

MODÈLE NON LINÉAIRE (G2, J1)

(12 / 09 / 2020, © Monfort, Dicostat2005, 2005-2020)

Une **représentation statistique** dans laquelle les variables ou les paramètres interviennent de façon non linéaire dans une **relation fonctionnelle** peut être qualifié de **modèle non linéaire**.

(i) Plus particulièrement (cf **régression**), un **modèle de régression** dont la **fonction de régression** dépend, de façon non linéaire, d'un **paramètre** explicite est le **modèle de régression non linéaire** (à erreur additive). Il constitue une généralisation du modèle de régression linéaire.

Dans l'**espace de variables**, ce modèle se présente sous la forme suivante :

$$(1) \quad \eta = f(\xi_1, \dots, \xi_k, b_1, \dots, b_Q) + \varepsilon = f(\xi, b) + \varepsilon,$$

dans laquelle η est une **variable endogène**, $\xi = (\xi_1, \dots, \xi_k)$ une liste (ou un vecteur) de **variables exogènes**, $b \in \mathbf{R}^Q$ un **paramètre d'intérêt**, ε une **perturbation aléatoire** et $f : \mathbf{R}^K \times \mathbf{R}^Q \mapsto \mathbf{R}$ une fonction numérique représentant analytiquement la « **relation** », ou la « **liaison** », entre ξ et η .

(ii) En général, f est supposée entièrement déterminée (ou connue analytiquement). L'équation (1) est donc entièrement connue (ou « spécifiée ») si b l'est (cf **spécification**). Ceci est le cas dans les exemples suivants :

$$f(\xi, b_1, b_2) = b_1 \cdot e^{-b(2) \cdot \xi} \quad (\text{forme exponentielle}),$$

$$f(\xi, b_1, b_2) = b_1 \cdot \xi^{b(2)} \quad (\text{forme potentielle ou puissance}),$$

$$f(\xi_1, \xi_2, b) = (\xi_1 + \xi_2)^{-b} \quad (\text{forme hyperbolique}),$$

$$f(\xi_1, \xi_2, b_1, b_2, b_3) = b_1^{\xi(1)} \cdot (b_2 \cdot \xi_2 + b_3 \cdot \xi_3) \\ (\text{forme mixte, puissance-linéaire}),$$

dans lesquels on note par commodité $b(q)$ pour désigner b_q ($\forall q$) et $x(k)$ pour ξ_k ($\forall k$), etc.

(iii) En notant u_n la n -ième **copie** de ε , on « observe » (1) dans l'**espace des observations** selon :

$$(2) \quad y_n = f(x_{n1}, \dots, x_{nK}, b) + u_n = f(X_n, b) + u_n, \quad \forall n \in \mathbf{N}_N^*,$$

où X_n est la n -ième ligne de la **matrice d'observation** $X \in M_{NK}(\mathbf{R})$ des **observations** de ξ et y_n la n -ième coordonnée du vecteur y des observations de η .

Si l'on pose :

$$(3) \quad F_n(b) = f(X_n, b), \quad \forall n \in \mathbf{N}_N^*,$$

expression dans laquelle F_n dépend donc de X_n , on obtient l'expression plus compacte :

$$(4) \quad y = F(b) + u,$$

avec $F(b) = (F_1(b), \dots, F_N(b))'$ aussi noté $(F_1(b) \text{ /// } F_N(b))$ (vecteur colonne), expression dans laquelle $F: \mathbf{R}^Q \mapsto \mathbf{R}^N$ est une fonction vectorielle dépendant de X et /// désigne un saut de ligne. Les notations majuscules F_n et F font référence au caractère aléatoire de ces fonctions.

(iv) En particulier :

(a) si $Q = K$ et si f est bilinéaire sur \mathbf{R}^K (ie $f(\xi, b) = \xi' b$) (cf **forme multilinéaire**), alors $f(X_n, b) = F_n(b) = X_n b$ et (4) se simplifie linéairement selon $y = X b + u$ (**modèle de régression linéaire** multiple) ;

(b) si f est tq $f(\xi, b) = \gamma(\xi)' b$, où $\gamma: \mathbf{R}^K \mapsto \mathbf{R}^Q$, on obtient $F_n(b) = G_n b$, avec $G_n = \gamma(X_n)$. On ne quitte cependant pas la classe des modèles linéaires pr aux paramètres, car ce modèle peut encore s'écrire sous la forme :

$$(5) \quad y = G b + u, \quad \text{avec } G = (G_1, \dots, G_N)';$$

(c) si f est tq $f(\xi, b) = \gamma(\xi)' \psi(b)$, où γ est définie comme précédemment et $\psi: \mathbf{R}^Q \mapsto \mathbf{R}^Q$ est une fonction vectorielle donnée, on rencontre un premier modèle non linéaire non élémentaire (cf **noyau** de GOURSAT). C'est aussi le « premier » modèle non linéaire (ou modèle le plus simple) « séparant » des fonctions non linéaires des variables ξ et des fonctions non linéaires des paramètres b .

