

Recherche opérationnelle Les modèles linéaires et les séries chronologiques

Alain Haurie

Volume 41, numéro 4, janvier–mars 1966

URI : <https://id.erudit.org/iderudit/1003129ar>

DOI : <https://doi.org/10.7202/1003129ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

HEC Montréal

ISSN

0001-771X (imprimé)

1710-3991 (numérique)

[Découvrir la revue](#)

Citer cet article

Haurie, A. (1966). Recherche opérationnelle : les modèles linéaires et les séries chronologiques. *L'Actualité économique*, 41(4), 693–718.
<https://doi.org/10.7202/1003129ar>

Analyse

Recherche opérationnelle

Les modèles linéaires et les séries chronologiques

L'étude des relations entre des variables économiques fait largement appel à des modèles construits sur la base de théories économiques. La spécification du modèle est un ensemble d'hypothèses sur la forme générale de la relation que l'on étudie, et sur les caractéristiques des éléments aléatoires qui y figurent. La statistique a alors pour rôle :

- d'estimer les paramètres du modèle ;
- d'indiquer la confiance que l'on peut mettre dans ces estimés et de tester les hypothèses que l'on peut faire.

C'est ainsi que se développe l'étude statistique des modèles de régression linéaire normale. Le temps n'apparaît pas dans ces modèles, ou, du moins, ces modèles ne traduisent pas l'évolution au cours du temps des variables. Sur des modèles où le temps intervient, les mêmes procédés statistiques sont souvent employés ; on peut s'interroger sur leur validité.

On rappellera donc les principes et les résultats de l'étude des modèles de régression linéaire normale, et ensuite on considèrera la valeur de ces résultats à propos de modèles faisant intervenir le temps (modèles dynamiques).

Modèle de régression linéaire normale. Une variable y , dépendante ou « endogène » dépend linéairement de k variables indépendantes ou « exogènes » : x_1, x_2, \dots, x_k

$$y = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_k x_k$$

Mais cette équation n'est pas toujours vérifiée : il y a des « erreurs sur l'équation », que l'on peut interpréter comme des perturbations ou des « chocs ». Ces erreurs u , sont aléatoires, ont une loi de probabilité normale de paramètres 0, σ_u

Le modèle est donc :

$$y = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_k x_k + u$$

$$E[u] = 0$$

$$D^2[u] = \sigma_u^2 \quad \text{loi de } u : \mathcal{N}(0, \sigma_u)$$

Un échantillon sera obtenu en mesurant les différentes valeurs $y^1 \dots y^m$ associées aux groupes de k variables :

$$1 : (x_1^1, x_2^1, \dots, x_k^1)$$

⋮

$$m : (x_1^m, x_2^m, \dots, x_k^m)$$

Ces observations seront résumées dans les matrices :

$$Y = \begin{bmatrix} y^1 \\ \vdots \\ y^m \end{bmatrix} \quad \text{et } M = \begin{bmatrix} x_1^1 & x_2^1 & \dots & x_k^1 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^m & x_2^m & \dots & x_k^m \end{bmatrix}$$

Lors des mesures, la perturbation u aura pris différentes valeurs u^1, \dots, u^m , chacune étant aléatoire.

Une hypothèse importante est celle de l'indépendance des différentes perturbations, qui entraîne :

$$E[u_i u_j] = 0 \quad \text{si } i \neq j$$

1) Estimation des paramètres du modèle.

Ces paramètres sont les k coefficients : $\alpha_1 \dots \alpha_k$ et la variance des perturbations : σ_u^2

On peut chercher des estimés des coefficients $\alpha_1 \dots \alpha_k$ en suivant l'un des deux critères d'optimalité :

- meilleurs estimés non biaisés, linéaires,
- estimés du maximum de vraisemblance.

Le meilleur estimé non biaisé de α_i sera la grandeur α_i^* calculée à partir de l'échantillon, donc aléatoire puisque l'échantillon l'est,

mais dont l'espérance mathématique est toujours $E[\alpha_i^*] = \alpha_i$ et dont la variance est minimale.

Toute autre estimation donnerait un estimé $\hat{\alpha}$ de variance supérieure ou égale à $D^2[\alpha_i^*]$.

L'estimé du maximum de vraisemblance est celui qui maximise la « fonction de vraisemblance » de l'échantillon : si u est $\mathcal{N}(0, \sigma_u)$, $y - \alpha_1 x_1 - \alpha_2 x_2 - \dots - \alpha_k x_k$ est aussi $\mathcal{N}(0, \sigma_u)$, c'est-à-dire que l'expression de la densité de probabilité conditionnelle de y , étant donné les valeurs de $x_1 \dots x_k$ est :

$$f(y^1 | x_1 \dots x_k) = \frac{1}{\sqrt{2\pi\sigma_u^2}} \exp \left[-\frac{(y - \alpha_1 x_1 - \dots - \alpha_k x_k)^2}{2\sigma_u^2} \right]$$

L'échantillon est constitué d'un ensemble de variables u indépendantes et a pour densité de probabilité

$$f(y^1 | x_1^1 \dots x_k^1) \dots f(y^m | x_1^m \dots x_k^m) = \frac{1}{(2\pi\sigma_u^2)^{\frac{m}{2}}} \exp \left[-\frac{(y^1 - \alpha_1 x_1^1 - \dots - \alpha_k x_k^1)^2 + \dots + (y^m - \alpha_1 x_1^m - \dots - \alpha_k x_k^m)^2}{2\sigma_u^2} \right]$$

La notation matricielle va simplifier l'écriture :

$$\text{si } \alpha = \begin{bmatrix} \alpha_1 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_k \end{bmatrix}$$

$$\text{si } u = \begin{bmatrix} u^1 \\ \cdot \\ \cdot \\ \cdot \\ u^m \end{bmatrix}$$

on a donc :

$$Y = M \cdot \alpha + u$$

$$u = Y - M \cdot \alpha$$

la fonction de densité de l'échantillon est alors

$$f(Y|M) = \frac{1}{(2\pi\sigma_u^2)^{\frac{m}{2}}} \exp \left[-\frac{(Y - M \cdot \alpha)^2 + (Y - M \cdot \alpha)^2}{2\sigma_u^2} \right]$$

On remarquera qu'on étudie en fait la distribution conditionnelle de l'échantillon des y , associée à la matrice M des observations des variables exogènes ($x_1 \dots x_k$). C'est ainsi que dans la suite, M ne sera pas considérée comme aléatoire, mais comme une matrice fixée.

