**CONDITIONS D'IDENTIFICATION (G, J)**
(18 / 06 / 2020, © Monfort, Dicostat2005, 2005-2020)

Une **condition d'identification** est un **critère** permettant de déterminer si un modèle est un **modèle identifiable** ou un **modèle identifié**, en totalité ou en partie. Il existe, en général, plusieurs telles conditions.

Ainsi, on peut chercher à identifier un **mélange de lois**, un **modèle d'analyse de la variance** ou un **modèle d'analyse de la covariance**, un **modèle d'interdépendance** ou l'une de ses équations, etc.

(i) Parmi les nombreuses conditions existantes, les suivantes sont relatives à un **modèle d'interdépendance linéaire**, écrit dans l'**espace des variables** $(\xi, \eta)$ sous la forme :

(1) $B\,\eta + C\,\xi = \varepsilon$, avec $E\,\varepsilon = 0$, $V\,\varepsilon = \Sigma$,

comportant G **variables endogènes** $\eta_g$, K **variables exogènes** $\xi_k$, et « observé » (dans l'**espace des observations** (X, Y)) selon :

(2) $Y B' + X C' = U$, avec $E\,U = 0$, $V\,U = \Sigma \otimes I_N$ .

Si la condition d'inversibilité Dét $B \neq 0$ est vérifiée, la **forme réduite** resp associée à chaque écriture (1) et (2) s'écrit :

(3) $\eta = A\,\xi + \varphi$, avec $E\,\varphi = 0$, $V\,\varphi = B^{-1}\,\Sigma\,B'^{-1}$,

où $A = -B^{-1}C$, $\varphi = B^{-1}\varepsilon$, dans le premier cas, et :

(4) $Y = X A' + V$, avec $E\,V = 0$, $V\,V = V\,(U\,B'^{-1})$,

où $A = -B^{-1}C$, $V = U\,B'^{-1}$ , dans le second cas.

Si A est connue (eg estimée) et que le couple (A, B) incorpore la convention de **normalisation** nécessaire (ie $b_{gg} = 1$, pour tout $g \in N_G{}^*$) et des **restrictions a priori** suffisantes sur leurs éléments, alors l'équation suivante en (B, C) :

(5) $B\,A + C = 0$

comprend un **paramètre** connu $A \in M_{GK}(\mathbf{R})$, donc G . K paramètres scalaires. Par ailleurs, la **forme structurelle** (1) comporte un paramètre $(B, C) \in M_G(\mathbf{R}) \times M_{GK}(\mathbf{R})$, soit, formellement G . (G + K - 1) paramètres scalaires.

On appelle parfois **degré d'identification** du modèle (1) (ou (2)) l'entier défini par la différence :

(6) $I = G\,.\,(G + K - 1) - G\,.\,K = G\,.\,(G - 1) > 0$,

en supposant que $G \geq 2$.

(ii) On dit alors que le modèle (1) (ou (2)) est :

(a) un **modèle juste identifié**, ou un **modèle exactement identifié**, ou simplement un **modèle identifié**, ssi (B, C) comporte exactement I restrictions a priori ;

(b) un **modèle sous-identifié** ssi (B, C) comporte moins de I restrictions a priori ;

(c) un **modèle sur-identifié** ssi (B, C) comporte plus de I restrictions a priori.

Lorsque le modèle est juste identifié et que (5) admet une solution, celle-ci est unique, ie $B^{\tilde{}} = \beta$ (A), $C^{\tilde{}} = \gamma$ (A). Autrement dit, l'application (B, C) $\mapsto$ A est une **application injective** (donc inversible).

Lorsque le modèle est sous-identifié, (5) admet une infinité de solutions (l'application (B, C) $\mapsto$ A est une **application surjective**).

Enfin, lorsque le modèle est sur-identifié, (5) n'admet pas de solution (A n'appartient pas à l'image de l'application définie ci-dessus).

On appelle **condition d'ordre** la condition nécessaire suivante : il existe exactement I restrictions a priori sur (B, C). Cette condition constitue un **critère d'identification**.

(iii) On dit qu'une équation n° g de la forme structurelle est une **équation sous-identifiée** (resp **juste identifiée**, resp **sur-identifiée**) ssi le nombre total de paramètres $(b_{gh} , c_{gk})_{(h,k)}$ qu'elle incorpore (compte tenu des restrictions a priori) est supérieur (resp égal, resp inférieur) au nombre K de variables exogènes $\xi_k$ du modèle, ou encore ssi le nombre total de **variables** $\xi_k$ ou $\eta_g$ absentes de l'équation est inférieur (resp égal, resp supérieur) à G - 1.

(iv) Il existe d'autres conditions que la condition d'ordre précédente, notamment la **condition de rang** suivante. Si le nombre total de variables absentes d'une équation n° g donnée de la forme structurelle est supérieur ou égal à G - 1, alors cette équation est exactement identifiée (ou sur-identifiée) si la **matrice** $[B_g , C_g]$ des coefficients de ces variables dans les G - 1 autres équations du modèle est une **matrice régulière** (ie une matrice de plein **rang**) :

(7) $\quad \mathrm{rg}\, [B_g , C_g] = G - 1.$

(v) Lorsque l'équation g est sur-identifiée, ce qui est fréquent en pratique, l'estimation des paramètres s'effectue à l'aide de méthodes adaptées : **méthode des doubles moindres carrés**, **méthode du maximum de vraisemblance** (à information complète), **maximum de vraisemblance à information limitée** (une équation n° g est considérée), **méthode des variables instrumentales**, ou encore **méthode des triples moindres carrés**.

(vi) Le cas du modèles d'interdépendance non linéaire est plus complexe : il existe des conditions d'identification locale, définies à partir d'un **modèle linéarisé** (cf **linéarisation**) au **voisinage** de certaines valeurs des paramètres : valeurs a priori, estimations préliminaires, valeurs « vraies » (dans une **simulation**).