(iv) On peut citer les exemples de modèles non linéaires suivants :

(a) le **modèle fractionnaire de H. WOLD** (rapport de formes linéaires) :

$$(6) \quad f(\xi_1, \xi_2, b_1, b_2) = (\xi_1)' b_1 / (\xi_2)' b_2,$$

avec $b_i \in \mathbf{R}^{K(i)}$ et $\xi_j: \Omega \mapsto \mathbf{R}^{K(j)}$, $\forall j \in \{1, 2\}$ (où $K(j)$ désigne la dimension K_j) ;

(b) le **modèle exponentiel** (exponentielle d'une forme linéaire) :

$$(7) \quad f(\xi, b) = b_0 \cdot e^{\xi' b};$$

(c) le **modèle logistique** (partie certaine bornée) :

$$(8) \quad f(\xi, b) = (1 + e^{-\xi' b})^{-1};$$

(d) le **modèle quadratique** :

$$(9) \quad f(\xi, b) = b' M(\xi) b,$$

où $M: \mathbf{R}^K \mapsto M_Q(\mathbf{R})$ est une fonction matricielle donnée.

(v) Avec les notations de (4), la fonction $F : \mathbf{R}^Q \mapsto \mathbf{R}^N$ définit l'équation « paramétrée » (au sens de l'analyse mathématique) par b d'une sous-variété \mathcal{V} de l'espace des observations ambiant \mathbf{R}^N . Dans les cas « réguliers », \mathcal{V} est de pleine dimension Q .

On impose, en général, des **conditions de régularité** à \mathcal{V} : conditions de **différentiabilité** ou conditions de **rang**, le plus souvent.

Les hypothèses (statistiques et mathématiques) les plus simples relatives au modèle précédent sont les suivantes :

(a) $E u = 0$ (perturbation d'effet nul en **moyenne**) ;

(b) $V u = \Sigma > 0$ (**matrice de dispersion** définie positive) ;

(c) $\text{rg } D F (b) = Q, \forall b \in \mathcal{V}_{b^*}$ (\mathcal{V} est de plein rang au **voisinage** de la **vraie valeur** b^* du paramètre $b \in \mathbf{R}^Q$).

(vi) La **méthode des moindres carrés généralisés** appliquée au modèle (4) équivaut à celle du **maximum de vraisemblance** (gaussienne) et revient à résoudre le **problème d'optimisation** suivant :

$$(10) \quad \min (y - F(b))' \Sigma^{-1} (y - F(b)), \quad \text{avec } b \in \mathbf{R}^Q,$$

ie le problème de **programmation mathématique** :

$$(10)' \quad \min (y - z)' \Sigma^{-1} (y - z), \quad \text{avec } z \in \mathcal{V},$$

où $\mathcal{V} = \{z \in \mathbf{R}^N : z = F(b), \forall b \in \mathbf{R}^Q\}$ désigne une représentation paramétrée (ou « **paramétrisation** ») de la sous-variété \mathcal{V} précédente.

Sous certaines hypothèses (et notamment si Σ est une **matrice régulière** et connue), b_F^\wedge est solution des **équations normales** (en b) (condition du premier ordre du problème (10)) :

$$(11) \quad (D F (b))' \Sigma^{-1} (D F (b)) = (D F (b))' \Sigma^{-1} y.$$

(vii) La solution b_F^\wedge de (10) ou (10)' vérifie des propriétés statistiques qui peuvent être analysées :

(a) soit à distance « finie » ($N \ll +\infty$), à l'aide des procédés électroniques de calcul numérique : eg calcul formel, simulations ;

(b) soit asymptotiquement ($N \gg 0$) (cf **propriété asymptotique**).

(viii) Un modèle non linéaire se présente aussi sous **forme implicite**. Dans l'espace des variables, cette forme est :

$$(12) \quad g(\xi, \eta, b) = \varepsilon, \quad \text{avec } E \varepsilon = 0.$$

Elle est « observée » dans l'espace des observations selon :

$$(13) \quad g(X_n, y_n, b) = u_n, \quad \text{avec } E u_n = 0, \quad \forall n \in \mathbb{N}_N^*,$$

où l'on suppose eg que les u_n forment une **suite iid** selon ε , que $g : \mathbf{R}^K \times \mathbf{R} \times \mathbf{R}^Q \mapsto \mathbf{R}$ est une fonction « régulière », que $b \in \mathbf{R}^Q$ est un **paramètre d'intérêt**, que X_n est le vecteur représentant la n -ième observation des variables exogènes (ou prédéterminées) ξ et que y_n est la n -ième observation de la variable endogène η .

