RAIRO. RECHERCHE OPÉRATIONNELLE

J. ABADIE
D. TRAVERS

Une approche simplifiée de la méthode de Box et Jenkins pour l'analyse et la prévision des séries temporelles unidimensionnelles

RAIRO. Recherche opérationnelle, tome 14, n° 4 (1980), p. 355-380

<http://www.numdam.org/item?id=RO_1980__14_4_355_0>

R.A.I.R.O. Recherche opérationnelle/Operations Research
(vol. 14, n° 4. novembre 1980, p. 355 à 380)

UNE APPROCHE SIMPLIFIÉE DE LA MÉTHODE DE BOX ET JENKINS POUR L'ANALYSE ET LA PRÉVISION DES SÉRIES TEMPORELLES UNIDIMENSIONNELLES (I) (*)

par J. ABADIE (1) et D. TRAVERS (2)

Résumé. — *L'analyse des séries temporelles unidimensionnelles par la méthode de Box et Jenkins comporte trois étapes : identification, estimation, validation.*

La phase initiale d'identification est souvent difficile dans la pratique. Nous proposons ici une approche interactive simplifiée, reposant sur des manipulations algébriques simples (factorisation de polynômes, recherche de racines communes ou voisines de deux polynômes, etc.). Cela permet de placer la méthode à la portée de l'utilisateur moyen.

Abstract. — *Time Series Analysis by the Box-Jenkins method involves three stages: identification, estimation, validation.*

The initial identification stage is often difficult in practice. We propose in this paper a simplified interactive approach based on simple algebraic manipulations (factorization and roots of polynomes, etc.). This enables the method to be handled by the average user.

1. PRÉSENTATION GÉNÉRALE

L'analyse des séries temporelles unidimensionnelles par la méthode de Box et Jenkins comporte traditionnellement trois étapes :

Identification de la forme de modèle la mieux adaptée à la série étudiée;

Estimation des paramètres du modèle;

Validation : le modèle retenu convient-il bien? Sinon pourquoi? Et comment l'améliorer.

Sous cette forme, la méthode est d'un emploi très général et son intérêt n'est plus à démontrer. Cependant, la phase initiale d'identification se révèle, en pratique, être une étape délicate. Elle repose sur l'analyse des fonctions

(*) Reçu octobre 1978.
(1) Université Pierre-et-Marie-Curie, Paris-VI, Institut de Programmation.
(2) C.E.S.A.

R.A.I.R.O. Recherche opérationnelle/Operations Research, 0399-0842/1980/355/© 5.00

d'autocorrélation et d'autocorrélation partielle de la série, et il arrive malheureusement assez souvent que les corrélogrammes observés soient relativement peu typiques. Une grande expérience est alors nécessaire pour reconnaître le modèle théorique auquel les apparenter, surtout dans le cas saisonnier.

Nous proposons ici une approche simplifiée qui doit permettre de placer la méthode à la portée de l'utilisateur moyen, tel qu'on peut le rencontrer dans le contexte de la gestion des entreprises, et non plus seulement du statisticien spécialisé. Dans cette optique, l'analyse des fonctions d'autocorrélation et d'autocorrélation partielle est remplacée par des manipulations algébriques simples sur les opérateurs du modèle (factorisation de polynômes, recherche de racines communes ou voisines, etc.). Cette approche a été rendue possible par la mise au point d'une nouvelle méthode d'estimation (§ 3) qui utilise un algorithme moderne d'optimisation, et dans laquelle les premiers résidus sont tout simplement considérés comme des paramètres supplémentaires à estimer en même temps que les paramètres proprement dits du modèle. Contrairement à la méthode habituelle d'estimation (estimation avec « prévisions à rebours », § 3.2.2), cette méthode ne présuppose pas la réversibilité du modèle et peut donc s'appliquer même à des modèles très mal identifiés, notamment en ce qui concerne le degré de différenciation.

Le schéma général de la méthode de Box et Jenkins se trouve alors profondément modifié, puisque l'identification n'est plus un préalable nécessaire à l'estimation. Bien au contraire, ce sont les estimations successives de différents modèles qui permettent de préciser peu à peu le modèle à retenir en définitive (§ 4). Les deux étapes identification et estimation ne constituent plus ainsi qu'une étape unique, que nous appelons « choix du modèle », et qui rend la méthode beaucoup plus accessible et souple. Elle conduit progressivement à l'identification et à l'estimation du modèle Arima ([1]) ou Sarima ([1]) le mieux adapté. Quant à l'analyse, si délicate, des autocorrélations et autocorrélations partielles, elle est maintenant réservée à l'étape de la validation. Elle est alors beaucoup plus facile à réaliser puisqu'il ne s'agit plus que de vérifier que la série des résidus possède bien les caractéristiques d'un bruit blanc (§ 5).

Enfin, nous n'avons pas voulu perdre de vue que l'objectif ultime de toute analyse de série temporelle reste la prévision. Nous examinons donc les conséquences du choix d'un modèle sur la prévision (§ 6).

([1]) Nous écrivons Arima pour A.R.I.M.A. (Sarima pour S.A.R.I.M.A.).

Dans un souci de clarté et de brièveté, notre approche est illustrée par un nombre restreint d'exemples; on trouvera en annexe une revue rapide des modèles retenus pour un nombre plus grand de séries.

Il nous a paru utile de donner en appendice une méthode de calcul numérique des dérivées premières et secondes qui s'introduisent dans notre exposé.

Ce premier article traite des paragraphes 1 à 4, et contient aussi les références. Les autres paragraphes, les annexes et les appendices feront l'objet d'un second article.

2. RAPPELS RAPIDES SUR LES MODÈLES UTILISÉS

2.1. Modèles Arima pour séries non saisonnières

Pour représenter une série temporelle non saisonnière $\{z_t\} = \{z_1, z_2, \ldots z_N\}$ on peut envisager un modèle Arima (p, d, q) de la forme

$$\varphi(B) w_t = \theta(B) a_t, \qquad t = t_0+1, \ldots, N, \tag{2.1.1}$$

où les a_t sont des variables aléatoires indépendantes normales $(0, \sigma_a)$ (les résidus, constituant ce qu'on appelle *le bruit blanc*), et où l'on pose

$$\varphi(B) = 1-\varphi_1 B-\varphi_2 B^2-\ldots-\varphi_p B^p,$$
$$\theta(B) = 1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q,$$
$$\bar{z} = \frac{1}{N}\sum_{t=1}^{N} z_t,$$
$$w_t = \begin{cases} \tilde{z}_t = z_t-\bar{z} & \text{si } d=0, \\ \nabla^d z_t & \text{si } d>0, \end{cases}$$

où $\{w_t\}$ est une série stationnaire, et où les opérateurs de retard B et de différence ∇ sont définis par

$$B z_t = z_{t-1}, \qquad B^k z_t = z_{t-k},$$
$$\nabla z_t = (1-B) z_t = z_t - z_{t-1}.$$

La valeur donnée à t_0 est celle qui fait utiliser tous les z_t, $t = 1, \ldots, N$, c'est-à-dire $t_0 = p+d$.

En pratique, des valeurs modérées de p, d et q suffisent pour les séries souvent courtes que l'on rencontre en économie et en gestion : $d \leqq 2$; $p + q \leqq 3$ ou parfois 4.

Le modèle Arima (2, d, 2) s'écrit, par exemple :

$$(1-\varphi_1 B-\varphi_2 B^2)\, w_t = (1-\theta_1 B-\theta_2 B^2)\, a_t\,, \tag{2.1.2}$$

soit

$$w_t = \varphi_1 w_{t-1}+\varphi_2 w_{t-2}+a_t-\theta_1 a_{t-1}-\theta_2 a_{t-2}. \tag{2.1.3}$$

Il peut être utile d'introduire une constante θ_0 : le modèle (2.1.2) devient alors le modèle Arima (p, d, q) *avec constante* :

$$\varphi(B)\, w_t = \theta_0+\theta(B)\, a_t. \tag{2.1.4}$$

2.2. Transformation initiale

Dans certains cas, l'opérateur ∇ n'est pas suffisant pour transformer la série $\{ z_t \}$ en une série $\{ w_t \}$ stationnaire, et on peut alors utiliser au préalable une transformation non linéaire du type logarithme ou fonction puissance (Box et Cox, 1964) : il est alors bon de corriger le biais introduit au stade de la prévision (Granger et Newbold, 1976).

