

J. ABADIE

D. TRAVERS

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

*Revue française d'automatique, d'informatique et de recherche
opérationnelle. Recherche opérationnelle*, tome 15, n° 1 (1981),
p. 51-71.

http://www.numdam.org/item?id=RO_1981__15_1_51_0

© AFCET, 1981, tous droits réservés.

L'accès aux archives de la revue « Revue française d'automatique, d'informatique et de recherche opérationnelle. Recherche opérationnelle » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/legal.php>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

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

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 avons proposé précédemment 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. Nous exposons ici les procédés de validation, qui sont très simples, et quelques expériences numériques, incluant la prévision.

Mots clés : séries temporelles, Box et Jenkins.

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 have previously described a simplified interactive approach based on simple algebraic manipulations (factorization and roots of polynomials, etc.). This enables the method to be handled by the average user. We explain here very simple validation procedures, and some numerical experiments, including forecasting.

Keywords: Time series, Box-Jenkins.

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 déjà d'un emploi très général et son intérêt n'est plus à démontrer. Cependant, la phase initiale d'identification se révèle, en pratique, d'être une étape délicate. Elle repose sur l'analyse des fonctions d'autocorrélation et d'autocorrélation partielle de la série et il arrive

(*) Reçu octobre 1978.

⁽¹⁾ Université Pierre-et-Marie-Curie, Paris-VI, Institut de Programmation, 4, place Jussieu, 75230 Paris Cedex 05.

⁽²⁾ C.E.S.A., 1, rue de la Libération, 78350 Jouy-en-Josas, France.

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 avons proposé dans un précédent article (Abadie et Travers, 1980, I, § 1 à 4) 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 [I, § 3 ⁽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 ») (I, § 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 (I, § 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 ou Sarima 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 : c'est ce que nous exposons au paragraphe 5 ci-dessous.

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 au paragraphe 6 les conséquences du choix d'un modèle sur la prévision.

⁽³⁾ Les paragraphes 1 à 4 figurent dans notre précédent article cité en référence.

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 certain nombre de séries.

Signalons enfin qu'un appendice (avant l'annexe) expose une méthode pratique de calcul en machine des dérivées de la fonction $S(x)$ (somme des carrés des écarts).

5 (4). VALIDATION

Nous avons vu comment il était possible d'éviter l'interprétation difficile des corrélogrammes au stade de l'identification en pratiquant l'identification par élimination progressive. Au stade de la validation, alors qu'il ne s'agit plus que de vérifier que les résidus constituent bien un bruit blanc, le corrélogramme redevient un outil simple et utile, que nous complétons par le test du chi 2.

5.1. Corrélogramme résiduel

Rappelons d'abord les définitions des grandeurs suivantes, indépendantes du temps (processus stationnaires). On pose :

$$\mu = E(z_t), \quad \tilde{z}_t = z_t - \mu.$$

(a) *Autocovariances* :

$$\left. \begin{aligned} \gamma_k = \gamma_{-k} = \text{cov}(\tilde{z}_t, \tilde{z}_{t-k}) &= E(\tilde{z}_t \tilde{z}_{t-k}), & k \neq 0, \\ \gamma_0 = E(\tilde{z}_t^2) &= \sigma_z^2. \end{aligned} \right\} \quad (5.1.1)$$

(b) *Autocorrélations* :

$$\left. \begin{aligned} \rho_k = \rho_{-k} &= \frac{E(\tilde{z}_t \tilde{z}_{t-k})}{\sqrt{E(\tilde{z}_t^2) \cdot E(\tilde{z}_{t-k}^2)}} = \frac{\gamma_k}{\gamma_0} \leq 1, & k \neq 0, \\ \rho_0 &= 1. \end{aligned} \right\} \quad (5.1.2)$$

(c) *Fonction d'autocorrélation ou corrélogramme*

Il s'agit de la fonction symétrique $\rho_k = f(k)$ dont on ne représente que la partie correspondant à $k \geq 0$. Pour un bruit blanc, le corrélogramme théorique se réduit à $\rho_k = 0$ pour tout entier $k \geq 1$.

(4) Les paragraphes 1 à 4 figurent dans notre précédent article, cité en référence.

Dans la pratique les coefficients γ_k et ρ_k sont estimés à partir des formules :

$$\hat{\gamma}_k = C_k = \frac{1}{N} \sum_{t=k+1}^N (z_t - \bar{z})(z_{t-k} - \bar{z}), \quad (5.1.3)$$

$$\hat{\rho}_k = r_k = \frac{C_k}{C_0}, \quad (5.1.4)$$

où l'on pose, comme d'habitude :

$$\bar{z} = \frac{1}{N} \sum_{t=1}^N z_t.$$

Même si les ρ_k théoriques sont nuls, les $\hat{\rho}_k$ calculés ne sont évidemment jamais exactement nuls. Il faut donc également calculer $\text{Var}(\hat{\rho}_k)$, de façon à déceler les $\hat{\rho}_k$ qui seraient significativement différents de zéro. On prend habituellement la formule approchée $\text{Var}(\hat{\rho}_k) = 1/n$, où, comme au paragraphe 3.4, on a posé $n = N - d - sD$.

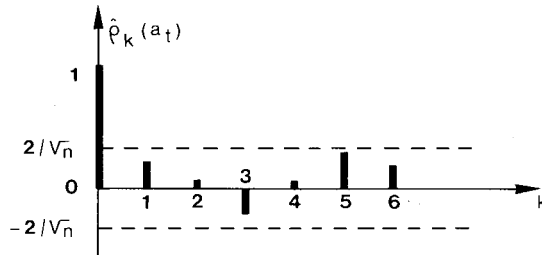


Figure 3

Le corrélogramme d'un bruit blanc a donc l'allure indiquée à la figure 3, où les lignes pointillées représentent un intervalle de confiance voisin de 95 % autour de la valeur théorique zéro. On devrait normalement s'attendre à trouver tous les $\hat{\rho}_k$ (sauf $\rho_0 = 1$) à l'intérieur de cet intervalle. Pourtant, la probabilité de 95 % s'applique à chaque coefficient pris individuellement et il serait donc tolérable que sur un ensemble, disons, d'une vingtaine de coefficients, un ou deux d'entre eux sortent des limites prévues.

