

TRUE POPULATION MODEL (UNKNOWN / NOT OBSERVABLE)

$$Y_n = a + bX_n + \epsilon_n \quad \text{REGRESSION MODEL}$$

Y_n, X_n : Random Variables - OBSERVED (FROM A SAMPLE)

ϵ_n : Random Variable - NOT OBSERVED (ERROR TERM)

a, b : UNKNOWN PARAMETERS TO BE ESTIMATED FROM A SAMPLE

GOAL: $\hat{a}, \hat{b}$: ESTIMATES OF a and b FROM A SAMPLE
 $n = 1, \dots, N$

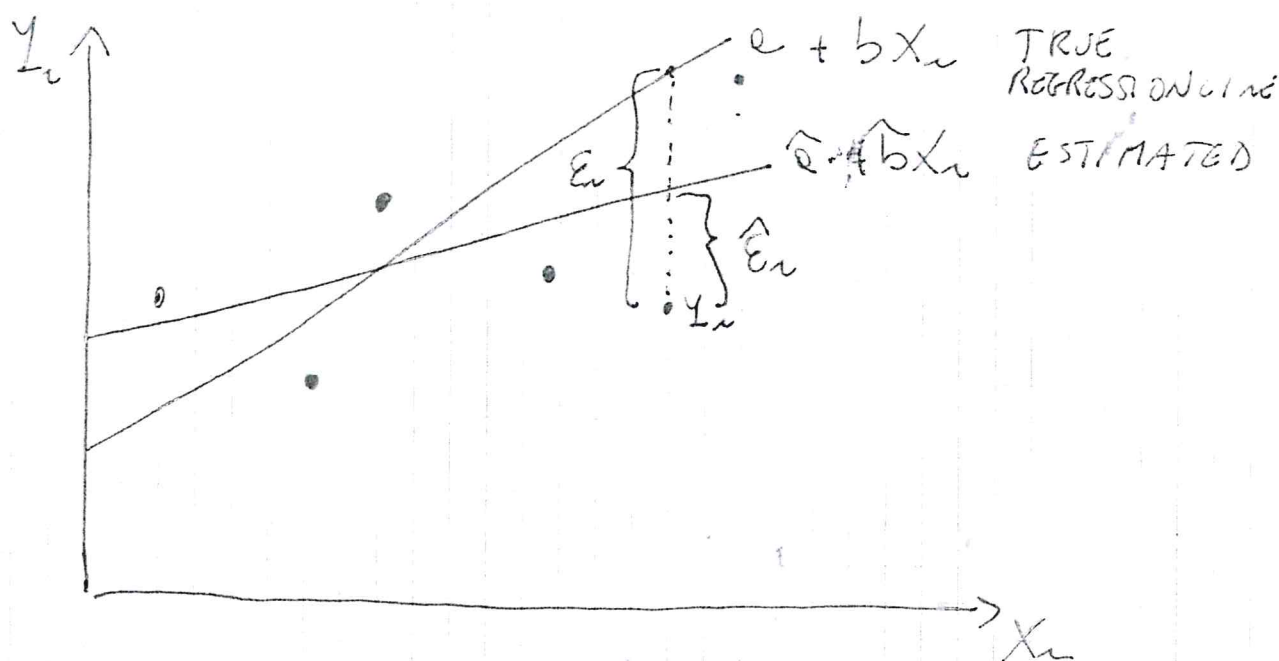
by definition: $\epsilon_n = Y_n - a - bX_n$

LET US DEFINE: $\hat{Y}_n = \hat{a} + \hat{b}X_n$

$$\hat{\epsilon}_n = Y_n - \hat{Y}_n$$

$$= Y_n - \hat{a} - \hat{b}X_n$$

= RESIDUAL



2
HOW TO ESTIMATE a and b ?

OLS
ORDINARY LEAST SQUARE

I have defined $\hat{e}_n = y_n - \hat{a} - \hat{b} x_n$

OLS are defined as:

$$\text{Min}_{\hat{a}, \hat{b}} \sum_{n=1}^N \hat{e}_n^2$$

$$\text{Min}_{\hat{a}, \hat{b}} \sum_{n=1}^N (y_n - \hat{a} - \hat{b} x_n)^2$$

F.O.C.

$$\left\{ \begin{array}{l} \textcircled{1} \frac{\partial \sum_{n=1}^N (y_n - \hat{a} - \hat{b} x_n)^2}{\partial \hat{a}} = 0 \\ \textcircled{2} \frac{\partial \sum_{n=1}^N (y_n - \hat{a} - \hat{b} x_n)^2}{\partial \hat{b}} = 0 \end{array} \right.$$

$$\textcircled{1} -2 \sum_{n=1}^N (y_n - \hat{a} - \hat{b} x_n) = 0$$

$$\textcircled{2} -2 \sum_{n=1}^N x_n (y_n - \hat{a} - \hat{b} x_n) = 0$$

$$\textcircled{1} \quad \sum_{i=1}^N (Y_i - \hat{\alpha} - \hat{\beta} X_i) = 0$$

$$\sum_{i=1}^N Y_i - \hat{\alpha} N - \hat{\beta} \sum_{i=1}^N X_i = 0$$

$$\hat{\alpha} = \frac{1}{N} \sum_{i=1}^N Y_i - \hat{\beta} \frac{1}{N} \sum_{i=1}^N X_i$$

$$\Rightarrow \hat{\alpha} = \overline{Y} - \hat{\beta} \overline{X}$$

MEAN Y MEAN X

$$\text{also: } \overline{Y} = \hat{\alpha} + \hat{\beta} \overline{X}$$

by (initial) definition:

$$Y_i = \hat{\alpha} + \hat{\beta} X_i + \hat{\epsilon}_i$$

from the definition of $\hat{\alpha} = \overline{Y} - \hat{\beta} \overline{X}$:

$$Y_i = \overline{Y} - \hat{\beta} \overline{X} + \hat{\beta} X_i + \hat{\epsilon}_i$$

$$(Y_i - \overline{Y}) = \hat{\beta} (X_i - \overline{X}) + \hat{\epsilon}_i$$

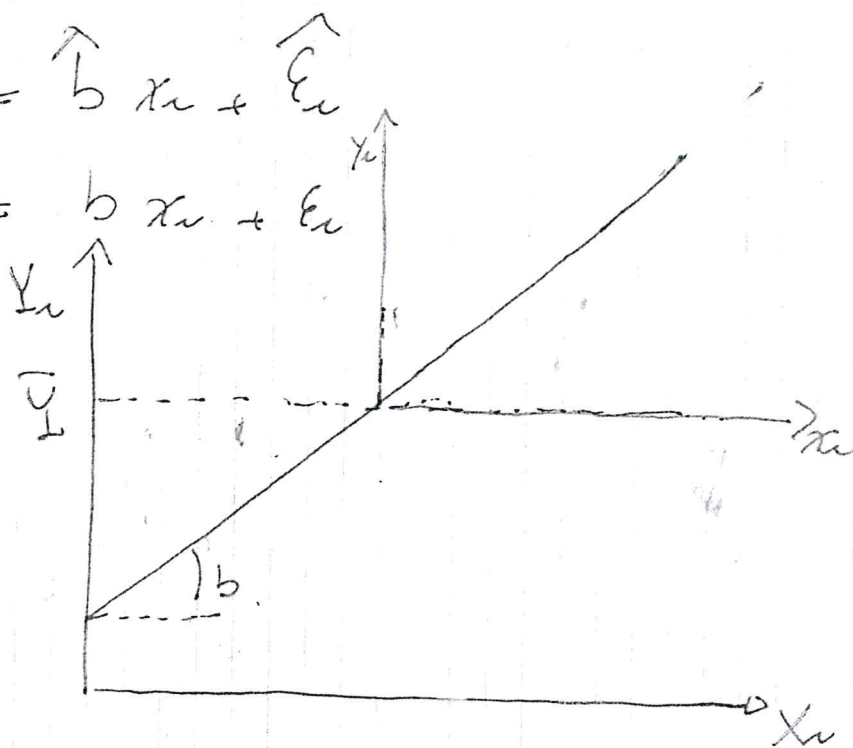
OXES TRANSLATION:

$$\text{and } Y_i = \hat{\beta} X_i + \hat{\epsilon}_i$$

$$y_i = b x_i + \epsilon_i$$

$$y_i = (Y - \overline{Y})$$

$$x_i = (X - \overline{X})$$



I

4

$$y_i = b x_i + \epsilon_i$$

$$(2) \quad -2 \sum_{i=1}^N x_i (y_i - \hat{b} x_i) = 0$$

$$\sum_{i=1}^N x_i (y_i - \hat{b} x_i) = 0$$

$$\sum_{i=1}^N x_i y_i - \hat{b} \sum_{i=1}^N x_i^2 = 0$$

$$\hat{b} = \frac{\sum_{i=1}^N x_i y_i}{\sum_{i=1}^N x_i^2}$$

$$= \frac{\frac{1}{N} \sum_{i=1}^N [(x_i - \bar{x})(y_i - \bar{y})]}{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$$

$$= \frac{\text{cov}(X, Y)}{\text{VAR}(X)}$$

I

F.O.C. OLS

5:

$$\textcircled{1} \quad \frac{\partial \sum_{i=1}^n (y_i - \hat{a} - \hat{b}x_i)^2}{\partial \hat{a}} = 0$$

$$\Rightarrow \sum_{i=1}^n (y_i - \hat{a} - \hat{b}x_i) = 0$$

$$\Rightarrow \sum_{i=1}^n \hat{e}_i = 0$$

$$\textcircled{2} \quad \frac{\partial \sum_{i=1}^n (y_i - \hat{a} - \hat{b}x_i)^2}{\partial \hat{b}} = 0$$

$$\Rightarrow \sum_{i=1}^n x_i (y_i - \hat{a} - \hat{b}x_i) = 0$$

$$\Rightarrow \sum_{i=1}^n x_i \hat{e}_i = 0$$

$$\left\{ \begin{array}{l} \sum_{i=1}^n \hat{e}_i = 0 \\ \sum_{i=1}^n x_i \hat{e}_i = 0 \end{array} \right.$$

NORMAL EQUATIONS
SYSTEM

S.O.C

6

HESSIAN MATRIX DEFINITE POSITIVE
(for being minimizing)

$$H = \begin{pmatrix} \frac{\partial^2 \sum_{i=1}^N \hat{\epsilon}_i^2}{\partial \hat{\alpha}^2} & \frac{\partial^2 \sum_{i=1}^N \hat{\epsilon}_i^2}{\partial \hat{\alpha} \hat{\beta}} \\ \frac{\partial^2 \sum_{i=1}^N \hat{\epsilon}_i^2}{\partial \hat{\alpha} \hat{\beta}} & \frac{\partial^2 \sum_{i=1}^N \hat{\epsilon}_i^2}{\partial \hat{\beta}^2} \end{pmatrix}$$

$$H = \begin{pmatrix} 2N & 2 \sum_{i=1}^N x_i \\ 2 \sum_{i=1}^N x_i & 2 \sum_{i=1}^N x_i^2 \end{pmatrix}$$

In general, we have to study eigenvalues,
they should be > 0 for a definite POSITIVE MATRIX

2x2 Matrix: — The elements in the diagonal > 0

— Det > 0

• Elements in the diagonal are > 0

$$\begin{aligned} |H| &= 2N \cdot 2 \sum_{i=1}^N x_i^2 - 4 \left(\sum_{i=1}^N x_i \right)^2 = 4N \sum_{i=1}^N x_i^2 - 4(N\bar{x})^2 \\ &= 4N \left(\sum_{i=1}^N x_i^2 - N\bar{x}^2 \right) = 4N \sum_{i=1}^N (x_i - \bar{x})^2 > 0 \end{aligned}$$

II

$$Y_i = a + b X_i + \epsilon_i \quad \dots \text{TRUE (POPULATION) REGRESSION MODEL}$$

$$\hat{Y}_i = \hat{a} + \hat{b} X_i \quad \dots \text{ESTIMATED REGRESSION}$$

$$\begin{aligned} \hat{\epsilon}_i &= Y_i - \hat{Y}_i \\ &= Y_i - \hat{a} - \hat{b} X_i \end{aligned} \quad \dots \text{RESIDUALS}$$

