

RÉGRESSION – APPROCHE MATRICIELLE

Source : D.N.Gujarati, "Basic Econometrics", Third Ed., McGraw Hill, 1995

1. Le modèle de base

On suppose le modèle suivant :

$$y_i = b_1 + b_2 x_{2i} + \dots + b_k x_{ki} + u_i$$

dont la représentation matricielle est la suivante :

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{21} & \dots & x_{k1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{2n} & \dots & x_{kn} \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_k \end{bmatrix} + \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$$

que l'on notera par la suite : $Y_{n1} = X_{nk} b_{k1} + U_{n1}$

2. Les hypothèses

- $E(U) = 0$: l'espérance des résidus est égale à 0
- $E(UU') = \sigma^2 I$: les résidus sont IID et de variance constante
- soit X est non-stochastique, soit $E(X'U) = 0$ (les variables indépendantes sont stochastiques mais non-corrélées avec les résidus)
- $r(X) = k$: aucune variable indépendante n'est une combinaison linéaire des autres variables indépendantes (pas de multicollinéarité parfaite)
- $U \approx N(0, \sigma^2 I)$: les résidus suivent une distribution normale (nécessaire pour les tests d'hypothèse sur les coefficients)

3. L'estimation par OLS

Note : $\hat{}$ est utilisé pour les estimés

$$Y = X\hat{b} + \hat{U}$$

$$\Rightarrow \hat{U} = Y - X\hat{b}$$

$$\hat{U}'\hat{U} = (Y - X\hat{b})'(Y - X\hat{b}) = Y'Y + \hat{b}'X'X\hat{b} - 2\hat{b}'X'Y$$

où $\hat{U}'\hat{U}$ est la somme des carrés des résidus.

On notera aussi que $\hat{b}'_{1k} X'_{kn} Y_{n1}$ est un scalaire et est donc égal à sa transposée ($Y' X \hat{b}$).

$$\text{Min } \hat{U}'\hat{U} = \text{Min}_{\hat{b}} Y'Y + \hat{b}'X'X\hat{b} - 2\hat{b}'X'Y$$

On cherche donc :

et on obtient :

$$\frac{\partial(\hat{U}'\hat{U})}{\partial \hat{b}} = -2X'Y + 2X'X\hat{b} = 0$$

$$X'X\hat{b} = X'Y$$

$$\hat{b} = (X'X)^{-1}X'Y$$

4. Matrice variance/covariance des estimateurs :

$$\hat{b} = (X'X)^{-1}X'Y$$

Or $Y = Xb + U$

$$\Rightarrow \hat{b} = (X'X)^{-1}X'(Xb + U)$$

$$= (X'X)^{-1}X'Xb + (X'X)^{-1}X'U$$

$$= b + (X'X)^{-1}X'U$$

$$\Rightarrow \hat{b} - b = (X'X)^{-1}X'U$$

$$\text{Or } \text{VarCo var}(\hat{b}) = E[(\hat{b} - b)(\hat{b} - b)']$$

$$\Rightarrow \text{VarCo var}(\hat{b}) = E[(X'X)^{-1}X'U][(X'X)^{-1}X'U]'$$

$$= E[(X'X)^{-1}X'UU'X(X'X)^{-1}]$$

Comme X est non stochastique :

$$= (X'X)^{-1}X'E(UU')X(X'X)^{-1}$$

$$= (X'X)^{-1}X's^2IX(X'X)^{-1}$$

$$= s^2I(X'X)^{-1}(X'X)(X'X)^{-1}$$

$$= s^2I(X'X)^{-1}$$

$$\Rightarrow \text{Var}(\hat{b}_i) = s^2[(X'X)^{-1}]_{i,i}$$

5. BLUE (Best Linear Unbiased Estimator) des estimateurs OLS

$$\hat{b} = (X'X)^{-1}X'Y$$

- estimateur linéaire : $(X'X)^{-1}X'$ est matrice de nombre et donc, \hat{b} est une combinaison linéaire des Y

- estimateur non-biaisé :

$$\begin{aligned}\hat{\mathbf{b}} &= (X'X)^{-1}X'Y \text{ et } Y = X\mathbf{b} + U \\ \Rightarrow \hat{\mathbf{b}} &= (X'X)^{-1}X'(X\mathbf{b} + U) \\ &= \mathbf{b} + (X'X)^{-1}X'U \\ E(\hat{\mathbf{b}}) &= E(\mathbf{b}) + (X'X)^{-1}X'E(U) \\ \text{Or } E(U) &= 0 \\ \Rightarrow E(\hat{\mathbf{b}}) &= E(\mathbf{b}) = \mathbf{b}\end{aligned}$$

- meilleur estimateur :
Supposons \mathbf{b}^* , un autre estimateur de \mathbf{b}
tel que $\mathbf{b}^* = [(X'X)^{-1}X' + C]Y$
où C est une matrice de constantes.