On peut écrire (équation « observée ») :

$$(14) \quad G_n(y_n, b) = u_n, \quad \text{avec } E u_n = 0, \quad \forall n \in \mathbb{N}_N^*,$$

où l'écriture G_n implique la prise en compte de X_n . En notation vectorielle compacte, on a :

$$(15) \quad G(b) = u, \quad \text{avec } E u = 0,$$

où G prend en compte (X, y) . Si l'on admet que les u_n sont homoscédastiques et indépendantes (ou seulement non corrélées) entre elles (cf **corrélacion, homoscédasticité**), alors :

$$(16) \quad C(u_\alpha, u_\beta) = \delta_{\alpha\beta} \cdot \sigma_\varepsilon^2, \quad \forall (\alpha, \beta) \in (\mathbb{N}_N^*)^2,$$

ie :

$$(17) \quad V u = \sigma_\varepsilon^2 \cdot I_N, \quad \text{avec } \sigma_\varepsilon^2 = V \varepsilon.$$

On pose :

$$(18) \quad \begin{aligned} D_{nq}(y_n, b) &= \partial G_n(y_n, b) / \partial b_q, \\ E_{nq}(y_n, b) &= (\partial / \partial b_q) \text{Log}(\partial G_n(y_n, b) / \partial y_n), \end{aligned} \quad \forall (n, q) \in \mathbb{N}_N^* \times \mathbb{N}_Q^*,$$

et l'on note $D(y, b)$ (resp $E(y, b)$) la (N, Q) -matrice d'élément générique $D_{nq}(y_n, b)$ (resp $E_{nq}(y_n, b)$).

La Log-**vraisemblance** (gaussienne) associée au modèle (14) s'écrit alors :

$$(19) \quad L(b) = \prod_{n=1}^N f(u_n),$$

dans laquelle f désigne la densité de la **loi normale** centrée $\mathcal{N}_1(0, \sigma_\varepsilon^2)$. Son **gradient** est donc :

$$(20) \quad \text{Grad } L(b) = -D(y, b)' G(b) + E(y, b)' \varepsilon,$$

où $D(y, b)'$ prend ses valeurs dans $M_{Q,N}(\mathbf{R})$, $G(b)$ dans $M_{N,1}(\mathbf{R})$, $E(y, b)'$ dans $M_{Q,N}(\mathbf{R})$ et ε dans \mathbf{R}^N .

On montre notamment (**propriété de convergence de R. DAVIDSON - J.G. MAC'KINNON**) que :

$$(21) \quad \text{plim}_{N \rightarrow +\infty} N^{-1} [D(y, b)' D(y, b) + E(y, b)' E(y, b)] = I(b),$$

où $I(b)$ est la **matrice d'information** (cf **information de FISHER**) du modèle. Par suite, on peut estimer $I(b)$ de façon convergente à l'aide de tout **estimateur** de la forme :

$$(22) \quad I_N(b) = N^{-1} \cdot \{D(y, b_N^\#)' D(y, b_N^\#) + E(y, b_N^\#)' E(y, b_N^\#)\},$$

où $b_N^\#$ est un estimateur convergent (quelconque) de b . La statistique associée au **test des multiplicateurs de LAGRANGE** s'écrit ici :

$$(23) \quad L_N = \{\text{Grad } L(b_N^\#_c)\}' \{N \cdot I(b)\} \{\text{Grad } L(b_N^\#_c)\},$$

où $b_N^\#_c$ est l'estimateur contraint (par l'hypothèse à tester) de b .

(ix) Ce qui précède s'étend au **modèle d'interdépendance non linéaire**. Son équation est de la forme « implicite » suivante :

$$(24) \quad g(\eta, \xi, \theta) = g(\eta_1, \dots, \eta_G, \xi_1, \dots, \xi_K, \theta_1, \dots, \theta_Q) = \varepsilon,$$

dans laquelle $g : \mathbf{R}^G \times \mathbf{R}^K \times \mathbf{R}^Q \mapsto \mathbf{R}^G$ est une fonction vectorielle donnée, les η_g des variables endogènes, les ξ_k des variables exogènes (ou des **variables prédéterminées**), les θ_q des paramètres d'intérêt et ε une perturbation à valeurs dans \mathbf{R}^G , appelée **erreur sur l'équation**.