L'Actualité économique

L'ACTUALITÉ ÉCONOMIQUE
REVUE D'ANALYSE ÉCONOMIQUE

Modèles pour l'étude de quelques types simples de séries chronologiques

R. Feron

Volume 40, Number 1, April–June 1964

URI: https://id.erudit.org/iderudit/1002823ar
DOI: https://doi.org/10.7202/1002823ar

See table of contents

Publisher(s)

HEC Montréal

ISSN

0001-771X (print)
1710-3991 (digital)

Explore this journal

Cite this article

Feron, R. (1964). Modèles pour l'étude de quelques types simples de séries chronologiques. *L'Actualité économique*, *40*(1), 40–58. https://doi.org/10.7202/1002823ar

Modèles pour l'étude de quelques types simples de séries chronologiques

Nous nous proposons ici d'étudier les phénomènes économiques sur lesquels les saisons, c'est-à-dire la période de l'année où on se trouve, ont une influence marquée. De tels phénomènes sont assez bien connus à l'aide des renseignements statistiques dont on dispose. On constate toujours que la grandeur étudiée est mesurée à intervalles réguliers. Cet intervalle est généralement le mois, comme c'est le cas pour le nombre de voyageurs-kilomètres pour les lignes autres que celles de banlieue, mais ce peut être aussi la semaine, comme dans les relevés du nombre de cartes hebdomadaires pour le métro de Paris, la journée, comme dans les relevés du nombre de centimètres de pluie dans un lieu déterminé, ou même l'heure, comme dans l'étude de la consommation de l'énergie électrique.

Ici, nous sommes placés dans de beaucoup moins bonnes conditions d'étude que ne l'est le physicien. De nombreuses raisons font que les causes dont nous ne tenons pas compte dans le modèle économique initial ont une très grosse importance. Ainsi, le nombre de voyageurs-kilomètres des lignes de chemin de fer autres que celles de banlieue est donné par le tableau suivant.

On constate que la courbe représentative a une allure grossièrement périodique, mais il y a de nombreuses causes de perturbations qui influent sur le nombre de voyageurs, causes dont il est difficile de tenir compte. Il semble, en effet, que c'est pendant les vacances que les gens voyagent le plus. Mais, si le désir de partir en congé est la cause dominante, il est clair que d'autres phénomènes secon-

Mois	1946	1947	1948	1949	Mois	1946	1947	1948	1949
Janvier	1,560	1,640	1,560	1,400	Juillet	3,000	3,300	3,450	3,400
Février	1,640	1,520	1,500	1,400	Août	3,300	3,400	3,400	3,500
Mars	1,560	1,800	2,000	1,500	Septembre	2,800	2,800	2,600	2,600
Avril	2,200	2,400	1,340	2,300	Octobre	2,000	2,000	1,900	1,900
Mai	1,960	2,300	2,200	1,800	Novembre	1,760	1,560	1,400	1,560
Juin	2,200	2,200	2,200	2,300	Décembre	1,960	1,500	1,760	1,760

daires, difficilement contrôlables, comme le temps qu'il fait au moment des vacances des intéressés et la variation de leur niveau de vie, ont aussi une grande influence. D'autres phénomènes dont on peut mathématiquement tenir compte comme la mobilité des fêtes de Pâques et de Pentecôte, ne laissent pas d'avoir une importance très grande sur le trafic des mois considérés.

Les mêmes phénomènes s'observent également dans le tableau suivant qui représente la consommation d'électricité d'un particulier.

À première vue, il semble curieux de constater que cette consommation présente une légère recrudescence en avril. Ceci est sans doute dû au fait que, le chauffage de l'immeuble ayant cessé, le particulier avait été amené à utiliser le chauffage électrique durant quelques journées encore froides.

L'importance des termes dont nous ne pouvons tenir compte lorsque nous proposons un modèle économique fait que si $X(t)$

Année	1948	1949	1950	1951	Totaux	Moyennes mensuelles	Ecart
Janvier	1,634	1,908	2,388	2,082	8,012	2,003	+ 296
Février	1,586	1,851	2,525	1,794	7,756	1,939	+ 232
Mars	1,321	1,709	2,284	1,722	7,036	1,759	+ 52
Avril	1,464	1,815	2,400	1,657	7,336	1,834	+ 127
Mai	1,316	1,640	2,019	1,285	6,260	1,565	− 142
Juin	1,315	1,493	1,986	1,322	6,116	1,529	− 178
Juillet	1,210	1,471	1,790	1,289	5,760	1,440	− 267
Août	1,200	1,349	1,487	1,272	5,308	1,327	− 380
Septembre	1,250	1,588	1,566	1,428	5,832	1,458	− 249
Octobre	1,545	1,825	1,918	1,572	6,860	1,715	+ 8
Novembre	1,766	2,178	1,978	1,722	7,644	1,911	+ 204
Décembre	1,767	1,477	1,875	1,873	7,992	1,998	+ 291
Totaux	17,374	21,304	24,216	19,018	81,912	20,478	
Moyenne annuelle	1,448	1,775	2,018	1,585	6,826	$\boxed{1,707}$	Moyenne générale

est la valeur observée et $a(t)$ la valeur donnée par la théorie, on aura $X(t) = a(t) + \varepsilon$ où ε est relativement grand. Ce fait entraîne que, quelle que soit la méthode employée, les paramètres qui figurent dans $a(t)$ seront estimés d'une manière peu précise et explique que l'étude détaillée des séries temporelles économiques ait été beaucoup plus tardive que, par exemple, celle des séries temporelles astronomiques.

