- The multiple regression model with k explanatory variables can be written as :

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \ldots + \beta_k x_{k,t} + \varepsilon_t,\ t = 1, \ldots, T.$$

- β_0, β_1, ..., β_k are called the partial regression coefficients.

Go back

- The matrix form of the model is :

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_t \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,k} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{t,1} & x_{t,2} & \cdots & x_{t,k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{T,1} & x_{T,2} & \cdots & x_{T,k} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_t \\ \vdots \\ \varepsilon_T \end{pmatrix} \quad (32)$$

$$Y = X \beta + \varepsilon \quad (33)$$

1. The regressors are non stochastics and the matrix of the explanatory variable X must be independent from the error term (exogeneity) : Go to the simple case

$$E(\varepsilon|X) = \begin{pmatrix} E(\varepsilon_1|X) \\ E(\varepsilon_2|X) \\ \vdots \\ E(\varepsilon_T|X) \end{pmatrix} = 0 \iff X'(Y - X\hat{\beta}) = \qquad (34)$$

2. The X matrix is of full rank :

$$Rank(X) = k + 1 \qquad (35)$$

   The explanatory variables are linearly independent (no multicollinearity). We need $T > k + 1$ for that assumption to hold.

What if $T < k + 1$?

3. The error term has zero mean : ‣ Go to the simple case

$$E(\varepsilon) = \begin{pmatrix} E(\varepsilon_1) \\ E(\varepsilon_2) \\ \vdots \\ E(\varepsilon_T) \end{pmatrix} = 0 \tag{36}$$

and

$$E(Y) = X\beta \tag{37}$$

4. Homoskedasticity and no serial correlation of the errors : ‣ Go to the simple case

$$E(\varepsilon\varepsilon') = \sigma_\varepsilon^2 \mathbf{I}, \tag{38}$$

In the general case,

$$\begin{aligned} E(\varepsilon\varepsilon') &= \begin{pmatrix} V(\varepsilon_1) & Cov(\varepsilon_1,\varepsilon_2) & \dots & Cov(\varepsilon_1,\varepsilon_T) \\ Cov(\varepsilon_2,\varepsilon_1) & V(\varepsilon_2) & \dots & Cov(\varepsilon_2,\varepsilon_T) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\varepsilon_T,\varepsilon_1) & Cov(\varepsilon_2,\varepsilon_T) & \dots & V(\varepsilon_T) \end{pmatrix} \qquad (39) \\ &= \begin{pmatrix} E(\varepsilon_1^2) & E(\varepsilon_1\varepsilon_2) & \dots & E(\varepsilon_1\varepsilon_T) \\ E(\varepsilon_2\varepsilon_1) & E(\varepsilon_2^2) & \dots & E(\varepsilon_2\varepsilon_T) \\ \vdots & \vdots & \ddots & \vdots \\ E(\varepsilon_T\varepsilon_1) & E(\varepsilon_2\varepsilon_T) & \dots & E(\varepsilon_T^2) \end{pmatrix} \end{aligned}$$

Assumptions of the multiple regression model III

5. Normality of the errors :
$$\varepsilon \sim N(0, \sigma_\varepsilon^2 \mathbf{I}) \tag{40}$$
6. The model is correctly specified $\Longrightarrow$ There is no specification bias in the model used for the empirical application.

From equation (33) :

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

we get that the OLS estimators minimize the sum of squared residual :

$$\min_{\beta} \sum_{t=1}^{T} \hat{\varepsilon}_t^2 = \min(Y - X\hat{\beta})'(Y - X\hat{\beta}) = \min(e'e)$$

$$(Y - X\hat{\beta})'(Y - X\hat{\beta}) =$$

with

$$= (-\hat{\beta}'X'Y) \qquad - (Y'X\hat{\beta})' =$$

Finally,

$$e'e =$$

We want to minimize $e'e \iff \frac{\partial e'e}{\partial \hat{\beta}} = 0$

$$\begin{aligned} \frac{\partial e'e}{\partial \hat{\beta}} &= \\ &= \end{aligned}$$

since $\frac{\partial A'B}{\partial B} = \frac{\partial B'A}{\partial B} = A$ and $\frac{\partial B'CB}{\partial B} = 2CB$, we have :

$$\frac{\partial e'e}{\partial \hat{\beta}} =$$

and since X is of full rank, its inverse exists and :

