

STATISTIQUE ET ANALYSE DES DONNÉES

DINH TUAN PHAM

Choix de modèle en analyse des séries chronologiques

Statistique et analyse des données, tome 15, n° 1 (1990), p. 61-90.

http://www.numdam.org/item?id=SAD_1990__15_1_61_0

© Association pour la statistique et ses utilisations, 1990, tous droits réservés.

L'accès aux archives de la revue « Statistique et analyse des données » 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/>

CHOIX DE MODELE EN ANALYSE DES SERIES CHRONOLOGIQUES

PHAM Dinh Tuan

Laboratoire de Modélisation et Calcul

IMAG

B.P. 53x

38041 Grenoble cedex

FRANCE

Résumé

Cet article fait le point des différentes méthodes du choix de l'ordre du modèle ARMA en séries chronologiques. Deux types d'approche sont décrits en détail : approche interactive et approche automatique.

Mots clefs : choix de l'ordre, identification de modèle, fonction de vraisemblance, modèle autorégressif moyenne mobile, test d'adéquation.

Classification AMS : 62 M 10

Abstract

Our subject concerns various methods for selecting the order of ARMA models in time series analysis. Two approaches are described : interactive and automatic approaches.

Keywords : order selection, identification, likelihood fonction, autoregressive-moving average model, diagnostic checking.

1. INTRODUCTION

L'inférence statistique paramétrique est basée sur la donnée d'un modèle comportant un petit nombre de paramètres. Sa validité dépend donc fortement du bon choix d'un modèle. Parfois le modèle peut être choisi a priori, indépendamment de l'observation, à l'aide des connaissances théoriques du phénomène aléatoire étudié. Mais dans beaucoup de cas, celles-ci sont insuffisantes pour formuler un modèle fiable, et on doit recourir aux données observées pour la construction de modèle. La modélisation statistique fait alors partie intégrante de l'inférence paramétrique. Curieusement, ce sont surtout les analystes de séries chronologiques qui s'intéressent à ce problème. Ceci peut s'expliquer en partie par le fait que dans ce domaine, on a souvent à traiter de masses importantes de données, parfois en temps réel, et un choix automatique de modèle sera très utile. Une autre raison est que l'on dispose déjà d'une classe de modèles très riche, à savoir les modèles autorégressifs moyenne mobile (ARMA - autoregressive moving average, en anglais), qui permettent de décrire la plupart des séries chronologiques rencontrées en pratique, au moins en première approximation, et le problème du choix d'un modèle se simplifie alors en le problème du choix de l'ordre du modèle ARMA. C'est ce dernier problème qui nous intéresse dans cet article. Nous exposerons différentes techniques proposées dans la littérature et au passage quelques-unes de nos contributions propres. Une synthèse de ces techniques peut être trouvée dans Shibata (1985) et De Gooijer et al. (1985). Le choix de l'ordre du modèle ARMA peut être aussi vu comme un problème d'approximation rationnelle de la structure au second ordre d'un processus stationnaire, considéré dans Hannan (1987).

2. GENERALITES SUR LES MODELES ARMA ET LEUR IDENTIFICATION

L'analyse statistique des séries chronologiques est essentiellement une analyse du second ordre, où on s'intéresse uniquement aux moyennes et covariances. On considère donc un processus stationnaire du second ordre X_t , $t = \dots, -1, 0, 1, \dots$, c'est-à-dire tel que $E(X_t)$ et $\text{cov}(X_t, X_{t+k})$ ne dépendent pas de t , pour tout k . Pour simplifier, nous nous restreindrons aux processus centrés, c'est-à-dire tels que $E(X_t) = 0$. Cela ne nuit pas à la généralité car en pratique, on peut se ramener à ce cas par simple soustraction de la série observée de sa moyenne arithmétique. Le processus X_t est appelé ARMA d'ordre (p, q) (au sens large) s'il admet une représentation de la forme

$$X_t = - \sum_{j=1}^p a_j X_{t-j} + e_t + \sum_{j=1}^q b_j e_{t-j} \quad (2.1)$$

où les a_j, b_j sont des constantes et e_t est une suite de variables aléatoires centrées non corrélées de même variance σ^2 . Cette classe de modèles contient deux sous-classes importantes : celle des modèles autorégressifs (AR) qui correspondent au cas $q = 0$ et celle des modèles moyenne mobile (MA - moving average en anglais -) qui correspondent au cas $p = 0$.

Il est facile de voir que le processus (1) a pour densité spectrale

$$f(\lambda) = \sigma^2 \left| \frac{\sum_{j=0}^p b_j e^{ij\lambda}}{\sum_{j=0}^q a_j e^{ij\lambda}} \right|^2 / (2\pi), \quad (a_0 = b_0 = 1),$$

pourvu que le dénominateur $\sum_{j=0}^q a_j e^{ij\lambda}$ du second membre précédent ne s'annule pas.