5.2. Test du chi 2

Pour construire un test global plus rigoureux, on peut calculer la quantité :

$$Q_k = n \sum_{i=1}^k \hat{\rho}_i^2. \quad (5.2.1)$$

On montre que, pour un modèle Sarima $(p, d, q) \times (P, D, Q)_s$, Q_k se comporte approximativement comme un χ^2 à $k - p - q - P - Q$ degrés de liberté, sous réserve que $n \gg k$ et que k soit lui-même suffisamment grand (par exemple $k = 25$ pour une série non saisonnière avec $n > 200$). On peut donc vérifier que la valeur de Q_k n'est pas trop élevée en la comparant aux valeurs théoriques lues dans une table du χ^2 .

En fait, ce test ne s'avère pas très sensible et on peut utiliser, à la place de Q_k (O. D. Anderson, 1977 a) :

$$Q'_k = n(n+2) \sum_{i=1}^k \frac{1}{n-i} \hat{\rho}_k^2,$$

qui possède une valeur supérieure.

5.3. Fonction d'autocorrélation partielle et périodogramme cumulé

L'étude de la fonction d'autocorrélation va de pair avec celle de la fonction d'autocorrélation partielle et nous avons vu comment on peut se passer de ces notions au stade de l'identification. Au stade de la validation, on peut aussi calculer la fonction d'autocorrélation partielle des résidus. C'est ce que nous avons fait pour un grand nombre de séries et de modèles sans que ce test n'apporte d'information qui ne soit déjà contenue dans le corrélogramme. En conséquence, nous proposons de nous passer complètement de la notion d'autocorrélation partielle que nous ne décrivons donc pas ici et nous renvoyons le lecteur intéressé à l'ouvrage de Box et Jenkins.

Dans le cas des séries saisonnières, on peut éventuellement effectuer aussi le test du périodogramme cumulé pour s'assurer que la périodicité de la série a été bien prise en compte. Nous n'avons pas utilisé ce test.

5.4. Exemples (séries Z et P)

Examinons, à titre d'exemple les corrélogrammes résiduels de quelques uns des modèles estimés pour les séries Z et P .

(a) Série Z , $N = 64$ observations (voir § 4.4.4) :

$$\begin{cases} (1 + 0,314 B^{12}) (1 - 0,918 B^{12}) \nabla z_t = (1 - 0,615 B) a_t, \\ (1 + 0,366 B^{12}) \nabla \nabla_{12} z_t = (1 - 0,633 B) a_t. \end{cases}$$

Le diagnostic est excellent dans les deux cas :

1^{er} cas :

$$(0, 1, 1) \times (2, 0, 0) \quad \begin{cases} \bar{a} = 62, \\ \hat{\sigma}_a = 579, \end{cases} \quad \begin{cases} Q_{48} = 17,8, \\ \chi^2_{5\%}(45) = 61,4. \end{cases}$$

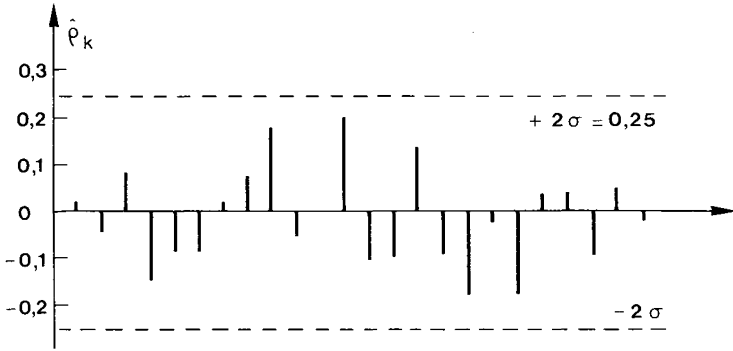


Figure 4. — Série Z, $(0, 1, 1) \times (2, 0, 0)$.

2^e cas :

$$(0, 1, 1) \times (1, 0, 0) \quad \begin{cases} \bar{a} = 58, \\ \hat{\sigma}_a = 581, \end{cases} \quad \begin{cases} Q_{50} = 13,3, \\ \chi^2_{5\%}(48) = 64,9. \end{cases}$$

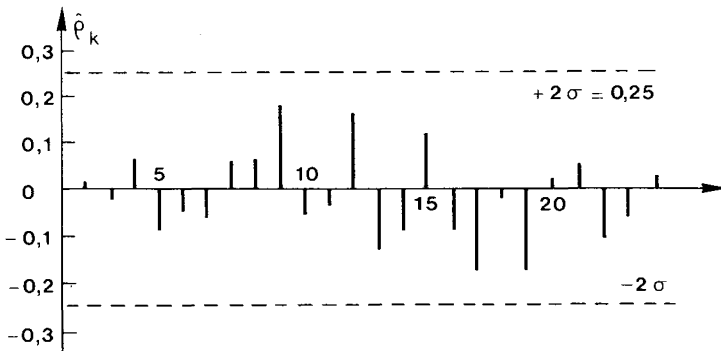


Figure 5. — Série Z, $(0, 1, 1) \times (1, 1, 0)$.

(b) Série P, $N=77$ observations (voir § 4.5).

Le diagnostic du modèle (g_1) est le suivant :

$$\begin{cases} \bar{a} = 0,003, \\ \hat{\sigma}_a = 0,071, \end{cases} \quad \begin{cases} Q_{50} = 42,3, \\ \chi^2_{5\%}(47) = 63,7. \end{cases}$$

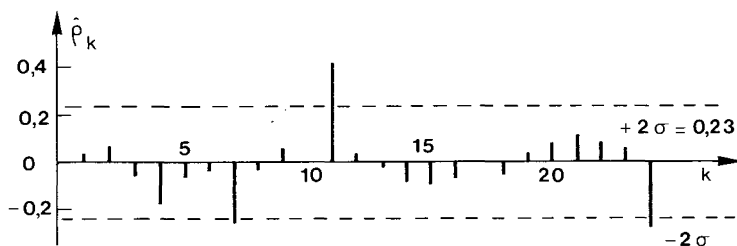


Figure 6. — Série P, modèle (g_1).

