MODÈLES EN ÉCONOMIE. NOTION DE MODÈLES. GÉNÉRALITÉS.

Lors de l'assemblée générale de la Régionale APMEP, René GIRAUD, président honoraire de l'Université de Poitiers, a prononcé une conférence, "Modèles en économie". Nous en publions ci-dessous l'introduction qu'il nous a aimablement communiquée.

1 - Introduction

Pour étudier un phénomène économique on essaie de représenter celui-ci par le comportement d'une variable. Cette variable économique dépend elle-même d'autres variables que l'on relie entre elles par une relation mathématique.

* Par exemple, si l'on étudie l'Offre O et la Demande D d'un certain bien sur un marché, on sait que D et O dépendent du prix p du bien. On peut écrire que O est une certaine fonction du prix, D une autre fonction de ce même prix et que l'équilibre sur le marché se traduit par O = D. On vient donc de construire un modèle élémentaire sous la forme :

$$(1)\quad \begin{cases} O = f\,(p) \\ D = g\,(p) \\ O = D. \end{cases}$$

Cependant, Offre et Demande dépendent d'autres variables que le prix.

* La demande de certains produits alimentaires par exemple dépend du revenu des ménages, du prix de produits analogues, etc... S'il s'agit de denrées agricoles l'offre dépend des prix de l'année précédente.

Il n'est donc pas inutile de préciser la relation donnée à un instant t et d'indexer les variables utilisées de la façon suivante :

$$(2)\quad \begin{cases} O_t = f\,(p_{t-1}, x_{1t}, x_{2t}, ..., x_{nt}) \\ D_t = g\,(p_t, x_{1t}, x_{2t}, ..., x_{pt}) \\ O_t = D_t \end{cases}$$

Dans le deuxième modèle ainsi obtenu, on a introduit différentes variables explicatives X_1, X_2, ..., X_n et on a considéré les réalisations de ces variables aux instants t et t - 1.

On remarquera ici que le modèle proposé comporte plusieurs équations : nous avons un modèle à équations simultanées.

Mais pour commencer il est plus simple de raisonner sur un modèle comportant une seule équation.

II - Notion de modèle aléatoire.

Proposons nous d'étudier la consommation C_i d'un certain bien par un ménage *i*. Cette consommation dépend entre autres du revenu r_i du ménage.

Le modèle le plus élémentaire consiste à expliciter C_i en fonction de r_i. D'autres facteurs, dont certains sont partiellement inconnus, déterminent également C_i. On peut condenser les effets de ces autres facteurs en un facteur aléatoire ε_i.

On obtiendra alors le modèle *aléatoire :* (1) $C_i = f(r_i) + \varepsilon_i$.

Les ε_i obéissent à une loi de probabilité P qu'il faudra préciser au cours des hypothèses faites sur le modèle. Très souvent, ces hypothèses qui portent sur les premiers moments des ε_i suffiront.

En présence du modèle (1) il faudra maintenant s'assurer que la classe de la fonction choisie pour f n'est pas en contradiction avec les résultats de l'expérience.

Par exemple, si on décide de choisir pour f une fonction du premier degré : (2) $C_i = a\, r_i + b + \varepsilon_i$

on s'assurera en faisant varier i selon les différents ménages considérés que la relation (2) est bien satisfaisante. On dit que l'on « *teste* » le modèle.

Si le résultat obtenu est convenable on « *estimera* » alors les paramètres a et b. Enfin on définira une *règle de prévision* permettant, connaissant r_i, de déterminer C_i.

III - Nature des variables figurant dans le modèle. Spécification et structure du modèle.

A) On distingue deux types de variables dans un modèle économique.

1) Les variables *exogènes.*

Ce sont des variables explicatives de la variable étudiée. Elles sont considérées comme des données autonomes. Ainsi dans le modèle (1) $C_i = a\, r_i + b + \varepsilon_i$ la grandeur r_i est la variable explicative ou variable exogène. C'est sa valeur pour i donné qui permet de déterminer C_i, à l'aléatoire près ε_i. Parfois même elles sont considérées comme prédéterminées. Rappelons que dans une fonction keynésienne r_i serait le revenu de l'individu i sur la période qui précède la période de référence de la consommation.

