

G. LIBERT

Construction automatique et continue de prévisions

RAIRO. Recherche opérationnelle, tome 16, n° 4 (1982),
p. 333-348

[<http://www.numdam.org/item?id=RO_1982__16_4_333_0>](http://www.numdam.org/item?id=RO_1982__16_4_333_0)

© AFCET, 1982, tous droits réservés.

L'accès aux archives de la revue « RAIRO. Recherche opérationnelle » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). 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/>

CONSTRUCTION AUTOMATIQUE ET CONTINUE DE PRÉVISIONS (*)

par G. LIBERT ⁽¹⁾

Résumé. — *La procédure de Box et Jenkins est une des meilleures méthodes de construction de prévisions pour le court terme. Elle présente toutefois un inconvénient important : elle n'est pas automatique.*

Cet article décrit le programme CAPRI qui construit automatiquement et rapidement un modèle, génère des prévisions précises et met à jour le modèle si de nouvelles données sont disponibles.

Mots clés : Automatisation; Box et Jenkins; Prévisions.

Abstract. — *The Box-Jenkins procedure is one of the best methods for building short-term forecasts. However, it has an important disadvantage: it isn't automatic.*

This paper describes the program CAPRI which constructs automatically and quickly one model, generates accurate forecasts and updates the model if new data are available.

Keywords: Automation; Box-Jenkins; Forecasts.

1. INTRODUCTION

De nombreuses méthodes de prévisions, des plus simples aux plus élaborées, existent. Parmi ces dernières, on peut classer la procédure développée par Box et Jenkins [7] qui, englobant un large éventail de modèles possibles, a très souvent la possibilité de sélectionner un modèle bien adapté à la série étudiée. Toutefois, il importe d'insister sur le fait que son domaine d'applicabilité doit être restreint aux prévisions à court terme et que dans certains cas, d'autres méthodes sont au moins tout aussi performantes [11, 17, 18, 19]. Notamment, il est tout à fait inutile d'employer une technique aussi complexe lorsque la série ne compte pas plus d'une soixantaine de valeurs. De plus, cette méthodologie ne trouve toute sa puissance que si les hypothèses utilisées pour la construction du modèle sont vérifiées.

Parmi celles-ci, il en est une qui a une importance essentielle. Par la procédure de Box et Jenkins, un modèle est choisi presque exclusivement en fonction des coefficients d'autocorrélation de la série. Dès lors, il est

(*) Reçu mai 1981.

(¹) Faculté polytechnique, rue de Houdain, 9, B-7000 Mons, Belgique.

indispensable pour la performance de la méthode que ces derniers caractérisent complètement la série et que, par conséquent, ils soient constants dans le temps. Or, en pratique, on ne teste jamais cette hypothèse. Cette lacune conduit alors à conclure que la méthode n'est pas adaptée alors qu'en fait, elle est mal utilisée [2].

Ce papier montre comment vérifier toutes les hypothèses de travail avec lesquelles il est permis d'affirmer que cette technique est une des plus puissantes méthodes d'élaboration de prévisions à court terme [13, 22].

Toutefois elle présente un inconvénient important : elle n'est pas automatique car l'ajustement du modèle nécessite l'intervention du prévisionniste. Celui-ci doit donc disposer d'un bagage statistique suffisant et consacrer du temps à cette analyse. Si le nombre de séries à traiter est élevé, le coût en temps devient rapidement prohibitif. Cette situation se produit notamment dans les problèmes de gestion de stocks où il convient de prévoir la demande de plusieurs milliers d'articles. Pour résoudre ce type d'application, il importe d'automatiser la procédure de Box et Jenkins en essayant de respecter au mieux ses conditions d'applicabilité afin de lui conserver toute sa puissance de prévision.

L'originalité de la méthode proposée dans ce papier (CAPRI : Construction Automatique de Prévisions Régies sans Intervention [17]) réside dans cette façon d'aborder la construction de modèles qui entraîne parfois l'élimination de données anciennes qui ne sont plus en accord avec l'évolution la plus récente du phénomène analysé.

De plus, l'algorithme qui en découle présente l'avantage par rapport à d'autres méthodes automatiques [14] de consommer peu de temps de calcul.

Enfin, il permet un contrôle constant du modèle basé sur la confrontation des prévisions avec la réalité et une mise à jour de ce dernier si cela s'avère utile.

Le souci de respecter les hypothèses de la méthode donne à celle-ci la possibilité de traiter tout type de séries. En effet, comme les prévisions constituent l'objectif essentiel du programme CAPRI, on peut abandonner des données anciennes qui perturberaient l'analyse et s'intéresser exclusivement aux données les plus récentes qui constituent une série stationnaire. On s'impose de travailler sur un minimum d'une soixantaine de valeurs pour obtenir une estimation du modèle statistiquement valable et dans le cas où cet ensemble de valeurs n'est pas stationnaire, le programme signale cette insuffisance.