On trouve trois valeurs fortes à son corrélogramme résiduel : $\hat{\rho}_7 = -0,25$; $\hat{\rho}_{11} = 0,42$; $\hat{\rho}_{24} = -0,27$; mais, sur un ensemble de 24 coefficients, ce n'est pas trop surprenant, et on peut donc considérer le modèle comme acceptable.

Le diagnostic du modèle (b_1) est très voisin de celui du modèle (g_1), avec $\bar{a} = 0,002$, $\hat{\rho}_a = 0,057$ et $Q_{50} = 34$. Son corrélogramme résiduel présente, lui aussi, deux valeurs fortes $\hat{\rho}_7 = -0,29$ et $\hat{\rho}_{11} = 0,40$.

Notons au passage que les modèles (a) et (e_1), par exemple, ont eux aussi des corrélogrammes résiduels qui comportent quelques valeurs fortes :

$$\begin{cases} \text{modèle } (a) : \hat{\rho}_7 = -0,29, & \hat{\rho}_{11} = 0,34, \\ \text{modèle } (e_1) : \hat{\rho}_3 = 0,25, & \hat{\rho}_{11} = 0,23, & \hat{\rho}_{12} = 0,23. \end{cases}$$

Quel que soit le modèle, on trouve toujours une valeur résiduelle $\hat{\rho}_{11}$ élevée.

5.5. Conclusion

L'utilisateur familier des notions de corrélogramme peut trouver dans un diagnostic négatif des informations quant à la direction dans laquelle la forme du modèle doit être modifiée. Mais il ne s'agit, en pratique, que d'un cas relativement rare, car la validation ne semble pas être une épreuve très difficile. Bien des modèles doivent être modifiés. Mais il ne s'agit, que d'un cas relativement rare, car la validation ne semble pas être une épreuve très difficile. Bien des modèles qui ont été rejetés au cours de notre processus d'identification par élimination progressive présentaient en fait un diagnostic favorable. La

Ceci renforce l'importance du choix du modèle : c'est de lui que dépend directement la forme de la fonction qui sera utilisée en prévision.

6. CONSÉQUENCES DU CHOIX D'UN MODÈLE SUR LA PRÉVISION

6.1. Calcul des prévisions

L'analyse d'une série temporelle (choix du modèle et validation) ne constitue que rarement une fin en soi. Le véritable objectif poursuivi est la prévision.

Les prévisions $\hat{z}_t(h)$ effectuées à la date t pour l'horizon h (c'est-à-dire pour la date $t+h$) s'obtiennent à partir du modèle écrit sous sa forme générale en remplaçant chaque terme par son espérance mathématique, soit :

$$\left. \begin{array}{l} E[z_j] = \hat{z}_t(j-t) \\ E[a_j] = 0 \end{array} \right\} \text{ pour } j > t,$$

$$\left. \begin{array}{l} E[z_j] = z_j \\ E[a_j] = a_j \end{array} \right\} \text{ pour } j \leq t,$$

Par a_j ($j \leq t$), on entend ici les valeurs $a_j = \hat{a}_j$ calculées par récurrence et qui correspondent à l'optimum $\hat{S} = \min(\sum a_t^2)$. Ces valeurs doivent donc être conservées en mémoire sur l'ordinateur. (Si on ne les a pas conservées, on pourra reconstituer les \hat{a}_j utiles à partir des \hat{a}_* optimisés en même temps que les autres paramètres.)

L'ensemble des prévisions effectuées à une même date pour différents horizons h s'appelle *fonction de prévision*. Chaque modèle possède une fonction de prévision qui lui est propre.

La prévision ainsi obtenue est sans biais. La variance de l'erreur de prévision d'horizon h est donnée par :

$$V(h) = (1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{h-1}^2) \hat{\sigma}_a^2 \quad (6.1.1)$$

(voir § 2.3 pour la définition des ψ_j). En particulier, les résidus a_t s'interprètent comme les erreurs de prévision d'horizon 1 commises avec le modèle considéré : la méthode des moindres carrés ($\min S = \sum a_t^2$) est optimale lorsqu'il s'agit de prévision à l'horizon 1.

6.2. Exemples de fonctions de prévision

Examinons quelques formes typiques de fonctions de prévision pour différents modèles Arima (p, d, q).

6.2.1. Arima (0, 1, 1)

Pour le modèle :

$$\nabla z_t = (1 - \theta B) a_t,$$

l'équation aux différences s'écrit :

$$z_{t+h} = z_{t+h-1} + a_{t+h} - \theta a_{t+h-1},$$

soit, en prenant l'espérance mathématique à la date t :

$$\begin{aligned} \hat{z}_t(1) &= z_t - \theta a_t, \\ \dots \dots \dots \\ \hat{z}_t(h) &= \hat{z}_t(h-1) = \hat{z}_t(1), \quad h \geq 2. \end{aligned}$$

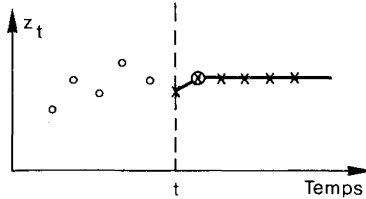


Figure 7

On voit que la fonction de prévision est constante :

$$\hat{z}_t(h) = z_t - \theta a_t = \hat{z}_t(1).$$

a_t est l'erreur de prévision d'horizon 1 faite à la date $t-1$:

$$a_t = z_t - \hat{z}_{t-1}(1) = z_t - \hat{z}_{t-1}(h).$$

On en déduit :

$$\hat{z}_t(h) = (1 - \theta) z_t + \theta \hat{z}_{t-1}(h).$$

On reconnaît la formule de mise à jour très simple caractéristique du lissage exponentiel (avec la valeur θ qui correspond à l'optimum pour l'horizon 1).

