MODÈLE D'INTERDÉPENDANCE LINÉAIRE (J1)
(19 / 12 / 2019, © Monfort, Dicostat2005, 2005-2019))

Le modèle d'interdépendance linéaire est le **modèle d'interdépendance** le plus simple. Dans certains cas, on peut le considérer comme un modèle d'interdépendance « linéarisé », ie comme une **approximation** du modèle d'interdépendance général (ie non linéaire) (cf **modèle linéarisé**). C'est aussi un instrument de base de l'analyse statistique multidimensionnelle (cf **analyse multidimensionnelle**). Il concerne des variables vectorielles, généralement réelles.

(i) Soit $\xi = (\xi_1, ..., \xi_K)$ une liste constituée de K **variables exogènes** (scalaires réelles) et $\eta = (\eta_1, ..., \eta_K)$ une liste constituée de G **variables endogènes** de même type (ie numériques).

On appelle **modèle d'interdépendance linéaire**, ou **modèle à équations simultanées linéaire**, ou encore **modèle linéaire à équations interdépendantes**, ou **système aléatoire linéaire**, un ensemble de relations linéaires (pr aux paramètres) qui relient η à ξ.

Ces relations sont souvent données sous la **forme implicite** suivante, dite **forme structurelle** du modèle, exprimée dans $\mathcal{X}$, x $\mathcal{Y}$ = $\mathbf{R}^G \times \mathbf{R}^K$, ie l'**espace des variables** (η, ξ) :

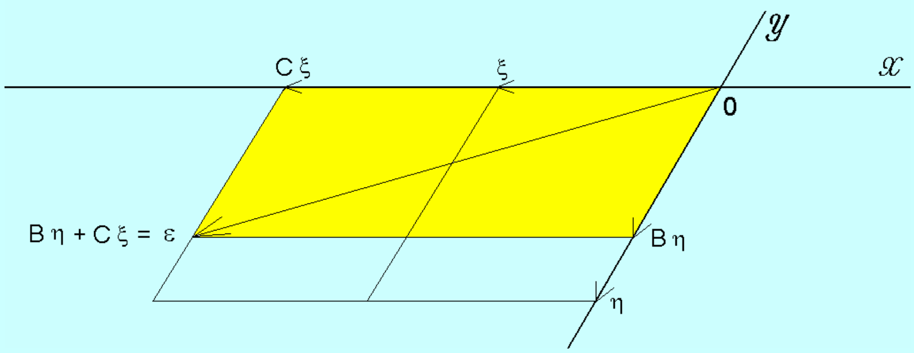
$$(1) \quad \begin{aligned} & b_{11}\,\eta_1 + ... + b_{1G}\,\eta_G + c_{11}\,\xi_1 + ... + c_{1K}\,\xi_K = \varepsilon_1 , \\ & b_{G1}\,\eta_1 + ... + b_{GG}\,\eta_G + c_{G1}\,\xi_1 + ... + c_{GK}\,\xi_K = \varepsilon_G , \end{aligned}$$

ie, sous forme matricielle (ou sous forme d'opérateurs) :

$$(2) \qquad B\,\eta + C\,\xi = \varepsilon,$$

dans laquelle $B \in M_G(\mathbf{R})$, $C \in M_{GK}(\mathbf{R})$ et ε est un **vecteur aléatoire** (**perturbation aléatoire** sur l'équation) à valeurs dans $\mathbf{R}^G$ (cf schéma formel ci-après).

représentation schématique d'un modèle d'interdépendance linéaire
dans l'espace $\mathcal{X} \times \mathcal{Y}$ des valeurs des variables (ξ, η)



D'autres présentations du même modèle sont les suivantes :

(3) $B\,\eta + C\,\xi + \varepsilon = 0$, où $0 \in \mathbf{R}^G$,

(4) $A\,\zeta = \varepsilon$,

ou encore :

(5) $A\,\zeta + \varepsilon = 0$,

avec $A = [B, C]$ (matrice constituée des « **blocs** » B et C) et $\zeta = (\eta\ ///\ \xi)$ (vecteur « colonne » constitué des « blocs » ou « listes » des variables).

(ii) L'observation d'un nombre (fini) N de **copies** de (η , ξ), ie des **va** (X_n , Y_n), n = 1 ,..., N, définit un couple (X , Y) constitué des **matrices d'observation** $X = (x_{nk})_{(n,k)} \in M_{NK}(\mathbf{R})$ et $Y = (y_{ng})_{(n,g)} \in M_{NG}(\mathbf{R})$ dont les lignes resp sont X_n et Y_n .

Pour la n-ième observation (X_n , Y_n) de (η , ξ) le modèle (1) est alors « observé » selon :

(6) $B\,Y_n' + C\,X_n' = U_n'$,

où, $\forall\ n \in N_N^*$, U_n est la n-ième ligne de la (N,G)-matrice U des perturbations (inobservables) :

(7) $U_n = (u_{n1} ,..., u_{nG})$, $\forall\ n \in N_N^*$.

De façon analogue, les représentations (3), (4) et (5) deviennent resp :

(8) $B\,Y_n' + C\,X_n' + U_n' = 0$,

(9) $A\,Z_n' = U_n'$,

et :

(10) $A\,Z_n' + U_n' = 0$.