On a donc :

$$\begin{aligned}\mathbf{b}^* &= [(X'X)^{-1}X' + C](X\mathbf{b} + U) \\ &= \mathbf{b} + CX\mathbf{b} + (X'X)^{-1}X'U + CU\end{aligned}$$

Si \mathbf{b}^* est sans biais :

$$\begin{aligned}E(\mathbf{b}^*) &= E\left[\left((X'X)^{-1}X' + C\right)(X\mathbf{b} + U)\right] \\ &= E\left[\left((X'X)^{-1}X'X\mathbf{b} + ((X'X)^{-1}X'U) + CX\mathbf{b} + CU\right)\right] \\ &= E(\mathbf{b}) + (X'X)^{-1}X'E(U) + CXE(\mathbf{b}) + CE(U) \\ &= \mathbf{b} + CX\mathbf{b}\end{aligned}$$

Pour que l'estimateur \mathbf{b}^* soit sans biais, il faut que CX soit égal à 0.

On obtient donc :

$$\begin{aligned}\Rightarrow \mathbf{b}^* - \mathbf{b} &= (X'X)^{-1}X'U + CU \\ \text{VarCo var}(\mathbf{b}^*) &= E[(\mathbf{b}^* - \mathbf{b})(\mathbf{b}^* - \mathbf{b})'] \\ &= E\left[\left[(X'X)^{-1}X'U + CU\right]\left[(X'X)^{-1}X'U + CU\right]'\right] \\ &= E\left[(X'X)^{-1}X'UU'X(X'X)^{-1} + (X'X)^{-1}X'UU'C' + CUU'X(X'X)^{-1} + CUU'C'\right]\end{aligned}$$

Or $E(UU') = \mathbf{s}^2I$. On a donc :

$$\text{VarCo var}(\mathbf{b}^*) = \mathbf{s}^2(X'X)^{-1} + \mathbf{s}^2(X'X)^{-1}X'C' + \mathbf{s}^2CX(X'X)^{-1} + \mathbf{s}^2CC'$$

Comme $CX = 0$, on obtient :

$$\text{VarCo var}(\mathbf{b}^*) = \mathbf{s}^2(X'X)^{-1} + \mathbf{s}^2CC'$$

Pour être de variance minimum, $C'C$ doit être égal à 0, ce qui revient à affirmer que $\mathbf{b}^* = \hat{\mathbf{b}}$.

6. Propriétés des estimateurs OLS

- Concernant la variable dépendant, la valeur moyenne des estimés est égale à la valeur moyenne des réalisés :

Si $Y_i = \mathbf{b}_1 + \mathbf{b}_2 X_{2i} + \mathbf{b}_3 X_{3i} + u_i$, on a : $\hat{\mathbf{b}}_1 = \bar{Y} - \mathbf{b}_2 \bar{X}_2 - \mathbf{b}_3 \bar{X}_3$ et on obtient donc le résultat suivant :

$$\begin{aligned}\hat{Y}_i &= \hat{\mathbf{b}}_1 + \hat{\mathbf{b}}_2 X_{2i} + \hat{\mathbf{b}}_3 X_{3i} \\ \hat{Y}_i &= (\bar{Y} - \hat{\mathbf{b}}_2 \bar{X}_2 - \hat{\mathbf{b}}_3 \bar{X}_3) + \hat{\mathbf{b}}_2 X_{2i} + \hat{\mathbf{b}}_3 X_{3i} \\ \hat{Y}_i &= \bar{Y} + \hat{\mathbf{b}}_2 (X_{2i} - \bar{X}_2) + \hat{\mathbf{b}}_3 (X_{3i} - \bar{X}_3) \\ \hat{Y}_i &= \bar{Y} + \hat{\mathbf{b}}_2 x_{2i} + \hat{\mathbf{b}}_3 x_{3i}\end{aligned}$$

où les minuscules expriment des déviations par rapport à la moyenne. En sommant la dernière équation sur l'échantillon et en divisant par n (le nombre d'observations), on constate que la moyenne des estimés est bien égale à la moyenne des réalisés (on notera que $\sum x_{2i} = 0$ par définition).

- Les résidus sont non-corrélés avec les variables indépendantes : $\sum \hat{u}_i X_{2i} = 0$, ce qui est une conséquence directement de la minimisation du carré des erreurs :

$$\text{Si } \sum \hat{u}_i^2 = \sum (Y_i - \hat{\mathbf{b}}_1 - \hat{\mathbf{b}}_2 X_{2i} - \hat{\mathbf{b}}_3 X_{3i})^2,$$

on obtient les paramètres de la droite de régression par :

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\mathbf{b}}_1} = 2 \sum (Y_i - \hat{\mathbf{b}}_1 - \hat{\mathbf{b}}_2 X_{2i} - \hat{\mathbf{b}}_3 X_{3i})(-1) = 0$$

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\mathbf{b}}_2} = 2 \sum (Y_i - \hat{\mathbf{b}}_1 - \hat{\mathbf{b}}_2 X_{2i} - \hat{\mathbf{b}}_3 X_{3i})(-X_{2i}) = 0$$