OLS Objective function:

$$\text{Min}_{\hat{a}, \hat{b}} \sum_{i=1}^N \hat{\epsilon}_i^2$$

$$\text{Min}_{\hat{a}, \hat{b}} \sum_{i=1}^N (Y_i - \hat{a} - \hat{b} X_i)^2$$

$$\text{① } \frac{\partial \sum_{i=1}^N \hat{\epsilon}_i^2}{\partial \hat{a}} = 0 \quad \Rightarrow \quad \sum_{i=1}^N \hat{\epsilon}_i = 0$$

$$\text{② } \frac{\partial \sum_{i=1}^N \hat{\epsilon}_i^2}{\partial \hat{b}} = 0 \quad \Rightarrow \quad \sum_{i=1}^N X_i \hat{\epsilon}_i = 0$$

NORMAL
EQUATIONS
SYSTEM

$$H = \begin{pmatrix} \frac{\partial^2 \sum_{i=1}^N \hat{\epsilon}_i^2}{\partial \hat{a}^2} & \frac{\partial^2 \sum_{i=1}^N \hat{\epsilon}_i^2}{\partial \hat{a} \partial \hat{b}} \\ \frac{\partial^2 \sum_{i=1}^N \hat{\epsilon}_i^2}{\partial \hat{a} \partial \hat{b}} & \frac{\partial^2 \sum_{i=1}^N \hat{\epsilon}_i^2}{\partial \hat{b}^2} \end{pmatrix} = \begin{pmatrix} N & \sum_{i=1}^N X_i \\ \sum_{i=1}^N X_i & \sum_{i=1}^N X_i^2 \end{pmatrix}$$

DIAGONAL ELEMENTS $(N, \sum_{i=1}^N X_i^2) > 0$

$$|H| > 0$$

So H IS DEFINITE POSITIVE

$$\hat{a} = \bar{Y} - \hat{b} \bar{X}$$

$$\hat{b} = \frac{\sum_{i=1}^N [(X_i - \bar{X})(Y_i - \bar{Y})]}{\sum_{i=1}^N (X_i - \bar{X})^2} = \frac{\sum_{i=1}^N X_i Y_i}{\sum_{i=1}^N Y_i^2}$$

SIMPLE LINEAR REGRESSION MODEL

$$(1) \quad Y_i = a + b X_i + \epsilon_i \quad \text{"TRUE" MODEL}$$

$$\Rightarrow y_i = b x_i + \epsilon_i$$

$$(2) \quad E(\epsilon_i | X_i) = E(\epsilon_i)$$

$$(3) \quad E(\epsilon_i) = 0$$

$$(2) + (3) \quad E(\epsilon_i | X_i) = 0 \quad \text{ZERO CONDITIONAL MEAN ASSUMPTION}$$

$$(4) \quad E(\epsilon_i)^2 = \sigma^2$$

HOMOGENEITY

(Variance does not depend on i)

$$(5) \quad E(\epsilon_i \epsilon_j) = 0 \quad \forall i \neq j \quad \text{(CROSS-SECTIONAL)} \\ \text{INDEPENDENCE}$$

Main goal: estimation of b , which is the SLOPE of the regression line and also

$$\frac{\partial Y_i}{\partial X_i} = b$$

CETERIS PARIBUS INTERPRETATION

The estimation of a is generally less important

WHAT ABOUT THE OLS ESTIMATOR $\hat{b}$?

Recap: $\hat{b} = \frac{\sum_{i=1}^N x_i y_i}{\sum_{i=1}^N x_i^2}$ is the OLS estimator of b

$\hat{b}$ = "a given value" is the OLS estimate of b

$$\hat{b} = \frac{\sum_{i=1}^N x_i y_i}{\sum_{i=1}^N x_i^2}$$

HIP ZBIS: X DETERMINISTIC VARIABLE
(NOW STOCHASTIC)

LET US DEFINE WEIGHTS: $w_i = \frac{x_i}{\sum_{i=1}^N x_i^2}$

I CAN REWRITE:

$$\hat{b} = \sum_{i=1}^N w_i y_i = \sum_{i=1}^N w_i (b x_i + \epsilon_i) \Rightarrow \text{from ①}$$

$\hat{b}$ = WEIGHTED SUMMATION OF THE OBSERVATIONS, y_i

$$\hat{b} = b \sum_{i=1}^N w_i x_i + \sum_{i=1}^N w_i \epsilon_i$$

$$\hat{b} = b \frac{\sum_{i=1}^N x_i^2}{\sum_{i=1}^N x_i^2} + \sum_{i=1}^N w_i \epsilon_i$$

$$\hat{b} = b + \sum_{i=1}^N w_i \epsilon_i$$

$\hat{b}$ is a random variable which - ex ante - depends on

- b (a parameter)

- a weighted summation of random variables

WHICH SHOWS UNBIASEDNESS OF $\hat{b}$

Recap : a "generic" estimator $\hat{b}$ is unbiased if

$$\mathbb{E}(\hat{b}) = b$$

$$\text{BIAS}(\hat{b}) = \mathbb{E}(\hat{b}) - b$$

$$\hat{b} = b + \sum_{i=1}^n w_i \epsilon_i$$

$$\mathbb{E}(\hat{b}) = b + \mathbb{E}\left(\sum_{i=1}^n w_i \epsilon_i\right)$$

* HP 2 BIS : $w_i = \frac{x_i}{\sum_{i=1}^n x_i^2}$ are FIXED WEIGHTS (NOT RANDOM)

$$\hookrightarrow \mathbb{E}(\hat{b}) = b + \sum_{i=1}^n w_i \mathbb{E}(\epsilon_i)$$

* HP 3 : $\mathbb{E}(\epsilon_i) = 0$

$$\hookrightarrow \mathbb{E}(\hat{b}) = b + \sum_{i=1}^n w_i \cdot 0$$

$$\mathbb{E}(\hat{b}) = b$$

$\hat{b}$ IS AN UNBIASED ESTIMATOR OF b

II

5

Variance of $\hat{b}$

$$\text{Var}(\hat{b}) = \mathbb{E}(\hat{b} - \mathbb{E}(\hat{b}))^2$$

$$= \mathbb{E}(\hat{b} - b)^2$$

We also know that: $\hat{b} = b + \sum_{i=1}^N w_i \epsilon_i$

$$(\hat{b} - b) = \sum_{i=1}^N w_i \epsilon_i$$

$$\text{Var}(\hat{b}) = \mathbb{E}\left(\sum_{i=1}^N w_i \epsilon_i\right)^2$$

$$= \mathbb{E}\left[w_1^2 \epsilon_1^2 + w_2^2 \epsilon_2^2 + \dots + w_N^2 \epsilon_N^2 + 2(w_1 \epsilon_1 w_2 \epsilon_2 + \dots + w_{N-1} \epsilon_{N-1} w_N \epsilon_N)\right]$$

HP 4 $\rightarrow \mathbb{E}(\epsilon_i)^2 = \sigma^2$

HP 5 $\rightarrow \mathbb{E}(\epsilon_i \epsilon_j) = 0$

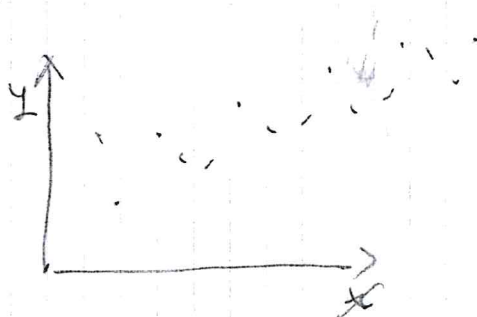
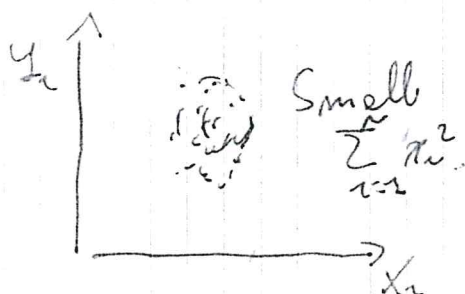
$$\text{Var}(\hat{b}) = \sigma^2 \sum_{i=1}^N w_i^2 + 0$$

$$\text{Var}(\hat{b}) = \sigma^2 \sum_{i=1}^N w_i^2$$

What about $\sum_{i=1}^N w_i^2$?

$$\sum_{i=1}^N w_i^2 = \sum_{i=1}^N \left(\frac{x_i}{\sum_{i=1}^N x_i^2} \right)^2 = \frac{\sum_{i=1}^N x_i^2}{\left(\sum_{i=1}^N x_i^2 \right)^2} = \frac{1}{\sum_{i=1}^N x_i^2}$$

$$\text{Var}(\hat{b}) = \frac{\sigma^2}{\sum_{i=1}^N x_i^2}$$



4

OLS

$$\text{Min} \sum_{i=1}^n \hat{\epsilon}_i^2$$

6

$$(A) \hat{b} = \sum_{i=1}^n w_i y_i$$

$$w_i = \frac{x_i}{\sum_{i=1}^n x_i^2}$$

$$(B) E(\hat{b}) = b \quad \text{HP 2+3}$$

$$(C) \text{Var}(\hat{b}) = \frac{\sigma^2}{\sum_{i=1}^n x_i^2} \quad \text{HP 4+5}$$

GAUSS-MARKOV THEOREM:

UNDER ALL THE HYPOTHESES WE HAVE DONE,
THE ESTIMATOR $\hat{b}$ IS BLUE

BEST LINEAR UNBIASED ESTIMATOR

BEST ??

↳ MINIMUM VARIANCE (AMONG THE UNBIASED ESTIMATORS)

II

ALTERNATIVE PROOF OF $E(\hat{\beta}) = \beta$

4

We already know that:

UNDER HP2 INSTEAD OF HP2BIS

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$$

$$y_i = \beta x_i + \varepsilon_i$$

$$\hat{\beta} = \beta + \frac{\sum_{i=1}^n x_i \varepsilon_i}{\sum_{i=1}^n x_i^2}$$

$$\varepsilon_i = \frac{x_i}{\sum_{i=1}^n x_i^2}$$

$$\hat{\beta} = \beta + \frac{\sum_{i=1}^n x_i \varepsilon_i}{\sum_{i=1}^n x_i^2}$$

$$E(\hat{\beta}) = \beta + E\left(\frac{\sum_{i=1}^n x_i \varepsilon_i}{\sum_{i=1}^n x_i^2}\right)$$

USE LAW OF ITERATED EXPECTATIONS (LIE)

 $X = \text{R.V.}$; $x_i = \text{REALISATION}$

$$E(Y) = \sum_{i=1}^n E(Y|X=x_i) \cdot P(X=x_i)$$

$E(Y) =$ WEIGHTED AVERAGE - WITH WEIGHTS EQUAL TO THE PROBABILITY OF A GIVEN REALISATION OF X - OF CONDITIONAL EXPECTED VALUES OF Y GIVEN $X=x_i$

$$E(Y) = E[E(Y|X)]$$

$E(Y) =$ EXPECTED VALUE OF THE EXPECTED VALUE OF $Y|X$

II

$$\text{LIE: } E(Y) = E[E(Y|X)]$$

8

Y is a "generic" random variable
I can write:

$$E(E) = E[E(E|X)]$$

Consequently, from $E(\hat{\beta}) = b + E\left(\frac{\sum_{i=1}^n x_i e_i}{\sum_{i=1}^n x_i^2}\right) :$

$$E\left(\frac{\sum_{i=1}^n x_i e_i}{\sum_{i=1}^n x_i^2}\right) = E\left(\frac{\sum_{i=1}^n x_i E(e_i|x_i)}{\sum_{i=1}^n x_i^2}\right)$$

From H.P (2) and (3) :

$$E(e_i | x_i) = 0$$

$$E(\hat{\beta}) = b + 0$$

$$E(\hat{\beta}) = b$$

II

PROOF THAT $\hat{b}$ IS BLUE

9

We have to show that $\hat{b}$ has the lowest variance• Let us define a "general" linear estimator:

$$\tilde{b} = \sum_{i=1}^n c_i y_i \quad y_i = x_i b + \epsilon_i$$

$$= b \sum_{i=1}^n c_i x_i + \sum_{i=1}^n c_i \epsilon_i$$

 c_i are deterministic weights (non stochastic) $\tilde{b}$ is a random variable !! Why??

$$\mathbb{E}(\tilde{b}) = b \sum_{i=1}^n c_i x_i + \sum_{i=1}^n c_i \mathbb{E}(\epsilon_i)$$

(c_i DETERMINISTIC WEIGHTS)

$\mathbb{E}(\epsilon_i) = 0$
HIP 3

So, To have $\mathbb{E}(\tilde{b}) = b$ We need: $\boxed{\sum_{i=1}^n c_i x_i = 1}$ CONDITION FOR UNBIASEDNESS

$$\text{Var}(\tilde{b}) = \mathbb{E} \left(\tilde{b} - \mathbb{E}(\tilde{b}) \right)^2$$

$$= \mathbb{E} \left(\sum_{i=1}^n c_i \epsilon_i \right)^2$$

$$= \sigma^2 \sum_{i=1}^n c_i^2$$

HIP (4)