Une représentation condensée du modèle linéaire peut alors, avec des notations évidentes, prendre l'une des formes suivantes :

(11) $B\,Y' + C\,X' = U'$, $B\,Y' + C\,X' + U' = 0$,

ou l'une des formes suivantes :

(12) $A\,Z' = U'$, $A\,Z' + U' = 0$.

(iii) On rencontre aussi les formes transposées des précédentes.

(iv) En particulier, un modèle linéaire est un **modèle récursif** (H. WOLD) ssi $B \in T_G^+(\mathbf{R})$ est une **matrice triangulaire** supérieure, ou ssi $B \in T_G^-(\mathbf{R})$ est une matrice triangulaire inférieure.

Le modèle (1), dans lequel il existe autant de variables endogènes que d'équations est appelé **modèle d'interdépendance complet** (cf **modèle complet**). Si l'on multiplie (2) par une **matrice régulière** R, on obtient un modèle équivalent au premier. Poser (2) revient implicitement à se placer dans la classe d'équivalence des modèles transformés de (2) par des matrices tq R.

On appelle **règle de normalisation** la règle (nécessité logique et pratique) qui consiste à poser :

(13) $b_{gg} = 1, \quad \forall g \in N_G^*$.

Cette convention de normalisation, généralement implicite, revient (à une permutation près) à prémultiplier (2) par la (G,G)-**matrice diagonale** :

(14) $D = (d_{gh})_{(g,h)}$, où $d_{gh} = \delta_{gh} / b_{gh}$ (avec $b_{gh} \neq 0$ si h = g).

Autrement dit, dans chaque équation g, la variable y_g se voit effecter un **coefficient** $b_{gg} = 1$. En pratique, y_g est souvent écrite à gauche de l'équation (avec ce coefficient muet), les autres variables (endogènes et exogènes) étant écrites à droite de cette même équation.

(v) La **forme réduite** du modèle d'interdépendance d'explicite selon :

(15) $\eta = A \xi + \varphi$, avec $A = - B^{-1} C$ et $\varphi = B^{-1} \varepsilon$,

ce qui suppose que $B \in R_G$ (**R**) (**matrice régulière**). Autrement dit, la forme réduite exprime l'ensemble des variables endogènes en fonction (ici linéaire) de l'ensemble des variables exogènes (cf aussi **régression**).

(vi) Dans l'étude du modèle linéaire précédent (sous forme structurelle ou sous forme réduite), certaines hypothèses sont souvent faites :

(a) le vecteur ξ des exogènes est observé sans **erreur** ;

(b) le vecteur ε des perturbations est en **moyenne** nul ($E \varepsilon = 0$) ;

(c) la **loi de probabilité** de U_n' ne dépend pas de X_n', ni des vecteurs X_α', $\forall \alpha < n$;

(d) les **intercovariance** sont nulles, ie :

(16) $C (U_\alpha', U_\beta') = E U_\alpha' U_\beta' = 0 \in M_G$ (**R**) ;

(e) la matrice des **intracovariances**, ie $V \varepsilon = E \varepsilon \varepsilon' = \Sigma$, ne dépend pas des observations $n \in N_N^*$;

(f) les **convergences en probabilité** suivantes sont vérifiées :

(17)
$$\bar{X}_N = N^{-1} \Sigma_{n=1}^{N} X_n' \to_{N \to +\infty} \mu \in \mathbf{R}^K,$$

$$S_N^2 = N^{-1} \Sigma_{n=1}^{N} (X_n - \bar{X}_N)(X_n - \bar{X}_N)' \to_{N \to +\infty} M_{xx} \in S_K(\mathbf{R}) ;$$

(g) les perturbations sont normales, ie :

(18) $\varepsilon \sim \mathcal{N}_G(0, \Sigma)$ (**loi normale multidimensionnelle** centrée).

Cette dernière hypothèse est utilisée pour mettre en oeuvre des méthodes d'estimation (eg la **méthode du mv**) ou des **tests d'hypothèses**.

Dans les cas les plus simples, et compte tenu de certaines des hypothèses précédentes, on peut estimer la forme réduite $Y_n' = A X_n' + U_n'$ à l'aide d'une **méthode de moindres carrés** (cf **méthode des moindres carrés indirects**).

(vii) Le formalisme de type implicite (1) ou (2) est souvent adopté dans des **sciences** (eg sociologie : économie) où l'**homme de l'art** définit les équations « scalaires » une à une : eg équation de comportement d'agents économiques, équations de type « identité », équations de type réglementaire (barèmes), etc.

Dans les formalismes impliquant des notions d'influence ou de **causalité**, le modèle d'interdépendance peut aussi s'écrire sous forme explicite (à perturbation additive) :

(19) $\eta = D \xi + \varepsilon,$

dans laquelle on exprime les endogènes η en fonction des exogènes ξ de façon globale.

Cette équation vectorielle (dans l'espace des variables) s'écrit alors dans l'espace $\mathbf{R}^N$ des observations (cf **régressions multiples**) :

(20) $Y_n' = C X_n' + U_n' , \ \forall n \in N_N^*.$