Topic: SUMMARY STATISTICS

- SUMMARY STATISTICS ARE USED TO SUMMARIZE A SET OF OBSERVATIONS IN ORDER TO COMMUNICATE THE LARGEST AMOUNT OF INFORMATION AS SIMPLY AS POSSIBLE

- A RANDOM VARIABLE X CAN BE CHARACTERIZED IN PART BY ITS MOMENT CHARACTERISTICS: MEAN, VARIANCE, SKEWNESS AND KURTOSIS

- IN APPLICATIONS MEAN, VARIANCE, SKEWNESS AND KURTOSIS CAN BE ESTIMATED FROM THE DATA. THE ESTIMATES ARE CALLED SUMMARY STATISTICS.

- THEY ARE MEASURES OF:
    - CENTRAL TENDENCY (mean, median, mode)
    - DISPERSION (range, variance, standard deviation)
    - SHAPE (skewness, kurtosis)

Exercise 2.1

a) SUMMARY STATISTICS AND HISTOGRAM CAN BE COMPUTED IN EVIEWS.

OPEN FILE d-ibm3dx7008 in Eviews
↓
DOUBLE CLICK ON THE VARIABLE rtn
↓
VIEW
↓
DESCRIPTIVE STATISTICS AND TESTS
↓
HISTOGRAM AND STATS

WE OBTAIN:

SAMPLE MEAN = 0.000402 → mean is close to zero

STANDARD DEVIATION = 0.016927

MINIMUM = -0.229630

MAXIMUM = 0.131636

SKEWNESS = 0.061387

KURTOSIS = 12.91636

↓

THIS IS AN INDICATION THAT THE DISTRIBUTION OF THESE RETURNS MAY BE NOT NORMAL (KURTOSIS IS 12.916 WHICH IS MUCH HIGHER THAN KURTOSIS OF THE NORMAL DISTRIBUTION, THAT IS 3)

NOW WE NEED TO DO RIGOROUS TESTING FOR SIGNIFICANCE OF MEAN, SKEWNESS AND EXCESS KURTOSIS.

b)

i) is the mean significant?

WE WANT TO TEST IF THE MEAN OF RETURN IS SIGNIFICANT OR NOT

- WE FORMALISE IT BY STATING H_0, H_1

$H_0: E(x) = 0$ ------→ mean is not significant

$H_1: E(x) \neq 0$ ------→ mean is significant. In other words: mean is statistically different from zero

- 5 % SIGNIFICANCE LEVEL

- RULE :

IF $|\bar{x}| > 2 \frac{SD}{\sqrt{n}}$ THEN REJECT H_0 (so mean is significant at 5 % level)

IF $|\bar{x}| \leq 2 \frac{SD}{\sqrt{n}}$ THEN FAIL TO REJECT H_0 (mean is not significant at 5 % level)

where $\bar{x}$ = SAMPLE MEAN
n = SAMPLE SIZE
SD = SAMPLE STANDARD DEVIATION (s or $\hat{\sigma}$)

HOW DO WE OBTAIN SUCH RESECTION RULE?

→ WE CONSTRUCT THE TEST STATISTICS
(t- statistic)

$$t = \frac{\bar{X} - \mu_{H_0}}{SD/\sqrt{n}} = \frac{\bar{X} - 0}{SD/\sqrt{n}} = \frac{\bar{X}}{\frac{SD}{\sqrt{n}}}$$

WE WANT TO TEST IF
$E(x) = 0$. THEREFORE
μ_{H0} IS ZERO.
μ_{H0} IS THE HYPOTHESIZED
POPULATION MEAN

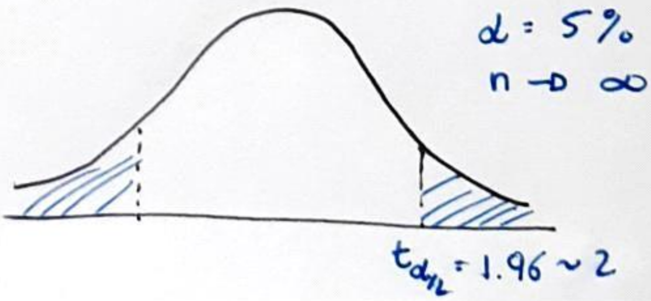
ACCORDING TO STATISTICAL THEORY,
WE RESECT H_0 AT 5% SIGNIFICANT LEVEL
IF $|t| > t_{\alpha/2}$