II

IN ORDER $\hat{b}$ TO BE BLUE:

10

$$\text{Var}(\hat{b}) = \sigma^2 \sum_{i=1}^n c_i^2$$

$$\text{Min} \sum_{i=1}^n c_i^2$$

$$\text{v.c.} \quad \sum_{i=1}^n c_i x_i = 1$$

CONDITION FOR
 $E(\hat{b}) = b$

$$\mathcal{L} = \sum_{i=1}^n c_i^2 - 2\lambda \left(\sum_{i=1}^n c_i x_i - 1 \right)$$

do only for simplify algebra

F.O.C

$$\frac{\partial \mathcal{L}}{\partial c_i} = 0 \Rightarrow 2c_i - 2\lambda x_i = 0$$

$$c_i = \lambda x_i$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = 0 \Rightarrow -2 \left(\sum_{i=1}^n c_i x_i - 1 \right) = 0$$

$$\sum_{i=1}^n c_i x_i = 1$$

Solving the system:

$$\sum_{i=1}^n c_i x_i = 1$$

$$\sum_{i=1}^n (\lambda x_i) x_i = 1$$

$$\lambda \sum_{i=1}^n x_i^2 = 1 \Rightarrow \lambda = \frac{1}{\sum_{i=1}^n x_i^2}$$

$$c_i = \frac{x_i}{\sum_{i=1}^n x_i^2}$$

$$\Rightarrow c_i = w_i \quad !!!$$

if $c_i = w_i \Rightarrow \hat{b}$ is BLUE
and $\text{Var}(\hat{b}) = \sigma^2 / \sum_{i=1}^n x_i^2$

GOAL: TO DEFINE OLS FROM ANOTHER POINT OF VIEW:
FROM THE SPACE $\mathbb{R}^N$ WHICH IS CALLED
THE SPACE OF OBSERVATIONS

Recap: let x be a vector belonging from $\mathbb{R}^N$

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix}$$

- $\|x\|^2$ is the square of the EUCLIDEAN NORM

$$\|x\| = (x'x)^{1/2} \quad \text{EUCLIDEAN NORM}$$

$$\|x\|^2 = x'x = \sum_{i=1}^N x_i^2$$

- TWO "NAIVE" DEFINITIONS:

→ VECTOR SPACE: COLLECTION OF VECTORS WHICH
CAN BE ADDED TOGETHER AND MULTIPLIED
BY SCALARS

$\mathbb{R}^N$ IS A VECTOR SPACE

→ VECTOR SUBSPACE: VECTOR SPACE THAT IS
A SUBSET OF ANOTHER - HIGHER DIMENSION -
VECTOR SPACE

THEOREM

Let L be a VECTOR SUBSPACE belonging from $\mathbb{R}^N$, the minimum distance in terms of euclidean norm between a vector $Y \in \mathbb{R}^N$ and L is given by the ORTHOGONAL PROJECTION OF Y ONTO L

- VECTOR REPRESENTATION OF THE SIMPLE LINEAR REGRESSION MODEL

Let define the following vectors belonging from $\mathbb{R}^N$

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix}_{N \times 1} \quad X = \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix}_{N \times 1} \quad E = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_N \end{pmatrix}_{N \times 1} \quad \hat{Y} = \begin{pmatrix} \hat{Y}_1 \\ \vdots \\ \hat{Y}_N \end{pmatrix}_{N \times 1} \quad \hat{E} = \begin{pmatrix} \hat{\epsilon}_1 \\ \vdots \\ \hat{\epsilon}_N \end{pmatrix}_{N \times 1}$$

Let define also:

$$1_N = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{(N \times 1)}$$

- I can rewrite $Y_i = a + b X_i + \epsilon_i$ as:

$$Y_{N \times 1} = 1_N a_{1 \times 1} + X_{N \times 1} b_{1 \times 1} + E_{N \times 1}$$

$$\hat{Y}_{N \times 1} = 1_N \hat{a}_{1 \times 1} + X_{N \times 1} \hat{b}_{1 \times 1}$$

$$Y_{N \times 1} = \hat{Y}_{N \times 1} + \hat{E}_{N \times 1}$$

$$\hat{E}_{N \times 1} = Y_{N \times 1} - \hat{Y}_{N \times 1}$$

$$\hat{a}, \hat{b} \quad \underline{\text{OLS ESTIMATES}}$$

- "Implicit" hypothesis of OLS (from $y_i = b x_i + \epsilon_i$)
 - $\sum_{i=1}^N x_i^2 \neq 0$ (for the existence of $\hat{b}$)

this because $\hat{b} = \frac{\sum_{i=1}^N x_i y_i}{\sum_{i=1}^N x_i^2}$

- Let us revisit this hypothesis in a different way:

~~$\exists$~~ $\lambda \in \mathbb{R}$ such that $x = \lambda x_m$

the vectors x and x_m are LINEARLY INDEPENDENT

■ If this hypothesis is true:

→ x and x_m define a vector subspace from $\mathbb{R}^N$ of dimension 2 that is, a plane noted as $\mathcal{L}(x, x_m)$

→ $\hat{y}$ is by definition a linear combination

$$\hat{y} = x_m \hat{\alpha} + x \hat{\beta}$$

thus:

$$\hat{y} \in \mathcal{L}(x, x_m)$$

$$y \in \mathbb{R}^N$$

- GOAL: TO MINIMIZE THE DISTANCE BETWEEN y and $\hat{y}$ USING THE METRIC OF THE EUCLIDEAN NORM (OR. EQUALLY, THE SQUARE OF THE NORM)

III

4

$$\text{Min } \| \underline{y} - \hat{\underline{y}} \|$$

$$\text{Min } \| \hat{\underline{e}} \|$$

or

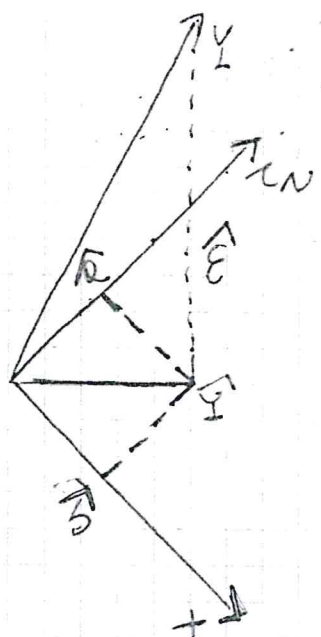
$$\text{Min } \| \underline{y} - \hat{\underline{y}} \|^2$$

$$\text{Min } \| \hat{\underline{e}} \|^2 \Leftrightarrow \text{Min } \sum_{i=1}^N \hat{e}_i^2$$

then, I can apply the previous THEOREM:

THE MINIMUM DISTANCE ($\text{Min } \| \hat{\underline{e}} \|^2$) is
OBTAINED DEFINING

$\hat{\underline{y}}$ as ORTHOGONAL PROJECTION OF $\underline{y}$ ONTO $\mathcal{L}(X, \lambda_N)$



The COORDINATES on
The PLANE $\mathcal{L}(X, \lambda_N)$
on $\hat{\lambda}_N$ and $\hat{X}$

$\hat{\underline{y}}$ is the ORTHOGONAL PROJECTION OF $\underline{y}$ ($\underline{y} \in \mathbb{R}^N$)
ONTO THE PLANE $\mathcal{L}(X, \lambda_N)$

$$\hat{\underline{y}} = \lambda_N \hat{\underline{e}} + X \hat{\underline{b}}$$

$$\hat{\underline{y}} = \hat{\underline{y}} + \hat{\underline{e}}$$

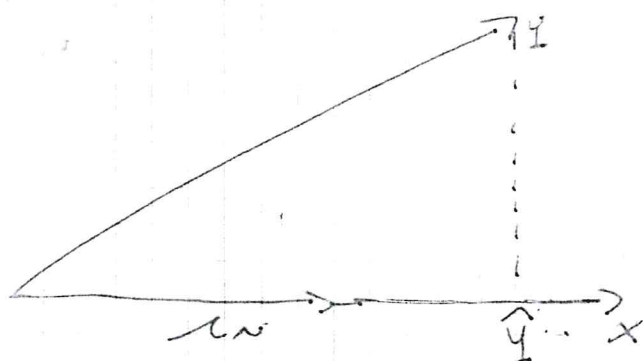
$$\hat{\underline{e}} \perp \mathcal{L}(X, \lambda_N) \Rightarrow \begin{matrix} \hat{\underline{e}} \perp X \\ \hat{\underline{e}} \perp \lambda_N \end{matrix}$$

$$\begin{cases} \hat{\underline{e}}' \lambda_N = 0 \\ \hat{\underline{e}}' X = 0 \end{cases} \quad \text{NORMAL EQUATIONS SYSTEM}$$

III

5

- What happens if $\exists \lambda$ such that $X = \lambda \mu$



$\hat{x}$ and $\hat{y}$ are NOT DETERMINED

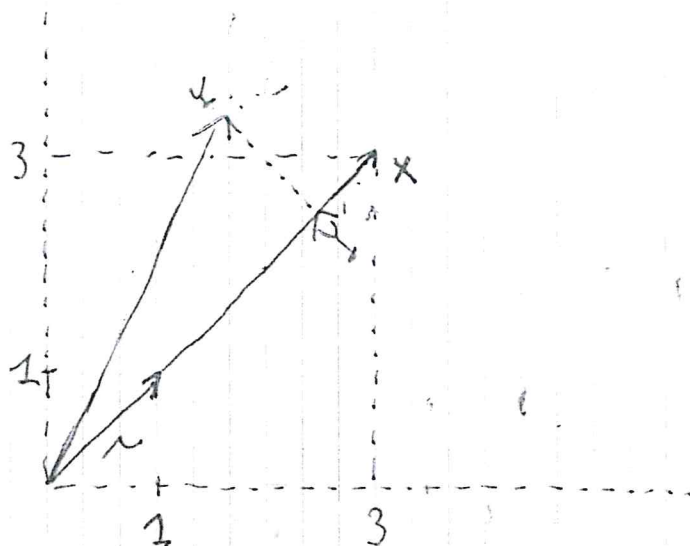
- WHEN IT HAPPENS??

X IS CONSTANT (Remember $\sum_{i=1}^n x_i^2 = 0 \Rightarrow x_i$ constant)

EXAMPLE in $\mathbb{R}^2$

$$\mu_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad X = \begin{pmatrix} 3 \\ 3 \end{pmatrix}$$

$$X = 3\mu_2$$



(iv)

 $\hat{b}$ (OLS) IS BLUE

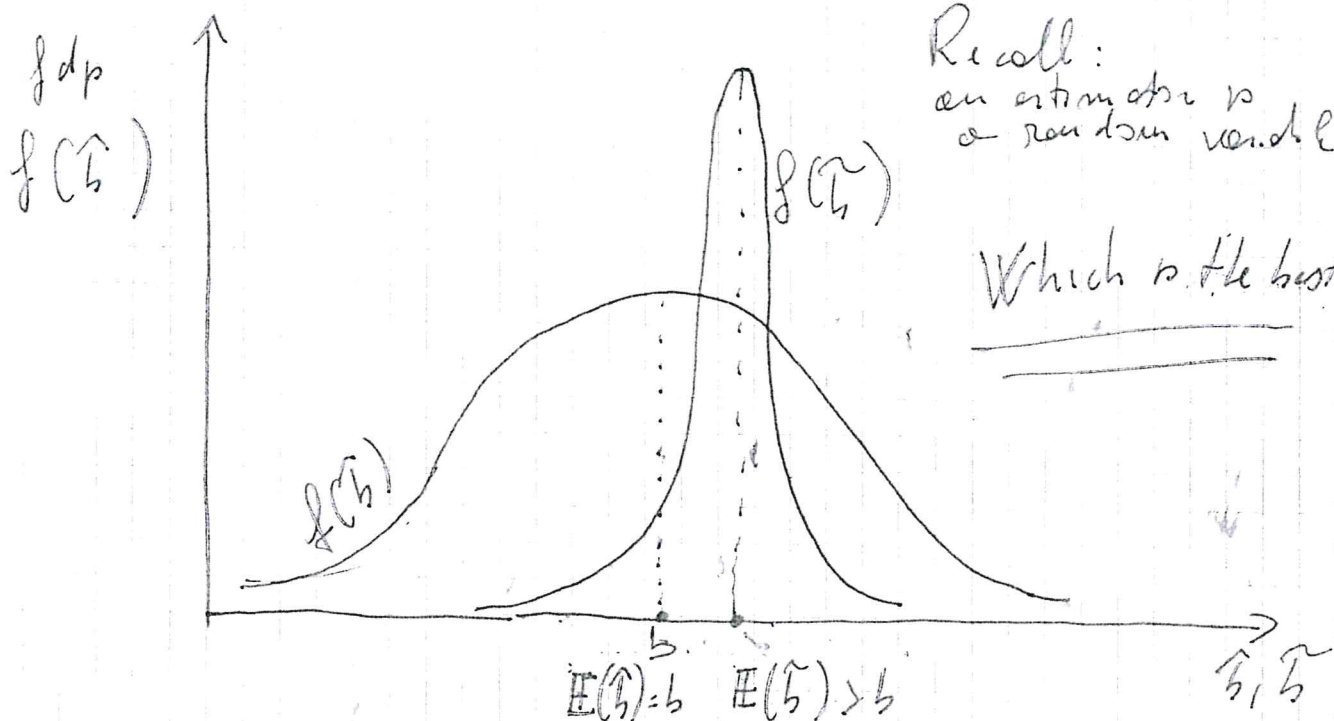
1

Min. Variance among all the UNBIASED estimators

□ POSSIBLE TRADE-OFF BETWEEN
UNBIASEDNESS - AN EFFICIENCY

UNBIASED ESTIMATOR: $E(\hat{b}) = b$ BIAS: $E(\hat{b}) - b \neq 0$

$E(\hat{b}) - b \begin{cases} > 0 \text{ OVER ESTIMATE} \\ < 0 \text{ UNDER ESTIMATE} \end{cases}$

