

Introduction à l'économétrie

Le modèle linéaire multiple

Ce cours vous est proposé par Olivier Baron, Maître de conférences, Université de Bordeaux et par AUNEGe, l'Université Numérique en Économie Gestion.

Table des matières

Préambule	3
Écriture du modèle et notation	3
Les hypothèses classiques du modèle	5
Estimation par les moindres carrés ordinaires (MCO)	7
Notations.....	7
Définition de l'estimateur des MCO.....	8
Interprétation géométrique.....	11
Équation d'analyse de la variance	13
Le coefficient de détermination	15
Propriétés statistiques des MCO	16
Propriétés désirées d'un estimateur	16
Espérance de β	17
Variance de β	17
Théorème de Gauss-Markov	19
Conséquences de l'hypothèse de normalité des perturbations	20
Loi de β	21
Loi de σ^2	21

Références **23**

Préambule

Objectifs :

- Présentation des notations et de la méthode des Moindres Carrés Ordinaires dans le cadre du modèle linéaire multiple.

On considère dans ce chapitre le cas d'un modèle linéaire comportant plusieurs variables explicatives.

Écriture du modèle et notation

Considérons le modèle suivant fourni par la théorie économique :

$$y_i = \beta_0 + \beta_1 \cdot x_{1,i} + \beta_2 \cdot x_{2,i} + \dots + \beta_{k-1} \cdot x_{k-1,i} + \varepsilon_i \quad (1)$$

On cherche à expliquer les variations du phénomène y par un modèle linéaire en fonction des variables explicatives x_1, x_2, \dots, x_{k-1} , des k paramètres $\beta_0, \beta_1, \dots, \beta_{k-1}$ et d'une perturbation aléatoire ε qui intervient de façon additive. Le modèle est qualifié de linéaire car y est une fonction linéaire des paramètres. La variable y est appelée variable « à expliquer » ou encore variable dépendante.

Parmi les k paramètres figure une constante β_0 .

La perturbation ε rend compte de l'ensemble des déterminants du phénomène y qui ne figurent pas dans la liste des variables explicatives. En effet, on ne connaît jamais de façon exhaustive l'ensemble des causes d'un phénomène socio-économique.

Pour estimer ce modèle, on dispose de n observations sur la variable dépendante et les $(k-1)$ variables explicatives ($i = 1, 2, \dots, n$).

À noter

La règle suivante **doit impérativement être respectée** : $n \geq k$.

Le nombre d'observations doit être supérieur ou égal au nombre de paramètres à estimer. Nous verrons plus tard que cette contrainte est une condition nécessaire de nature essentiellement technique.

Pour aboutir à l'écriture matricielle du modèle, on doit empiler les n observations. Le modèle se traduit ainsi par un système de n équations linéaires à k inconnues.

$$\begin{cases} y_1 = \beta_0 + \beta_1 \cdot x_{1,1} + \beta_2 \cdot x_{2,1} + \dots + \beta_{k-1} \cdot x_{k-1,1} + \varepsilon_1 \\ \vdots \\ y_i = \beta_0 + \beta_1 \cdot x_{1,i} + \beta_2 \cdot x_{2,i} + \dots + \beta_{k-1} \cdot x_{k-1,i} + \varepsilon_i \\ \vdots \\ y_n = \beta_0 + \beta_1 \cdot x_{1,n} + \beta_2 \cdot x_{2,n} + \dots + \beta_{k-1} \cdot x_{k-1,n} + \varepsilon_n \end{cases}$$

On peut ensuite écrire l'ensemble de ces n équations sous la forme réduite suivante :

$$y_{(n,1)} = \beta_0 \cdot e_{n,(n,1)} + \beta_1 \cdot x_{1,(n,1)} + \dots + \beta_{k-1} \cdot x_{k-1,(n,1)} + \varepsilon_{(n,1)}$$

où :

$y_{(n,1)} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ est le vecteur des n observations de la variable dépendante ;

$x_{1,(n,1)} = \begin{pmatrix} x_{1,1} \\ \vdots \\ x_{1,n} \end{pmatrix}$ est le vecteur des n observations de la première variable explicative ;

de même $x_{2,(n,1)}, \dots, x_{k-1,(n,1)}$ sont les vecteurs des n observations de la deuxième, ..., la $(k-1)$ ème variable explicative ;

$e_{n,(n,1)} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ est le vecteur lié à l'introduction de la constante β_0 dans la spécification du modèle ;

$\varepsilon_{(n,1)} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$ est le vecteur des n réalisations, non observées, de la perturbation.

En notant : $X_{(n,k)} = \begin{bmatrix} e_{n,(n,1)} & x_{1,(n,1)} & \dots & x_{k-1,(n,1)} \end{bmatrix}$ et $\beta_{(k,1)} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{k-1} \end{pmatrix}$, le modèle peut alors s'écrire sous la forme

matricielle suivante :

$$y_{(n,1)} = X_{(n,k)} \cdot \beta_{(k,1)} + \varepsilon_{(n,1)} \quad (2)$$

La matrice $X_{(n,k)}$ est la matrice des variables explicatives et le vecteur $\beta_{(k,1)}$ est le vecteur des paramètres à estimer.