6.2.2. Arima (1, 1, 0)

Le modèle est :

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

(avec un terme constant θ_0 pour plus de généralité, § 2.1). L'équation aux différences s'écrit :

$$z_{t+h} = (\varphi + 1) z_{t+h-1} - \varphi z_{t+h-2} + \theta_0 + a_{t+h},$$

soit, en prenant l'espérance mathématique à la date t :

$$\begin{aligned} \hat{z}_t(1) &= (\varphi + 1) z_t - \varphi z_{t-1} + \theta_0, \\ \hat{z}_t(2) &= (\varphi + 1) \hat{z}_t(1) - \varphi z_t + \theta_0, \\ \dots \dots \dots \\ \hat{z}_t(h) &= (\varphi + 1) \hat{z}_t(h-1) - \varphi \hat{z}_t(h-2) + \theta_0, \quad h \geq 3. \end{aligned}$$

Ces équations aux différences finies du second ordre (en h), où les deux premières spécifient les conditions initiales, ont pour solution unique :

$$\hat{z}_t(h) = z_t + \frac{\varphi(1-\varphi^h)}{1-\varphi}(z_t - z_{t-1}) + \frac{h\theta_0}{1-\varphi} - \frac{\varphi(1-\varphi^h)\theta_0}{(1-\varphi)^2}.$$

Comme $|\varphi| < 1$, la fonction de prévision est asymptote à la droite :

$$y_t(h) = z_t + \frac{\varphi}{1-\varphi}(z_t - z_{t-1}) - \frac{\varphi\theta_0}{(1-\varphi)^2} + \frac{\theta_0}{1-\varphi}h.$$

Les figures 8 et 9 montrent deux exemples.

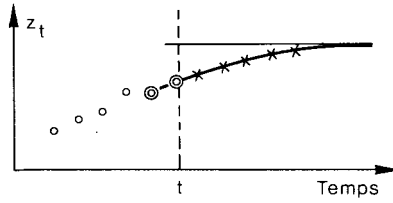


Figure 8

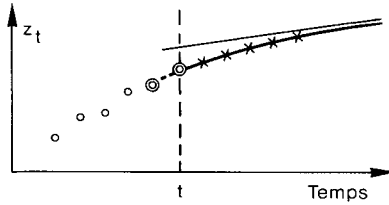


Figure 9

6.2.3. *Arima* (1, 0, 1)

Considérons le modèle :

$$(1 - \varphi B)\tilde{z}_t = (1 - \theta B)a_t \quad \text{avec} \quad \tilde{z}_t = z_t - \bar{z}.$$

L'équation aux différences s'écrit :

$$\tilde{z}_{t+h} = \varphi \tilde{z}_{t+h-1} + a_{t+h} - \theta a_{t+h-1},$$

c'est-à-dire en prenant l'espérance mathématique à la date t :

$$\tilde{z}_t(1) = \varphi \tilde{z}_t - \theta a_t,$$

.....,

$$\tilde{z}_t(h) = \varphi \tilde{z}_t(h-1), \quad h \geq 2$$

La fonction de prévision est donc :

$$\tilde{z}_t(h) = \varphi^h \tilde{z}_t - \theta \varphi^{h-1} a_t.$$

Elle se rapproche exponentiellement de la moyenne \bar{z} à partir de la première prévision $\tilde{z}_t(1)$. Par exemple, pour le modèle $(1 - 0,8B)\tilde{z}_t = (1 - 0,4B)a_t$, on a typiquement la figure 10.

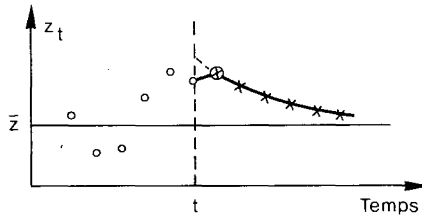


Figure 10

6.3. Comparaison des prévisions

Une méthode d'estimation ne se juge pas sur le nombre de décimales qu'elle permet de retenir pour chaque paramètre. En fait, du point de vue de la prévision, une estimation avec un ou deux chiffres significatifs suffit le plus souvent : on constate en effet que les prévisions sont très stables par rapport au choix des valeurs des paramètres. Cette stabilité est d'ailleurs un trait fondamental de la méthode de Box et Jenkins. Elle permet de ne pas avoir besoin d'estimer à nouveau le modèle chaque fois qu'une nouvelle valeur de la série devient disponible.

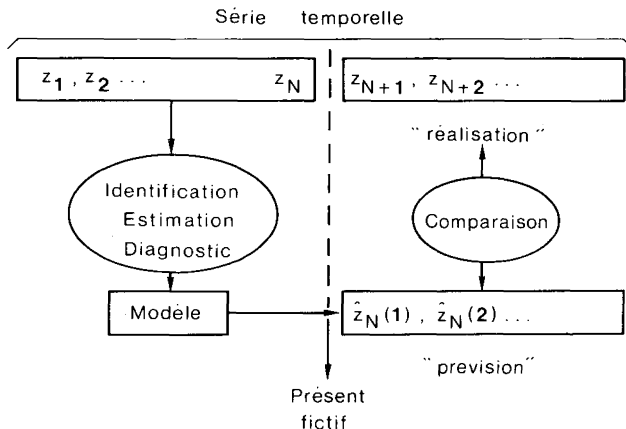


Figure 11

C'est plutôt la forme du modèle qui conditionne l'allure de la fonction de prévision. Il est donc intéressant d'effectuer des comparaisons sur les séries pour lesquelles plusieurs modèles sont restés en concurrence (séries Z et P).

Jusqu'ici, les différents modèles envisagés pour décrire une même série étaient comparés sur la base de leur adéquation à la série : somme résiduelle faible, indépendance des résidus, etc. Pour comparer les performances en prévision de différents modèles, il faut pouvoir rapprocher les prévisions et la réalité.

On utilisera donc un artifice qui consiste à n'employer qu'une partie de la série pour la phase d'analyse et à simuler une prévision qui pourra être comparée aux valeurs réelles gardées en réserve (fig. 11).

Cette comparaison peut se faire par inspection graphique, ou de manière plus formelle, à partir des écarts prévisions/réalisations $\hat{z}_N(h) - z_{N+h}$, par exemple en calculant l'écart quadratique moyen :

$$E.Q.M. = \sqrt{\frac{\sum_{h=1}^H (\hat{z}_N(h) - z_{N+h})^2}{H}} \quad (6.3.1)$$

C'est ce que nous allons faire pour la série P prise pour exemple.