Let suppose we have two alternative estimators of b : $\hat{b}$ is UNBIASED $\Rightarrow E(\hat{b}) = b$ $\tilde{b}$ is BIASED (OVER ESTIMATE) $\Rightarrow E(\tilde{b}) > b$ but $\text{Var}(\hat{b}) > \text{Var}(\tilde{b})$ 

Let introduce a Loss function

I call this loss function as Mean Square Error

$$MSE(\tilde{b}) = E(\tilde{b} - b)^2$$

Please note that it is close to the variance:

$$Var(\tilde{b}) = E(\tilde{b} - E(\tilde{b}))^2$$

A possible way to choose between $\hat{b}$ and $\tilde{b}$ is choosing the one with lower MSE

According to the MSE a biased estimator can be better than an unbiased one but with lower variance:

I can rewrite:

$$MSE(\tilde{b}) = E(\underbrace{\tilde{b} - E(\tilde{b})}_{\text{variance}} + \underbrace{E(\tilde{b}) - b}_{\text{bias}})^2$$

$$= Var(\tilde{b}) + bias^2(\tilde{b}) + 2bias(\tilde{b}) \underbrace{[E(\tilde{b}) - E(\tilde{b})]}_{=0}$$

$$MSE(\tilde{b}) = Var(\tilde{b}) + bias^2(\tilde{b})$$

IV

 $\hat{\beta}_{OLS}$

3

$$MSE(\hat{\beta}) = \text{Var}(\hat{\beta}) + 0$$

$\hat{\beta}$ has minimum Variance among all the unbiased estimators

$\Rightarrow \hat{\beta}$ has minimum MSE among all the unbiased estimators

However...

there may exist an estimator $\tilde{\beta}$ which is biased but:

$$MSE(\tilde{\beta}) < MSE(\hat{\beta})$$

this happens if

$$\text{Var}(\tilde{\beta}) < \text{Var}(\hat{\beta}) - \text{bias}^2(\tilde{\beta})$$

$$(1) Y_i = a + b X_i + \epsilon_i \quad \text{or} \quad y_i = b x_i + \epsilon_i$$

$$(2) (3) (4) (5) \Rightarrow \hat{b} \text{ (OLS) BLUE}$$

$$(6) \epsilon_i \sim N(0, \sigma^2) \quad \text{NORMALITY}$$

$\hookrightarrow$ INDEPENDENTLY

It means that every ϵ_i is ..

Normal on all the i are

independent, that is $E(\epsilon_i, \epsilon_j) = 0$

We already know that:

$$\hat{b} = b + \sum_{i=1}^N w_i \epsilon_i \quad w_i = \frac{x_i}{\sum_{i=1}^N x_i^2}$$

A feature of the Normal distribution is that:

A linear combination of Normal Random Variables is also a Normal distribution:

$$\hookrightarrow \text{If } \epsilon_i \sim N \Rightarrow \hat{b} = b + \sum_{i=1}^N w_i \epsilon_i \sim N$$

$$\hat{b} \sim N\left(b, \frac{\sigma^2}{\sum_{i=1}^N x_i^2}\right)$$

$$(\hat{b} - b) \sim N\left(0, \frac{\sigma^2}{\sum_{i=1}^N x_i^2}\right)$$

IV

WHAT ABOUT σ^2

3

 σ^2 is the variance of ϵ_i

$$\begin{aligned} E(\epsilon_i)^2 &= E(\epsilon_i - E(\epsilon_i))^2 \\ &= E(\epsilon_i - 0)^2 \\ &= \sigma^2 \end{aligned}$$

HOWEVER ϵ_i IS NOT KNOWN
 thus σ^2 is not observable

A UNBIASED ESTIMATOR OF σ^2 is s^2

$$s^2 = \frac{\sum_{i=1}^N \hat{\epsilon}_i^2}{N-2}$$

$N-2$ degrees of freedom
 2 constraints

$$\sum_{i=1}^N \hat{\epsilon}_i = 0$$

$$\sum_{i=1}^N x_i \hat{\epsilon}_i = 0$$

$$E(s^2) = \sigma^2$$

s $\Rightarrow$ standard error of regression

IV

TEST OF HYPOTHESIS

6

about b

$$H_0 : b = b_0$$

$$H_1 : b \neq b_0$$

Assumptions:

Random error.

Normality ε

OLS ESTIMATION

 $\Rightarrow$

$$\hat{b}$$

$$\hat{b} =$$

$$\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$$

 $\Rightarrow$

$$\hat{\sigma}_{\hat{b}}^2$$

$$\hat{\sigma}_{\hat{b}}^2 = \frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2}$$

is not known, while

$$\sigma_{\hat{b}}^2 = \frac{\sigma^2}{\sum_{i=1}^n x_i^2}$$

is known

then, the test is

$$\frac{\hat{b} - b_0}{\hat{\sigma}_{\hat{b}}} \sim t_{n-2}$$

Generally, we want to test the following

$$H_0 : b = 0$$

$$H_1 : b \neq 0$$

$$\Rightarrow \frac{\hat{b}}{\hat{\sigma}_{\hat{b}}} \sim t_{n-2}$$

EXTENDING THE SIMPLE LINEAR REGRESSION MODEL (SLRM) TO THE CASE OF MULTIPLE EXPLANATORY VARIABLES

1

■ LET US START WITH 2 EXPLANATORY VARIABLES
→ RELEVANT OMITTED VARIABLES IN THE ESTIMATED MODEL

① → IRRELEVANT INCLUDED VARIABLES IN THE ESTIMATED MODEL

② EASIER UNDERSTANDING OF THE OLS ESTIMATOR WITH RESPECT TO THE CASE OF MORE THAN 2 X, WHILE KEEPING THE MAIN FEATURES

Let us write the model:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \epsilon_i$$

Define - $y_i = Y_i - \bar{Y}$; $x_{2i} = X_{2i} - \bar{X}_2$; $x_{3i} = X_{3i} - \bar{X}_3$

$$\bar{Y} = \beta_1 + \beta_2 \bar{X}_2 + \beta_3 \bar{X}_3 + 0$$

$$Y_i - \bar{Y} = \cancel{\beta_1 - \beta_1} + \beta_2 (X_{2i} - \bar{X}_2) + \beta_3 (X_{3i} - \bar{X}_3) + \epsilon_i$$

$$y_i = \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon_i$$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i}$$

$\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$: OLS ESTIMATES

$$\hat{\epsilon}_i = Y_i - \hat{Y}_i$$

V

$$Y_{2i} = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \epsilon_i$$

2

$$(Y_{2i} = \beta_2 X_{2i} + \beta_3 X_{3i} + \epsilon_i)$$

Objective function: $\min_{\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3} \sum_{i=1}^N \hat{\epsilon}_i^2$

$$\left(\hat{\beta}_j = \underset{\beta_j}{\text{ARG MIN}} \sum_{i=1}^N (Y_{2i} - \beta_1 - \beta_2 X_{2i} - \beta_3 X_{3i})^2 \right)_{j=1,2,3} \quad 0$$

F.O.C.

$$1) \frac{\partial \sum_{i=1}^N \hat{\epsilon}_i^2}{\partial \hat{\beta}_1} = 0$$

$$\Rightarrow -2 \sum_{i=1}^N (Y_{2i} - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}) = 0$$

$$\Rightarrow \sum_{i=1}^N \hat{\epsilon}_i = 0 \iff \hat{\epsilon}_{2N} = 0$$

As in the SLRM: EX POST residuals sum to Zero !!

$$\sum_{i=1}^N Y_{2i} - N \hat{\beta}_1 - \hat{\beta}_2 \sum_{i=1}^N X_{2i} - \hat{\beta}_3 \sum_{i=1}^N X_{3i} = 0$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^N Y_{2i}}{N} - \hat{\beta}_2 \frac{\sum_{i=1}^N X_{2i}}{N} - \hat{\beta}_3 \frac{\sum_{i=1}^N X_{3i}}{N}$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3$$

Recap: In the SLRM $\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X}$

✓

 $\hat{\beta}_2$ and $\hat{\beta}_3$?

3

To simplify calculation, let us focus on y_i

$$2) \frac{\partial \sum_{i=1}^n \hat{\epsilon}_i^2}{\partial \hat{\beta}_2} = 0 \Rightarrow -2 \sum_{i=1}^n (y_i - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i}) x_{2i} = 0$$

$$\Rightarrow \sum_{i=1}^n \hat{\epsilon}_i x_{2i} = 0 \Leftrightarrow \hat{\epsilon}' x_2 = 0$$

$$3) \frac{\partial \sum_{i=1}^n \hat{\epsilon}_i^2}{\partial \hat{\beta}_3} = 0 \Rightarrow -2 \sum_{i=1}^n (y_i - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i}) x_{3i} = 0$$

$$\Rightarrow \sum_{i=1}^n \hat{\epsilon}_i x_{3i} = 0 \Leftrightarrow \hat{\epsilon}' x_3 = 0$$

$$\begin{cases} \sum_{i=1}^n \hat{\epsilon}_i = 0 \\ \sum_{i=1}^n \hat{\epsilon}_i x_{2i} = 0 \\ \sum_{i=1}^n \hat{\epsilon}_i x_{3i} = 0 \end{cases}$$

Normal equations system

$$2) \sum_{i=1}^n y_i x_{2i} - \hat{\beta}_2 \sum_{i=1}^n x_{2i}^2 - \hat{\beta}_3 \sum_{i=1}^n x_{2i} x_{3i} = 0$$

$$3) \sum_{i=1}^n y_i x_{3i} - \hat{\beta}_2 \sum_{i=1}^n x_{2i} x_{3i} - \hat{\beta}_3 \sum_{i=1}^n x_{3i}^2 = 0$$

$$2) \hat{\beta}_2 \sum_{i=1}^n x_{2i}^2 = \sum_{i=1}^n y_i x_{2i} - \hat{\beta}_3 \sum_{i=1}^n x_{2i} x_{3i}$$

$$\Rightarrow \hat{\beta}_2 = \frac{\sum_{i=1}^n y_i x_{2i} - \hat{\beta}_3 \sum_{i=1}^n x_{2i} x_{3i}}{\sum_{i=1}^n x_{2i}^2}$$

✓

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n y_i x_{2i} - \hat{\beta}_3 \sum_{i=1}^n x_{2i} x_{3i}}{\sum_{i=1}^n x_{2i}^2} \quad *$$

4

We can put now the expression above of $\hat{\beta}_2$ in equation 3) $\sum_{i=1}^n y_i x_{3i} - \hat{\beta}_2 \sum_{i=1}^n x_{2i} x_{3i} - \hat{\beta}_3 \sum_{i=1}^n x_{3i}^2 = 0$

