URCA Hugo Harari-Kermadec

2008-2009 harari@ecogest.ens-cachan.fr

Économétrie 1 : Modèle linéaire

I Outils d'algèbre linéaire

I.1 Projections

Soit X une matrice à n lignes et p colonnes avec $p \leq n$. On note classiquement $X \in \mathcal{M}_{np}(\mathbb{R})$. $\mathbb{R}$ est sous-entendu dans la suite.

Proposition I.1 (Projection) *Si X est de rang p alors*

1. *la projection sur Vect(X) a pour matrice*

$$P_X = X(X'X)^{-1}X'$$

appartient à $\mathcal{M}_{nn}$,

2. *la projection sur l'orthogonal de Vect(X) est*

$$P_{X^\perp} = I_n - P_X = I_n - X(X'X)^{-1}X',$$

3. *le rang de P_X est égal à la dimension de Vect(X) et est donné par la trace de P_X,*
4. *P_X est une matrice de projection donc*

$$P_X^2 = P_X \qquad\qquad P_X' = P_X$$

I.2 Dérivation matricielle

Soit $A \in \mathcal{M}_{np}$ et $X \in \mathcal{M}_{p1}$ alors

$$\frac{\partial}{\partial X} AX = A'.$$

Si A est symétrique (et donc $n = p$),

$$\frac{\partial}{\partial X} X'AX = 2AX.$$

II Modèle et estimation

Le modèle linéaire pour explique Y à l'aide de K variables explicatives et d'une constante s'écrit pour une observation i :

$$Y_i = \theta_{K+1} + \sum_{k=1}^{K} \theta_k x_{i,k} + \varepsilon_i,$$

c'est-à-dire

$$Y_i = X_i\theta + \varepsilon_i,$$

avec $X_i = (x_{i,1}, \ldots, x_{i,K}, 1) \in \mathcal{M}_{1,K+1}$ et $\theta = (\theta_1, \ldots, \theta_K, \theta_{K+1})' \in \mathbb{R}^{K+1}$.
On a alors pour n observations :

$$Y = X\theta + \mathcal{E},$$

avec $Y = (Y_1, \ldots, Y_n)' \in \mathbb{R}^n$, $\mathcal{E} = (\varepsilon_1, \ldots, \varepsilon_n)' \in \mathbb{R}^n$ et $X = (X_1, \ldots, X_n)' \in \mathcal{M}_{n,K+1}$.

II.1 Estimation

On va utiliser les hypothèses suivantes :
H_1 : les colonnes de X sont linéairement indépendantes,
H_2 : $\mathbb{E}[\mathcal{E}|X] = 0 \in \mathbb{R}^n$.
H_3 : $\mathrm{Var}(\mathcal{E}) = \sigma^2 I_n$.

Définition II.1 (Estimateur des moindres carrés) *L'estimateur des moindres carrés est défini par*

$$\hat{\theta} = arginf_{\theta \in \mathbb{R}^{K+1}} ||Y - X\theta||^2.$$

Proposition II.1 *Sous l'hypothèse H_1 il existe un unique estimateur des moindres carrés :*

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

Preuve II.1 *On pose $L(\theta) = ||Y - X\theta||^2 = (Y - X\theta)'(Y - X\theta)$ et on annule la dérivée :*

$$\begin{aligned} \frac{\partial}{\partial \theta} L(\theta)|_{\hat{\theta}} &= -2X'(Y - X\hat{\theta}) = 0 \\ 2X'Y &= 2X'X\hat{\theta} \\ (X'X)^{-1}X'Y &= \hat{\theta} \end{aligned}$$

Définition II.2 (Prédictions associées) *On pose $\widehat{Y} = X\hat{\theta}$ et $\widehat{\mathcal{E}} = Y - \widehat{Y}$.*

Proposition II.2 *Sous H_1,*

1. $\widehat{Y} = P_X Y$,
2. $X'\widehat{\mathcal{E}} = 0$,
3. $\frac{1}{n}\sum_i \hat{\varepsilon}_i = 0$,
4. $\widehat{Y}'\widehat{\mathcal{E}} = 0$

Proposition II.3 *Sous H_1 et H_2, $\hat{\theta}$ est sans biais :*

$$\mathbb{E}[\hat{\theta}] = \theta_0,$$

où θ_0 est la vraie valeur du vecteur de paramètres θ.
Sous H_1, H_2 et H_3,

$$\widehat{\sigma^2} = \frac{\widehat{\mathcal{E}}'\widehat{\mathcal{E}}}{n - K - 1}$$

est un estimateur sans biais de σ^2,

Théorème II.1 (Gauss-Markov) *Sous H_1, H_2 et H_3, $\hat{\theta}$ est l'estimateur linéaire sans biais optimal de θ. Sa variance vaut $Var(\hat{\theta}) = \sigma^2(X'X)^{-1}$.*

III Tests d'utilité des régresseurs

Dans cette section, on fait l'hypothèse suivante :
H_G : La loi de $\mathcal{E}$ sachant X est normale, centrée, de variance $\sigma^2 I_n$:

$$\mathcal{E}|X \sim \mathcal{N}(0, \sigma^2 I_n).$$

Cette hypothèse entraine H_2 et H_3.