2) Les variables *endogènes* ou variables à expliquer.

C_i est la variable endogène du modèle précédent. C_i est maintenant devenue une variable aléatoire par l'intermédiaire de ε_i.

Elle ne sera pas traitée comme une variable exogène ; cette distinction entre nature des variables est très importante et devra toujours être précisée avant l'étude d'un modèle.

B) Qu'entend-on maintenant par spécification du modèle ?

On dit que l'on *spécifie* un modèle quand on donne à celui-ci sa formulation mathématique définitive. Le modèle (1) ci-dessus est spécifié. On connaît la forme de la fonction f dans l'expression (2) $C_i = f(r_i) + \varepsilon_i$.

(3) $f(r_i) = a r_i + b$.

L'adjonction de la variable aléatoire ε_i donne au modèle sa formulation définitive. La *structure* du modèle, elle, est constituée par l'ensemble des paramètres qui définissent complétement le modèle.

Par exemple, supposons que : $a = 0{,}3$ et $b = 15$ et que les ε suivent une loi normale d'espérance mathématique nulle et de variance égale à 3 ; alors l'ensemble $\left\{\begin{array}{l} a = 0{,}3 \\ b = 15 \\ \sigma = \sqrt{3} \end{array}\right.$ constitue la structure du modèle.

Le but de l'économétrie sera alors, connaissant les couples (C_i, r_i) associés aux différents ménages i de déterminer la structure vraie du modèle ; c'est-à-dire, à partir d'un *espace échantillon* défini par l'ensemble des (C_i, r_i) de *déterminer la structure vraie* du modèle dans *l'espace* à 3 dimensions des *structures* (a, b, σ).

Le modèle étant choisi une fois pour toutes on admet qu'il existe un triplet (a, b, σ) qui permet de représenter exactement le processus par lequel les valeurs des variables observées ont été déterminées.

La procédure choisie, comme nous le verrons ultérieurement, consistera à obtenir des estimateurs des paramètres a et b permettant de déterminer au mieux les valeurs réelles de ces paramètres. On les appréciera en général à l'aide d'intervalles de confiance choisis à un niveau de signification donné α.

Par exemple dans le modèle : $C_i = a\, r_i + b + \varepsilon_i$, on trouvera que a appartient à l'intervalle [0,27 ; 0,33] avec une probabilité de 95 %. De même b appartiendra à l'intervalle [14; 16] avec la même probabilité.

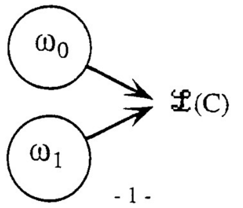
Il sera possible également d'estimer l'écart-type σ de la variable aléatoire ε. On va voir maintenant le rôle important joué par cette variable aléatoire dans le modèle.

IV- L'identification du modèle. Structures non identifiables - structures équivalentes.

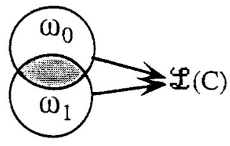
Considérons toujours notre modèle : (1) $C_i = a\, r_i + b + \varepsilon_i$ et imaginons que la procédure utilisée en partant des informations recueillies sur (C_i, r_i) nous conduise non pas à une solution unique mais à deux structures distinctes :

$$\omega_0 = \begin{cases} a_0 \\ b_0 \\ \sigma_0 \end{cases} \qquad \text{et} \quad \omega_1 = \begin{cases} a_1 \\ b_1 \\ \sigma_1 \end{cases}$$

Il est évident que la loi de probabilité définie sur ε précise la loi de la variable endogène C.
Donc chaque structure, compte tenu des valeurs données aux exogènes et de la loi de ε, conduit à une loi de probabilité de C.
Imaginons que ω_0 et ω_1 conduisent à la même loi $\mathcal{L}$ de C :
Deux cas de figure sont possibles :
- ou ω_0 et ω_1 sont distinctes et on ne peut choisir ω_0 plutôt que ω_1. Dans ce cas les structures considérées ne sont pas identifiables et, par là même, le modèle n'est *pas identifiable*. On ne peut pas déterminer les valeurs des paramètres qui figurent dans le modèle.