We get:

$$\sum_{i=1}^n y_i x_{3i} - \left(\frac{\sum_{i=1}^n y_i x_{2i} - \hat{\beta}_3 \sum_{i=1}^n x_{2i} x_{3i}}{\sum_{i=1}^n x_{2i}^2} \right) \sum_{i=1}^n x_{2i} x_{3i} - \hat{\beta}_3 \sum_{i=1}^n x_{3i}^2 = 0$$

Solving for $\hat{\beta}_3$ we get:

$$\hat{\beta}_3 = \frac{\sum_{i=1}^n y_i x_{3i} \sum_{i=1}^n x_{2i}^2 - \sum_{i=1}^n y_i x_{2i} \sum_{i=1}^n x_{2i} x_{3i}}{\sum_{i=1}^n x_{2i}^2 \sum_{i=1}^n x_{3i}^2 - \left(\sum_{i=1}^n x_{2i} x_{3i} \right)^2}$$

Substituting $\hat{\beta}_3$ in *

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n y_i x_{2i} \sum_{i=1}^n x_{3i}^2 - \sum_{i=1}^n y_i x_{3i} \sum_{i=1}^n x_{2i} x_{3i}}{\sum_{i=1}^n x_{2i}^2 \sum_{i=1}^n x_{3i}^2 - \left(\sum_{i=1}^n x_{2i} x_{3i} \right)^2}$$

$\swarrow \text{Cov}(y, x_2)$ $\swarrow \text{Var}(x_2)$ $\swarrow \text{Var}(x_3)$ $\swarrow (\text{Cov}(x_2, x_3))^2$

So we can also write: (just notation)

$$\hat{\beta}_2 = \frac{\hat{\Gamma}_{12} \hat{\Gamma}_{33} - \hat{\Gamma}_{13} \hat{\Gamma}_{23}}{\hat{\Gamma}_{22} \hat{\Gamma}_{33} - \hat{\Gamma}_{23}^2}$$

$\hat{\Gamma}_{12}$ = sample covariance (y, x_2)

$\hat{\Gamma}_{22}$ = sample variance (x_2)

...

POPULATION PARAMETERS AND SAMPLE STATISTICS

$$1) Y_n = \beta_1 + \beta_2 X_{2n} + \beta_3 X_{3n} + \epsilon_n$$

TRUE POPULATION MODEL

$$Y_n = \beta_2 \pi_n + \beta_3 \lambda_n + \epsilon_n$$

2) JOINT NORMALITY of Y, X_2, X_3 or alternatively of Y, π, λ

$$\begin{bmatrix} Y_n \\ X_{2n} \\ X_{3n} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix} \right)$$

We also know that the conditional expectation $E(Y_n | X_{2n}, X_{3n})$:

$$E(Y_n | X_{2n}, X_{3n}) = \beta_2 X_{2n} + \beta_3 X_{3n}$$

Recap:

$$E(Y) = \int Y f(Y) dY$$

 $f(Y)$: p.d.f.

$$E(Y|X) = \int Y \frac{f(X, Y)}{f(X)} dY$$

joint distribution

$$\text{Where } f(X) = \int f(X, Y) dY$$

marginal distribution of X

Using the above definition of conditional expectation it can shown that:

$$E(y_i | x_{i2}, x_{i3}) = \underbrace{\begin{pmatrix} \sigma_{12} & \sigma_{13} \end{pmatrix}}_{1 \times 2} \underbrace{\begin{pmatrix} \sigma_{22} & \sigma_{23} \\ \sigma_{23} & \sigma_{33} \end{pmatrix}^{-1}}_{2 \times 2} \underbrace{\begin{pmatrix} x_{i2} \\ x_{i3} \end{pmatrix}}_{2 \times 1} \quad \xrightarrow{\Sigma} \quad \Sigma$$

Theorem:

The matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible only if $|A| \neq 0$.

In such a case: $A^{-1} = \frac{1}{|A|} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$

With $|A| = ad - bc$

So: $\Sigma^{-1} = \frac{1}{\sigma_{22}\sigma_{33} - \sigma_{23}^2} \begin{pmatrix} \sigma_{33} & -\sigma_{23} \\ -\sigma_{23} & \sigma_{22} \end{pmatrix}$

Since we have that:

$$E(y_i | x_{i2}, x_{i3}) = \beta_2 x_{i2} + \beta_3 x_{i3}$$

and also:

$$E(y_i | x_{i2}, x_{i3}) = \begin{pmatrix} \sigma_{12} & \sigma_{13} \end{pmatrix} \left(\frac{1}{\sigma_{22}\sigma_{33} - \sigma_{23}^2} \begin{pmatrix} \sigma_{33} & -\sigma_{23} \\ -\sigma_{23} & \sigma_{22} \end{pmatrix} \right) \begin{pmatrix} x_{i2} \\ x_{i3} \end{pmatrix}$$

So

$$\beta_2 x_{i2} + \beta_3 x_{i3} =$$

Finally, I can get the expressions of β_2 on β_3

V. Population PARAMETERS

7

$$\beta_2 = \frac{\sigma_{12}\sigma_{33} - \sigma_{13}\sigma_{23}}{\sigma_{23}\sigma_{33} - \sigma_{23}^2}$$

$$\beta_3 = \frac{\sigma_{13}\sigma_{22} - \sigma_{12}\sigma_{23}}{\sigma_{23}\sigma_{33} - \sigma_{23}^2}$$

β_2, β_3 : UNKNOWN POPULATION PARAMETERS
they are expressed as LINEAR COMBINATIONS
of other unknown population parameters,
variances and covariances - $\sigma_{12}, \sigma_{13}, \sigma_{23}, \sigma_{22}, \sigma_{33}$

OLS ESTIMATORS

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n y_i x_{2i} \sum_{i=1}^n x_{3i}^2 - \sum_{i=1}^n y_i x_{3i} \sum_{i=1}^n x_{2i} x_{3i}}{\sum_{i=1}^n x_{2i}^2 \sum_{i=1}^n x_{3i}^2 - \left(\sum_{i=1}^n x_{2i} x_{3i} \right)^2}$$

$$\hat{\beta}_3 = \frac{\sum_{i=1}^n y_i x_{3i} \sum_{i=1}^n x_{2i}^2 - \sum_{i=1}^n y_i x_{2i} \sum_{i=1}^n x_{2i} x_{3i}}{\sum_{i=1}^n x_{2i}^2 \sum_{i=1}^n x_{3i}^2 - \left(\sum_{i=1}^n x_{2i} x_{3i} \right)^2}$$

OLS estimators $\hat{\beta}_2, \hat{\beta}_3$ are expressed as
LINEAR COMBINATIONS of SAMPLE STATISTICS -
- Variances and covariances - $\hat{\sigma}_{12}, \hat{\sigma}_{13}, \hat{\sigma}_{23}, \hat{\sigma}_{22}, \hat{\sigma}_{33}$

Estimated model: $y_i = \beta_2^* x_{i2} + \epsilon_i$

"True" model: $y_i = \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i$

We know that the OLS estimator of β_2^* is

$$\hat{\beta}_2^* = \frac{\sum_{i=1}^n y_i x_{i2}}{\sum_{i=1}^n x_{i2}^2}$$

What about $\hat{\beta}_2^*$?

Let us substitute y_i with its "TRUE" expression

$$y_i = \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i$$

$$\hat{\beta}_2^* = \frac{\sum_{i=1}^n x_{i2} (\beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i)}{\sum_{i=1}^n x_{i2}^2}$$

$$\hat{\beta}_2^* = \beta_2 \frac{\sum_{i=1}^n x_{i2}^2}{\sum_{i=1}^n x_{i2}^2} + \beta_3 \frac{\sum_{i=1}^n x_{i2} x_{i3}}{\sum_{i=1}^n x_{i2}^2} + \frac{\sum_{i=1}^n x_{i2} \epsilon_i}{\sum_{i=1}^n x_{i2}^2}$$

$$E(\hat{\beta}_2^*) = \beta_2 + \beta_3 \frac{\sum_{i=1}^n x_{i2} x_{i3}}{\sum_{i=1}^n x_{i2}^2} + 0$$

↳ x_{i2}, x_{i3} are "FIXED" VARIABLES" so:

$$E(x_{i2}) = x_{i2}$$

$$E(x_{i3}) = x_{i3}$$

$$\text{BIAS}(\hat{\beta}_2^*) = E(\hat{\beta}_2^*) - \beta_2$$

$$= \beta_3 \frac{\sum_{i=1}^n x_{i2} x_{i3}}{\sum_{i=1}^n x_{i2}^2}$$

$$\text{BIAS}(\hat{\beta}_2^*) = \beta_3 \frac{\sum_{i=1}^n x_{2i} x_{3i}}{\sum_{i=1}^n x_{2i}^2}$$

What this expression is saying to us?

$$\text{BIAS}(\hat{\beta}_2^*) = f(\beta_3^+, \text{Cov}(x_2, x_3), \text{Var}(x_2))$$

- If $\beta_3 = 0 \Rightarrow \text{BIAS}(\hat{\beta}_2^*) = 0$

TRIVIAL: In such a case, there is not an omission of a relevant variable

More the omitted variable is "relevant" (high β_3) and more the bias will be high

- If $\text{Cov}(x_2, x_3) = 0 \Rightarrow \text{BIAS}(\hat{\beta}_2^*) = 0$

In such a case, x_{3i} "enters" E_{it}

And we know that if $E(E_{it} | x_{2i}) = 0$

$\Rightarrow \hat{\beta}_2^*$ has no bias

Generally it is not the case

- Again having a big variance of x_{2i} "help" the estimation; in this specific case, it helps to reduce the omitted variable bias

Example: Wage equation

Estimated model: $WAGE_i = \beta_1 + \beta_2^* EDUCATION_i + \epsilon_i$

"TRUE" Model: $WAGE_i = \beta_1 + \beta_2 EDUCATION_i + \beta_3 ABILITY_i + \epsilon_i$

$$BIAS(\hat{\beta}_2^*) = \beta_3 \frac{\sum_{i=1}^n education_i ability_i}{\sum_{i=1}^n education_i^2}$$

EX ANTE, what is the "expected" sign of $BIAS(\hat{\beta}_2^*)$?
Think a bit...

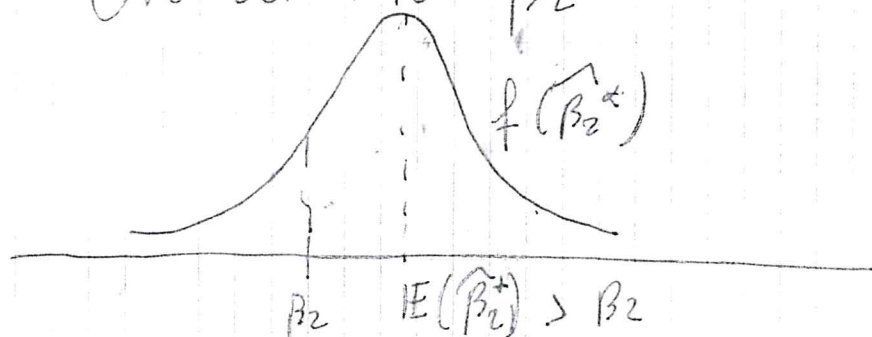
I am expecting that:

$$\bullet \beta_3 > 0$$

$$\bullet \sum_{i=1}^n education_i ability_i > 0$$

$$\Rightarrow BIAS(\hat{\beta}_2^*) > 0$$

OVERESTIMATE β_2



VI INCLUSION OF AN IRRELEVANT VARIABLE 4

ESTIMATED MODEL : $y_i = \beta_2^* x_{2i} + \beta_3^* x_{3i} + \epsilon_i$

"TRUE" MODEL : $y_i = \beta_2 x_{2i} + \epsilon_i$

We already know that:

$$\hat{\beta}_2^* = \frac{\sum_{i=1}^n y_i x_{2i} \sum_{i=1}^n x_{2i}^2 - \sum_{i=1}^n y_i x_{3i} \sum_{i=1}^n x_{2i} x_{3i}}{\sum_{i=1}^n x_{2i}^2 \sum_{i=1}^n x_{3i}^2 - \left(\sum_{i=1}^n x_{2i} x_{3i} \right)^2}$$

