Econometrics – Massimiliano Marcellino

What is Econometrics?

Econometrics deals with the **quantitative study of economic relations**. It is a tool to interpret reality in the light of economic theory, using statistical techniques.
The starting point is always a question that requires a quantitative answer.

- **Example 1**, which fraction of aggregate disposable income is consumed, and which spared?
- **Example 2**, what is the effect of a reduction in the rate of pollution on health spending?

Once we have identified a problem we want to tackle, we must assess whether and at what level of accuracy economic theory provides guidance for its solution.

Elements of an Econometric Study

Example 1, which fraction of aggregate disposable income is consumed, and which spared?

- Keynesian theory links the aggregate consumption level to that of aggregate disposable income: $C = C_0 + c \times Y_d$
- Having established an economic theory of reference, the next step is to develop an **econometric model**.
- Taking the Keynesian theory as valid, an econometric model is expressible as:

$$E(C) = C_0 + c \times Y_d$$

$$C - E(C) \sim N(0, \sigma_C^2)$$

Combining,

$$C \sim N(C_0 + c \times Y_d, \sigma_C^2)$$

The **econometric model**

$$C \sim N(C_0 + c \times Y_d, \sigma_C^2)$$

- Allows a **more accurate description of the economic reality** (by providing a value for the parameters C_0 and c);
- Allows us to **test hypotheses about the validity or otherwise of the underlying economic theory**: there is no consensus about the fact that Keynesian theory is correct; for instance, the alternative life cycle theory suggests that consumption depends not only on income but also on wealth.

Example 2, what is the effect of a reduction in the rate of pollution on health spending?

- In this case, there is no specific economic theory to which we can refer to.
- The role of econometrics is, yet, all the more important, because it allows to **derive empirical regularities** from the analysis of the economic reality, which can then **provide an opportunity to develop an appropriate economic theory**.
- Although economic theory does not help us, we can assume that the expected value of health spending (SS) is still linearly related to the pollution level (LI).

$$E(SS) = a + b \times LI$$

$$SS - E(SS) \sim N(0, \sigma_{SS}^2)$$

Combining,

$$SS \sim N(a + b \times LI, \sigma_{LI}^2)$$

In order to apply **statistical theory** to the **econometric model**, it is necessary to collect a sample of **data** that provides relevant information about the **parameters** of the model.

- **Example 1**, In order to determine which fraction of disposable income is consumed and which spared (i.e., estimate the parameters C_0 and c), we need data on aggregate consumption, on disposable income and, possibly, on wealth.
- **Example 2**, In order to assess the effect of a reduction in the rate of pollution on health spending (i.e., estimate the parameters a and b), we need measurements of the rate of pollution and data on health care expenditure`s.

Once we have estimated the model parameters, it is necessary to interpret them correctly.
- The parameter C_0 indicates that if the disposable income is 0, then the expected consumption will be equal to C_0.
- If instead there is a marginal change in disposable income, then there will be a corresponding variation in the expected value of consumption given by $c \times \Delta Y_d$.

It is good to clarify 3 concepts:
- We do not know the model parameters, rather we estimate them using statistical procedures, so there is a more or less broad uncertainty around their values. For example, if the estimated value for the marginal propensity to consume, c, is 0.8, with a confidence interval at 90% for c of [0.4, 1.2], then we need to keep in mind that the effects of a change in disposable income on consumption are very uncertain.
- When the explanatory variable changes substantially, the model parameters could change too. The parameters could also change for other reasons, such as institutional changes, extreme events, or technological change.
- The casual interpretation of the results of an econometric model is sustainable only when the model is based on economic theory. If, instead, the econometric model is not based on a valid and shared economic theory, the fact that a variable x has an estimated coefficient significantly different from 0 to explain a variable y does not imply that x cause y: the link between x and y might be spurious and due for example to a third variable z which is not included in the model.

Data
In order to apply statistical techniques to the econometric model, it is necessary to have a sample of data.
- Datasets composed of a single data point for many units are referred to as **cross-sections**.
- Datasets composed of many temporal observations for the same variable are known as **time series**.
- Datasets with both longitudinal and temporal dimension are defined as **panels**.

In addition, variable can be **continuous** (as consumption, income, the rate of pollution or health spending), or **discrete** (as in the case of binary variables, e.g. the choice of whether undertake graduate studies or not).

In economics the data are usually not the result of experiments (as in physics or biology) but choices and outcomes of real actions of economic actors.
For example, we cannot ask ourselves what Mr. Smith's consumption would be if he had any level of income, we can only observe what his consumption is, given his actual level of income.
- The problem might be partly compensated by having data on many individuals (cross-section).
- Or to observe the features and choices of the same individual or unit over an extended period of time (time-series).
- Or we could observe the consumption and income of several individuals for several months (panel).

Bottom line, when we use cross-section, time-series or panel data we implicitly assume homogeneity of the econometric model across units and/or over time.
For example, it is assumed that several individuals have the same marginal propensity to consume, or that this remains unchanged over time for the same individual, or across both dimensions in the case of panel data. This assumption is required to have a sufficiently large sample of observations to ensure that statistical procedures (e.g., parameter estimation or hypothesis testing) give reliable results.

The Linear Regression Model

Definitions and Notation

We want to explain the behaviour of an economic variable Y according to that of another variable X, for a set of periods or units i = 1, …, T. We assume that this relationship is linear,

$$Y_i = \beta_1 + \beta_2 X_i \qquad i = 1, \dots, T$$

Note that we are also assuming that the relationship does not change at different times or for different units: the parameters β_1 and β_2 do not depend on the index i.
From a statistical point of view, we think of Y_i as a random variable, whose T realizations y_i, $i = 1, \dots, T$, we can observe.

The explanatory variable X can be deterministic or stochastic. In practice, in economic applications, the X variable is almost always stochastic. From the statistical point of view, it is easier to analyze the regression model by assuming that X is deterministic (many of the results obtained in the deterministic case are also valid in the stochastic case, under a proper set of additional assumptions).

Due to the randomness of Y, the linear relationship will not hold exactly: each Y_i will be only explained up to an error term E_i

$$Y_i = \beta_1 + \beta_2 X_i + E_i \qquad i = 1, \dots, T$$

Since $E_i = Y_i - \beta_1 - \beta_2 x_i$, E_i is also a random variable. Alternatively, we may think of E_i itself as a random variable, that contributes to explaining the variable Y. In this case, the randomness of the Y variable results from that of the error E_i.

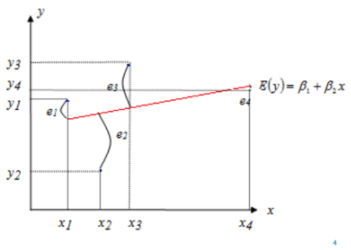
If the linear relationship is correct, we can require that, for every i, $E(E_i) = 0$.
When the variable X is deterministic $(X = x)$, it follows that:

$$E(Y_i) = \beta_1 + \beta_2 x_i \text{ and } y_i = E(Y_i) + e_i$$

Where we remind that lower case letters indicate realizations of the variables.
Therefore, the X variable helps explain the expected value of the Y variable, while the difference between the realization y_i and its expected value corresponds to the realization of the error term, e_i.

Notation:

- Y, y → dependent variable;
- X, x → independent variable / regressor / explanatory variable;
- E → error;
- e → residual / realization of the error;
- $E(Y_i)$ → systematic component / expected value of Y;
- $\beta_1 + \beta_2 x_i$ → regression function.

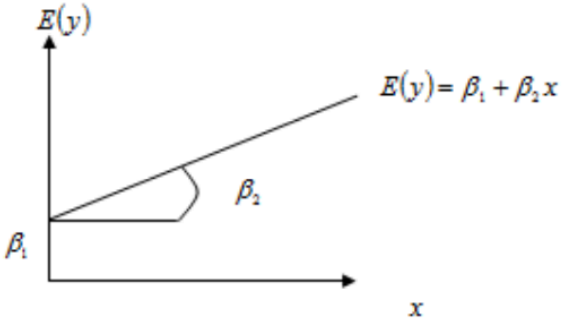


Interpretation of the Regression Function

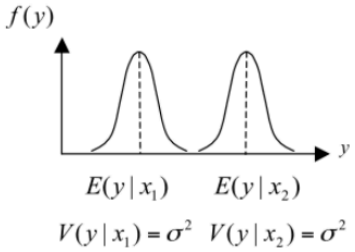
$\beta_2 = {}^{\partial E(y|X)}/_{\partial X}$

and, only if $x = 0, E(y) = \beta_1$.

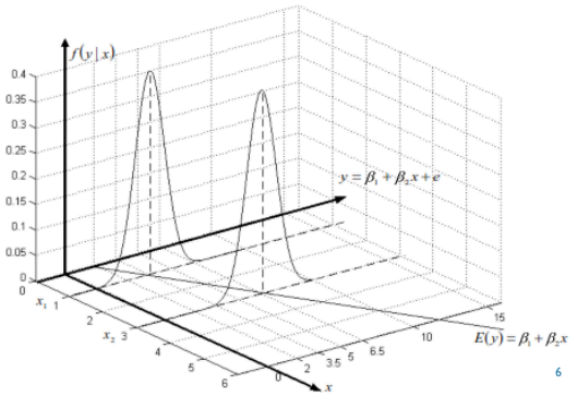
- The *intercept* β_1 measures the expected value of the dependent variable when the explanatory variable is 0.
- The *slope* β_2 indicates how the expected value of Y changes when there is a marginal variation of the regressor.



Density function of Y for given values of X.



Joint representation of x, y and density function of Y.



The Assumptions of the Linear Regression Model

- **Linearity and Stability of Parameters**: $E(Y_i \mid X_i) = \beta_1 + \beta_2 x_i$
- **Deterministic Regressors**: $E(E_i \mid X_i) = E(E_i)$
- **Homoscedasticity**: $V(E_i) = V(Y_i - \beta_1 - \beta_2 X_i) = V(Y_i) = \sigma^2$
- **No Correlation**: $cov(E_i, E_j) = cov(Y_i, Y_j) = 0,\ i \neq j$
- **There exist** i *and* j **such that** $x_i \neq x_j$
- **Normality**: $E_j \sim N(0, \sigma^2)$ or, equivalently, $Y_i \sim N(\beta_1 + \beta_2 x_i,\ \sigma^2)$

Assumptions (1) to (5) are referred to as **weak OLS hypotheses**.
Assumptions (1) to (6) are referred to as **strong OLS hypotheses**.