Par la suite, on préférera parfois une écriture « en ligne » de la matrice $X_{(n,k)}$ des variables explicatives, sous la forme suivante :

$$X_{(n,k)} = \begin{pmatrix} X'_{1(1,k)} \\ X'_{2(1,k)} \\ \vdots \\ X'_{n(1,k)} \end{pmatrix}$$

Dans la première expression de $X_{(n,k)}$, on considèrerait une écriture en colonnes, chaque colonne j regroupant les n observations de la $j^{\text{ème}}$ variable explicative alors que la seconde expression de $X_{(n,k)}$ décompose la même matrice en lignes. Chaque ligne $X'_{i(1,k)}$ correspond aux observations, pour l'individu i (ou à la date i), des $(k-1)$ variables explicatives.

Ainsi, $X'_{i(1,k)} = (1, x_{1,i}, x_{2,i}, \dots, x_{k-1,i})$.

À partir de cette dernière écriture, l'expression (1) du modèle linéaire multiple, pour l'observation i , peut être formalisée de la façon suivante :

$$y_i = X'_{i(1,k)} \cdot \beta_{(k,1)} + \varepsilon_i \quad (3)$$

Désormais, on raisonnera directement sur la forme matricielle générale du modèle linéaire multiple (2) ou sur l'expression (3) lorsque nous aurons besoin de nous focaliser sur une observation particulière.

Les hypothèses classiques du modèle

Il faut avoir à l'esprit que les hypothèses qui vont être détaillées dans cette section ne sont pas nécessairement vérifiées. Nous verrons dans la suite de ce cours les conséquences de la non vérification de certaines d'entre elles.

Considérons le modèle linéaire multiple écrit sous forme matricielle : $y_{(n,1)} = X_{(n,k)} \cdot \beta_{(k,1)} + \varepsilon_{(n,1)}$

$$. H_1 : E \begin{pmatrix} \varepsilon \\ (n,1) \end{pmatrix} = \begin{pmatrix} 0 \\ (n,1) \end{pmatrix} \Leftrightarrow E(\varepsilon_i) = 0 \quad \forall i = 1, 2, \dots, n$$

La perturbation est **d'espérance nulle**. Cette hypothèse est d'autant plus réaliste que la taille de l'échantillon étudié est importante. Par ailleurs, le fait que la spécification retenue pour le modèle comporte une constante, rend cette hypothèse plausible même pour de petits échantillons.

. H₂ : La matrice des variables explicatives $X_{(n,k)}$ est une matrice certaine.

Sous cette hypothèse, les $k-1$ variables contenues dans $X_{(n,k)}$ sont des variables économiques au même titre que $y_{(n,1)}$. Retenir H₂ signifie que l'on cherche à modéliser le phénomène

$y_{(n,1)}$ conditionnellement aux réalisations $X_{i'}'_{(1,k)}$ observées dans l'échantillon. Cette hypothèse

implique que $X_{i'}'_{(1,k)}$ et $\varepsilon_{i'}$ sont indépendantes pour tout i et i' . Elle conduit à des estimateurs sans

biais.

. H₃ : La matrice $X_{(n,k)}$ est de rang égal à k . On fait ici l'hypothèse que les k colonnes e_n, x_1, \dots, x_{k-1}

qui composent la matrice $X_{(n,k)}$ sont des vecteurs linéairement indépendants. Si le rang $X_{(n,k)}$ était

strictement inférieur à k , cela signifierait qu'il existe au moins une variable explicative dont le vecteur colonne de ses observations pourrait s'écrire comme une combinaison linéaire des vecteurs colonnes des observations d'autres variables explicatives. Cette variable serait alors superflue. Cette hypothèse signifie donc que le modèle est correctement spécifié ; il n'y a pas de redondance dans la liste des variables explicatives.

On peut remarquer que la règle $n \geq k$ imposée dès le départ, est une condition nécessaire pour que H₃ soit vérifiée.

$$. H_4 : E \begin{pmatrix} \varepsilon, \varepsilon' \\ (n,n) \end{pmatrix} = \sigma^2 \cdot I_n$$

La matrice de variances-covariances des perturbations est une matrice scalaire. Elle s'écrit comme le produit de la matrice identité par un nombre réel positif. Cette hypothèse est en fait double :

- On suppose tout d'abord que la variance de la perturbation est constante (hypothèse d'homoscédasticité des perturbations) : $E(\varepsilon_i^2) = Var(\varepsilon_i) = \sigma^2 \quad \forall i$
- On suppose ensuite que la covariance entre deux réalisations différentes de la perturbation est nulle (hypothèse d'absence d'autocorrélation des erreurs) :

$$E(\varepsilon_i \varepsilon_{i'}) = Cov(\varepsilon_i, \varepsilon_{i'}) = 0 \quad \forall i, \forall i', i \neq i'$$

Lorsque les hypothèses H_1 et H_4 sont simultanément vérifiées, les erreurs sont qualifiées de sphériques.

. H_5 : La matrice de variances-covariances empirique des variables explicatives, S_X , converge $_{(k,k)}$ vers une matrice finie et définie-positive lorsque n tend vers l'infini. L'idée importante est ici que les $x_{j,i}$ conservent toujours une certaine variance quand n devient grand : des observations supplémentaires améliorent l'information.

. H_6 : $\varepsilon_{(n,1)} \rightarrow N(0, \sigma^2 \cdot I_n)$

Cette hypothèse signifie que les perturbations ε_i sont indépendantes et identiquement distribuées, telles que $\varepsilon_i \rightarrow N(0, \sigma^2)$ (perturbations i.i.d.).