Again, let us substitute y_i with its true expression,
now: $y_i = \beta_2 x_{2i} + \epsilon_i$

$$\hat{\beta}_2^* = \frac{\sum_{i=1}^n x_{2i} (\beta_2 x_{2i} + \epsilon_i) \sum_{i=1}^n x_{2i}^2 - \sum_{i=1}^n x_{2i} x_{3i} \sum_{i=1}^n x_{3i} (\beta_2 x_{2i} + \epsilon_i)}{\sum_{i=1}^n x_{2i}^2 \sum_{i=1}^n x_{3i}^2 - \left(\sum_{i=1}^n x_{2i} x_{3i} \right)^2}$$

$$\hat{\beta}_2^* = \frac{\beta_2 \left[\sum_{i=1}^n x_{2i}^2 \sum_{i=1}^n x_{2i}^2 - \left(\sum_{i=1}^n x_{2i} x_{3i} \right)^2 \right] + \sum_{i=1}^n x_{2i}^2 \sum_{i=1}^n \epsilon_i - \sum_{i=1}^n x_{2i} x_{3i} \sum_{i=1}^n \epsilon_i}{\sum_{i=1}^n x_{2i}^2 \sum_{i=1}^n x_{3i}^2 - \left(\sum_{i=1}^n x_{2i} x_{3i} \right)^2}$$

$$\hat{\beta}_2^* = \beta_2 \frac{\sum_{i=1}^n x_{2i}^2 \sum_{i=1}^n x_{2i}^2 - \left(\sum_{i=1}^n x_{2i} x_{3i} \right)^2}{\sum_{i=1}^n x_{2i}^2 \sum_{i=1}^n x_{3i}^2 - \left(\sum_{i=1}^n x_{2i} x_{3i} \right)^2} + \frac{\sum_{i=1}^n x_{2i}^2 \sum_{i=1}^n \epsilon_i - \sum_{i=1}^n x_{2i} x_{3i} \sum_{i=1}^n \epsilon_i}{\sum_{i=1}^n x_{2i}^2 \sum_{i=1}^n x_{3i}^2 - \left(\sum_{i=1}^n x_{2i} x_{3i} \right)^2}$$

$E(\hat{\beta}_2^*) = \beta_2 + 0 \rightarrow$ Simply because $E(\epsilon_i) = 0$

$E(\hat{\beta}_3^*) = 0$ Please show this result at home

VI

there is no bias from including an irrelevant variable,⁵
 but there is still a PRICE to pay:

$$\text{Var}(\hat{\beta}_2^*) = \frac{\sigma^2}{\sum_{i=1}^n x_{2i}^2 (1 - \hat{\rho}^2)}$$

$$\text{with } \hat{\rho}^2 = \left(\frac{\sum_{i=1}^n x_{1i} x_{2i}}{\sqrt{\sum_{i=1}^n x_{1i}^2 \sum_{i=1}^n x_{2i}^2}} \right)^2$$

While, as we know, when estimating the "true" regression model $y_i = \beta_2 x_{2i} + \epsilon_i$, we have:

$$\text{Var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum_{i=1}^n x_{2i}^2}$$

▣ Including an irrelevant variable $\rightarrow$ LOSS EFFICIENCY

$$\text{Var}(\hat{\beta}_2^*) \geq \text{Var}(\hat{\beta}_2)$$

$$\text{Var}(\hat{\beta}_2^*) = \text{Var}(\hat{\beta}_2) \text{ if } \rho = 0$$

▣ Including an irrelevant variable is a Less important SPECIFICATION ERROR than omitting a relevant variable

$\hookrightarrow$ MAIN MESSAGE:

Melius est abstinere quam deficere !!

~~///~~ $K-1$ EXPLANATORY VARIABLES

$$Y_n = \beta_1 + \beta_2 X_{2n} + \beta_3 X_{3n} + \dots + \beta_K X_{Kn} + \epsilon_n$$

$$n = 1, \dots, N$$

$\beta_2, \beta_3, \dots, \beta_K$: UNKNOWN PARAMETERS TO BE ESTIMATED

ϵ_n : RANDOM VARIABLE - NOT OBSERVABLE

Y_n : RANDOM VARIABLE - OBSERVABLE

$X_{2n}, \dots, X_{Kn}$: RANDOM - OR FIXED - VARIABLES - OBSERVABLE

REWRITE THE MODEL:

$$n=1 : Y_1 = \beta_1 + \beta_2 X_{21} + \dots + \beta_K X_{K1} + \epsilon_1$$

$$n=2 : Y_2 = \beta_1 + \beta_2 X_{22} + \dots + \beta_K X_{K2} + \epsilon_2$$

$$\vdots$$

$$n=N : Y_N = \beta_1 + \beta_2 X_{2N} + \dots + \beta_K X_{KN} + \epsilon_N$$

$Y, X_2, \dots, X_K$ VARY ACROSS CROSS-SECTIONS

$\beta_1, \dots, \beta_K$ ARE THE SAME FOR ALL n

WHY?

• This "MAKES SENSE" ??

• SOLUTIONS : USE OF DUMMY VARIABLES
PANEL DATA

$$Y = \alpha \beta_1 + X \beta_2 + \epsilon$$

$$\begin{pmatrix} Y \\ Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{pmatrix}_{N \times 1} = \alpha \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{N \times 1} \beta_1 + \begin{pmatrix} X_{21} & X_{k1} \\ X_{22} & X_{k2} \\ \vdots & \vdots \\ X_{2N} & X_{kN} \end{pmatrix}_{N \times (k-1)} \begin{pmatrix} \beta_2 \\ \beta_3 \\ \vdots \\ \beta_k \end{pmatrix}_{(k-1) \times 1} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{pmatrix}_{N \times 1}$$

ALTERNATIVE 4:

$$Y = X \beta + \epsilon$$

$$\begin{pmatrix} Y \\ Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{pmatrix}_{N \times 1} = \begin{pmatrix} 1 & X_{21} & X_{k1} \\ 1 & X_{22} & X_{k2} \\ \vdots & \vdots & \vdots \\ 1 & X_{2N} & X_{kN} \end{pmatrix}_{N \times k} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}_{k \times 1} + \begin{pmatrix} \epsilon \\ \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{pmatrix}_{N \times 1}$$

$$Y = X\beta + \varepsilon$$

$N \times 1$ $N \times K$ $K \times 1$ $N \times 1$

$$E(\varepsilon) = 0 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$\underbrace{N \times 1}_{N \times 1}$ $N \times 1$

$E(\varepsilon_1) = 0$
 $E(\varepsilon_2) = 0$
 $\vdots$
 $E(\varepsilon_N) = 0$

$$E(\varepsilon|X) = 0 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$\underbrace{N \times 1}_{N \times 1}$ $N \times 1$

$E(\varepsilon_1|X) = 0$
 $E(\varepsilon_2|X) = 0$
 $\vdots$
 $E(\varepsilon_N|X) = 0$

$$E(\varepsilon \varepsilon') = \sigma^2 I$$

$N \times 1$ $1 \times N$ 1×1 $N \times N$
 $N \times N$

I : IDENTITY MATRIX, $I = \begin{pmatrix} 1 & 0 & & 0 \\ 0 & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix}$

$$E(\varepsilon \varepsilon') = \sigma^2 I = \begin{pmatrix} \sigma^2 & 0 & & 0 \\ 0 & \sigma^2 & & \\ & & \ddots & \\ 0 & & & \sigma^2 \end{pmatrix}$$

$E(\varepsilon \varepsilon') = \sigma^2 I$ IS EQUIVALENT TO

$$E(\varepsilon_i)^2 = \sigma^2$$

HOMOGENEITY

$$E(\varepsilon_i \varepsilon_j) = 0$$

CROSS-SECTIONAL
INDEPENDENCE

New Hypothesis

- $\text{Rank}(X) = k$ (full rank)
(and $k \leq n$)

Recap. $\text{Rank}(X)$: maximum number of (column) vectors linearly independent

2 vectors x_1 and x_2 are linearly independent if:

$$\nexists \lambda \in \mathbb{R} \text{ such that } x_1 = \lambda x_2$$

3 vectors x_1, x_2, x_3 are linearly independent if:

$$\nexists \lambda \in \mathbb{R} \text{ such that } x_1 = \lambda_1 x_2 + \lambda_2 x_3$$

Link with the SLRM

$$\underset{n \times 1}{Y} = \underset{1 \times 1}{\beta_1} \underset{n \times 1}{x_1} + \underset{1 \times 1}{\beta_2} \underset{n \times 1}{x_2} + \underset{n \times 1}{\epsilon}$$

$$\nexists \lambda \in \mathbb{R} \text{ such that } x = \lambda x_1$$

(or equivalently: $\sum_{i=1}^n x_{2i}^2 \neq 0$ $y_i = \beta_1 x_{1i} + \epsilon_i$)

$$\underset{n \times 1}{Y} = \underset{n \times 2}{X} \underset{2 \times 1}{\beta} + \underset{n \times 1}{\epsilon}$$

$$X = \begin{pmatrix} 1 & x_{21} \\ 1 & x_{22} \\ \vdots & \vdots \\ 1 & x_{2n} \end{pmatrix}$$

$$\text{Rank}(X) = 2 \text{ if}$$

$$\nexists \lambda \in \mathbb{R} \text{ such that } x_2 = \lambda x_1$$

$$\text{if } x_2 = \lambda x_1 \Rightarrow \text{Rank}(X) = 1$$

$$Y = X\beta + \varepsilon$$

$N \times 1$ $N \times K$ $K \times 1$ $N \times 1$

$$\hat{\varepsilon} = Y - X\hat{\beta}$$

OLS Min $\hat{\varepsilon}'\hat{\varepsilon}$

$1 \times N$ $N \times 1$
 1×1

DEFINE. $S(\hat{\beta}) = \hat{\varepsilon}'\hat{\varepsilon} = (Y - X\hat{\beta})' (Y - X\hat{\beta})$

$$\begin{aligned}
 S(\hat{\beta}) &= Y'Y - \hat{\beta}'X'Y - Y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta} \\
 &= Y'Y - 2\hat{\beta}'X'X + \hat{\beta}'X'X\hat{\beta}
 \end{aligned}$$

$1 \times N$ $N \times 1$ $1 \times K$ $K \times N$ $N \times 1$ $1 \times N$ $N \times K$ $K \times 1$ $1 \times K$ $K \times N$ $N \times K$ $K \times 1$
 1×1 1×1 1×1 1×1 1×1

Recap DERIVATIVES OF MATRICES

Let X and c be two column vectors $N \times 1$

$c'X$ is a scalar and $\frac{\partial c'X}{\partial X} = c$

$1 \times N$ $N \times 1$ 1×1 1×1 $N \times 1$

also $\frac{\partial c'X}{\partial c} = X$

1×1 $N \times 1$

If A is a matrix

$$\frac{\partial AX}{\partial X} = A'$$

$$\text{Min } S(\hat{\beta})$$

$$\text{with } S(\hat{\beta}) = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}$$

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

$$\frac{\partial S(\hat{\beta})}{\partial \hat{\beta}} = \frac{\partial (Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta})}{\partial \hat{\beta}}$$

$$= 0 - 2 \frac{\partial \hat{\beta}'X'Y}{\partial \hat{\beta}} + \frac{\partial \hat{\beta}'X'X\hat{\beta}}{\partial \hat{\beta}}$$

$$= 0 - 2 \underset{\substack{K \times N \quad N \times 1 \\ K \times 1}}{X'Y} + 2 \underset{\substack{K \times N \quad N \times K \quad K \times 1 \\ K \times K \quad K \times 1 \\ K \times 1}}{X'X\hat{\beta}}$$

$$- 2X'Y + 2X'X\hat{\beta} = 0$$

$$X'Y - X'X\hat{\beta} = 0$$

$$\hat{\beta} = \underset{\substack{K \times N \quad N \times K \\ K \times K}}{(X'X)^{-1}} \underset{\substack{K \times N \quad N \times 1 \\ K \times 1}}{X'Y}$$

$$\hat{\beta} \text{ is a } K \times 1 \text{ vector : } \hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_K \end{pmatrix}$$

NOTE:

$$\underset{\substack{1 \times K \quad K \times N \quad N \times K \\ 1 \times N \quad N \times 1 \\ 1 \times 1}}{\hat{\beta}'X'X\hat{\beta}}$$

IS A QUADRATIC FORM,

AND BY DEFINITION IS A SCALAR.