SINCE n IS BIG ($n = 9843$), THEN $t_{2.5\%} = 1.96$

WE CAN WRITE:

$$|t| > t_{2.5\%}$$

$$|t| > 1.96 \simeq 2$$



WE SUBSTITUTE THE t-statistic IN THE REJECTION RULE

$$|t| > 2$$

$$\left| \frac{\bar{x}}{SD/\sqrt{n}} \right| > 2$$

$$|\bar{x}| > 2 \frac{SD}{\sqrt{n}}$$ → this is the rejection rule

→ IN OUR CASE:

WE HAVE $\bar{x} = 0.000402$

$SD = 0.016927$

$n = 9843$

SO $2 \frac{SD}{\sqrt{n}} = 2 \cdot \frac{0.016927}{\sqrt{9843}} = 0.00034$

WE CONCLUDE THAT

$$|\bar{x}| = 0.000402 > 0.00034$$

$$\Downarrow$$

- WE REJECT H_0
- IT MEANS THAT THE MEAN IS STATISTICALLY DIFFERENT FROM ZERO

ii) IS SKEWNESS EQUAL TO ZERO?

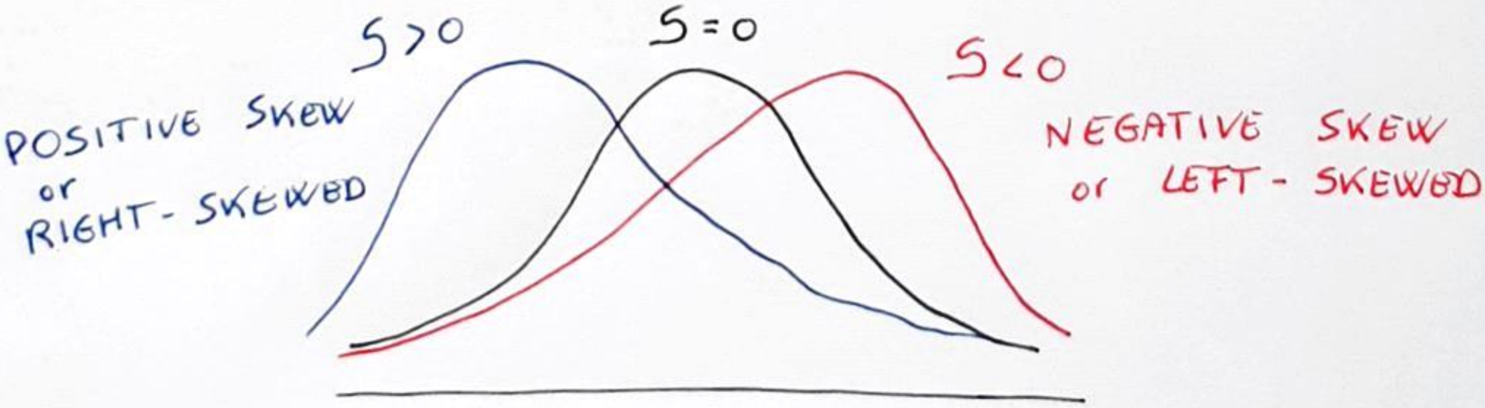
SKEWNESS IS THE MEASURE OF THE ASYMMETRY OF THE PROBABILITY DISTRIBUTION OF A RANDOM VARIABLE ABOUT ITS MEAN.

IN OTHER WORDS, SKEWNESS IS A MEASURE OF HOW MUCH THE PROBABILITY DISTRIBUTION OF A RANDOM VARIABLE DEVIATES FROM THE NORMAL DISTRIBUTION.

IF THE RANDOM VARIABLE IS NORMALLY DISTRIBUTED, THEN SKEWNESS IS EQUAL TO ZERO.

HOWEVER WE CAN OBSERVE:

- POSITIVE SKEWNESS
- NEGATIVE SKEWNESS



→ NOW WE WANT TO TEST IF THE SKEWNESS $S(x)$ OF THE DISTRIBUTION OF RETURNS IS SIGNIFICANTLY DIFFERENT FROM ZERO OR NOT.

- WE KNOW THAT THE SAMPLE SKEWNESS
  $\hat{S}(x) = 0.061387$

- WE PERFORM THE TEST AT 5% SIGNIFICANCE LEVEL

$$H_0: S(x) = 0$$
$$H_1: S(x) \neq 0$$

- WE CONSTRUCT THE TEST STATISTIC:

$$t = \frac{\hat{S}(x)}{\sqrt{6/n}} = \frac{0.061387}{\sqrt{6/9843}} = 2.4864$$

- <u>RULE:</u> WE REJECT H_0 AT 5% SIGNIFICANCE LEVEL IF:

$$|t| > t_{2.5\%} \sim 2$$

- IN OUR CASE:

