

NOTION SUR LA METHODE

BOX-JENKINS

Par : GEORGE ROTIER

Généralités

BOX et JENKINS (1970) ont développé des méthodes qui permettent l'ajustement, à des fins concrètes, en particulier de prévision à court terme, des processus linéaires que nous avons étudiés au chapitre II. Ils ont défini une méthode générale d'étude (cf. 31 et 32) et l'ont étendue à certains types de séries non stationnaires (§ 33) et de séries ayant une composante saisonnière. Nous n'étudierons pas ce dernier cas, pour lequel on se reportera à différents manuels disponibles (O.D. ANDERSON ; NELSON ; GRANGER et NEWBOLD).

La mise en oeuvre de la méthode dépend largement de la qualité du programme informatique disponible ; elle exige des séries assez longues : au moins 60 observations pour un modèle non saisonnier, notablement plus pour un modèle saisonnier. L'intérêt de la méthode vient surtout de ce qu'il semble qu'un nombre assez restreint de modèles permet de représenter correctement la plupart des séries économiques. Dans le cas de séries stationnaires non saisonnières, O.D. ANDERSON indique que les cinq modèles étudiés chap. 2 (AR1 ; AR2 ; MA. 1 ; ARMA 1.1) suffiraient dans tous les cas. Il ne s'agit cependant que d'un avis personnel de l'auteur, que nous suivrons pour des raisons de simplicité.

L'étude d'une série comprend quatre étapes :

- a) L'identification du modèle. On entend par ce terme ce que l'économètre appellerait la spécification du modèle. Dans le cas d'une série stationnaire, c'est, par exemple, le choix entre les cinq processus linéaires énumérés ci-dessus. La phase d'identification est la plus délicate à mettre en oeuvre et peut conduire à des incertitudes que seule l'expérience du praticien peut réduire. L'identification repose en grande partie sur des estimations initiales des paramètres.
- b) L'estimation - A partir des estimations initiales, différentes méthodes itératives permettent d'approcher l'estimateur du maximum de vraisemblance. Nous n'en traiterons pas ici (cf. GRANGER et NEWBOLD).
- c) La vérification - Les résultats de l'estimation permettent divers tests, que nous ne développerons pas, de la spécification initiale du modèle. Dans le cas où la spécification initiale, retenue au terme de la phase d'identification, est remise en cause, on revient à la phase a) pour procéder ensuite à de nouvelles estimations, puis à une nouvelle étape de vérification.

l'estimation est toujours faite à partir d'un grand échantillon, nous admettons que l'erreur d'échantillonnage sur les $\hat{\psi}$ est négligeable. Nous admettrons qu'on connaît également, de façon exacte, toutes les valeurs de x_{t-j} et u_{t-j} de la période initiale $t-j = 0$ jusqu'en t .

On remarquera que, dans le cas d'un modèle non stationnaire, la série $\varphi(B)u_t$ n'est pas convergente ; cela

ne présente pas d'inconvénient puisque, admettre que la série a commencé en un instant donné $t-j = 0$ revient à admettre que u_t est identiquement nul pour toutes

les valeurs de $t-j$ entre $-\infty$ et -1 .

Connaissant les ψ_j , les u_{t-j} et les x_{t-j} , on cherche à prévoir le plus efficacement possible x_{t+h} où h désigne l'horizon de la prévision.

Les méthodes que nous allons développer s'appliquent à la prévision à court terme : h est donc faible par rapport à t .

Notations

La prévision de x_{t+h} faite en t s'écrira $\hat{x}_h(t)$. (remarque : certains auteurs utilisent la notation $\hat{x}(t+h, t)$)

L'erreur de prévision est

$$(3.13) \quad e_t(h) = x_{t+h} - \hat{x}_h(t)$$

$E[\hat{x}_h(t) | x_t]$ et $E[e_t(h) | x_t]$ désignent l'espérance conditionnelle de la prévision et de l'erreur de prévision connaissant la suite $u_0 \dots u_t$ (ou, de façon équivalente à partir de 3.11, x_t)

Prévision linéaire - Nous admettons que

$$(3.14) \quad \hat{x}_t(h) = \psi_h^* u_t + \psi_{h+1}^* u_{t-1} + \psi_{h+2}^* u_{t-2} + \dots + \psi_{h+t}^* u_0$$

Nous chercherons à déterminer les ψ_{h+j}^* de telle façon que l'erreur quadratique moyenne de prévision soit minimale

et que la prévision soit sans biais :

$$(3.15) \quad \begin{cases} E [e_t(h) | x_t] = 0 \\ E [e_t^2(h) | x_t] \text{ minimum} \end{cases}$$

On remarquera que la dernière condition, bien qu'apparemment naturelle, ne conduit pas obligatoirement à des prévisions optimales pour le décideur. Dans de nombreuses situations concrètes, le coût d'une erreur par excès est différent de celui d'une erreur par défaut et une fonction de perte quadratique n'est pas justifiée. (cf. GRANGER et NEWBOLD ; pp. 117-119).