$N=77$ observations sont utilisées pour l'estimation de la série P . Les prévisions sont effectuées à la date 77 pour 6 périodes; nous négligeons le biais introduit par la transformation logarithmique (environ 1 % pour $h=1$). Tous les modèles comparés ont été introduits au paragraphe 4.5.

TABLEAU VI

Modèles estimés par la méthode des prévisions à rebours

h	1	2	3	4	5	6	E.Q.M.
$(a_0) : (1+0,47 B) \nabla \nabla_{12} \log z_t = (1-0,81 B^{12}) a_t$. . .	305	482	673	990	1297	1387	362
$(a) : idem$ avec $a_* \neq 0$	-	-	-	-	-	1221	-
$(b) : (1+0,51 B) (1+0,47 B^{12}) \nabla \nabla_{12} \log z_t = a_t$	-	-	452	-	-	990	-
$(c) : (1+0,56 B^{12}) \nabla \nabla_{12} \log z_t = (1-0,49 B) a_t$	-	-	-	-	-	940	-
$(d) : \nabla \nabla_{12} \log z_t = (1-0,44 B) (1-0,85 B^{12}) a_t$	-	-	-	-	-	1225	-
$(a') : (1+0,5 B) \nabla \nabla_{12} z_t^{1/4} = (1-0,8 B^{12}) a_t$	286	409	511	761	966	1091	147
$(f) : (1-0,50 B) \nabla \nabla_{12} z_t = 0,50 \exp(0,016 t) a_t$	262	368	428	711	887	926	69
Réalisation	260	304	390	614	783	872	

Pour le premier modèle, les prévisions notées (a_0) (voir tableau VI) sont effectuées en prenant $a_* = 0$. Cette approximation est bien sûr insuffisante, surtout avec $\hat{\theta}$ proche de 1, et les autres prévisions utilisent les valeurs a_* prévues à rebours pour les premiers résidus. Les modèles et les prévisions sont ceux de Chatfield et Prothero (1973) [modèles (a_0) , (a) , (b) , (c) , (d)], Box et Jenkins (1973) [modèle (a')], et Melard (1977) [modèle (f)].

TABEAU VII

Modèles estimés avec optimisation simultanée des premiers résidus

h	1	2	3	4	5	6	E.Q.M.
(b_1) : $(1+0,60 B)(1+0,23 B^{12}) \nabla \nabla_{12} \log z_t = a_t$	249	384	438	748	941	1 007	94
(c_1) : $(1+0,35 B^{12}) \nabla \nabla_{12} \log z_t = (1-0,51 B) a_t$	269	345	447	694	926	967	83
(e_1) : $\nabla_{12} \log z_t = (1+0,43 B+0,72 B^2) a_t$	248	402	404	677	858	895	58
(g_1) : $(1+0,56 B)(1-0,63 B^{12}-0,24 B^{24}) \nabla \log z_t = a_t$	254	362	416	646	801	843	32
Réalisation.....	260	304	390	614	783	872	

Ici (tableau VII), les prévisions sont toutes effectuées en utilisant les valeurs \hat{a}_* des premiers résidus à l'optimum. Les modèles (a_1) , (a'_1) et (d_1) de mêmes formes que a_0 ou a , a' , d) ont été rejetés car ils correspondent après calculs à $\hat{\theta} = 1$. Le modèle (e_1) est très proche du modèle (e) retenu par Anderson (1975 a) et le modèle (g_1) est le modèle auquel on aboutit en pratiquant l'identification par élimination progressive (tableau V).

Chatfield et Prothero n'indiquent pas toutes les prévisions de $h = 1$ à 6, mais les valeurs de \hat{z}_{77} (6) pour novembre 1971 suffisent pour permettre la comparaison de tous les modèles.

Les prévisions (a) , (a_0) , (a') et (d) sont manifestement « fortes ». Elles proviennent toutes de modèles que nous avons trouvés mal estimés, puisque l'estimation avec optimisation simultanée des premiers résidus conduit à $\hat{\theta} = 1$. Les autres prévisions sont plus « raisonnables » et, en tous cas, plus proches de la réalisation. Le modèle (g_1) que nous avons identifié et estimé fournit les meilleures prévisions si on s'en tient au critère de l'écart quadratique moyen prévision/réalisation (voir fig. 12).

L'écart quadratique moyen n'est bien sûr pas un critère définitif, et la réalisation elle-même peut sembler un peu « faible ». Ceci pose le problème de la prévision à plusieurs périodes : les modèles (b_1) et (c_1) ont un facteur $\nabla \nabla_{12} = (1-B)(1-B^{12})$ qui produit des oscillations saisonnières entretenues, alors que pour le modèle (g_1) le facteur correspondant est $(1-B)(1-0,9 B^{12})$ (voir tableau V) qui produit des oscillations amorties (fig. 13). Il n'est donc pas

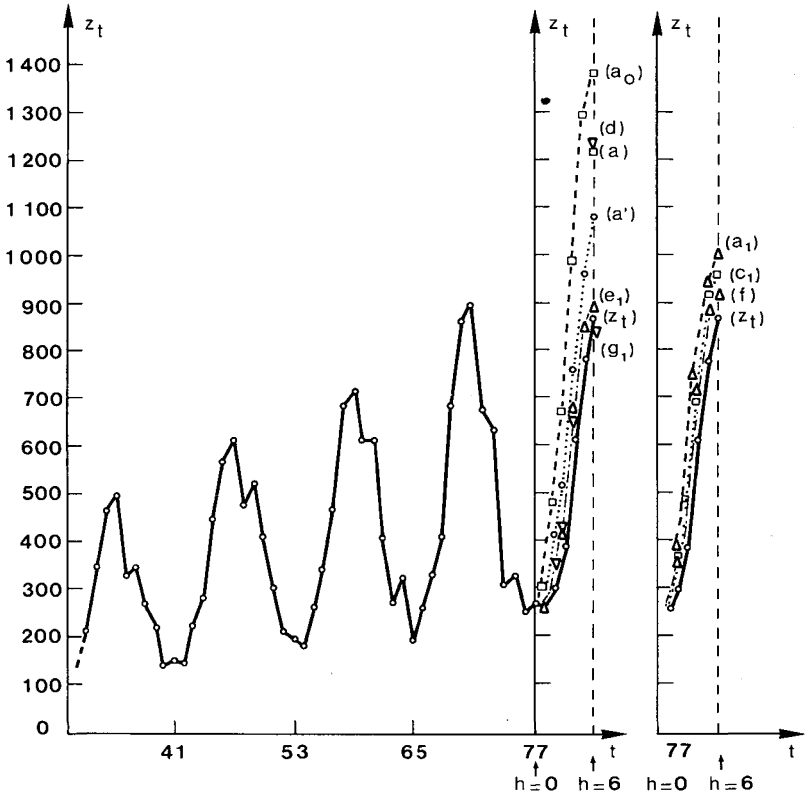


Figure 12. - Série P.