L'explication en vidéo de la matrice de variance-covariance



Estimation par les moindres carrés ordinaires (MCO)

Notations

Soit β_j un coefficient que l'on cherche à estimer. Pour ne pas confondre le coefficient inconnu β_j et son estimation (réalisée sur la base de l'échantillon de données dont dispose l'économètre), on note cette dernière $\hat{\beta}_j$. Cette écriture désigne à la fois l'estimation obtenue et l'estimateur utilisé pour obtenir cette estimation.

Définition de l'estimateur des MCO

Considérons le modèle linéaire multiple (2) :

$$y_{(n,1)} = X_{(n,k)} \cdot \beta_{(k,1)} + \varepsilon_{(n,1)}$$

Le but que se fixe l'économètre est de connaître $\beta_{(k,1)}$ le mieux possible. Il s'agira donc de construire un estimateur $\hat{\beta}$ du vecteur β qui possède un certain nombre de propriétés voulues.

Pour une observation i , on a vu que le modèle linéaire multiple s'écrivait :

$$y_i = \beta_0 + \beta_1 \cdot x_{1,i} + \beta_2 \cdot x_{2,i} + \dots + \beta_{k-1} \cdot x_{k-1,i} + \varepsilon_i = X_i' \cdot \beta_{(k,1)} + \varepsilon_i$$

On dispose pour estimer le vecteur β , d'un échantillon de n observations. Chaque observation i est caractérisée par la réalisation y_i de la variable expliquée et les réalisations des variables explicatives $x_{1,i}, x_{2,i}, \dots, x_{k-1,i}$. Notre échantillon correspond donc à un nuage de points M_i ($i = 1 \dots n$), chaque point ayant pour coordonnées le k -uplet $(y_i, x_{1,i}, \dots, x_{k-1,i})$. Il s'agira donc de trouver les valeurs $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_{k-1}$ telles que l'hyperplan d'équation :

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot x_{1,i} + \hat{\beta}_2 \cdot x_{2,i} + \dots + \hat{\beta}_{k-1} \cdot x_{k-1,i} = X_i' \cdot \hat{\beta}_{(k,1)}$$

passse le plus près possible de tous les points M_i du nuage.

Comme dans le cas du modèle linéaire simple, le critère retenu consiste à choisir $\hat{\beta}_{(k,1)}$ de façon à minimiser la somme des carrés des distances entre l'hyperplan et les observations caractérisées par les points M_i et où $\hat{\beta}$ représente le vecteur dont les composantes sont les estimateurs $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_{k-1}$ des composantes du vecteur β .

Par définition, $\hat{\beta}$ est solution du programme suivant :

$$\text{Min}_{\hat{\beta}} \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \text{Min}_{\hat{\beta}} \sum_{i=1}^n \left(y_i - X_i' \cdot \hat{\beta}_{(k,1)} \right)^2$$

On montre simplement que le programme précédent s'écrit sous forme matricielle :

$$\text{Min}_{\hat{\beta}} (y - X\hat{\beta})' \cdot (y - X\hat{\beta})$$

où $(y - X\hat{\beta})'$ est le vecteur transposé de $(y - X\hat{\beta})$.



Une condition nécessaire pour que $\hat{\beta}$ soit solution du programme précédent est qu'il vérifie les conditions de premier ordre. On montre que $\hat{\beta}$ doit alors vérifier :

$$\begin{matrix} X' & \cdot & X & \cdot & \hat{\beta} & = & X' & \cdot & y \\ (k,n) & & (n,k) & & (k,1) & & (k,n) & & (n,1) \end{matrix}$$

En examinant les formats des matrices et vecteurs figurant dans l'expression précédente, on constate qu'il s'agit d'un système de k équations linéaires à k inconnues. Ces k inconnues sont les composantes du vecteur $\hat{\beta}$.

Ce système d'équations est appelé **système des équations normales**. Il admet une solution unique si les k équations sont linéairement indépendantes, c'est-à-dire si la matrice $(X'X)$ est régulière. Dans ce cas le système ci-dessus est un système de Cramer.

Cette condition est vérifiée dès lors que l'on retient l'hypothèse H_3 . En effet, sous cette hypothèse, $\text{rang } X_{(n,k)} = k$ donc $\text{rang } X'_{(n,k)} = k$ et ainsi : $\text{rang } X'X_{(k,k)} = k$. Or, une matrice carrée d'ordre k dont le rang est égal à k est régulière (inversible). On peut alors déterminer $\hat{\beta}$ de manière unique :

$$\hat{\beta}_{(k,1)} = (X'X)_{(k,k)}^{-1} \cdot X'_{(k,n)} \cdot y_{(n,1)} \quad (4)$$

$\hat{\beta}$ est l'estimateur des MCO de β dans le modèle linéaire multiple (2).

L'explication en vidéo des conditions de premier ordre



On montre que les conditions de second ordre sont aussi vérifiées dès lors que la matrice $X_{(n,k)}$ est de plein rang ($\text{rang } X_{(n,k)} = k$). Ainsi, la solution des MCO, $\hat{\beta}_{(k,1)}$, minimise la somme des carrés des résidus estimés.

La valeur ajustée du vecteur y est définie par : $\hat{y}_{(n,1)} = X_{(n,k)} \cdot \hat{\beta}_{(k,1)}$.