Comme la structure de covariance d'un processus dépend uniquement de sa densité spectrale, la représentation ARMA (1) n'est pas unique. Cela provient du fait que (i) les polynômes $\sum_{j=0}^p a_j z^j$ et $\sum_{j=0}^q b_j z^j$, appelés respectivement polynôme autorégressif et

moyenne mobile, peuvent avoir un facteur commun, et que (ii) la factorisation d'un polynôme trigonométrique positif en λ en le carré d'un polynôme en $e^{i\lambda}$ n'est pas unique. Pour enlever la première ambiguïté, il suffit d'imposer la condition que les polynômes autorégressif et moyenne mobile n'ont pas de racine commune. Quant à l'unicité de la factorisation, elle sera réalisée si on se restreint aux polynômes n'ayant pas de racines à l'intérieur du cercle unité du plan complexe (on les appelle polynômes stables). On supposera donc que les polynômes autorégressif et moyenne mobile sont stables. Cette dernière condition est équivalente à ce que e_t soit l'innovation du processus, c'est-à-dire $e_t = X_t - X_{t|t-1}$ où $X_{t|t-1}$ désigne le meilleur prédicteur linéaire de X_t basé sur les observations passées jusqu'à l'instant $t - 1$.

Jusqu'à présent, nous avons seulement supposé que les innovations e_t sont non corrélées. En pratique, pour obtenir des résultats intéressants comme la normalité asymptotique des estimateurs, on a besoin d'hypothèses plus fortes sur ceux-ci. L'hypothèse minimale semble être la suivante :

$$(i) \quad E(e_t | X_s, s < t) = 0$$

et

$$(ii) \quad E(e_t^2 | X_s, s < t) = \sigma^2 \text{ (non aléatoire).}$$

Toutefois, dans la littérature, on travaille souvent sous l'hypothèse que les e_t sont indépendantes et équidistribuées. Nous appellerons dans ce cas le processus X_t ARMA au sens strict.

Box et Jenkins (1970) ont été les premiers à populariser l'usage des modèles ARMA dans l'étude des séries chronologiques. Ils ont proposé une méthodologie simple pour sa modélisation. Celle-ci comporte trois étapes :

- (i) **Identification** : on essaie d'identifier le type de modèle, soit AR ($q = 0$) soit MA ($p > 0$ et $q > 0$), ainsi que son ordre. On construit aussi un estimateur préliminaire (non efficace) des paramètres.
- (ii) **Ajustement du modèle** : on estime les paramètres par les méthodes classiques (maximum de vraisemblance exacte ou approchée).
- (iii) **Validation du modèle** : on teste le modèle proposé pour voir si celui-ci est en accord avec les observations. Si le modèle est rejeté, on revient à (i) pour une nouvelle identification. Sinon, le modèle peut être retenu.

On voit que la partie la plus délicate est l'étape (i), l'étape (ii) et (iii) étant des problèmes classiques bien étudiés. Pour l'étape (i), Box et Jenkins ont proposé une méthode d'identification de modèle MA et AR, basée sur l'examen des fonctions d'autocorrélation empirique et d'autocorrélation partielle empirique, respectivement. En effet, on sait que les autocorrélations ρ_k , de retard k , d'un processus MA d'ordre q , sont nulles pour $k > q$. On peut donc reconnaître un tel processus en examinant les autocorrélations empiriques qui, pour $k > q$, sont asymptotiquement normales

centrées de variance $\sum_{j=-q}^q \rho_j^2/n$, n étant la taille de l'échantillon. Dans la pratique, on cherchera l'existence d'un rang faible q , à partir duquel la fonction d'autocorrélation empirique peut être considérée comme nulle, en tenant compte de son écart type. De même, les autocorrélations partielles d'un processus AR d'ordre p sont nulles pour les retards plus grands que p , et leurs homologues empiriques basés sur une série chronologique provenant d'un tel processus sont, pour les retards supérieurs à p , asymptotiquement indépendants, de distribution normale-centrée de variance $1/n$. Toutefois, Box et Jenkins n'ont pas proposé de méthode pour identifier le modèle ARMA général, et ont seulement suggéré d'adopter ce dernier quand la série observée ne semble pas suivre ni un modèle MA ni un modèle AR.

L'approche de Box et Jenkins a été généralisée par la suite au cas des modèles ARMA par divers auteurs. L'idée de base est de construire une famille de statistiques indexée par deux indices, qui seront faibles (plus exactement qui tendent vers zéro quand n tend vers l'infini), à partir de certains indices directement liés aux ordres du vrai modèle. L'examen d'un tel tableau de statistiques permettra au statisticien de choisir un couple d'ordres probable pour une modélisation préliminaire, qui sera validé par la suite par un test d'adéquation.

L'approche précédente sera appelée interactive par opposition à l'approche automatique décrite plus loin. Cette approche a l'avantage de laisser le statisticien maître de son choix. Elle lui permet de prendre en considération des informations a priori sur le phénomène aléatoire étudié. Dans le cas où le choix n'est pas clair, c'est-à-dire où il existe deux modèles (ou plus) qui sont apparemment aussi bons l'un que l'autre, le statisticien a l'opportunité d'examiner ceux-ci en détail avant de fixer son choix. Toutefois, l'approche interactive n'est possible que si on n'a qu'une ou quelques séries à analyser. Dans le cas où on doit traiter rapidement un grand nombre de séries (comme dans certains problèmes de traitement du signal), une approche qui fournit automatiquement l'ordre du modèle sera utile. L'intérêt d'une telle approche réside aussi du fait qu'elle est objective, car n'ayant fait intervenir que les observations, tandis que l'approche interactive contient toujours un élément subjectif.

