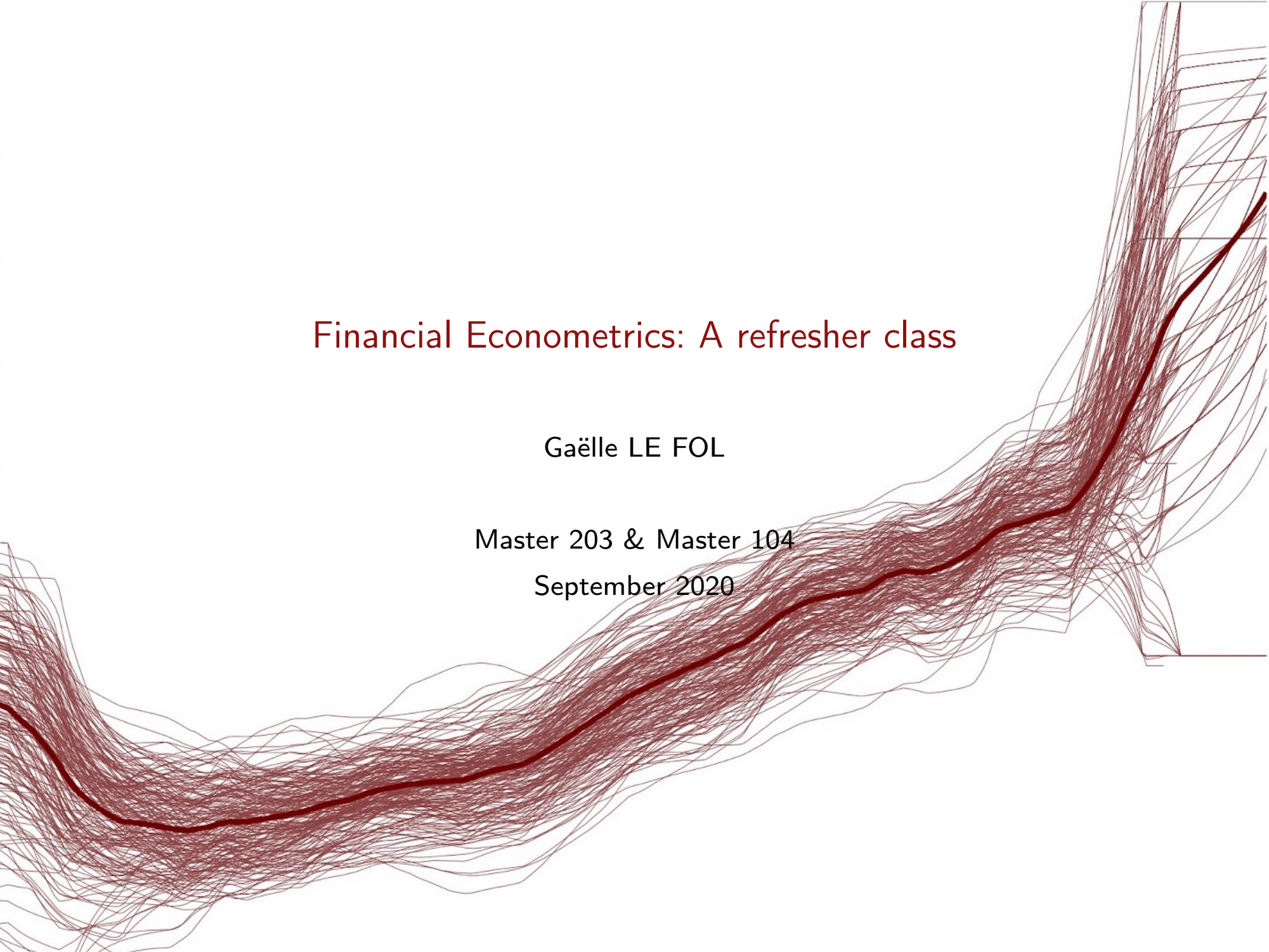
Financial Econometrics: A refresher class

Gaëlle LE FOL

Master 203 & Master 104

September 2020



Outline

Statistical inference : the idea I

- Suppose that there is a *Data Generating Process* (DGP), which generates the sample data. This is called the population.
- The sample data is then summarized in an econometric model for which key parameters of the (conditionnal) distribution of the observations are estimated.
- We further want to assign some confidence to these estimates.

Example :

The mean of the observations is estimated to range from 3% to 5% with 95% of confidence $\Longleftrightarrow$ the probability that the mean does not lie between 3 and 5 % is 5% (risk tolerance).

- The smaller the risk tolerance, the more precise the prediction of the parameter but the higher the possibility for the parameter to be out of the confidence interval.

Statistical inference allows us to answer questions such as :

- Since the estimated parameters are calculated from the sample, how are these parameters going to change if I change the sample ?
- What exactly is the (statistical) link between the "true"' parameters and the estimated ones ?
- How to test the significance of one or more parameters ?

Example : The fund example (ctd)

Fund example

- The estimated regression line for that fund was :

$$\hat{y}_t = 0.154 + 1.2578 x_t$$

- We can calculate the standard errors (See equations) of the estimated parameters and we get $\hat{\sigma}_{\hat{\beta}_0} = 0.087$ and $\hat{\sigma}_{\hat{\beta}_1} = 0.5886$.
- $\hat{\beta}_1$ is a point estimation of the true parameter β_1 of the population. Is this estimation reliable ? Ask the standard deviation of the estimated parameters.
- We use the information from the sample to draw conclusions (inferences) about the population.
- We run hypothesis tests characterized by a hypothesis called the null hypothesis H_0 and an alternative hypothesis H_a.