Cette fonction de densité est fonction de la colonne α . On prendra pour estimé $\hat{\alpha}$ de α , la colonne qui maximise la probabilité élémentaire de l'échantillon, donc qui maximise la fonction de densité, ce qui revient à rendre maximum son logarithme: $\text{Log}(f(Y|M))$ est appelé la fonction de vraisemblance de l'échantillon. $\hat{\alpha}$ sera déterminée par :

$$\frac{\partial}{\partial \alpha} \text{Log} (f(Y|M)) = 0$$

La fonction de vraisemblance sera maximale si la quantité :

$$(Y - M \hat{\alpha})^2 + (Y - M \hat{\alpha})^2$$

est « minimale ». Le développement de ce produit scalaire serait :

$$(y^1 - \hat{\alpha}_1 x_1^1 - \dots - \hat{\alpha}_k x_k^1)^2 + \dots + (y^m - \hat{\alpha}_1 x_1^m - \dots - \hat{\alpha}_k x_k^m)^2$$

On cherche donc à minimiser la somme des carrés des différences entre y_i et une combinaison linéaire

$$\hat{\alpha}_1 x_1^i \dots \hat{\alpha}_k x_k^i$$

Cette méthode d'estimation s'appelle donc méthode des moindres carrés. La colonne $\hat{\alpha}$ sera déterminée alors par l'équation matricielle :

$$M^+ (Y - M \hat{\alpha}) = 0$$

d'où :

$$\hat{\alpha} = (M^+ M)^{-1} M^+ Y$$

qui constitue l'estimé par les moindres carrés des coefficients de régressions.

Il serait facile de voir que l'estimé $\hat{\alpha}$, qui est aléatoire puisque lié à un échantillon aléatoire, a pour espérance mathématique :

$$E [\hat{\alpha}] = \alpha$$

On dit alors que l'estimé n'est pas biaisé. La colonne

$$\hat{\alpha} = \begin{bmatrix} \hat{\alpha}_1 \\ \vdots \\ \hat{\alpha}_k \end{bmatrix}$$

est donc constituée de k éléments aléatoires dont on connaît l'espérance mathématique, la matrice des variances et covariances de ces k variables sera :

$$E \left[(\hat{\alpha} - \alpha) (\hat{\alpha} - \alpha)^+ \right] = \sigma_u^2 (M + M)^{-1}$$

Enfin on démontre que les estimés par les moindres carrés sont aussi les estimés de variance minimale ; ceci restant vrai même quand les perturbations ne sont pas distribuées suivant une loi normale.

En résumé, l'estimation des coefficients du modèle de régression par la méthode des moindres carrés, conduit aux meilleurs estimés linéaires ; si la loi des perturbations est normale ce sont aussi les estimés du maximum de vraisemblance. Quand la taille de l'échantillon augmente, la précision dans la détermination des α_j s'accroît, et on montre que : quel que soit $\varepsilon > 0$

$$P \left[|\hat{\alpha}_j - \alpha_j| > \varepsilon \right] \longrightarrow 0$$

quand le nombre d'observations m tend vers l'infini ; on dit alors que $\hat{\alpha}_j$ tend en probabilité vers la valeur α_j .

Ayant les estimés des coefficients de régression, on peut calculer les m résidus

$$\begin{aligned} e^1 &= y^1 - \hat{\alpha}_1 x_1^1 - \hat{\alpha}_2 x_2^1 - \dots - \hat{\alpha}_k x_k^1 \\ &\vdots \\ e^m &= y^m - \hat{\alpha}_1 x_1^m - \hat{\alpha}_2 x_2^m - \dots - \hat{\alpha}_k x_k^m \end{aligned}$$

ou, sous forme matricielle :

$$e = Y - M \hat{\alpha}$$

ces résidus permettent d'estimer la variance σ_u^2 des perturbations u par :

$$\hat{\sigma}_u^2 = \frac{e'e}{m-k} = \frac{\sum e^{i(2)}}{m-k}$$

Intervalles de confiance ; tests d'hypothèse. Sous l'hypothèse de normalité, pour les perturbations, la loi des estimés est une loi à k dimensions, normale, de matrice de covariance :

$$\alpha \text{ est } \mathcal{N}\left(\alpha; \sigma_u^2 (M+M)^{-1}\right)$$

La loi du rapport :

$$\frac{\sum e^{i(2)}}{\sigma_u^2}$$

est une loi du χ^2 à $m-k$ degrés de liberté.

Définir des intervalles de confiance pour α_j , nécessite une loi ne contenant plus la variance inconnue σ_u^2 :

le rapport :

$$t = \frac{\hat{\alpha}_j - \alpha_j}{\sqrt{\sum \frac{e^{i(3)}}{m-k}}} \frac{1}{m_{jj}}$$

où m_{jj} est le j^{me} élément de la diagonale principale de $(M+M)^{-1}$ ne dépend plus de σ_u^2 et suit une loi de Student à $m-k$ degrés de liberté. Si $\varphi(t)$ est la loi de Student à $m-k$ degrés de liberté, on appelle $t_{0.025}$ la valeur de t telle que :

$$\int_{t_{0.025}}^{\infty} \varphi(t) dt = 0.025$$

Au cours d'épreuves répétées constituant des échantillons de taille m , 95 p.c. des estimations du coefficient α_j par $\hat{\alpha}_j$ sont telles que l'intervalle :

$$\hat{\alpha}_j \pm t_{0.025} \sqrt{\frac{\sum e^{i(3)}}{m-k}} \sqrt{m_{jj}}$$

contienne la valeur exacte α_j .

Enfin, l'analyse de variance permet de tester la contribution d'un groupe de variables exogènes à la « variance expliquée » de la variable endogène.