Le vecteur des résidus estimés est égal à : $\hat{\varepsilon}_{(n,1)} = y_{(n,1)} - \hat{y}_{(n,1)}$.

L'explication en vidéo des conditions de second ordre



Remarque

1. L'expression (4) montre que l'estimateur $\hat{\beta}$ est un vecteur aléatoire, défini comme une fonction des variables aléatoires y_i .
2. $\hat{\beta}$ est un estimateur linéaire : il est égal au produit du vecteur y par la matrice $(X'X)^{-1}X'$.
3. Quand H_3 n'est pas vérifiée ($\text{rang}_{(n,k)} X < k$) on dit que le modèle n'est pas identifiable. Dans ce cas, le système des équations normales est un système à k inconnues avec un nombre d'équations indépendantes inférieur à k . Il admet donc une infinité de solutions et les composantes du vecteur $\hat{\beta}$ ne sont pas définies de manière unique. L'influence de chacune des $(k-1)$ variables explicatives sur y ne peut pas être identifiée.

Interprétation géométrique

Lorsque l'échantillon utilisé comprend n observations, les observations de la variable expliquée et des variables explicatives forment des vecteurs de R^n .

Reconsidérons le modèle linéaire multiple défini par :

$$\underset{(n,1)}{y} = \underset{(n,k)}{X} \cdot \underset{(k,1)}{\beta} + \underset{(n,1)}{\varepsilon}$$

Si H_3 est vérifiée, la matrice $\underset{(n,k)}{X}$ de rang égal à k , définit un sous espace vectoriel de R^n de dimension k . Cet espace, noté $\mathcal{L}(X)$, est engendré par les k vecteurs linéairement indépendants qui composent la matrice $\underset{(n,k)}{X}$.

Par définition, $\underset{(n,1)}{\hat{y}} \in \mathcal{L}(X)$ puisque \hat{y} est une combinaison linéaire des vecteurs de X ($\underset{(n,1)}{\hat{y}} = \underset{(n,k)}{X} \cdot \underset{(k,1)}{\hat{\beta}}$).

Le principe des MCO consiste donc à déterminer $\underset{(n,1)}{\hat{y}} \in \mathcal{L}(X)$, tel que la distance entre y et \hat{y} soit minimale. En mesurant la distance entre les vecteurs y et \hat{y} par la norme euclidienne $\|y - \hat{y}\|$, cette distance minimale est obtenue en définissant \hat{y} comme la projection orthogonale de y sur $\mathcal{L}(X)$.

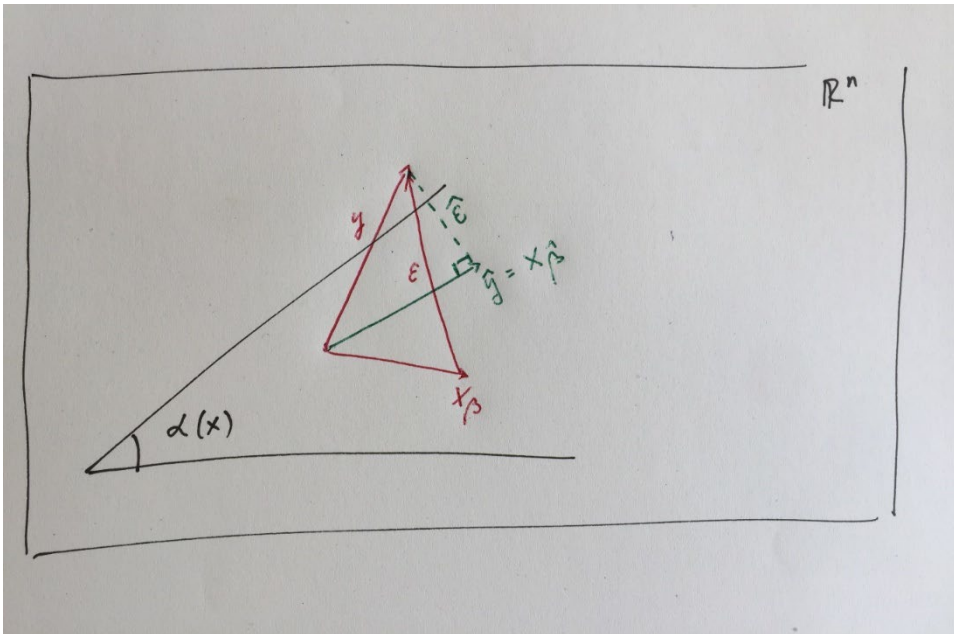


Figure 1 : approche géométrique - Courbe d'illustration des vecteurs

Pour obtenir \hat{y} , on projette y , vecteur de \mathbb{R}^n sur $\mathcal{L}(X)$, sous espace de dimension k engendré par les vecteurs de X . Le vecteur y est égal à la somme $\hat{y} + \hat{\varepsilon}$.

De plus, \hat{y} étant égal à la projection orthogonale de y sur $\mathcal{L}(X)$, on en déduit que $\hat{\varepsilon}$ est un vecteur orthogonal à $\mathcal{L}(X)$.

Cette approche géométrique permet d'interpréter certains vecteurs en terme de projections.