Lemme 3.1

$$E [u_{t-j} | x_t] = \begin{cases} u_{t-j} & j \geq 0 \\ 0 & j < 0 \end{cases}$$

Dém. Posons $t-j = t'$ - Les $u_{t'}$ sont sans corrélation

Donc :

$$t' > t \Rightarrow E [u_{t'} | x_t] = E (u_{t'}) = 0$$

$$t' \leq t \quad u_{t'} \text{ se confond avec sa réalisation observée} \quad *$$

Théorème 3.1 : La prévision optimale au sens de (3.15) s'obtient en posant

$$\psi_j^* = \psi_j \quad \forall j \in \{h, h+1, \dots, h+t\}$$

Dém. :

$$(3.16) \quad \hat{e}_t(k) = x_{t+k} - \hat{x}_t(k)$$

$$= u_{t+k} + \psi_1 u_{t+k-1} + \dots + \psi_{h-t} u_{t+1} + (\psi_{h-t} - \psi_{h-t}^*) u_t + (\psi_{h-t-1} - \psi_{h-t-1}^*) u_{t-1} + \dots + (\psi_{h+t} - \psi_{h+t}^*) u_0$$

D'après le Lemme 3.1

$$E [\hat{e}_t(h) | x_t] = 0 \text{ si et seulement si } \psi_j^* = \psi_j \quad \forall j \in \{h, h+1, \dots, h+t\}$$

$$E [\hat{e}_t^2(h)] = \sigma^2 (1 + \psi_1 + \dots + \psi_{h-1}) + \sigma^2 \sum_{j=h}^{h+t} (\psi_j - \psi_j^*)^2$$

$$\text{minimale si } \psi_j^* = \psi_j \quad *$$

En effet, en t : $\hat{x}_t = x_t$ $E(u_{t+1} | x_t) = 0$ $E_t(u_t | x_t) = u_t$
 (Lemme 3.1)

$$h \geq 2 \quad \hat{x}_t(h) = \phi \hat{x}_t(1)$$

$$\hat{x}_t(h) = \phi^h \hat{x}_t(1) \quad \forall h > 1$$

Donc, une fois connus x_t et u_t , on en déduit $\hat{x}_t(1)$, d'où

$$\hat{x}_t(h) = \phi^{h-1} \hat{x}_t(1)$$

La mise à jour des prévisions se fait sans difficulté.
 Nous l'exposerons dans le cas général

$$\hat{x}_{t+1}(h) = \psi_h u_{t+1} + \psi_{h+1} u_t + \psi_{h+2} u_{t-1} + \dots$$

$$\hat{x}_t(h+1) = \psi_{h+1} u_t + \psi_{h+2} u_{t-1} + \dots$$

$$\text{Donc : } \hat{x}_{t+1}(h) = \hat{x}_t(h+1) + \psi_h u_{t+1}$$

$$= \hat{x}_t(h+1) + \psi_h [x_{t+1} - \hat{x}_t(1)]$$

On remarquera que la mise à jour des prévisions implique le calcul des ψ_h . Cependant, comme on reste dans le domaine de la prévision à court terme, seules les premières valeurs du développement de $\hat{\theta}(B) \cdot \hat{\phi}(B)^{-1}$ ont besoin d'être calculées.

b) Prévision d'un ARIMA (0,1,1)

$$x_t = x_{t-1} + u_t + \theta u_{t-1}$$

On déduit du a), en posant $\phi = 1$

$$(3.20) \quad \hat{x}_t(1) = x_t + \theta u_t$$

$$\hat{x}_t(h) = \hat{x}_t(1) \quad \forall h > 1$$

(3.20) peut s'écrire

$$\hat{x}_t(1) = (1 + \theta) x_t - \theta (x_t - u_t)$$

Corollaire 3.1 : La prévision $\hat{x}_t(h)$ est égale à l'espérance conditionnelle en t de x_{t+h}

Dém. : Découle directement du Lemme 3.1 et de

$$(3.17) \quad x_t(h) = \Psi_h u_t + \Psi_{h+1} u_{t-1} + \dots + \Psi_{h+t} u_0 \quad *$$