2.3. Stationnarité et inversibilité

L'utilisation des polynômes $\varphi(B)$ et $\theta(B)$ permet d'écrire facilement le modèle (2.1.1) sous la forme dite du filtre linéaire

$$z_t = \Psi(B)\, a_t\,, \tag{2.3.1}$$

avec

$$\Psi(B) = \frac{\theta(B)}{\varphi(B)\nabla^d} = 1-\psi_1 B-\psi_2 B^2-\ldots \tag{2.3.2}$$

ou sous forme inversée

$$\pi(B)\nabla^d z_t = a_t\,, \tag{2.3.3}$$

avec

$$\pi(B) = \frac{\varphi(B)}{\theta(B)} = 1-\pi_1 B-\pi_2 B^2-\ldots \tag{2.3.4}$$

Le passage entre les formes (2.1.1) et (2.3.1) ou (2.3.3) du modèle Arima suppose que les paramètres φ_i et θ_i remplissent respectivement, les conditions de stationnarité et d'inversibilité, c'est-à-dire que les racines des polynômes $\varphi(B)$ et $\theta(B)$ soient situées à l'extérieur du cercle unité dans le plan complexe.

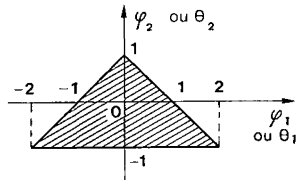

Figure 1.

Pour un modèle Arima (2, *d*, 2) par exemple, les domaines de stationnarité et d'inversibilité sont définis par les formules suivantes (*fig.* 1) :

$$|\varphi_2| < 1, \qquad \varphi_2+\varphi_1 < 1, \qquad \varphi_2-\varphi_1 < 1,$$
$$|\theta_2| < 1, \qquad \theta_2+\theta_1 < 1, \qquad \theta_2-\theta_1 < 1.$$

2.4. Parcimonie

Il est utile de considérer des modèles ayant le plus petit nombre possible de paramètres à estimer (principe de parcimonie).

2.5. Modèles Sarima pour séries saisonnières

Pour les séries saisonnières, on définit un modèle multiplicatif Sarima $(p, d, q)\times(P, D, Q)$:

$$\varphi(B)\Phi(B^s)w_t = \theta(B)\Theta(B^s)a_t, \tag{2.5.1}$$

avec

$$\Phi(B^s) = 1-\Phi_1 B^s-\Phi_2 B^{2s}-\ldots-\Phi_P B^{Ps},$$
$$\Theta(B^s) = 1-\Theta_1 B^s-\Theta_2 B^{2s}-\ldots-\Theta_Q B^{Qs},$$
$$\nabla_s = 1-B^s \quad \text{(opérateur de différence saisonnière)},$$
$$w_t = \begin{cases} \tilde{z}_t = z_t-\bar{z} & \text{si} \quad d = D = 0, \\ \nabla^d \nabla_s^D z_t & \text{si} \quad d>0 \quad \text{ou} \quad D>0. \end{cases}$$

Tout ce qui a été dit pour les modèles non saisonniers s'applique encore ici, notamment ce qui concerne l'introduction éventuelle d'une constante θ_0 et les conditions de stationnarité [pour $\Phi(B^s)$] et d'inversibilité [pour $\Theta(B^s)$].

2.6. Séries doublement saisonnières

Les valeurs à donner à t dans (2.5.1) sont celles qui font utiliser toutes les données z_t, $t = 1, \ldots, N$. A cause de l'opérateur $\nabla^d \nabla_s^D$ l'indice t de w_t

varie de $d + s\,D + 1$ à N. Au premier membre de la formule apparaît le terme de plus bas indice $w_{t-p-s\,P}$ ce qui donne pour plus petite valeur de t dans (2.5.1) le nombre $d + s\,D + p + s\,P + 1$. Nous poserons

$$t_0 = d + s\,D + p + s\,P,$$

et ferons varier t, dans (2.5.1), de $t_0 + 1$ à N.

Tout ce qui a été dit pour les modèles non saisonniers s'applique encore ici, notamment ce qui concerne l'introduction éventuelle d'une constante θ_0 et les conditions de stationnarité [pour $\Phi(B^s)$] et d'inversibilité [pour $\Theta(B^s)$].

Des séries avec deux périodicités se rencontrent parfois : périodicité hebdomadaire et annuelle, par exemple, dans la consommation journalière d'énergie électrique (Abadie et Meslier, 1977 et 1979). Ces séries comportent alors souvent des perturbations, dues à la présence de fêtes fixes ou mobiles et aux périodes de vacances (mois d'août). Nous ne les aborderons pas ici.

3. ESTIMATION DES PARAMÈTRES

3.1. Moindres carrés

Le principe du maximum de vraisemblance permet d'estimer les paramètres du modèle le mieux adapté pour représenter une série donnée $\{ z_t \}$, ceci au sein d'une classe particulière de modèles Arima ou Sarima, c'est-à-dire pour p, d, q, P, D, Q, s fixés.

Pour des séries suffisamment longues ($N \geqq 40$ dans le cas non saisonnier ou 60 dans le cas saisonnier), ce « meilleur modèle » peut être obtenu par la méthode des moindres carrés. Le problème à résoudre est le suivant :

$$\text{(P)} \qquad \min S(x \mid w), \qquad \text{avec} \quad S = \Sigma a_t^2,$$

où w représente la série différenciée et x le vecteur des paramètres

$$x = (x_1, \ldots, x_m) = (\varphi_1 \ldots \varphi_p, \theta_0, \theta_1 \ldots \theta_q, \Phi_1 \ldots \Phi_P, \Theta_1 \ldots \Theta_Q).$$

3.2. Le problème de l'initialisation

Dans la somme S, les résidus sont calculés par récurrence, par exemple à partir de la formule suivante, dans le cas non saisonnier :

$$a_t = w_t - \varphi_1 w_{t-1} - \ldots - \varphi_p w_{t-p} - \theta_0 + \theta_1 a_{t-1} + \ldots + \theta_q a_{t-q}, \quad (3.2.1)$$

et par une formule analogue dans le cas saisonnier.

Pour la plus petite valeur de t dans (2.5.1), c'est-à-dire $t = t_0 + 1$, la formule de récurrence (3.2.1), ou son analogue pour le cas saisonnier, laisse $q + Q\,s$ valeurs de a_t qui ne peuvent pas être définies par récurrence et doivent l'être autrement. Nous poserons

$$r = q + s\,Q,$$

$$a_* = (a_{t_0+1-r}, \ldots, a_{t_0}).$$

a_*, ensemble des résidus dont la connaissance est nécessaire à l'initialisation de (3.2.1), s'appelle par la suite ensemble des *premiers résidus*.

3.2.1. *Premiers résidus nuls*

Une première méthode consiste à considérer ces premiers résidus comme nuls et à résoudre le problème approché

$$(P_0) \qquad \min S(x \mid w,\ a_* = 0).$$

La perturbation introduite n'est cependant pas toujours suffisamment vite absorbée au travers de l'équation de récurrence, surtout pour les séries saisonnières.

3.2.2. *Prévisions à rebours*

Une deuxième méthode consiste à effectuer des sortes de « prévisions à rebours » (back-forecasting) à l'aide du même modèle écrit en remontant dans le temps avec l'opérateur d'avance F ($F\,z_t = z_{t+1}$). Dans le cas non saisonnier on écrit, par exemple

$$\varphi(F)\,w_t = \theta(F)\,e_t, \qquad (3.2.2)$$

où e_t est un terme d'erreur.

On effectue alors une suite de prévisions en arrière et en avant. Le processus converge normalement assez rapidement et permet d'estimer les premiers résidus a_*. Le problème à résoudre peut ainsi s'écrire :

$$(P_R) \qquad \min S(x \mid w,\ a_* = a_{\text{Rebours}}).$$

Bien que cette méthode soit la plus généralement employée, nous ne la décrirons pas plus en détail ici car nous ne l'utilisons pas nous-mêmes. Le lecteur en trouvera l'exposé dans l'ouvrage de référence de Box et Jenkins (1970, p. 209-220). Notons que la méthode des prévisions à rebours présente l'inconvénient sérieux de reposer entièrement sur la parfaite stationnarité

de la série différenciée $\{ w_t \}$ et donc en particulier sur l'identification exacte des degrès de différenciation d et D.

3.2.3. *Une troisième voie : l'optimisation simultanée des premiers résidus* a_*

Nous proposons d'utiliser plutôt une méthode plus naturelle qui consiste à considérer les premiers résidus a_* comme de simples paramètres supplémentaires à optimiser

$$(P_1) \qquad \min S(x, a_* \mid w), \qquad \text{avec} \quad S = \Sigma' a_t^2 + \Sigma^* a_*^2,$$

où, dans Σ^*, l'indice de sommation varie de $t_0 + 1 - r$ à t_0, et dans Σ' de $t_0 + 1$ à N (les premiers résidus a_* sont inclus dans le calcul de la somme S). La variance résiduelle à l'optimum σ_a^2 peut alors être estimée en divisant la somme résiduelle $\hat{S}$ à l'optimum par le nombre ν de résidus calculés par récurrence, c'est-à-dire non compris les premiers résidus a_* :

$$\hat{\sigma}_a^2 = \frac{\hat{S}}{\nu} \qquad \text{avec} \quad \nu = N - d - sD - p - sP = N - t_0$$

3.3. Le programme d'optimisation

Le problème (P_1) peut comporter un assez grand nombre de paramètres en raison des $q + Q\,s$ premiers résidus a_* introduits.