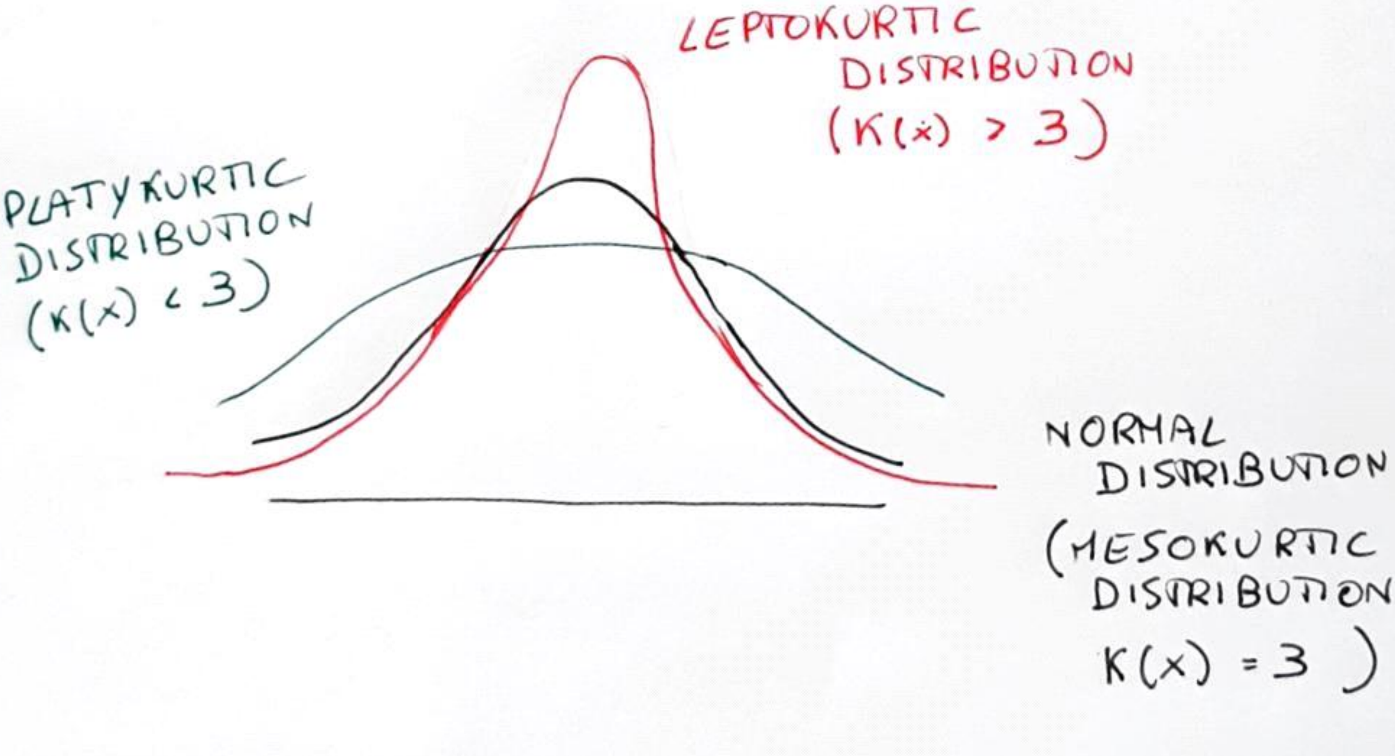
$t = 2.4864 > 2$ THEREFORE THE TEST SHOWS THAT THERE IS EVIDENCE IN THE DATA TO REJECT THE NULL HYPOTHESIS OF ZERO SKEWNESS.

Comment: THE DISTRIBUTION HAS SOME SMALL POSITIVE SKEWNESS. THE NUMBER OF OBSERVATIONS IS VERY LARGE, SO THIS COULD NOT BE SEEN FROM THE HISTOGRAM

iii) IS THE EXCESS KURTOSIS EQUAL TO ZERO?

KURTOSIS IS A STATISTICAL MEASURE THAT DEFINES HOW HEAVILY THE TAILS OF A DISTRIBUTION DIFFER FROM THE TAILS OF A NORMAL DISTRIBUTION.

IN A NORMAL DISTRIBUTION, KURTOSIS $K(x)$ IS EQUAL TO 3.