2. LA PROCÉDURE DE BOX ET JENKINS

Box et Jenkins considèrent pour le processus Z_t le modèle général :

$$(1 - \varphi_1 B - \dots - \varphi_p B^p) (1 - \Phi_1 B^s - \dots - \Phi_P B^{sP}) W_t \\ = \theta_0 + (1 - \theta_1 B - \dots - \theta_q B^q) (1 - \Theta_1 B^s - \dots - \Theta_Q B^{sQ}) a_t, \\ W_t = \nabla^d \nabla_s^D Z_t^\lambda,$$

$$Z_t^\lambda = \begin{cases} Z_t^\lambda & \text{si } \lambda \neq 0, \\ \log_e Z_t & \text{si } \lambda = 0, \end{cases}$$

$$\nabla^d = \nabla (\nabla^{d-1}), \quad BZ_t = Z_{t-1}, \quad \nabla Z_t = (1 - B) Z_t = Z_t - Z_{t-1}, \\ \nabla_s^D = \nabla_s (\nabla_s^{D-1}), \quad B^s Z_t = Z_{t-s}, \quad \nabla_s Z_t = (1 - B^s) Z_t = Z_t - Z_{t-s},$$

$a_t : N(0, \sigma_a)$ et $\text{cov}[a_t, a_{t+k}] = 0$ pour tout $k \neq 0$.

$\varphi_1, \dots, \varphi_p, \Phi_1, \dots, \Phi_P, \theta_0, \theta_1, \dots, \theta_q, \Theta_1, \dots, \Theta_Q$: paramètres du modèle.

s , période du phénomène saisonnier.

Souvent en pratique, on a $d \leq 2, D \leq 1, p + q \leq 3, P + Q \leq 2$.

Les transformations $\nabla^d, \nabla_s^D, Z_t^\lambda$ sont réalisées pour que le processus W_t soit gaussien et stationnaire. De cette façon, il est entièrement caractérisé par sa moyenne μ , sa variance σ_Z^2 et sa fonction d'autocorrélation ρ_k ($k = 1, 2, \dots$) :

$$\mu = E[W_t], \\ \sigma_Z^2 = E[(W_t - \mu)^2], \\ \rho_k = \frac{E[(W_t - \mu)(W_{t+k} - \mu)]}{\sigma_Z^2}, \quad k = 1, 2, \dots$$

Un modèle adapté à la série z_t ($t = 1, 2, \dots, N$) peut alors être construit. On vérifie tout d'abord les hypothèses en recherchant d, D, λ et s tels que la série :

$$w_t = \nabla^d \nabla_s^D z_t^\lambda,$$

soit stationnaire.

Ensuite on calcule :

$$m = \sum_{t=1}^n w_t/n, \quad n = N - d - Ds, \\ s_z^2 = \sum_{t=1}^n (w_t - m)^2/n, \\ r_k = \frac{\sum_{t=1}^{n-k} (w_t - m)(w_{t+k} - m)/n}{s_z^2}, \quad k = 1, 2, \dots$$

et on compare m, s_z^2 et r_k ($k = 1, 2, \dots$) avec μ, σ_Z^2 et ρ_k ($k = 1, 2, \dots$) de divers processus.

μ dépend surtout de θ_0 et σ_z^2 de σ_a^2 . Dès lors le choix d'un modèle se base essentiellement sur la comparaison des r_k et des ρ_k .

L'identification du modèle terminée, il reste à estimer les paramètres, à valider le modèle et à construire les prévisions. Si ces dernières sont confrontées continuellement à la réalité, on utilise une procédure de mise à jour du modèle qui tient compte des réalisations supplémentaires enregistrées.

3. VÉRIFICATION DES HYPOTHÈSES

La première étape de la procédure automatique consiste à rechercher λ , d , D et s afin de pouvoir travailler sur une série gaussienne et stationnaire. Cette étape est la plus importante car elle détermine les caractéristiques fondamentales de la série. On détermine alors sous quelles conditions on obtient une série dont les coefficients d'autocorrélation sont constants dans le temps. Dès ce moment, le choix des paramètres du modèle devient bien souvent un problème de moindre importance [10] et simple à résoudre.

3.1. Paramètre de transformation λ

λ qui transforme z_t en z_t^λ distribué suivant une loi normale est souvent recherché par la méthode du maximum de vraisemblance préconisée par Box et Cox [6]. Toutefois, cette transformation coûteuse en temps de calcul ne fournit pas toujours des résidus gaussiens. On appelle résidus les estimations des a_t résultant de l'identification et de l'estimation du modèle.

On envisage dès lors une transformation moins satisfaisante d'un point de vue théorique mais dont les résultats permettent d'atteindre l'objectif poursuivi. Bartlett a montré que la transformation qui stabilise la variance d'une série en fonction de la moyenne fournit souvent la normalité [16].

Cependant la méthode graphique proposée ne convient pas pour une procédure automatique. Dès lors on travaille de la façon suivante :

- on découpe la série z_t^λ en k sous-séries de même longueur k' ;
- on calcule moyenne et variance de chaque sous-série;
- on calcule le coefficient de corrélation, $rms(\lambda)$, entre les moyennes et les variances;
- la transformation cherchée est celle qui fournit une valeur minimale de $|rms(\lambda)|$.