Nous utilisons un programme d'optimisation non linéaire (Himmelblau, 1972), qui comporte une recherche unidimensionnelle par la méthode de Coggins et l'approximation de l'inverse du Hessien par la méthode de Davidon-Fletcher-Powell. L'optimisation se fait en deux temps : résolution du problème (P_0) puis résolution du problème (P_1) à partir de la solution approchée précédente. Ce programme se comporte bien, même au voisinage des limites du domaine d'inversibilité et de stationnarité. Il faut remarquer qu'aucune hypothèse n'a été faite sur la stationnarité de la série $\{ w_t \}$ et que l'estimation des paramètres s'effectue correctement même dans le cas où les polynômes $\varphi(B)$, $\theta(B)$, $\Phi(B^s)$ ou $\Theta(B^s)$ possèdent des racines proches de 1 ou égales à 1.

Le programme ne calcule jamais la somme S en des points de l'espace des paramètres situés à l'extérieur des domaines de stationnarité et d'inversibilité du modèle. Si le cas se présente, on calcule S sur la frontière du domaine et on ajoute une pénalité proportionnelle à la distance du point considéré à la frontière; sans cette précaution, des dépassements de capacité (overflow) pourraient se présenter à l'exécution des calculs (notons qu'il y a d'autres procédés pour traiter cette difficulté d'overflow).

3.4. Précision des estimations

Pour des séries suffisamment longues, on montre (Box et Jenkins, 1970, p. 227) que les valeurs estimées des paramètres sont des variables aléatoires approximativement distribuées selon une loi multinormale de moyennes égales aux vraies valeurs des paramètres, avec une matrice de variance-covariance :

$$V(\hat{x}) = \begin{pmatrix} V(\hat{x}_1) & \dots & \mathrm{Cov}(\hat{x}_1, \hat{x}_m) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}(\hat{x}_m, \hat{x}_1) & \dots & V(\hat{x}_m) \end{pmatrix} = \frac{2}{n} S(\hat{x}) . S''(\hat{x})^{-1}, \qquad (3.4.1)$$

avec

$$n = N - d - s D \quad (*),$$

$$S(\hat{x}) = \min \Sigma a_t^2,$$

$$S''(\hat{x}) = \left[\begin{array}{ccc} \dfrac{\partial^2 S}{\partial x_1^2} & \dots & \dfrac{\partial^2 S}{\partial x_1 \partial x_m} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 S}{\partial x_m \partial x_1} & \dots & \dfrac{\partial^2 S}{\partial x_m^2} \end{array} \right]_{x=\hat{x}}$$

On peut estimer $S''(\hat{x})$ en calculant les dérivées secondes par différences finies. Dans les cas simples, on peut utiliser les formules analytiques suivantes :

— *Arima* (0, *d*, 1) :

$$\mathrm{Var}(\hat{\theta}) \simeq \frac{1}{n}(1-\hat{\theta}^2);$$

— *Arima* (0, *d*, 2) :

$$V(\hat{\theta}_1, \hat{\theta}_2) \simeq \frac{1}{n} \begin{pmatrix} 1-\hat{\theta}_2^2 & -\hat{\theta}_1(1+\hat{\theta}_2) \\ -\hat{\theta}_1(1+\hat{\theta}_2) & 1-\hat{\theta}_2^2 \end{pmatrix};$$

— *Arima* (1, *d*, 0) :

$$\mathrm{Var}(\hat{\varphi}) \simeq {}^{1}(1-\hat{\varphi}^2);$$

— *Arima* (2, *d*, 0) :

$$V(\hat{\varphi}_1, \hat{\varphi}_2) \simeq \frac{1}{n} \begin{pmatrix} 1-\hat{\varphi}_2^2 & -\hat{\varphi}_1(1+\hat{\varphi}_2) \\ -\hat{\varphi}_1(1+\hat{\varphi}_2) & 1-\hat{\varphi}_2^2 \end{pmatrix};$$

(*) D'autres valeurs de n ont été proposées dans la littérature; nous utilisons parfois nous-mêmes deux autres valeurs : $n = N - t_0 - r$ et $n = \nu = N - t_0$. Cela n'a d'ailleurs guère d'importance pratique pour des échantillons assez grands.

— *Arima* (1, *d*, 1) :

$$V(\hat{\varphi}, \hat{\theta}) \simeq \frac{1}{n} \frac{1-\hat{\varphi}\hat{\theta}}{(\hat{\varphi}-\hat{\theta})^2} \begin{pmatrix} (1-\hat{\varphi}^2)(1-\hat{\varphi}\hat{\theta}) & (1-\hat{\varphi}^2)(1-\hat{\theta}^2) \\ (1-\hat{\varphi}^2)(1-\hat{\theta}^2) & (1-\hat{\varphi}\hat{\theta})(1-\hat{\theta}^2) \end{pmatrix}$$

On constate que la précision des estimations ne dépend que du nombre de termes de la série différenciée $\{ w_t \}$ et des valeurs $\hat{\varphi}_i$ et $\hat{\theta}_i$ des paramètres à l'optimum. Ceci permet de construire facilement des intervalles de confiance (par exemple à 95 %) autour des valeurs estimées (ou même une ellipse de confiance, que nous n'utilisons pas).

Remarquons que dans tous les cas une formule analytique approximative peut être calculée (*voir* : Box et Jenkis, 1970, paragraphes 7.2.6, A.7.5, 9.2.4, 9.3.3; Kendall et Stuart, 1976, paragraphe 50.12). On trouvera une méthode de calcul numérique en appendice.

4. CHOIX DU MODÈLE

4.1. Identification à l'aide des corrélogrammes

Avant d'estimer les paramètres, il faut spécifier les valeurs de p, d, q, P, D, Q, s. La méthode habituelle, appelée identification, repose sur l'analyse des corrélogrammes et fonctions d'autocorrélation partielle de la série étudiée $\{ z_t \}$ et de ses différences $\{ \nabla^d \nabla_s^D z_t \}$ (ces notions sont définies plus loin aux paragraphes 5.1 et 5.3). Les fonctions observées sont rapprochées des formes des fonctions théoriques connues correspondant aux modèles Arima ou Sarima simples et on retient finalement le modèle qui semble le mieux approprié. Il s'agit en pratique d'une opération assez délicate, car les fonctions observées sont rarement typiques et l'identification nécessite une assez grande expérience, surtout pour les modèles mixtes ($p \neq 0$ et $q \neq 0$) et les modèles saisonniers.

Pour identifier les degrés de différenciation d et D, certains auteurs (O. D. Anderson, 1976, p. 125, et Lenz, 1978, p. 7) proposent simplement de retenir le ou les couples (d, D) correspondent à la série différenciée de variance minimale

$$\min \left[\hat{\sigma}^2_{w_t}(d, D) \,\middle|\, w_t = \nabla^d \nabla_s^D z_t \right],$$

avec

$$0 \leqq d \leqq 2, \qquad 0 \leqq D \leqq 2.$$

Nous allons voir que l'utilisation d'un programme d'estimation avec optimisation simultanée des premiers résidus permet d'envisager une toute autre approche sans faire appel aux notions de corrélogramme et des fonctions d'autocorrélation partielle. Tout ce qui suit ne serait pas applicable avec la méthode des prévisions à rebours puisque celle-ci, basée sur le fait que $\{ w_t \}$ est stationnaire, suppose que le degré de différenciation a déjà été correctement identifié.

4.2. Équivalence des modèles

Étant donnée une série observée $\{ z_t \}$, l'estimation des paramètres d'un modèle quelconque (donc *a priori* mal identifié) a-t-elle un sens?

Examinons, par exemple, les estimations correspondant à différents modèles pour une série de cours d'actions IBM (Box et Jenkins, 1970). Les résultats sont rassemblés dans le tableau I. Les chiffres entre crochets désignent l'écart-type de l'estimation au-dessous de laquelle ils se trouvent.

Chaque modèle peut être considéré comme la « meilleure » approximation du « vrai modèle » sous-jacent à l'intérieur d'une classe particulière Arima (p, d, q).

Quelques opérations simples (factorisations, simplifications, abandon des paramètres non significativement différents de zéro) permettent de se rendre compte que la plupart de ces modèles sont équivalents. Les quelques exceptions se reconnaissent facilement à une somme résiduelle $\hat{S}$ nettement plus élevée.

