

ÉCOLE NATIONALE DE LA STATISTIQUE
ET DE L'ADMINISTRATION ÉCONOMIQUE

METHODES DE PREVISION DES
SERIES TEMPORELLES

par

A. MONFORT

DIRECTION GÉNÉRALE



METHODES DE PREVISION DES
SERIES TEMPORELLES

par

A. MONFORT

CHAPITRE I

LE LISSAGE EXPONENTIEL SIMPLE

A. DEFINITION

On dispose d'une série temporelle, c'est-à-dire de T nombres réels $(x_1, \dots, x_T)$ indicés par les dates $t = 1, \dots, T$.

On suppose qu'on se trouve à la date T et que l'on désire prévoir la valeur de x_{T+h} ; on notera $\hat{x}_T(h)$ une telle prévision et $h \in \mathbb{N}$ sera appelé l'horizon de la prévision.

Définition

La valeur de $\hat{x}_T(h)$ fournie par la méthode du lissage exponentiel simple, avec la constante de lissage β ($0 < \beta < 1$) est :

$$(1) \quad \hat{x}_T(h) = (1 - \beta) \sum_{j=0}^{T-1} \beta^j x_{T-j}$$

(Notons que, dans certains ouvrages, c'est $1 - \beta$ que l'on appelle la constante de lissage).

La formule (1) repose sur une idée simple :

on suppose que les observations influencent d'autant moins la prévision qu'elles sont éloignées de la date T à laquelle on fait la prévision ; en outre, on suppose que cette influence décroît exponentiellement.

On voit facilement que plus la constante de lissage est proche de 1 plus l'influence des observations passées remontera loin dans le temps et plus la prévision sera "rigide", c'est-à-dire peu sensible aux fluctuations "conjoncturelles" ; au contraire

plus la constante β est proche de zéro plus la prévision sera "souple", c'est-à-dire influencée par les observations récentes. Enfin, notons que $\hat{x}_T(h)$ ne dépend pas de h.

B. AUTRES INTERPRETATIONS

Comme $\hat{x}_T(h)$ ne dépend pas de h, on utilisera dans ce paragraphe la notation $\hat{x}_T$.

De la formule (1) on déduit immédiatement :

(2)

$$\hat{x}_T = \beta \hat{x}_{T-1} + (1-\beta) x_T$$

La prévision $\hat{x}_T$ apparaît donc comme une moyenne pondérée entre la prévision $\hat{x}_{T-1}$ faite à la date T-1 et la dernière observation x_T , le poids donné à cette observation étant d'autant plus fort que β est faible.

La formule (2) peut aussi s'écrire :

(3)

$$\hat{x}_T = \hat{x}_{T-1} + (1-\beta) (x_T - \hat{x}_{T-1})$$

où $\hat{x}_T$ s'interprète comme la prévision à la date précédente $\hat{x}_{T-1}$ corrigée d'un terme proportionnel à la dernière erreur de prévision.

Notons que cette formule (3) est celle utilisée en économétrie pour définir des "anticipations glissantes" qui, résolue en $\hat{x}_T$ implique une formulation à retards échelonnés de type Koyck (formule (1)).

On voit aussi que les formules (2) ou (3) fournissent une méthode de mise à jour très simple de la prévision, dans laquelle l'information apportée par le passé est résumée dans $\hat{x}_{T-1}$.

Si on veut utiliser uniquement la formule (2) ou (3) il est nécessaire d'initialiser l'algorithme, le plus simple est de prendre $\hat{x}_1 = x_1$; de toutes façons, il est clair que la valeur choisie pour $\hat{x}_1$ a peu d'influence sur $\hat{x}_T$ quand T est assez grand.

La simplicité de calcul a été longtemps un argument décisif en faveur du lissage exponentiel simple et, même avec les facilités de calcul apportées par l'informatique, c'est toujours un argument important lorsque le nombre de séries à traiter est élevé.

On va maintenant examiner l'interprétation de la méthode de lissage exponentiel simple qui est sans doute la plus utile car elle permettra de proposer des généralisations.

Considérons le programme :

$$(3) \quad \min_a \sum_{j=0}^{T-1} \beta^j (x_{T-j} - a)^2$$

la solution de (3) est immédiate, on trouve :

$$\hat{a} = \frac{1 - \beta}{1 - \beta^T} \sum_{j=0}^{T-1} \beta^j x_{T-j}$$

et donc, pour T assez grand :

$$(4) \quad \hat{x}_T \approx \hat{a}$$

Par conséquent, $\hat{x}_T$ s'interprète comme la constante qui s'ajuste le mieux à la série "au voisinage" de T, l'expression "au voisinage" traduisant le fait que, dans la minimisation (3), l'influence des observations décroît lorsqu'on s'éloigne de la date T ; notons que (3) peut s'interpréter comme une régression pondérée sur une constante, les pondérations décroissant exponentiellement lorsqu'on remonte le temps à partir de T.