On choisit k = partie entière de \sqrt{N} et k' = partie entière de N/k . De cette façon, on réalise un compromis entre bonne estimation des moyennes et variances et bonne estimation de $rms(\lambda)$.

La transformation obtenue par minimisation du *rms* jouit de qualités intéressantes. Elle est rapide et automatique. Elle fournit des résidus distribués suivant une loi normale et tend à minimiser le rapport $\hat{\sigma}_a^2/\sigma_w^2$ qui mesure, après estimation, la part de variance inexploitée par le modèle.

Ces conclusions sont illustrées au tableau I pour la série « International Airline Passengers » [9]. On a repris pour un même modèle mais pour un paramètre de transformation différent, les principaux résultats d'estimation : valeurs des paramètres et leur erreur-standard, coefficient de corrélation entre paramètres, $\hat{\sigma}_a^2/\sigma_w^2$, vraisemblance (1), coefficients d'autocorrélations totale et partielle significatifs pour la série résiduelle, valeur du test de Box et Pierce [8] calculée sur 6, 12, 18, 24 coefficients d'autocorrélation totale, valeur du χ^2 de Pearson pour le test de normalité des résidus.

On constate aisément que la transformation retenue par le *rms*, $\lambda = -0.25$, répond exactement au problème posé et que la transformation résultant de la méthode du maximum de vraisemblance, $\lambda = 0.0$, ne donne pas des résidus gaussiens.

3.2. Degrés de différenciation

Pour obtenir la stationnarité de la moyenne, on différencie d fois la série z_t^λ . La recherche de d peut être entreprise de diverses façons.

Box et Jenkins signalent qu'une série est stationnaire si sa fonction d'autocorrélation totale décroît rapidement vers 0. Dans une procédure automatique, tester une décroissance rapide vers 0 n'est pas chose aisée. De plus, Anderson [3] a montré que certains processus non stationnaires ont une fonction d'autocorrélation totale qui tend vite vers zéro.

Une autre procédure est celle préconisée par Anderson [4]. Dans l'analyse des séries chronologiques, un des buts poursuivis est l'explication d'un maximum de la variance de la série étudiée. Ceci peut déjà être entamé lors des différenciations en recherchant la série $\nabla^d z_t^\lambda$ dont la variance est minimale. Toutefois il convient de remarquer que si la série z_t est stationnaire et que son premier coefficient de corrélation totale est supérieur à 0.5, ce critère donnera une valeur de d trop grande. En effet :

$$\begin{aligned} w_t &= z_t - z_{t-1}, \\ \sigma_w^2 &= 2\sigma_z^2 - 2r_1\sigma_z^2, \\ \frac{\sigma_w^2}{\sigma_z^2} &= 2(1-r_1). \end{aligned}$$

Bien que ce fait soit plutôt favorable pour des prévisions adaptatives [7], des différenciations inutiles compliquent l'identification du modèle. Dès lors, on

TABLEAU I
International Airline Passengers

λ	-1	-0,8	-0,6	-0,4	-0,2	0,0
$\theta(E.S.)$	0,40(0,08)	0,41(0,08)	0,43(0,08)	0,44(0,08)	0,43(0,08)	0,39(0,08)
$\Theta(E.S.)$	0,40(0,08)	0,47(0,07)	0,55(0,07)	0,62(0,07)	0,65(0,06)	0,61(0,07)
Cor. (θ, Θ)	0,07	0,06	0,05	0,02	0,00	0,08
s_a^2/s_w^2 (%)	67,75	63,97	59,94	56,40	54,98	57,27
1	-382,76	-365,70	-349,57	-335,29	-325,28	-321,91
Coefficients d'autocorrélation (r)	3 : -0,22 6 : 0,17 23 : 0,20	3 : -0,22 6 : 0,18 23 : 0,20	3 : -0,21 6 : 0,18 23 : 0,19	3 : -0,19 6 : 0,17 23 : 0,18	23 : 0,19	23 : 0,22
Test de Box et Pierce	$\begin{cases} 4 & 10,36(^a) \\ 10 & 16,06 \\ 16 & 22,02 \\ 22 & 31,28 \end{cases}$	$\begin{cases} 4 & 15,46 \\ 10 & 20,94 \\ 16 & 29,86 \\ 22 & 31,28 \end{cases}$	$\begin{cases} 4 & 10,21(^a) \\ 10 & 15,30 \\ 16 & 20,22 \\ 22 & 28,56 \end{cases}$	$\begin{cases} 4 & 8,72 \\ 10 & 13,83 \\ 16 & 17,97 \\ 22 & 25,39 \end{cases}$	$\begin{cases} 4 & 6,63 \\ 10 & 10,63 \\ 16 & 14,26 \\ 22 & 21,45 \end{cases}$	$\begin{cases} 4 & 5,07 \\ 10 & 8,76 \\ 16 & 12,68 \\ 22 & 22 \end{cases}$
Coefficients d'autocorrélation partielle (ϕ)	3 : -0,21 7,91(^b)	3 : -0,21 7,39(^b)	3 : -0,20 4,68(^a)	3 : -0,18 2,67	0,08	20 : -0,16 4,18(^a)
Normalité						