La notion d'espace de probabilité. Dans ce modèle de régression, $x_1, x_2 \dots x_k$ étant fixés, y est aléatoire : son espérance mathématique est une combinaison linéaire des valeurs exogènes, la déviation de y par rapport à cette valeur est due à des phénomènes assez complexes pour qu'on les appelle aléatoires.

Il est commode d'imaginer une variable aléatoire comme étant une fonction :

$$y = y(\omega)$$

$$\omega \in \Omega$$

définie sur « un espace de probabilité » Ω ; un élément ω de cet espace abstrait Ω , figure la réalisation de certains phénomènes entraînant pour y la valeur : $y(\omega)$.

Toutes les réalisations possibles sont ainsi les éléments de Ω .

Ω est un ensemble mesurable. De même que la longueur d'un segment est définie comme une mesure sur une droite, la surface comme une mesure dans le plan, le volume, dans l'espace à trois dimensions, une mesure est définie sur l'ensemble Ω , avec les conventions :

- 1) $\text{mes}(\Omega) = 1$
- 2) si $\Omega_1, \Omega_2, \dots, \Omega_i, \dots$ sont des sous-ensembles disjoints de Ω

$$\text{mes}(\Omega_1 \cup \Omega_2 \dots \cup \Omega_i \dots) = \text{mes}(\Omega_1) + \text{mes}(\Omega_2) + \dots + \text{mes}(\Omega_i) + \dots$$

alors l'ensemble des ω entraînant pour $y(\omega)$ des valeurs inférieures à y^1 , que l'on note :

$$\{\omega \mid y(\omega) < y^1\}$$

a pour mesure :

$$\text{mes} \{\omega \mid y(\omega) < y^1\} = P[y < y^1] = F(y^1)$$

Cette mesure correspond aux probabilités intervenant dans la définition d'une variable aléatoire ; on voit en particulier la signification de la fonction de répartition : $F(y)$.

Ensemble d'observations et suite d'observations. On observe m réalisations : $y^1 \dots y^m$ de la variable aléatoire y , on obtient un ensemble de m valeurs :

$$\{y^1 \dots y^m\}$$

si ces observations ont été effectuées au cours du temps : il y a une première observation, puis une seconde et une $m^{\text{ième}}$. On a alors une suite de valeurs y_t ($t = 1, 2, \dots, m$).

La valeur y_t peut être influencée par les valeurs précédentes, il y aura alors un certain modèle d'évolution de y au cours du temps. Supposons que les valeurs successives y_t sont prises suivant un schéma autorégressif :

$$y_t = ay_{t-1} + u_t$$

où u_t est une variable aléatoire :

$$E[u_t] = 0$$

$$E[u_t u_{t'}] = 0 \quad \text{si } t \neq t'$$

Il y a autocorrélation entre les valeurs y_t qui se succèdent. L'espérance mathématique de y_t est :

$$E[y_t] = ay_{t-1}$$

donc, le coefficient a est tel que :

$$E[|y_t - ay_{t-1}|^2]$$

soit minimum ¹.

Donc

$$\frac{dE[|y_t - ay_{t-1}|^2]}{da} = 0$$

$$a = \frac{E[y_t y_{t-1}^2]}{E[y_{t-1}]^2}$$

qui constitue le coefficient d'autocorrélation de la suite y_t ($t = 1, 2, \dots$).

L'interdépendance entre les valeurs successives de y_t peut prendre bien des formes ; celle-ci était très simple. La considération de l'évo-

1. On démontre, en théorie des probabilités, que si X est une variable aléatoire, $E[|X - c|^2]$ est minimum si $c = E[X]$.

lution au cours du temps de quantités aléatoires dans des modèles économiques, se fait en y introduisant des processus aléatoires au lieu de variables aléatoires.

Les processus aléatoires. Considérons une fonction périodique, du temps :

$$y = A \sin vt \quad t \in \mathcal{T}$$

le graphe de cette fonction sera une sinusoïde d'amplitude A , de pulsation v .

Si A est une variable aléatoire : $A = A(\omega)$ $\omega \in \Omega$ espace des probabilités, la fonction devient aussi aléatoire, et à un point ω de l'espace des probabilités Ω correspond une fonction sinus, de pulsation v , et d'amplitude : $A(\omega)$.

On a donc autant de réalisations différentes de la fonction que de points ω dans l'espace des probabilités.

Cet exemple est très simple, on peut compliquer un peu la structure de la fonction aléatoire.

Soit :

$$y = A_1 \sin vt + A_2 \sin \frac{v}{2} t + A_3 \sin \frac{v}{3} t$$

où $\{A_1, A_2, A_3\}$ constitue un triplet aléatoire, c'est-à-dire qu'à un point ω de l'espace des probabilités Ω correspondront trois valeurs :

$$\{A_1(\omega), A_2(\omega), A_3(\omega)\}$$

qui constitueront une réalisation particulière du triplet.

À une réalisation particulière pour $\{A_1, A_2, A_3\}$ correspond une fonction liant y à t .

Dans les deux cas, on peut considérer y comme fonction de deux variables :

$$y = f(t, \omega)$$

— chaque réalisation ω entraîne une liaison particulière entre y et t ,

— à chaque époque t , différentes valeurs de y peuvent être associées suivant les différentes réalisations possibles ω .

Les deux exemples donnés sont loin d'être généraux. On gardera, pour définir une fonction aléatoire, la forme :

$$y = f(t, \omega)$$

$$t \in \mathcal{T} \quad \omega \in \Omega$$

— \mathcal{T} est l'ensemble des temps.

— Ω est un espace de probabilités.

Rien n'a été précisé sur l'ensemble \mathcal{T} , il peut être continu :

$$\mathcal{T} : \{t \mid t_a \leq t \leq t_b\}$$

il peut être aussi un ensemble discret :

$$\mathcal{T} : \{t \mid t = 0, 1, 2, \dots, n, \dots\}$$

ensemble d'époques successives : le processus autorégressif dont on a déjà parlé était défini sur un tel ensemble.

Certains auteurs distinguent les deux cas, en parlant respectivement de processus aléatoire ou de suite aléatoire ; dans la suite on ne fera pas de différence entre ces deux appellations.