Cette étude est même beaucoup plus délicate que celle de certains phénomènes physiques comme, par exemple, celle du bruit de fond. Car en économétrie on ne dispose jamais, dans les phénomènes présentant certaines causes de périodicité, que de séries s'étendant sur un nombre assez faible de périodes ou de pseudopériodes, alors qu'en physique on pourra généralement faire porter les observations sur un nombre de périodes pratiquement très grand.

En effet, il est clair, par exemple, que durant la période de guerre qui a précédé de peu les données relatives au trafic ferroviaire et à la consommation d'électricité, ces dernières données étaient soumises à d'autres espèces de contraintes et n'obéissaient plus à la même loi. De ces considérations résulte qu'on ne pourra pas appliquer à l'économétrie, comme on le fait couramment en physique, les théorèmes asymptotiques du calcul des probabilités.

Aussi, ne pouvons-nous espérer obtenir des modèles efficaces que si ces derniers sont extrêmement simples.

En ce qui concerne les phénomènes étudiés, nous examinerons successivement les modèles suivants : modèle périodique certain, modèle périodique stochastique, modèle périodique à tendance, modèles markoviens de Morlat et autres.

*
* *

Modèle périodique certain. — De même que c'était le cas en astronomie, on peut tout d'abord supposer que le phénomène peut être représenté par une fonction périodique $a(t)$ certaine de période T (généralement connue et égale à une année). Cette manière d'interpréter les faits, la plus simple qu'on puisse concevoir, semble être indispensable lorsqu'on a affaire à un système économique assez complexe dans lequel la variable envisagée peut être considérée comme une variable exogène.

Il n'y aura évidemment pas d'inconvénient à supposer que la fonction $a(t)$ est à variation totale bornée dans un intervalle période et, dès lors, on sait que $a(t)$ est développable en série de Fourier que l'on peut écrire :

(1)

$$a(t) = A_0 + A_1 \cos \frac{2\pi t}{T} + B_1 \sin \frac{2\pi t}{T} + \dots + A_n \cos \frac{2\pi nt}{T} + B_n \sin \frac{2\pi nt}{T} + \dots$$

avec :

$$A_0 = \frac{1}{T} \int_0^T a(t)dt$$

$$A_n = \frac{2}{T} \int_0^T a(t) \cos \frac{2\pi nt}{T}\, dt$$

$$B_n = \frac{2}{T} \int_0^T a(t) \sin \frac{2\pi nt}{T}\, dt$$

ou sous la forme équivalente :

$$a(t) = \sum_{-\infty}^{+\infty} C_n e^{\frac{2i\pi nt}{T}}$$

avec évidemment :

$$C_n = \frac{1}{T} \int_0^T a(t)\, e^{-\frac{2i\pi nt}{T}} dt$$

On montre que ces séries de Fourier sont semi-convergentes en tout point t de la droite réelle et que leur somme a pour valeur la moyenne arithmétique de la valeur limite à gauche $f(x - 0)$ et de la valeur limite à droite $f(x + 0)$:

(2)
$$\lim_{N \to \infty} \sum_{n = N} C_n e^{\frac{2i\pi nt}{T}} = \frac{1}{2} [a(t - 0) + a(t + 0)]$$

En particulier si $a(t)$ (toujours supposée à variation bornée) est continue la série de Fourier en ce point semi-converge vers $a(t)$.

De plus, si en tout point t d'un intervalle fermé $[\alpha, \beta]$, $a(t)$ est continue, la semi-convergence est uniforme dans $[\alpha, \beta]$.

Il est remarquable de constater que les $2n + 1$ premiers coefficients A_i et B_i (resp. c_i) peuvent être obtenus en minimisant la quantité.

$$\int_0^T \left[a(t) - \left(A_0 + A_1 \cos\frac{2\pi t}{T} + B_1 \sin\frac{2\pi t}{T} + \ldots \right.\right.$$
$$\left.\left. + A_n \cos\frac{2\pi nt}{T} + B_n \sin\frac{2\pi nt}{T}\right)\right]^2 dt$$

$$\left(\text{resp.} \int_0^T \left[a(t) - \sum_{-n}^{n} C_k e^{\frac{2i\pi kt}{T}}\right]^2 dt,\right.$$

c'est-à-dire en essayant d'ajuster $a(t)$ par la méthode des moindres carrés au moyen du début d'un développement de la forme (1) ou (3). Il paraît donc normal de considérer la quantité :

$$d(a,b) = \int_0^T [a(t) - b(t)]^2 dt$$

comme une mesure de l'écart entre les fonctions périodiques de période T, $a(t)$ et $b(t)$.

D'autre part, on sait d'après l'égalité de Bessel-Parseval que si $a(t)$ est de carré sommable, alors la série de ses coefficients de Fourier est de carré sommable et on a :

$$(3) \qquad \int_0^T a^2(t)\frac{dt}{T} = \sum_{-\infty}^{+\infty} C_K^2 = A_O^2 + \sum_{K=1}^{+\infty} \frac{A_K^2 + B_K^2}{2}$$

Ceci nous permet de donner un indice indiquant la qualité d'un ajustement de $2n + 1$ termes ; posons :

$$S_n = \sum_{-n}^{+n} C_K^2 = A_0^2 + \sum_{1}^{n} \frac{A_K^2 + B_K^2}{2}$$

(4) Alors l'indice :

$$I_n = \frac{S_n}{\int_0^T a^2(t)\frac{dt}{t}}$$

nous donne une idée de la qualité de l'ajustement au moyen de $2n + 1$ termes. On pourra s'arrêter dans le développement si I_n est suffisamment voisin de 1.

