CONTRAINTE SUR LES VARIABLES, SUR LES OBSERVATIONS (J)
(04 / 08 / 2020, © Monfort, Dicostat2005, 2005-2020)

Un **modèle statistique** peut comporter des **variables** sujettes à diverses **contraintes** : ces variables ne peuvent parcourir entièrement leur **espace de variables**.

Ceci est eg le cas d'un modèle comportant des **variables endogènes** et des **variables exogènes**.

(i) On considère un **modèle image** ($\mathcal{X}$, $\mathcal{B}$, $\mathcal{P}^X$) et un **espace mesurable** auxiliaire ($\mathcal{S}$, $\mathcal{C}$). La **variable d'intérêt** (ou **statistique** d'intérêt) est ici X. S'il existe une **application mesurable** $\phi : \mathcal{X} \mapsto \mathcal{S}$ et une valeur $s \in \mathcal{S}$ telles que ϕ (X) = s, on dit que la va X est liée par la **contrainte** ϕ. Souvent, ($\mathcal{S}$, $\mathcal{C}$) = ($\mathcal{X}$, $\mathcal{B}$).

La contrainte ϕ peut notamment être vectorielle si les espaces précédents sont dotés d'une structure d'**espace vectoriel**.

Si X est un N-**échantillon** (X_1 ,..., X_N), cette contrainte induit une contrainte sur les coordonnées X_n , ie sur les **observations** (cf **contrainte sur les observations**).

(ii) L'approche s'étend directement à un **modèle produit** ($\mathcal{X}$ x $\mathcal{Y}$, $\mathcal{B} \otimes \mathcal{C}$, $\mathcal{P}^X \otimes \mathcal{P}^Y$), la contrainte étant définie par une application $\phi : \mathcal{X}$ x $\mathcal{Y} \mapsto \mathcal{S}$ et une valeur $s \in \mathcal{S}$ tq ϕ (X, Y) = s.

(iii) On considère un **modèle de régression multidimensionnel** linéaire exprimé dans l'espace des variables (ξ_1 ... ξ_1 , η_1 ... η_G) :

$$(1) \quad \begin{array}{l} \eta_1 = \xi_1' \, b_1 + \varepsilon_1', \\ \ldots \\ \eta_G = \xi_G' \, b_G + \varepsilon_G', \end{array}$$

écrit sous forme condensée :

(1)' $\quad \eta' = \xi' B + \varepsilon'$, $\qquad$ avec $B = [b_1 ,..., b_G] \in M_{KG}(\mathbf{R})$,

et satisfaisant à des hypothèses standards :

(2) $\quad E\,\varepsilon = 0$, $\qquad V\,\varepsilon = \Sigma$.

On suppose ce modèle « observé » dans l'**espace des observations** (X, Y) selon :

(3) $\quad Y_n' = X_n' B + U_n'$, $\qquad \forall n \in N_N^*$,

où les variables Y_n' sont sans **corrélation** entre elles.

On appelle **contrainte** (linéaire) **sur les variables endogènes** η la donnée de **matrices** C_y et C_x tq :

(4) $\eta' C_y = \xi' C_x$.

On montre que l'**estimateur des mco** de B et celui de Σ sont identiques à ceux des mco du modèle (1) sans les contraintes (4), ie :

(5) $B^{\wedge} = (X' X)^{-1} X' Y$,

ces matrices étant tq (les X_n et Y_n en étant les vecteurs lignes) :

$$X = [X_1 : X_n : X_N] \in M_{NK} (\mathbf{R}),$$

$$Y = [Y_1 : Y_n : Y_N] \in M_{NG} (\mathbf{R}),$$

et :

(6) $\Sigma^{\wedge} = N^{-1} . U^{\wedge} ' U^{\wedge}$, avec $U^{\wedge} = Y - X B^{\wedge}$ et $\text{rg } \Sigma^{\wedge} = G$.

On en déduit un **estimateur sans biais** $\Sigma^{\sim} = N (N-1)^{-1} \Sigma^{\wedge}$ de Σ.

La **contrainte** (4) porte ici à la fois sur les variables endogènes et sur les variables exogènes.

Il existe des situations dans lesquelles elle ne porte que sur les endogènes : eg **équation de partage**, ou **équation de répartition** (dans laquelle les va endogènes η_g sont contraintes à un total égal à 1) (cf **estimateur par règle de trois**).

(iii) Les contraintes sur les variables engendrent, ipso facto, des **contraintes sur les observations** de ces variables.

(iv) Dans certaines **situations statistiques**, seules certaines observations sont contraintes : eg condition de **signe**, **censure**, existence de **frontières**, etc.

Dans ces cas, une contrainte sur une variable implique une contrainte sur l'une de ses observations, et vice versa (cf **contrainte sur les observations**).