λ	0,2	0,4	0,6	0,8	1	-0,25
$\theta(E.S.)$	0,37(0,08)	0,36(0,08)	0,34(0,08)	0,32(0,08)	0,32(0,08)	0,44(0,08)
$\Theta(E.S.)$	0,52(0,08)	0,41(0,09)	0,30(0,09)	0,21(0,10)	0,08(0,10)	0,65(0,06)
Cor. (θ, Θ)	0,08	0,07	0,09	0,04	0,04	0,03
s_a^2/s_w^2 (%)	62,24	68,08	73,54	77,94	81,32	55,01
1	-323,85	-328,79	-335,57	-343,80	-353,51	-327,07
Coefficients d'autocorrélation (r)	23 : 0,25 4 : 4,67 10 : 9,09 16 : 13,25 22 : 25,80	23 : 0,27 4 : 4,66 10 : 9,63 16 : 13,64 22 : 28,85	23 : 0,28 4 : 4,71 10 : 9,92 16 : 13,65 22 : 30,54	23 : 0,29 4 : 4,85 10 : 10,36 16 : 13,92 22 : 32,17	23 : 0,28 4 : 4,70 10 : 9,99 16 : 13,34 22 : 31,68	23 : 0,19 4 : 7,16 10 : 11,45 16 : 15,08 22 : 22,14
Test de Box et Pierce	$\begin{cases} 4 & 9,16 \\ 10 & 20 : -0,18 \\ 16 & 23 : 0,16 \end{cases}$	$\begin{cases} 4 & 9,16 \\ 10 & 19 : -0,16 \\ 16 & 20 : -0,20 \\ 22 & 23 : 0,17 \end{cases}$	$\begin{cases} 4 & 9,16 \\ 10 & 19 : -0,17 \\ 16 & 20 : -0,21 \\ 22 & 23 : 0,19 \end{cases}$	$\begin{cases} 4 & 4 : -0,16 \\ 10 & 19 : -0,18 \\ 16 & 20 : -0,22 \\ 22 & 23 : 0,19 \end{cases}$	$\begin{cases} 4 & 19 : -0,19 \\ 10 & 20 : -0,22 \\ 16 & 22 : -0,17 \\ 22 & 23 : 0,19 \end{cases}$	0,04
Coefficients d'autocorrélation partielle (ϕ)						
Normalité	1,50	6,07(^a)	11,30(^a)	8,40(^b)	4,10(^a)	

peut compléter la recherche de d en réalisant un test de tendance sur $\nabla^{d-1} z_t^\lambda$. Si une tendance existe, on retient d . Si ce n'est pas le cas, on considère $\nabla^{d-2} z_t^\lambda$. Si on détecte une tendance, on retient $d-1$. Sinon, etc.

Le test de tendance retenu est celui de Mann basé sur le coefficient de Kendall [16]. Pour cela, on utilise les données $z_1, z_{1+c}, z_{1+2c}, \dots$ afin d'éviter que le test de tendance ne réagisse à cause des coefficients d'autocorrélation de rang faible.

D est déterminé de la même manière mais pour cela il faut connaître s .

3.3. Périodicité

Dans la procédure de Box et Jenkins, la périodicité est déterminée à partir des coefficients d'autocorrélation ou est supposée connue. Comme cette hypothèse ne peut être formulée dans une procédure automatique, contrairement à ce qui est fait dans la méthode automatique de Hill et Woodworth [14], il est nécessaire de mettre au point une technique qui permette de résoudre ce problème.

Les techniques classiques de recherche de périodicité ne conviennent pas dans cette étude car elles ne sont adaptées qu'aux processus purement périodiques [15, 23]. Des résultats plus récents ne sont guère plus utiles car coûteux en temps de calcul et valables uniquement pour certains modèles [12, 24].

La procédure employée dans le programme CAPRI suit de très près la démarche du prévisionniste interprétant les fonctions d'autocorrélations totale et partielle. On utilise la série $\nabla^d z_t^\lambda$ qui est partiellement stationnaire et dont la variance est minimale. On détecte les coefficients d'autocorrélation les plus importants. On abandonne ceux qui sont expliqués par la partie non saisonnière du modèle. $s=k$ si r_k et r_{2k} sont significatifs. Si un tel k n'existe pas, $s=k$ avec r_k le plus grand coefficient, s'il en reste.

Cette technique est particulièrement satisfaisante étant donné que les séries étudiées, économiques pour la plupart, ne sont pas purement périodiques mais saisonnières et que si une saisonnalité existe, elle est fortement marquée.

3.4. Stationnarité d'ordre 2

Après avoir transformé la série étudiée avec λ , d , D et s , on dispose d'une série dont la moyenne et la variance sont constantes dans le temps. Contrairement à la pratique courante, on ne peut pas encore identifier un

modèle car on n'a pas vérifié la constance dans le temps des coefficients d'autocorrélation. Or ceci est fondamental puisque l'identification repose exclusivement sur la fonction d'autocorrélation. Si celle-ci évolue au cours du temps, le modèle ajusté à la série entière constituera un « modèle moyen » résultant de plusieurs phénomènes différents [2, 7]. Dès lors, il ne sera certainement pas le mieux adapté à l'évolution la plus récente de la série dont dépendent surtout les prévisions.