Dans la pratique, on ne connaît d'ailleurs que les valeurs prises par $a(t)$ en un nombre fini de points. Supposons, par exemple, que nous sachions que $a(t)$ prend aux $2n + 1$ instants $t_0, t_1, \ldots, t_{2n}$ les valeurs $y_0, y_1, \ldots, y_{2n}$. Alors, il est clair que nous pourrons faire passer une infinité de fonctions périodiques de période 2π par ces $2n + 1$ points.

Il en est toutefois une qui est plus simple que toutes les autres : celle qui est de la forme :

(5)
$$f_n(t) = a_0 + a_1 \cos t + b_1 \sin t + \ldots + a_n \cos nt + b_n \sin nt$$

Je dis, en effet, que si les t_k sont compris entre 0 et 2π et distincts, il existe une et une seule fonction $f_n(t)$, c'est-à-dire que le système d'équations :

(6)
$$\begin{aligned} a_0 + a_1 \cos t_0 + b_1 \sin t_0 + \ldots + a_n \cos nt_0 + b_n \sin nt_0 &= y_0 \\ a_0 + a_1 \cos t_1 + b_1 \sin t_1 + \ldots + a_n \cos nt_1 + b_n \sin nt_1 &= y_1 \\ a_0 + a_1 \cos t_{2n} + b_1 \sin t_{2n} + \ldots + a_n \cos nt_{2n} + b_n \sin nt_{2n} &= y_{2n} \end{aligned}$$

a toujours une solution unique.

Ou, ce qui revient au même, que le déterminant :

$$\Delta = \begin{vmatrix} 1 \cos t_0 \sin t_0 \ldots\ldots \cos nt_0 \sin nt_0 \\ 1 \cos t_1 \sin t_1 \ldots\ldots \cos nt_1 \sin nt_1 \\ 1 \cos t_{2n} \sin t_{2n} \ldots \cos nt_{2n} \sin nt_{2n} \end{vmatrix}$$

est toujours différent de zéro.

En effet, posons :

(7)
$$\varphi_k(t) = \frac{\sin \frac{1}{2}(t-t_0) \ldots \sin \frac{1}{2}(t-t_{k-1}) \sin \frac{1}{2}(t-t_{k+1}) \ldots \sin \frac{1}{2}(t-t_{2n})}{\sin \frac{1}{2}(t_k-t_0) \ldots \sin \frac{1}{2}(t_k-t_{k-1}) \sin \frac{1}{2}(t_k-t_{k+1}) \ldots \sin \frac{1}{2}(t_k-t_{2n})}$$

on a : $\varphi_k(t_r) = 0$ si $r \neq k$ et $\varphi_k(t_r) = 1$ si $r = k$,

donc : (8)
$$g_{2n}(t) = \sum_{k=0}^{2n} y_k \varphi_k(t)$$

qui est, comme on le montre facilement, de la forme (5) est telle que : $g_{2n}(t_r) = y_r$.

Le système a donc une solution quelles que soient les valeurs y_r donc Δ est toujours différent de 0 et la solution est unique.

*
* *

Modèle périodique stochastique. — Deux cas peuvent se présenter : ou bien la période est connue (généralement l'année), ou bien elle est inconnue et on se propose précisément de la déterminer.

Si la période T est connue. Dans ce cas, on n'observe généralement pas la variable étudiée à tout instant au cours du temps, mais à intervalles réguliers qui sont le plus souvent des sous-multiples de la période T. Ces dernières circonstances ne sont évidemment pas nécessaires pour que l'on puisse mener à bien les calculs qui sont faits sur de telles séries, mais elles permettent de les simplifier grandement ; aussi nous placerons-nous d'emblée dans le cas où elles ont lieu.

Alors, si nous prenons pour unité l'intervalle de temps qui sépare 2 observations consécutives, nous aurons des observations $X(1)$, $X(2)$, ..., $X(n)$ aux époques entières.

Il est commode de grouper les données en table de Buys-Ballot. On a :

$$\begin{vmatrix} X(1) \ldots\ldots X(2) & \ldots\ldots\ldots\ldots\ldots\ldots & X(T) \\ X(T+1)\ X(T+2) & \ldots\ldots\ldots\ldots\ldots\ldots & X(2T) \\ \vdots & & \vdots \\ X[(i-1)T+1] \ldots X[(i-1)T+j] & \ldots\ldots & X(iT) \\ \vdots & & \\ X[(s-1)T+1] \ldots X[(s-1)T+p] & & \end{vmatrix}$$

L'analyse du phénomène est évidemment encore un peu simplifiée lorsque les observations ont porté sur un nombre entier de périodes, la dernière observation étant $X[sT]$.

Nous supposerons toutes ces conditions réalisées et poserons :

$$[X_{ij} = X[(i-1)T + j]$$

On peut évidemment toujours écrire :

$$X_{ij} = y_j + \varepsilon_{ij} \tag{9}$$

où y_j est une quantité certaine qui dépend uniquement de la période considérée et où ε_{ij} est une certaine variable aléatoire de moyenne nulle.