Cette dernière interprétation montre clairement que la méthode ne pourra être utilisable que si la série peut être approchée par une constante au voisinage de T ; en particulier la méthode est à éviter lorsque la série présente un trend ou des fluctuations saisonnières.

Exercice

On suppose que les x_t , $t = 1, \dots, T$, sont les réalisations des variables aléatoires X_t satisfaisant :

$$X_t = \alpha_1 + \alpha_2 t + \varepsilon_t \quad E\varepsilon_t = 0 \quad \forall t$$

Montrer que

$$E \left[X_{t+h} - \hat{X}_t \right] = \frac{\alpha_2 \beta}{1-\beta} + \alpha_2 h$$

Un dernier problème dont il faut parler est le choix de β .

En pratique, le choix de β est souvent fondée sur les critères subjectifs de "rigidité" ou de "souplesse" de la prévision (voir le paragraphe A). Il existe cependant une méthode plus objective qui consiste à choisir la constante qui minimise la somme des carrés des erreurs de prévision aux dates 1, ..., T-h (pour un horizon h donné).

Ainsi, pour $h = 1$, il faut minimiser

$$(6) \quad \sum_{t=1}^{T-1} \left[x_{t+1} - (1-\beta) \sum_{j=0}^{t-1} \beta^j x_{t-j} \right]^2$$

Notons pour terminer que les études faites sur le choix de β montrent qu'en général son influence reste faible dans des plages assez grandes (voir par exemple Cox [1961]).

Les valeurs de β les plus couramment utilisées sont comprises entre 0,7 et 0,9 (voir à ce sujet Brown (1962) chapitre 8)

CHAPITRE II

LE LISSAGE EXPONENTIEL DOUBLE

A. PRINCIPE

On a vu dans le chapitre précédent que la méthode du lissage exponentiel simple était adaptée au cas où la série pouvait être ajustée à une droite horizontale au voisinage de T . Une façon naturelle de généraliser cette méthode est de supposer seulement que la série peut être ajustée à une droite quelconque au voisinage de T . Cela conduit à proposer comme prévision la fonction

(1)

$$\hat{x}_T(h) = \hat{a}_1(T) + h \hat{a}_2(T)$$

telle que $\hat{a}_1(T)$ et $\hat{a}_2(T)$ réalisent le minimum en a_1, a_2 de

(2)

$$Q = \sum_{j=0}^{T-1} \beta^j (x_{T-j} - a_1 + a_2 j)^2$$

Cherchons les expressions de $\hat{a}_1(T)$ et $\hat{a}_2(T)$. On a :

$$\frac{\partial Q}{\partial a_1} = -2 \sum_{j=0}^{T-1} \beta^j (x_{T-j} - a_1 + a_2 j)$$

$$\frac{\partial Q}{\partial a_2} = 2 \sum_{j=0}^{T-1} j \beta^j (x_{T-j} - a_1 + a_2 j)$$

En égalant ces dérivées partielles à zéro on obtient :

$$\begin{cases} \sum_{j=0}^{T-1} \beta^j x_{T-j} - a_1 \sum_{j=0}^{T-1} \beta^j + a_2 \sum_{j=0}^{T-1} j \beta^j = 0 \\ \sum_{j=0}^{T-1} j \beta^j x_{T-j} - a_1 \sum_{j=0}^{T-1} j \beta^j + a_2 \sum_{j=0}^{T-1} j^2 \beta^j = 0 \end{cases}$$

Ou, en remplaçant les séries tronquées par les séries complètes, c'est-à-dire

$$\sum_{j=0}^{T-1} \beta^j \text{ par } \frac{1}{1-\beta}, \quad \sum_{j=0}^{T-1} j \beta^j \text{ par } \frac{\beta}{(1-\beta)^2} \quad \text{et} \quad \sum_{j=0}^{T-1} j^2 \beta^j \text{ par } \frac{\beta(1+\beta)}{(1-\beta)^3}$$

$$(3) \quad \begin{cases} (1-\beta) \sum_{j=0}^{T-1} \beta^j x_{T-j} - a_1 + a_2 \frac{\beta}{1-\beta} = 0 \\ (1-\beta)^2 \sum_{j=0}^{T-1} j \beta^j x_{T-j} - a_1 \beta + a_2 \frac{\beta(1+\beta)}{1-\beta} = 0 \end{cases}$$