On recherche donc un sous-ensemble de la série transformée qui vérifie la stationnarité d'ordre 2. Ce sous-ensemble comprend les dernières données, les plus intéressantes pour un modèle actuel, et est de longueur maximale pour une estimation valable des paramètres.

On procède de la façon suivante. On calcule la fonction d'autocorrélation de la sous-série la plus récente (FA 60). Elle comprend une soixantaine de valeurs (M') afin de réaliser une bonne estimation. On allonge cette série de f valeurs et on calcule la fonction d'autocorrélation des $M' + f$ dernières données. On la compare à FA 60. Si elle est dissemblable, on retient uniquement les $M = M'$ dernières valeurs. Si elle est semblable, on calcule la fonction d'autocorrélation des $M' + 2f$ dernières données. On compare celle-ci à FA 60. Si elles sont différentes, on retient les $M = M' + f$ dernières valeurs. Sinon, ... f réalise un compromis entre la détection « exacte » de M (calcul avec $f = 1$) et un temps de calcul raisonnable (maximum pour $f = 1$).

Pour comparer les fonctions d'autocorrélation, on calcule :

$$\varepsilon = \frac{\sum_{i=1}^K (R_i - r_i)^2}{\sum_{i=1}^K R_i^2},$$

$R_i (i = 1, \dots, K)$ sont les coefficients de FA 60 et $r_i (i = 1, \dots, K)$ ceux de l'autre fonction. On choisit K pour tenir compte des coefficients les plus intéressants. Diverses applications [17] montrent que l'on peut admettre que les fonctions sont différentes si ε est supérieur à 0,2.

Le tableau II donne les valeurs de ε trouvées pour la série « International Airline Passengers » avec $K = 16$ et $f = 3$. On retient $M = 93$. Le tableau III montre les différences entre les fonctions d'autocorrélations totale et partielle

TABLEAU II

Valeurs des ϵ pour la série International Airline Passengers

N	n	$\epsilon(\%)$	N	n	$\epsilon(\%)$
144	131	58,47	102	89	21,28
141	128	53,99	99	86	21,04
138	125	53,23	96	83	23,68
135	122	48,17	93	80	12,15
132	119	51,04	90	77	11,79
129	116	33,22	87	74	19,46
126	113	25,67	84	71	17,12
123	110	22,36	81	68	8,08
120	107	40,27	78	65	4,74
117	104	23,06	75	62	4,33
114	101	26,75	72	59	0,94
111	98	37,23	69	56	0,56
108	95	37,11	66	53	0
105	92	31,58			

TABLEAU III

Fonctions d'autocorrélations pour $N=144$ et $N=93$

Fonctions d'autocorrélation totale								
N	1	2	3	4	5	6	7	8
144	-0,34	0,11	-0,20	0,02	0,06	0,03	-0,06	-0,00
93	-0,33	0,14	-0,06	-0,18	0,14	-0,02	0,09	0,03
N	9	10	11	12	13	14	15	16
144	0,18	-0,08	0,06	-0,39	0,15	-0,06	0,15	-0,14
93	0,13	-0,15	0,05	-0,14	0,02	-0,02	0,04	-0,00
Fonctions d'autocorrélation totale								
N	1	2	3	4	5	6	7	8
144	-0,34	-0,01	-0,19	-0,12	0,03	0,03	-0,06	-0,02
93	-0,33	0,04	-0,00	-0,23	0,03	0,08	0,09	0,05
N	9	10	11	12	13	14	15	16
144	0,23	0,04	0,05	-0,34	-0,11	-0,08	-0,02	-0,14
93	0,21	-0,05	-0,02	-0,11	-0,04	-0,12	-0,02	-0,05

pour $N=144$ et $M=93$. Les résultats d'estimation repris au tableau IV illustrent clairement que le modèle valable pour $N=144$ ne l'est pas du tout pour $M=93$. On remarque surtout que le paramètre saisonnier Θ a disparu.

TABLEAU IV
Modèles retenus pour $N = 144$ et $N = 93$

$N = 144$	$N = 93$	$N = 93$	$N = 93$
$\theta = 0,39(0,08)$ $\Theta = 0,61(0,07)$ $\text{cor}(\theta, \Theta) = -0,08$	$\theta = 0,36(0,10)$ $\Theta = 0,76(0,05)$ $\text{cor}(\theta, \Theta) = -0,21$	$\varphi = -0,34(0,10)$ — —	$\varphi = -0,33(0,11)$ $\theta_4 = 0,21(0,10)$ $\text{cor}(\varphi, \theta_4) = -0,17$
$r_{16} = -0,16$ — $r_{23} = 0,22$ — — $\varphi_{19,19} = -0,16$ — —	$r_4 = -0,25$ $r_{10} = -0,20$ $r_{12} = 0,20$ $r_{19} = -0,20$ $r_{23} = 0,26$ $\varphi_{4,4} = -0,24$ $\varphi_{9,9} = 0,24$ $\varphi_{12,12} = 0,20$	$r_4 = -0,19$ — — — — $\varphi_{4,4} = -0,19$ — —	— — — — — — —
Test de Box et Pierce :			
χ^2_4 5,07	χ^2_4 5,83	χ^2_5 4,37	χ^2_4 1,88
10 8,76	10 14,64	11 9,32	10 6,27
16 12,68	16 20,38	17 11,22	16 8,04
22 22	22 32,21	23 16,12	22 11,27
$2/\sqrt{n} = 0,17$	$2/\sqrt{n} = 0,22$	$2/\sqrt{n} = 0,22$	$2/\sqrt{n} = 0,22$