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\mathbf{b}}_3} = 2 \sum (Y_i - \hat{\mathbf{b}}_1 - \hat{\mathbf{b}}_2 X_{2i} - \hat{\mathbf{b}}_3 X_{3i})(-X_{3i}) = 0$$

On peut réécrire les trois équations de la manière suivante :

$$\begin{aligned}\sum \hat{u}_i &= 0 \\ \sum \hat{u}_i X_{2i} &= 0 \\ \sum \hat{u}_i X_{3i} &= 0\end{aligned}$$

- Les résidus sont non-corrélés avec les estimés de la variable dépendante. Sachant en effet que $\hat{Y}_i = \bar{Y} + \hat{\mathbf{b}}_2 x_{2i} + \hat{\mathbf{b}}_3 x_{3i}$, on peut affirmer que $\hat{y}_i = \hat{\mathbf{b}}_2 x_{2i} + \hat{\mathbf{b}}_3 x_{3i}$. En multipliant par \hat{u}_i à gauche et à droite et en sommant sur l'échantillon, on obtient sur $\sum \hat{u}_i \hat{Y}_i = 0$.

7. Tests d'hypothèses et statistiques classiques

Coefficient de détermination : il s'agit du rapport entre la somme des carrés expliquées par le modèle (ESS pour Explained Sum of Squares) sur la somme total des carrés (TSS) :

$$R^2 = \frac{ESS}{TSS} = \frac{\hat{\mathbf{b}}\mathbf{X}'\mathbf{Y} - n\bar{Y}^2}{\mathbf{Y}'\mathbf{Y} - n\bar{Y}^2}$$

Il s'interprète comme le % de la variation de la variable dépendante qui est expliquée par le modèle de régression.

Tests d'hypothèses sur les paramètres : Si les $U \approx N(0, \mathbf{S}^2 I)$, alors :

$\Rightarrow \hat{\mathbf{b}} \approx N(\mathbf{b}, \mathbf{S}^2 (\mathbf{X}'\mathbf{X})^{-1})$ et l'on construit le test de student de la manière suivante :

$$\Rightarrow t = \frac{\hat{b} - b}{SE(\hat{b})} \text{ avec } n \text{ (nombre d'observations)} - k \text{ (nombre de variables indépendantes + 1 pour}$$

la constante) degrés de liberté.

Test d'hypothèse sur le modèle : la degré de significativité global du modèle peut être testé à l'aide d'une analyse de la variance (ANOVA), ce qui revient à tester l'hypothèse nulle selon laquelle les coefficients du modèle sont simultanément égaux à 0. On obtient le test suivant :

$$F = \frac{(\hat{\mathbf{b}}\mathbf{X}'\mathbf{Y} - n\bar{Y}^2) / (k-1)}{(\mathbf{Y}'\mathbf{Y} - \hat{\mathbf{b}}\mathbf{X}'\mathbf{Y}) / (n-k)}$$

Le tableau ANOVA correspondant est :

Source des variations	Somme de carrés	Degrés de liberté	Variance
Modèle	$\hat{\mathbf{b}}\mathbf{X}'\mathbf{Y} - n\bar{Y}^2$	k-1	$\hat{\mathbf{b}}\mathbf{X}'\mathbf{Y} - n\bar{Y}^2 / k-1$
Résidus	$\mathbf{Y}'\mathbf{Y} - \hat{\mathbf{b}}\mathbf{X}'\mathbf{Y}$	n-k	$\mathbf{Y}'\mathbf{Y} - \hat{\mathbf{b}}\mathbf{X}'\mathbf{Y} / n-k$
Total	$\mathbf{Y}'\mathbf{Y} - n\bar{Y}^2$		

Intervalles de confiance autour des prévisions

On peut, à l'aide d'un modèle de régression donné, chercher soit à prévoir l'espérance de la valeur de la variable dépendante (prévision en moyenne), soit à prévoir une valeur spécifique de Y (prévision ponctuelle). Cela conduit à construire deux types d'intervalle de confiance différents :

- intervalle de confiance autour d'une prévision en moyenne :

La variance de l'estimation est : $\text{var}(\hat{Y}|x) = \hat{\mathbf{S}}^2 x'(X'X)^{-1}x$ où $\hat{\mathbf{S}}$ est l'estimation de l'écart-type des résidus sur les données et x' est le vecteur des réalisations pour les variables indépendantes. Cela conduit à un intervalle de confiance de la forme :

$$\hat{Y} \pm t_{\alpha/2} \sqrt{\hat{\mathbf{S}}^2 x'(X'X)^{-1}x} \text{ avec } n-(k+1) \text{ degrés de liberté.}$$

○ intervalle de confiance autour d'une prévision ponctuelle :

La variance de l'estimation est : $\text{var}(\hat{Y}|x) = \hat{\mathbf{S}}^2 [1 + x'(X'X)^{-1}x]$ et l'intervalle de

confiance à la forme : $\hat{Y} \pm t_{\alpha/2} \sqrt{\hat{\mathbf{S}}^2 [1 + x'(X'X)^{-1}x]}$, toujours avec $n-(k+1)$ degrés de liberté.