The Role of the Error Term

One reason for including the error term in the linear model is the randomness of the dependent variable, which allows explaining its expected value but not its actual realizations. Other (not mutually exclusive) reasons are:

- **Measurement errors in variables**
  For example, if aggregate private consumption depends on the unobservable permanent income (X) rather than on the observable current income (X^*), and if the relationship between the two notions of income is $X_i^* = X_i + V_i$, where V_i is the measurement error, then, even if the relationship between Y and X is deterministic, we have:

$$Y_i = \beta_1 + \beta_2 X_i^* - \beta_2 V_i = \beta_1 + \beta_2 X_i^* + E_i$$

- **Heterogeneity**
  Suppose that we want to analyze the consumption behaviour (Y) of two families with different marginal propensities to consume.

$$Y_1 = \alpha + \beta_1 x_1 \qquad Y_2 = \alpha + \beta_2 x_2$$

  By mistake, we assume that the propensities to consume are the same and equal to β. Then, the models become:

$$y_1 = \alpha + \beta x_1 + (\beta_1 - \beta)x_1 = \alpha + \beta x_1 + e_1 \quad y_2 = \alpha + \beta x_2 + (\beta_2 - \beta)x_2 = \alpha + \beta x_2 + e_2$$

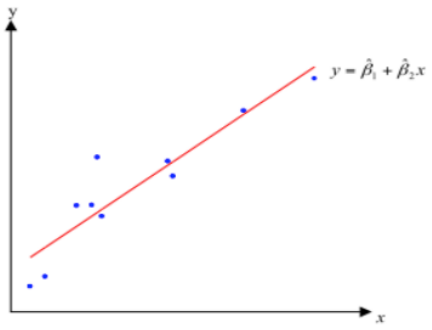
  And the errors e_1 and e_2 are cause by the heterogeneity of the coefficients.

- **Omitted Variables**
  In the example about aggregate consumption (Y), if income (X) is not the only significant explanatory variable for Y but, for example, wealth (Z) also amtters, then the error term reflects the consequences of the omission of Z from the model of Y.

- **Specification Error**
  If the relationship between consumption and income is not linear but governed by a generic function, such as $Y = g(X)$, then the error in the linear model captures the approximation error we make when using a linear relationship to model the dependence of Y on X instead of the proper nonlinear function.

Regardless of the source of the error term, the relevant question is whether the weak or strong OLS assumptions are met. We suppose that this is the case, for now.

OLS Estimators of the Regression Function Parameters
The β_1 and β_2 parameters of the linear regression function are not known and must be estimated.



For this, we use the assumption of **stable parameters** over time or across different sample units, so that we can use the available T observations on Y and X to gather information on β_1 and β_2.

Our aim is to choose β_1 and β_2 in order to "draw" a line that passes as close as possible to all the observations points in the graph.

OLS Estimators of the Regression Function Parameters – Derivation
We want to **minimize the sum of squared residuals**:

$$\min_{\beta_1 \beta_2} S = \sum_{t=1}^{T} e_t^2 = \sum_{t=1}^{T} (Y_t - \beta_1 - \beta_2 X_t)^2$$

- We first obtain the **first-order conditions** by differentiating the objective function w.r.t. β_1 and β_2.

$$\partial S / \partial \beta_1 = \sum_{t=1}^{T} 2(Y_t - \beta_1 - \beta_2 X_t)(-1) = 2T\beta_1 - 2\sum_{t=1}^{T} Y_t + 2\beta_2 \sum_{t=1}^{T} X_t$$

$$\partial S / \partial \beta_2 = \sum_{t=1}^{T} 2(Y_t - \beta_1 - \beta_2 X_t)(-X_t) = -2\sum_{t=1}^{T} X_t Y_t + 2\beta_1 \sum_{t=1}^{T} X_t + 2\beta_2 \sum_{t=1}^{T} X_t^2$$

- By equating the F.O.C.s to 0 we have:

$$T\hat{\beta}_1 + \hat{\beta}_2 \sum_{t=1}^{T} X_t = \sum_{t=1}^{T} Y_t$$
$$\hat{\beta}_1 \sum_{t=1}^{T} X_t + \hat{\beta}_2 \sum_{t=1}^{T} X_t^2 = \sum_{t=1}^{T} X_t Y_t$$

- If we multiply the first equation by $\sum_{t=1}^{T} X_t$ and the second equation by T, and we subtract the second from the first we obtain:

$$\hat{\beta}_2 \sum_{t=1}^{T} X_t \sum_{t=1}^{T} X_t - \hat{\beta}_2 T \sum_{t=1}^{T} X_t^2 = \sum_{t=1}^{T} Y_t \sum_{t=1}^{T} X_t - T \sum_{t=1}^{T} X_t Y_t$$

- Therefore, the **estimators $\widehat{\boldsymbol{\beta}}_1$** and **$\widehat{\boldsymbol{\beta}}_2$** are equal to:

$$\hat{\beta}_2 = \frac{T \sum_{t=1}^{T} X_t Y_t - \sum_{t=1}^{T} X_t \sum_{t=1}^{T} Y_t}{T \sum_{t=1}^{T} X_t^2 - \left(\sum_{t=1}^{T} X_t\right)^2}$$

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{T} Y_t}{T} - \hat{\beta}_2 \frac{\sum_{t=1}^{T} X_t}{T}$$

OLS Estimators of the Regression Function Parameters: Discussion

$\widehat{\boldsymbol{\beta}}_1$ and **$\widehat{\boldsymbol{\beta}}_2$** depend on the random variable Y and therefore are themselves **random variables**.

- If we replace the random variable Y by its realizations, the estimators become estimates (observed values):

$$\hat{\beta}_2 = \frac{T \sum_{t=1}^{T} x_t y_t - \sum_{t=1}^{T} x_t \sum_{t=1}^{T} y_t}{T \sum_{t=1}^{T} x_t^2 - \left(\sum_{t=1}^{T} x_t\right)^2}$$
$$\hat{\beta}_1 = \frac{\sum_{t=1}^{T} y_t}{T} - \hat{\beta}_2 \frac{\sum_{t=1}^{T} x_t}{T} = \bar{y} - \hat{\beta}_2 \bar{x}$$

  Where $\bar{x}$ indicates the sample mean.

Note that if $X_t = x$, i.e. there is not enough variability in the X variable, **$\widehat{\boldsymbol{\beta}}_2$** is not properly defined (**hypothesis 5**). Furthermore, as **$\widehat{\boldsymbol{\beta}}_2$** is based on cross moments of X and Y, a **casual interpretation of the parameters is not granted**.

Use of Different Loss Functions

What would happen if we didn't use the **quadratic loss function** $\sum_{j=1}^{T} e_t^2$?

- If we used the **simple sum of errors terms** $\sum_{j=1}^{T} e_j$, we could get a **compensation of positive and negative errors**, which we want to avoid.
- If we used the **absolute values** $\sum_{j=1}^{T} |e_j|$, the compensation would not occur, but the derivation of the estimators would be **excessively complicated** since the function is not differentiable in 0.

- More meaningfully, we could use a **loss function** such as $\sum_{j=1}^{T} e_j^4$ which gives more weight to large mistakes, or a composite function such as $a\sum_{j=1}^{T} e_j^2$ when $e < 0$ and $b\sum_{j=1}^{T} e_j^2$ when $e > 0$ which gives different importance to positive and negative errors.

However, as these cases **substantially complicate** the derivation of the parameter estimators, we will stick to the assumption of a **quadratic loss function**.

The Linear Model in Vector Notation

The **vector notation** is a useful tool for the analysis of more complex models.

- We can write the linear model in expanded form:

$$\begin{aligned} y_1 &= \beta_1 + \beta_2 x_1 + e_1 \\ y_2 &= \beta_1 + \beta_2 x_2 + e_2 \\ &\cdots \\ y_T &= \beta_1 + \beta_2 x_T + e_T \end{aligned} \quad \Rightarrow \quad \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \beta_1 + \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{bmatrix} \beta_2 + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_T \end{bmatrix}$$

$$\underset{T\times 1}{y} = \underset{T\times 2}{W}\ \underset{2\times 1}{\beta} + \underset{T\times 1}{e}, \quad W = (v : X)$$

The Linear Model in Vector Notation - OLS Assumptions

The **OLS Assumptions** can also be rewritten in vector notation:

- **Linearity and Stability of Parameters**: $E(Y|W) = W\beta$
- **Deterministic Regressors**: $E(e|W) = e$
- **Homoscedasticity and No Correlation of the Error Term**: $V(e) = E[(y - W\beta)(y - W\beta)'] = E(ee') = \sigma^2 I_T = \sigma^2 \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}$
- **No Multicollinearity**: $Rank(W'W) = 2$
- **Normality of the Error Term**: $e \sim N(0, \sigma^2 I)$, where $V(e) = E(ee')$

The Linear Model in Vector Notation – Derivation of the Estimators

The **Loss Function** to minimize is now:

$$S = \sum_{t=1}^{T} e_t^2 = e'e = [e_1 \quad \dots \quad e_T] \begin{bmatrix} e_1 \\ \vdots \\ e_T \end{bmatrix}$$

$$= (y - W\beta)'(y - W\beta) = y'y - y'W\beta - \beta'W'y + \beta W'W\beta$$

$$= y'y - 2\beta'W'y + \beta'W'W\beta$$

The **First-Order Conditions** are then: $\partial S/\partial \beta = \partial/\partial \beta\,(y'y - 2\beta'W'y + \beta'W'W\beta) = -2W'y + 2W'W\beta = 0$

And the **OLS Estimators** in vector notation are: $\hat{\beta} = (W'W)^{-1}W'y$

Finally, note that the **Second-Order Conditions** of the optimization problem are: $\partial^2 S/\partial\beta\partial\beta' = W'W$

Since $\boldsymbol{W'W}$ **is positive definite**, the OLS estimators are indeed the **minimizers** of the loss function.

Properties of OLS Estimators

- **Unbiasedness** – $\boldsymbol{E(\hat{\beta}) = \beta}$
  - Intuitively, unbiasedness means that, if we could obtain several samples of Y and X, and for each sample we estimated the parameters of the linear model by OLS, then the **average of the OLS estimators across all samples** would coincide with the **true of the parameters**.