4. CONSTRUCTION DU MODÈLE

Ayant trouvé une série stationnaire, la construction du modèle est entreprise. Elle comprend trois phases : l'identification du modèle, l'estimation des paramètres et la validation du modèle.

4.1. Identification du modèle

Les techniques récentes développées par Akaike [1], Parzen [21] et les procédures automatiques de construction de modèle qui en résultent apparaissent peu intéressantes car le choix du meilleur modèle résulte de l'essai de plusieurs modèles, en moyenne 5 pour Hill et Woodworth [14], ce qui est peu satisfaisant du point de vue temps de calcul.

Dans le but de choisir rapidement un bon modèle, il suffit d'observer soigneusement les fonctions d'autocorrélations révélatrices de toute l'information caractéristique de la série vu qu'elles sont invariantes au cours du temps. La vérification des hypothèses de stationnarité et d'inversibilité exprimées en fonction des coefficients d'autocorrélation [5, 17] permet déjà d'éliminer un certain nombre de modèles. Des tests de signification appliqués à ces mêmes coefficients conduisent alors à la sélection d'un modèle.

4. 2. Estimation des paramètres

On estime les paramètres grâce à l'algorithme de Marquardt [20] en prenant comme valeurs initiales des paramètres, les estimateurs préliminaires calculés à partir des coefficients d'autocorrélation. De cette façon, le temps de calcul est minimisé.

4. 3. Validation du modèle

On teste l'égalité à zéro des paramètres et si un paramètre est égal à zéro, on l'élimine et on réestime les autres.

On teste l'égalité à zéro des coefficients d'autocorrélation des résidus. Si certains coefficients sont différents de zéro, on introduit un ou des paramètres supplémentaires expliquant cette ou ces autocorrélations. On veille à ne pas créer une redondance entre les paramètres et on retourne à la phase d'estimation.

5. PRÉVISIONS

Le modèle étant entièrement déterminé, les prévisions et leurs intervalles de confiance sont obtenus directement à partir des formules classiques.

6. MISE A JOUR DU MODÈLE

Lorsque le modèle de prévision est confronté continuellement à la réalité, il importe d'utiliser les nouvelles données pour améliorer la qualité des prévisions. Des formules de mise à jour des prévisions résultent de la méthode de Box et Jenkins. Elles sont toutefois insuffisantes car elles supposent que le modèle et ses paramètres ne changent pas au cours du temps. Pour tenir compte de cette éventualité, une procédure originale de mise à jour du modèle complète le programme de génération de prévisions.

Lorsqu'une nouvelle réalisation apparaît dans l'intervalle de confiance à 95 % de la prévision correspondante, on admet que le modèle est toujours valable et on réestime uniquement les paramètres. En utilisant leurs anciennes valeurs comme estimateurs préliminaires, cette opération est très rapide.

Si la nouvelle réalisation sort de l'intervalle de confiance, on décrète que le modèle n'est plus adapté et on refait l'identification après avoir vérifié les hypothèses.

Cette méthode permet de suivre de très près l'évolution d'une série contrairement aux techniques classiques de contrôle de processus qui ne réagissent qu'assez longtemps après les changements.

7. VALIDATION DE LA MÉTHODE

Le programme CAPRI a été utilisé pour l'étude de nombreuses séries et surtout des séries économiques. Même pour ces dernières qui sont parfois particulièrement perturbées, les prévisions obtenues sont toujours satisfaisantes et comparables à celles fournies par la procédure non automatique. Cette propriété est normale compte tenu du fait qu'on s'efforce de vérifier au mieux les conditions de stationnarité. Ainsi pour les séries économiques mensuelles, la procédure rejette presque toujours les valeurs antérieures au début de la crise pétrolière vu qu'elles ont un comportement différent de celui des données actuelles.

Au lieu de commenter les résultats issus de diverses applications de la méthode, il apparaît plus intéressant d'illustrer complètement son fonctionnement pour une série particulière.

On considère les 96 premières données de la série « International Airline Passengers » qui en comprend 144. On identifie le modèle et on obtient $\lambda = 0.3$, $s = 12$, $d = D = 1$, $\theta = 0.40$, $\Theta = 0.65$, $N = 96$.

Hormis le paramètre de transformation, ce modèle est tout à fait semblable à celui retenu par Box et Jenkins pour la série complète [7].

