TEST DE SPÉCIFICATION (H2, 12)
(28 / 07 / 2020, © Monfort, Dicostat2005, 2005-2020)

Un **test de spécification** a pour objet de discriminer entre deux (voire plusieurs) **représentations statistiques** et, notamment, entre des **modèles de régression** ou entre des **modèle d'interdépendance**.

Ce problème de **discrimination** est d'importance pour choisir entre des théories « alternatives » ou encore pour étayer une « nouvelle » théorie avancée par l'**homme de l'art**. C'est pourquoi un test de spécification est aussi appelé **test de sélection entre modèles**, ou **test de choix entre modèles**, ou encore **test de comparaison entre modèles**.

(i) Dans le cas de deux modèles, les données d'un test de spécification sont constituées des **modèles images** ($\mathcal{Y}$, $\mathcal{C}$, $\mathcal{Q}^{\eta}$) et ($\mathcal{Z}$, $\mathcal{D}$, $\mathcal{R}^{\zeta}$), éventuellement issus d'un même modèle fondamental (Ω, $\mathcal{T}$, $\mathcal{P}$), les **variables aléatoires** considérées étant donc $\eta : \Omega \mapsto \mathcal{Y}$ et $\zeta : \Omega \mapsto \mathcal{Z}$, avec des observations Y et Z correspondantes.

Il s'agit alors de déterminer laquelle des spécifications précédentes est la plus « satisfaisante ».

Ce problème ne suppose pas a priori que les deux spécifications sont comparables. Souvent, une partie (au moins) des observations disponibles de η et de ζ est commune aux deux spécifications considérées. On ne restreint donc pas la généralité du problème en supposant que la « liste » ξ des (observations des) **va** à considérer est la même dans les deux modèles, ce qui implique $\mathcal{Z} = \mathcal{Y} = \mathcal{X}$. On se ramène ainsi au problème du choix entre un modèle ($\mathcal{X}$, $\mathcal{B}$, $\mathcal{Q}^{\xi}$) et un modèle ($\mathcal{X}$, $\mathcal{B}$, $\mathcal{R}^{\xi}$), avec $\mathcal{B} = \mathcal{C} \otimes \mathcal{D}$.

(ii) On peut distinguer deux catégories usuelles de tests :

(a) celle des **modèles emboîtés** : on a alors soit $\mathcal{Q}^{\xi} \subset \mathcal{R}^{\xi}$, soit $\mathcal{R}^{\xi} \subset \mathcal{Q}^{\xi}$. Si eg $\mathcal{Q}^{\xi} \subset \mathcal{R}^{\xi}$, il s'agit de tester, dans le modèle ($\mathcal{X}$, $\mathcal{B}$, $\mathcal{R}^{\xi}$), l'hypothèse $H_0 : P^{\xi} \in \mathcal{Q}^{\xi}$ contre l'alternative générale $H_a : P^{\xi} \in \mathcal{R}^{\xi}$, où P^{ξ} désigne la loi de ξ ;

(b) celle des **modèles non emboîtés** : on n'a alors ni $\mathcal{Q}^{\xi} \subset \mathcal{R}^{\xi}$, ni $\mathcal{R}^{\xi} \subset \mathcal{Q}^{\xi}$. On peut avoir eg $\mathcal{Q}^{\xi} \cap \mathcal{R}^{\xi} = \varnothing$ (hypothèses disjointes) ou $\mathcal{Q}^{\xi} \cap \mathcal{R}^{\xi} \neq \varnothing$ (hypothèses non disjointes).

Une approche possible consiste à « plonger » les modèles appartenant aux catégories précédentes dans un modèle plus général de la forme ($\mathscr{X}$, $\mathscr{B}$, $\mathscr{P}^{\xi}$), dans laquelle $\mathscr{Q}^{\xi} \subset \mathscr{P}^{\xi}$ et $\mathscr{R}^{\xi} \subset \mathscr{P}^{\xi}$. Par exemple, on pose $\mathscr{P}^{\xi} = \mathscr{Q}^{\xi} \cup \mathscr{R}^{\xi}$, ce qui permet de se ramener au cadre usuel de la **théorie des tests** (cf **plongement**).

(iii) En pratique, on se limite souvent à des spécifications moins générales que les précédentes. On ne considère pas les modèles ($\mathscr{X}$, $\mathscr{B}$, $\mathscr{Q}^{\xi}$) et ($\mathscr{X}$, $\mathscr{B}$, $\mathscr{R}^{\xi}$) dans leur intégralité. Ainsi, si ζ est un **vecteur aléatoire** dont on distingue G coordonnées η, dites **variables endogènes**, et K coordonnées ξ, dites **variables exogènes**, on peut considérer eg un **modèle de régression multiple** de la forme (non linéaire) :

(2) $\eta = g(\xi, \theta) + \varepsilon$,

où g est une **fonction de régression** associée à une loi $Q^{\xi} \in \mathscr{Q}^{\xi}$, et l'on teste cette hypothèse contre celle d'un modèle de régression multiple de la forme :

(3) $\eta = h(\xi, \theta) + \varepsilon$,

dans laquelle h est la fonction de régression associée à l'une des lois $R^{\xi} \in \mathscr{R}^{\xi}$.

Dans ce cadre, une spécification peut porter eg :

(a) au « premier ordre », sur g et h, ie sur les fonctions de régression (ou **espérances conditionnelles**), elles-mêmes ;

(b) au « second ordre », sur la **dispersion** conditionnelle de η pr à ξ dans chacun des deux cas.

(iv) La notion de test de spécification est à distinguer de celle de **test d'adéquation** ou de **test d'ajustement** (cf aussi **adéquation d'un ajustement**, **qualité d'un ajustement**). En effet :

(a) le premier compare des modèles entre eux au vu de l'observation disponible : le modèle retenu est celui que l'observation confirme le mieux ;

(b) tandis que le second vérifie le « degré » de compatibilité d'un modèle avec l'observation.