Dire que nous avons un modèle périodique stochastique revient à dire que ε_{ij} dépend peut être de j mais pas de i.

On suppose d'ailleurs généralement que ε_{ij} ne dépend ni de i, ni de j. Nous dirons dans ce cas que nous avons un modèle stochastique *du type A*.

1) *Modèle stochastique le plus général.* Le problème consistera essentiellement à estimer les y_j au moyen des observations X_{ij}. Or, nous avons vu dans le cas certain que la méthode des moindres carrés était particulièrement bien adaptée aux séries de Fourier. Aussi paraît-il naturel d'ajuster les y_i par la méthode des moindres carrés et de prendre pour y_i les valeurs $\hat{y}_j$ données par :

$$\sum_{i=1}^{s} \sum_{j=1}^{T} (X_{ij} - \hat{y}_j)^2 = \text{minimum}, \quad \hat{y}_j = \frac{1}{s} \sum_{i=1}^{s} X_{ij}$$

On peut aisément estimer la variance des ε_j par :

(10)
$$\sigma^2_{\varepsilon_j} = \frac{1}{s-1} \sum_i [X_{ij} - \hat{y}_j]^2$$

et celle des y_i s'en déduit. Il est dès lors aisé, si nous supposons que les ε_j obéissent à une loi de Laplace-Gauss de trouver des limites de confiance pour les y_i en appliquant le test t.

Il est clair que l'on peut tester diverses hypothèses, par exemple, que les y_i ne sont pas stochastiquement identiques. On peut aussi tester ce modèle contre des modèles plus généraux.

On peut se proposer de faire passer une fonction périodique continue le plus près possible des points-observations, c'est-à-dire que l'on cherchera à minimiser :

(11)
$$\mathcal{Q} = \sum_l \sum_j \left\{ X[(l-1)T + j] - \sum_{-\infty}^{+\infty} C_k e^{\frac{2ik\pi}{T}[(l-1)T+j]} \right\}^2$$

ou :
$$\sum_j \left\{ \sum_l \left[[X(l-1)T + j] - \sum_{-\infty}^{+\infty} C_k e^{\frac{2ik\pi}{T} j} \right]^2 \right\}$$

d'où l'on tire :

(12)
$$\hat{y}_j - \sum_{k:-\infty}^{k:+\infty} C_k e^{\frac{2i\pi kj}{T}} = 0,$$

ce qui montre que l'on obtient par cette méthode précisément une courbe qui passe par les $\dot{y}_j$. Or, nous savons qu'il existe une infinité de telles courbes. Nous avons vu toutefois que si T est impair ($T = 2p + 1$) l'une d'elles joue un rôle privilégié, celle qui est de la forme :

$$a(t) = \sum_{-p}^{+p} C_k e^{\frac{2i\pi kt}{T}} \tag{13}$$

Si T est pair ($T = 2p$), nous aurons aussi une courbe de la forme :

$$a(t) = \sum_{-p}^{+p} C_k e^{\frac{2i\pi kt}{T}}.$$

Les séries $a(t)$ se calculent très aisément au moyen de (12) dans lesquels k varie seulement de $-p$ à $+p$.

À cet effet, on remarque que :

$$\sum_{1}^{T} e^{2i\pi k \frac{j}{T}} = e^{2i\pi \frac{k}{T}} \frac{e^{2i\pi k \frac{T}{T}} - 1}{e^{\frac{2i\pi k}{T}} - 1}$$

de sorte que si $k < T$ cette somme est nulle.

Il en résulte que :

$$C_0 = \frac{1}{T} \sum \dot{y}_j ; C_1 = \frac{1}{T} \sum y_j \, e^{\frac{-2i\,y_j}{T}} \text{ etc....}$$

De même, les a et les b seraient donnés par des analogues des formules de Fourier.

Il y a toutefois une petite difficulté dans le cas où T est pair pour $C_{-1/2}$ et $C_{+1/2}$ car en multipliant toutes les équations (12) par $e^{-2i\pi \frac{T}{2T}}$ (resp. $e^{2i\pi \frac{T}{2T}}$ le coefficient de $C_{-\frac{T}{2}}$ prend la forme $\sum_{1}^{T} e^{-2i\pi j}$ et vaut T et nous constatons qu'il y a une certaine indétermination dans la valeur de $C_{\frac{T}{2}}$ et $C_{-\frac{T}{2}}$.

Ceci est d'ailleurs normal puisque nous avons vu qu'en choisissant convenablement ces deux coefficients nous pouvons faire passer la courbe par un point supplémentaire arbitraire. Si T est relativement grand par rapport à l'unité choisie (ce sera le cas si, par exemple, la période étant l'année, on possède des points-observations tous les mois ou tous les jours), on ne prendra généralement pas les p $\left(= \frac{T}{2} \text{ ou } \frac{T-1}{2}\right.$ termes du développement), mais on prendra :

$$a_N(t) = \sum_{-N}^{+N} C_k e^{\frac{2ik\pi t}{T}}$$

avec $N < p$ et on se considérera comme satisfaits quand $\sum [a_N(j) - y_j]^2$ sera petit par rapport à $\mathcal{Q}$.

2) *Modèle stochastique du type A.* On suppose alors que $x_{ij} = x\,[(i-1)\,T + j] = y_i + \varepsilon$ où ε est cette fois une variable aléatoire de moyenne nulle qui ne dépend pas de j.