Statistical inference : the idea III

If the CAPM is true, the excess return of any fund only depends on the excess return of the benchmark and the constant is null.

- Two-side test :

$$\begin{cases} H_0 : \beta_0 = 0 \\ H_a : \beta_0 \neq 0 \end{cases}$$

- One-side tests :

$$\begin{cases} H_0 : \beta_0 = 0 \\ H_a : \beta_0 > 0 \end{cases} \quad \text{or} \quad \begin{cases} H_0 : \beta_0 = 0 \\ H_a : \beta_0 < 0 \end{cases}$$

- Significativity test or confidence intervals.

Distribution of the estimated parameters I

- Suppose that the errors are distributed as a normal $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. From page 73 See equations, we know :

$$\hat{\beta}_0 \sim N\left(\beta_0, \sigma_\varepsilon^2 \left\{ \frac{\sum x_t^2}{T \sum (x_t - \bar{x})^2} \right\}\right), \text{ and } \hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma_\varepsilon^2}{\sum (x_t - \bar{x})^2}\right)$$

The estimated parameters have a normal distribution :

$$\hat{\beta}_0 \sim N\left(\beta_0, \sigma_{\hat{\beta}_0}^2\right), \text{ and } \hat{\beta}_1 \sim N\left(\beta_1, \sigma_{\hat{\beta}_1}^2\right)$$

The variables

$$\frac{\hat{\beta}_0 - \beta_0}{\sigma_{\hat{\beta}_0}} \sim N(0,1) \text{ and } \frac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} \sim N(0,1)$$

However, since the variances σ_ε^2 is unknown, it has to be replaced by its empirical counterpart :

$$\widehat{\sigma}_\varepsilon^2 = \frac{1}{T-2} \sum_{t=1}^{T} \hat{\varepsilon}_t^2 \text{ with } \hat{\varepsilon}_t \sim N(0, \sigma_\varepsilon^2),$$

$$(T-2)\frac{\widehat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \sim \chi_{T-2}^2$$