- The OLS estimation method ensures that, if the experiment were repeated many times, on average, the estimators would give the correct answer about the parameter values.
- Under the weak OLS assumptions, and in particular assumption (I) $(E(e) = 0)$, we indeed have that:

$$\hat{\beta} = (W'W)^{-1}W'(W\beta + e) = \beta + (W'W)^{-1}W'e$$
$$\Rightarrow E(\hat{\beta}) = E[\beta + (W'W)^{-1}W'e]$$
$$= \beta + (W'W)^{-1}W'\mathrm{E(e)} = \beta$$

- **Variance – $V(\widehat{\boldsymbol{\beta}}) = (\sigma^2 W'W)^{-1}$**
  - In matrix notation, the variance can be expressed as:

$$var(\hat{\beta}) = \begin{pmatrix} var(\hat{\beta}_1) & cov(\hat{\beta}_1, \hat{\beta}_2) \\ cov(\hat{\beta}_1, \hat{\beta}_2) & var(\hat{\beta}_2) \end{pmatrix}$$
$$= E(\hat{\beta} - \beta)(\hat{\beta} - \beta)' = (W'W)^{-1}W'\underbrace{E(ee')}_{=\sigma^2 I}W(W'W)^{-1} = \sigma^2(W'W)^{-1}\underbrace{W'W(W'W)^{-1}}_{=I}$$

  - Where:

$$\sigma^2(W'W)^{-1} = \sigma^2 * inv\begin{bmatrix} T & \sum_{t=1}^{T} x_t \\ \sum_{t=1}^{T} x_t & \sum_{t=1}^{T} x_t^2 \end{bmatrix}$$

  - Therefore, we can compute the elements of the **variance-covariance matrix** as:

$$var(\hat{\beta}_1) = \sigma^2\left[\frac{\sum_{t=1}^{T} x_t^2}{T\sum_{t=1}^{T}(x_t - \bar{x})^2}\right]$$
$$var(\hat{\beta}_2) = \left[\frac{\sigma^2}{\sum_{t=1}^{T}(x_t - \bar{x})^2}\right]$$
$$cov(\hat{\beta}_1, \hat{\beta}_2) = \sigma^2\left[\frac{-\bar{x}}{\sum_{t=1}^{T}(x_t - \bar{x})^2}\right]$$

  - The greater the variance of $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$, the greater the **uncertainty about the estimated values**.
  - Hence, we would like the variance of the OLS estimators to be low, which happens when:
    - $\sum_{t=1}^{T}(X_t - \bar{x})^2$ is large, meaning that there is a lot of variation in the independent variable and so a lot of information.
    - σ^2 is small, meaning that the variability of the error term is small.
    - The sample size T increases.

- **Consistency**
  - $\lim_{T \to \infty} var(\hat{\beta}_1) = \lim_{T \to \infty} var(\hat{\beta}_2) = 0$, meaning, as it is also $E(\hat{\beta}_1) = \hat{\beta}_1$ and $E(\hat{\beta}_2) = \hat{\beta}_2$, that $\hat{\beta}_1$ and $\hat{\beta}_2$ **converge in probability towards the true values** of the parameters.
  - Differently from unbiasedness, consistency can be achieved in a **single sample**, as long as it is an almost infinitely large one.

- **Efficiency**
  - An additional property of OLS estimators under the weak OLS assumptions (1 to 5) is that it is the **Best Linear Unbiased Estimator (BLUE)**, meaning it has the **highest precision** or **lowest variance** in the set of the linear and unbiased estimators of β. This result is known as the **Gauss-Markov Theorem** (proof sketch in the next slides).
  - Interestingly, the OLS estimator could also be obtained by imposing the condition that it must be B.L.U.E instead of the one minimizing the sum of squared residuals. Indeed, **narrowing the search to the set of linear and unbiased estimators**, i.e. such that:
    - $\tilde{\beta} = [(W'W)^{-1}W' + C]y$

- $CW = 0$

We can find $\widehat{\boldsymbol{\beta}}_{\boldsymbol{OLS}}$ **as the one with minimum possible variance**.

Properties of OLS Estimators – Proof Sketch of the Gauss-Markov Theorem
Consider a generic linear combination of parameters: $a'\beta = a_1\beta_1 + a_2\beta_2$. Then for any a and for every unbiased linear estimator $\tilde{\beta}$:

$$a'var(\hat{\beta})a \leq a'var(\tilde{\beta})a$$

i.e., $var(\hat{\beta}) - var(\tilde{\beta})$ is **negative semi-definite**.
Let us now take a **generic linear estimator**:

$$\tilde{\beta} = [(W'W)^{-1}W' + C]y = [(W'W)^{-1}W' + C][W\beta + e] = \beta + (W'W)^{-1}W'e + CW\beta + Ce$$

Where we have plugged in the expression for y.
We can then compute its expected value as:

$$E(\tilde{\beta}) = \beta + 0 + CW\beta + 0$$

As $\tilde{\beta}$ must be unbiased, it must then be $CW\beta = 0$, so $CW = 0$ and therefore:

$$\tilde{\beta} - \beta = (W'W)^{-1}W'e + Ce$$

Then, the variance covariance matrix of $\tilde{\beta}$ is:

$$var(\tilde{\beta} - \beta) = E(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)' = E[(W'W)^{-1}W'e + Ce][e'C' + e'W(W'W)^{-1}]$$
$$= E[(W'W)^{-1}W'ee'C'] + E[(W'W)^{-1}W'ee'W(W'W)^{-1}] + E[Cee'C'] + E[Cee'W(W'W)^{-1}]$$
$$= \underbrace{\sigma^2(W'W)^{-1}W'C'}_{=0} + \sigma^2(W'W)^{-1} + \sigma^2CC' + \underbrace{\sigma^2CW(W'W)^{-1}}_{=0} = \sigma^2(W'W)^{-1} + \sigma^2CC'$$
$$= var(\hat{\beta} - \beta) + \sigma^2CC'$$

Therefore, $var(\hat{\beta}) - var(\tilde{\beta}) = -\sigma^2CC'$, which is negative semi-definite as required.

The Distribution of the OLS Estimator
The results seen so far **do not require** that Y or the error term are normally distributed. However, if this is the case, meaning $Y \sim N(W\beta, \sigma^2)$:

$$\hat{\beta} \sim N(\beta, \sigma^2(W'W)^{-1})$$

Even if Y is not normal, the **Central Limit Theorem** ensures that $\hat{\beta}$ is normally distributed when the sample size T is very large (in practice, T > 50 is usually enough):

$$\frac{1}{\sqrt{T}}W'e \xrightarrow{d} N(0, \sigma^2\Sigma_{WW})$$

Proof Sketch:
From: $\hat{\beta} = (W'W)^{-1}W'Y = \beta + (W'W)^{-1}W'e$.
It follows that: $\sqrt{T}(\hat{\beta} - \beta) = \left(\frac{W'W}{T}\right)^{-1}\frac{1}{\sqrt{T}}W'e \rightarrow \quad \Sigma_{WW}^{-1}N(0, \sigma^2\Sigma_{WW}) = N(0, \sigma^2\Sigma_{WW}^{-1})$

Orthogonality between OLS Fitted Values and Residuals
An additional property of the OLS estimator it is **orthogonal to the residual component** of the regression.

- We can decompose the dependent variable y into the explained component or fitted value, $\hat{y}$, and the residual component $\hat{e}$:

$$y = W\hat{\beta} + \hat{e}$$

- Then, since $\hat{e} = [I - W(W'W)^{-1}W']e$:

$$\hat{y}'\hat{e} = \hat{\beta}'W'\hat{e} = \hat{\beta}'W'[I - W(W'W)^{-1}W']e = \hat{\beta}'[W' - W']e = 0$$

- It can be shown that the OLS estimator is the **only one** for which $\hat{y}'e = 0$, i.e., the residuals cannot be used to increase the informativeness of the fitted values.

An Estimator of the Error Variance
Ideally, we would like to estimate the variance of the error term σ^2 as:

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T} e_t^2 = \frac{1}{T}e'e$$

However, the **errors are unobservable** since they depend on the unknown parameters β.
We then replace β with $\hat{\beta}$ and use the **residuals**:

$$\hat{e} = y - W\hat{\beta} = [I - W(W'W)^{-1}W']y = [I - W(W'W)^{-1}W'][W\beta + e] = Me$$

Properties of Matrix M:
- **Symmetric**: M = M'
- **Idempotent**: MM' = MM = M

$$\sum_{t=1}^{T}\hat{e}_t^2 = \hat{e}'\hat{e} = E(e'M'Me) = E(e'Me)$$
$$= E[tr(eMe')] = E[tr(Mee')] = tr[E(Mee')] = \sigma^2 tr(M)$$
$$= \sigma^2[tr(I_T) - tr(\underbrace{W'W(W'W)^{-1}}_{I_2})] = \sigma^2[T-2]$$

$\hat{\sigma}^2$
$\hat{\sigma}^2 = \frac{\hat{e}'\hat{e}}{T-2} = \frac{\sum_{t=1}^{T}\widehat{e_t^2}}{T-2}$
$\widehat{Var}(\hat{\beta})$
$\widehat{Var}(\hat{\beta}) = \hat{\sigma}^2(W'W)^{-1}$

Under the **strong OLS assumptions**:

$$\sum_{t=1}^{T}\left(\frac{e_t}{\sigma}\right)^2 \sim \chi^2(T) \Rightarrow V = \frac{\sum_{t=1}^{T}\hat{e}_t^2}{\sigma^2} = \frac{(T-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(T-2)$$

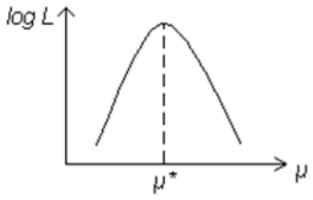
Then:
- $E(V) = E\left(\frac{(T-2)\hat{\sigma}^2}{\sigma^2}\right) = \frac{(T-2)E(\hat{\sigma}^2)}{\sigma^2} = T-2 \Rightarrow \hat{\sigma}^2$ is **unbiased**.
- $Var(V) = 2(T-2) = \frac{(T-2)^2 var(\hat{\sigma}^2)}{\sigma^4} \Rightarrow Var(\hat{\sigma}^2) = \frac{2\sigma^4}{T-2} \Rightarrow \lim_{T\to\infty} Var(\hat{\sigma}^2) = 0 \Rightarrow \hat{\sigma}^2$ is **consistent**.
- $E(\hat{e}(\hat{\beta} - \beta)') = E[(I - W(W'W)^{-1}W')ee'W(W'W)^{-1}] = 0 \Rightarrow \hat{\sigma}^2$ is **uncorrelated** with $\hat{\beta}$.