Si on note $S_1(t) = (1-\beta) \sum_{j=0}^{t-1} \beta^j x_{t-j}$ la série lissée et

$S_2(t) = (1-\beta) \sum_{j=0}^{t-1} \beta^j S_1(t-j)$ la série doublement lissée, on voit d'abord que

$$\begin{aligned} S_2(T) &= (1-\beta)^2 \sum_{j=0}^{T-1} \beta^j \sum_{i=0}^{T-j-1} \beta^i x_{T-j-i} \\ &= (1-\beta)^2 \sum_{j=0}^{T-1} \sum_{i=0}^{T-j-1} \beta^{i+j} x_{T-j-i} \\ &= (1-\beta)^2 \sum_{k=0}^{T-1} (k+1) \beta^k x_{T-k} \\ &= (1-\beta)^2 \sum_{k=0}^{T-1} k \beta^k x_{T-k} + (1-\beta) S_1 \end{aligned}$$

Donc les équations (3) s'écrivent :

$$(4) \quad \begin{cases} S_1(T) - a_1 + a_2 \frac{\beta}{1-\beta} = 0 \\ S_2(T) - (1-\beta) S_1(T) - a_1 \beta + \frac{\beta(1+\beta)}{1-\beta} a_2 = 0 \end{cases}$$

En résolvant par rapport à a_1 et a_2 :

$$(5) \quad \boxed{\begin{aligned} \hat{a}_1(T) &= 2 S_1(T) - S_2(T) \\ \hat{a}_2(T) &= \frac{1-\beta}{\beta} [S_1(T) - S_2(T)] \end{aligned}}$$

Notons qu'on peut aussi exprimer $S_1(T)$ et $S_2(T)$ en fonction de $\hat{a}_1(T)$ et $\hat{a}_2(T)$:

$$(6) \quad \begin{cases} S_1(T) = \hat{a}_1(T) - \frac{\beta}{1-\beta} \hat{a}_2(T) \\ S_2(T) = \hat{a}_1(T) - \frac{2\beta}{1-\beta} \hat{a}_2(T) \end{cases}$$

B. FORMULES DE MISE A JOUR

On a exprimé $\hat{a}_1(T)$ et $\hat{a}_2(T)$ en fonction de $S_1(T)$ et $S_2(T)$ [formules (5)] de façon à obtenir des formules de mise à jour simples. En effet, on a immédiatement des formules de mise à jour pour $S_1(T)$ et $S_2(T)$:

de la définition de $S_1(T)$ et $S_2(T)$ on tire :

$$S_1(T) = (1-\beta) x_T + \beta S_1(T-1)$$

et

$$\begin{aligned} S_2(T) &= (1-\beta) S_1(T) + \beta S_2(T-1) \\ &= (1-\beta)^2 x_T + \beta(1-\beta) S_1(T-1) + \beta S_2(T-1) \end{aligned}$$

En utilisant (5) :

$$\begin{aligned} \hat{a}_1(T) &= 2(1-\beta) x_T + 2\beta S_1(T-1) \\ &\quad - (1-\beta)^2 x_T - \beta(1-\beta) S_1(T-1) - \beta S_2(T-1) \\ &= x_T(1-\beta^2) + \beta(\beta+1) S_1(T-1) - \beta S_2(T-1) \end{aligned}$$

d'après (6) :

$$\hat{a}_1(T) = x_T(1-\beta^2) + \beta^2[\hat{a}_1(T-1) + \hat{a}_2(T-1)]$$

ou en utilisant (1) avec $h = 1$:

$$\hat{a}_1(T) = \hat{a}_1(T-1) + \hat{a}_2(T-1) + (1-\beta^2)[x_T - \hat{x}_{T-1}(1)]$$