L'étude de phénomènes économiques se faisant généralement au cours du temps, on peut toujours faire correspondre son époque à chaque valeur y observée. La correspondance : temps, valeur observée, définit une fonction

$$y = f(t)$$

La forme de cette fonction dépend de causes trop complexes pour être explicitées, on préfère dire qu'il y a un certain hasard dans la réalisation d'une telle liaison, ou bien que la liaison observée était une réalisation particulière d'une fonction aléatoire :

$$y = f(t, \omega) \quad \omega \in \Omega, t \in \mathcal{T}$$

Il faut bien insister sur le fait qu'à un ω correspond une réalisation complète du processus, on peut alors s'interroger sur la structure de l'espace de probabilités Ω .

Pour la plupart des grandeurs aléatoires dépendant du temps, la suite des valeurs prises dépend de réalisations successives, au cours des époques, de phénomènes aléatoires.

On considère ω comme un point dans un espace à plusieurs dimensions, dont les coordonnées sont $\omega(t)$.

$$\omega, \{\omega(t) \mid t \in \mathcal{T}\}$$

Le nombre de dimensions sera égal au nombre d'époques t figurant dans l'ensemble \mathcal{T} . Fixer ω dans Ω , c'est fixer la suite des réalisations $\omega(t)$, $t \in \mathcal{T}$ entraînant la liaison

$$y = f(t)$$

Cette façon de concevoir l'espace des probabilités entraîne des difficultés d'interprétations, surtout lorsque l'ensemble \mathcal{T} est continu :

$$\mathcal{T} : \{t \mid t_a \leq t \leq t_b\}$$

Une étude statistique d'un processus aléatoire devant explorer les réalisations possibles de fonctions du temps exigera en général des échantillons importants. De ce qui vient d'être dit on peut déduire que beaucoup d'observations de grandeurs économiques ne constituent en fait que des échantillons de taille un : on n'observe qu'une seule évolution au cours du temps.

Des hypothèses restrictives sur la nature du processus sont alors faites, une des plus importantes est celle de la stationnarité.

La stationnarité. Une variable aléatoire est entièrement définie par sa fonction de répartition.

Le processus aléatoire y_t , $t \in \mathcal{T}$ est constitué d'une suite de variables aléatoires. Si $t = t_0$, $y_t = y_0$ est une variable aléatoire dont on note la fonction de répartition :

$$F(y_0; t_0) = P[y_0 < y_0]$$

à deux époques : t_0, t_1 correspond le couple aléatoire

$$\{y_0, y_1\}$$

dont on note la fonction de répartition :

$$F(y_0, y_1; t_0, t_1) = P[y_0 < y_0, y_1 < y_1]$$

si on peut définir la fonction de répartition de toute suite :

correspondant aux époques $t_1, t_2, \dots, t_n, \dots$

$$F(y_0, y_1, \dots, y_n, \dots; t_0, t_1, \dots, t_n, \dots) =$$

$$P[y_0 < y_0, y_1 < y_1, \dots, y_n < y_n, \dots]$$

on a alors défini le processus aléatoire.

Le processus sera stationnaire si ces fonctions de répartition ont la propriété :

$$F(y_0, y_1, \dots, y_n; t_0, t_1, \dots, t_n) = F(y_0, y_1, \dots, y_n; t_0 + e, t_1 + e, \dots, t_n + e)$$

où e est une constante quelconque ajoutée à toutes les époques t_0, \dots, t_n .

En particulier on aura

$$F(y_0; t) = P[y_t < y_0] = F(y_0; t + l) = P[y_{t+l} < y_0]$$

y_0 fixé, la probabilité d'avoir une valeur observée du processus à la date t inférieure à y_0 , est égale à la probabilité d'avoir une valeur observée à la date $t + l$ inférieure à y_0 .

Une grandeur dont l'évolution est caractérisée par une croissance constante ne peut donc être stationnaire ; c'est le cas de grandeur possédant un *trend* ; les déviations par rapport à ce *trend* peuvent, elles, être stationnaires.

Si le processus est stationnaire, le temps n'intervient plus de façon absolue, mais relative, la corrélation entre deux valeurs du processus ne dépendra pas des deux dates d'observation de ces valeurs mais de la durée entre les deux observations.

On a coutume d'appeler

$$E[(y_t - E[y_t])(y_{t'} - E[y_{t'}])] = \Gamma(t, t')$$

la fonction de covariance du processus ; dans le cas de la stationnarité :

$$\Gamma(t, t') = \Gamma(|t - t'|)$$

On peut dire que les propriétés du processus ne changent pas si on fait subir au temps une translation.

De ce fait les caractéristiques de y_t , comme l'espérance mathématique ou la variance, sont constantes au cours du temps.

L'observation d'une seule réalisation d'un processus stationnaire permet une étude statistique, elle renseigne sur l'ensemble des réalisations possibles. Cette propriété est illustrée par l'ergodicité.

L'espérance mathématique $E[y_t]$ est égale à la limite :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n y_t$$

La moyenne d'une suite de valeurs appartenant à la réalisation d'un processus estime l'espérance mathématique d'une valeur du processus à un instant fixé.

Modèles linéaires et processus aléatoires. Quand on a défini un processus suivant un schéma autorégressif, on a utilisé un modèle linéaire très simple :

$$y_t = ay_{t-1} + u_t$$

$$E[u_t] = 0 \quad E[u_t u_{t'}] = 0 \quad t \neq t'$$

y_t est un processus dont chaque valeur ne dépend que de la valeur précédente. On peut faire dépendre y_t de plusieurs valeurs retardées :

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_h y_{t-h} + u_t$$

$$E[u_t] = 0 \quad E[u_t u_{t'}] = 0 \quad t \neq t'$$

Ces deux modèles rappellent celui de la régression, mais les variables exogènes sont remplacées par des valeurs retardées des variables endogènes.