Maximum Likelihood Estimators for the Linear Model
Consider N independent and identically distributed (iid) Normal variables, with unit variance and unknown mean: $Z_i \overset{iid}{\sim} N(\mu, 1)$ We can compute their joint density function as:

$$f(Z_1, Z_2, \dots, Z_N) = f(Z_1)f(Z_2)\dots f(Z_N) = \left(\frac{1}{\sqrt{2\pi}}\right)^N e^{-\frac{1}{2}\sum_{i=1}^{N}(Z_i-\mu)^2}$$

If we consider $f(Z_1, Z_2, \dots, Z_N)$ as a function of the unknown parameter μ, for given realizations of $Z_1, Z_2, \dots, Z_N$, and we apply the log transformation, we obtain the log-likelihood function:

$$logL(\mu) = Nlog\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2}\sum_{i=1}^{N}(Z_i - \mu)^2$$



The **Maximum Likelihood Estimator (MLE)** of the parameter μ is the maximizer of the log-likelihood.
Solving analytically, the MLE coincides with the sample mean of the observations.

$$\frac{\partial logL}{\partial \mu} = \sum_{i=1}^{N}(z_i - \mu) \Rightarrow \mu^* = \frac{\sum_{i=1}^{N} z_i}{N}$$

By extending this procedure, we can obtain MLEs of the parameters of the linear regression model. Under the strong OLS assumptions, $Y \sim N(W\beta, \sigma^2 I_T)$, and therefore:

$$L(\alpha, \beta, \sigma^2) = f(Y_i)\dots f(Y_T) = \frac{1}{(2\pi\sigma^2)^{T/2}} exp\left\{-\sum_{i=1}^{T}\frac{(Y_i - \alpha - \beta X_i)^2}{2\sigma^2}\right\}$$

Or in logarithmic form:

$$\log L(\alpha, \beta, \sigma^2) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{T}(Y_i - \alpha - \beta X_i)^2$$

From the **first-order conditions**:

$$\partial \log L / \partial \alpha = \frac{1}{\sigma^2}\sum_{i=1}^{T}(Y_i - \alpha - \beta X_i) = 0$$

$$\partial \log L / \partial \beta = \frac{1}{\sigma^2}\sum_{i=1}^{T} X_i(Y_i - \alpha - \beta X_i) = 0$$

$$\partial \log L / \partial \sigma^2 = -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{T}(Y_i - \alpha - \beta X_i)^2 = 0$$

We obtain the MLEs estimators as:

$$\hat{\alpha}_{ML} = \bar{Y} - \bar{X}\hat{\beta}_{ML}; \qquad \hat{\beta}_{ML} = \frac{\sum_{i=1}^{T}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{T}(X_i - \bar{X})^2}; \qquad \hat{\sigma}^2_{ML} = \frac{\sum_{i=1}^{T}\hat{e}_i^2}{T}$$

Maximum Likelihood Estimators: Properties
The MLEs of α and β are the **same as the OLS estimators**. However, $\hat{\sigma}^2_{ML}$ **is slightly biased**, due to its denominator being T as opposed to T – 2.
The MLEs estimators in general are:

- **Consistent**, they converge in probability to the true values of the parameters under correct model specification.
- **Efficient**, with asymptotic variance-covariance matrix which for $\theta = \alpha, \beta$ is equal to:

$$E\left[\frac{\partial^2 logL}{\partial\theta\partial\theta'}\right]^{-1} = \sigma^2(W'W)^{-1}$$

- **Asymptotically Normally Distributed** when multiplied by $\sqrt{T}$.

The Coefficient of Determination R^2

The coefficient of determination provides a measure of the **goodness of fit** of the explanatory model: $y = W\hat{\beta} + \hat{e} = \hat{y} + \hat{e}$.

As $\hat{y}'\hat{e} = 0$, multiplying by y' and subtracting $T\bar{y}^2$ from both sides we can rewrite the previous expression as:

$$y'y - T\bar{y}^2 = \sum_{t=1}^{T}(y_t - \bar{y})^2 = \hat{y}'\hat{y} - T\bar{y}^2 + \hat{e}'\hat{e}$$

It can be shown that, as there is an intercept x_1 in the model, the sample mean of y is the same as that of $\hat{y}$: $\bar{y} = \bar{\hat{y}}$.

In fact, since $x_1' = (1, \dots, 1)$, it is: $\bar{\hat{y}} = \frac{\sum_{t=1}^{T}\hat{y}_t}{T} = \frac{x'_1\hat{y}}{T} = \frac{x'_1 W}{T}\hat{\beta} = \frac{x'_1(W\hat{\beta}+\hat{e})}{T} = \frac{x'_1 y}{T} = \bar{y}$ as $x_1'\hat{e} = 0$

Therefore:

$$\sum_{t=1}^{T}(y_t - \bar{y})^2 = \sum_{t=1}^{T}(\hat{y}_t - \bar{\hat{y}})^2 + \sum_{t=1}^{T}(\hat{e}_t - 0)^2$$

Where:

- $\sum_{t=1}^{T}(y_t - \bar{y})^2$ is the **Total Sum of Squares (TSS)**, measuring the variability for y.
- $\sum_{t=1}^{T}(\hat{y}_t - \bar{\hat{y}})^2$ is the **Explained Sum of Squares (ESS)**, measuring the variability of the model fit.
- $\sum_{t=1}^{T}(\hat{e}_t)^2$ is the **Residual Sum of Squares (RSS)**, measuring the variability of the residuals.

The **coefficient of determination** $0 \leq R^2 \leq 1$ represents the percentage of variability of y explained by the model:

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

R^2 is sensible only when using OLS estimators for the model parameters, otherwise $\tilde{y}'\tilde{e} \neq 0$.

$0 \leq R^2 \leq 1$ only when the model has an intercept, otherwise $\bar{y} \neq \bar{\hat{y}}$ and it can be $R^2 < 0$.

R^2 always increases if explanatory variables are added to the model, even if they are irrelevant. To avoid this issue, the **adjusted coefficient of determination** can be used. With k regressors:

$$R^2 = 1 - \frac{\frac{RSS}{T-k}}{\frac{TSS}{T-1}}$$

As an alternative measure of the explanatory power of a model, we could compute the **correlation between the dependent variable y and the model fit $\hat{y}$**, which should be close to 1 for a very accurate model.

However, it can be shown that this measure provides exactly the same information as the coefficient of determination. Indeed:

$$corr^2(y,\hat{y}) = \frac{cov^2(y,\hat{y})}{var(y)var(\hat{y})} = \frac{cov^2(\hat{y}+\hat{e},\hat{y})}{var(y)var(\hat{y})} = \frac{var^2(\hat{y})}{var(y)var(\hat{y})} = \frac{var(\hat{y})}{var(y)} = \frac{\sum_{t=1}^{T}(\hat{y}_t - \bar{y})^2}{\sum_{t=1}^{T}(y_t - \bar{y})^2} = \frac{ESS}{TSS} = R^2$$

Linear Transformations of Variables

In economics it is common to work with **transformed** (usually re-scaled) variables.

If the **independent** variable is rescaled: $y = W\beta + e \quad \Rightarrow \quad y = \frac{W}{a}\gamma + u = \omega\gamma + u$

The OLS estimator of this transformed model is then equal to the one we obtained in the original model multiplied by the a constant:

$$\hat{\gamma} = (\omega'\omega)^{-1}\omega' y = a(W'W)^{-1}W' y = a\hat{\beta}$$

Its variance is instead: $var(\hat{\gamma}) = \sigma^2(\omega'\omega)^{-1} = a^2\sigma^2(W'W)^{-1} = a^2 var(\hat{\beta})$

Finally, the residuals are not affected by the linear variable transformation: $\hat{\sigma}_u^2 = \hat{\sigma}_e^2$ and the coefficient of determination R^2 is also unchanged

$$\hat{u} = y - \omega\hat{\gamma} = y - \frac{W}{a}a\hat{\beta} = y - W\hat{\beta} = \hat{e}$$

If the **dependent** variable is rescaled: $\frac{y}{a} = W\gamma + u$

The OLS estimator becomes the same as in the original model but divided by a: $\hat{\gamma} = (W'W)^{-1}W' y/a = \frac{1}{a}\hat{\beta}$

Its variance is instead: $var(\hat{\gamma}) = \frac{1}{a^2}var(\hat{\beta})$

Now, the error term is also re-scaled by a^2, but it does not affect R^2 as the variance of y decreases by the same amount

$$\hat{u} = \frac{y}{a} - \frac{x}{a}\hat{\beta} = \frac{\hat{e}}{a} \qquad \hat{\sigma}_u^2 = \hat{\sigma}_e^2 / a^2$$

The Multiple Regression Model

Generally, more than one explanatory variable are needed in economics, whereby the **multiple regression model**:

$$y_t = \beta_1 + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + e_t$$

Or, equivalently, in **vector form**:

$$\underset{T\times 1}{y} = \underset{T\times k}{X}\ \underset{k\times 1}{\beta} + \underset{T\times 1}{e}.$$

The usual OLS assumptions apply, with the exception of assumption (5):

- Linearity and Stability of Parameters
- Deterministic Regressors
- Homoscedasticity
- No Correlation
- **The Regressors are not Linearly Dependent**, i.e. (X'X) has rank k.