Le modèle AR fut le premier à bénéficier des procédures de choix automatique de l'ordre. Ce modèle a attiré beaucoup d'attention vu la facilité de son ajustement, de son réalisme pour représenter des phénomènes cycliques (les racines du polynôme AR, quand elles sont proches du cercle unité, correspondent aux modes de vibration propre du système) et de son utilisation simple en prédiction. Il est bien connu que tout processus stationnaire de densité spectrale continue strictement positive peut être approché en moyenne quadratique par un processus AR d'ordre suffisamment grand. Cela donne lieu à la procédure d'estimation de densité spectrale par ajustement d'un modèle AR d'ordre élevé. La question est de savoir quel ordre il faut prendre. Parzen (1974, 1977) a considéré ce problème et a proposé un critère baptisé CAT (Criterion Autoregressive Transfer Function). Parallèlement, Akaike (1969, 1970) a considéré l'utilisation des modèles AR pour la prédiction et a proposé un critère de choix de l'ordre basé sur l'erreur de prédiction finale (FPE - Final Prediction Error, en anglais). Il est clair qu'un modèle d'ordre plus élevé permet de mieux rendre compte de la réalité complexe du phénomène aléatoire considéré, mais cet argument seul conduit à choisir l'ordre le plus grand possible. En fait, quand on augmente l'ordre du modèle, l'estimation de ses paramètres se dégrade, et, à un certain point, le gain en fidélité du modèle est annulé par la mauvaise estimation de ses paramètres. L'erreur de prédiction finale d'Akaike tient compte à la fois de l'erreur de prédiction due à la mauvaise adéquation du modèle et de celle due à la mauvaise estimation de ses paramètres (d'où l'attribut final). Par la suite Akaike (1973) a introduit un critère plus général, applicable pour une classe de modèles quelconques, appelée AIC (Akaike Information Criterion). La philosophie de ce critère est basée sur une fonction de perte liée à la quantité d'information de Kullback-Leibler et est donc différente de la précédente, quoique dans le cas des modèles ARMA, la considération du FPE conduit à un critère asymptotiquement équivalent (voir Pham, 1983). D'autres procédures de choix automatique de l'ordre du modèle ont été proposées par la suite. Elles sont toutes basées sur un critère, qui associe à chaque modèle (dans une classe donnée a priori) une valeur numérique, et on choisira le modèle qui minimise le critère en question. Les avantages de cette approche ont déjà été mentionnés. L'inconvénient est qu'elle peut être très coûteuse sur le plan numérique, car il faut examiner tous les modèles possibles. De plus, l'évaluation du critère nécessite l'estimation des paramètres du modèle par une méthode optimale (maximum de vraisemblance), de sorte que l'on est obligé d'ajuster un très grand nombre de modèles pour n'en retenir qu'un seul. Toutefois, Hannan et Rissanen (1982) ont proposé une procédure rapide pour choisir l'ordre du modèle. Celle-ci peut être utilisée pour un choix préliminaire qui sera raffiné par la suite. Un tel procédé à plusieurs étapes a été suggéré dans Hannan et Kavalieris (1984a, 1984b) et Hannan et Deistler (1988).

Nous avons distingué deux types d'approche pour le choix de l'ordre du modèle ARMA, et mentionné succinctement les travaux qui sont à leur origine. Dans la suite nous examinerons plus en détail les développements récents de ces deux approches. Citons toutefois une autre démarche consistant à considérer le problème comme un problème de test d'hypothèse multiple où il faut choisir une parmi plusieurs hypothèses possibles (McClave, 1978, Duong, 1984). On peut aussi envisager de sélectionner le modèle par une succession de tests d'adéquation. Pour éviter le problème d'identification, on commence par les modèles d'ordre faible et on augmente l'ordre tant que le test rejette le modèle (Pötscher, 1983).

3. APPROCHE INTERACTIVE POUR LE CHOIX DE L'ORDRE

La définition (2.1) montre que la fonction d'autocovariance d'un processus ARMA satisfait une relation de récurrence linéaire. En effet, multiplions les deux membres de (2.1) par X_{t-m} et prenons l'espérance mathématique de chaque membre; nous obtenons,

$$\gamma_m = - \sum_{j=1}^p a_j \gamma_{m-j}, \quad m > q, \quad (3.1)$$

où $\gamma_m = E(X_t X_{t-m})$ désigne la fonction d'autocovariance du processus. Evidemment, la fonction d'autocorrélation $\rho_k = \gamma_k / \gamma_0$ satisfait la même relation de récurrence. Gray, Kelley et McIntyre (1978) et Gray, Houston et Morgan (1978) sont les premiers à exploiter cette propriété pour déterminer l'ordre du modèle ARMA via ce qu'on appelle l'epsilon-algorithme. Cet algorithme (voir Shanks, 1955, Wynn, 1956) a été inventé par les numériciens pour la sommation des séries numériques et on sait que si le terme général d'une série tend vers zéro géométriquement, alors l'algorithme permet de la sommer en un nombre fini de pas. L'application de cette technique à la série $\sum \gamma_k$ ou $\sum \rho_k$ permet de déterminer l'ordre du modèle, car γ_k tend vers zéro géométriquement à partir du rang q . Différentes variantes de l'epsilon-algorithme ont été proposées (voir Berline, 1984, 1985), en particulier la méthode des tableaux R-S de Gray, Kelley et McIntyre (1978). Certains éléments du tableau S sont en fait des autocorrélations partielles (Woodward et Gray, 1981). Parallèlement, Béguin, Gouriéroux et Montfort (1980) ont proposé une méthode dite "du coin" pour déterminer l'ordre du modèle ARMA. Elle est basée sur le tableau des $\Delta(j, i)$ qui sont les déterminants des matrices de Toeplitz