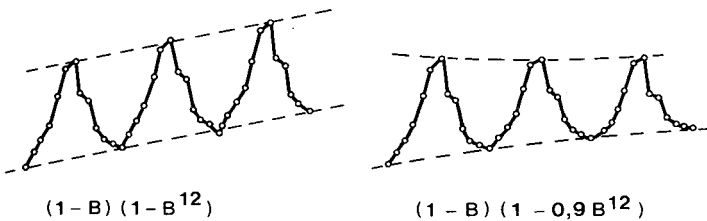


Figure 13

étonnant que les prévisions du modèle (g_1) soient systématiquement plus faibles que celles des modèles (b) et (c_1) , même pour des horizons modérés. Cette seule présence, ou absence, d'amortissement peut faire préférer l'un ou l'autre modèle, lorsque les valeurs correspondantes de S sont voisines l'une de l'autre (voir g_1 et b_1 , tableau V par exemple).

APPENDICE
DÉRIVÉES DE S

Nous avons employé, pour estimer le vecteur des paramètres :

$$x = (x_1, \dots, x_m) = (\varphi_1, \dots, \varphi_p, \theta_0, \theta_1, \dots, \theta_q, \Phi_1, \dots, \Phi_p, \Theta_1, \dots, \Theta_q), \quad (A.1)$$

une procédure d'optimisation basée sur le calcul du gradient de la fonction :

$$S(x) = \sum_{t=t_0+1-r}^N a_t^2, \quad (A.2)$$

où a_t est défini par :

$$\theta_0 + \theta(B)\Theta(B^s)a_t = \varphi(B)\Phi(B^s)w_t, \quad t = t_0 + 1, \dots, N \quad (A.3)$$

et où :

$$t_0 = d + sD + p + sP, \quad r = q + sQ. \quad (A.4)$$

Rappelons que a_* , défini par :

$$a_* = (a_{t_0+1-r}, \dots, a_{t_0}) \quad (A.5)$$

est un vecteur inconnu à estimer en même temps que x .

Étant données les formes particulières des opérateurs :

$$\left. \begin{aligned} \varphi(B) &= 1 - \varphi_1 B - \dots - \varphi_p B^p \\ \theta(B) &= 1 - \theta_1 B - \dots - \theta_q B^q \\ \Phi(B^s) &= 1 - \Phi_1 B^s - \dots - \Phi_p B^{ps} \\ \Theta(B^s) &= 1 - \Theta_1 B^s - \dots - \Theta_q B^{qs} \end{aligned} \right\} \quad (A.6)$$

la formule (A.3) constitue une formule itérative pour le calcul de a_t , $t = t_0 + 1, \dots, N$, à partir de la série $\{w_t\}$, et de x, a_* . Il en sera de même pour le calcul des dérivées partielles par les formules qui vont suivre.

(A.3) donne la formule de récurrence suivante pour la dérivée partielle de a_t par rapport à φ_i :

$$\theta(B)\Theta(B^s)\frac{\partial a_t}{\partial \varphi_i} = -B^i \Phi(B^s)w_t.$$

Cette formule se simplifie si $\theta_0 = 0$:

$$\varphi(B)\frac{\partial a_t}{\partial \varphi_i} = -B^i a_t = -a_{t-1}. \quad (A.7)$$

On obtient de la même façon :

$$\theta(B) \frac{\partial a_t}{\partial \theta_j} = B^j a_t = a_{t-j}, \tag{A.8}$$

$$\Phi(B^s) \frac{\partial a_t}{\partial \Phi_k} = -B^{ks} a_t = -a_{t-ks}, \tag{A.9}$$

$$\Theta(B^s) \frac{\partial a_t}{\partial \Theta_l} = B^{ls} a_t = a_{t-ls}, \tag{A.10}$$

$$\theta(B) \Theta(B^s) \frac{\partial a_t}{\partial \theta_0} = -1. \tag{A.11}$$

Ces formules permettent le calcul successif des dérivées de a_{t_0+1}, \dots, a_N , à partir de x, a_* , et de :

$$\frac{\partial a_t}{\partial \varphi_i} = \frac{\partial a_t}{\partial \theta_j} = \frac{\partial a_t}{\partial \Phi_k} = \frac{\partial a_t}{\partial \Theta_l} = \frac{\partial a_t}{\partial \theta_0} = 0, \quad t < t_0. \tag{A.12}$$

En ce qui concerne les dérivées partielles par rapport aux composantes de a_* , on a, pour $t = t_0 + 1, \dots, N$,

$$\theta(B) \Theta(B^s) \frac{\partial a_t}{\partial a_\alpha} = 0, \quad \alpha = t_0 + 1 - r, \dots, t_0, \tag{A.13}$$

avec, pour $t = t_0 + 1 - r, \dots, t_0$ et $\alpha = t_0 + 1 - r, \dots, t_0$:

$$\frac{\partial a_t}{\partial a_\alpha} = \begin{cases} 0, & t \neq \alpha, \\ 1, & t = \alpha. \end{cases} \tag{A.14}$$

Une dérivée partielle quelconque de $S(x)$ en (A.2) se calcule alors très simplement par :

$$\frac{\partial S}{\partial x_\beta} = 2 \sum_{t=t_0+1-r}^N a_t \frac{\partial a_t}{\partial x_\beta}, \quad \beta = 1, \dots, m. \tag{A.15}$$

Remarquons que le nombre des mémoires nécessaires au calcul n'est pas élevé : pour chaque valeur de β , on fait croître t de $t_0 + 1$ à N ; il faut retenir en mémoire un certain nombre de valeurs antérieures de $\partial a_t / \partial x_\beta$. Comme on peut utiliser le même groupe de mémoires, le nombre des mémoires à réserver est le plus grand des deux nombres $p + sP, q + sQ$.