On sait que : $\hat{y} = X \cdot \hat{\beta} = X \cdot (X'X)^{-1} \cdot X' \cdot y = P_X \cdot y$

$P_X = X \cdot (X'X)^{-1} \cdot X'$ est la matrice de projection orthogonale sur $\mathcal{L}(X)$. Comme toute matrice de projection orthogonale elle est symétrique et **idempotente**¹. Son rang est égal à la dimension de l'espace de projection : $\text{rang } P_X = k$.

De même, on sait que $\hat{\varepsilon} = y - \hat{y} = y - P_X \cdot y = (I_n - P_X) \cdot y = M_X \cdot y$ où I_n est la matrice identité d'ordre n .

$M_X = I_n - P_X = I_n - X \cdot (X'X)^{-1} \cdot X'$ est la matrice de projection orthogonale sur $\mathcal{L}^\perp(X)$ où $\mathcal{L}^\perp(X)$ désigne l'espace orthogonal à $\mathcal{L}(X)$. Cette matrice est aussi symétrique et idempotente et son rang est égal à la dimension de l'espace de projection soit : $\text{rang } M_X = n - k$.

¹ Une matrice M est idempotente si et seulement si $M^2 = M$.

M_X et P_X étant des matrices de projection sur des espaces orthogonaux elles vérifient :

$$M_X \cdot P_X = O_n \text{ où } O_n \text{ est la matrice nulle d'ordre } n.$$

Si $\hat{\varepsilon}$ correspond à la projection orthogonale de y sur $\mathcal{L}^\perp(X)$, ce vecteur peut aussi être interprété comme la projection orthogonale de ε sur $\mathcal{L}^\perp(X)$.

En effet, $\hat{\varepsilon} = M_X \cdot y = M_X \cdot (X \cdot \beta + \varepsilon) = M_X \cdot \varepsilon$, puisque comme X est orthogonal à $\mathcal{L}^\perp(X)$, $M_X \cdot X = O_{(n,n)}$.

Équation d'analyse de la variance

Avant de démontrer cette équation, nous allons montrer deux résultats préliminaires importants.

Résultat 1 : Si le modèle comporte une constante, la moyenne des résidus estimés est égale à 0.

Preuve : Puisque e_n est l'un des vecteurs colonnes de la matrice X , $e_n \in \mathcal{L}(X)$. Le vecteur des résidus estimés $\hat{\varepsilon}$ est donc, par définition des MCO, orthogonal à e_n . On en déduit donc que :

$$e_n' \cdot \hat{\varepsilon} = 0$$

⇔

$$\sum_{i=1}^n \hat{\varepsilon}_i = 0$$

⇔

$$\bar{\hat{\varepsilon}} = \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i = 0$$

Ce résultat implique que la moyenne des observations de la variable expliquée y est égale à la moyenne des valeurs ajustées de cette variable (les \hat{y}_i). En effet, puisque pour toute observation i on a l'égalité : $y_i = \hat{y}_i + \hat{\varepsilon}_i$, cette égalité est toujours vraie en sommant chaque membre de l'équation et donc :

$$\sum_{i=1}^n y_i = \sum_{i=1}^n \hat{y}_i + \sum_{i=1}^n \hat{\varepsilon}_i = \sum_{i=1}^n \hat{y}_i$$

puisque la somme des résidus estimés est nulle. On déduit donc que $\bar{y} = \bar{\hat{y}}$.

Résultat 2 : Le modèle estimé passe par le point moyen de l'échantillon.

Preuve : En développant l'expression de $\bar{\hat{y}}$ on obtient :

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n \hat{y}_i = \frac{1}{n} \sum_{i=1}^n (\hat{\beta}_0 + \hat{\beta}_1 \cdot x_{1,i} + \hat{\beta}_2 \cdot x_{2,i} + \dots + \hat{\beta}_{k-1} \cdot x_{k-1,i})$$

⇔

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot \bar{x}_1 + \hat{\beta}_2 \cdot \bar{x}_2 + \dots + \hat{\beta}_{k-1} \cdot \bar{x}_{k-1} = \bar{y}$$

compte tenu du premier résultat. On voit donc que le modèle estimé est exactement vérifié au point moyen de l'échantillon et donc que l'hyperplan de régression passe par ce point.

On peut maintenant écrire l'équation d'analyse de la variance en procédant de la façon suivante :

On sait que :

$$y_{(n,1)} = \hat{y}_{(n,1)} + \hat{\varepsilon}_{(n,1)}$$

d'où, en retranchant le vecteur $(\bar{y} \cdot e_n)_{(n,1)}$ de chaque membre de cette équation, on obtient :

$$(y_{(n,1)} - \bar{y} \cdot e_n)_{(n,1)} = (\hat{y}_{(n,1)} - \bar{y} \cdot e_n)_{(n,1)} + \hat{\varepsilon}_{(n,1)}$$

On a vu que $\hat{y} \in \mathcal{L}(X)$. Il en est de même pour $\bar{y} \cdot e_n$ puisque $e_n \in \mathcal{L}(X)$ et \bar{y} est un scalaire. Ainsi

$(\hat{y}_{(n,1)} - \bar{y} \cdot e_n)_{(n,1)}$ est un vecteur de $\mathcal{L}(X)$ et donc orthogonal à $\hat{\varepsilon}_{(n,1)}$.