Distribution of the estimated parameters II

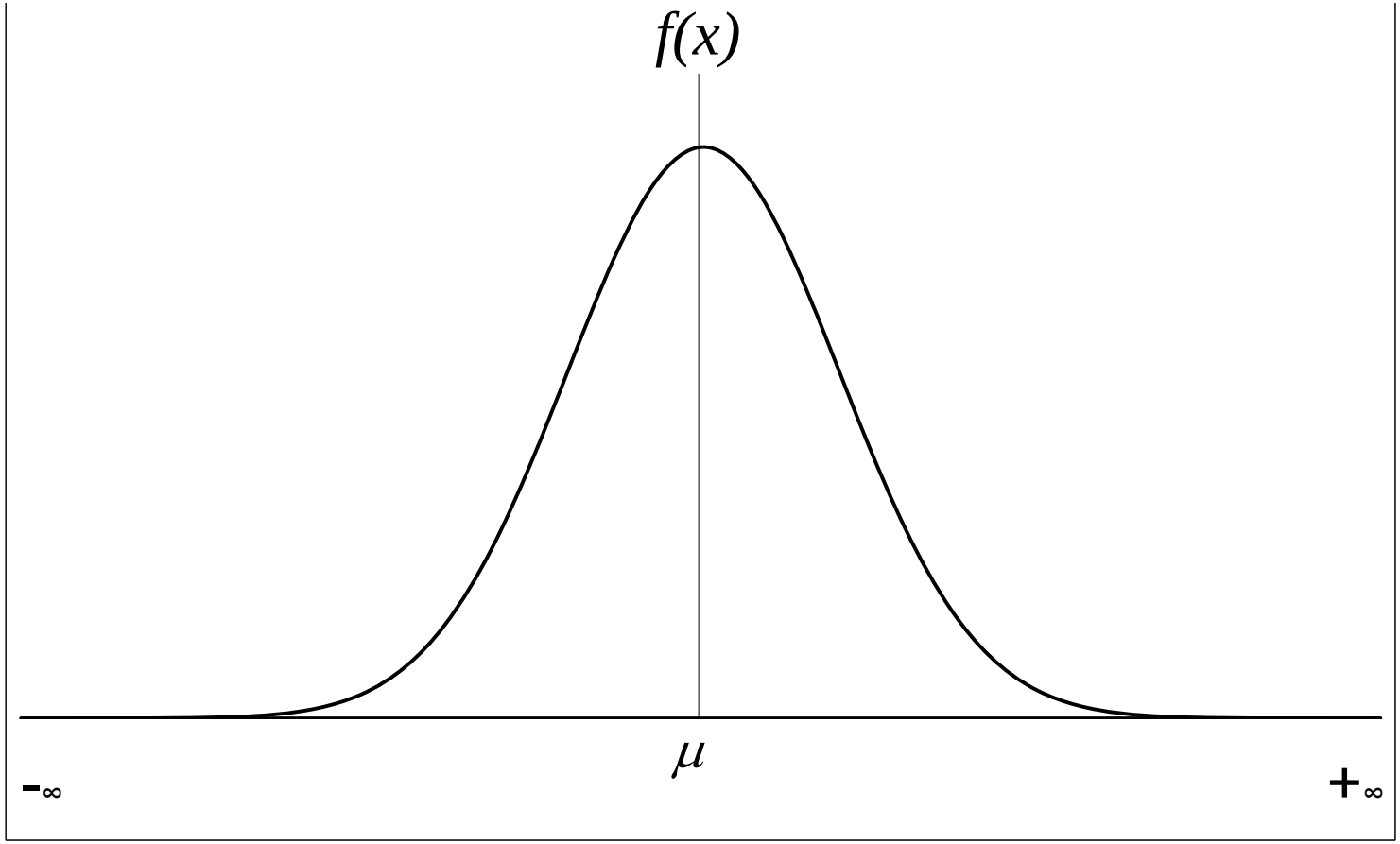
- The variables

$$\frac{\hat{\beta}_0 - \beta_0}{\widehat{\sigma}_{\hat{\beta}_0}} \sim t_{T-2} \text{ and } \frac{\hat{\beta}_1 - \beta_1}{\widehat{\sigma}_{\hat{\beta}_1}} \sim t_{T-2}$$

- We need some tabulated distribution (t-tables critical values) with which to compare the estimated test statistics.
- We need to choose a "significance level", often denoted α. This is also sometimes called the size of the test and it determines the region where we will reject, or not reject, the null hypothesis that we are testing.

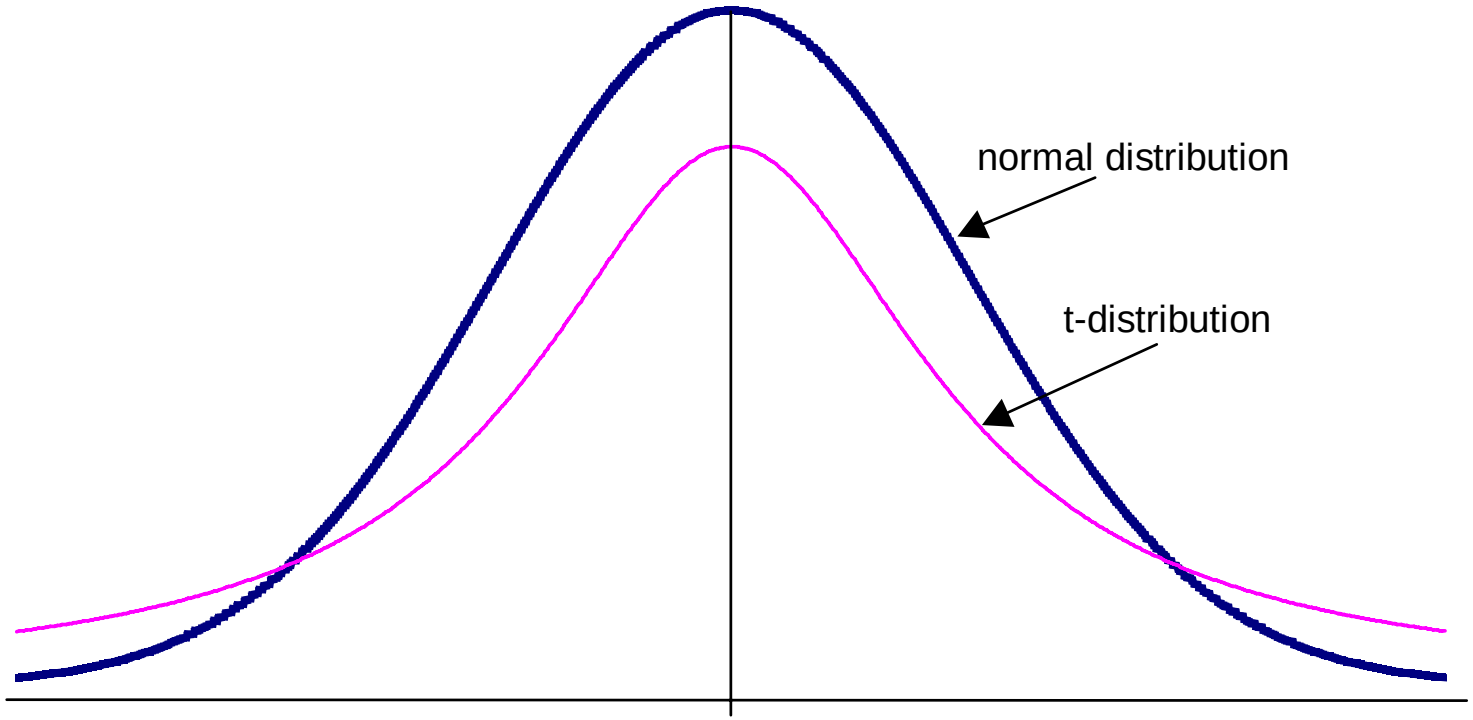
Distribution of the estimated parameters III

Figure : Normal distribution



Distribution of the estimated parameters IV

Figure : Normal and Student distributions



Significativity test I

Conducting a significativity test :