Remarque : On déduit de (3.16) que

$$(3.18) \quad \hat{\varepsilon}_t(1) = u_{t+1}$$

On établira à titre d'exercice

$$\text{Cov} [\hat{\varepsilon}_t(h), \hat{\varepsilon}_t(h+m)]$$

$$\text{et Cov} [\hat{\varepsilon}_t(h); \hat{\varepsilon}_{t+k}(h)]$$

On en déduira que $\hat{\varepsilon}_t(h)$ et $\hat{\varepsilon}_t(h+m)$ sont corrélés ainsi que $\hat{\varepsilon}_t(h)$ et $\hat{\varepsilon}_{t+k}(h)$

Calcul concret des prévisions

Les prévisions se calculent aisément à partir de la forme (3.10). Nous le montrerons sur deux cas particuliers :

$$a) \text{ ARMA } (1,1) \quad x_t = \phi x_{t-1} + u_t + \theta u_{t-1}$$

$$\text{d'où (3.19)} \quad x_{t+h} = \phi x_{t+h-1} + u_{t+h} + \theta u_{t+h-1}$$

Nous avons établi que

$$\hat{x}_t(h) = E(x_{t+h} | x_t)$$

En prenant l'espérance conditionnelle des deux membres de (3.19), il vient

$$\hat{x}_t(h) = \phi \hat{x}_t(h-1) + E(u_{t+h} | x_t) + \theta E(u_{t+h-1} | x_t)$$

Cette formule de récurrence s'utilise aisément :

$$h = 1 \Rightarrow \hat{x}_t(1) = \phi x_t + \theta u_t$$

Un exemple de modèle ARIMA :

Le lissage exponentiel : ARIMA (0, 1, 1)

$$(3.6) \quad x_t - x_{t-1} = u_t + \theta u_{t-1} \Leftrightarrow \Delta x_t = u_t(1 + \theta B)$$

Comme $\Delta = 1 - B$

on peut écrire le processus sous la forme autorégressive

$$(3.7) \quad (1 - B)(1 + \theta B)^{-1} x_t = u_t$$

$$\left[1 - (1 + \theta)B + \theta B^2 \right] x_t = u_t$$

qui s'écrit, en posant $1 + \theta = \lambda \Rightarrow \theta = \lambda - 1$

$$x_t = \tilde{x}_{t-1} + u_t$$

où

$$(3.8) \quad \tilde{x}_{t-1} = \sum_{j=1}^{\infty} (1 - \lambda)^{j-1} x_{t-j}$$

\tilde{x}_{t-1} est une moyenne mobile des valeurs passées de x_t , avec des coefficients qui décroissent exponentiellement. On vérifie directement, à partir de (3.6) que

$$(3.9) \quad \tilde{x}_t = \lambda x_t + (1 - \lambda) \tilde{x}_{t-1}$$

Cette formule de récurrence est à la base d'une méthode classique de prévision à court terme que nous retrouverons dans la section suivante

La prévision à partir d'un processus linéaire

le processus ARIMA

$$(3.10) \quad \varphi(B)x_t = \theta(B)u_t$$

s'écrit, de façon équivalente

$$(3.11) \quad x_t = \theta(B) \cdot \varphi(B)^{-1} u_t = \psi(B) u_t$$

$$(3.12) \quad \varphi(B) \cdot \theta(B)^{-1} x_t = \pi(B) x_t = u_t$$

La forme (3.10) est la plus commode pour calculer concrètement les prévisions. L'étude théorique est, par contre, plus aisée en partant de la forme (3.11). Nous admettrons que tous les ψ_j de la fonction $\psi(B)$ sont connus exactement. En fait, ils ne sont qu'estimés, mais comme

Or, $x_t - u_t = x_{t-1} - \theta u_{t-1}$ d'après la définition du processus.

$$\text{Donc } \hat{x}_t(1) = (1 + \theta)x_t - \theta(x_{t-1} + \theta u_{t-1}) = (1 + \theta)x_t - \theta \hat{x}_{t-1} \quad (1)$$

$$\text{D'où } \hat{x}_t(1) = (1 + \theta) \sum_{j=0}^{\infty} (-\theta)^j x_{t-j}$$

qu'on peut écrire, en posant $1 + \theta = \lambda$

$$\hat{x}_t(1) = \lambda \sum_{j=0}^{\infty} (1 - \lambda)^j x_{t-j}$$

On retrouve la formule classique du lissage exponentiel.