On cherchera encore des estimations $\hat{y}_j$ de y, par la méthode des moindres carrés et on trouvera $\hat{y}_j = \frac{1}{s} \sum_{i=1}^{s} X_{ij}$, et on prendra comme estimation de σ_ε^2 la quantité :

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{T(s-1)} \sum_i \sum_j (X_{ij} - \hat{y}_j)^2 = \frac{1}{T} \sum \sigma_{\varepsilon_j}^2$$

Il est alors aisé, si nous supposons que ε obéit à une loi de Laplace-Gauss, de trouver des limites de confiance pour les y_i en appliquant le test t à $(s - 1)\,T$ degrés de liberté.

L'ajustement d'une série de Fourier se fait exactement comme précédemment.

Il est, en effet, fort encombrant de traîner dans tous les calculs des ε_j et on a généralement tendance à adopter un modèle du type A. Cette manière de voir est-elle justifiée par les faits ?

Pour nous en rendre compte, calculons les $\hat{y}_j$ et $\hat{\sigma}_{\varepsilon_j}^2$ sur la série donnant le nombre de voyageurs-kilomètre pour les lignes autres que les lignes de banlieue.

On trouve le tableau suivant :

Mois	Janvier	Février	Mars	Avril	Mai	Juin
Y_j	1,540	1,515	1,715	2,050	2,065	2,225
σ_j^2	10,133	9,700	52,900	237,066	51,566	2,500

Mois	Juillet	Août	Septembre	Octobre	Novembre	Décembre
Y_j	3,285	3,400	2,700	1,950	1,570	1,745
σ_j^2	40,633	6,666	13,333	3,333	21,733	35,566

Il n'y a pas lieu ici de faire le test d'hétérogénéité des variances.

Le tableau confirme notre opinion à priori, à savoir que la variance varie avec le mois envisagé et que ce fait est dû à la variabilité de la date de Pâques. Le même phénomène s'observe en ce qui concerne la quantité d'électricité consommée par un particulier.

Le modèle du type *A* nous apparaîtra donc le plus souvent comme une simplification difficilement acceptable et il sera souhaitable de creuser le modèle le plus général, ce qui ne veut pas dire que l'on ne continuera pas encore longtemps à l'appliquer. Le même phénomène a d'ailleurs lieu également en ce qui concerne la plupart des applications pratiques de l'analyse de la variance.

Si la période T est inconnue. Supposons que nous sachions que $X(t)$ est de la forme $X(t) = C(t) + \varepsilon_t =$

$$\sum C_k e^{it\lambda_k} + \varepsilon_t$$

Il est clair que :

$$(15) \qquad f(\lambda) = \frac{1}{\sqrt{n}} \sum_{1}^{n} X(t) e^{-i\lambda t}$$

$$= \frac{1}{\sqrt{n}} \sum_{-\infty}^{+\infty} C_k e^{i(\lambda_k - \lambda)} \frac{e^{in(\lambda_k - \lambda)} - 1}{e^{i(\lambda_k - \lambda)} - 1} + \frac{1}{\sqrt{n}} \sum_{1}^{n} \varepsilon_t e^{-it\lambda}$$

si $\lambda \neq \lambda_k$.

Le premier terme tend vers 0 comme $\frac{1}{\sqrt{n}}$ alors que le second a pour moyenne 0 et une variance de l'ordre de la variance des ε_t au contraire, si $\lambda = \lambda_k$ cette quantité est de l'ordre de $\sqrt{n}\ C_k$. Il en résulte que si n est assez grand (ce qui a presque toujours lieu en physique, mais beaucoup plus rarement en économétrie) $f(\lambda)$ présentera des maxima marqués pour les $\lambda = \lambda_k$. Ceci permet de déterminer les périodes et, pour n assez grand, un test de Fisher permet de savoir si on peut admettre que les nombres obtenus pourraient être observés dans un processus de la forme $X(t) = \varepsilon(t)$ où ε obéit à une loi de Laplace-Gauss.

Il nous paraîtrait souhaitable de tester directement que $a(t)$ a pour période T et non que son développement contient un terme sinusoïdal de période T.

Autrement dit, nous croyons qu'il conviendrait de calculer la fonction $\hat{\sigma}^2_\varepsilon(T)$ qui donne la variance de ε sous l'hypothèse que la période est T.

Si cette quantité est petite par rapport à la variance des $X(t)$ pour une certaine valeur T_0 de T et ceci bien que les données s'étendent sur plusieurs périodes, on devra rechercher si certaines raisons économiques ne nous incitent pas à penser que, soit T_0 soit un sous-multiple de T_0 est bien une période.