1. We calculate $\hat{\beta}_0$, $\hat{\beta}_1$, $\widehat{\sigma}_{\hat{\beta}_0}$ and $\widehat{\sigma}_{\hat{\beta}_1}$.
2. We calculate the statistics of the test

$$\frac{\hat{\beta}_1 - \beta_1^*}{\widehat{\sigma}_{\hat{\beta}_1}},$$

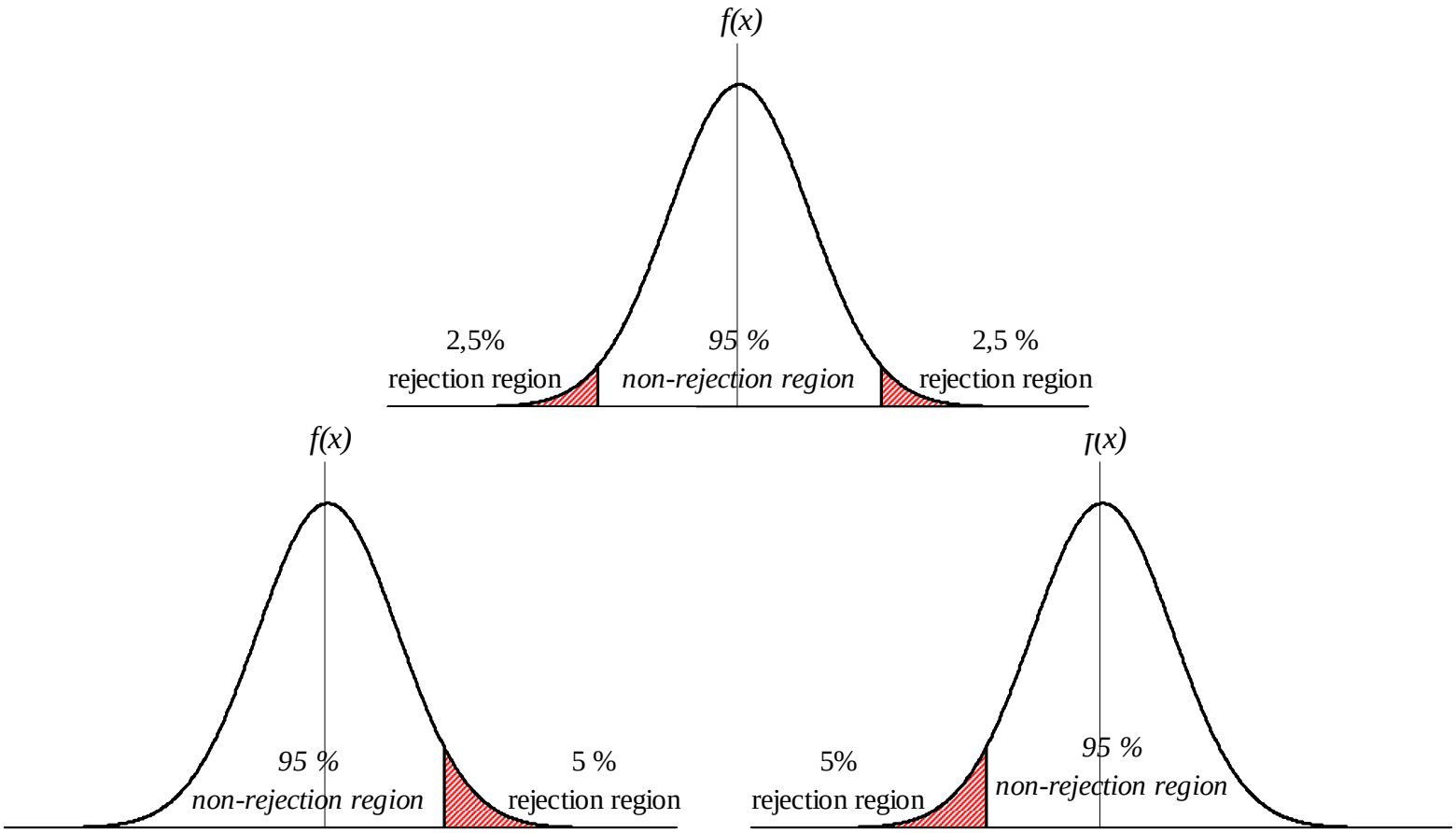
where β_1^* is the value of β_1 under the null H_0. For example, if $\beta_1^* = 0$, we have :

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_a : \beta_1 \neq 0 \end{cases}$$

3. We choose a significativity level (usually 5% in economics and 1% in finance).
4. We compare to a tabulated Student distribution with T-2 (2 estimated parameters) and read in the corresponding table, the critical value.
5. We reject the null if the statistics is greater then the critical value.

Significativity test II

Figure : Hypothesis testing



Interval estimation : some basics ideas

- We have seen that the point estimate of β_1 can differ from its true value because of sampling fluctuations but in repeated sampling we have $E[\hat{\beta}_1] = \beta_1$.
- The reliability of the point estimator is measured by its standard error.
- We may construct an interval around the estimator which has a high probability of containing β_1 : interval estimation.
- The probability that the random interval $(\hat{\beta}_1 - \delta, \hat{\beta}_1 + \delta)$ contains the true β_1 is $1 - \alpha$:

$$Pr(\hat{\beta}_1 - \delta \leq \beta_1 \leq \hat{\beta}_1 + \delta) = 1 - \alpha \tag{25}$$

- This confidence interval is caracterized by its level of significance (α), and its confidence limits $\hat{\beta}_1 - \delta$ and $\hat{\beta}_1 + \delta$.

