CH2 : Le modèle de régression linéaire multiple

I) Présentation du modèle et hypothèses :

1 Ecriture du modèle sous forme matricielle :

Le modèle de régression multiple comporte k variables explicatives et s'écrit comme suit :

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_k x_{kt} + \varepsilon_t$$

Avec $t = 1, \ldots, T$.

* y_t : variable endogène ou à expliquer ;
* $x_{1t}, x_{2t}, \ldots, x_{kt}$: variables exogènes ou explicatives ;
* ε_t : terme d'erreur ou variable aléatoire ;
* $\beta_0, \beta_1, \ldots, \beta_k$: les paramètres à estimer ou coefficients de régression.

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_k x_{kt} + \varepsilon_t$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & & x_{k2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1T} & x_{2T} & & x_{kT} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix}$$

$$\boxed{\Rightarrow \underset{(T,1)}{Y} = \underset{(T,\underbrace{k+1}_{K})}{X}\ \underset{(\underbrace{k+1}_{K},1)}{\beta} + \varepsilon \longrightarrow (T,1)}$$

Avec :

* Y : C'est le vecteur composé par les valeurs de la variable endogène.

* X : C'est la matrice des variables explicatives y compris les constantes.
* β : C'est le vecteur des paramètres à estimer.
* ε : C'est le vecteur relatif au terme d'erreur.

② Les hypothèses du modèle:
Le modèle de régression multiple repose sur les hypothèses suivantes:

② Les hypothèses du modèle :

Le modèle de régression multiple repose sur les hypothèses suivantes :

* H_1 : La matrice X est non aléatoire.
* H_2 : La matrice X est de plein rang $\Rightarrow$ $Rang(X) = K = k+1$ ~~des variables explicatives y compris les constantes~~

$\Rightarrow$ Absence de colinéarité entre les variables explicatives (indépendance entre les variables explicatives)

$\Rightarrow$ $(X'X)$ est symétrique et régulière et $(X'X)^{-1}$ existe.

Avec : X' = transposée de X.

* H_3 : $E(\varepsilon) = 0$.
* H_4 : hypothèses d'homoscédasticité et d'absence d'autocorrélation des erreurs.

$Var(\varepsilon) = \Omega_\varepsilon = \sigma_\varepsilon^2 . I$; I : matrice identité

avec : Ω_ε = la matrice de variances-covariances des erreurs.

$$Var(\varepsilon) = \Omega_\varepsilon = \begin{pmatrix} V(\varepsilon_1) & cov(\varepsilon_1, \varepsilon_2) & \cdots & cov(\varepsilon_1, \varepsilon_T) \\ cov(\varepsilon_2, \varepsilon_1) & & & \vdots \\ \vdots & & & \\ cov(\varepsilon_T, \varepsilon_1) & \cdots & & V(\varepsilon_T) \end{pmatrix}$$

$$\Rightarrow \Omega_\varepsilon = \begin{pmatrix} \sigma_\varepsilon^2 & & (0) \\ & \ddots & \\ (0) & & \sigma_\varepsilon^2 \end{pmatrix} = \sigma_\varepsilon^2 \cdot I$$

$\Rightarrow$ $Var(\varepsilon) = \Omega_\varepsilon = \sigma_\varepsilon^2 \cdot I$.

* H_5 : Hypothèse de normalité des erreurs : $\varepsilon \rightsquigarrow \mathcal{N}(0, \sigma_\varepsilon^2 \cdot I)$

(2) C. ML

II] Estimation par la méthode des MCO :

1) Déf :

On considère ci-dessous le modèle de régression multiple sous sa forme matricielle avec k variables exogènes et T observations.

$$Y = X.\beta + \varepsilon \iff \varepsilon = Y - X.\beta$$

Pour estimer le vecteur β, on applique la méthode des MCO qui consiste à minimiser la somme des carrées des erreurs.

$$\Rightarrow \min_{\beta} \sum_{t=1}^{T} \varepsilon_t^2 = \min_{\beta} \varepsilon'\varepsilon$$

2) Résolution :

$$\min_{\beta} \sum_{t=1}^{T} \varepsilon_t^2 = \min_{\beta} \varepsilon'\varepsilon = \min_{\beta} \underbrace{(Y - X\beta)'(Y - X\beta)}_{Q}$$

$$= \min_{\beta} Q$$

Avec $Q = (Y - X\beta)'(Y - X\beta)$.

$$= (Y' - \beta'X')(Y - X\beta)$$

$$= Y'Y - Y'X\beta - \beta'X'Y + \beta'X'X\beta$$

Mq (mathématiquement).

$$Y'X\beta = (\beta'X'Y)' = \beta'X'Y .$$

$$\Rightarrow Q = Y'Y - 2\beta'X'Y + \beta'X'X\beta ,$$

$$\Rightarrow \frac{\partial Q}{\partial \beta} = 0 \Leftrightarrow -\cancel{2}X'Y + \cancel{2}X'X\beta = 0$$

$$\Leftrightarrow X'Y - X'X\beta = 0$$

$\Leftrightarrow X'Y = X'X\beta$: Equations normales puisque $(X'X)^{-1}$ existe d'après H_2 donc on obtient :

$$\boxed{\hat{\beta} = (X'X)^{-1}.X'Y} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} .$$

<u>Notations</u> :

* $\hat{Y} = X.\hat{\beta}$: le vecteur des valeurs ajustées de la variable endogène.
* $\hat{\varepsilon} = Y - \hat{Y}$: le ~~résidu~~ vecteur résidu de l'estimation.

3) Equation d'analyse de la variance :

Comme dans le cas du modèle de régression simple l'équation d'analyse de la variance s'écrit comme suit : SCT = SCE + SCR

(8) C.M.L

Avec:

* $SCT = \sum (y_t - \bar{y})^2 = \sum y_t^2 - T\bar{y}^2 = Y'Y - T\bar{y}^2$
* $SCE = \sum (\hat{y}_t - \bar{y})^2 = \sum \hat{y}_t^2 - T\bar{y}^2 = \hat{Y}'\hat{Y} - T\bar{y}^2$.
* $SCR = \sum \hat{\varepsilon}_t^2 = \hat{\varepsilon}'\hat{\varepsilon}$

Théorème:

$$\boxed{SCR = Y'Y - \hat{\beta}'X'Y}$$

Démonstration:

$SCR = SCT - SCE$

$= Y'Y - T\bar{y}^2 - \hat{Y}'\hat{Y} + T\bar{y}^2$.

$\Rightarrow SCR = Y'Y - \hat{\beta}'\underbrace{X'X\hat{\beta}}_{=X'Y}$.

Equations normales: $X'Y = X'X\hat{\beta}$.

$$\Rightarrow \boxed{SCR = Y'Y - \hat{\beta}'X'Y}$$