4.3. Identification par élimination progressive

4.3.1 *Principe général*

L'exemple de la série B suffit à montrer qu'une « erreur » dans le choix du degré de différenciation d se traduit simplement par l'estimation d'un facteur $1 - B$ du côté autorégressif ou du côté des moyennes mobiles, ce qui permet une rectification très facile. On voit alors qu'il est possible de concevoir une démarche progressive qui permette de trouver le modèle idéal sans pour autant estimer tous les modèles possibles, ce qui serait coûteux en temps de calcul.

Une première idée consisterait à partir d'un modèle largement suridentifié pour n'avoir plus qu'à le simplifier progressivement pour arriver au modèle définitif. En fait une telle approche est un peu brutale : l'interprétation d'un modèle Arima $(2, d, 2)$ n'est pas toujours facile et le temps de calcul nécessaire

TABLEAU I

Série B. Modèles estimés, factorisés, simplifiés (série IBM, 369 observations)

Arima (p, d, q)	$\hat{S} = \Sigma a_t^2$ à l'optimum	Modèles
(0, 0, 2)	274572	$\tilde{z}_t = (1 + 1{,}5\,B + 0{,}87\,B^2)\,a_t$ (S très grand) [0,03] [0,03]
(2, 0, 0)	19201	$(1 - 1{,}09\,B + 0{,}09\,B^2)\,\tilde{z}_t = a_t$ [0,05] [0,05] $(1 - B)(1 - 0{,}09\,B)\,\tilde{z}_t = a_t$
(1, 0, 1)	19211	$(1 - 0{,}999\,B)\,\tilde{z}_t = (1 + 0{,}09\,B)\,a_t$ [0,002] [0,05]
(2, 0, 2)	19067	$(1 - 0{,}056\,B - 0{,}94\,B^2)\,\tilde{z}_t = (1 + 1{,}04\,B - 0{,}1\,B^2)\,a_t$ $(1 + 0{,}9\,B)(1 - 0{,}99\,B)\,\tilde{z}_t = (1 + 0{,}9\,B)(1 + 0{,}13\,B)\,a_t$ (1)
(0, 1, 2)	19215	$\nabla z_t = (1 + 0{,}088\,B + 0{,}009\,B^2)\,a_t$ (2) [0,05] [0,05]
(2, 1, 0)	19184	$(1 - 0{,}087\,B + 0{,}007\,B^2)\,\nabla z_t = a_t$ (2) [0,05] [0,05]
(1, 1, 1)	19207	$(1 - 0{,}09\,B)\,\nabla z_t = (1 - 0{,}005\,B)\,a_t$ (2) [0,60] [0,60]
(2, 1, 2)	18161	$(1 - 0{,}92\,B + 0{,}92\,B^2)\,\nabla z_t = (1 - 0{,}85\,B + 0{,}87\,B^2)\,a_t$ (1) racines complexes racines complexes
(0, 2, 2)	19224	$\nabla^2 z_t = (1 - 0{,}9\,B - 0{,}08\,B^2)\,a_t$ [0,05] [0,05] $\nabla^2 z_t = (1 - 0{,}99\,B)(1 + 0{,}08\,B)\,a_t$ (1)
(2, 2, 0)	25762	$(1 + 0{,}58\,B + 0{,}28\,B^2)\,\nabla^2 z_t = a_t$ (S très grand) [0,05] [0,05]
(1, 2, 1)	19215	$(1 - 0{,}08\,B)\,\nabla^2 z_t = (1 - 0{,}99\,B)\,a_t$ (1), (2) [0,05] [0,07]
(2, 2, 2)	19101	$(1 + 0{,}84\,B - 0{,}1\,B^2)\,\nabla^2 z_t = (1 - 0{,}07\,B - 0{,}91\,B^2)\,a_t$ $(1+0{,}94\,B)(1-0{,}1\,B)\,\nabla^2 z_t = (1+0{,}92\,B)(1-0{,}99\,B)\,a_t$ (1)

(1) *Facteur commun aux deux membres*; (2) *un de plusieurs termes négligeables.*

peut être assez long, surtout si un même facteur est présent des deux côtés à la fois. L'explication en est la suivante : les méthodes de minimisation modernes ne sont rapides que si la matrice des dérivées secondes $S''(\hat{x})$

est *strictement* définie positive pour la solution minimisante. Or dans un modèle (1, *d*, 1) par exemple, les formules du paragraphe (3.4) donnent

$$S''(\hat{x}) = 2S(\hat{x}) \begin{pmatrix} \dfrac{1}{1-\hat{\varphi}^2} & -\dfrac{1}{1-\hat{\varphi}\,\hat{\theta}} \\ -\dfrac{1}{1-\hat{\varphi}\,\theta} & \dfrac{1}{1-\hat{\theta}^2} \end{pmatrix},$$

formule qui montre que $S''(\hat{x})$ est une matrice singulière si $\hat{\varphi} = \hat{\theta}$.

Finalement, nous proposons d'adopter la ligne de conduite suivante, pour les séries non saisonnières, en partant d'une valeur de d donnée quelconque ($0 \leqq d \leqq 2$, par exemple 0).

(*a*) Estimer le modèle (1, *d*, 1) :

(a_1) si $\hat{\varphi} \simeq 1$, remplacer d par $d + 1$ et retourner en (*a*); sinon aller en (a_2);

(a_2) si $\hat{\theta} \simeq 1$, remplacer d par $d - 1$ et retourner en (*a*); sinon aller en (*b*).

(*b*) Suridentifier ensuite dans la direction suggérée par le modèle retenu précédemment, c'est-à-dire :

(b_1) si $\hat{\theta} \simeq 0$ estimer (2, *d*, 0);

(b_2) si $\hat{\varphi} \simeq 0$ estimer (0, *d*, 2);

(b_3) si $\hat{\varphi} \simeq \hat{\theta}$ estimer (2, *d*, 0) *et* (0, *d*, 2) puis retenir le côté correspondant à la plus petite somme S;

(b_4) si $\hat{\varphi} \neq 0$, $\hat{\theta} \neq 0$, $\hat{\varphi} \neq \hat{\theta}$ estimer (2, *d*, 2).

Factoriser, si possible, les polynômes et aller en (*c*).

(*c*) Réduire (c'est-à-dire diminuer p ou q) ou développer (c'est-à-dire augmenter p ou q) progressivement le modèle en tenant compte :

(c_1) de la précision des estimations pour décider si un paramètre est non significativement différent de zéro ou si deux facteurs sont identiques;

(c_2) de la valeur de $\hat{S}$ à l'optimum et du principe de parcimonie pour le nombre de paramètres utilisés.

En cas de doute, on peut toujours essayer quelques modèles supplémentaires, mais généralement le choix final s'opère assez rapidement. Dans le cas relativement fréquent où certains paramètres prennent des valeurs à peine significatives, c'est à l'analyste de décider si l'abaissement de la somme $\hat{S}$ est suffisamment important pour justifier l'emploi de ces paramètres supplémentaires.

4.3.2. *Premier exemple : série IBM (suite)*

En suivant la méthode ci-dessus pour la série IBM nous aurions successivement essayé

$$(1, 0, 1): \underset{[0{,}002]}{(1 - 0{,}999\,B)}\,\tilde{z}_t = \underset{[0{,}05]}{(1 + 0{,}09\,B)}\,a_t, \qquad \hat{S} = 19211,$$

$$(1, 1, 1): \underset{[0{,}6]}{(1 - 0{,}09\,B)}\,\nabla z_t = \underset{[0{,}6]}{(1 - 0{,}005\,B)}\,a_t, \qquad \hat{S} = 19207,$$

$$(2, 1, 0): \underset{[0{,}05]\ [0{,}05]}{(1 - 0{,}09\,B + 0{,}007\,B^2)}\,\nabla z_t = a_t, \qquad \hat{S} = 19184,$$

$$(1, 1, 0): \underset{[0{,}05]}{(1 - 0{,}09\,B)}\,\nabla z_t = a_t, \qquad \hat{S} = 19207.$$

Le modèle (1, 1, 1) comporte un paramètre φ voisin de zéro et on peut donc également essayer successivement

$$(0, 1, 2): \nabla z_t = \underset{[0{,}05]\ [0{,}05]}{(1 + 0{,}09\,B + 0{,}009\,B^2)}\,a_t \qquad S = 19215,$$

$$(0, 1, 1): \nabla z_t = \underset{[0{,}05]}{(1 + 0{,}09\,B)}\,a_t \qquad S = 19217.$$

Remarquons que les deux modèles finaux (1, 1, 0) et (0, 1, 1) sont pratiquement indiscernables malgré la longueur de la série étudiée; l'explication en est que l'on a

$$(1 - 0{,}09\,B)^{-1} \simeq (1 + 0{,}09\,B).$$

Ces modèles sont d'ailleurs très proches du modèle de marche au hasard $\nabla z_t = a_t$ (puisque $\hat{\varphi} \simeq 0$ et $\hat{\theta} \simeq 0$) que l'on rencontre souvent pour les séries de cours boursiers.

