ESTIMATEUR DE HOERL-KENNARD (H, J1)
(12 / 04 / 2020, © Monfort, Dicostat2005, 2005-2020)

En présence de **colinéarité** approchée, l'**estimateur « raccourci » de HOERL - KENNARD** constitue une méthode d'estimation du paramètre d'un **modèle de régression linéaire** : cette méthode recherche un compromis entre le **biais** et la **variance** de l'**estimateur des mco** de ce paramètre.

(i) Soit, dans un **espace de variables** :

(1) $\eta = \xi' b + \varepsilon$, avec $E\,\varepsilon = 0$, $V\,\varepsilon = \sigma^2$,

un **modèle de régression multiple** linéaire, « observé » (ie appliqué aux observations (X, y) d'un **espace d'observation**) selon :

(2) $y = X b + u$, avec $E\,u = 0$, $V\,u = \sigma^2 . I_N$.

On suppose que rg X = K mais que $X \in M_{NK}(\mathbf{R})$ est une **matrice** quasi-colinéaire (ou « mal conditionnée »), ie que Dét (X' X) n'est pas nul mais quasi-nul (ie $|\text{Dét}(X' X)| << + \infty$) (cf **déterminant**, **rang**, **matrice singulière**, **quasi-colinéarité**).

L'estimateur des mco $b^\wedge$ de b, dont la **matrice de dispersion** est $V\, b^\wedge = \sigma^2 (X' X)^{-1}$, possède un **écart quadratique moyen** (scalaire) tq :

(3) $E\, ||b^\wedge - b||^2 = \sigma^2 \text{ tr } (X' X)^{-1} = \sigma^2\, \Sigma_k\, \lambda_k^{-1}$,

où les λ_k sont les **valeurs propres** de X' X.

Par suite :

(4) $E\, ||b^\wedge - b||^2 \geq \sigma^2 \{\min_{k=1}^K \lambda_k\}^{-1}$, ou $E\, ||b^\wedge||^2 \geq ||b||^2 + (\min_{k=1}^K \lambda_k)^{-1} \sigma^2$.

Si $\min_{k=1}^K \lambda_k$ est « négligeable » (ie $\min_{k=1}^K \lambda_k << + \infty$), cet écart quadratique moyen peut prendre des valaurs « importantes » (ie $E\, ||b^\wedge - b||^2 >> 0$). Autrement dit, l'une (au moins) des coordonnées $b_k^\wedge$ de $b^\wedge$ peut être éloignée de la coordonnée correspondante b_k de b.

(ii) En pratique, si X est une matrice « contrôlée » (eg dans le cadre d'un **plan d'expérience**), il est souvent possible de redéfinir le plan et de recommencer l'expérience (supposée renouvelable) pour corriger le défaut précédent (cf **renouvellement**, **plan répété**).

Dans le cas contraire, on peut décider qu'un estimateur biaisé de b est préférable à $b^\wedge$ s'il n'est pas trop sensible à la quasi-colinéarité de X, ou à la quasi-singularité de X' X, ie à la quasi-nullité de Dét (X' X) (cf **estimateur sans biais**).

Par suite, **« raccourcir » la régression** (2) consiste à remplacer $b^{\wedge}$ par une **famille** $(b_{\alpha}^{\#})_{\alpha \geq 0}$ d'estimateurs biaisés de b possédant cependant un écart quadratique moyen inférieur à celui de $b^{\wedge}$. Si X' X est quasi-singulière, alors $X' X + \alpha I_K$ sera « moins singulière » et l'**estimateur « raccourci » de A.E. HOERL - R.W. KENNARD** de b est défini en remplaçant l'**équation normale** usuelle :

(5) $X' X b = X' y$

par l'**équation « perturbée »** (ie « à matrice X' X perturbée ») suivante :

(6) $(X' X + \alpha I_K) b = X' y$,

d'où la famille d'estimateurs de b suivante :

(7) $b_{\alpha}^{\wedge} = (X' X + \alpha I_K)^{-1} X' y, \quad \forall \alpha \in \mathbf{R}_+$.

Ces **estimateurs raccourcis** $(b^{\wedge}_{\alpha})_{\alpha \geq 0}$ possèdent les propriétés suivantes :

(a) $b_0^{\wedge} = b^{\wedge}$ (estimateur des mco) ;

(b) $\forall \alpha > 0$, $b_{\alpha}^{\#}$ est biaisé et $E \, ||b_{\alpha}^{\wedge}|| < E \, ||b^{\wedge}||$;

(c) si $\alpha \to +\infty$, on a $V b_{\alpha}^{\wedge} \to 0$ mais $E b_{\alpha}^{\wedge} - b \to +\infty$. De plus, $b_{\alpha}^{\wedge} \to 0$.

(iii) Le problème consiste alors à choisir un **facteur de raccourcissement** α qui soit « optimal ». La **représentation grahique** de la fonction vectorielle $\alpha \mapsto b_{\alpha}^{\wedge}$ peut être utilisée pour déterminer une valeur α^* de α à partir de laquelle toutes les (fonctions) coordonnées $\alpha \mapsto b_{k,\alpha}$ de $\alpha \mapsto b_{\alpha}^{\wedge}$ commencent à se stabiliser.

On montre que la valeur optimale de α, ie celle qui minimise l'**écart quadratique moyen** $E \, ||b_{\alpha}^{\wedge} - b||^2$, est tq $(\alpha^*)^2 = K \, . \, \sigma^2 / ||b||^2$. Cette valeur peut être estimée par de l'estimateur suivant :

(8) $((\alpha^*)^2)^{\wedge} = K \, . \, (\sigma^2)^{\wedge} / ||b^{\wedge}||^2$,

où $(\sigma^2)^{\wedge} = ||y - X b^{\wedge}||^2 / (N - K) = ||u^{\wedge}||^2 / (N - K)$ est l'**estimateur des mco** usuel de σ^2, obtenu par la **méthode des mco**, et où $b^{\wedge}$ est l'estimateur des mco de b (estimateur qui peut éventuellement être remplacé par l'estimateur sur composantes principales de b) (cf **régression sur composantes principales**).

Si $K > 2$ et si $u \sim \mathcal{N}_N (0, \sigma^2 \, . \, I_N)$ (**loi normale** centrée), on montre que :

(9) $E \, (\alpha^*)^{\wedge} \in [0, K / (K - 2)]$.

Par ailleurs, on montre qu'il existe une valeur $\alpha^{\sim} \in \mathbf{R}_+$ tq le problème d'optimisation (programme mathématique) suivant :

(10) inf $||y - X \beta||^2$ sous $||\beta||^2 = r^2$ (donné)

admette pour solution l'estimateur $b_{\alpha\sim}\hat{}$.

(iv) Dans ce qui précède, on peut supposer que X est une **matrice d'observation** de K **variables exogènes** normalisées, ie que $X = P\ T\ S^{-1}$, où T est la matrice d'observation des variables initiales (ie non normalisées), P la **matrice de centrage par rapport à la moyenne** empirique et S la **matrice diagonale** dont chaque terme s_k est l'**écart-type** empirique (non corrigé), ie tq $s_k^2 = N^{-1} \Sigma_{n=1}^{N} (t_{nk} - \bar{t}_k)^2$, avec $\bar{t}_k = N^{-1} \Sigma_n t_{nk}$ (**moyenne empirique**), $\forall$ k $\in N_K$*.

Un **estimateur raccourci généralisé** est défini en remplaçant dans (6) la matrice α I_K par une **matrice définie positive** D tq X' X + D soit une **matrice régulière**.