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{41}$$

OLS estimator properties I

- $\hat{\beta}$ is a linear estimator of β :

$$\hat{\beta} \quad = \quad (X'X)^{-1}X'Y = \tag{42}$$

- $\hat{\beta}$ is an unbiaised estimator of β :

$$E(\hat{\beta}) \quad = \tag{43}$$

- The OLS coefficients variance-covariance matrix is :

$$\Omega_{\hat{\beta}} = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] \tag{44}$$

$$\Omega_{\hat{\beta}} = \begin{pmatrix} V(\hat{\beta}_0) & Cov(\hat{\beta}_0, \hat{\beta}_1) & Cov(\hat{\beta}_0, \hat{\beta}_2) & \dots & Cov(\hat{\beta}_0, \hat{\beta}_k) \\ Cov(\hat{\beta}_1, \hat{\beta}_0) & V(\hat{\beta}_1) & Cov(\hat{\beta}_1, \hat{\beta}_2) & \dots & Cov(\hat{\beta}_1, \hat{\beta}_k) \\ Cov(\hat{\beta}_2, \hat{\beta}_0) & Cov(\hat{\beta}_2, \hat{\beta}_1) & V(\hat{\beta}_2) & \dots & Cov(\hat{\beta}_2, \hat{\beta}_k) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Cov(\hat{\beta}_k, \hat{\beta}_0) & Cov(\hat{\beta}_k, \hat{\beta}_1) & Cov(\hat{\beta}_k, \hat{\beta}_2) & \dots & V(\hat{\beta}_k) \end{pmatrix}$$

Using equation (42), we get :

$$\Omega_{\hat{\beta}} = \tag{45}$$

- $\hat{\beta}$ is the best estimator : among the set of unbiaised linear estimators of β, $\hat{\beta}$ is the one with the smalest variance.

OLS estimator properties III

Property

The OLS (Ordinary Least Square) estimator $\hat{\beta}$ of β is the Best Linear Unbiased Estimator (BLUE)

Gauss-Markov theorem

In a linear model where the errors are uncorrelated with zero mean and homoscedastic with finite variance, the Ordinary Least Squares (OLS) estimator has the lowest sampling variance within the class of linear unbiased estimators.

$$e \quad = \quad Y - X\hat{\beta}$$

Using $P =$ and the properties of the *trace* operator, we get :

$$E(e'e) \quad =$$

$$\Rightarrow \widehat{\sigma}_{\varepsilon}^2 \quad = \quad \frac{e'e}{T-k-1} = \frac{1}{T-k-1}\sum_{t=1}^{T} e_t^2.$$

Estimator of the variance of the errors

$$\widehat{\sigma}_{\varepsilon}^2 = \frac{1}{T-k-1}\sum_{t=1}^{T} \hat{\varepsilon}_t^2 \tag{46}$$

is an unbiased estimator of σ_{ε}^2.

Exercise 9 : Estimator of the variance of the errors

$$e = \hat{\varepsilon} = \varepsilon - X(X'X)^{-1}X'\varepsilon$$

1. *Rewrite this equation as a function of* $P = I - X(X'X)^{-1}X'$.
2. *Recall that* P *is symmetric* ($P = P'$) *and idempotent* ($P'P = PP' = P$). *Write* $e'e$ *as a function of* ε *and* P.
3. *What is the dimension of* $e'e$ *? What is the dimension of* P *?*
4. *Use the properties of the trace operator to show that* $E\hat{\varepsilon}'\hat{\varepsilon} = \sigma_\varepsilon^2 \; Tr(P)$
5. *Calculate the trace of* P.
6. *Give the expression of an unbiaised estimator of the variance of the errors.*

Outline

- Suppose that the errors are distributed as a normal $\varepsilon \sim N(0, \sigma_\varepsilon^2 \mathbf{I})$ and following equation (42) page 112 :

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

  $\hat{\beta}$ is linear in ε and is normally distributed as a normal $\hat{\beta} \sim N(\beta, \Omega_{\hat{\beta}})$, where $\Omega_{\hat{\beta}} = \sigma_\varepsilon^2 (X'X)^{-1}$ (see equation 45).