Signalons enfin que l'estimation du modèle (1, 2, 1) au départ, nous aurait immédiatement conduit à envisager $d = 1$. On trouve en effet :

$$(1, 2, 1): \underset{[0{,}05]}{(1 - 0{,}09\,B)}\,\nabla^2 z_t = \underset{[0{,}002]}{(1 - 0{,}999\,B)}\,a_t \qquad \hat{S} = 19215.$$

4.3.3. *Modèles comportant un terme constant*

Généralement, la moyenne résiduelle est très petite par rapport à l'écart-type : $\bar{a} \ll \hat{\sigma}_a$. Dans certains cas, il peut pourtant arriver que la moyenne résiduelle ne soit pas suffisamment petite. Il est alors préférable d'utiliser un modèle de la forme (2.1.4) comportant un terme constant θ_0 :

$$\varphi(B)\,\nabla^d z_t = \theta_0 + \theta(B)\,a_t.$$

TABLEAU II

Série N. Choix du modèle (80 observations)

	Arima	$\hat{S} = \Sigma a_t^2$	Modèles estimés, factorisés, simplifiés	Remarques
$d = 0$	(1, 0, 1)	3593,0	$(1 - 0{,}999\,97\,B)\,\tilde{z}_t = (1 + 0{,}64\,B)a_t$ [0,001] [0,09]	$\hat{\sigma}_a \simeq \bar{a} \rightarrow$ introduire θ_0
	avec Cte θ_0	2011,4	$(1 - 0{,}999\,98\,B)\,\tilde{z}_t = 6{,}98 + (1 + 0{,}47\,B)\,a_t$ [0,001] [0,1]	$\varphi = 1 \rightarrow d > 0$
	(1, 1, 1)	1959,7	$(1 - 0{,}89\,B)\nabla z_t = (1 - 0{,}16\,B)\,a_t$ [0,06] [0,12]	$\hat{\sigma}_a \simeq 5\,\bar{a} \rightarrow$ intorduire θ_0
	avec Cte θ_0	1738,9	$(1 - 0{,}62\,B)\nabla z_t = 2{,}71 + (1 - 0{,}01\,B)\,a_t$ (2) [0,14] [0,2]	$\left.\begin{matrix}\varphi \neq 1 \\ \theta \simeq 0\end{matrix}\right\} \rightarrow \left\{\begin{matrix}\text{développer} \\ \text{à gauche}\end{matrix}\right.$
$d = 1$	(2, 1, 0) avec Cte θ_0	1737,9	$(1 - 0{,}61\,B - 0{,}01\,B^2)\nabla z_t = 2{,}74 + a_t$ (2) [0,11] [0,11] $(1 + 0{,}02\,B)(1 - 0{,}62\,B)\nabla z_t = 2{,}74 + a_t$ (2)	$\varphi_2 \simeq 0 \rightarrow p = 1$
	(1, 1, 0) avec Cte θ_0	1739,1	$(1 - 0{,}62\,B)\nabla z_t = 2{,}78 + a_t$ [0,09]	choix final
$d = 2$	(1, 2, 1)	1754,3	$(1 - 0{,}56\,B)\nabla^2 z_t = (1 - 0{,}95\,B)\,a_t$ [0,11] [0,04]	$\hat{\sigma}_a \simeq 5\,\bar{a} \rightarrow$ introduire θ_0
	avec Cte θ_0	1660,5	$(1 - 0{,}55\,B)\nabla^2 z_t = 0{,}04 + (1 - 0{,}999\,9\,B)\,a_t$ [0,09] [0,001]	$\theta \simeq 1 \rightarrow d < 2$

(2) *Un terme négligeable.*

Si l'adoption d'un tel modèle est justifiée, elle se traduit alors par une réduction très sensible de $\bar{a}$ et de la somme $\hat{S}$.

Pour ces modèles, les simplifications s'opèrent de la même manière que pour les autres modèles, à condition de modifier la constante au passage.

C'est ainsi que

$$(1-\alpha B)(1-\varphi B)\nabla z_t = \theta_0 + (1-\alpha B)\, a_t,$$

se simplifie en

$$(1-\varphi B)\nabla z_t = \frac{\theta_0}{1-\alpha} + a_t, \qquad \alpha \neq 1.$$

4.3.4. *Deuxième exemple : série N*

Un deuxième exemple permettra de mieux se familiariser avec la méthode d'identification par élimination progressive (tableau II). Il s'agit du produit national brut aux U.S.A. (Nelson, 1973).

4.4. Identification des séries saisonnières

4.4.1. *Les modèles envisageables*

Pour les séries saisonnières, le problème de l'identification à l'aide des corrélogrammes devient encore plus difficile que dans le cas non saisonnier. L'identification par élimination progressive peut donc rendre des services encore plus grands. Mais la méthode doit être appliquée de manière judicieuse, car les modèles envisageables sont beaucoup plus nombreux : déjà 213 modèles Sarima $(p, d, q)\times(P, D, Q)_s$ en se limitant pourtant à $p \leqq 2$, $d \leqq 2$, $q \leqq 2$, $P \leqq 1$, $D \leqq 1$, $Q \leqq 1$.

De plus, des modèles très complexes du genre $(2, d, 2)\times(2, D, 2)_s$ sont à proscrire car ils sont trop peu parcimonieux et la convergence du programme d'estimation peut être très lente.

Dans le cas de séries mensuelles, il faut éviter d'envisager des modèles qui réduisent trop le nombre de points utilisables pour l'estimation. En pratique on peut se limiter à $P + D \leqq 2$.

Quant à l'opérateur de moyennes mobiles saisonnières $\Theta\,(B^{12})$, on se limite aussi à $Q \leqq 1$ pour ne pas introduire trop de paramètres supplémentaires à optimiser. En effet, l'estimation se fait avec optimisation simultanée des premiers résidus et $Q = 2$ correspondrait à 24 paramètres supplémentaires.

On se reportera à l'appendice pour le calcul numérique de la matrice de covariance des paramètres.

4.4.2. *Degré de différenciation*

Pour les séries qui nous intéressent, le choix de s ne pose généralement pas de problème. Il s'agira le plus souvent de $s = 12$ (séries mensuelles) ou $s = 4$ (séries trimestrielles), et parfois $s = 5$, 6, ou 7 (séries hebdomadaires).

Le choix de l'opérateur de différences $\nabla^d \nabla_s^D$ est plus délicat. Comme pour les séries non saisonnières, il se fait à partir de l'estimation des paramètres de différents modèles $(1, d, 1)\times(1, D, 1)_s$. Plusieurs cas se présentent :

$\hat{\varphi} \simeq 1$: différencier une fois de plus ($d \to d + 1$);

$\hat{\theta} \simeq 1$: différencier une fois de moins ($d \to d - 1$);

$\hat{\Phi} \simeq 1$: différencier une fois de plus ($D \to D + 1$);

$\hat{\Theta} \simeq 1$: différencier une fois de moins ($D \to D - 1$).

Malheureusement, on peut se retrouver dans un cercle apparemment vicieux, comme le montre la figure 2.

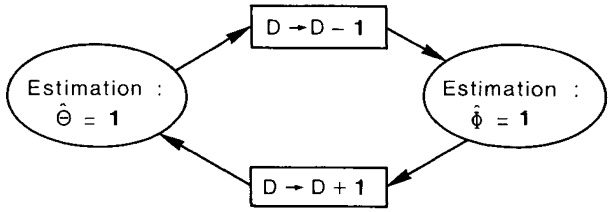

Figure 2.

En fait, l'opérateur $\nabla^d \nabla_s^D$ ne correspond alors vraiment à une surdifférenciation que si l'estimation du modèle $(1, d, 1)\times(0, D, 1)_s$ conduit aussi à $\hat{\Theta} = 1$.

De même, l'opérateur $\nabla^d \nabla_s^D$ ne correspond vraiment à une sous-différenciation que si l'estimation du modèle $(1, d, 1)\times(1, D, 0)_s$ conduit aussi à $\hat{\Phi} = 1$.

4.4.3. *Forme du modèle*

L'identification consiste à passer progressivement d'un modèle à l'autre jusqu'à arriver au modèle définitif. Ce passage s'effectue en tenant compte des résultats de chaque estimation à l'aide des opérations suivantes :

(*a*) modification de l'opérateur de différence chaque fois qu'apparaît un facteur proche de $(1 - B)$ ou $(1 - B^s)$ dans l'estimation [s'il apparaît des deux côtés à la fois, *voir* (*j*)];

(*b*) abandon des paramètres voisins de 0 (compte tenu de la précision des estimations, paragraphe 3.4);

(*c*) factorisation des polynômes et simplifications éventuelles chaque fois qu'apparaissent des facteurs égaux (compte tenu de la précision des estimations) des deux côtés à la fois;

(*d*) développement du modèle du côté où les paramètres ne sont pas nuls.