On peut donc appliquer le théorème de Pythagore au triangle rectangle formé par les trois vecteurs apparaissant dans l'égalité précédente. On aura donc :

$$\|y - \bar{y} \cdot e_n\|^2 = \|\hat{y} - \bar{y} \cdot e_n\|^2 + \|\hat{\varepsilon}\|^2$$

soit :

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (\hat{\varepsilon}_i)^2 \quad (5)$$

$\sum_{i=1}^n (y_i - \bar{y})^2$ est égal, au facteur $\frac{1}{n}$ près, à la variance empirique de la variable y .

Comme d'après le second résultat préliminaire $\bar{y} = \bar{\hat{y}}$, $\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ est égal, au facteur $\frac{1}{n}$ près, à la variance empirique de la variable \hat{y} .

De la même façon, comme d'après le premier résultat préliminaire $\bar{\hat{\varepsilon}} = 0$, la quantité $\sum_{i=1}^n (\hat{\varepsilon}_i)^2$ est égale, au facteur $\frac{1}{n}$ près, à la variance empirique de la variable $\hat{\varepsilon}$.

L'équation (5) est l'équation d'analyse de la variance qui s'interprète de la façon suivante :

Variance de y = Variance de \hat{y} + Variance de $\hat{\varepsilon}$

ou encore :

Variance de y = Variance expliquée par le modèle + Variance résiduelle

Le coefficient de détermination

Un modèle est une représentation simplifiée des déterminants du phénomène économique étudié. Il est nécessairement réducteur. Le coefficient de détermination permet d'évaluer la part de la variance de y expliquée par ce modèle. Cet indicateur, noté R^2 , est défini de la façon suivante :

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^n (\hat{\varepsilon}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (6)$$

Calculé après chaque régression, le R^2 permet, dans certaines limites, d'apprécier la qualité du résultat obtenu. Il prend ses valeurs dans l'intervalle $[0,1]$. Au pire, le modèle n'explique rien. Au mieux il explique 100% de la variance de la variable dépendante.

Remarques critiques :

1 – Le R^2 est aisément manipulable. On peut l'améliorer ou le détériorer en modifiant la forme fonctionnelle dans laquelle la variable expliquée est spécifiée (niveaux, logarithmes, taux de croissance, ...).

2 – Si le modèle ne comporte pas de terme constant, l'équation d'analyse de la variance (5) n'est plus vérifiée, en général. On ne peut donc plus décomposer la variance de la variable expliquée y sous une forme intéressante pour l'interprétation.

3 – Le coefficient de détermination augmente mécaniquement quand on ajoute une variable explicative, même si cette dernière n'a pas beaucoup de rapport avec y . A la limite, quand le nombre de variables explicatives est égal au nombre d'observations moins une ($k = n$), on obtient un $R^2 = 1$, si l'hypothèse H_3 est vérifiée.

Dans le but de corriger cet effet mécanique, on construit une statistique alternative, appelée R^2 ajusté ou R^2 corrigé, noté \bar{R}^2 , et défini par :

$$\bar{R}^2 = 1 - \frac{\frac{\sum_{i=1}^n (\hat{\varepsilon}_i)^2}{n-k}}{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}} \quad (7)$$

Lorsqu'on ajoute une variable explicative, $\sum_{i=1}^n (\hat{\varepsilon}_i)^2$ diminue, mais $(n - k)$ aussi, puisque k augmente. On n'obtient donc pas d'augmentation systématique du \bar{R}^2 comme pour le R^2 .

Au total, on constate le caractère peu satisfaisant d'une statistique comme le R^2 , si elle est utilisée de façon brute, sans être complétée par une approche en termes de tests. Par exemple, on peut examiner la variation du R^2 découlant de l'ajout ou de la suppression d'une ou plusieurs variables explicatives. La démarche est alors pertinente sur un plan scientifique car on teste la significativité de cette variation.

Propriétés statistiques des MCO

Reconsidérons le modèle linéaire multiple (2) :

$$y_{(n,1)} = X_{(n,k)} \cdot \beta_{(k,1)} + \varepsilon_{(n,1)}$$

On a déjà souligné que l'estimateur $\hat{\beta}$ est un vecteur aléatoire, puisqu'il est défini comme une fonction du vecteur aléatoire y . Il semble donc logique de s'intéresser à ses propriétés statistiques. Au cours des calculs qui suivront, l'hypothèse H_3 sera toujours supposée vérifiée car elle permet de définir $\hat{\beta}$ de manière unique. Par contre, on n'adopte, pour l'instant, aucune hypothèse sur la loi suivie par la perturbation ε .

Propriétés désirées d'un estimateur

Les estimateurs sont des variables aléatoires qui possèdent des distributions d'échantillonnage.

Pour pouvoir utiliser un estimateur $\hat{\theta}$ pour estimer un certain paramètre inconnu θ , on demande qu'il satisfasse les deux propriétés principales suivantes :

P1 - Absence de biais : $E(\hat{\theta}) = \theta$. L'espérance mathématique de l'estimateur doit être égale au paramètre qu'il est sensé estimer. Cela signifie qu'on ne fait pas d'erreur systématique relativement à la vraie valeur du paramètre à estimer.

P2 - Variance minimale : $\lim_{n \rightarrow \infty} Var(\hat{\theta}) = 0$. La dispersion autour de la valeur centrale de l'estimateur se réduit lorsque la taille de l'échantillon augmente.

On peut remarquer que ces deux propriétés sont des conditions suffisantes de convergence en probabilité.