THE $K \times 1$ VECTOR OF DERIVATIVES OF THIS TERM WITH RESPECT TO $\hat{\beta}$ IS EQUAL TO $2X'X\hat{\beta}$ (PROOF NOT PROVIDED HERE)

- $\hat{\beta}$ exists if $(X'X)$ can be inverted
 $(X'X)$ can be inverted if $\text{Rank}(X) = k$

$$\text{Rank}(X) = k \begin{cases} \rightarrow X \text{ invertible} \\ \rightarrow (X'X) \text{ INVERTIBLE} \end{cases}$$

- PROOF OF MINIMUM

If the MATRIX X HAS RANK k , the HESSIAN MATRIX:

$$\frac{\partial^2 S(\hat{\beta})}{\partial \hat{\beta} \partial \hat{\beta}'} = 2 X'X$$

IS A POSITIVE DEFINITE MATRIX

$X'X$ is a $k \times k$ MATRIX
 $k \times n \quad n \times k$

DEFINE THE QUADRATIC FORM

$$q = \underset{1 \times n}{C'} \underset{n \times n}{X'} \underset{n \times k}{X} \underset{n \times 1}{C}$$

C is a non-zero vector $k \times 1$

If $q > 0 \Rightarrow X'X$ DEFINITE POSITIVE

DEFINE $V = XC$ (JUST NOTATION)
 $n \times 1 \quad n \times 1$

$$q = V'V$$

$\rightarrow q > 0$ in all cases except when all the elements of $V = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$ are ZERO
 $\rightarrow$ this happens only if $\text{Rank}(X) < k$

VII

OLS

8

$$E(\hat{\beta}) = ?$$

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

$$E(\hat{\beta}) = E\left[(X'X)^{-1}X'(X\beta + \varepsilon)\right]$$

$$= E\left[\cancel{(X'X)^{-1}}\cancel{X'X}\beta + (X'X)^{-1}X'\varepsilon\right]$$

$$= E\left[\beta + (X'X)^{-1}X'\varepsilon\right]$$

$$= \beta + E(X'X)^{-1}X'\varepsilon$$

$$E(\varepsilon) = 0$$

$$E(\hat{\beta}) = \beta$$

VII

OLS

8

$$\text{Var}(\hat{\beta}) = ?$$

$$= \mathbb{E} \left[(\hat{\beta} - \mathbb{E}(\hat{\beta})) (\hat{\beta} - \mathbb{E}(\hat{\beta}))' \right]$$

$$= \mathbb{E} \left[(\hat{\beta} - \beta) (\hat{\beta} - \beta)' \right]$$

$$= \mathbb{E} \left[\left((X'X)^{-1} X' \underbrace{(X\beta + \varepsilon)}_{\hat{y}} - \beta \right) \left((X'X)^{-1} X' (X\beta + \varepsilon) - \beta \right)' \right]$$

$$= \mathbb{E} \left[\left(\cancel{(X'X)^{-1} X' \beta} + (X'X)^{-1} X' \varepsilon - \beta \right) \left(\cancel{(X'X)^{-1} X' \beta} + (X'X)^{-1} X' \varepsilon - \beta \right)' \right]$$

$$= \mathbb{E} \left[\left(\cancel{\beta - \beta} + (X'X)^{-1} X' \varepsilon \right) \left(\cancel{\beta - \beta} + (X'X)^{-1} X' \varepsilon \right)' \right]$$

$$= \mathbb{E} \left[(X'X)^{-1} X' \varepsilon \varepsilon' X (X'X)^{-1} \right]$$

$$\mathbb{E}(\varepsilon \varepsilon') = \sigma^2 \mathbf{I}$$

$$= \cancel{(X'X)^{-1} X' X} (X'X)^{-1} \sigma^2$$

$$= \sigma^2 (X'X)^{-1}$$

True $gdp_x = \beta_2 corruption_x + \beta_3 institution_x + \epsilon_x$

Estimated
model
case 1

$$gdp_x = \beta_2^* corruption_x + \epsilon_x$$

$$BIAS(\hat{\beta}_2^*) ?$$

Estimated
model
case 2

$$gdp_x = \beta_3^* institution_x + \epsilon_x$$

$$BIAS(\hat{\beta}_3^*) ?$$

Expected signs: $\beta_2 < 0$, $\beta_3 > 0$

~~Expected signs:~~

$$\sum_{x=1}^n corruption_x institution_x < 0$$

CASE 1

$$BIAS(\hat{\beta}_2^*) = \beta_3$$

> 0

< 0

$$\frac{\sum_{x=1}^n corruption_x institution_x < 0}{\sum_{x=1}^n corruption_x^2}$$

CASE 2

$$BIAS(\hat{\beta}_3^*) = \beta_2$$

< 0

> 0

$$\frac{\sum_{x=1}^n corruption_x institution_x < 0}{\sum_{x=1}^n institution_x^2}$$

$$y'y$$

$$y = X\beta + \epsilon$$

$$y' = \beta'X' + \epsilon'$$

$$(\beta'X' + \epsilon') (X\beta + \epsilon)$$

$$\frac{\partial}{\partial \beta} \frac{\beta'X'X\beta + \beta'X'\epsilon + \epsilon'X\beta + \epsilon'\epsilon}{2}$$

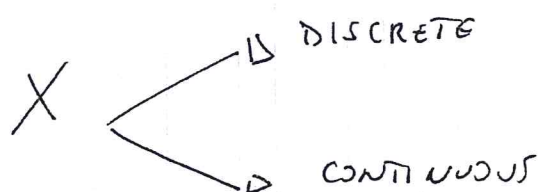
$$2X'X\beta + X'\epsilon + \epsilon'X + \cancel{\epsilon'\epsilon}$$

RANDOM VARIABLE

2

X : RANDOM VARIABLE

x : REALIZATION OF X



X DISCRETE

X : x_1 x_2 $x_3 \dots x_m$

Probability: p_1 p_2 $p_3 \dots p_m$

PROBABILITY FUNCTION $p_n = P(X = x_n)$

$$\bullet \sum_{n=1}^m p_n = 1$$

$$\bullet p_n \geq 0 \quad \forall n$$

IT IS A FUNCTION ASSOCIATING AT EACH VALUE OF X THE CORRESPONDING PROBABILITY

CUMULATIVE PROBABILITY FUNCTION

(FUNZIONE DI RIPARTIZIONE)

$$F(x) = P(X \leq x)$$

$$F(x) = \sum_{n: x_n \leq x}$$

EXPECTATION of a R.V.

3

$$\mathbb{E}(X) = \sum_{i=1}^n x_i p_i$$

Recall: "descriptive mean"

$$M(X) = \sum_{i=1}^n x_i \cdot x_i$$

~~Simple Arithmetic Mean~~ SIMPLE ARITHMETIC MEAN: $w_i = \frac{1}{n}$

$\mathbb{E}(X)$ is the EXPECTED VALUE,
"expected" if the experiment
will be performed many times

VARIANCE of a R.V.

$$\begin{aligned} V(X) &= \mathbb{E} [X - \mathbb{E}(X)]^2 \\ &= \sum_{i=1}^n [\cancel{x_i} - \mathbb{E}(X)]^2 p_i \end{aligned}$$

Sometime useful to know:

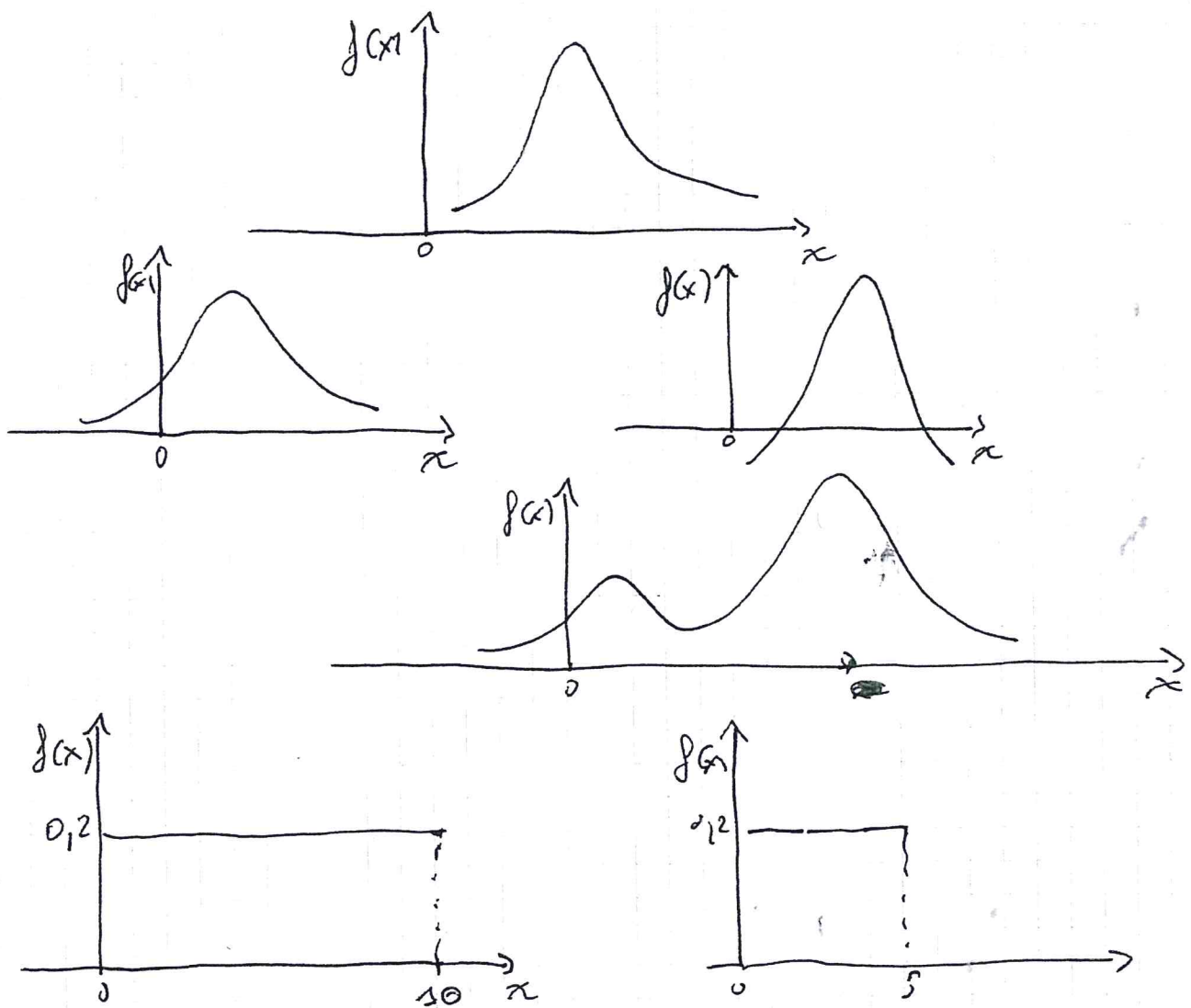
$$\begin{aligned} V(X) &= \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 \\ &= \sum_{i=1}^n x_i^2 p_i - [\mathbb{E}(X)]^2 \end{aligned}$$

CONTINUOUS RANDOM VARIABLES

$f(x)$: (probability) density function (pdf)

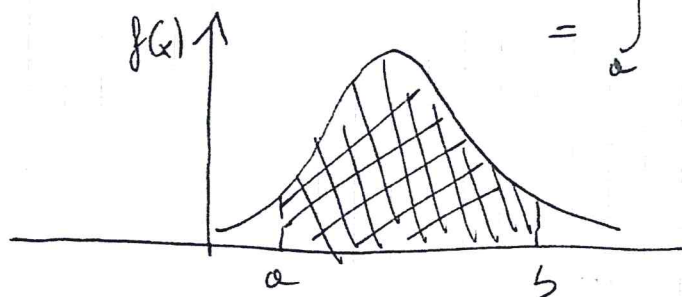
- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x) dx = 1$

Question: the following functions could be (in theory) pdfs??



$$F(x) = P(X \leq x) \\ = \int_{-\infty}^x f(x) dx$$

$$P(a \leq X \leq b) = F(b) - F(a) \\ = \int_a^b f(x) dx$$



$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

$$V(X) = E[X - E(X)]^2 \\ = E(X^2) - [E(X)]^2 \\ = \int_{-\infty}^{\infty} x^2 f(x) dx - [E(X)]^2$$