The β_j coefficients now represent the change in the expected value of the dependent variable for a marginal variation of regressor j, **assuming that all other explanatory variables remain constant**:

$$\beta_j = \partial E(y_t) / \partial x_j \,|_{x_1,\ldots x_{j-1},x_{j+1},\ldots x_k}$$

The **OLS estimators** are still determined as the minimizers of the sum of squared errors, so we get:

$$\underset{k\times 1}{\hat{\beta}} = \underset{k\times k}{(X'X)^{-1}}\ \underset{k\times 1}{X'y}, \qquad var(\hat{\beta}) = \sigma^2(X'X)^{-1}$$

Where:

$$X'X = \begin{bmatrix} \sum_{t=1}^{T} x_{1t}^2 & \sum_{t=1}^{T} x_{1t}x_{2t} & \cdots & \sum_{t=1}^{T} x_{1t}x_{kt} \\ \sum_{t=1}^{T} x_{1t}x_{2t} & \sum_{t=1}^{T} x_{2t}^2 & \cdots & \sum_{t=1}^{T} x_{2t}x_{kt} \\ \vdots & & \ddots & \\ \sum_{t=1}^{T} x_{1t}x_{kt} & \sum_{t=1}^{T} x_{2t}x_{kt} & \cdots & \sum_{t=1}^{T} x_{kt}^2 \end{bmatrix} \qquad X'y = \begin{bmatrix} \sum_{t=1}^{T} x_{1t}y_t \\ \sum_{t=1}^{T} x_{2t}y_t \\ \vdots \\ \sum_{t=1}^{T} x_{kt}y_t \end{bmatrix}$$

If we introduce the additional **normality assumption**, $\hat{\beta}$ is normally distributed also in finite samples:

$$\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1}$$

An unbiased and consistent **estimator of the variance** of the error term is then given by:

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^{T} \hat{e}_t^2}{T-k}$$

The estimators $\hat{\beta}$ are still **uncorrelated** with $\hat{\sigma}^2$, or independent if the hypothesis of Normality of the errors holds.
As there are no substantial differences from the case with one explanatory variable, the **Gauss-Markov** theorem still holds, and $\hat{\beta}$ is **B.L.U.E.** and **consistent**.
The overall model explanatory power can still be measured by the $\boldsymbol{R^2}$ or its adjusted version.

The Multiple Regression Model – Alternative Interpretation
We will now consider an **alternative interpretation** of the model coefficients.
For simplicity, consider a model with two explanatory variables:

$$y = x_1\beta_1 + x_2\beta_2 + e$$

And **premultiply** both sides by the **matrix** $[I - x_2(x'_2x_2)^{-1}x'_2]$:

$$[I - x_2(x'_2x_2)^{-1}x'_2]y$$
$$= [I - x_2(x'_2x_2)^{-1}x'_2]x_1\beta_1 + \underbrace{[I - x_2(x'_2x_2)^{-1}x'_2]x_2\beta_2}_{=0} + \underbrace{[I - x_2(x'_2x_2)^{-1}x'_2]e}_{=u}$$

Therefore:

$$[y - x_2(x'_2x_2)^{-1}x'_2y] = [x_1 - x_2(x'_2x_2)^{-1}x'_2x_1]\,\beta_1 + u$$

We can now consider the dependent and independent variables in this modified model. If we regress y on x_2, we get:

$$y = x_2\gamma + e_1, \qquad \hat{\gamma} = (x'_2x_2)^{-1}x'_2y$$

And also:

$$\hat{e}_1 = y - x_2\hat{\gamma} = (y - x_2(x'_2x_2)^{-1}x'_2y)$$

Which corresponds to the **dependent variable** in the modified model.
If we instead regress x_1 on x_2:

$$x_1 = x_2\eta + e_2, \qquad \hat{\eta} = (x'_2x_2)^{-1}x'_2x_1$$

And also:

$$\hat{e}_2 = x_1 - x_2\hat{\eta} = x_1 - x_2((x'_2x_2)^{-1}x'_2x_1)$$

Which corresponds to the **independent variable** in the modified model.
To summarize:

- The **dependent** variable in the modified model is the **residual of a regression of $\boldsymbol{y}$ on $\boldsymbol{x_2}$**.
- The **independent** variable in the modified model is the **residual of a regression of $\boldsymbol{x_1}$ on $\boldsymbol{x_2}$**.
- We can then rewrite the model as: $\hat{e}_1 = \hat{e}_2\beta_1 + u$

Where β_1 measures the ability of the part of x_1 unexplained by x_2 to explain the part of y unexplained by x_2, i.e., the **contribution of** x_1 in explaining y net of effects of the presence of x_2 in the model. In formula:

$$\beta_1 = {}^{\partial E(\hat{e}_1)}/_{\partial \hat{e}_2}$$

- More in general, i.e., in a model with k independent variables, β_j measures the explanatory capacity of the part of x_j unexplained by $x_1, \ldots, x_{j-1}, x_j, \ldots, x_k$ for the part of y unexplained by $x_1, \ldots, x_{j-1}, x_j, \ldots, x_k$.

Multicollinearity
Let us consider the model with two explanatory variables:

$$y = x_1\beta_1 + x_2\beta_2 + \varepsilon$$

If $x_1 = x_2$, i.e., the explanatory variables are **perfect collinear**, we can no longer separately identify β_1 and β_2, as:

$$y = (\beta_1 + \beta_2)x_2 + \varepsilon$$

Similarly, for the case with k regressors, multicollinearity results in the matrix $X'X$ having reduced rank (< k), so that the system $(X'X)\hat{\beta} = X'y$ admits an infinite number of solutions.
In practice, collinearity is usually high but not perfect, so that the coefficients of the OLS model can still be identified, but their precision is much lower, indeed:

$$|X'X| \sim 0 \Rightarrow \sigma^2(X'X)^{-1} \sim \infty$$

Methods to **detect** collinearity are:
- Computing the correlations between all possible pairs of regressors.
- Computing the t and F tests for the non-significance of the regressors. If they give contrasting results, there may be problems of collinearity.
- Computing the determinant of the $(X'X)$ matrix.

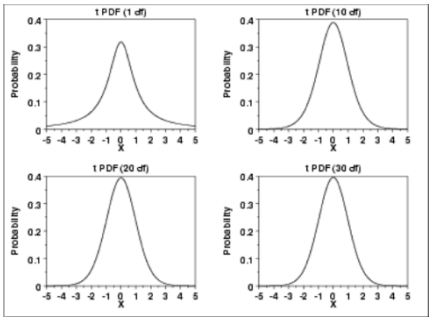
Possible **solutions** are:
- Model re-parametrization.
- Replacement of the regressors with their first principal components.
- Increasing the sample size.

Inference in the Linear Regression Model

Interval Estimators
In the previous chapter we have defined (point) estimators for the parameters of the linear regression model. We now want to derive **interval estimators** (or **confidence intervals**) for β, i.e., define a range containing β, the true value of the parameters, with a fixed probability.

The Student – t Distribution
Consider two independent random variables:

$$Z = \frac{q - \mu_q}{\sigma_q} \sim N(0,1), \qquad V \sim \chi^2(p)$$



It can be shown that:

$$t = \frac{z}{\sqrt{V/p}} \sim t(p)$$

Where p are the **degrees of freedom (df)**, that completely characterize the t-distribution, as

$$E(t) = 0, \qquad Var(t) = \frac{p}{p-2}$$

And t → N(0,1) when p → ∞.
Otherwise, the t-density has more probability in the tails than the Normal density.

The t-Distribution in the context of the Linear Regression Model

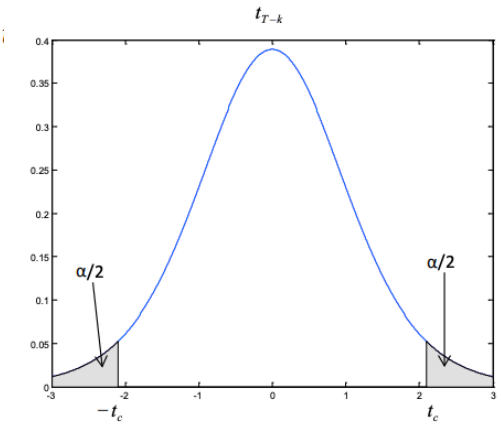
We know that, when the errors are normally distributed $z = \frac{\hat{\beta}_i - \beta_i}{\sqrt{Var(\hat{\beta}_i)}} \sim N(0,1)$

Moreover, z is independent of $V = \frac{(T-k)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(T-k)$, where T is the sample size and k the number of regressors. It follows that $\frac{z}{\sqrt{V/T-k}} \sim t(T-k)$

In the simple model with one explanatory variable (and the intercept)

$$\frac{z}{\sqrt{V/T-2}} = \frac{\hat{\beta}_2 - \beta_2}{\sqrt{\frac{\sigma^2}{\sum(x_t - \bar{x})^2}}} \Big/ \sqrt{\frac{(T-2)\hat{\sigma}^2}{\sigma^2}/T-2} = \frac{\hat{\beta}_2 - \beta_2}{\sqrt{\frac{\hat{\sigma}^2}{\sum(x_t - \bar{x})^2}}} = \frac{\hat{\beta}_2 - \beta_2}{\sqrt{V\hat{a}r(\hat{\beta}_2)}} = \frac{\hat{\beta}_2 - \beta_2}{s.\hat{e}.(\hat{b}_2)} \sim t(T-2)$$

In the general model with k regressors we have, for q = 1,2, …, k.



If we fix a probability level α (= 0.1, or 0.05, or 0.01), we can find the **critical value** t_c such that:

- $\Pr\{t \geq t_c\} = \frac{\alpha}{2}$
- $\Pr\{t \leq -t_c\} = \frac{\alpha}{2}$

The Student-t distribution is symmetric, so that the left critical value is equal in absolute value to the right critical value).
Combining we have: $\Pr\{-t_c \leq t \leq t_c\} = 1 - \alpha$.

Given α and the number of degrees of freedom $T-k$, we can obtain the critical value t_c.
We can now replace t with its definition in terms of $\hat{\beta}_q$, which yields: $\Pr\left\{-t_c \leq \frac{\hat{\beta}_q - \beta_q}{\widehat{se}(\hat{\beta}_q)} \leq t_c\right\} = 1 - \alpha$

From $\Pr\{\hat{\beta}_q - t_c \cdot \widehat{se}(\hat{\beta}_q) \leq \beta_q \leq \hat{\beta}_q + t_c \cdot \widehat{se}(\hat{\beta}_q)\} = 1 - \alpha$, it follows that $\hat{\beta}_q \pm t_c \cdot \widehat{se}(\hat{\beta}_q)$

is an interval estimator or confidence interval for the parameter β_q at the $(1-\alpha)\%$ level, while $\hat{\beta}_q$ is a point estimator.
In the simple case of the model with intercept and one explanatory variable (k = 2),

$$\hat{\beta}_q \pm t_c \cdot \widehat{se}(\hat{\beta}_q) = \frac{T\sum x_t y_t - \sum x_t \sum y_t}{T\sum x_t^2 - (\sum x_t)^2} \pm t_c \cdot \sqrt{\frac{\hat{\sigma}^2}{\sum(x_t - \bar{x})^2}}$$

Model with Intercept and One Explanatory Variable (K = 2)

- **Example:** $\hat{\beta}_2 = 2.5,\ \widehat{se}(\hat{\beta}_2) = 1,\ T = 32$

  then $p = T - 2 = 30$, and $t_c = 1.697$.