$$D(j, i) = \begin{pmatrix} p_j & \cdots & p_{j-i+1} \\ \vdots & \cdots & \vdots \\ p_{j+i-1} & \cdots & p_j \end{pmatrix},$$

et sont donc, d'après (3.1), nuls pour $i > p$ et $j > q$. Notons que des méthodes semblables ont été proposées par Woodside (1971) et Chow (1972). Une autre méthode, proposée par Graupe, Krause et Moore (1975) procède de façon analogue, mais utilise la réponse impulsionnelle, c'est-à-dire les coefficients w_k de l'expansion

$$\left(1 + \sum_{j=1}^q b_j z^j \right) / \left(1 + \sum_{j=1}^p a_j z^j \right) = \sum_{j \geq 0} w_j z^j,$$

à la place des autocorrélations. La méthode du coin est d'autre part directement liée à l'epsilon algorithm car les quantités générées par ce dernier sont fonctions du déterminant des matrices de Hankel

$$H(k, m) = \begin{pmatrix} p_m & \cdots & p_{m+k-1} \\ \vdots & \cdots & \vdots \\ p_{m+k-1} & \cdots & p_{m+2k-2} \end{pmatrix}$$

Il est clair que $H(k, m) = \det D(m+k-1, k)$. La méthode du coin nous semble beaucoup plus simple que l'epsilon algorithm. L'intérêt de ce dernier provient du fait qu'il est plus rapide sur le plan du calcul, mais récemment Pham (1984) a obtenu des relations de récurrence pour calculer les quantités $\Delta(j, i)$ pour la méthode du coin. La plus intéressante est

$$\Delta(j+1, i) \Delta(j-1, i) = \Delta(j, i)^2 - \Delta(j, i+1) \Delta(j, i-1).$$

Notons que si les méthodes précédentes appliquées à la fonction d'autocorrélation permettent de déterminer exactement l'ordre du modèle, dans la pratique cette fonction est inconnue. On doit donc la remplacer par son estimateur, la fonction d'autocorrélation

empirique, et il se pose le problème de prendre en compte l'erreur due à l'estimation. Dans la méthode du coin par exemple; si les $\Delta(j, i)$ sont nuls pour $i > p, j > q$, les $\hat{\Delta}(j, i)$, obtenus comme $\Delta(j, i)$ mais avec les autocorrélations empiriques r_k à la place des autocorrélations théoriques ρ_k , ne le seront pas. Il faut donc interpréter une petite valeur de $\hat{\Delta}(j, i)$ comme pouvant provenir d'un $\Delta(j, i)$ nul, mais pour cela il faut connaître la loi de la statistique $\hat{\Delta}(j, i)$. Cette dernière est asymptotiquement normale, pour $i \leq p$ ou $j \leq q$, mais sa variance asymptotique est assez compliquée et se prête mal au calcul numérique. Toutefois Mareschal et Mélard (1988) ont fourni un procédé de calcul basé sur les relations de récurrence de Pham (1984). On rencontre la même difficulté dans l'épsilon-algorithme. La méthode ci-dessous, due à Glasbey (1982) a l'avantage que les statistiques considérées ont une variance asymptotique très simple à calculer. Ces statistiques sont données par

$$y_{ij} = r_{j+i+1} + \sum_{k=1}^i \hat{a}_k(i, j) r_{i+j+1-k}$$

où les $\hat{a}_k(i, j)$ sont solution du système d'équations

$$r_l + \sum_{k=1}^i \hat{a}_k(i, j) r_{l-k} = 0, \quad l = j+1, \dots, j+i. \tag{3.2}$$

On peut interpréter y_{ij} comme la différence entre le premier et le second membre de (3.1) quand $p = i, m = j + i + 1$, avec γ_m remplacés par r_m et a_k remplacés par les estimateurs $\hat{a}_k(i, j)$. Par suite pour $i \geq p, j \geq q, y_{ij} \rightarrow 0$ quand $n \rightarrow \infty$ (n étant la taille de l'échantillon). La méthode de Glasbey est en fait équivalente à la méthode du coin car on peut montrer que $y_{ij} = (-1)^i \hat{\Delta}(j+1, i+1) / \hat{\Delta}(i, j)$, (voir Pham, 1984). L'avantage des statistiques y_{ij} est que leur loi asymptotique est très simple. Glasbey (1982) a montré que les statistiques $n^{1/2} y_{pq}, \dots, n^{1/2} y_{p+1, q}$ et $n^{1/2} y_{p, q+1}, \dots, n^{1/2} y_{p, q+J}$ (I, J fixés) convergent en loi vers des variables $Y_{00}, \dots, Y_{I0}, Y_{10}, \dots, Y_{J0}$ de loi conjointe normale centrée de structure de covariance, donnée par

$$E(Y_{oi} Y_{oj}) = \sum_m v_m v_{m+i-j}$$

où

$$v_m = \sum_{k=0}^p \sum_{l=0}^p a_k a_l p_{m+k-l} = \sigma^2 \sum_{k=0}^{q-|m|} b_k b_{k+|m|}, \quad (a_0 = b_0 = 1). \tag{3.3}$$

Notons que $v_m = 0$ pour $m > q$. Par la suite, on examinera les statistiques normalisées

$$z_{ij} = y_{ij} \left[n / \hat{v}_m(i, j) \right]^{1/2}.$$

où

$$\hat{v}_m = \sum_{k=0}^p \sum_{l=0}^p a_k(i, j) a_l(i, j) r_{m+k-l},$$

qui pour $i = p, j \geq q$ ou $i \geq p, j = q$, sont asymptotiquement normales centrées de variance unité.