Espérance de $\hat{\beta}$

On montre facilement le résultat suivant : Soit le modèle linéaire multiple (2). Sous les hypothèses H_1 , H_2 et H_3 l'estimateur des MCO $\hat{\beta}$ est défini de manière unique et est sans biais. On montre donc que :

$$E(\hat{\beta}) = \beta \quad (8)$$

L'explication en vidéo de l'estimateur sans biais



Variance de $\hat{\beta}$

Pour que la connaissance de $\beta_{(k,1)}$ apportée par l'estimation soit satisfaisante, il ne suffit pas que $\hat{\beta}_{(k,1)}$ soit sans biais, mais il faut aussi que la précision de l'estimateur soit satisfaisante, c'est-à-dire que sa variance soit la plus faible possible.

On montre que sous les hypothèses H_1 , H_2 , H_3 et H_4 :

$$\text{Var}_{(k,k)} \hat{\beta} = \sigma^2 \cdot (X'X)^{-1} \quad (9)$$



Remarques :

1 – Dans l'expression (8) de la matrice de variances-covariances du vecteur $\hat{\beta}$, on reconnaît σ^2 , variance de la perturbation, et la matrice $(X'X)$, que l'on peut interpréter comme la quantité d'information apportée par les variables explicatives.

2 – On peut donner à l'expression (8) la signification suivante : la connaissance de β apportée par $\hat{\beta}$ sera d'autant plus précise que la quantité d'information apportée par X sera élevée, relativement à la variance de la perturbation (c'est-à-dire de l'ensemble des facteurs explicatifs de y non retenus dans le modèle).

Cependant, dans l'expression précédente figure σ^2 , qui est en général un paramètre inconnu. Dans la pratique, $\text{Var}_{(k,k)} \hat{\beta}$ est donc inconnue et doit donc être estimée. On utilise pour cela l'estimateur $\hat{\sigma}^2$ défini par :

$$\hat{\sigma}^2 = \frac{SCR}{n - k} = \frac{\hat{\varepsilon}' \hat{\varepsilon}}{(1,n)(n,1)} = \frac{\sum_{i=1}^n \hat{\varepsilon}_i^2}{n - k}$$

où SCR représente la somme des carrés des résidus estimés.

On montre que l'estimateur $\hat{\sigma}^2$ définit ci-dessus est un estimateur sans biais de σ^2 .

Pour estimer la matrice de variances-covariances de $\hat{\beta}$, on utilisera alors l'estimateur :

$$\widehat{\text{Var}}_{(k,k)} \hat{\beta} = \hat{\sigma}^2 \cdot (X'X)^{-1} \quad (10)$$

L'explication en vidéo de l'estimateur de la variance de la perturbation



Théorème de Gauss-Markov

Un bon estimateur est un estimateur qui délivre, avec une probabilité élevée, des valeurs proches de la valeur que l'on cherche à estimer. Autrement dit, un bon estimateur est un estimateur dont la distribution d'échantillonnage est, d'une part, centrée sur la valeur qu'on cherche à estimer, et d'autre part, peu dispersée autour de cette valeur.

L'estimateur des MCO $\hat{\beta}$, étant non biaisé, il est bien centré sur la valeur inconnue du vecteur β . Par ailleurs, sa dispersion est donnée par sa matrice de variances-covariances, $\widehat{\text{Var}}_{(k,k)} \hat{\beta}$.

La question qui se pose est donc la suivante : Est-il possible de trouver un autre estimateur non biaisé de β , dont la dispersion serait plus faible que celle de l'estimateur des MCO, $\hat{\beta}$?

Le théorème de Gauss-Markov indique que non, à tout le moins si on se restreint à considérer la classe des estimateurs linéaires sans biais de β .

Ce théorème s'énonce de la façon suivante :

Soit le modèle de régression linéaire multiple (2). Sous les hypothèses H_1 , H_2 , H_3 et H_4 , l'estimateur des MCO $\hat{\beta}$ est l'estimateur le plus précis dans l'ensemble des estimateurs linéaires sans biais de β .

On dit que l'estimateur des MCO $\hat{\beta}$ est le meilleur estimateur linéaire sans biais de β (Best Linear Unbiased Estimator) ou encore qu'il est l'estimateur BLUE.

Explication en vidéo du théorème de Gauss-Markov [ici](#).



Il est important de remarquer que le théorème de Gauss-Markov assure que $\hat{\beta}$ est le meilleur estimateur (variance minimale) parmi seulement les estimateurs linéaires et non biaisés de β , pas parmi tous les estimateurs possibles. Cependant, si aux hypothèses H_1 , H_2 , H_3 et H_4 , on ajoute l'hypothèse optionnelle de normalité H_6 , on peut montrer que $\hat{\beta}$ est alors le meilleur estimateur (variance minimale) parmi tous les estimateurs non biaisés de β , qu'ils soient linéaires ou non.

Conséquences de l'hypothèse de normalité des perturbations

Sur le modèle linéaire multiple (2), on adopte non seulement les hypothèses H_1 , H_2 , H_3 et H_4 mais aussi l'hypothèse supplémentaire :

$$H_6 : \underset{(n,1)}{\varepsilon} \rightarrow N \left(\underset{(n,1)}{0}, \sigma^2 \cdot I_n \right)$$

Le modèle est à présent qualifié de modèle paramétrique car l'ensemble des hypothèses retenues permettent de spécifier exactement la loi générant les variables y_i . Puisque X est une matrice certaine, l'hypothèse H_6 implique en effet que :

$$y_{(n,1)} \rightarrow N \left(\begin{matrix} X \\ (n,k) \end{matrix} \beta_{(k,1)}, \sigma^2 \cdot I_n \right)$$

L'hypothèse de normalité des perturbations va aussi permettre de définir les lois de probabilité suivies par les variables aléatoires $\hat{\beta}$ et $\hat{\sigma}^2$. La connaissance de ces lois permettra dans le prochain chapitre d'effectuer des tests sur la valeur des coefficients composant le vecteur $\beta_{(k,1)}$.