Proposition III.1 *Sous H_1 et H_G,*

1. $\hat{\theta} \sim \mathcal{N}(\theta_0, \sigma^2(X'X)^{-1})$,
2. $\widehat{\sigma^2}$ *et* $\hat{\theta}$ *sont indépendants,*
3. $\widehat{\sigma^2}\frac{n-K-1}{\sigma^2} \sim \chi^2_{n-K-1}$,
4. *Pour chaque coordonnée $k \in [|1, K+1|]$,*

$$\frac{\hat{\theta}_k - \theta_k}{\hat{\sigma}_k} \sim \mathcal{T}(n - K - 1), \text{ où } \hat{\sigma}_k = \sqrt{\widehat{\sigma^2}(X'X)^{-1}_{kk}}.$$

On peut alors tester la pertinence d'une variable explicative : si l'hypothèse "$\theta_k = 0$" est rejetée, c'est que la k-ième variable est utile.

Proposition III.2 (Test de nullité d'un coefficient) *Sous H_1 et H_G, le test de "$\theta_k = 0$" contre "$\theta_k \neq 0$" de niveau α a pour zone critique :*

$$W_n = \left\{ |\hat{\theta}_k| \geq \hat{\sigma}_k q_{n-K-1}(1 - \frac{\alpha}{2}) \right\}$$

où $q_{n-K-1}(1 - \frac{\alpha}{2}) =$ est le quantile de niveau $1 - \frac{\alpha}{2}$ d'une loi de Student à $n - K - 1$ degrés de liberté.

Remarque III.1 (p-value) *L'interprétation de la p-value est très importante dans ce contexte et parfois un peu déroutante. On rappelle que la p-value est l'erreur de première espèce limite.*
Si la p-value du test est par exemple de 37%, ca veut dire qu'avec une erreur de première espèce même grande (jusqu'à 37% donc), on accepte l'hypothèse nulle. Donc on est vraiment très convaincu que le coefficient soit nul, c'est-à-dire que la variable n'ait pas d'effet (linéaire) sur Y.
Si la p-value vaut 10%, on accepte donc à 90% (et a fortiori à 95%) que le coefficient soit nul, c'est le cas où l'interprétation est un peu ambivalente. Enfin si la p-value est très petite (disons inférieure à 1%) alors on est pratiquement sûr qu'il faut rejeter l'hypothèse nulle : la variable est utile pour expliquer (linéairement) Y.

IV Prévisions

Définition IV.1 (Prévisions) *Pour une nouvelle valeur du vecteur des variables explicatives X_{n+1}, on prédit $\widehat{Y} = X_{n+1}\hat{\theta}$.*

Proposition IV.1 *Sous H_1 et H_G, $\widehat{Y}$ est un estimateur sans biais de Y_{n+1} et*

$$\widehat{Y} \sim \mathcal{N}\left(X_{n+1}\theta_0, \sigma^2 X_{n+1}(X'X)^{-1}X'_{n+1}\right).$$

V Analyse de la variance

Pour un vecteur $Z \in \mathbb{R}^n$, on pose $\overline{Z} = \frac{1}{n}\sum_{i=1}^n Z_i$ et $S_n^2(Z) = \sum_{i=1}^n (Z_i - \overline{Z})^2$.

Théorème V.1 (Décomposition de la variance) *Sous H_1,*

$$S_n^2(Y) = S_n^2(\widehat{Y}) + S_n^2(\widehat{\mathcal{E}})$$

Preuve V.1 *On a d'après la proposition II.1, $Y = P_X Y + P_{X^\perp} Y = \widehat{Y} + \widehat{\mathcal{E}}$. De plus, comme la dernière colonne de X est composée de 1, $\overline{Y} = (1, \ldots, 1) \cdot Y$ appartient à Vect(X). Donc $Y - \overline{Y}$ se décompose en la somme orthogonale de $\widehat{Y} - \overline{Y}$ et $\widehat{\mathcal{E}}$. On applique ensuite le théorème de Pythagore :*

$$||Y - \overline{Y}||^2 = ||\widehat{Y} - \overline{Y}||^2 + ||\widehat{\mathcal{E}}||^2.$$

On note ces sommes respectivement
SCT : Somme des Carrés Totale
SCE : Somme des carrés Expliquée
SCR : Somme des Carrés des Résidus.
On a donc

$$\boxed{SCT = SCE + SCR}$$

.

Définition V.1 (Qualité de l'ajustement) *On définit alors le R^2 qui mesure la qualité de l'ajustement :*

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} = 1 - \frac{(n-K-1)\widehat{\sigma^2}}{SCT}$$

Il est trivial que le R^2 appartient à $[0,1]$. Plus le R^2 est proche de 1, plus l'ajustement est bon. S'il vaut 1, c'est que toutes les valeurs Y_i sont parfaitement prédites par $\widehat{Y}_i = X_i\hat{\theta}$.
Il faut néanmoins bien se garder d'interpréter trop hâtivement le R^2 : un R^2 de 0.3 peut être très bon si on a peu de variables explicatives et que Y a une grande variance. Il permet par contre de comparer deux modèles à conditions qu'ils aient le même nombre de variables explicatives. Si le nombre de variables explicative (K) change, il faut introduire le R^2 ajusté.

Définition V.2 (R^2 **ajusté**)

$$R^2_{aj} = 1 - \frac{SCR/n-K-1}{SCT/n-1} = 1 - (1-R^2)\frac{n-1}{n-K-1}.$$

Le R^2_{aj} est à nouveau entre 0 et 1. Il permet de comparer des modèles linéaires entre eux, même si les nombres de variables explicatives diffèrent. En particulier, il permet de dire s'il faut retirer une variable explicative : si le R^2_{aj} diminue lorsqu'on retire une variable explicative, il faut la remettre.