Des valeurs données à des variables exogènes peuvent influencer aussi sur le processus, le modèle a alors la forme :

$$y_t = \sum_{j=1}^h a_j y_{t-j} + \sum_{i=1}^k b_i z_i + a_0 + u_t$$

u_t est un processus aléatoire ; dans les deux exemples précédents on supposait que ce processus était constitué d'une suite de réalisations indépendantes (on ne peut rien imaginer de plus simple) ; cette hypothèse peut être abandonnée, il faut alors spécifier le processus u_t . Par exemple : u_t est autorégressif :

$$u_t = \rho u_{t-1} + \eta_t$$

$$E[\eta_t] = 0 \quad E[\eta_t \eta_{t'}] = 0 \quad t \neq t'$$

En général, même si on observe des grandeurs à intervalles réguliers, l'évolution de ces grandeurs au cours du temps se fait

sur un ensemble continu de valeurs. Il serait donc plus précis de considérer des fonctions :

$$y = y(t) \quad t \in T$$

et des modèles de la forme :

$$y(t) = \int_{t-h}^t f(t-\tau)y(\tau)d\tau + u(t)$$

L'intégrale qui y figure est une intégrale de fonction aléatoire, ce qui diffère un peu de la notion classique d'intégrale. Si des quantités exogènes interviennent au cours des mêmes époques, on peut imaginer une relation :

$$y(t) = \int_{t-h}^t f(t-\tau)y(\tau)d\tau + \int_{t-h'}^t g(t-\tau)x(\tau)d\tau + u(t)$$

$u(t)$ est une fonction aléatoire dont il reste à préciser les caractéristiques.

Les fonctions $f(t-\tau)$ et $g(t-\tau)$ ont des rôles analogues à ceux des coefficients d'autorégression et de régression. La différence $t-\tau$ apparaît au lieu des deux variables séparées, t , τ , à cause de l'hypothèse de stationnarité. L'abandon de cette hypothèse conduirait à des modèles :

$$y(t) = \int_{t-h'}^t f(t,\tau)y(\tau)d\tau + \int_{t-h'}^t g(t,\tau)x(\tau)d\tau + u(t)$$

De tels modèles ont été très peu étudiés. On a davantage étudié les modèles d'autorégression, ou qui en dérivent ; à cause de leur ressemblance avec les modèles de régression dont l'étude est très complète, il est tentant d'appliquer les mêmes méthodes à ces modèles autorégressifs ; on ignore alors le fait que l'on travaille sur des processus aléatoires et non sur des variables aléatoires. On s'expose aussi à des erreurs.

La méthode des moindres carrés et le modèle :

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_n y_{t-n} + b_1 x_{1t} + b_2 x_{2t} + \dots + b_k x_{kt} + u_t$$

$$E[u_t] = 0 \quad D^2[u_t] = \sigma_u^2$$

$$E[u_t u_{t'}] = 0 \quad \text{si } t \neq t'$$

$$E[u_t y_{t-i}] = 0 \quad \text{si } i > 0$$

On suppose aussi que u_t est $\mathcal{N}(0, \sigma_u^2)$.

LES MODÈLES LINÉAIRES ET LES SÉRIES CHRONOLOGIQUES

Appliquer les méthodes du modèle de régression sur ce modèle, c'est associer les valeurs retardées de y à des variables exogènes :

$$y_{t-1} = x_{k+1, t}$$

$$y_{t-2} = x_{k+2, t}$$

$$y_{t-h} = x_{k+h, t}$$

et après m observations on a la relation matricielle :

$$Y = Xa + u$$

où

$$Y = \begin{bmatrix} y_m \\ \vdots \\ y_1 \end{bmatrix}$$

$$X = \begin{bmatrix} x_{1m} & \dots & x_{k+h, m} \\ \vdots & & \vdots \\ x_{11} & \dots & x_{k+h, 1} \end{bmatrix}$$

$$a = \begin{bmatrix} b_1 \\ \vdots \\ b_k \\ a_1 \\ \vdots \\ a_h \end{bmatrix}$$

La loi Φ de cet échantillon est conditionnée par la donnée de m valeurs initiales :

$$y_0, y_{-1}, \dots, y_{-m+1}$$

c'est une loi normale, et comme les u_i sont indépendants :

$$\Phi = \frac{1}{(2\pi\sigma_u^2)^{\frac{m}{2}}} \exp \left[-\frac{1}{2\sigma^2} (Y - Xa)'(Y - Xa) \right]$$

La méthode du maximum de vraisemblance conduit à la méthode des moindres carrés et on estime a par a^* solution de l'équation

$$X + Xa^* - X + Y = 0$$

$$a^* = (X + X)^{-1}X + Y$$

a^* sera cependant un estimateur biaisé. La présence de la variable endogène dans la matrice X entraîne :

$$E[a^*] \neq a$$

Pour mettre en évidence ce biais il est commode d'utiliser un modèle d'autorégression très simple :

$$y_t = ay_{t-1} + u_t \quad t = 1, \dots, m$$

u_t vérifiant les mêmes hypothèses, modèle que l'on compare au modèle de régression correspondant :

$$y_t = bx_t + u_t$$

l'estimé :

$$\hat{b} = \frac{\sum x_t y_t}{\sum x_t^2}$$

par les moindres carrés est une forme linéaire en y_t . En remplaçant y_t par sa valeur suivant le modèle :

$$\hat{b} = \frac{\sum x_t (bx_t + u_t)}{\sum x_t^2} = b + \frac{\sum x_t u_t}{\sum x_t^2}$$

comme les x_t ne sont pas considérés comme aléatoires

$$E\left[\frac{\sum x_t u_t}{\sum x_t^2}\right] = 0$$

$$E[\hat{b}] = b$$

par contre l'estimé :

$$a^* = \frac{\sum y_t y_{t-1}}{\sum y_{t-1}^2}$$

de a par les moindres carrés est un rapport de formes quadratiques en y , ce rapport de quantités aléatoires ne peut être manié facile-

ment. Rien ne permet de démontrer que $E[a^*] = a$, en général il y a un biais.

Ce biais pourrait infirmer dès à présent la méthode des moindres carrés, tout au moins sur des petits échantillons. L. Hurwicz dans une monographie de la Cowles Commission a étudié ce biais ; il a trouvé une erreur systématique de l'ordre de $\frac{a}{m}$ si m est la taille de l'échantillon.