* * *

Modèle périodique stochastique avec tendance. — Supposons que $X(t)$ soit de la forme :

(16) $$X(t) = a(t) + bt + \varepsilon_t$$

où les ε_t sont des variables aléatoires indépendantes de moyenne nulle et de variance σ^2 et $a(t)$ une fonction périodique de période T supposée connue. Nous pourrons encore grouper les données suivant la table de Buys-Ballot et poser :

$$X_{ij} = X\,[(i-1)\,T + j] = a_j + b\,[(i-1)\,T + j] + \varepsilon_{ij}.$$

$X_{11}\ X_{12}\ \ldots\ldots\ldots\ldots\ X_{1T}$	l_1
$X_{21}\ \ldots\ldots\ldots\ldots\ldots\ X_{2T}$	l_2
$\ldots\ldots\ X_{ij}\ \ldots\ldots$	l_i
$X_{s1}\ \ldots\ldots\ldots\ldots\ldots\ X_{sT}$	l_s
$m_1 \qquad m_j \qquad m_T$	

Et on pose :

$$m_j = \frac{1}{3} \sum_i X_{ij}, \quad l_i = \frac{1}{T} \sum X_{ij}.$$

On trouve comme estimation de b et des $\hat{a}_j$:

$$\hat{b} = \frac{12}{T^3 s(s^2-1)} \sum_i \left(i - \frac{s+1}{2}\right) l_i \tag{17}$$

$$\hat{a}_j = m_j - \hat{b}\left[j + \frac{s-1}{2} T\right] \tag{18}$$

et on montre que :

$$\sigma^2_{\hat{b}} = \frac{12}{T^3 s(s^2-1)} \sigma^2_{\varepsilon} \tag{19}$$

$$\sigma^2_{\hat{a}_j} = \frac{\sigma^2_{\varepsilon}}{N} + \frac{12}{T^3 s^3(s^2-1)} \left(j + \frac{s-1}{2} T\right)^2 \sigma^2_{\varepsilon} \tag{20}$$

comme exemple prenons $T = 52$, $N = 5$ (cas où on a une observation par semaine pendant 5 ans). On trouve :

$$\sigma_{\hat{b}} = 0.8433 \times 10^{-3}\ \sigma_{\varepsilon}$$
$$\sigma_{\hat{a}_1} = 0.4429\ \sigma_{\varepsilon}$$
$$\sigma_{\hat{a}_{52}} = 0.4473\ \sigma_{\varepsilon}$$

On voit que $\sigma_{\hat{b}}$ devient rapidement très petit quand s croît. Par contre, les $\sigma_{\hat{a}_j}$ sont toujours supérieurs à $\frac{\sigma_{\varepsilon}}{\sqrt{s}}$ de sorte que même si le modèle restait identique à lui-même pendant un siècle on aurait encore sur les $\sigma_{\hat{a}_j}$ une erreur légèrement supérieure à $\frac{\sigma_{\varepsilon}}{10}$.

Les calculs se font d'ailleurs tout aussi bien si la fonction $b(t)$ est remplacée par une fonction plus compliquée avec des paramètres inconnus et on peut évaluer ces paramètres avec une bonne précision.

De tels modèles me paraissent particulièrement bien adaptés à l'étude des séries chronologiques ; car, sans être aussi compliqués que les schémas auto-régressifs, ils tiennent compte à la fois des facteurs généraux qui ont une action durable sur le fait étudié et des causes de fluctuations cycliques qui, comme en physique, peuvent être des périodes propres d'oscillation du système ou à des oscillations forcées du système de causes extérieures.

Quoi qu'il en soit, il est aisé de tester si on peut statistiquement considérer qu'on a affaire à un modèle de ces types ou à un modèle stochastique périodique du type A, dans le cas où on sait que les ε obéissent à une loi de Laplace-Gauss.

Il suffit, à cet effet, de tester si les l_i, peuvent être considérés comme extraits d'une population de Laplace-Gauss d'écart-type $\frac{\sigma_\varepsilon}{\sqrt{n}}$. On y arrivera généralement au moyen d'un test F.

* * *

Modèles markoviens. — a) Reprenons le modèle II (période t connue) (p. 46) le plus général, on peut encore l'écrire en remplaçant $X(t)$ par Z_t :

(21) $Z_t = a_{[t]} + \varepsilon_{[t]}$ où $[t]$ est l'entier modulo T correspondant à t, ou encore :

$Z_t = a_{[t]} + \omega_t\,\sigma_{[t]}$ où les ω_t sont des variables aléatoires indépendantes de moyenne nulle et de variance unité $E(\omega_t) = 0$ et $\sigma^2\,\omega_t = 1$.

b) On peut transformer ce modèle de manière à ce que les ω_t ne soient plus indépendantes de la manière suivante en supposant qu'on a toujours (21) mais que :

$$\omega_t = \rho_{[t]}\,\omega_{t-1} + \sqrt{1 - \rho^2_{[t]}}\,u_t.$$

c) Si on veut supposer que l'état du système se conserve à un terme aléatoire près, on a à considérer comme variable non plus Z_t mais la quantité $X_t = X_{t-1} + Z_t$ où Z_t est donné par la forme b) précédente.

On a donc le modèle suivant :

$$X_t = X_{t-1} + Z_t$$
$$Z_t = a_{[t]} + \omega_t \, \sigma_{[t]}$$
$$\omega_t = \rho_{[t]} \, \omega_{t-1} + \sqrt{1 - \rho^2_{[t]}} \; u_t$$

d) Enfin, on peut remarquer que, dans un processus de production d'une quantité y_t, c'est le rapport $\frac{y_t}{y_{t-1}}$ qui intervient. Ce qui amène à poser : $\log y_t = X_t$, où X_t vérifie c).

