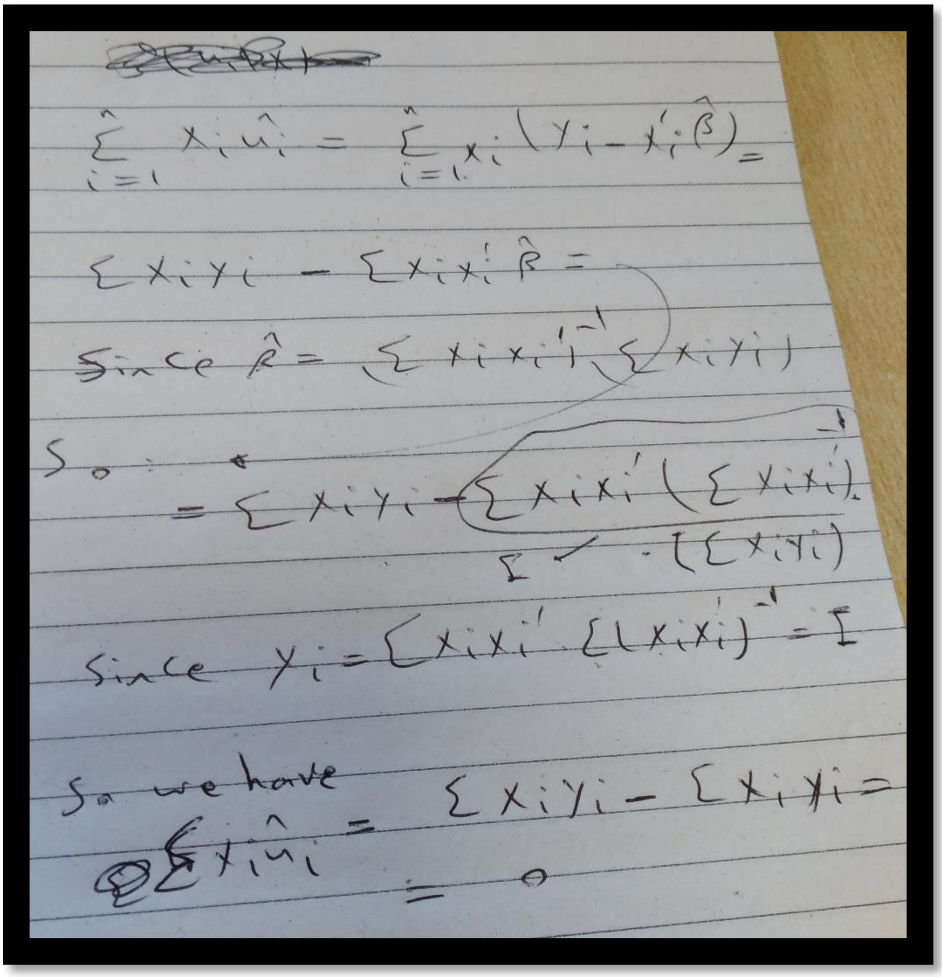
- **Check out the slide deck on IV:**
  - **Give the definition of the simple IV in the univariate case:**
    - ✓ Because of the problem of endogeneity, we should find an instrument (Zi) that solves this problem. With the IV, we have 2sls (two step least square). IV framework involves an additional step in causal reasoning. We can think of the randomization Zi as initiating causal chain: Zi operates on Di which in turn affects outcome Yi.
  - **List the two key assumptions that have to hold for an IV?**
    - ✓ The instrument variable (Zi) should be selected randomely (random assignment assumption) which means the correlation between zi and ei is zero.
    - ✓ The second assumption limits the causal channel of Zi on Yi to only operate through Di, such that $\mathrm{Y}_i^{\mathrm{D}=d,\mathrm{Z}=1} = \mathrm{Y}_i^{\mathrm{D}=d,\mathrm{Z}=0}$

PART B: Formal Exercise

1. **Revisit the OLS-slidedeck (UEA_ecoR2PhD_CoreLectA01_OLS_stkm)**

- **First, make sure you understand the minimization of the CEF.**
  - **Now enumerate all the rules from stats primer slidedeck 1 that have been used and indicate in which line of the proof:**
    - ✓ Justifying Linear CEF: 1. True model is linear, 2. Best predictor, 3. Best approximation
    - ✓ Key property of CEF error is (conditional) mean independence (also in some circumstances called strict exogeneity) E(u|x) = 0
    - ✓ The best linear predictor of y given x is found by selecting the vector β to minimize mean squared prediction error S(β) = E([y − x 0β] 2).
    - ✓ Linear CEF Model: $$y = \mathbf{x}'\beta + u$$ $$\mathrm{E}(u|\mathbf{x}) = 0$$
    - ✓ Linear Projection Model: $$y = \mathbf{x}'\beta + u$$ $$\mathrm{E}(\mathbf{x}u) = 0$$ $$\beta = (\mathrm{E}(\mathbf{x}\mathbf{x}'))^{-1}\mathrm{E}(\mathbf{x}y)$$
    - ✓ We use Central Tendency Measures, properties of expected value, Measures of Variability: Variance, Standard Deviation and covariance and properties of conditional expectation for minimizing CEF.
- When deriving the OLS-estimator (in the sample analogue), the optimization was skipped.