Les prévisions sont générées et le modèle est mis à jour en injectant progressivement les données 97, 98, . . . , 144. Jusqu'en $t = 134$, la réestimation suffit. En $t = 135$, une réidentification est réalisée. Il vient :

$$\lambda = -0.05, \quad s = 12, \quad d = D = 1, \quad \varphi = -0.31, \quad N = 81.$$

Il en est de même en $t = 136$ et l'on a :

$$\lambda = -0.1, \quad s = 12, \quad d = D = 1, \quad \varphi = -0.44, \quad N = 82.$$

Le modèle est ensuite maintenu et en $t = 144$, on a :

$$\lambda = -0.1, \quad s = 12, \quad d = D = 1, \quad \varphi = -0.40, \quad N = 90.$$

On constate que la mise à jour a bien enregistré la modification de stationnarité puisque l'étude des 144 données avait fourni précédemment $N = 93$.

TABLEAU V
Validation des prévisions

N	Réalizations	Intervalles de confiance à horizon		
		1	2	3
97	315	289-348	—	—
98	301	283-340	280-346	—
99	356	327-395	328-410	324-418
100	348	319-385	315-393	317-407
101	355	320-386	316-393	313-401
102	422	374-454	365-459	361-467
103	465	428-525	415-526	406-530
104	467	417-509	414-524	402-524
105	404	369-447	360-451	359-463
106	347	319-383	315-390	308-393
107	305	276-329	273-335	270-340
108	336	321-385	314-388	311-395
109	340	322-386	326-403	319-405
110	318	306-366	308-379	312-395
111	362	350-421	356-443	359-459
112	348	331-397	338-419	344-439
113	363	326-391	329-409	337-431
114	435	386-467	375-471	380-492
115	491	439-535	424-536	412-539
116	505	439-534	427-540	413-540
117	404	393-476	376-471	367-474
118	359	326-390	333-414	321-409
119	310	286-340	280-344	287-363
120	337	323-386	319-394	313-398
121	360	321-383	325-402	321-409
122	342	312-371	302-371	306-388
123	406	364-438	357-444	346-442
124	396	358-430	349-432	342-437
125	420	370-444	361-449	352-450
126	472	449-545	431-541	422-546
127	548	491-599	497-630	477-624
128	559	497-607	485-614	492-644
129	463	423-511	411-514	403-518
130	407	368-440	362-448	354-450
131	362	321-381	314-384	310-391
132	405	367-439	353-436	346-439
133	417	385-461	376-465	363-461
134	391	364-435	360-444	353-448
135	419	<u>426-512</u>	<u>423-527</u>	419-538
136	461	<u>397-448</u>	<u>407-505</u>	406-519
137	472	435-495	412-477	417-535
138	535	505-576	496-576	458-546
139	622	581-663	578-673	562-674
140	606	595-678	587-684	582-699
141	508	479-545	487-565	479-573
142	461	417-473	413-479	422-503
143	390	380-430	368-425	364-433
144	432	418-475	423-490	405-483
145	—	419-476	423-490	428-511
146	—	—	389-450	392-467

De plus, en analysant les prévisions, le dernier modèle apparaît meilleur que le précédent. Le tableau V signale les limites de l'intervalle de confiance à 95 % des prévisions à horizon 1, 2, 3 et la valeur de la réalisation correspondante. Toutes les observations hormis la 135^e et la 136^e sont situées dans cet intervalle. Les prévisions semblent donc aussi valables avant et après modification du modèle.

La largeur de l'intervalle de confiance permet ensuite d'affirmer qu'elles sont meilleures avec le dernier modèle. Le tableau VI présente les largeurs des intervalles de confiance à 95 % pour les horizons 1 et 2. On constate que pour un même moment dans les périodes successives, ces dernières croissent avant la révision du modèle et diminuent fortement ensuite.

TABLEAU VI
Largeur (L) de l'intervalle de confiance à 95 % des prévisions

A horizon 1							
N	L	N	L	N	L	N	L
100	66	112	66	124	72	136	51
101	66	113	65	125	74	137	60
102	80	114	81	126	96	138	71
103	97	115	96	127	98	139	82
104	92	116	95	128	110	140	83
105	78	117	83	129	88	141	66
106	62	118	76	130	72	142	56
107	53	119	54	131	60	143	50
108	64	120	63	132	72	144	57
109	64	121	62	133	76	145	57
110	70	122	59	134	71	—	—
111	71	123	74	135	86	—	—
A horizon 2							
N	L	N	L	N	L	N	L
101	77	113	80	125	88	137	65
102	94	114	96	126	110	138	80
103	111	115	112	127	133	139	95
104	110	116	113	128	129	140	97
105	91	117	95	129	103	141	78
106	75	118	81	130	86	142	66
107	62	119	64	131	70	143	57
108	74	120	75	132	83	144	73
109	77	121	77	133	89	145	73
110	71	122	71	134	84	146	61
111	87	123	87	135	104	—	—
112	81	124	83	136	98	—	—

Exemple : l'observation 100 doit se trouver dans un intervalle large de 66 unités. Il en est de même pour la réalisation 112. L'intervalle passe alors à 72 pour la 124 et se rétrécit à 51 pour la 136. La correction du modèle engendre donc des prévisions plus fiables.