Le calcul des dérivées secondes se fait suivant le même schéma.

Ces procédures de calcul, facilement programmables, permettent l'utilisation des méthodes les plus récentes pour la minimisation de $S(x)$ (voir par exemple Abadie, 1979, pour une vue d'ensemble).

ANNEXE

REVUE DES MODÈLES RETENUS POUR QUELQUES SÉRIES

Nous présentons une revue rapide des modèles que nous avons retenus pour un certain nombre de séries, y compris celles qui ont été examinées ici à titre d'exemples. Les chiffres entre crochets au-dessous des paramètres d'un modèle indiquent l'écart-type de l'estimation correspondante. La variance résiduelle à l'optimum $\hat{\sigma}_a^2$ est calculée comme indiqué au paragraphe 3.2.3 (du moins pour les modèles que nous avons nous-mêmes estimés). Nous citons aussi les résultats obtenus par d'autres auteurs.

Série A : Concentrations chimiques relevées toutes les 2 heures (197 observations).

Référence : Box et Jenkins, 1970.

Deux modèles retenus :

$$\left\{ \begin{array}{l} (1-0,91 B) \tilde{z}_t = (1-0,60 B) a_t, \quad \hat{\sigma}_a^2 = 0,097, \\ \quad \quad \quad [\pm 0,04] \quad \quad \quad [\pm 0,08] \\ \quad \quad \quad \nabla z_t = (1-0,71 B) a_t, \quad \hat{\sigma}_a^2 = 0,101. \\ \quad \quad \quad \quad \quad \quad \quad \quad \quad [\pm 0,05] \end{array} \right.$$

Modèles de Box et Jenkins :

$$\left\{ \begin{array}{l} (1-0,92 B) z_t = 1,45 + (1-0,58 B) a_t, \quad \hat{\sigma}_a^2 = 0,097, \\ \quad \quad \quad [\pm 0,04] \quad \quad \quad [\pm 0,08] \\ \quad \quad \quad \nabla z_t = (1-0,70 B) a_t, \quad \hat{\sigma}_a^2 = 0,101. \\ \quad \quad \quad \quad \quad \quad \quad \quad \quad [\pm 0,05] \end{array} \right.$$

Série B : Cours journaliers (fermeture) des actions IBM : 17 mai 1961-2 novembre 1962 (369 observations).

Référence : Box et Jenkins, 1970.

Deux modèles retenus (§ 4.2 et 4.3.2) :

$$\left\{ \begin{array}{l} (1-0,09 B) \nabla z_t = a_t, \quad \hat{\sigma}_a^2 = 52,3, \\ \quad \quad \quad [\pm 0,05] \\ \quad \quad \quad \nabla z_t = (1+0,09 B) a_t, \quad \hat{\sigma}_a^2 = 52,2. \\ \quad \quad \quad \quad \quad \quad \quad \quad \quad [\pm 0,05] \end{array} \right.$$

Modèle de Box et Jenkins :

$$\nabla z_t = (1+0,09 B) z_t = a_t, \quad \hat{\sigma}_a^2 = 52,2. \\ \quad \quad \quad \quad \quad \quad \quad \quad \quad [\pm 0,05]$$

Série E : Taux d'épargne des ménages en France (%) : trimestriel, 1962 à 1974 (52 observations).

Référence : Indicateurs du 7^e plan, avril 1977 (I.N.S.E.E.).

Modèle retenu :

$$(1 - 0,72 B) \tilde{z}_t = a_t, \quad \hat{\sigma}_a^2 = 0,80. \\ [\pm 0,10]$$

Série F : Matériel ferroviaire : indice mensuel corrigé des variations saisonnières : janvier 1963 à septembre 1975 (153 observations).

Référence : Bulletin mensuel de statistique (I.N.S.E.E.); Les indices de la production industrielle base 100 en 1970, Les Collections de l'I.N.S.E.E., E 35, février 1976.

Modèle retenu :

$$\nabla z_t = (1 - 0,74 B) a_t, \quad \hat{\sigma}_a^2 = 153. \\ [\pm 0,05]$$

Série G : Trafic aérien international mensuel (milliers de passagers) : janvier 1949-décembre 1960 (144 observations).

Référence : Box et Jenkins, 1970.

Modèle retenu (§ 4.4.4.a) :

$$\nabla \nabla_{12} \log z_t = (1 - 0,396 B) (1 - 0,614 B^{12}) a_t, \quad \hat{\sigma}_a^2 = 0,001 39. \\ [\pm 0,08] \quad [\pm 0,07]$$

Modèle de Box et Jenkins :

$$\nabla \nabla_{12} \text{Log } z_t = (1 - 0,396 B) (1 - 0,614 B^{12}) a_t, \quad \hat{\sigma}_a^2 = 0,001 34 (*)$$

Modèle d'Anderson, 1976 :

$$\nabla \nabla_{12} \text{Log } z_t = (1 - 0,362 B) (1 - 0,624 B^{12}) a_t, \quad \hat{\sigma}_a^2 = 0,001 32.$$

Modèle de Lenz, 1978 :

$$\nabla \nabla_{12} \text{Log } z_t = (1 - 0,4 B - 0,05 B^2 - 0,2 B^3) (1 - 0,65 B^{12}) a_t, \\ \sigma_a^2 = 0,001 29.$$

Série H : Immatriculations automobiles mensuelles aux U.S.A. (véhicules neufs) : janvier 1947-décembre 1968 (264 observations).

(*) On remarquera les valeurs différentes de $\hat{\sigma}_a^2$ pour un même modèle : cela tient aux façons différentes de calculer la variance S/v (différents v).