Les trois derniers modèles ont en général $3T$ paramètres à déterminer qui se réduisent à $2T + 1$ *si tous les* $\rho_{[t]}$ sont égaux et à $T + 2$ si tous les $\sigma_{[t]}$ sont également égaux $a_{[t]}$ et $\sigma_{[t]}$ où σ se détermine comme dans II (p. 46) et $\sigma_{[t]}$ comme un coefficient de corrélation. Ce modèle a été utilisé par Morlat pour l'étude de la production électrique. À son sujet se posent essentiellement 2 problèmes :

1) *Un problème d'agrégation.* De la production mensuelle d'énergie électrique, on peut déduire la production annuelle $Y_i = y_{12\,i+1} + y_{12\,i+2} + \ldots + y_{12i+12}$. Les variables Y_i déterminent un nouveau processus obéissant à des lois fort complexes mais que l'on peut en première approximation remplacer par un processus à accroissements autocorellés.

2) *Un problème de prévision à court terme.* On a évidemment :

$$E(\omega_{t+1} \mid \omega_t) = \rho\omega_t.$$
$$E(\omega_{t+h} \mid \omega_t) = \rho^h\omega_t.$$

d'où :

$$E(Z_{t+h} \mid t) = a_{[t+h]} + \rho^h (Z_{[t]} - a_{[t]})$$
$$E(X_{t+h} \mid t)\, X_t + \Sigma \; E(Z_{t+j})$$
$$= X_t + a_{t+1} + \ldots + a_{t+h} + (\rho + \rho^2 + \ldots + \rho^h)(Z_t - a_t)$$

De même, $\operatorname{var}(X_{t+h} \mid t)$ peut se calculer ; on trouve :

$$\operatorname{var}(X_{t+h} \mid t) = \sigma^2 \left[h + 2(h-1)\frac{\rho}{1-\rho} - 2\left(\frac{\rho}{1-\rho}\right)^2 \frac{1-\rho^{h-1}}{1-\rho} - \left(\frac{\rho}{1-\rho}\right)^2 (1-\rho^h)^2 \right]$$

Morlat a appliqué ce modèle à l'étude de la production d'énergie électrique en France. Il trouve pour les a_j des valeurs fortes pendant les mois de travail (surtout septembre), faibles pendant les mois de repos (surtout août). σ_j et ρ_j sont nettement variables avec le mois considéré. Le modèle simplifié donne $\sigma = 0.012$, $\rho = -0.50$. Donc, tant qu'il s'agit de décrire la réalité, on ne rencontre guère de difficultés. Il en va autrement lorsqu'on veut travailler effectivement sur un tel modèle et résoudre grâce à lui des problèmes concrets. On se trouve assez *rapidement alors* devant des calculs inextricables.

Examen d'un problème concret. Certes le modèle précédent peut paraître beaucoup trop simple. En réalité, un peu de réflexion montre que tel quel il est déjà assez complexe, parce que faire un modèle n'est pas une fin en soi, mais est, au contraire, un moyen pour donner une réponse à une question déterminée. Si le modèle proposé est trop complexe, il devient impossible de l'intégrer facilement dans un schéma donné. Par contre, nous verrons sur un exemple concret qu'il n'est pas toujours très adapté à la réalité.

Examinons, en effet, un problème concret, et cherchons à déterminer l'élasticité de la demande de transport par métro et par autobus par rapport au prix.

À cet effet, on considère 2 périodes pendant chacune desquelles les prix sont restés constants et on ajuste a_j et a'_j à chacune de ces 2 périodes ce qui permet de connaître la diminution de trafic correspondant à une augmentation relative de prix $\frac{\Delta p}{p}$. On peut ainsi estimer quelle devrait être l'augmentation des prix métro-autobus pour combler le déficit de ces régies. Les augmentations de tarif métro-autobus ont lieu simultanément et dans les mêmes proportions, ce qui évite les phénomènes de substitution.

On voit donc que toute la difficulté consiste à évaluer a_j et a'_j mais en même temps apparaissent de nombreuses complications pour appliquer le schéma ci-dessus, complications inhérentes à la nature du problème posé.

1) On ne peut avoir ni statistiques mensuelles ni statistiques journalières, mais seulement des statistiques hebdomadaires (cartes hebdomadaires de transport) et il n'y a pas un nombre exact de semaines dans l'année.

2) Il y a des grèves, d'où la nécessité d'employer des techniques analogues à celles du *missing-plot*.

3) Les fêtes mobiles appellent une correction.

4) Le coût de la vie et les salaires ne sont pas fixes ; aussi est-il nécessaire de faire intervenir dans les calculs l'indice du coût de la vie et celui des salaires.

5) La réaction à une augmentation de tarif n'est pas instantanée et la variation du prix du ticket de métro s'accompagne, en général, d'une proche variation de salaire. Aussi peut-on obtenir l'élasticité cherchée.

Pour tenir compte de tous ces faits, on est amené à poser :

$$\mathcal{X}(t) = A(t) \,.\, F(t) \,.\, \mathcal{N}(t, E) \;\; (1 + \varepsilon(t))$$

où : $A(t)$ est une fonction pérodique de période $\mathcal{T}$,

$F(t)$ une fonction égale à 1 partout sauf au moment des fêtes mobiles où elle prend une valeur connue d'avance,

$\mathcal{N}(t, E)$ le nombre de voyageurs moyen à la période t dans un environnement économique E (nombre calculé abstraction faite du facteur saisonnier, donc ne faisant intervenir que la tendance à long terme),

$\varepsilon(t)$ un terme aléatoire.