Ainsi les prévisions fournies par CAPRI sont plus précises que celles obtenues par Box et Jenkins qui utilisent la série complète.

8. CONCLUSIONS

La procédure automatique développée est performante du point de vue statistique puisqu'elle fournit de bons modèles et que les prévisions qui en résultent sont au moins comparables avec celles issues de la procédure non automatique

Elle est aussi performante du point de vue temps de calcul. L'application présentée qui comprend la construction d'un premier modèle, 46 réestimations et 2 réidentifications consomme 58 secondes sur un ordinateur VAX-11. On peut donc considérer que la mise à jour du modèle en fonction d'une nouvelle observation requiert au maximum 1 seconde.

BIBLIOGRAPHIE

- J. ABADIE et D. TRAVERS, *Une approche simplifiée de la méthode de Box et Jenkins pour l'analyse et la prévision des séries temporelles unidimensionnelles*, R.A.I.R.O., Recherche Opérationnelle, vol. 14, 1980, p. 355-380 et vol. 15, 1981, p. 51-71.
1. H. AKAIKE, *A New Look at the Statistical Model Identification*, I.E.E.E., AC, vol. 19, 1974, p. 716-723.
 2. A. P. ANDERSEN et H. NELSON, *A Look at the Stationarity of 14 Quarterly U.S. Economic Time Series*, J. Op. Res. Soc., vol. 31, 1980, p. 353-358.
 3. O. D. ANDERSON, *On Realisations From Nonstationary Time Processes*, J. Op. Res. Soc., vol. 30, 1979, p. 253-258.
 4. O. D. ANDERSON, *Time Series Analysis and Forecasting*, Butterworths, London, 1976, 182 p.
 5. O. D. ANDERSON, *Feasible Regions for the First Pair of Autocorrelations of Moving Average Processes*, Math. Operationsforsch. u. Statist., vol. 7, 1976, p. 85-93.
 6. G. E. P. BOX et D. R. COX, *An Analysis of Transformations*, J. Roy. Stat. Soc., B, vol. 26, 1964, p. 211-243.
 7. G. E. P. BOX et G. M. JENKINS, *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, 1976, 575 p.
 8. G. E. P. BOX et D. A. PIERCE, *Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models*, J. Amer. Stat. Ass., vol. 65, 1970, p. 1509-1526.
 9. R. G. BROWN, *Smoothing, Forecasting and Prediction of Discrete Time Series*, Prentice Hall, New Jersey, 1962, 468 p.

10. C. CHATFIELD, *The Analysis of Time Series: Theory and Practice*, Chapman and Hall, London, 1975, 263 p.
11. R. FILDES, *Quantitative Forecasting, the State of the Art: Extrapolative Models*, J. Op. Res. Soc., vol. 30, 1979, p. 691-710.
12. E. J. GODOLPHIN, *A Procedure for Estimating Seasonal Moving Average Models Based on Efficient Estimation of the Correlogram*, J. Roy. Stat. Soc., B, vol. 39, 1977, p. 237-248.
13. C. W. J. GRANGER et P. NEWBOLD, *Forecasting Economic Time Series*, Academic Press, New York, 1977, 333 p.
14. G. W. HILL et D. WOODWORTH, *Automatic BOX-JENKINS Forecasting*, J. Op. Res. Soc., vol. 31, 1980, p. 413-422.
15. G. M. JENKINS et D. G. WATTS, *Spectral Analysis and its Applications*, Holden-Day, San Francisco, 1969, 525 p.
16. M. G. KENDALL et A. STUART, *The Advanced Theory of Statistics*, vol. II, Griffin, London, 1961, 676 p.
17. G. LIBERT, *Analyse de séries chronologiques. Élaboration automatique et continue de prévisions*, Thèse de Doctorat, Faculté Polytechnique de Mons, Belgique, 1981.
18. S. MAKRIDAKIS et M. HIBON, *Accuracy of Forecasting: An Empirical Investigation (with Discussion)*, J. Roy. Stat. Soc., A (General), vol. 142, 1979, p. 97-145.
19. S. MAKRIDAKIS et al., *The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition*, J. of Forecasting, vol. 1, n° 2, 1982, p. 111-153.
20. D. W. MARQUARDT, *An Algorithm for Least Squares Estimation of Non-Linear Parameters*, J. Soc. Ind. Applied Math., vol. 11, 1963, p. 431.
21. E. PARZEN, *Some Recent Advances in Time Series Modeling*, I.E.E.E., AC, vol. 19, 1974, p. 723-730.
22. D. J. REID, *A Comparative Study of Time Series Prediction Techniques on Economic Data*, Ph. D. Thesis, Dept. of Mathematics, Univ. of Nottingham, 1969.
23. R. SNEYERS, *Sur l'analyse statistique des séries d'observations*, Note technique n° 143, O.M.M., Genève, 1975, 192 p.
24. A. S. SAINT-LÉGER, *Comparison of two Tests of Seasonality in Epidemiological Data*, Applied Stat., vol. 25, 1976, p. 280-286.