Confidence intervals for regression coefficients

- We have seen that

$$\frac{\hat{\beta}_1 - \beta_1^*}{\hat{\sigma}_{\hat{\beta}_1}} \sim t_{T-2}$$

- We can use the t-distribution to establish a confidence interval for β_1 :

$$Pr(-t_{\alpha/2} \leq t \leq t_{\alpha/2}) = 1 - \alpha$$

$$Pr\left[-t_{\alpha/2} \leq \frac{\hat{\beta}_1 - \beta_1^*}{\hat{\sigma}_{\hat{\beta}_1}} \leq t_{\alpha/2}\right] = 1 - \alpha, \tag{26}$$

where β_1^* is the value of β_1 under the null, and where $-t_{\alpha/2}$ and $t_{\alpha/2}$ are the values of t obtained from the t-table for $(\alpha/2)$ level of significance and $T-2$ degrees of freedom.

- Equivalently, we have

$$Pr\left[\hat{\beta}_1 - t_{\alpha/2}\hat{\sigma}_{\hat{\beta}_1} \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2}\hat{\sigma}_{\hat{\beta}_1}\right] = 1 - \alpha, \tag{27}$$

which provides a $100(1-\alpha)$ confidence interval for β_1 :

$$\hat{\beta}_1 \pm t_{\alpha/2}\hat{\sigma}_{\hat{\beta}_1}$$

Hypothesis testing using confidence interval approach I

- The idea is to construct an interval of numerical values that contains the unknown parameter.
- Calculate $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\sigma}^2_{\hat{\beta}_0}$ and $\hat{\sigma}^2_{\hat{\beta}_1}$ as before.
- Choose a significance level, say 5% $\Longleftrightarrow$ choosing a 95% confidence interval.
- Use the t-tables to find the appropriate critical value, which will again have $T-2$ degrees of freedom.
- The confidence interval is such that :

$$Pr\left[-t_{\alpha/2} \leq \frac{\hat{\beta}_1 - \beta_1^*}{\hat{\sigma}_{\hat{\beta}_1}} \leq t_{\alpha/2}\right] = 1-\alpha, \tag{28}$$

where β_1^* is the value of β_1 under the null, and where $-t_{\alpha/2}$ and $t_{\alpha/2}$ are the values of t obtained from the t-table for $(\alpha/2)$ level of significance and $T-2$ degrees of freedom.

- Equivalently, we have

$$Pr\left[\hat{\beta}_1 - t_{\alpha/2}\hat{\sigma}_{\hat{\beta}_1} \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2}\hat{\sigma}_{\hat{\beta}_1}\right] = 1-\alpha, \tag{29}$$

which gives the interval in wich β_1 will fall with $1-\alpha$ probability. If β under H_0 (β_{H_0}) is contained in the interval, we cannot reject the null of $\beta_1 = \beta_{H_0}$.

Hypothesis testing using confidence interval approach II

- This interval corresponds to the region of non-rejection of the null and the confidence limits are the critical values.

Example : Coming back to the Fund Example

The estimated regression line for that fund was :

$$\hat{y}_t = 0.154 + 1.2578 x_t,$$

with $\hat{\sigma}_{\hat{\beta}_0} = 0.087$ and $\hat{\sigma}_{\hat{\beta}_1} = 0.5886$.

- The t-statitics under the null of H_0 : The coefficient is zero, is $t = 1.7727$ for $\hat{\beta}_0$ and $t = 2.1369$ for $\hat{\beta}_1$.
- $t_{2.5\%} \times \hat{\sigma}_{\hat{\beta}_1} = 3.182 \times 0.5886 = 1.87$.
- $\hat{\beta}_1 - t_{2.5\%} \times \hat{\sigma}_{\hat{\beta}_1} = 1.2578 - 1.87 = -0.6122$ and $\hat{\beta}_1 + t_{2.5\%} \times \hat{\sigma}_{\hat{\beta}_1} = 1.2578 + 1.87 = 3.1278$

Hypothesis testing using confidence interval approach III

Figure : t-distribution, source Basic Econometrics, Gujarati.