Pourtant certaines situations peuvent sembler assez difficiles au premier abord et les indications suivantes peuvent s'avérer utiles :

(*e*) partir d'un modèle $(1, d, 1)\times(1, D, 1)_s$, par exemple $(1, 0, 1)\times(1, 0, 1)_s$;

(*f*) ne pas envisager inutilement de modèles correspondant à $P > 1$, $D > 1$ ou $Q > 1$. En pratique, on se limitera à $P + D \leqq 2$ et $Q \leqq 1$;

(*g*) ne pas tout changer à la fois : modifier la partie saisonnière *ou* la partie non saisonnière du modèle;

(*h*) si plusieurs possibilités s'offrent, les explorer toutes;

(*i*) en cas de simplification [voir (*c*)] essayer des modèles dissymétriques $(0, d, q)\times(P, D, Q)$ et $(p, d, 0)\times(P, D, Q)$ ou $(p, d, q)\times(0, D, Q)$ et $(p, d, q)\times(P, D, O)$;

(*j*) en cas de cercle vicieux (§ 4.4.2) essayer aussi des modèles dissymétriques. Il s'agit en effet, le plus souvent, d'un cas particulier de (*i*) où des termes voisins de $(1 - B)$ ou $(1 - B^s)$ apparaissent simultanément des deux côtés);

(*k*) si l'opérateur $(1 + B)$ ou $(1 + B^s)$ apparaît d'un côté, essayer de développer le modèle de l'autre côté. Le résultat est en général un abaissement sensible de la somme S;

(*l*) si plusieurs modèles restent finalement en concurrence, on pourra éventuellement les départager en comparant les sommes S et les diagnostics.

4.4.4. *Exemples (séries G et Z)*

Deux exemples vont éclairer le genre de cheminement qu'il faut réaliser.

(*a*) *Série G : Trafic aérien* (Box et Jenkins, 1970)

L'identification de la série transformée $\{\operatorname{Log} z_t\}$ est résumée dans le tableau III. Elle ne soulève aucune difficulté.

(*b*) *Série Z : ventes mensuelles d'une entreprise.*

L'identification de la série Z (les données seront publiées en Annexe) est résumée dans le tableau IV. Elle conduit au modèle

$$(0, 1, 1)\times(2, 0, 0): \quad \underset{[0,1]}{(1-0{,}604\,B^{12}-0{,}289\,B^{24})}\,\nabla z_t = \underset{[0,1]}{(1-0{,}615\,B)}\,a_t.$$

L'opérateur saisonnier se factorise

$$(1+0{,}314\,B^{12})(1-0{,}918\,B^{12})\,\nabla z_t = (1-0{,}615\,B)\,a_t,$$

et on remarque alors la présence de l'opérateur $(1 - 0{,}918\,B^{12})$ très proche de $\nabla_{12} = 1 - B^{12}$. De fait, l'estimation d'un modèle Sarima $(0, 1, 1)\times(1, 1, 0)$ aboutit à un modèle très voisin

$$(0, 1, 1)\times(1, 1, 0): \quad \underset{[0,1]}{(1+0{,}366\,B^{12})}\,\nabla\nabla_{12}\,z_t = \underset{[0,1]}{(1-0{,}633\,B)}\,a_t.$$

4.5. Une identification controversée : la série P

4.5.1. *Historique*

La série *P* représente les ventes mensuelles d'une certaine entreprise *X*. Cette série a été choisie à dessein pour faire l'objet d'une étude plus détaillée car son identification est assez difficile et a déjà fait couler beaucoup d'encre à la suite de l'article initial de Chatfield et Prothero (1973). Résumons les faits.

Chatfield et Prothero utilisent une transformation en logarithmes décimaux et identifient puis estiment le modèle

$$(1+0{,}47\,B)\,\nabla\nabla_{12}\log_{10} z_t = (1-0{,}81\,B^{12})\,a_t. \qquad (a)$$

Ce modèle se comporte manifestement assez mal en prévision, même en prenant soin d'utiliser les premiers résidus fournis par la méthode des prévisions à rebours. Chatfield et Prothero estiment alors trois autres modèles en $\nabla\nabla_{12}$:

$$\begin{cases} (1+0{,}51\,B)(1+0{,}47\,B^{12})\,\nabla\nabla_{12}\log_{10} z_t = a_t, & (b)\\ (1+0{,}56\,B^{12})\,\nabla\nabla_{12}\log_{10} z_t = (1-0{,}49\,B)\,a_t, & (c)\\ \nabla\nabla_{12}\log_{10} z_t = (1-0{,}44\,B)(1-0{,}85\,B^{12})\,a_t, & (d) \end{cases}$$

pour finalement préférer le modèle (*b*). Compte tenu de toutes les difficultés rencontrées dans le choix entre les quatre modèles (*a*), (*b*), (*c*), (*d*), ils concluent en apportant de sérieuses réserves quant à l'utilisation de la méthode de Box et Jenkins.

Box et Jenkins (1973) répondent en incriminant la transformation logarithmique. Ils suggèrent plutôt une transformation en puissance et identifient la même forme de modèle qu'ils estiment à leur tour

$$(1+0{,}5\,B)\,\nabla\nabla_{12}\,z_t^{1/4} = (1-0{,}8\,B^{12})\,a_t. \qquad (a')$$

O. D. Anderson (1975 *a*) propose le modèle

$$\nabla_{12}\,\mathrm{Log}\,z_t = (1+0{,}467\,B+0{,}715\,B^2)\,a_t, \qquad (e)$$

TABLEAU III

Série G. Choix du modèle (144 observations)

	Sarima	$\hat{S} = \Sigma a_t^2$	Modèles estimés, factorisés, simplifiés	Remarques
1	$(1, 0, 1)\times(1, 0, 1)_{12}$	0,182	$(1 - 0{,}998\,B)(1 - 0{,}999\,9\,B^{12})\,\mathrm{Log}\,z_t$ [0,005] [0,000 9] $= (1 - 0{,}381\,B)(1 - 0{,}568\,B^{12})\,a_t$ [0,08] [0,07]	$\varphi \simeq 1 \rightarrow d > 0$ $\Phi \simeq 1 \rightarrow D > 0$
2	$(1, 0, 1)\times(1, 1, 1)_{12}$	0,163	$(1 - 0{,}990\,B)(1 + 0{,}213\,B^{12})\,\nabla_{12}\,\mathrm{Log}\,z_t$ [0,01] [0,15] $= (1 - 0{,}466\,B)(1 - 0{,}412\,B^{12})\,a_t$ [0,08] [0,14]	$\varphi \simeq 1 \rightarrow d > 0$
3	$(1, 1, 1)\times(1, 1, 1)_{12}$	0,163	$(1 - 0{,}177\,B)(1 + 0{,}119\,B^{12})\,\nabla\nabla_{12}\,\mathrm{Log}\,z_t$ [0,19] [0,14] $= (1 - 0{,}592\,B)(1 - 0{,}533\,B^{12})\,a_t$ [0,15] [0,12]	$\Phi \simeq 0 \rightarrow$ développer à droite
4	$(1, 1, 1)\times(0, 1, 1)_{12}$	0,175	$(1 - 0{,}299\,B)\,\nabla\nabla_{12}\,\mathrm{Log}\,z_t$ [0,18] $= (1 - 0{,}665\,B)(1 - 0{,}628\,B^{12})\,a_t$ [0,14] [0,07]	$\varphi \simeq 0 \rightarrow$ développer à droite
5	$(0, 1, 2)\times(0, 1, 1)_{12}$	0,175	$\nabla\nabla_{12}\,\mathrm{Log}\,z_t$ $= (1 - 0{,}391\,B - 0{,}047\,B^2)(1 - 0{,}616\,B^{12})\,a_t$ [0,09] [0,09] [0,07]	$\theta_2 \simeq 0 \rightarrow q = 1$
6	$(0, 1, 1)\times(0, 1, 1)_{12}$	0,182	$\nabla\nabla_{12}\,\mathrm{Log}\,z_t = (1 - 0{,}396\,B)(1 - 0{,}614\,B^{12})\,a_t$ [0,08] [0,07]	choix final

TABLEAU IV

Série Z. Choix du modèle (64 observations)