Notons que toutes les méthodes reposent sur la propriété que les $\Delta(j, i)$ associés à un modèle ARMA d'ordre (p, q) sont nuls pour $i > p$ et $j > q$, mais il n'est pas clair que l'inverse soit vrai. La méthode de Glasbey repose aussi sur le fait que le système (3.2) ne converge pas vers un système singulier, c'est-à-dire $\Delta(j, i) \neq 0$, pour $i = p, j \geq q$ ou $i \geq p, j = q$. Béguin, Gouriéroux et Montfort (1980) ont montré que c'est le cas si (p, q) est le couple d'ordre minimal du modèle (en fait minimum, cf. Théorème 3.2). Le résultat suivant (Pham, 1984) précise les positions possibles des zéros dans le tableau des Δ .

Théorème 3.1

Les zéros dans le tableau des Δ (resp. $\hat{\Delta}$) se groupent toujours en blocs carrés de taille k (k peut être infini). Si $\Delta(i, j)$ (resp. $\hat{\Delta}(i, j)$), $i = r+1, \dots, r+k, j = s+1, \dots, s+k$ est un tel bloc, alors il existe un vecteur unique $(\alpha_0, \dots, \alpha_r)'$ avec $\alpha_0 = 1$, tel que

$$\sum_{l=0}^r \alpha_l p_{j-l} = 0, \text{ (resp. } \sum_{l=0}^r \alpha_l r_{j-l} = 0), \quad j = s+1, \dots, s+k.$$

Notons que pour $i > p, j > q$, le système (3.2) converge vers un système singulier, et les statistiques $n^{1/2} y_{ij}$ ainsi que z_{ij} ne seront plus asymptotiquement

normales. Pham (1982) montre que celles-ci convergent encore en loi, mais la loi limite est assez compliquée pour être d'une grande utilité. Evidemment, dans le cas où $i \leq p$ ou $j \leq q$, y_{ij} et $z_{ij} / n^{1/2}$ convergent vers une limite finie, sauf si $\Delta(j-1, i-1) = 0$.

Les statistiques y_{ij} précédentes peuvent avoir une variance importante, ce qui limite la capacité de discerner des modèles d'ordre voisin de la méthode précédente. En effet, en considérant $\hat{a}_k(i, j)$ comme des constants, y_{ij} apparaît comme une combinaison linéaire des autocorrélations empiriques de retard $j+1, \dots, j+i$. Or, les autocorrélations de grands retards sont généralement mal estimées. En plus si $\Delta(j, i)$ est proche de zéro, les $\hat{a}_k(i, j)$ seront très instables. Pour cette raison, Pham (1988) a introduit une nouvelle méthode où les statistiques considérées sont calculées à partir d'un grand nombre d'autocorrélations estimées, ce qui améliore leur précision. Cette méthode est basée sur la caractérisation suivant du modèle ARMA.

Théorème 3.2

Notons $X_{t|s}$ la projection de X_t sur le sous-espace de Hilbert engendré par les $X_u, u \leq s$ (avec le produit scalaire $X, Y \rightarrow E(XY)$). Le processus X_t est ARMA d'ordre (p, q) si et seulement si $X_{t|t-q-1}$ est une combinaison linéaire de $X_{t-j|t-q-1}, j = 1, \dots, p$. Dans ce cas, il existe p_0, q_0 tel que le processus est ARMA d'ordre (p', q') pour tout $p' \geq p_0, q' \geq q_0$, mais n'est pas un processus ARMA d'ordre (p', q') si $p' < p_0$ ou $q' < q_0$. En plus, pour $p \leq p_0$ ou $q \leq q_0$, les $X_{t-j|t-q-1}, j = 1, \dots, p$ sont linéairement indépendants, et en notant $a_j(p, q)$ les coefficients de la régression de $X_{t|t-q-1}$ sur ceux-ci, le polynôme

$$1 + \sum_{j=1}^p a_j(p, q) z^j \text{ est stable.}$$

Note

La définition du processus ARMA dans le Théorème précédent n'exclut pas le cas où le polynôme autorégressif a des racines sur le cercle unité pourvu que celles-ci soient annulées par des racines identiques du polynôme moyenne mobile. Ainsi la classe de ces processus ARMA contient les processus harmoniques avec bruit additif :

$$X_t = \sum_{k=1}^r A_k \cos(w_k t + \psi_k) + e_t$$

où e_t un bruit blanc (ou plus généralement un processus stationnaire de densité spectrale rationnelle) et w_k sont des nombres réels et $A_k \exp(i\psi_k)$ sont des variables aléatoires complexes non corrélées, c'est-à-dire $E[A_j A_k \exp(i\psi_j - i\psi_k)] = 0$ pour $j \neq k$. Il est facile de voir que ce processus admet la représentation ARMA (2.1) avec $p = q = 2r$ et $b_j = a_j$, $j = 1, \dots, 2r$. Toutefois, de tels processus sont souvent exclus dans la littérature, par l'adjonction de la condition que le polynôme autorégressif n'ait pas de racines de module unité.