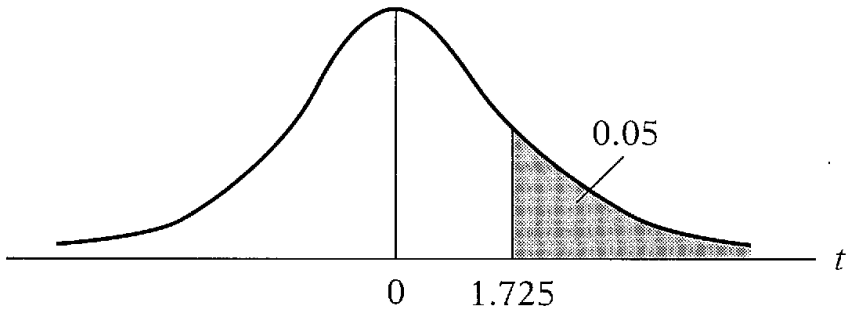
PERCENTAGE POINTS OF THE t DISTRIBUTION

Example

$\Pr(t > 2.086) = 0.025$

$\Pr(t > 1.725) = 0.05$ for df $= 20$

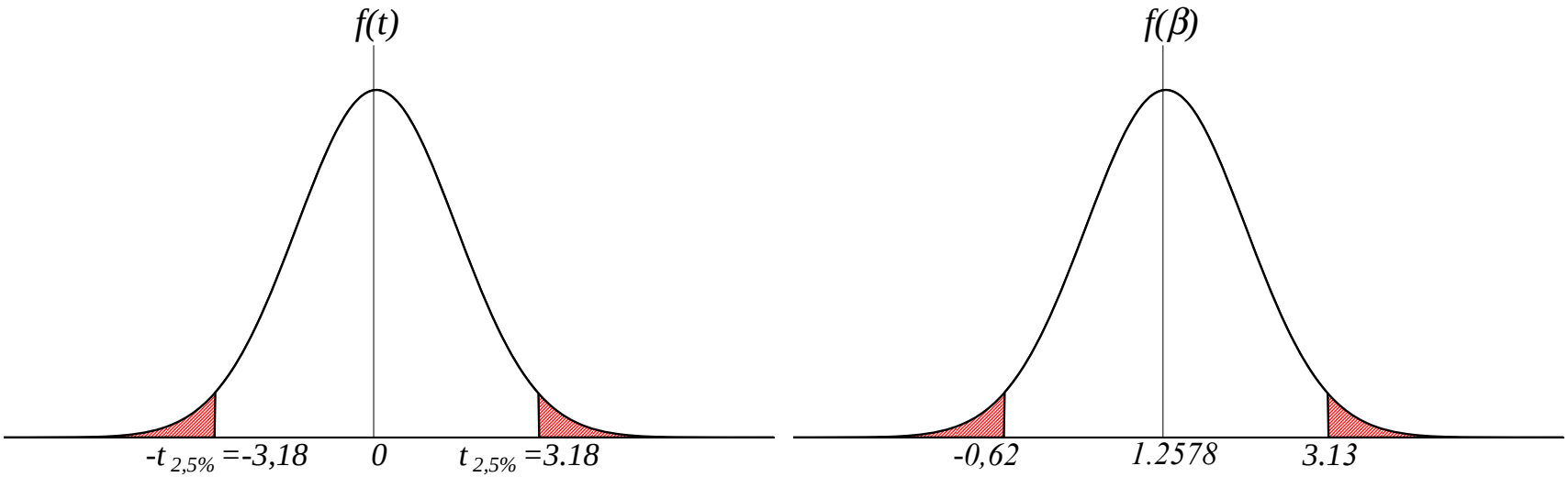
$\Pr(|t| > 1.725) = 0.10$



df \ Pr	0.25 0.50	0.10 0.20	0.05 0.10	0.025 0.05	0.01 0.02	0.005 0.010	0.001 0.002
1	1.000	3.078	6.314	12.706	31.821	63.657	318.31
2	0.816	1.886	2.920	4.303	6.965	9.925	22.327
3	0.765	1.638	2.353	3.182	4.541	5.841	10.214
4	0.741	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893
6	0.718	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297

Hypothesis testing using confidence interval approach IV

Figure : 95% confidence interval for t (3 df) - left graph-, for β_1 - right graph-.



Confidence interval for the explained variable I

- The idea is to construct an interval of numerical values that contains the value predicted/estimated by the model (Confidence interval for a prediction or an "estimation") or the future value of the explained variable (Prediction interval).
- We have $y_{t+1} = y(x_{t+1}^p) + \varepsilon_{t+1}$ where $y(x_{t+1}^p) = \beta_0 + \beta_1 x_{t+1}$ is unknown.
- The model predicted/estimated value is $y(x_{t+1}^p)$ and the OLS predicted/estimated value from our sample is $\hat{y}_{t+1} = \hat{\beta}_0 + \hat{\beta}_1 x_{t+1}$.
- The estimated error is $y(x_{t+1}^p) - \hat{y}_{t+1}$. We need to get the distribution of this error in order to build a confidence interval.
- The future value of y is noted y_{t+1} and the value predicted by our linear model from our sample is $\hat{y}_{t+1} = \hat{\beta}_0 + \hat{\beta}_1 x_{t+1}$.
- The prediction error is then $y_{t+1} - \hat{y}_{t+1}$. We need to get the distribution of this error if we want to build a prediction interval.

The confidence interval for the explained variable is :

$$\hat{y}_t \pm t_{\alpha/2} \left[\hat{\sigma}_\varepsilon \sqrt{\frac{1}{T} + \frac{(x_t - \bar{x})^2}{\sum (x_t - \bar{x})^2}} \right] \tag{30}$$

The prediction interval for the explained variable is :

$$\hat{y}_{t+1} \pm t_{\alpha/2} \left[\hat{\sigma}_\varepsilon \sqrt{1 + \frac{1}{T} + \frac{(x_{t+1} - \bar{x})^2}{\sum (x_t - \bar{x})^2}} \right] \tag{31}$$

Exercise 8 : Confidence and prediction intervals

We want to determine the distribution of the prediction error.

1. *Inject* $y(x^p_{t+1})$ *in* $y_{t+1} - \hat{y}_{t+1}$.
2. *Write the prediction error as a linear function of* $\varepsilon_{t+1}, \hat{\beta}_0$ *et* $\hat{\beta}_1$.
3. *Calculate the expectation of this error.*
4. *Calculate* $\hat{y}_{t+1} - \bar{y}$ *and rewrite what* $\hat{y}_{t+1}$ *stands for from this equation.*
5. *Calculate the variance of* $\hat{y}_{t+1}$. *Deduce the variance of the prediction error.*
6. *Give the distribution of*
$$Z = \frac{y_{t+1} - \hat{y}_{t+1}}{\sqrt{var\left(y_{t+1} - \hat{y}_{t+1}\right)}}$$
7. *Recall that if* $Z \sim N(0,1)$ *and* $v \sim \chi^2_T$, *then* $\frac{Z\sqrt{T}}{\sqrt{v}} \sim t(T)$. *Show that*
$$\hat{Z} = \frac{y_{t+1} - \hat{y}_{t+1}}{\hat{\sigma}_\varepsilon \sqrt{1 + \frac{1}{T} + \frac{(x_{t+1} - \bar{x})^2}{\sum (x_t - \bar{x})^2}}} \sim t(T-2)$$
8. *Use a similar approach to determine the distribution of the estimation error.*

The level of significance : choosing α

- Accepting or rejecting the null hypothesis depends on the α.
- α is the level of significance also called the probability of committing a type I error - the probability of rejecting the true hypothesis.
- α is linked to the probability of committing a type II error - the probability of not rejecting the false hypothesis, called β.