Cependant, J. Durbin montre que la méthode des moindres carrés est encore la meilleure méthode d'estimation, car elle établit « l'équation d'estimation linéaire non biaisée de variance minimale ».

Il définit :

$$a^* \sum y_{t-1}^2 - \sum y_t y_{t-1} = 0$$

comme une « équation d'estimation linéaire non biaisée » puisque :

$$E \left[a \sum y_{t-1}^2 - \sum y_t y_{t-1} \right] = 0$$

où a est la valeur exacte du coefficient.

En effet :

$$y_t y_{t-1} = (a y_{t-1} + u_t) y_{t-1} = a y_{t-1}^2 + u_t y_{t-1}$$

donc :

$$a \sum y_{t-1}^2 - \sum y_t y_{t-1} = a (\sum y_{t-1}^2 - \sum y_{t-1}^2) + \sum u_t y_{t-1} = \sum u_t y_{t-1}$$

et :

$$E \left[a \sum y_{t-1}^2 - \sum y_t y_{t-1} \right] = E \left[\sum u_t y_{t-1} \right]$$

d'après nos hypothèses u_t est indépendant des valeurs passées de y_t d'où le résultat.

L'équation d'estimation est donc « non biaisée » ; puisque a^* y figure au premier degré, elle est appelée linéaire (par rapport à a^* et non par rapport aux valeurs observées).

Cette notion est analogue à celle d'estimateur non biaisé.

J. Durbin développe aussi une notion analogue à celle d'estimateur non biaisé, de variance minimale.

Il définit la variance de l'équation d'estimation par les moindres carrés par :

$$D^2 \left[a \frac{\sum y_{t-1}^2}{E[\sum y_{t-1}^2]} + \frac{\sum y_t y_{t-1}}{E[\sum y_{t-1}^2]} \right]$$

le coefficient de a , a dès lors une espérance mathématique égale à 1.

Toute autre équation d'estimation linéaire non biaisée sera de la forme :

$$C_1 a^* + C_2 = 0$$

où C_1 et C_2 sont des fonctions des valeurs observées.

$$E[C_1 a + C_2] = 0$$

si on impose à C_1 d'avoir une espérance mathématique égale à 1

$$E[C_1] = 1$$

on impose un choix, parmi toutes les équations équivalentes qui se déduisent par multiplication par une constante.

Sans cette façon de faire, la variance

$$D^2[C_1 a + C_2] = E[(C_1 a + C_2)^2]$$

ne serait plus définie de façon unique.

On comparera, alors, deux équations d'estimations linéaires non biaisées :

$$C_1 a^* + C_2 = 0 \quad E[C_1 a + C_2] = 0$$

$$C'_1 a^* + C'_2 = 0 \quad E[C'_1 a + C'_2] = 0$$

$$E[C_1] = E[C'_1] = 1$$

en comparant les variances :

$$D^2[C_1 a + C_2]$$

et

$$D^2[C'_1 a + C'_2]$$

on retiendra comme meilleure équation, celle dont la variance sera minimale.

J. Durbin montre que

$$D^2 \left[a \frac{\sum y_{t-1}^2}{E[\sum y_{t-1}^2]^2} - \frac{\sum y_t y_{t-1}}{E[\sum y_{t-1}^2]} \right]$$

est égale à la variance minimale que peut avoir une équation d'estimation linéaire, non biaisée.

La densité de probabilité des y_t est donnée par :

$$\Phi = \frac{1}{(2\pi\sigma_a^2)^{\frac{m}{2}}} \exp\left[-\frac{1}{2} \sum (y_t - a y_{t-1})^2\right]$$

L'espérance mathématique de l'équation s'écrit alors :

$$E[C_1 a + C_2] = 0 = \int (C_1 a + C_2) \Phi dy$$

où \int est une intégrale multiple, $dy = dy_1 \cdot dy_2 \cdot \dots \cdot dy_n$.

On dérive cette expression par rapport à a qui figure aussi dans la fonction Φ ;

$$\int C_1 \Phi dy + \int (C_1 a + C_2) \frac{\partial \Phi}{\partial a} dy = 0$$

or

$$\int C_1 \Phi dy = E[C_1] = 1$$

par hypothèse donc :

$$\int (C_1 a + C_2) \frac{\partial \Phi}{\partial a} dy = -1$$

ce qu'on peut écrire :

$$-1 = \int (C_1 a + C_2) \frac{\partial \Phi}{\partial a} \frac{\Phi}{\Phi} dy = \int (C_1 a + C_2) \frac{\partial \text{Log } \Phi}{\partial a} \Phi dy$$

La dernière intégrale est de la forme :

$$\int A \cdot B \cdot \Phi dy = E[A \cdot B]$$

$$A = C_1 a + C_2$$

où :

$$B = \frac{\partial \text{Log } \Phi}{\partial a}$$

or on sait, par l'inégalité de Schwartz, que :

$$(E[A \cdot B])^2 \leq E[A^2] \cdot E[B^2]$$

c'est-à-dire aussi que :

$$E[(C_1 a + C_2)^2] \cdot E\left[\left(\frac{\partial \text{Log } \Phi}{\partial a}\right)^2\right] \geq \left[\int [C_1 a + C_2] \frac{\partial \text{Log } \Phi}{\partial a} \Phi dy\right]^2$$

donc

$$E[(C_1 a + C_2)^2] \cdot E\left[\left(\frac{\partial \text{Log } \Phi}{\partial a}\right)^2\right] \geq + 1$$

$$E[(C_1 a + C_2)^2] \geq \frac{+ 1}{E\left[\left(\frac{\partial \text{Log } \Phi}{\partial a}\right)^2\right]}$$

mais

$$\frac{\partial \text{Log } \Phi}{\partial a} = \frac{1}{\Phi} \frac{\partial \Phi}{\partial a}$$

$$\left[\frac{\partial \text{Log } \Phi}{\partial a}\right]^2 = \frac{1}{\Phi^2} \left(\frac{\partial \Phi}{\partial a}\right)^2$$

tandis que

$$\frac{\partial^2 \text{Log } \Phi}{\partial a^2} = -\frac{1}{\Phi^2} \left(\frac{\partial \Phi}{\partial a}\right)^2 + \frac{1}{\Phi} \frac{\partial^2 \Phi}{\partial a^2}$$

or

$$\begin{aligned} \frac{\partial \Phi}{\partial a} &= \frac{+ 1}{(2\pi\sigma_u^2)^{\frac{m}{2}}} \exp\left[\Sigma (y_t - ay_{t-1})y_{t-1}\right] \left[-\frac{1}{2} \Sigma (y_t - ay_{t-1})^2\right] \\ &= \Phi \cdot [\Sigma y_t y_{t-1} - a \Sigma y_{t-1}^2] \end{aligned}$$

donc :

$$\int \frac{\partial \Phi}{\partial a} dy = \int \left[\Sigma y_t y_{t-1} - a \Sigma y_{t-1}^2\right] \Phi dy = 0$$

comme étant l'espérance mathématique de l'équation d'estimation des moindres carrés.