Les entiers p_0, q_0 du théorème 3.2 sont les ordres minimum du modèle. Dans la suite, l'ordre sera sous-entendu l'ordre minimum et sera noté (p_0, q_0) . Le théorème précédent montre que les $a_j(p, q)$ sont uniquement définis pour $p \leq p_0$ ou $q \leq q_0$ et coïncident avec les coefficients a_j du modèle (2.1). Toutefois, les $a_j(p, q)$ font intervenir toute la fonction d'autocorrélation tandis qu'en pratique, sur un échantillon de taille finie, seule un petit nombre des autocorrélations peuvent être raisonnablement estimées. On est donc amené à remplacer les $X_{t-j|t-q-1}$ par $X_{t-j|t-q-1, t-q-m}$ où m est un entier donné et $X_{t|t, s}$ désigne la projection de X_t sur le sous-espace engendré par X_t, \dots, X_s . D'après le théorème 3.1, pour $p = p_0$ ou $q = q_0$ et $m \geq p$, les $X_{t-j|t-q-1, t-q-m}$, $j = 1, \dots, p$ sont linéairement indépendants. Ce résultat n'est pas forcément vrai si $p < p_0$ ou $q < q_0$ mais d'après le théorème 3.2, ce sera le cas si m est suffisamment grand. Dans ces conditions, on peut définir uniquement les coefficients $-a_1(p, q, m), \dots, -a_p(p, q, m)$ de la régression de $X_{t+q|t-1, t-m}$ sur $X_{t-q|t-1, t-m}, \dots, X_{t|t-q-p, t-q-m}$. Ces coefficients dépendent uniquement des autocorrélations jusqu'au retard $m+q$, et en remplaçant ces dernières par les autocorrélations empiriques, on obtient les estimateurs $\hat{a}_j(p, q, m)$ de $a_j(p, q, m)$. Pham (1989) a obtenu des relations de récurrence pour calculer rapidement les $a_j(p+1, q, m)$ en fonction de $a_j(p, q, m)$ et des coefficients $-a_j^*(p, q, m)$ de la régression de $X_{t+q-p-1|t-1, t-m}$ sur $X_{t+q-j|t-1, t-m}$, $j = 1, \dots, p$. Les $a_j^*(p+1, q, m)$, de leurs côtés, peuvent être calculés à partir de $a_j(p, q-1, m)$ et $a_j^*(p, q-1, m)$. Ces relations de récurrence font intervenir les variances $v^*(p, q, m)$ des résidus des régression en question, c'est-à-dire de

$$e_t(p, q, m) = X_{t+q|t-1, t-m} + \sum_{j=1}^p a_j(p, q, m) X_{t+q-j|t-1, t-m},$$

$$e_t^*(p, q, m) = X_{t+q-p|t-1, t-m} + \sum_{j=1}^p a_{p+1}^*(p, q, m) X_{t+q-j|t-1, t-m},$$

ainsi que la covariance partielle $\pi(p, q, m)$ entre $X_{t+q|t-1, t-m}$ et $X_{t+q-p|t-1, t-m}$ en gardant fixe $X_{t+q-j|t-1, t-m}$, $j = 1, \dots, p-1$, c'est-à-dire la covariance entre $e_t(p-1, q, m)$ et $e_t^*(p-1, q, m)$. Ces variances résiduelles et covariance partielle seront évidemment estimées par les variances résiduelles et covariance partielle empiriques $\hat{v}(p, q, m)$, $\hat{v}^*(p, q, m)$ et $\hat{\pi}(p, q, m)$, obtenues de la même façon avec les autocorrélations empiriques à la place des autocorrélations théoriques. D'après le théorème 3.2, pour $q = q_0$ et $p > p_0$, $v(p-1, q, m)$ et donc $\pi(p, q, m)$ sont nuls et on peut montrer (Pham, 1989) que $n^{1/2} \hat{\pi}(p, q, m)$ est asymptotiquement normale centrée réduite de variance approximativement (pour grand m) égale à $v_0 v^*(p, q, m)$, où v_0 est définie dans (3.3). On examine donc les statistiques $n^{1/2} \hat{\pi}(p, q, m) / [\hat{v}_0 \hat{v}^*(p, q, m)]$, où \hat{v}_0 est un estimateur de v_0 , pour la détermination de l'ordre du modèle. Pour $p > p_0$, $q = q_0$, ces statistiques sont asymptotiquement normales centrées de variance unité, tandis que pour $p > p_0$, $q > q_0$, on peut montrer qu'elles sont au plus 0 (m) en probabilité (Pham, 1988).