	Sarima	$\hat{S} = \Sigma a_t^2$	Modèles estimées, factorisés, simplifiés	Remarques
1	$(1, 0, 1)\times 1, 0, 1)_{12}$	$28,06.10^6$	$(1 - 0{,}970\,B)(1 - 0{,}999\,B^{12})\,\tilde{z}_t = (1 - 0{,}570\,B)(1 - 0{,}581\,B^{12})\,a_t$ [0,03] [0,001] [0,11] [0,1]	$\Phi \simeq 1$ / $\Theta \neq 1$ \ → essayer $D = 1$
2	$(1, 0, 1)\times(1, 1, 1)_{12}$	$14,54.10^6$	$(1 - 0{,}933\,B)(1 + 0{,}494\,B^{12})\,\nabla_{12}\,z_t = (1 - 0{,}532\,B)(1 - 0{,}999\,B^{12})\,a_t$ [0,06] [0,12] [0,15] [0,002]	$\Theta \simeq 1$: essayer $P = 0$ pour voir si ∇_{12} est vraiment inadéquat
3	$(1, 0, 1)\times(0, 1, 1)_{12}$	$26,03.10^6$	$(1 - 0{,}900\,B)\,\nabla_{12}\,z_t = (1 - 0{,}601\,B)(1 - 0{,}999\,9\,B^{12})\,a_t$ [0,02] [0,12] [0,001 5]	$\Theta \simeq 1$: ∇_{12} ne convient pas essayer $d = 1$, $D = 0$
4	$(1, 1, 1)\times(1, 0, 1)_{12}$	$28,40.10^6$	$(1 + 0{,}084\,B)(1 - 0{,}999\,9\,B^{12})\,\nabla z_t = (1 - 0{,}541\,B)(1 - 0{,}573\,B^{12})\,a_t$ [0,2] [0,001] [0,2] [0,1]	$\Phi \simeq 1$: essayer $Q = 0$ pour voir si ∇_{12} apparaît toujours
5	$(1, 1, 1)\times 1, 0, 0)_{12}$	$29,75.10^6$	$(1 + 0{,}077\,B)(1 - 0{,}652\,B^{12})\,\nabla z_t = (1 - 0{,}492\,B)\,a_t$ (2) [0,2] [0,1] [0,1]	$\varphi \simeq 0$: essayer p = 0
6	$(0, 1, 2)\times(1, 0, 0)_{12}$	$29,76.10^6$	$(1 - 0{,}657\,B^{12})\,\nabla z_t = (1 - 0{,}569\,B + 0{,}050\,B^2)\,a_t$ (2) [0,1] [0,1] [0,1]	$\theta_2 \simeq 0$: essayer $q = 1$
7	$(0, 1, 1)\times(1, 0, 0)_{12}$	$29,81.10^6$	$(1 - 0{,}647\,B^{12}\,\nabla z_t = (1 - 0{,}549\,B)\,a_t$ [0,1] [0,1]	$P + D < 2$: on peut essayer encore suridentifier en essayant $P = 2$
8	$(0, 1, 1)\times(2, 0, 0)_{12}$	$21,70.10^6$	$(1 - 0{,}604\,B^{12} - 0{,}289\,B^{24})\,\nabla z_t = (1 - 0{,}615\,B)\,a_t$ [0,1] [0,1] [0,1] $(1 + 0{,}314\,B^{12})(1 - 0{,}918\,B^{12})\,\nabla z_t = (1 - 0{,}615\,B)\,a_t$	Factorisation choix final
A titre indicatif, citons aussi :				
9	$(0, 1, 1)\times(1, 1, 0)_{12}$	$21,85.10^6$	$(1 + 0{,}366\,B^{12})\,\nabla\nabla_{12}\,z_t = (1 - 0{,}633\,B)\,a_t$ [0,1] [0,1]	Modèle très voisin du précédent
10	$(0, 1, 1)\times(0, 1, 1)_{12}$	$22,13.10^6$	$\nabla\nabla_{12}\,z_t = (1 - 0{,}634\,B)(1 - 0{,}999\,9\,B^{12})\,a_t$ [0,1] [0,002]	$\Theta \simeq 1$: essayer $D = 0$

(2) *Un terme négligeable.*

et Melard (1977) propose une transformation sous forme d'une fonction exponentielle du temps

$$(1-0{,}50\,B)\,\nabla\nabla_{12}\,z_t = 0{,}50\exp.(0{,}016\,t)\,a_t. \qquad (f)$$

4.5.2. *Comparaison des estimations*

Ces auteurs ont estimé tous ces modèles par la méthode des prévisions à rebours. De notre côté, l'estimation avec optimisation simultanée des premiers résidus, au lieu de l'équation (*a*), nous conduit à l'équation

$$\underset{[0,11]}{(1+0{,}423\,B)}\,\nabla\nabla_{12}\log_{10} z_t = \underset{[0,000\,01]}{(1-0{,}999\,999\,B^{12})}\,a_t, \qquad (a_1)$$

de sorte que l'opérateur $\nabla\nabla_{12}$ paraît surdifférencié.

La transformation logarithmique n'est pas responsable de cette surdifférenciation. En effet, au lieu de l'équation (*a'*), nous trouvons encore

$$\underset{[0,11]}{(1+0{,}475\,B)}\,\nabla\nabla_{12}\,z_t^{1/4} = \underset{[0,002]}{(1-0{,}999\,9\,B^{12})}\,a_t. \qquad (a_1')$$

Pour les trois autres modèles de Chatfield et Prothero, nous trouvons respectivement :

$$\begin{cases} \underset{[0,10]}{(1+0{,}601\,B)}\,\underset{[0,12]}{(1+0{,}229\,B^{12})}\,\nabla\nabla_{12}\log_{10} z_t = a_t, & (b_1) \\ \underset{[0,11]}{(1+0{,}349\,B^{12})}\,\nabla\nabla_{12}\log_{10} z_t = \underset{[0,12]}{(1-0{,}508\,B)}\,a_t, & (c_1) \\ \nabla\nabla_{12}\log_{10} z_t = \underset{[0,11]}{(1-0{,}373\,B)}\,\underset{[0,001]}{(1-0{,}999\,9\,B^{12})}\,a_t. & (d_1) \end{cases}$$

Quant au modèle d'Anderson nous trouvons (en logarithmes décimaux pour faciliter les comparaisons) :

$$\nabla_{12}\log_{10} z_t = \underset{[0,09]\quad[0,09]}{(1+0{,}432\,B+0{,}717\,B^2)}\,a_t. \qquad (e_1)$$

4.5.3. *Identification*

La situation apparaît pour le moins confuse. Essayons donc la méthode par élimination progressive à partir de la série transformée en logarithmes décimaux (tableau V). Elle conduit à l'identification et à l'estimation du modèle suivant (modèle g_1) :

$$\underset{[0,09]}{(1+0{,}561\,B)}\,\underset{[0,11]\qquad[0,11]}{(1-0{,}631\,B^{12}-0{,}237\,B^{24})}\,\nabla\log_{10} z_t = a_t$$

TABLEAU V

Série P. Choix du modèle (77 observations)

	Sarima	$\hat{S} = \Sigma a_t^2$	Modèles estimés, factorisés, simplifiés	Remarques
1	$(1, 0, 1)\times(1, 0, 1)_{12}$	0,245	$(1 - 0{,}983\,B)(1 - 0{,}927\,B^{12}) \log_{10} \tilde{z}_t = (1 - 0{,}467\,B)(1 - 0{,}999\,9\,B^{12})\, a_t$ [0,02] [0,04] [0,11] [0,001]	$\varphi \simeq 1 \rightarrow d > 0$ $\Phi \simeq \Theta \simeq 1 \rightarrow$ garder $D = 0$
2	$(1, 1, 1)\times(1, 0, 1)_{12}$	0,235	$(1 + 0{,}642\,B)(1 - 0{,}921\,B^{12}) \nabla \log_{10} z_t = (1 + 0{,}234\,B)(1 - 0{,}999\,9\,B^{12})\, a_t$ [0,18] [0,04] [0,23] [0,000 7]	$\Phi \simeq \Theta \simeq 1 \rightarrow$ essayer $P = 0$
3	$(1, 1, 1)\times(0, 0, 1)_{12}$	0,861	$(1 - 0{,}434\,B) \nabla \log_{10} z_t = (1 - 0{,}484\,B)(1 + 0{,}999\,9\,B^{12})\, a_t$ (1) [1,6] [1,6] [0,000 1]	$\Theta \simeq -1$ et $\hat{S}$ grand $\rightarrow$ essayer plutôt $Q = 0$
4	$(1, 1, 1)\times(1, 0, 0)_{12}$	0,388	$(1 + 0{,}683\,B)(1 - 0{,}841\,B^{12}) \nabla \log_{10} z_t = (1 + 0{,}065\,B)\, a_t$ [0,12] [0,06] [0,18]	$\theta \simeq 0 \rightarrow q = 0$
5	$(2, 1, 0)\times(1, 0, 0)_{12}$	0,384	$(1 + 0{,}609\,B - 0{,}059\,B^2)(1 - 0{,}832\,B^{12}) \nabla \log_{10} z_t = a_t$ [0,11] [0,11] [0,06]	$\varphi_2 \simeq 0 \rightarrow p = 1$
6	$(1, 1, 0)\times(1, 0, 0)_{12}$	0,389	$(1 + 0{,}643\,B)(1 - 0{,}844\,B^{12}) \nabla \log_{10} z_t = a_t$ [0,09] [0,06]	$P + D < 2$: on peut encore suridentifier en essayant $P = 2$
7	$(1, 1, 0)\times(2, 0, 0)_{12}$	0,263	$(1 + 0{,}561\,B)(1 - 0{,}631\,B^{12} - 0{,}237\,B^{24}) \nabla \log_{10} z_t = a_t$ [0,09] [0,11] [0,11], $(1 + 0{,}561\,B)(1 + 0{,}265\,B^{12})(1 - 0896\,B^{12}) \nabla \log_{10} z_t = a_t$	Factorisation choix final (g_1)