On a, en prenant les logarithmes népériens :

$$\log \mathcal{X}(t) = \log A(t) + \log F(t) + \log \mathcal{N}(t,E) + \log (1 + \varepsilon(t))$$

$$y_t = a_t + f_t + b \log \frac{U}{P} + c \log \frac{S}{P} + K + \varepsilon_t$$

en posant : $\operatorname{Log} \mathcal{X}(t) = y_t$, $\log A(t) = a_t$, $\log F(t) = f_t$,

$$\log \mathcal{N}(t,E) = b \log \frac{U}{P} + c \log \frac{S}{P} + K,$$

où U est un indice de valeur du titre de transport, S un indice de salaire et P un indice des prix, ces 3 indices étant rapportés à la base 100 en 1946. $\frac{U}{P}$ et $\frac{S}{P}$ sont donc des indices de valeur réelle du titre de transport et de valeur réelle du prix. a représente l'élasticité du salaire réel et b celle du tarif réel. La méthode d'étude de modèles à tendance généralisés qui va suivre montre comment on peut calculer des estimations $\hat{a}$ et $\hat{b}$ de a et b. Fourgeaud trouve $\hat{a} = -0.104$, $\hat{b} = +0.185$ en ce qui concerne les cartes hebdomadaires, pour la période 1946-1951. En ce qui concerne les billets, il trouve pour cette même période $\hat{a} = -0.223$, $\hat{b} = +0.209$. Contrairement à une opinion répandue, l'octroi de cartes hebdomadaires à un tarif de faveur est une opération ruineuse pour la compagnie. Car, non seulement elle accorde un tarif très bas à des gens qui circuleraient de toute façon, mais surtout, les titulaires de cartes hebdomadaires circulant généralement aux heures de pointe, elle augmente de façon considérable le coût marginal du transport.

Modèle à tendance généralisé

Le modèle précédent est un cas particulier du modèle plus général :

$$X(t) = a_{[t]} + \sum_{1}^{K} b_K \, \varphi_K(t) + \omega_t . \sigma_\varepsilon$$

où (t) est la classe modulo la période T de t (donc égal au reste de la division de t par T, $\varphi_K(t)$ des fonctions connues de t et ω_t une variable aléatoire normée.

On se propose de calculer des estimations de $a_{[t]}$, b_K, et σ_ε. Pour ce faire, on minimise la quantité :

$$X^2 = \sum \left[X(t) - a_{[t]} - \sum_{1}^{K} b_K \, \varphi_K(t) \right]^2$$

$$= \sum_{ij} \left(X_{ij} - a_{[j]} - b_1 y'_{ij} - \ldots - b_p y^p_{ij} \right)^2$$

ce qui, en écrivant : $\dfrac{\partial X^2}{\partial a_{[i]}} = 0$, $\dfrac{\partial X^2}{\partial b_K} = 0$, donne $T + p$ équations linéaires dont les solutions sont les estimations $\hat{a}_j$ $\hat{b}_K$ des paramètres du modèle :

$$\frac{\partial X^2}{\partial a_{j_0}} = \sum_{i=1}^{N} X_{ij_0} - N\hat{a}_{[j_0]} - \hat{b}_1 \sum_{1}^{N} y^1_{ij_0} - \ldots - \hat{b}_p \sum_{1}^{N} y^p_{ij_0}$$

et permettra de trouver les $a_{[j_0]}$ quand on connaîtra les $\hat{b}_K$. Pour estimer la loi des $\hat{b}_K$ on élimine les $\hat{a}_{j_0}$ entre les équations précédentes et les $\dfrac{\partial X^2}{\partial b_K}$. Il reste un système de la forme :

$$\hat{b}_1 A_{11} + \dots + \hat{b}_p A_{1p} = B_1$$
$$\hat{b}_1 A_{p1} + \dots + \hat{b}_p A_{pp} = B_p$$

où :

$$A\vec{b} = \vec{B}$$
$$\vec{b} = A^{-1} \vec{B}$$

On peut chercher la loi de probabilité des $\hat{b}_K$ qui sont évidemment des v.a. et on montre que :

$$E(\hat{b}_K) = b_K$$
$$E(\hat{b}_K - b_K)(\hat{b}_i - b_i) = \sigma_\varepsilon^2 A_{ki}^{-1}$$

où A_{ki} est le mineur de A obtenu en barrant la $k^{ème}$ ligne et la $i^{ème}$ colonne.

$$E(\hat{b}_K - b_K)^2 = \sigma_\varepsilon^2 A_{kk}^{-1}$$

de même :

$$E\hat{c}_j = c_j$$
$$E(\hat{c}_j - c_j)^2 = \frac{\sigma_\varepsilon^2}{N} + \frac{\sigma_\varepsilon^2}{N^2} \sum_{kl} A_{kl}^{-1} \Upsilon_{0j}^K \Upsilon_{0j}^c$$

cov $\hat{c}_j \hat{a}_K$ et cov $\hat{c}_j \hat{c}_K$ ont également des expressions simples [1].

R. FERON,
professeur à la Faculté des Sciences de Lyon.

1. Références :
Feron et Fourgeaud, *Quelques remarques sur l'estimation des variations saisonnières*, publ. I.S.U.P., vol. 1, fascicule 3, p. 21-25, 1952.
Fourgeaud, Thèse.
Morlat, *Modèle pour les chronologiques économiques mensuelles*, publ. Colloque T.I.M.S., Dublin, septembre 1962.