Notons que pour $m = p$, les $\hat{a}_j(p, q, m)$ se réduisent aux $\hat{a}_j(p, q)$ dans la méthode de Glasbey précédente, et les $\hat{\pi}(p+1, q, m)$ se réduisent aux y_{pq} . L'utilisation de $m > p$ permet d'avoir des estimateurs plus précis des paramètres a_j du modèle, et réduit la variabilité des statistiques indicatrices de l'ordre (y_{ij} ou $\hat{\pi}(p, q, m)$). Notons également que la quantité $\sigma^2(p, q, m) = \sigma^2 + v(p, q, m)$ s'interprète comme la variance de la différence entre X_t et sa projection sur le sous-espace linéaire engendré par X_{t-1}, \dots, X_{t-p} et e_{t-1}, \dots, e_{t-q} . Hannan et Rissanen (1982) ont proposé d'effectuer la régression de X_t sur X_{t-1}, \dots, X_{t-p} et \hat{e}_{t-1} , où les \hat{e}_t sont les résidus de l'ajustement d'un modèle AR d'ordre élevé m , pour obtenir les estimateurs des coefficients du modèle ARMA. Les estimateurs de a_j ainsi obtenus sont en fait asymptotiquement équivalents à nos $\hat{a}_j(p, q, m)$, dans le cas où m tend vers l'infini avec n . L'algorithme proposée par ces auteurs, pour calculer leur estimateurs, ainsi que celui de Franke (1985), est toutefois moins efficace sur le plan numérique que celui mentionné plus haut pour le calcul de $\hat{a}_j(p, q, m)$. Cet algorithme fournit aussi les estimateurs $\hat{\sigma}^2(p, q, m)$ de $\sigma^2(p, q, m)$, qui sont à la base d'une procédure du choix automatique de l'ordre du modèle de Hannan

et Rissanen (1982) (ces auteurs ont fait tendre m vers l'infini avec n). Nous reviendrons sur cette procédure dans le paragraphe suivant.

Récemment, Chaverie, Szpiro et Topol (1990) ont proposé une autre méthode d'identification de modèle ARMA basé sur la transformée en z de la réponse impulsionnelle $H(z) = \sum_{j \geq 0} w_j z^j$. On cherche, par la méthode de Padé, une approximation rationnelle de degrés (i, j) de cette fonction dans un voisinage d'un point z_0 (à choisir). Il est vrai que l'approximation est exacte quand le processus est ARMA d'ordre (p, q) avec $p \geq i, q \geq j$, et par suite l'examen des suites approximantes de Padé permet de déceler les ordres du modèle. Cette méthode diffère des précédentes essentiellement du fait qu'elle est basée sur les réponses impulsionnelles w_j au lieu des autocorrélations, et en plus comme on considère l'expansion de Taylor de $H(z)$ autour de z_0 , ce sont en fait des combinaisons linéaires de w_j qui jouent le rôle des autocorrélations (dans le même ordre d'idée, on peut travailler avec une certaine transformation de la séquence des autocorrélations comme dans la note ci-dessous). Le choix de z_0 , n'est toutefois pas clair et aucun résultat concernant le comportement probabiliste des statistiques considérées n'est disponible.

Note

1) Si la fonction d'autocorrélations ρ_j du processus satisfait une récurrence linéaire, la fonction $z^j \rho_j$ satisfait aussi une récurrence de même ordre. Pour des raisons de stabilité, on se restreint à $|z| = 1$ (z est complexe); alors $z_j \rho_j$ s'interprète comme la fonction d'autocorrélation du processus (complexe) $z^t X_t$. On peut donc appliquer les méthodes précédentes avec $z^j \rho_j$ à la place de ρ_j pour la détermination de l'ordre du modèle (on prend souvent $z = -1$, pour rester dans le domaine réel). Notons toutefois que le processus $z^t X_t$ est seulement ARMA au sens large, ce qui peut invalider les résultats asymptotiques concernant les statistiques indicatrices de l'ordre. En tous cas, ces résultats ne sont applicables que si la constante z est choisie indépendamment de l'observation.

2) Les autocorrélations partielles empiriques (à partir du rang $p_0 + 1$) ont la propriété intéressante d'être asymptotiquement indépendantes. On ne retrouve pas cette propriété ni pour les autocorrélations empiriques, ni pour les statistiques introduites pour la détermination de l'ordre du modèle ARMA. Or si X_t suit un modèle MA, l'inverse de sa densité spectrale est celle d'un processus AR. D'où l'idée de considérer les

autocorrélations partielles associées à cette inverse, appelées autocorrélations partielles inverses, pour la détermination de l'ordre du modèle. Ces corrélations ont été introduites par Cleveland (1972). Leur estimation et leur utilisation pour la détermination de l'ordre du modèle MA ont été considérées par Chatfield (1979) et Bhansali (1980, 1983). La méthode toutefois ne s'applique qu'à ce modèle.

4. CHOIX AUTOMATIQUE DE L'ORDRE

Il existe de nombreux travaux consacrés au problème de choix automatique de l'ordre du modèle AR. Le premier critère proposé par Akaike est basé sur la notion d'erreur de prédiction finale. Soient a_{1p}, \dots, a_{pp} les coefficients de la régression de X_t sur X_{t-1}, \dots, X_{t-p} ; adopter un modèle AR d'ordre p revient à utiliser

$-\sum_{j=1}^p a_j X_{t-j}$ comme prédicteur de X_t . L'erreur de prédiction (à un pas) est donc