- For any $\hat{\beta}_i$ associated to i^{st} explanatory variable x_{it}, we have :

$$\hat{\beta}_i \sim N(\beta_i, \sigma_\varepsilon^2 a_{i+1,i+1}),$$

  where $a_{i+1,i+1}$ is the $(i+1)^{st}$ diagonal element of $(X'X)^{-1}$. We have :

$$\frac{\hat{\beta}_i - \beta_i}{\sigma_\varepsilon \sqrt{a_{i+1,i+1}}} \sim N(0,1).$$

Distribution of the estimated parameters II

From equation (46), we know :

$$\widehat{\sigma}_\varepsilon^2 = \frac{e'e}{T-k-1}.$$

Recall that if $w \sim N(0, \sigma_w^2 \mathbf{I})$ then $\frac{w'w}{\sigma_w^2}$ is distributed as a χ^2 and

$$\frac{e'e}{\sigma_\varepsilon^2} \sim \chi^2_{T-k-1}$$

$$\frac{(T-k-1)\widehat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \sim \chi^2_{T-k-1}$$

Recall that for any $z \sim N(0,1)$ and $\nu \sim \chi^2_r$, $\frac{Z\sqrt{r}}{\sqrt{\nu}} \sim t(r)$. We then have :

$$\frac{\frac{\hat{\beta}_i - \beta_i}{\sigma_\varepsilon \sqrt{a_{i+1,i+1}}} \sqrt{T-k-1}}{\sqrt{\frac{(T-k-1)\widehat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2}}}$$

Finally, we have :

$$\frac{\hat{\beta}_i - \beta_i}{\widehat{\sigma}_\varepsilon \sqrt{a_{i+1,i+1}}} \sim t(T-k-1) \tag{47}$$

- Testing the null hypothesis that β_i equals the particular value of β^* :

$$\begin{cases} H_0 : \beta_i = \beta^* \\ H_a : \beta_i \neq \beta^* \end{cases}$$

Under H_0,

$$\frac{\hat{\beta}_i - \beta^*}{\widehat{\sigma}_\varepsilon \sqrt{a_{i+\mathbf{1},i+\mathbf{1}}}} \sim t(T-k-1)$$

and the decision rule gives :

- If $\left|\frac{\hat{\beta}_i - \beta^*}{\widehat{\sigma}_\varepsilon \sqrt{a_{i+\mathbf{1},i+\mathbf{1}}}}\right| \leq t_{p/\mathbf{2}}$ we do not reject $H_\mathbf{0}$ that $\beta_i = \beta^*$ at a $100\alpha\%$ significance level.
- If $\left|\frac{\hat{\beta}_i - \beta^*}{\widehat{\sigma}_\varepsilon \sqrt{a_{i+\mathbf{1},i+\mathbf{1}}}}\right| > t_{p/\mathbf{2}}$ we reject $H_\mathbf{0}$ and $\beta_i \neq \beta^*$ at a $100\alpha\%$ significance level.

Testing individually the estimated parameters II

The most common test is the significance of the parameters :

- Testing the null hypothesis that $\beta_i = 0$:

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

Under H_0,

$$t_{\hat{\beta}_i} = \frac{\hat{\beta}_i}{\widehat{\sigma}_{\hat{\beta}_i}} \sim t(T-k-1) \text{ where } \widehat{\sigma}_{\hat{\beta}_i} = \widehat{\sigma}_{\varepsilon}\sqrt{a_{i+1,i+1}}$$

- If $\left|\frac{\hat{\beta}_i}{\widehat{\sigma}_{\hat{\beta}_i}}\right| \leq t_{p/2}$ we do not reject H_0 that $\beta_i = 0$ at a $100\alpha\%$ significance level $\Longleftrightarrow$ the variable X_{it} is not significant and does not influence Y_t.
- If $\left|\frac{\hat{\beta}_i}{\widehat{\sigma}_{\hat{\beta}_i}}\right| > t_{p/2}$ we reject H_0 and $\beta_i \neq 0$ at a $100\alpha\%$ significance level $\Longleftrightarrow$ the variable X_{it} is significant and does influence Y_t.