  Therefore, the confidence interval at the 90% level for β_2 is

  $$\hat{\beta}_2 \pm 1.697 \cdot s.\hat{e}(\hat{\beta}_2) = 2.5 \pm 1.697 \cong [0.8, 4.2]$$

Note that the table shows the probability to the right of the critical value. Hence, if the chosen significance level is $\alpha = 0.1$, the relevant column in the table is that for $\frac{\alpha}{2} = 0.05$ *(table in the next page).*

In practice, the length of the confidence interval gives an idea of the reliability of the estimated parameter. If, in fact, $\widehat{se}(\hat{\beta}_2)$ is large, meaning there is a lot of uncertainty, then the confidence interval is large, and vice versa if $\widehat{se}(\hat{\beta}_2)$ is small.

	$P(t \geq t_c)$					
df	0.10	0.05	0.025	0.01	0.005	0.001
1.	3.078	6.314	12.706	31.821	63.657	318.313
2.	1.886	2.920	4.303	6.965	9.925	22.327
3.	1.638	2.353	3.182	4.541	5.841	10.215
4.	1.533	2.132	2.776	3.747	4.604	7.173
5.	1.476	2.015	2.571	3.365	4.032	5.893
6.	1.440	1.943	2.447	3.143	3.707	5.208
7.	1.415	1.895	2.365	2.998	3.499	4.782
8.	1.397	1.860	2.306	2.896	3.355	4.499
9.	1.383	1.833	2.262	2.821	3.250	4.296
10.	1.372	1.812	2.228	2.764	3.169	4.143
11.	1.363	1.796	2.201	2.718	3.106	4.024
12.	1.356	1.782	2.179	2.681	3.055	3.929
13.	1.350	1.771	2.160	2.650	3.012	3.852
14.	1.345	1.761	2.145	2.624	2.977	3.787
15.	1.341	1.753	2.131	2.602	2.947	3.733
16.	1.337	1.746	2.120	2.583	2.921	3.686
17.	1.333	1.740	2.110	2.567	2.898	3.646
18.	1.330	1.734	2.101	2.552	2.878	3.610
19.	1.328	1.729	2.093	2.539	2.861	3.579
20.	1.325	1.725	2.086	2.528	2.845	3.552
21.	1.323	1.721	2.080	2.518	2.831	3.527
22.	1.321	1.717	2.074	2.508	2.819	3.505
23.	1.319	1.714	2.069	2.500	2.807	3.485
24.	1.318	1.711	2.064	2.492	2.797	3.467
25.	1.316	1.708	2.060	2.485	2.787	3.450
26.	1.315	1.706	2.056	2.479	2.779	3.435
27.	1.314	1.703	2.052	2.473	2.771	3.421
28.	1.313	1.701	2.048	2.467	2.763	3.408
29.	1.311	1.699	2.045	2.462	2.756	3.396
30.	1.310	1.697	2.042	2.457	2.750	3.385
40.	1.303	1.684	2.021	2.423	2.704	3.307
50.	1.299	1.676	2.009	2.403	2.678	3.261
60.	1.296	1.671	2.000	2.390	2.660	3.232
70.	1.294	1.667	1.994	2.381	2.648	3.211
80.	1.292	1.664	1.990	2.374	2.639	3.195
90.	1.291	1.662	1.987	2.368	2.632	3.183
100.	1.290	1.660	1.984	2.364	2.626	3.174
∞ .	1.282	1.645	1.960	2.326	2.576	3.090

Testing Hypotheses on the Parameters of the Linear Regression Model

An econometric model is based on economic theory. For example, basic consumption theory suggests that $C = c_0 + c_1 Y$, with $c_0 > 0$ and $0 < c_1 < 1$, where C indicates consumption and Y disposable income.

In order to test whether a theory is correct, or at least compatible with the data, we can verify whether its implications for the econometric model parameters are valid. For example, it should be $0 < c_1 < 1$.

We must therefore consider how to statistically test this kind of hypotheses on the model parameters. To construct a **hypothesis test** we need 4 components:

- The **null hypothesis**: what we want to test (and assume to be valid);
- The **alternative hypothesis**: what (we assume) is true if the null hypothesis is false;
- A **test statistic**: a random variable which allows us to discriminate between (1) and (2);
- A **rejection region**: a set of values such that if the test statistic in 3 takes one of these values we can reject (1).

Example, regression model of aggregate consumption on disposable income: $C_t = \beta_1 + \beta_2 Y_t + \varepsilon_t$

- The null hypothesis specifies a value for a parameter, for example, $H_0 : \beta_2 = 0.9$, or more generally $H_0 : \beta_2 = c$. This is what we believe is the true value of the parameter β_2 and we want to test whether we can reject our guess or not.
- The alternative hypothesis tells us what we expect to be true about β_2 if our null hypothesis is incorrect. For example, it can be $H_1 : \beta_2 \neq c$, or $H_1 : \beta_2 < c$.

Remarks:

- Though testing H_0 against H_1 it turns out that H_0 is rejected, we do not necessarily accept H_1. The testing procedure is meant to reject or not H_0, not to accept or reject H_1. If H_1 is of particular interest, such as $H_1 : \beta_2 < c$, then we should repeat the procedure by putting $\beta_2 < c$ as the null.
- Typically, in the statistical literature, the null hypothesis is the most important one, such as a new drug has no side effects (H_0), against the same drug has side effects (H_1). IN econometrics things are a little reversed, e.g., income (Y) should affect consumption (C), but often the null hypothesis is formulated as $H_0 : \beta_2 = c$ (Y has no effect on C), versus $H_0 : \beta_2 > 0$ (Y has a positive effect on C). This is because often we are not able to formulate a precise null hypothesis (e.g. $\beta_2 = 0.77$) and therefore we are satisfied with less, namely, if we can at least reject that an explanatory variable has no effects on the dependent variable ($H_0 : \beta_2 = 0$).
- We should keep in mind that hypothesis testing is built assuming that H_0 is the key hypothesis of interest and hence it is more problematic to reject H_0 by mistake than not to reject it by mistake (the procedure is conservative in this sense, it favors H_0).

The **Test Statistic** is a random variable that must be informative on the null hypothesis of interest. **Example**, If $H_0 : \beta_2 = c$ against $H_1 : \beta_2 \neq c$, then we need something that summarizes the information in the sample on β_2. For this we can use $\hat{\beta}_2$, the OLS estimator of β_2.

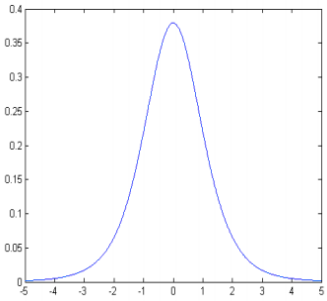
- Since we know that under the assumptions of the linear regression model, $\hat{\beta}_2$ is a good estimator of β_2, we could use $\hat{\beta}_2 - c$ to estimate the distance between β_2 and c. If $\hat{\beta}_2 - c$ is large, probably $\beta_2 - c$ is also large, then we can reject H_0.
- But we must consider that $\hat{\beta}_2$ is an estimator and has a variance, so that $\hat{\beta}_2 - c$ could be large just by randomness. To take this feature into account, we could use $\frac{\hat{\beta}_2 - c}{se(\hat{\beta}_2)}$ as a statistic.
    - If the parameter estimator is accurate $se(\hat{\beta}_2)$ is small and therefore even a small difference between $\hat{\beta}_2$ and c is magnified.
    - The opposite is true if $se(\hat{\beta}_2)$ is great, i.e., if there is a lot of uncertainty.

- A second important characteristic of a **test statistic** is that its distribution should be fully known when the null hypothesis is true. In this regard, we know that $t = \frac{\hat{\beta}_2 - \beta_2}{\widehat{se}(\hat{\beta}_2)} \sim t(T-2)$.
- Therefore, if $H_0 : \beta_2 = c$ is true, it is: $t = \frac{\hat{\beta}_2 - c}{\widehat{se}(\hat{\beta}_2)} \sim t(T-2)$.
- Note that if H_0 is false, t doe not have the $t(T-2)$ distribution. In fact, if $\beta_2 \neq c$, then

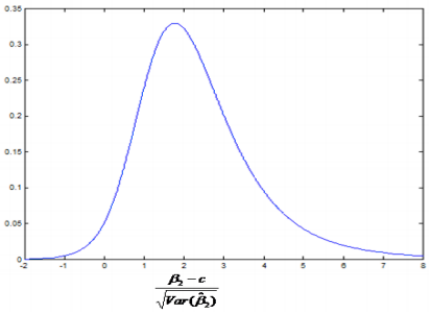
from $t = \frac{\hat{\beta}_2 - \beta_2}{\sqrt{Var(\hat{\beta}_2)}} \sim N(0,1)$ it follows that $\frac{\hat{\beta}_2 - c}{\sqrt{Var(\hat{\beta}_2)}} + \frac{c - \beta_2}{\sqrt{Var(\hat{\beta}_2)}} \sim N(0,1)$

and then $\frac{\hat{\beta}_2 - c}{\sqrt{Var(\hat{\beta}_2)}} \sim N\left(\frac{\beta_2 - c}{\sqrt{Var(\hat{\beta}_2)}}, 1\right)$.

- If H_0 is true, the density of the t-statistic is similar to:



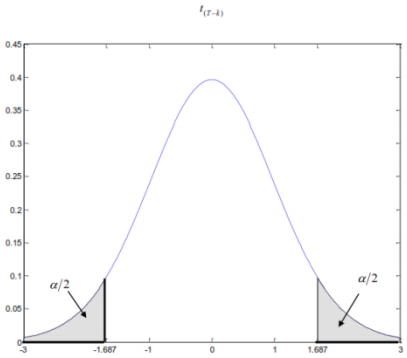
- If H_0 is false then the density of the t-statistic is centered not on 0 but on $\frac{\beta_2 - c}{\sqrt{Var(\hat{\beta}_2)}}$.



In other words, if H_0 is true, it is likely that t is close to 0. If H_0 is false it is likely that t is positive (if $\beta_2 > c$), or negative (if $\beta_2 < c$). The idea is not to reject H_0 if the calculated value for the t-statistic is close to 0, and to reject it otherwise.