Donc, en dérivant par rapport à a :

$$\int \frac{\partial^2 \Phi}{\partial a^2} dy = 0 = \int \left(\frac{1}{\Phi} \frac{\partial^2 \Phi}{\partial a^2}\right) \Phi dy = E\left[\frac{1}{\Phi} \frac{\partial^2 \Phi}{\partial a^2}\right]$$

on peut alors voir que :

$$E \left[\frac{\partial \text{Log } \Phi}{\partial a} \right]^2 = - E \left[\frac{\partial^2 \text{Log } \Phi}{\partial a^2} \right]$$

et donc que :

$$D^2 [(C_1 a + C_2)^2] \geq \frac{-1}{E \left[\frac{\partial^2 \text{Log } \Phi}{\partial a^2} \right]}$$

comme

$$-\frac{\partial^2 \text{Log } \Phi}{\partial a^2} = \sum y_{t-1}^2$$

$$D^2 [(C_1 a + C_2)^2] \geq \frac{1}{\sum y_{t-1}^2}$$

Revenons à l'expression :

$$D^2 \left[\frac{\sum y_{t-1}^2}{E[\sum y_{t-1}^2]} a - \frac{\sum y_t y_{t-1}}{E[\sum y_{t-1}^2]} \right] = d^2$$

qui est l'expression de la variance de l'équation des moindres carrés.

On peut l'écrire aussi

$$d^2 = \frac{1}{\{E[\sum y_{t-1}^2]\}^2} E[(\sum y_{t-1}^2 a - \sum y_t y_{t-1})^2]$$

or

$$a \sum y_{t-1}^2 - \sum y_t y_{t-1} = - \frac{\partial \text{Log } \Phi}{\partial a}$$

par définition, donc

$$\begin{aligned} d^2 &= \frac{1}{\{E[\sum y_{t-1}^2]\}^2} E \left[\left(\frac{\partial \text{Log } \Phi}{\partial a} \right)^2 \right] \\ &= \frac{1}{\{E[\sum y_{t-1}^2]\}^2} E[\sum y_{t-1}^2] = \frac{1}{E[\sum y_{t-1}^2]} \end{aligned}$$

Ce qui est justement la plus petite valeur que peut prendre la variance. Les moindres carrés donnent alors ce que Durbin appelle : « la meilleure équation d'estimation linéaire, non biaisée ».

Si la méthode garde ce caractère d'optimalité, il n'en reste pas moins que a^* est un estimé biaisé de a .

Cette notion d'équation linéaire non biaisée est étendue au cas plus général, constitué par le modèle qui a été présenté au début du paragraphe.

$$X^+Xa - X^+Y = 0$$

est un ensemble d'équations d'estimations linéaires, non biaisées et de variances minimales.

Asymptotiquement, quand le nombre des observations augmente indéfiniment, si X^+X tend à devenir une matrice régulière, le biais des estimés de a tend à disparaître, et a^* tend en probabilité vers a .

Aussi quand on dispose d'un échantillon important, l'ensemble de la méthode des moindres carrés est applicable.

Avoir un large échantillon demande des observations sur une période assez longue, l'hypothèse de stationnarité est alors difficile à maintenir.

En pratique des échantillons de taille 20 ou 30 sont très fréquents ; dans quelle mesure les procédés appliqués à un modèle de régression, donnent-ils encore des résultats fiables pour un modèle autorégressif, (ou qui en dérive), c'est la question qui se pose.

Devant les difficultés d'une approche théorique, E. Malinvaud a étudié un modèle

$$y_t = ay_{t-1} + bx_t + c + u_t$$

où x_t est une variable exogène, de façon expérimentale. Ce modèle est de la même forme que le précédent. Des échantillons artificiels de taille réduite ont été formés suivant une méthode de Monte Carlo.

— Les paramètres du modèle sont précisés.

— La loi de u_t aussi.

Une suite de 25 valeurs x_t étant fixée, 200 échantillons de taille 25 de valeurs y_t , y correspondant suivant le modèle, ont été construits en donnant aux u_t des valeurs vraisemblables. À chaque fois, les 20 premières valeurs de l'échantillon servaient à estimer les paramètres, les 5 dernières à éprouver la précision qui était faite sur la foi de ces estimés.

Une autre suite de 25 valeurs x_t , et d'autres échantillons construits à partir d'elle, permet de tester l'importance des valeurs de la variable exogène sur les propriétés des estimés.

— Trois lois de probabilité pour les u_t ont été choisies :
 les deux premières supposent que les u_t sont indépendants et suivent
 soit :

— une loi normale : $\mathcal{N}(0, 6)$

soit :

— une loi binomiale à deux valeurs :

$$\begin{array}{c|cc} u_t & -3\sqrt{2} & 6\sqrt{2} \\ \hline \rho u_t & \frac{2}{3} & \frac{1}{3} \end{array}$$

La troisième suppose que les u_t constituent un processus autorégressif simple :

$$u_t = \rho u_{t-1} + \eta_t$$

où $\rho = \frac{1}{2}$ (l'autocorrélation est moyenne)

les η_t sont indépendants et suivent une loi normale :

$$\mathcal{N}(0, 3\sqrt{3})$$

De ces trois lois, seule la première correspond aux hypothèses faites par J. Durbin. La seconde est très éloignée de la loi normale, on peut donc voir si les résultats sont affectés par des changements dans la forme de la distribution des perturbations.