- Testing the significance of an isolated parameter corresponds to a Student t-test.
- Testing the significance of several (all) the estimated parameters at the same time corresponds to a Fisher test. Suppose that the parameters are subject to r constraints :

$$R\beta = r,$$

where R is a $(q, k+1)$ matrix and r a vector of size q. We suppose $q \leq k+1$ and R to be full rank $\iff$ the constraints are linearly independent.

  - The case where $R = [0 \dots 010 \dots 0]$ and $r = 0$ corresponds to
  - The case where $R = [01-1 \dots 0]$ and $r = 0$ corresponds to

- The case where

$$R = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix} \text{ and } r = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

corresponds to

Exercise 10 : Test examples

Considering the model

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \ldots + \beta_k x_{k,t} + \varepsilon_t,\ t = 1, \ldots, T.$$

give R and r in the following cases :

1. $H_0 : \beta_3 = 1$; $H_a : \beta_3 \neq 1$.
2. $H_0 : \beta_0 + \beta_1 = 1$; $H_a : \beta_0 + \beta_1 \neq 1$.
3. $H_0 : \beta_2 \times \beta_1 = 0$; $H_a : \beta_2 \times \beta_1 \neq 0$.

- Aiming at implementing the previous test, we need to consider two models :
  - The unconstrained model $\mathcal{M}_{uc}$ corresponds to the regression of the dependent variable on the set of all explanatory variables without imposing any contraint on the parameters.
  - The constrained model $\mathcal{M}_c$ corresponds to the regression of the dependent variable ona sub-set of explanatory variables that account for the linear contraints on the parameters.

  - The constrained model $\mathcal{M}_c$ is a special case of the unconstrained model $\mathcal{M}_{uc}$.
  - The constrained $\mathcal{M}_c$ and unconstrained $\mathcal{M}_{uc}$ models must be estimated to compute the statistic of the test F.

- The statistic of the test is :

$$F = \frac{(ESS_c - ESS_{uc})/q}{ESS_{uc}/(T-k-1)} \sim F(q, T-k-1) \tag{48}$$

- The Fisher test is testing that the relative gap between the two ESS is significant $H_0 : ESS_c - ESS_{uc} = 0$ and thus, that the constaint is statistically justified.

What is the alternative hypothesis of that test ?

- If $F \leq F(q, T-k-1)$, we do not reject the null hypothesis H_0. In the contrary case, on reject the null.
- The Student-t test of the significativity of β_i can be seen as a Fisher test, where the constrained model resums to the regression of the dependent variable on a set of explanatory variables including all the explanatory variables but X_{it}. In that case, the Student-t statistic, t-stat is such that $t^2(T-k-1) = F(1, T-k-1)$.
- The null hypothesis of the overall significance test of a multiple regression, is $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$.
- The unconstrained model is :

Testing simultaneously some estimated parameters V

- The constrained model is
- We have $\hat{y}_t =$

$$ESS_c =$$

- Replacing, ESS_c and ESS_{uc} in

$$F = \frac{(ESS_c - ESS_{uc})/q}{ESS_{uc}/(T-k-1)} \sim F(q, T-k-1), \quad (49)$$

we get,

- The Student test of the significance of β_i can be interpreted as a Fisher test, where the constraint model resumes to regressing the dependant variable over the set of explanatory variables but x_{it}.

Exercise 11 : Constrained vs unconstrained model

Consider the model

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{3,t} + u_t,\ t = 1, \dots, T$$

Describe the method (set of regressions included) to test $\{\beta_1 = 1; \beta_3 = -1\}$.

- To perform multiple tests,
  - we estimate the two models $\mathcal{M}_c$ and $\mathcal{M}_{nc}$ and each time, we extract the *ESS*,
  - we compute the statistic of the test,
  - we compare it to a critical value of the corresponding distribution,
  - we reject H_0 whenever the statistic is greater than the critical value (tabulated value).
- Equivalently,
  - we can calculate the risk of rejecting the true hypothesis (type-I error),
  - compare it to a risk tolerence level we are willing to bear (α),
  - we reject H_0 if the p-value is lower than the risk we are willing to bear.