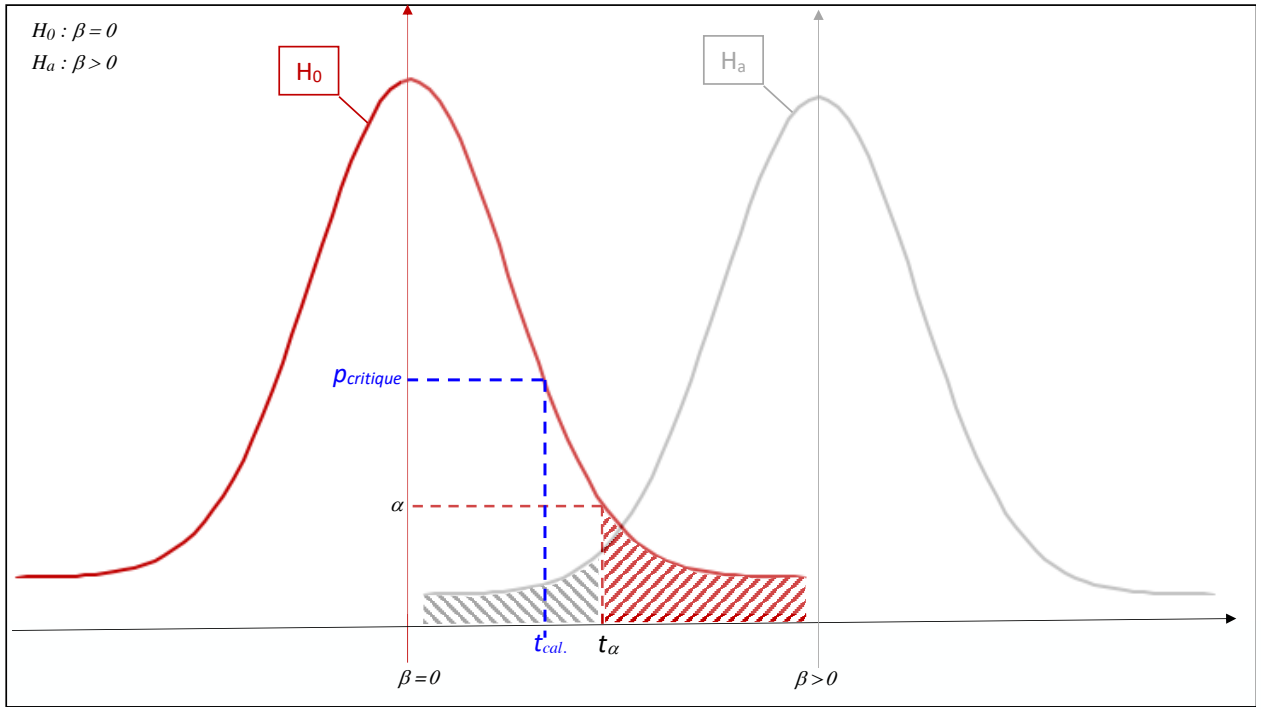
	State of nature	
	S_0	S_1
Decision	H_0 is true	H_0 is false
D_0 = Accept H_0	No error	Type II error
D_1 = Reject H_0	Type I error	No error

- $\alpha = Pr\,[\text{Reject } H_0 | H_0 \text{ is true }] = Pr\,[D_1|S_0]$ and $\beta = Pr\,[\text{Accept } H_0 | H_0 \text{ is false }] = Pr\,[D_0|S_1]$
- The probability of not committing a type II error, $1-\beta$, is called the power of the test.

The exact level of significance : the p-value I

- In hypothesis testing :
  - Calculate the statistics under the null hypothesis and compare it to the appropriate critical value (c_α) with a confidence probability of $(1 - \alpha)$.
  - Reject the null at α% if the statistics is greater than the critical value.
- Working with p-values :
  - Calculate the statistics under the null hypothesis, get the corresponding confidence level (p-value) and compare it to the risk you are willing to bear (α).
  - Reject the null at α% if the p-value is lower than your own α.