De même, pour $\hat{a}_2(T)$, on a :

$$\begin{aligned} \hat{a}_2(T) &= \frac{1-\beta}{\beta} [\beta(1-\beta)x_T + \beta^2 S_1(T-1) - \beta S_2(T-1)] \\ &= (1-\beta)^2 x_T + \beta(1-\beta) S_1(T-1) - (1-\beta) S_2(T-1) \\ &= (1-\beta)^2 x_T + \beta(1-\beta) \hat{a}_1(T-1) - \beta^2 \hat{a}_2(T-1) - (1-\beta) \hat{a}_1(T-1) \\ &\quad + 2\beta \hat{a}_2(T-1) \\ &= (1-\beta)^2 x_T - (1-\beta)^2 \hat{a}_1(T-1) + (2\beta - \beta^2) \hat{a}_2(T-1) \end{aligned}$$

ou

$$\hat{a}_2(T) = \hat{a}_2(T-1) + (1-\beta)^2 [x_T - \hat{x}_{T-1}(1)]$$

En résumé,

$$(7) \quad \begin{aligned} \hat{a}_1(T) &= \hat{a}_1(T-1) + \hat{a}_2(T-1) + (1-\beta)^2 [x_T - \hat{x}_{T-1}(1)] \\ \hat{a}_2(T) &= \hat{a}_2(T-1) + (1-\beta)^2 [x_T - \hat{x}_{T-1}(1)] \end{aligned}$$

Exercice

Interpréter les formules (7) lorsque $x_T - \hat{x}_{T-1}(1)$, l'erreur de prévision à la date $T-1$, est nulle.

CHAPITRE III

LE LISSAGE EXPONENTIEL GENERALISE

A. MATRICES DE TRANSITION FIXES

On conçoit aisément que l'on ait essayé de généraliser les méthodes des deux premiers chapitres en cherchant à ajuster au voisinage de T , une fonction plus compliquée qu'une fonction affine (en particulier des fonctions ayant une composante périodique pour pouvoir traiter les séries saisonnières). La difficulté dans cette généralisation était de conserver des méthodes de calcul simples ; ce problème a été résolu par BROWN (1962), à l'aide de l'introduction de la notion de vecteur de fonctions à matrice de transition fixe.

Définition

On dira que le vecteur $f(t) = \begin{bmatrix} f_1(t) \\ \vdots \\ f_n(t) \end{bmatrix}$, $t \in \mathbb{Z}$ est à matrice de transition fixe, s'il existe une matrice L régulière telle que $f(t) = L f(t-1)$.

La méthode du lissage exponentiel généralisé consiste à ajuster au voisinage de T une fonction $\phi(t)$ du type $\sum_{i=1}^n a_i f_i(t)$ et à utiliser cette fonction ajustée pour la prévision.

Montrons d'abord que la classe des fonctions $\phi(t)$ ainsi définies comprend les fonctions usuelles.

1) Fonction constante : $\phi(t) = a$

On l'obtient avec $f(t) = 1$ et, dans ce cas, $L = 1$.

Avec une telle fonction la méthode du lissage exponentiel généralisé coïncidera avec le lissage exponentiel simple.

2) Fonction linéaire : $\phi(t) = a_1 + a_2 t$

On l'obtient avec $f(t) = \begin{bmatrix} 1 \\ t \end{bmatrix}$ et, dans ce cas,

$$L = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad \text{puisque} \quad \begin{bmatrix} 1 \\ t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ t-1 \end{bmatrix}$$

3) Polynômes de degré m

On prend comme base les polynômes

$$f_1(t) = 1, f_2(t) = t, f_3(t) = \frac{t(t-1)}{2}, \dots, f_{m+1}(t) = \frac{t(t-1)\dots(t-m+1)}{m!}$$

D'après la généralisation du triangle de Pascal, on a :

$$f_k(t) = f_{k-1}(t-1) + f_k(t-1) \quad \forall k > 1$$

donc le vecteur $f(t) \begin{bmatrix} f_1(t) \\ \vdots \\ f_{m+1}(t) \end{bmatrix}$ est à matrice de transition fixe,

cette matrice étant :

$$L = \begin{bmatrix} 1 & 0 & & & 0 \\ & 1 & 1 & 0 & & 0 \\ & 0 & 1 & 1 & & \\ & & & \ddots & \ddots & \ddots \\ & & & & 0 & 0 \\ 0 & & & & & 1 & 1 \end{bmatrix}$$

4) Fonctions sinusoïdales

Les fonctions $\phi(t) = a_1 \sin \omega t + a_2 \cos \omega t$ sont obtenues en prenant

$$f(t) = \begin{bmatrix} \sin \omega t \\ \cos \omega t \end{bmatrix}$$

et, dans ce cas, on a :

$$L = \begin{bmatrix} \cos \omega & \sin \omega \\ -\sin \omega & \cos \omega \end{bmatrix}$$

5) Fonction exponentielle

$\phi(t) = a e^{\alpha t}$ est obtenue avec $f(t) = e^{\alpha t}$ et, dans ce cas, $L = e^{\alpha}$.