To formalize the meaning of "close to 0" we introduce the notion of **rejection region**.
The rejection region is defined as the set of values such that, if the test statistic is one of these values, then H_0 is rejected. In the case of the t-test, the rejection region is represented by values larger and smaller than the critical value t_c.
To determine the critical value t_c we must define the **significance level** of the test, α (typically = 0.05, or 0.01, 0.10). Then, we choose t_c such that $\Pr\{t < -t_c\} = \Pr\{t > t_c\} = \frac{\alpha}{2}$.
If $t < -t_c$ or $t > t_c$ then we reject the null hypothesis. The rejection region is then $(-\infty, -t_c) \cup (t_c, \infty)$. The critical value t_c is again determined using tables, for given values of α and $T - k$.

Graphical Example:
Rejection region of the t-test, for $\alpha = 0.1$ & $T - k = 37$.



- What happens if $-t_c < t < t_c$? We do not reject H_0.
- Can we then say that we accept H_0, namely, that H_0 is true? No, strictly speaking.
- For example, if $H_0: \beta_2 = 0.9$, but the true value of the parameter is $\beta_2 = 0.88$ or $\beta_2 = 0.92$, then in practice we never reject H_0 unless the number of available observations, T, is very large.

Example of Hypothesis Testing
Suppose that we want to test $H_0: \beta_2 = 1$ against $H_0: \beta_2 \neq 1$ and, we have estimated $\hat{\beta}_2 = 2.5$, with $se(\hat{\beta}_2) = 1, T = 39, T - k = 37$ *and* $\alpha = 0.1$.

The value of the t-statistic is $t = \frac{\hat{\beta}_2 - 1}{1} = 1.5$ while $t_c = 1.687$.

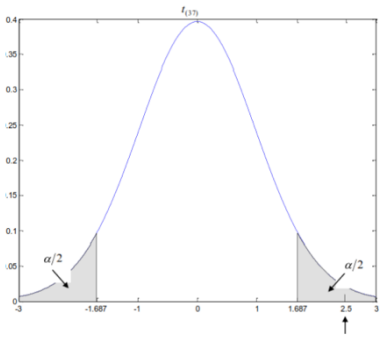
Then the null hypothesis $H_0: \beta_2 = 1$ is not rejected since the t-test does not fall in the rejection region $(-\infty, -1.687) \cup (1.687, \infty)$.

If the null hypothesis is $H_0: \beta_2 = 0$, then $t = \frac{\hat{\beta}_2 - 0}{1} = 2.5$ and therefore H_0 is rejected since $t > t_c$.

Type I and Type II Errors

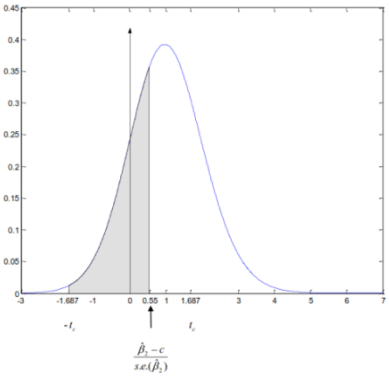
	Test does not reject	Test rejects
H_0 true	right conclusion	type I error
H_0 false	type II error	right conclusion

Type I Error – Reject H_0 when True
Suppose the test value falls in the rejection region (grey area) when H_0 is true.
As a consequence, the probability of making a type I error is equal to α, the significance level of the test.
Remember that we set α, and the fact that it usually takes low values such as 0.10, 0.05 or 0.01 reflects the choice of having a low probability of committing a type I error.



Type II Error – Do Not Reject H_0 when False
Suppose the test value falls in the acceptance region when H_0 is false.
The probability of making a type II error cannot be controlled. It depends **inversely** from:

- **The significance level of the test**: when α is reduced, the acceptance region widens, and the probability of a type II error increases. The interval $[-t_c, t_c]$ widens, therefore also the grey area, representing the probability of type II error, increases.



- **The degree of the violation of the null hypothesis**: the more different β_2 is from c, the more different the distribution of the test statistic under H_1 is from that under H_0, and, therefore, the probability of type II error is reduced. If β_2 is very close to c, the distribution of the test statistic is centered at a point close to 0, thus the grey area increases.

- **The sample size**: when T increases the estimator becomes more accurate, $\widehat{se}(\hat{\beta}_2)$ is reduced, and therefore the (absolute value of the) statistic increases. In the limiting case where $T \to \infty$, consistency of the OLS estimator implies $\hat{\beta}_2 \to \beta_2$ and $\widehat{se}(\hat{\beta}_2) \to 0$. Hence, if $\beta_2 \neq c$, it is $\frac{\hat{\beta}_2 - c}{s.\hat{e}.(\hat{\beta}_2)} \to \pm\infty$. Therefore, the null hypothesis is properly rejected, and the probability of type II error goes to 0 (so the test is said to be **consistent**). When T increases, the distribution of the test statistic becomes more and more concentrated, and centered on a value that tends to move rightward of t_c (or leftward of $-t_c$). Then the grey area shrinks.

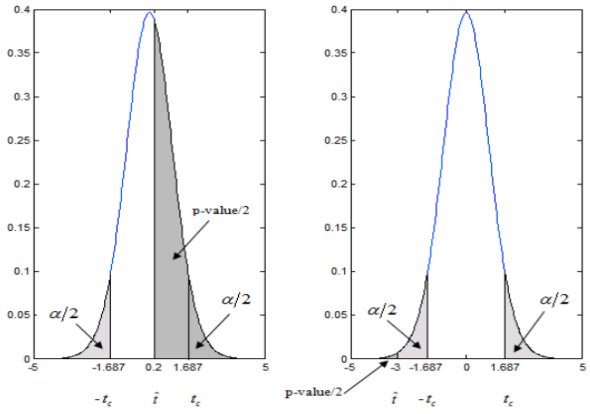
- It can be shown that the t-test is the one that minimizes the probability of type II error for a given probability of type I error, and in this sense it is optimal.
- **Intuition**: the test is based on $\hat{\beta}_{OLS}$, which is BLUE.

The P-Value

Consider the calculated value of the statistic, $\hat{t}$, and calculate the probability:

$$\Pr\{t \geq \hat{t}\} + \Pr\{t \leq -\hat{t}\}$$

This probability is called **p-value** (probability value) of the test.



- If p > α we do not reject H_0.
- If p < α we reject H_0.

The p-value is greater than the significance level of the test if and only if $\hat{t}$ falls in the acceptance region.

Significance Testing

A very common null hypothesis in empirical applications is $\beta_2 = 0$ against the alternative $\beta_2 \neq 0$, i.e., under H_0 the explanatory variable X is not significant (for this reason this is called **significance testing**). In this case, the expression of the t-test simplifies to:

$$t = \frac{\hat{\beta}_2 - \beta_2}{\widehat{se}(\hat{\beta}_2)} = \frac{\hat{\beta}_2}{\widehat{se}(\hat{\beta}_2)}$$

The statistical significance of a coefficient does not necessarily mean that it is also economically relevant, and viceversa …

Example, we want to estimate Phillips curve to evaluate the existence of a trade-off between inflation and unemployment, having a sample of T = 100 observations. In this case, the dependent variable is the rate of inflation and the explanatory variable is the rate of unemployment.

- If the estimated coefficient is $\hat{\beta} = -0.001$, with $\widehat{se} = 0.0001$, then the null hypothesis of no significance of the unemployment rate is rejected for most significance levels, since the calculated value of the statistic is equal to -10. From an economic point of view, however, the unemployment rate is pretty much irrelevant to explain inflation, given the estimated value of the parameter.

- If, instead, $\hat{\beta} = -0.5$, with $\widehat{se} = 0.5$, then the null hypothesis of no significance of the unemployment rate would not be statistically rejected since the t-test is -1, while the estimated value of the parameter would imply a significant economic importance of unemployment.

The example also emphasizes the importance of the choice of the null hypothesis that we want to test.

- If we set $\beta_2 = -1$ instead of $\beta_2 = 0$ as the null hypothesis, implying a 1-to-1 trade-off between inflation and unemployment, the value of the t-test is 1 and the null hypothesis is not rejected. On the other hand, even a null hypothesis such as $\beta_2 = 0.4$, a positive relationship between inflation and unemployment as in stagflation periods, would not be rejected at a 5% significance level, being the t-test equal to -1.8 and the critical value -1.984.
- If instead we set the significance level of the test at 10%, then the null hypothesis $\beta_2 = 0.4$ is rejected, because the critical value is equal to -1.660. Therefore, the type I error probability we are willing to accept is also an important choice from an economic point of view, because it can determine whether the hypothesis of interest is rejected or not.

The Relationship Between Confidence Intervals and Hypothesis Testing

Recall the expression for the confidence interval for β at the $(1 - \alpha)$% level:

$$\hat{\beta} - t_c \times \widehat{se}(\hat{\beta}) \leq \beta \leq \hat{\beta} + t_c \times \widehat{se}(\hat{\beta})$$

With

$$Pr\{-t_c \leq \frac{\hat{\beta} - \beta}{\widehat{se}(\hat{\beta})} \leq t_c\} = 1 - \alpha$$

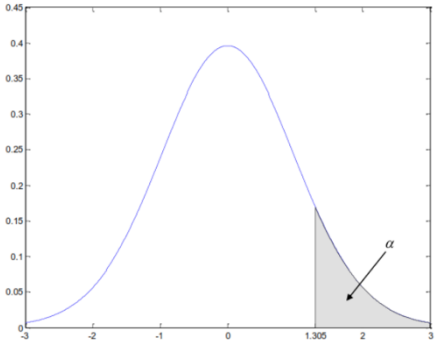
This expression can be used to derive the acceptance (more properly, non-rejection) region of a test. In fact, if we replace β with c, then to calculate the acceptance region for $H_0: \beta = c$ we use:

$$Pr\{-t_c \leq \frac{\hat{\beta} - c}{\widehat{se}(\hat{\beta})} \leq t_c\} = 1 - \alpha$$

As a result, if $\hat{t}$ falls in the acceptance region for $H_0: \beta = c$ at the α% level, and therefore H_0 is not rejected, then $\hat{t}$ is part of a (1 - α)% confidence interval for β.
Conversely, if c is part of a (1 - α)% confidence interval for β, then the null hypothesis $H_0: \beta = c$ is not rejected at a significance level α.