Loi de $\hat{\beta}$

Puisque $\hat{\beta}_{(k,1)} = (X'X)^{-1} \cdot X'_{(k,n)} \cdot y_{(n,1)}$, $\hat{\beta}$ est par définition une combinaison linéaire certaine des y_i . Le vecteur $\hat{\beta}_{(k,1)}$ suit donc, comme le vecteur $y_{(n,1)}$, une loi normale, dont l'espérance et la matrice de variances-covariances sont fournies par les expressions (8) et (9) établies précédemment :

$$\hat{\beta}_{(k,1)} \rightarrow N \left(\begin{matrix} \beta \\ (k,1) \end{matrix}, \sigma^2 \cdot (X'X)^{-1}_{(k,k)} \right) \quad (11)$$

En particulier, pour une composante quelconque $\hat{\beta}_j$ du vecteur $\hat{\beta}_{(k,1)}$, on aura :

$$\hat{\beta}_j \rightarrow N(\beta_j, \sigma^2 \cdot x^{jj}) \quad (12)$$

où l'on désigne par x^{jj} le j ième élément de la diagonale principale de la matrice $(X'X)^{-1}_{(k,k)}$.

Loi de $\hat{\sigma}^2$

Sous les hypothèses retenues, on peut établir le résultat suivant :

$$(n - k) \cdot \frac{\hat{\sigma}^2}{\sigma^2} \rightarrow \chi_{n-k}^2 \quad (12)$$

Compte tenu de la définition de $\hat{\sigma}^2$ donnée au point 3 de la précédente section, ce résultat s'écrit de la manière équivalente suivante :

$$\frac{SCR}{\sigma^2} \rightarrow \chi_{n-k}^2$$

Preuve : L'hypothèse H_0 peut s'écrire sous la forme :

$$\frac{\varepsilon}{\sigma} \rightarrow N \left(\begin{matrix} 0 \\ (n,1) \end{matrix}, I_n \right)$$

On a vu que $\hat{\varepsilon} = M_X \cdot \varepsilon$ où $M_X = I_n - X \cdot (X'X)^{-1} \cdot X'$ est la matrice de projection orthogonale sur $\mathcal{L}^\perp(X)$. On sait que $\text{rang } M_X_{(n,n)} = n - k$.

En appliquant un théorème relatif aux formes quadratiques de vecteurs gaussiens², on déduit que :

$$\frac{\varepsilon' M_X \varepsilon}{\sigma^2} \rightarrow \chi_{n-k}^2$$

On sait que $(n - k) \cdot \frac{\hat{\sigma}^2}{\sigma^2} = \frac{SCR}{\sigma^2} = \frac{\hat{\varepsilon}' \hat{\varepsilon}}{\sigma^2} = \frac{(M_X \cdot \varepsilon)' (M_X \cdot \varepsilon)}{\sigma^2} = \frac{\varepsilon' M_X' M_X \cdot \varepsilon}{\sigma^2}$

M_X est une matrice symétrique ($M_X' = M_X$) et idempotente ($M_X^2 = M_X$), on déduit immédiatement le résultat (12), à savoir :

$$(n - k) \cdot \frac{\hat{\sigma}^2}{\sigma^2} = \frac{\varepsilon' M_X \varepsilon}{\sigma^2} \rightarrow \chi_{n-k}^2$$

² Théorème : Soit y un vecteur aléatoire de \mathbb{R}^n , tel que $y \rightarrow N(0, I_n)$ et $X_{(n,k)}$ une matrice certaine de rang k . Soit $P_X = X \cdot (X'X)^{-1} \cdot X'$ la matrice de projection orthogonale sur l'espace $\mathcal{L}(X)$ engendré par les vecteurs de X . Soit $M_X = I_n - X \cdot (X'X)^{-1} \cdot X'$ la matrice de projection orthogonale sur $\mathcal{L}^\perp(X)$. Les variables $y'P_X y$ et $y'M_X y$ sont indépendantes et ont pour lois :

$$y'P_X y \rightarrow \chi_k^2 \quad \text{et} \quad y'M_X y \rightarrow \chi_{n-k}^2$$

Références

Comment citer ce cours ?

Introduction à l'économétrie, Olivier Baron, AUNEGe (<http://aunega.fr>), CC – BY NC ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



Cette œuvre est mise à disposition dans le respect de la législation française protégeant le droit d'auteur, selon les termes du contrat de licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International (<http://creativecommons.org/licenses/by-nc-nd/4.0/>). En cas de conflit entre la législation française et les termes de ce contrat de licence, la clause non conforme à la législation française est réputée non écrite. Si la clause constitue un élément déterminant de l'engagement des parties ou de l'une d'elles, sa nullité emporte celle du contrat de licence tout entier.

Table des illustrations

Figure 1 : approche géométrique - Courbe d'illustration des vecteurs 12