Recap: GAUSSIAN DISTRIBUTION

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2} (x - \mu)^2\right]$$

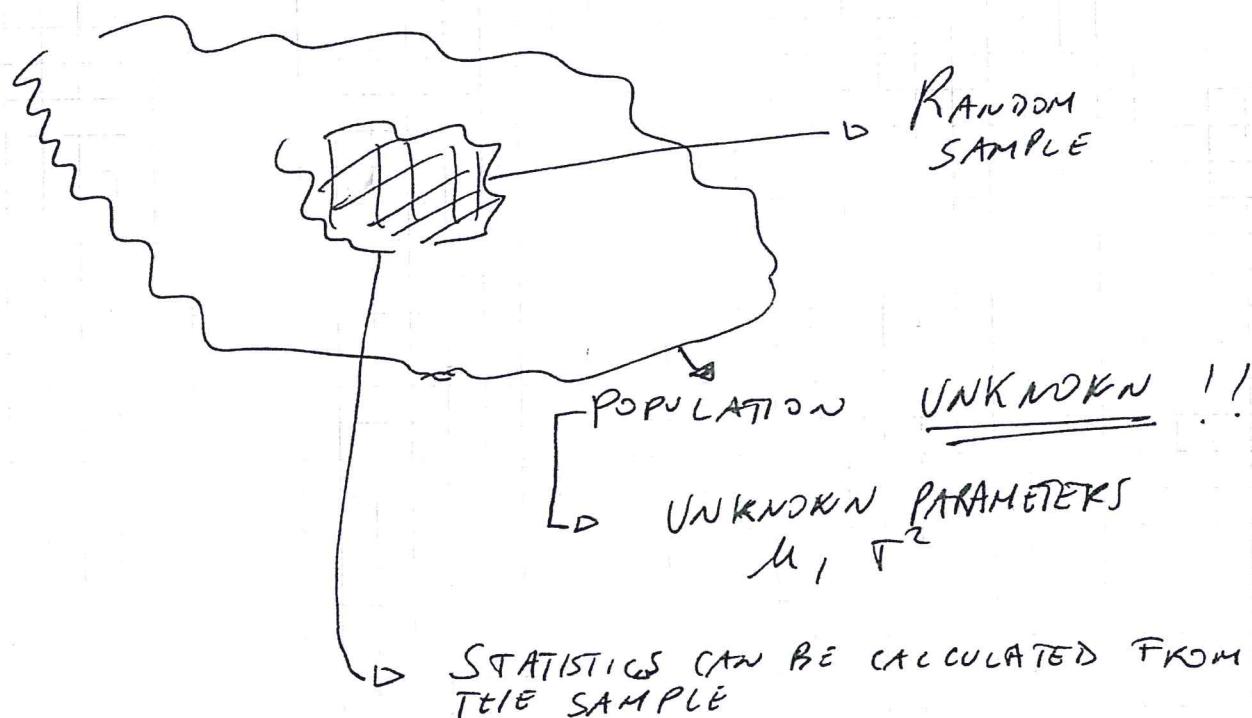
$$E(X) = \mu$$

$$V(X) = \sigma^2$$

$$X \sim N(\mu, \sigma^2)$$

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

RANDOM SAMPLING AND CLT



$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2$$

CLT : IF THE POPULATION IS NORMAL (GAUSSIAN)
OR IF WE HAVE A BIG SAMPLE
 (even if the population is not Gaussian),
 we have:

$$\bar{X} \sim N \left(\mu, \left(\frac{\sigma}{\sqrt{n}} \right)^2 \right)$$

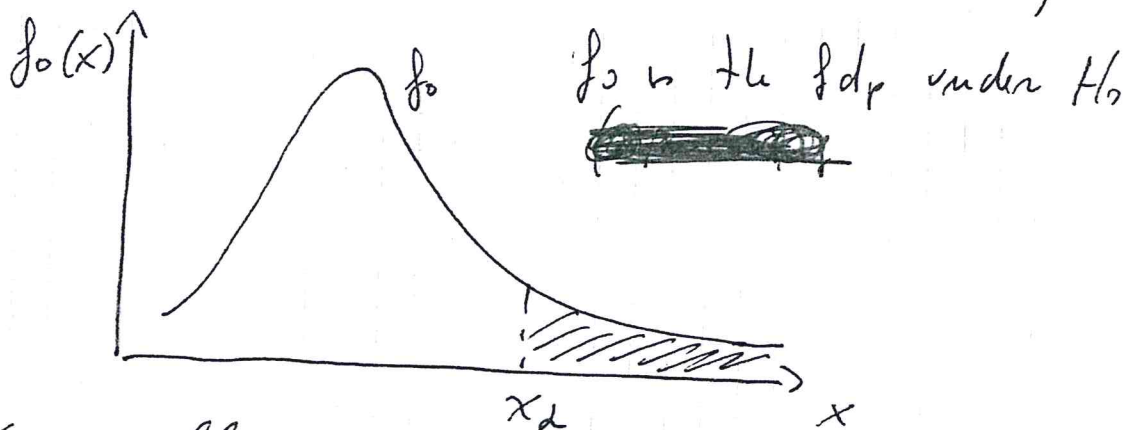
TEST OF HYPOTHESIS

6

$$H_0 : \mu = c$$

$$H_1 : \mu > c$$

Fix α , which is called SIGNIFICANT LEVEL, ex: 0.05



x_α is called CRITICAL VALUE. It is the value such that:

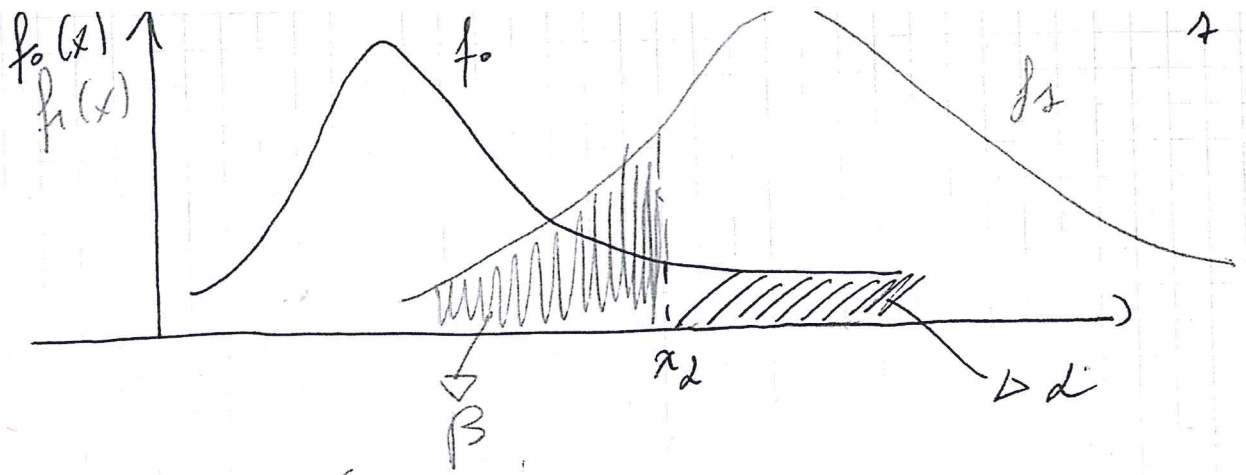
$$\int_{x_\alpha}^{\infty} f_0(x) dx = \alpha$$

$$= P(\text{reject } H_0 \mid H_0)$$

$$\int_{-\infty}^{x_\alpha} f(x) dx = (1 - \alpha)$$

$$= P(\text{not reject } H_0 \mid H_0)$$

α error of first type



$$\int_{-\infty}^{x_2} f_1(x) dx = \beta$$

$$= P(\text{not reject } H_0 \mid H_1)$$

$$\int_{x_2}^{\infty} f_1(x) dx = 1 - \beta$$

$$= P(\text{reject } H_0 \mid H_1)$$

β : Second type error

$(1 - \beta)$: POWER OF A TEST

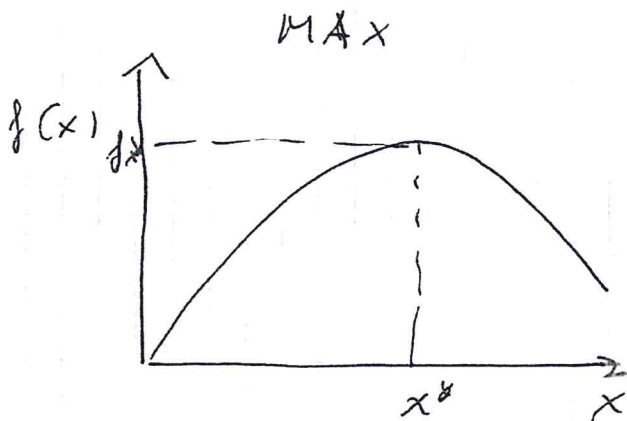
MAX / MIN

8

1 VARIABLE

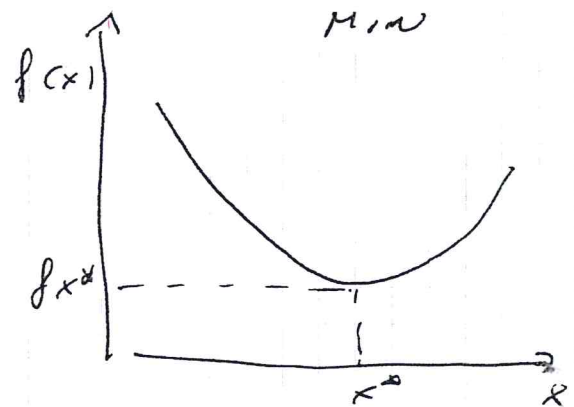
$$f: \mathbb{R} \rightarrow \mathbb{R}$$

x^* : max/min



F.o.c. $f'(x) = 0$

S.o.c. $f''(x) < 0$



F.o.c. $f'(x) = 0$

S.o.c. $f''(x) > 0$

MULTIVARIATE CASE

~~2~~ 2 VARIABLES

• MAX $f(x_1, x_2)$

• MIN $f(x_1, x_2)$

F.o.c. :

$$\frac{\partial f(x_1, x_2)}{\partial x_1} = 0$$

$$\frac{\partial f(x_1, x_2)}{\partial x_2} = 0$$

n Variables

9

GRADIENT VECTOR (vector of partial derivatives)

$$Df(x) = \left(\frac{\partial f(x_1, \dots, x_n)}{\partial x_1}, \dots, \frac{\partial f(x_1, \dots, x_n)}{\partial x_n} \right)$$

$$\text{F.O.C : } Df(x) = 0$$

$$\text{: } \frac{\partial f(x_1, \dots, x_n)}{\partial x_1} = 0, \dots, \frac{\partial f(x_1, \dots, x_n)}{\partial x_n} = 0$$

S.O.C HESSIAN MATRIX (second derivatives)

2 variables case

$$f_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

$$H = \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix}$$

MAX : H definite NEGATIVE

MIN : H = POSITIVE

$$\begin{pmatrix} A \end{pmatrix} \begin{matrix} X \\ \end{matrix} \begin{matrix} X^T \\ \end{matrix}$$

$m \times m \quad \quad \quad m \times 1 \quad \quad \quad 1 \times m$

QUADRATIC FORM

$$\begin{matrix} X^T & A & X \\ 1 \times m & m \times m & m \times 1 \end{matrix} = \text{scalar } 1 \times 1$$

$$\begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= a_{11} x_1^2 + (a_{12} + a_{21}) x_1 x_2 + a_{22} x_2^2$$

$X^T A X > 0 \rightarrow$	A Definite POSITIVE
$X^T A X \geq 0 \rightarrow$	A SEMI DEFINITE POSITIVE
$X^T A X < 0 \rightarrow$	A DEFINITE NEGATIVE
$X^T A X \leq 0 \rightarrow$	A SEMI DEFINITE NEGATIVE

$A : \text{MATRIX } 2 \times 2$

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

A DEFINITE POSITIVE IF $\det A > 0$

$$\det(A) = a_{11}a_{22} - a_{21}a_{12}$$

In general, to determine if a symmetric matrix is definite POSITIVE or NEGATIVE, we have to study EIGENVALUES

Given a symmetric matrix $A_{n \times n}$

$\vec{z}$ ($n \times 1$ vector) and λ (scalar) are the EIGENVECTOR or EIGENVALUE of A if

$$\begin{matrix} A \vec{z} \\ n \times n & n \times 1 \\ n \times 1 \end{matrix} = \begin{matrix} \lambda \vec{z} \\ 1 \times 1 & n \times 1 \\ n \times 1 \end{matrix}$$

The Vector $A\vec{z}$ is proportional to $\vec{z}$ according to λ .
We search all the λ s for which

$$A\vec{z} - \lambda \vec{z} = 0$$

$$(A - \lambda I) \vec{z} = 0$$

Theorem:

$\lambda_1, \dots, \lambda_n > 0 \rightarrow A$ definite POSITIVE

$\lambda_1, \dots, \lambda_n \geq 0 \rightarrow A$ SEMI definite POSITIVE

$\lambda_1, \dots, \lambda_n < 0 \rightarrow A$ definite NEGATIVE

$\lambda_1, \dots, \lambda_n \leq 0 \rightarrow A$ SEMI definite NEGATIVE