A titre indicatif, citons ausssi :

8	$(1, 1, 0)\times(1, 1, 0)_{12}$	0,290	$(1 + 0{,}601\,B)(1 + 0{,}229\,B) \nabla\nabla_{12} \log_{10} z_t = a_t$ [0,10] [0,12]	Modèle (b_1) très voisin du précédent
9	$(0, 1, 1)\times(0, 1, 1)_{12}$	0,290	$\nabla\nabla_{12} \log_{10} z_t = (1 - 0{,}373\,B)(1 - 0{,}999\,9\,B^{12})\, a_t$ [0,11] [0,001]	$\Theta \simeq 1$ essayer $D = 0$

(1) *Facteur commun aux deux membres.*

dans lequel l'opérateur saisonnier se factorise

$$(1+0{,}561\,B)(1+0{,}265\,B^{12})(1-0{,}896\,B^{12})\nabla \log_{10} z_t = a_t. \qquad (g_1)$$

Ce modèle est très proche du modèle (b_1). Ce sont donc ces deux modèles (g_1) et (b_1) que nous retiendrons au stade de l'identification.

Nous comparerons les validations et les prévisions correspondant à ces modèles aux paragraphes 5.5 et 6.2.

4.5.4. *Bilan*

L'identification à partir de divers modèles $(1, d, 1)\times(1, D, 1)_{12}$ nous permet de conclure que seul l'opérateur ∇ convient vraiment. L'opérateur ∇_{12} est inadéquat et le modèle (e_1) présente, de toutes façons, une somme résiduelle beaucoup trop forte. Enfin, parmi tous les modèles en $\nabla\nabla_{12}$, on peut éventuellement retenir le modèle (b_1) qui constitue une bonne approximation de notre modèle identifié (g_1).

Les difficultés antérieures paraissent provenir de ce que la méthode d'estimation avec prévisions à rebours n'est pas suffisamment précise lorsque $\hat{\Theta}$ est proche de 1. En effet, le modèle n'est alors plus vraiment réversible et les allers-retours caractéristiques de la méthode des prévisions à rebours convergent trop lentement.

C'est ce que confirment Box et Jenkins eux-mêmes dans leur réponse à Chatfield et Prothero (Box et Jenkins, 1973, remarque en bas de page 339) en signalant qu'une estimation plus poussée du modèle (a') conduit à $\hat{\Theta} = 0{,}9$.

BIBLIOGRAPHIE

J. ABADIE et F. MESLIER, *Présentation synthétique de modèles de prévision à très court terme de l'énergie journalière produite par Électricité de France et de la température moyenne journalière relevée à Paris-Montsouris*, Cahier du Lamsade, n° 10, 1977, Université Paris-IX - Dauphine.

J. ABADIE et F. MESLIER, *Étude de l'utilisation des modèles A.R.I.M.A. pour la prévision à court terme de l'énergie journalière produite par Électricité de France*, R.A.I.R.O. Recherche opérationnelle, vol. 13, n° 1, février 1979, p. 37-54.

O. D. ANDERSON, *On the Collection of Time Series Data*, Oper. Res. Quart., vol. 26, 1975, p. 331-335.

O. D. ANDERSON, *Time Series Analysis and Forecasting: The Box-Jenkins Approach*, 1976, Butterworths, Londres.

O. D. ANDERSON, *A Commentary on "A Survey of Time Series"*, Inst. Stat. Rev., vol. 45, 1977 *a*, p. 273-297.

O. D. ANDERSON, *The Interpretation of Box-Jenkins Time Series Models*, The Statistician, vol. 26, 1977 *b*, p. 127-145.

O. D. ANDERSON, *Box-Jenkins Rules, O.K. — Provided you Know it Makes Sense*, Management Science, vol. 24, n° 11, 1978 *a*, p. 1199.

O. D. ANDERSON, *Orthodox Box-Jenkins Versus "Dynamic Data System Modelling" Secession-a Rejoinder*, Management Science, vol. 24, n° 11, 1978 *b*, p. 1203-1205.

T. W. ANDERSON, *The Statistical Analysis of Time Series*, 1958, Wiley, Londres.

G. E. P. BOX and D. R. COX, *An Analysis of Transformations*, J. Roy. Stat. Soc., vol. B 26, 1964, p. 211-241.

G. E. P. BOX and G. M. JENKINS, *Time Series Analysis, Forecasting and Control*, 1970, Holden Day, San Francisco.

G. E. P. BOX and G. M. JENKINS, *Some Comments on a Paper by Chatfield and Prothero and a Review by Kendall*, J. Roy. Stat. Soc., vol. A 135, 1973, p. 337-352.

G. E. P. BOX and D. A. PIERCE, *Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models*, J. Amer. Stat. Assoc., vol. 65, 1970, p. 1509-1526.

C. CHATFIELD, *The Analysis of Time Series: Theory and Practice*, 1975, Chapman and Hall, Londres.

C. CHATFIELD and D. L. PROTHERO, *Box-Jenkins Seasonal Forecasting: Problems in a Case-Study*, J. Roy. Stat. Soc., vol. A 136, 1973, p. 295-336.

G. V. GLASS, V. L. WILLSON and J. M. GOTTMAN, *Design and Analysis of Time Series Experiments*, 1975, Colorado Associated University Press, Boulder, Colorado.

C. W. J. GRANGER and P. NEWBOLD, *Forecasting Transformed Series*, J. Roy. Stat. Soc., vol. B 38, 1976, p. 189-203.

D. M. HIMMELBLAU, *Applied Nonlinear Programming*, 1972, McGraw-Hill, New York.

M. G. KENDALL, *Review of Box and Jenkins*, J. Roy. Stat. Soc., vol. A 134, 1971, p. 450-453.

M. G. KENDALL, *Time Series*, 1973, Griffin, Londres.

M. G. KENDALL and A. STUART, *The Advanced Theory of Statistics* (Vol. III ch. 45-50), 1976, Griffin, Londres.

H. J. LENZ, *Strategies for Implementation of a Fully Automatic Box and Jenkins Forecasting Technique*, Discussionsarbeiten n° 7/78, 1978, Institut für Quantitative Okonomik und Statistik, Freie Universität, Berlin.

S. MAKRIDAKIS, *A Survey of Time Series*, Int. Stat. Rev., vol. 44, 1976, p. 29-70.

S. MAKRIDAKIS, *Time-Series Analysis and Forecasting: an Update and Evaluation*, Int. Stat. Rev., vol. 46, 1978, p. 255-278.

S. MAKRIDAKIS and S. C. WHEELWRIGHT, *Forecasting Methods for Management*, 1973, Wiley, New York.

S. MAKRIDAKIS and S. C. WHEELWRIGHT, *Interactive Forecasting*, 1978, Holden Day, San Francisco.

G. MELARD, *Prévision des ventes*, Communication au 4e Colloque international d'Économétrie appliquée, 1977, Strasbourg, Faculté des Sciences sociales et politiques, Institut de Statistique, Campus Plaine, Université libre de Belgique.

G. V. MOGHA, *Sur un problème d'estimation du modèle Arima*, E.D.F., Bulletin de la Direction des Études et Recherches, série C, suppl. n° 1, 1977, p. 87-132.

T. H. NAYLOR, T. G. SEAKS and D. W. WICHERN, *Box-Jenkins Methods: An Alternative to Econometric Models*, Inst. Stat. Rev., vol. 40, 1972, p. 123-137.

C. R. NELSON, *Applied Time Series Analysis for Managerial Forecasting*, 1973, Holden Day, San Francisco.

P. NEWBOLD, *The Principles of the Box-Jenkins Approach*, Oper. Res. Quart., vol. 26, 1975, p. 397-412.

H. J. STEUDEL and S. M. WU, *A Time Series Approach to Queueing Systems with Applications for Modelling Job-Shop in -Process Inventories*, Management Science, vol. 7, 1977, p. 745-755.

H. J. STEUDEL and S. M. WU, *Dynamic Data System Modelling-Revisited*, Management Science, vol. 24, 1978, p. 1200-1202.