→ Be coreful TO THE TERMINOLOGY USED:

- THE KURTOSIS OF A NORMAL DISTRIBUTION IS EQUAL TO 3
- THE EXCESS KURTOSIS OF A NORMAL DISTRIBUTION IS EQUAL TO ZERO
- THEREFORE, THE EXCESS KURTOSIS IS FOUND BY USING THE FORMULA:

$$\text{excess Kurtosis} = \text{Kurtosis} - 3$$

→ NOW WE WANT TO TEST IF THE EXCESS KURTOSIS OF THE DISTRIBUTION OF RETURNS IS SIGNIFICANTLY DIFFERENT FROM ZERO

- WE KNOW THAT THE SAMPLE KURTOSIS $\hat{K}(x) = 12.91636$
- WE PERFORM THE TEST AT 5% SIGNIFICANCE LEVEL

$$H_0 : K(x) - 3 = 0$$

$$H_1 : K(x) - 3 \neq 0$$

- WE CONSTRUCT THE TEST STATISTIC:

$$t = \frac{\hat{K}(x) - 3}{\sqrt{24/n}} = \frac{12.91 - 3}{\sqrt{24/9843}} = 200.6928$$

- RULE: WE REJECT H_0 AT 5% SIGNIFICANCE LEVEL IF:

$$|t| > t_{2.5\%} \sim 2$$

- IN OUR CASE:

$t = 200.6928 > 2$ THEREFORE WE REJECT H_0.

WE REJECT THE NULL HYPOTHESIS OF NO EXCESS KURTOSIS

- comment: WE FOUND OUT THAT

$K(x) - 3 > 0$ or $K(x) > 3$.

A POSITIVE EXCESS KURTOSIS INDICATES A LEPTOKURTIC DISTRIBUTION.

THE LEPTOKURTIC DISTRIBUTION SHOWS HEAVY TAILS ON EITHER SIDE, INDICATING LARGE OUTLIERS.

THEREFORE, RETURNS ARE NOT NORMALLY DISTRIBUTED.

iv) IS THE DISTRIBUTION OF THE RETURNS NORMAL ?

THE <u>JARQUE - BERA</u> TEST IS A TEST FOR NORMALITY.
IN PARTICULAR, THE TEST MATCHES THE SKEWNESS AND KURTOSIS OF DATA TO SEE IF THEY ARE NORMALLY DISTRIBUTED.

→ $H_0: S(x) = 0$ AND $K(x) - 3 = 0$

$H_1: S(x) \neq 0$ OR $K(x) - 3 \neq 0$

IN OTHER WORDS:

H_0: NORMAL DISTRIBUTION

H_1: DISTRIBUTION IS NOT NORMAL

- EVIEWS PROVIDES THE CALCULATED TEST STATISTIC (which is 40335.5) AND THE CORRESPONDING P. VALUE (also called probability)

- <u>RULE</u>:

IF P-VALUE $< \alpha$ → REJECT H_0

IF P. VALUE $\geq \alpha$ → FAIL TO REJECT H_0

→ where α is THE SIGNIFICANCE LEVEL

- WE HAVE P_VALUE = 0.0000 AND WE CONSIDER 5% SIGNIFICANTE LEVEL ($\alpha = 5\%$)

THEREFORE:

P_VALUE $< \alpha$

$0.0000 < 0.05 \rightarrow$ REJECT H_0

$\downarrow$

THE TEST INDICATES THAT THE DISTRIBUTION IS NOT NORMAL

c) HISTOGRAM AND THE NORMAL PROBABILITY DISTRIBUTION SHOW THAT THE NORMAL DENSITY DOES NOT FIT WELL THE HISTOGRAM IN THE PEAK REGION

THE TAILS OF THE HISTOGRAM ARE SAID TO BE HEAVIER THAN THE TAILS OF THE NORMAL DISTRIBUTION.

- THE PROBABILITY DENSITY FUNCTION FOR GAUSSIAN (NORMAL) RANDOM VARIABLE WITH MEAN (μ) AND VARIANCE (σ^2)

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

d)

	DAILY RTN	AGGREGATE RTN
MEAN	0.000402	0.008448
ST. DEV.	0.016927	0.073452
SKEWNESS	0.061387	0.052078
KURTOSIS	12.91636	4.689412
MIN	-0.22963	-0.28805
MAX	0.131636	0.326204

- THE MEAN OF DAILY RETURNS IS CLOSE TO ZERO WHILE FOR AGGREG. RETURNS IS SLIGHTLY HIGHER.
- AGGR. RETURNS HAVE HIGHER ST. DEVIATION THAN DAILY RETURNS.
- SKEWNESS IS CLOSE TO ZERO (AND POSITIVE) FOR DAILY AND AGGR. RETURNS
- DAILY RETURNS TEND TO HAVE HIGHER EXCESS KURTOSIS THAN AGGREGATE RETURNS
- BOTH RETURNS ARE NOT NORMALLY DISTRIBUTED
- THE MINIMUM AND MAXIMUM OF RETURN SERIES CAN BE SUBSTANTIAL.

THE NEGATIVE EXTREME RETURNS ARE IMPORTANT IN RISK MANAGEMENT.
POSITIVE EXTREME RETURNS ARE CRITICAL IN HOLDING A SHORT POSITION