- 1 -



- 2 -

- ou ω_0 et ω_1 ne sont pas distinctes, leur intersection n'est pas vide. Les structures ω_0 et ω_1 permettent d'identifier une partie des paramètres du modèle, ceux précisément qui appartiennent à l'intersection de ω_0 et ω_1.

Ces deux structures seront dites *équivalentes*, mais ne permettront pas une identification complète du modèle.
Ce problème très important de l'identification se posera en particulier lors de l'étude de modèles à équations simultanées.

V- Prévision des variables endogènes à l'aide de valeurs fixées pour les exogènes.

L'intérêt d'un modèle dont la structure est déterminée consiste à l'utiliser pour prévoir, à une époque future ou dans une circonstance fixée s'il s'agit d'observations prises au même instant, les valeurs des variables endogènes lorsque celles des exogènes sont fixées.
Soit le modèle suivant. On étudie l'évolution des importations Y en fonction de la production intérieure brute X_1 et de la formation des stocks X_2. On aura :

(1) $y_t = a_1 x_{1t} + a_2 x_{2t} + b + \varepsilon_i$ $\quad t = 1, 2, \ldots T$, t représente ici le temps.

L'historique permettant d'estimer les paramètres porte sur la période 1960 - 1985. Les observations étant annuelles.
On a trouvé comme estimations ponctuelles des paramètres : $\hat{a}_1 = 0{,}140$; $\hat{a}_2 = 0{,}6$; $\hat{b} = 6$.
Le *modèle estimé* s'écrit alors : (2) $\underset{}{y_t} = \underset{(0{,}004)}{(0{,}140)}\, x_{1t} + \underset{(0{,}10)}{(0{,}60)}\, x_{2t} + \underset{(1{,}1).}{6}$

Les chiffres entre parenthèses sous les valeurs estimées représentent ici les écarts-types des estimateurs $\hat{a}_1$, $\hat{a}_2$ et $\hat{b}$ afin de déterminer des intervalles de confiances pour a_1, a_2, b. Supposons que l'on veuille faire une prévision pour l'année 1987.
Il faudra alors fixer la production intérieure brute X_1, et la formation des stocks X_2 en 1987.
Si on impose $x_1 = 1030$ en milliards de francs au prix de 1960 et $x_2 = 12{,}7$ on aura comme prévision pour Y :

$$(3) \quad \hat{y}_{1987} = (0{,}140)(1030) + (0{,}60)(12{,}7) + 6$$

ou d'une façon générale : $\quad (4) \quad \hat{y}_\theta = \hat{a}_1 x_{1\theta} + \hat{a}_2 x_{2\theta} + \hat{b}$, θ étant l'époque de la prévision.

Il est évident que la valeur prévue pour Y en 1987 appelle des remarques.
1. On a choisi arbitrairement les valeurs $x_{1\theta}$ et $x_{2\theta}$ compte tenu de leur évolution passée. Première source d'erreur.
2. La spécification du modèle n'est peut-être pas parfaite. La forme de la fonction choisie pour expliquer l'évolution de Y n'est pas suffisamment précise.
3. Il est possible enfin que les variables qui expliquent, d'après notre modèle, l'évolution de Y n'interviennent plus de la même façon que lorsque l'on a étudié notre phénomène sur la période 1960-1985. Autrement dit, il peut y avoir rupture d'équilibre entre les variables qui expliquent le phénomène au moment de la prévision. Les parts respectives qu'elles représentent dans la variation de Y ne sont plus les mêmes. Il est évident que la prévision dans ce cas sera sérieusement biaisée.

Les trois causes évoquées ci-dessus sont sources d'erreurs pour la prévision et nous verrons les méthodes permettant de minimiser ces erreurs.

Pour résumer cette introduction, retenons les étapes suivantes dans la construction et l'utilisation du modèle :
1) spécification du modèle ;
2) estimation des paramètres et test du modèle à l'aide de statistiques déjà connues ;
3) prévision de la variable endogène.