Il est apparu que pour les deux premières de ces lois, a^* estimé par les moindres carrés de a , était « indiscutablement biaisé » ; alors que a avait pour valeur 0.6, sur la base des 200 échantillons, on pouvait dire que à 95 p.c.

$$0.500 \leq E[a^*] \leq 0.535$$

et que a^* sous-estimait a à trois fois sur quatre.

— L'estimation de σ_u^2 par σ^{*2}_u , estimé des moindres carrés, ne comportait pas de biais valables.

— L'estimation de la variance de l'estimé a^* , σ_a^{*2} semblait biaisée légèrement supérieurement et semblait aussi dépendre des valeurs de la suite x_t .

— Le biais de a^* , celui de σ_a^{*2} , le fait que cette variance estimée peut même ne plus suivre une loi du χ^2 , entraîne que le seuil de

probabilité d'un intervalle de confiance construit par le t de Student est supérieur à 5 p.c. : plus d'une fois sur vingt, la valeur exacte a se trouve hors de cet intervalle.

Nous renvoyons le lecteur intéressé à l'article de E. Malinvaud pour l'étude des erreurs de prévision, mais il est intéressant de savoir qu'en guise de conclusion celui-ci donne la méthode des moindres carrés comme approximativement valable, dans le cas d'un modèle représentant les situations habituelles dans les études économiques ; une hypothèse est cependant fondamentale, c'est l'indépendance des perturbations.

On se souvient que la troisième loi des perturbations les définissait comme un processus autorégressif d'ordre 1. Ce processus est un des plus simples imaginables ; cependant, une telle liaison entre les erreurs successives entraîne généralement une grande perte d'efficacité pour la méthode des moindres carrés.

Si on considère le modèle :

$$y_t = ay_{t-1} + u_t$$

$$u_t = \rho u_{t-1} + \eta_t$$

on peut aussi écrire :

$$u_{t-1} = y_{t-1} - ay_{t-2}$$

donc

$$y_t = ay_{t-1} + \rho u_{t-1} + \eta_t = (a + \rho)y_{t-1} - \rho ay_{t-2} + \eta_t$$

$$y_t = (a + \rho)y_{t-1} - \rho ay_{t-2} + \eta_t$$

on a fait une erreur de spécification du modèle en le considérant comme autorégressif d'ordre 1.

L'estimé des moindres carrés de a , à partir du premier modèle :

$$a^* = \frac{\sum y_t y_{t-1}}{\sum y_{t-1}^2}$$

s'écrit en tenant compte du modèle autorégressif d'ordre 2.

$$a^* = a + \rho - \rho a \frac{\sum y_{t-1} y_{t-2}}{\sum y_{t-1}^2} + \frac{\sum \eta_t y_{t-1}}{\sum y_{t-1}^2}$$

quand le nombre m d'observations tend vers l'infini

$$\frac{\sum y_{t-1} y_{t-2}}{\sum y_{t-1}^2} \longrightarrow \frac{\sum y_t y_{t-1}}{\sum y_{t-1}^2} \longrightarrow \text{Lim } a^*$$

donc :

$$\text{Lim } a^* = a + \rho - a\rho \text{ Lim } a^* + \text{Lim } \frac{\sum \eta_t y_{t-1}}{\sum y_{t-1}^2}$$

La limite est prise au sens de limite en probabilité. La dernière limite est nulle et :

$$\text{Lim } a^* = \frac{a + \rho}{1 + a\rho}$$

valeur différente de a si $\rho \neq 0$

le biais subsiste asymptotiquement. Ce biais asymptotique rend illusoire la formation d'intervalles de confiance en suivant la loi de Student. On peut estimer simultanément a et ρ , les méthodes proposées conduisent à des estimés convergents vers les valeurs exactes. Ces méthodes séduisantes sont souvent peu utiles car il suffit que la spécification du processus des perturbations soit mauvaise (qu'il ne soit pas autorégressif d'ordre 1, donc que les η_t soient encore autocorrélés), pour que ces méthodes mènent à des erreurs aussi considérables. L'étude expérimentale d'un tel modèle, entreprise par E. Malinvaud, montre que les estimés de a sur de petits échantillons par a^* sont mauvais. Les méthodes estimant simultanément a et ρ n'ont pas été plus fructueuses, sauf dans l'application à la prévision.

Les moindres carrés restent cependant utilisables, l'ajustement direct par cette méthode ne pouvant être remplacé par un meilleur procédé.

En conclusion, des modèles linéaires dynamiques sont souvent soumis aux mêmes méthodes d'estimations que des modèles de régression, bien que ces méthodes soient approximativement valables on doit faire des réserves sur la confiance que l'on peut mettre dans les résultats.

De nombreux modèles particuliers ont été étudiés, dont nous n'avons pas parlé. On peut consulter la monographie de la Cowles

Commission pour se rendre compte de la complexité du sujet. Enfin, on n'a pas fait mention de l'analyse spectrale des processus aléatoires, cela nous aurait entraîné trop loin. C'est cependant une façon prometteuse d'aborder les problèmes de liaisons entre processus aléatoires.

Alain HAURIE,
professeur à l'École des
Hautes Études commerciales
(Montréal)

NOTE BIBLIOGRAPHIQUE

- MURRAY ROSENBLATT, *Random Processes*, Oxford University Texts in Mathematical Science.
- E. MALINVAUD, *Méthodes statistiques de l'économétrie*, Dunod, Paris.
- E. MALINVAUD, « Estimation et prévisions dans les modèles économiques autorégressifs », *Revue de l'Institut international de Statistique*, vol. 29, no 2.
- J. DURBIN, « Estimation of Parameters in Time Series Regression Models », *Journal of the Royal Statistical Society, série B*, vol. 22, no 1.
- J. JOHNSTON, *Econometric Methods*, McGraw-Hill Book Company, New York, 1963.
- T.-C. KOOPMANS (éd.), *Statistical Inference in Dynamic Economic Models*, Cowles Commission, John Wiley & Sons, Inc., New York, 1950.