Exercice 1

On peut combiner les résultats précédents et montrer par exemple qu'on peut obtenir les fonctions :

$$\phi(t) = a_1 \sin \omega_1 t + a_2 \cos \omega_1 t + a_3 \sin \omega_2 t + a_4 \cos \omega_2 t$$

ou

$$\phi(t) = (a_1 + a_3 t) \sin \omega t + (a_2 + a_4 t) \cos \omega t$$

Le vérifier et trouver les matrices L correspondantes.

B. DESCRIPTION DE LA METHODE

L'ajustement au voisinage de T , de la fonction $\phi(t) = f'(t)a$ [$f'(t)$ désignant le transposé de $f(t)$ et a le vecteur colonne formé par les a_i] consiste à minimiser en a la quantité :

$$(1) \quad \sum_{j=0}^{T-1} \beta_j [x_{T-j} - f'(-j)a]^2$$

Introduisons les notations

$$Y = \begin{bmatrix} x_T \\ \vdots \\ x_1 \end{bmatrix}, \quad F = \begin{bmatrix} f_1(0) \dots f_n(0) \\ \vdots \\ f_1(-T+1) \dots f_n(-T+1) \end{bmatrix} = \begin{bmatrix} f'(0) \\ \vdots \\ f'(-T+1) \end{bmatrix}$$

$$\Omega = \begin{bmatrix} 1 & & & \\ & 1/\beta & & \\ & \vdots & \ddots & \\ & & & i/\beta^{T-1} \end{bmatrix}$$

Il est clair, d'après la théorie de la régression, que la régression de y sur les colonnes de F associée à la matrice de covariance Ω conduit formellement au même problème, à savoir la minimisation de (1). On sait donc que la solution unique de la minimisation est :

$$(2) \quad \hat{a}(T) = (F' \Omega^{-1} F)^{-1} F' \Omega^{-1} y$$

En notant $M(T) = F' \Omega^{-1} F$ et $z(T) = F' \Omega^{-1} y$, on a

$$(3) \quad \hat{a}(T) = [M(T)]^{-1} z(T)$$

or :

$$(4) \quad M(T) = \sum_{j=0}^{T-1} \beta^j f(-j) f'(-j)$$

$$(5) \quad z(T) = \sum_{j=0}^{T-1} \beta^j f(-j) x_{T-j}$$

D'après la formule (4), on voit qu'en général $M(T)$ converge vers une matrice M finie lorsque $T \rightarrow \infty$; si on se restreint aux fonctions ϕ usuelles les seuls termes de $M(T)$ posant un problème pour la convergence sont les termes en $\beta^j e^{-2\alpha j}$ qui forment une série divergente si $\beta > e^{2\alpha}$. Notons toutefois que, dans les séries économiques, les fonctions du type $e^{\alpha t}$ sont introduites pour exprimer des trends croissants, donc $\alpha > 0$ et, par suite $\beta < e^{2\alpha}$.

...

On remplace donc $M(T)$ par sa limite M , supposée régulière, et on obtient :

$$(6) \quad \boxed{\hat{a}(T) = M^{-1} z(T)}$$

avec
$$M = \sum_{j=0}^{\infty} \beta^j f(-j) f'(-j)$$

La prévision de x_{T+h} , faite à la date T , est, rappelons le :

$$(7) \quad \boxed{\hat{x}_T(h) = f'(h) \hat{a}(T)}$$

Examinons maintenant le problème de la mise à jour de $\hat{a}(T)$.

On a :

$$\begin{aligned} z(T) &= \sum_{j=0}^{T-1} \beta^j x_{T-j} f(-j) \\ &= x_T f(0) + \sum_{j=1}^{T-1} \beta^j x_{T-j} f(-j) \\ &= x_T f(0) + \beta \sum_{j=0}^{T-2} \beta^j x_{T-j-1} f(-j-1) \\ &= x_T f(0) + \beta L^{-1} \sum_{j=0}^{T-2} \beta^j x_{T-j-1} f(-j) \\ &= x_T f(0) + \beta L^{-1} z(T-1) \end{aligned}$$

Donc, en utilisant (6) :

$$\hat{a}(T) = x_T M^{-1} f(0) + \beta M^{-1} L^{-1} M \hat{a}(T-1)$$

ou

(8)

$$\boxed{\begin{aligned} \hat{a}(T) &= g x_T + C \hat{a}(T-1) \\ \text{avec} \\ g &= M^{-1} f(0) \\ C &= \beta M^{-1} L^{-1} M \end{aligned}}$$