Référence : Nelson, 1975.

Modèle retenu :

$$\begin{aligned} \nabla \nabla_{12} \text{Log } z_t = & (1 - 0,211 B - 0,261 B^2) (1 - 0,845 B^{12}) a_t, \\ & [\pm 0,061] \quad [\pm 0,061] \quad [\pm 0,034] \\ \hat{\sigma}_a^2 = & 0,012\,26. \end{aligned}$$

Modèle de Nelson :

$$\begin{aligned} \nabla \nabla_{12} \text{Log } z_t = & (1 - 0,211 B - 0,261 B^2) (1 - 0,847 B^{12}) a_t, \\ & [\pm 0,061] \quad [\pm 0,061] \quad [\pm 0,034] \\ \hat{\sigma}_a^2 = & 0,012\,28. \end{aligned}$$

Série I : Matériel agricole : indice mensuel corrigé des variations saisonnières ; janvier 1963 à mai 1976 (161 observations).

Référence : voir série F.

Deux modèles retenus :

$$\left\{ \begin{array}{l} (1 - 0,93 B) \tilde{z}_t = (1 - 0,45 B) a_t, \quad \hat{\sigma}_a^2 = 52, \\ \quad [\pm 0,035] \quad [\pm 0,09] \\ \nabla z_t = (1 - 0,49 B) a_t, \quad \hat{\sigma}_a^2 = 54. \\ \quad [\pm 0,07] \end{array} \right.$$

Série M : Consommations électriques françaises du jour ouvrable moyen mensuel corrigées de l'effet de température : janvier 1951-juin 1958 (90 observations).

Référence : Mogha, 1977.

Modèle retenu :

$$\begin{aligned} \nabla z_t = & 0,85 + (1 - 0,26 B) a_t, \quad \hat{\sigma}_a^2 = 2,2. \\ & [\pm 0,1] \end{aligned}$$

Série N : Produit national brut aux U.S.A. aux prix courants du marché : 1^{er} trimestre 1947-4^e trimestre 1966; milliards de dollars, taux annuels [80 observations (corrigé des variations saisonnières)].

Référence : Nelson, 1975.

Modèle retenu (§ 4.3.4) :

$$\begin{aligned} (1 - 0,62 B) \nabla z_t = & 2,78 + a_t, \quad \hat{\sigma}_a^2 = 223. \\ & [\pm 0,09] \end{aligned}$$

Modèle de Nelson :

$$(1 - 0,62 B) \nabla z_t = 2,69 + a_t, \quad \hat{\sigma}_a^2 = 224.$$

$$[\pm 0,09] \quad [\pm 0,08]$$

Série P : Ventes mensuelles d'une entreprise X : janvier 1965-mai 1971 (77 observations).

Référence : Chatfield et Prothero, 1973.

Se reporter pour le choix du modèle : § 4. 5; pour la validation : § 5. 5. b; pour les prévisions : § 6. 3.

Série T : Trafic marchandises S.N.C.F. en tonnes-kilomètres : janvier 1971-décembre 1976 (72 observations).

Référence : Tendances de la conjoncture (I.N.S.E.E.).

Modèle retenu :

$$(1 - 0,61 B^{12}) \nabla z_t = (1 - 0,54 B) a_t, \quad \hat{\sigma}_a^2 = 198 \cdot 10^3.$$

$$[\pm 0,09] \quad [\pm 0,10]$$

Série X : Taux de couverture des échanges extérieurs en France (%) : trimestriel, 1962 à 1974 (52 observations).

Référence : Indicateurs du 7^e plan, avril 1977 (I.N.S.E.E.).

Modèle retenu :

$$(1 - 0,66 B) \tilde{z}_t = a_t, \quad \hat{\sigma}_a^2 = 13 \cdot 6.$$

$$[\pm 0,10]$$

Série Z : Ventes mensuelles d'une entreprise française : janvier 1973-avril 1978 (64 observations).

Pas de référence : l'entreprise a tenu à rester anonyme.

1786.3	2920.5	2473.5	2196.4	2418.1	2471.3	1899.7	1348.4	2372.6	2307.3
2028.4	2872.8	2853.0	3440.6	3525.9	3655.2	4476.3	2593.0	3395.4	2466.1
3723.1	3844.0	2348.2	3551.1	2125.6	2967.9	3265.9	2486.4	2759.0	2674.2
2538.6	1701.7	2770.2	2684.4	3728.7	2986.4	3211.8	4419.8	3815.2	4218.6
5015.0	4612.9	3565.6	2636.1	3957.6	3771.6	5654.9	4894.1	5788.4	6803.2
5437.7	4567.0	6251.2	4611.8	4128.5	4003.5	4803.1	5209.1	6048.5	4838.2
5712.9	8148.1	5807.2	5143.0	→					

Deux modèles retenus (§ 4. 4. 4. b et 5. 5. a) :

$$\left\{ \begin{array}{l} (1 - 0,60 B^{12} - 0,29 B^{24}) \nabla z_t = (1 - 0,62 B) a_t, \quad \hat{\sigma}_a^2 = 556 \cdot 10^3, \\ \quad [\pm 0,1] \quad [\pm 0,1] \quad [\pm 0,1] \\ (1 + 0,37 B^{12}) \nabla \nabla_{12} z_t = 1 - 0,63 B) a_t, \quad \hat{\sigma}_a^2 = 560 \cdot 10^3. \\ \quad [\pm 0,2] \quad [\pm 0,2] \end{array} \right.$$

BIBLIOGRAPHIE

- J. ABADIE, *Advances in Nonlinear Programming*, p. 900-930, in K. B. HALEY, ed., *Operational Research'78, Proceedings of the Eighth I.F.O.R.S. Conference on Operational Research*, Toronto, June 19-23, 1978, North-Holland Publishing Company, Amsterdam-New York-Oxford, 1979.

Les autres références se trouvent dans :

- J. ABADIE et D. TRAVERS, 1980, *Une approche simplifiée de la méthode de Box et Jenkins pour l'analyse et la prévision des séries temporelles unidimensionnelles (I)*, R.A.I.R.O., Recherche Opérationnelle. vol. 14, n° 4, p. 355-380.