Example, $\hat{\beta}_2 = 2.5,\ se(\hat{\beta}_2) = 1,\ T = 32,\ T - k = 30,\ and\ \alpha = 0.1$

- Then a 90% confidence interval for $\hat{\beta}_2$ is [0.8, 4.2].
- Thus, the null hypothesis $H_0: \beta = 0$ would be rejected at a 10% significance level, while $H_0: \beta = 1$ would not be rejected.



One-Way Tests

So far, we have always considered alternative hypotheses of the type $\beta \neq c$.
We now extend the testing procedure to allow for a different formulation of the alternative hypothesis, maintaining the same null hypothesis:

$$H_0: \beta = c \text{ and } H_1: \beta > c$$

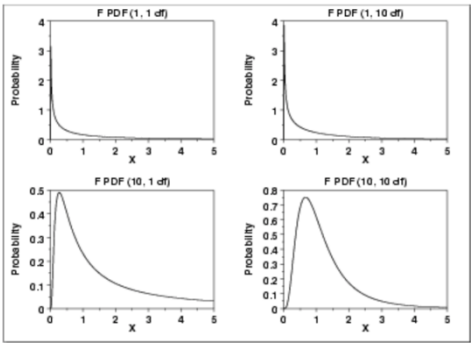
The only difference with respect to the testing procedure we have used so far is in the choice of the rejection region, which is now in the right tail of the distribution of the statistics (left if $H_1: \beta < c$). In fact, in this case, if H_0 is false we expect very large positive values for the t-statistic ($\beta > c$).

Example $H_0: \beta_2 = 1, H_1: \beta_2 > 1, \hat{\beta}_2 = 2.5,\ se(\hat{\beta}_2) = 1,\ T = 39,\ T - k = 37,\ and\ \alpha = 0.1$

- Then $t = (\hat{\beta}_2 - 1)/1 = 1.5$ and $t_c = 1.310$, therefore H_0 is rejected.
- Note that if instead it is $H_1: \beta_2 \neq 1$ then H_0 is not rejected.

The F-Distribution

So far, we have learned to test hypotheses on individual parameters of the linear regression model, for example, $\beta_k = 0$ or $\beta_k = 1$. However, there are hypotheses of interest that involve multiple parameters, such as $\beta_1 = 0$ and $\beta_2 = 1$. These are often called **composite hypotheses**.



To test composite hypotheses, it is useful to introduce the **F-Distribution**.

Given two random variables: $v_1 \sim \chi^2(p), v_2 \sim \chi^2(q)$ with v_1 and v_2 independent, we have: $\frac{v_1/p}{v_2/q} \sim F(p, q)$

The F-Test

Let us suppose that we want to test the hypotheses:

$$H_0: \beta_1 = c_1, \beta_2 = c_2, \ldots, \beta_k = c_k$$
$$H_1: \beta_1 \neq c_1, \beta_2 \neq c_2, \ldots, \beta_k \neq c_k$$

Or, using the vector notation,

$$H_0: \beta_{kx1} = c_{kx1}$$
$$H_1: \beta_{kx1} \neq c_{kx1}$$

In the model

$$y_{Tx1} = X_{Txk}\, \beta_{kx1} + \varepsilon$$

Note, the alternative hypothesis states that **at least one** of the components of H_0 is false. In addition, the alternative hypothesis is formulated in terms of "different from", it cannot be formulated in terms of "greater than" or "less than", as we will see shortly.

Again, the idea is to use the OLS estimator of the parameters, $\hat{\beta}$, and assess if they are "very different" from the values assumed under the null hypothesis, grouped in the vector c. Consider the variable:

$$v_1: \underset{1\times k}{(\hat{\beta} - c)'}[\sigma^2 (X'X)^{-1})]^{-1}(\hat{\beta} - c) = \frac{(\hat{\beta} - c)' X'X (\hat{\beta} - c)}{\sigma^2}$$

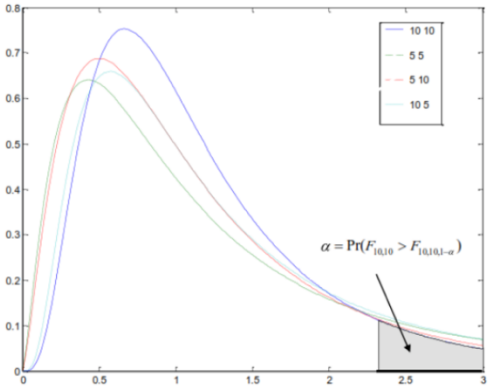
It can be shown that, if H_0 is true, $v_1 \sim \chi^2(k)$.

The variable v_1 cannot be used directly as a test statistic because the σ^2 parameter in the denominator is unknown. However, we know that: $v_2: (T-k)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2(T-k)$

Moreover, since $\hat{\beta}$ and $\hat{\sigma}^2$ are independent, v_1 and v_2 are also independent. Therefore, we can construct the **F-test Statistic** as:

$$\lambda = \frac{v_1/k}{v_2/(T-k)} = \frac{(\hat{\beta}-c)'(X'X)(\hat{\beta}-c)}{k\hat{\sigma}^2} \sim F(k, T-k)$$

We have three of the four elements that we need to construct the test: the **null hypothesis** H_0, the **alternative hypothesis** H_1, the **test statistic** λ. The missing element is the **region of rejection** (or its complement, the acceptance region).



If H_0 is false, λ is expected to be large and positive. Hence, the acceptance region is $\{\lambda < F_c\}$, where F_c is the critical value such that $P_{H_0}\{\lambda \geq F_c\} = \alpha$ and α is the significance level of the test.

Note that λ assumes positive values for both $\beta > c$ and $\beta < c$, so that we cannot conduct directional tests in the case of composite hypotheses. Therefore, the alternative hypothesis should be formulated as $\beta \neq c$.

Testing General Hypotheses with the F-Test

We now introduce a general way to express hypotheses about the parameters of interest. For example, the hypothesis $H_0: \beta_1 = \beta_2$, $H_1: \beta_1 \neq \beta_2$ can be rewritten as:

$$H_0: \beta_1 - \beta_2 = 0 \Leftrightarrow [1 \quad -1]\begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix} = 0 \Leftrightarrow R\beta = r$$

As another example, if $H_0: \frac{\beta_1}{\beta_2} = 4$, $H_1: \frac{\beta_1}{\beta_2} \neq 4$, then we can re-write H_0 as

$$H_0: \beta_1 = \beta_2 4 \Leftrightarrow [1 \quad -4]\begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix} = 0 \Leftrightarrow R\beta = r$$

Also, if $H_0: \beta_1 = \beta_2 = \beta_3 = 3$, $H_1: \beta_1 \neq \beta_2 \neq \beta_3 \neq 3$, then: $H_0: \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\beta_1\\ \beta_2\\ \beta_3\end{bmatrix} = \begin{bmatrix}3\\ 3\\ 3\end{bmatrix} \Leftrightarrow \underset{3\times3}{R}\ \underset{3\times1}{\beta} = \underset{3\times1}{r}$

Therefore, a general hypothesis of interest is: $H_0: \underset{j\times k}{R}\ \underset{k\times1}{\beta} = \underset{j\times1}{r}$, $H_1: R\beta \neq r$

To propose a test statistic for the null hypothesis, note that if $\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$ then $R\hat{\beta} \sim N(R\beta, \sigma^2 R(X'X)^{-1}R')$, and it can be shown that: $v^*_1: (R\hat{\beta}-R\beta)'[\sigma^2 R(X'X)^{-1}R]^{-1}(R\hat{\beta}-R\beta) \sim \chi^2(j)$

Hence, if H_0 holds, $\lambda = \frac{v^*_1/j}{v_2/(T-k)} = \frac{(R\hat{\beta}-r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}-r)}{j\hat{\sigma}^2} \sim F(j, T-k)$

The rejection region is one-sided, as before: $\{\lambda \geq F_c\}$, where F_c is the critical value such that $P_{H_0}\{\lambda \geq F_c\} = \alpha$.

It can be shown that the λ-statistic can be re-written as:

$$\lambda = \frac{(RSS_R - RSS_U)/j}{RSS_U/(T-k)}$$

Where $RSS_U = \sum_{t=1}^{T} \hat{e}_t^2$, $RSS_R = \sum_{t=1}^{T} \tilde{e}_t^2$, $\hat{e}_t$ are the OLS residuals of the unrestricted model, while $\tilde{e}_t$ are the residuals from the model whose coefficients are estimated under the constraints in $H_0: R\beta = r$, namely, they are the residuals of the restricted model.
In general, $RSS_R \geq RSS_U$ but, if the restrictions in the null hypothesis hold, the difference is expected to be small and the test statistic falls in the acceptance region.

If we test a single restriction on the model parameters, so that $j = 1$, then: $\lambda = t^2 = (\frac{\hat{\beta} - c}{\widehat{se}(\hat{\beta})})^2$.
Moreover, if $v \sim t(p)$ then $v^2 \sim F(1, p)$. Hence, $H_0: \underset{1\times1}{\beta} = \underset{1\times1}{c}$ under,
$H_1: \underset{1\times1}{\beta} \neq \underset{1\times1}{c}$

$t^2 = \lambda \sim F(1, T - k)$ and the t-test and F-test have the same p-value.

The single restriction of interest can be formulated more generally. For example, to test: $H_0: \underset{1\times k}{R}\ \underset{k\times1}{\beta} = \underset{1\times1}{r}$
$H_1: R\beta \neq r$

We can use the test statistic: $\frac{R\hat{\beta} - r}{\widehat{se}(R\hat{\beta})} \overset{H_0}{\sim} t(T - k)$

An important application of the F-statistic is to test for the **model significance**, namely, the joint significance of the explanatory variables in a model.
Given the model:

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \ldots + \beta_k x_{kt} + e_t$$

The hypothesis of interest is:

$$H_0: \beta_2 = 0, \beta_3 = 0, \ldots, \beta_k = 0$$
$$H_1: at\ least\ one\ element\ of\ \beta \neq 0$$

Why do we use the F-statistic to verify this composite hypothesis instead of a set of k – 1 t-statistics? Because $\hat{\beta}_2, \ldots, \hat{\beta}_k$ are generally correlated and this feature is taken into account in the construction of the F-test, but not in the set of t-tests.
As a result, it may happen that, for the same significance level, the hypothesis $H_0: \beta_2 = 0, \beta_3 = 0, \ldots, \beta_k = 0$ is not rejected by the F-test, while some of its components are rejected by the associated t-tests.