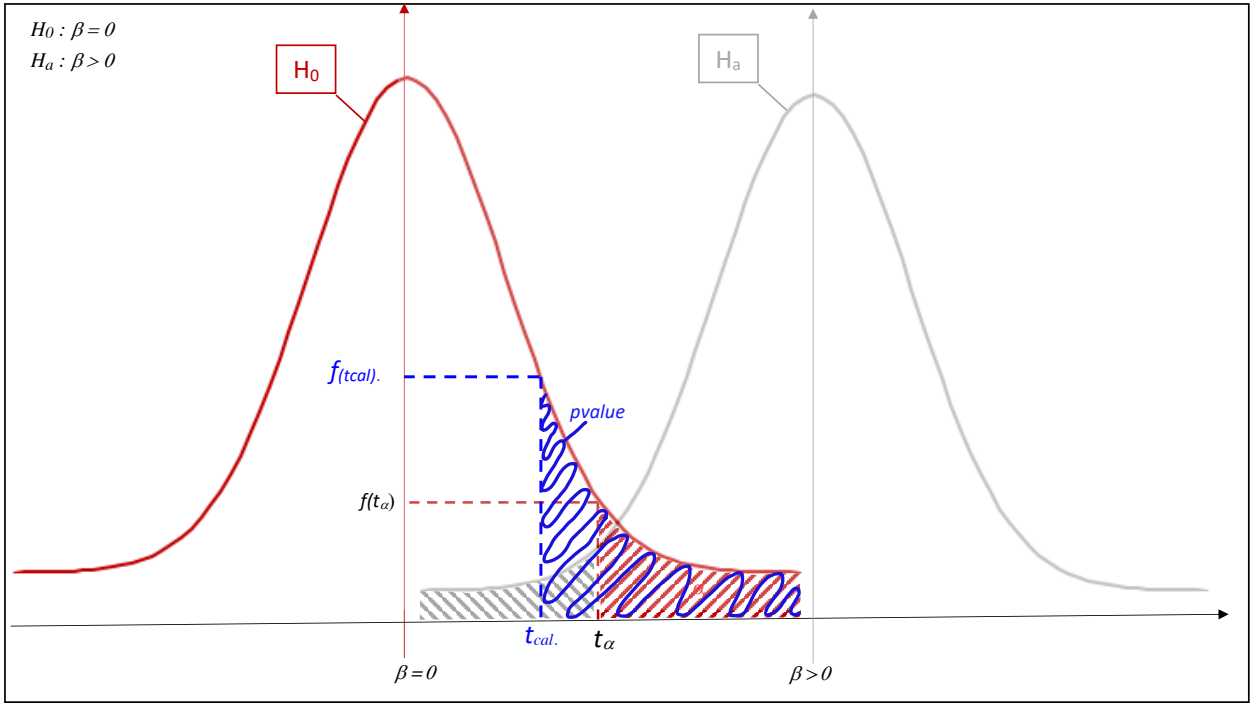
Figure : Relationship between critical value and critical probability



The exact level of significance : the p-value I

- In hypothesis testing :
  - Calculate the statistics under the null hypothesis and compare it to the appropriate critical value (c_α) with a confidence probability of $(1 - \alpha)$.
  - Reject the null at α% if the statistics is greater than the critical value.
- Working with p-values :
  - Calculate the statistics under the null hypothesis, get the corresponding confidence level (p-value) and compare it to the risk you are willing to bear (α).
  - Reject the null at α% if the p-value is lower than your own α.

Figure : Relationship between critical value and critical probability



The exact level of significance : the p-value II

Figure : t-distribution, source Basic Econometrics, Gujarati.

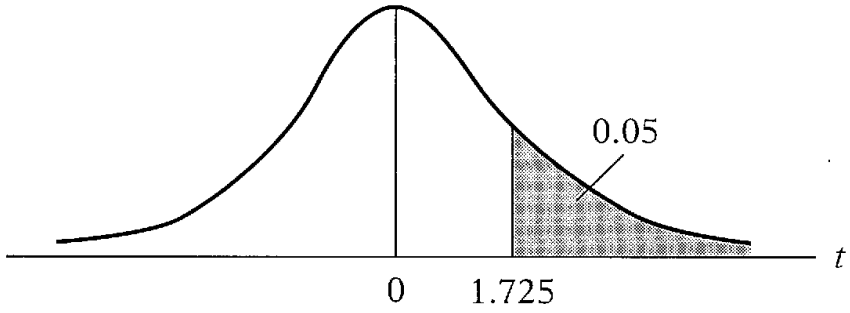
PERCENTAGE POINTS OF THE t DISTRIBUTION

Example

$\Pr(t > 2.086) = 0.025$

$\Pr(t > 1.725) = 0.05$ for df $= 20$

$\Pr(|t| > 1.725) = 0.10$



df \ Pr	0.25 0.50	0.10 0.20	0.05 0.10	0.025 0.05	0.01 0.02	0.005 0.010	0.001 0.002
1	1.000	3.078	6.314	12.706	31.821	63.657	318.31
2	0.816	1.886	2.920	4.303	6.965	9.925	22.327
3	0.765	1.638	2.353	3.182	4.541	5.841	10.214
4	0.741	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893
6	0.718	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297

Example : Coming back to the Fund Example

(Again) The estimated regression line for that fund was :

$$\hat{y}_t = 0.154 + 1.2578 x_t,$$

with $\hat{\sigma}_{\hat{\beta}_0} = 0.087$ and $\hat{\sigma}_{\hat{\beta}_1} = 0.5886$.

- The t-statitics under the null of H_0 : The coefficient is zero, is $t = 1.7727$ for β_0 and $t = 2.1369$ for β_1.
- The t-tabulated value for a confidence level of 5% and $T - 2$ degrees of freedom is $t = 3.182$.
- $t_{2.5\%} \times \sigma_{\hat{\beta}_1} = 3.182 \times 0.5886 = 1.87$.
- The associated p-value for β_1 is in between 10 and 20 %(12.22%).
- If our risk tolerance is 5%, we do not reject the null hypothesis of $\beta_1 = 0$.