- Try to do the minimization (the sum of squared residuals) for the sample analogue and derive beta_hat.
  - ✓ This is written in the slide 23 in “Part II.1: Linear Model and OLS Regression”



$$\sum_{i=1}^{n} x_i \hat{u}_i = \sum_{i=1}^{n} x_i (y_i - x_i' \hat{\beta}) =$$

$$\sum x_i y_i - \sum x_i x_i' \hat{\beta} =$$

Since $\hat{\beta} = (\sum x_i x_i')^{-1} (\sum x_i y_i)$

So:

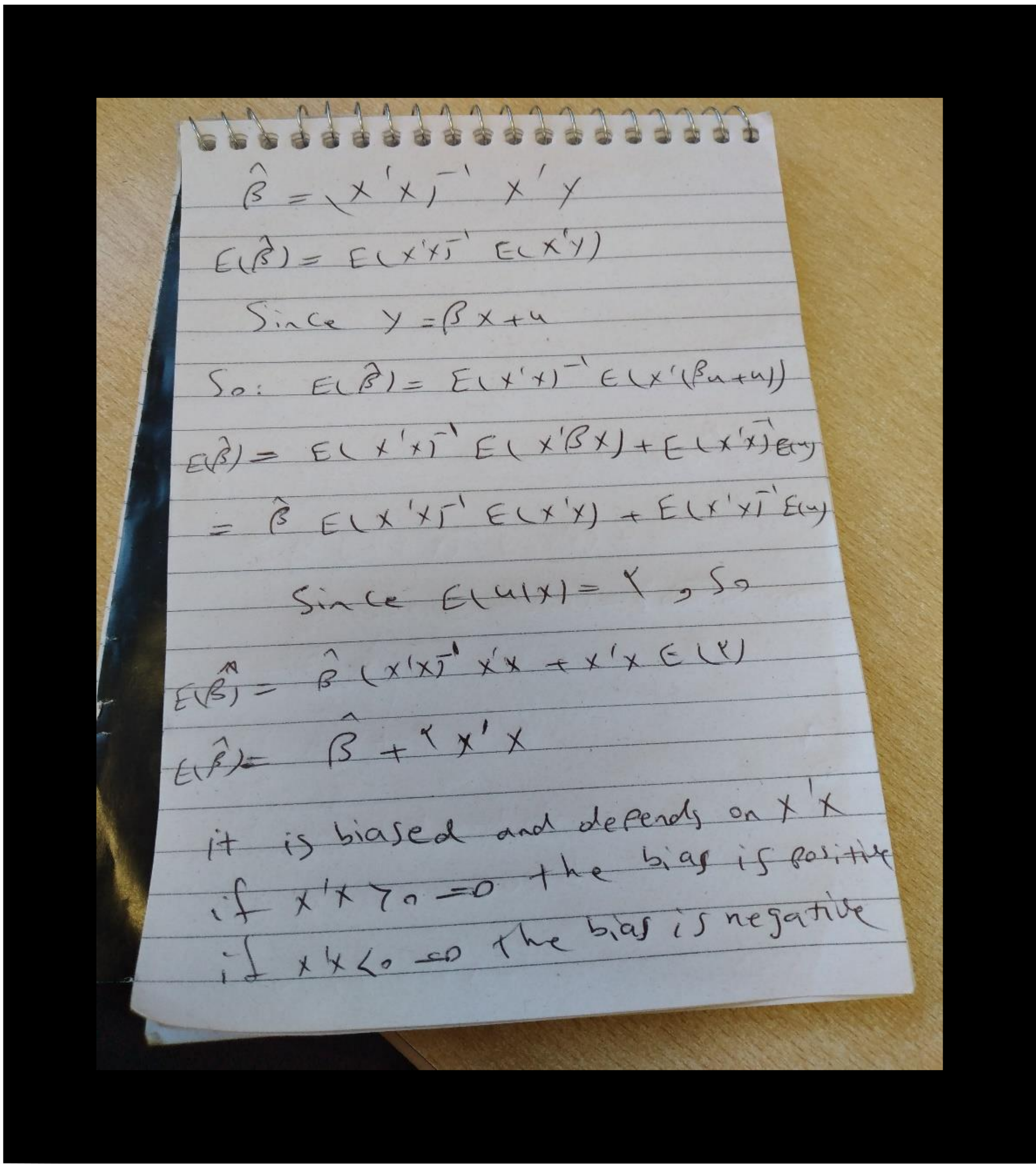
$$= \sum x_i y_i - \sum x_i x_i' (\sum x_i x_i')^{-1} \cdot (\sum x_i y_i)$$

Since $\sum x_i x_i' (\sum x_i x_i')^{-1} = I$

So we have $\sum x_i \hat{u}_i = \sum x_i y_i - \sum x_i y_i = 0$

2. Assume you are a senior economist and your intern/research assistant is showing you an OLS model ….

- Try to derive a formal expression for E(beta_hat) under E(u|X)= 2



PART C: Coding

1.) Do at least two of the exercises from Lab 06 (preferably all). I did 1 and 2 and posted in the blog, but for being sure, I put the Markdown of the two problems in the attachments. (LAB-06-problem-set1 and LAB06-problem2)
2.) The same (LAB 07)
3.) Part 2: Testing: (The same) (PartC-Q3-Part2)