$e_t(p) = X_t + \sum_{j=1}^p a_j X_{t-j}$, et sa variance est notée σ_p^2 . Si on tient compte de

l'erreur d'estimation des coefficients du modèle, on arrive à une erreur de prédiction finale dont la variance est $\sigma_p^2(1+p/n)$. Mais, comme on ne connaît pas σ_p^2 , on doit l'estimer, par exemple, par

$$\hat{\sigma}_p^2 = \sum_{t=p+1}^n \left(X_t + \sum_{j=1}^p \hat{a}_{jp} X_{t-j} \right)^2$$

où \hat{a}_{jp} sont des estimateurs des moindres carrés de a_{jp} . Or, cet estimateur $\hat{\sigma}_p^2$ est biaisé et, pour corriger son biais, on doit le multiplier par $n / (n - p)$. On est donc amené au critère

$$\text{FPE}(p) = \hat{\sigma}_p^2 \left(1 + \frac{p}{n} \right) / \left(1 - \frac{p}{n} \right) \approx \hat{\sigma}_p^2 \left(1 + \frac{2p}{n} \right),$$

introduit par Akaike (1969) sous le nom FPE (final prediction error en anglais). Notons que les calculs précédents concernent un modèle centré; dans le cas non centré, le facteur p/n sera remplacé par $(p+1)/n$. D'un autre côté, l'ajustement du modèle autorégressif d'ordre p permet de construire l'estimateur de la densité spectrale du processus par

$$\hat{f}(\lambda) = \hat{\sigma}_p^2 / \left(2\pi \left| 1 + \hat{a}_{1p} e^{-i\lambda} + \dots + \hat{a}_{pp} e^{-ip\lambda} \right|^2 \right).$$

Un critère raisonnable pour le choix de l'ordre p du modèle est donc l'erreur intégrée de cet estimateur de densité spectrale :

$$\int_{-\infty}^{\infty} E \left\{ \left[\hat{f}(\lambda) - f(\lambda) \right]^2 / f(\lambda) \right\} d\lambda .$$

Parzen (1974, 1976) a montré que minimiser cette erreur est asymptotiquement équivalent à minimiser le critère CAT (Autoregressive Transfer Criterion)

$$\text{CAT}(p) = \frac{n-j}{n^2} \sum_{j=1}^p \hat{\sigma}_j^{-2} - \frac{n-p}{n} \hat{\sigma}_p^{-2}, \quad p > 0, \quad (\text{CAT}(0) = -(1 + 1/n)).$$

En pratique, les critères FPE et CAT choisissent très souvent le même ordre. Ceci se voit en remarquant que pour p voisin de l'ordre vrai p_0 du modèle,

$(p - p_0) / n$ et $\hat{\sigma}_p^2 - \hat{\sigma}_{p_0}^2$ sera petite et par suite

$$\text{FPE}(p) \approx \text{Const.} + \hat{\sigma}_p^2 - \hat{\sigma}_{p_0}^2 + \frac{2}{n} (p - p_0) \hat{\sigma}_{p_0}^2$$

$$\text{CAT}(p) \approx \text{Const.} + \left(\hat{\sigma}_p^2 - \hat{\sigma}_{p_0}^2 \right) / \hat{\sigma}_{p_0}^4 + \frac{2}{n} (p - p_0) / \hat{\sigma}_{p_0}^2.$$

Plus récemment, Akaike (1973) a introduit un critère très général pour le choix d'un modèle basé sur la notion d'information. Considérons un modèle spécifié par un paramètre vectoriel $\theta \in \Theta$, un ouvert de \mathbb{R}^p . Soit $L_n(\theta)$ la fonction log-vraisemblance normalisée par la taille n de l'échantillon. Alors, sous des conditions très générales, $L_n(\theta)$ converge quand n tend vers infini vers $L_\infty(\theta) = -H(P_\theta, P)$ où P_θ désigne la loi de l'observation associée au modèle et P la vraie loi, et où $H(Q, P)$ est l'entropie relative de Q par rapport à P (au sens de Kullback-Leibner). On sait que $H(Q, P)$ est minimum quand $Q = P$, et par suite $H(P_\theta, P) - H(P, P)$ peut être utilisé comme une fonction de perte quand le modèle spécifié par θ est utilisé à la place du "vrai modèle" (ce dernier peut ne correspondre à aucun θ dans Θ). Soient maintenant deux classes de

modèles spécifiées par deux sous-ensembles de Θ' et Θ'' de Θ . Il est naturel de préférer la première classe si $L_\infty(\hat{\theta}')$ > $L_\infty(\hat{\theta}'')$ où $\hat{\theta}'$, $\hat{\theta}''$ sont les estimateurs du maximum de vraisemblance (ou d'une méthode asymptotiquement équivalente) relative à ces deux classes. La difficulté est qu'on ne connaît pas la fonction $L_\infty(\cdot)$.

L'idée est de remplacer $L_\infty(\hat{\theta}') - L_\infty(\hat{\theta}'')$ par $L_n(\hat{\theta}') - L_n(\hat{\theta}'')$, mais le premier est un estimateur biaisé du second. Pour le voir, considérons le cas où Θ' est isomorphe à un ouvert de $\mathbf{R}^{p'}$, Θ'' est contenu dans Θ' et isomorphe à un ouvert de $\mathbf{R}^{p''}$ et la vraie loi est donnée par P_θ où θ_0 est un point de Θ'' .

Il est bien connu que $2n \left[L_n(\hat{\theta}') - L_n(\hat{\theta}'') \right]$ converge en loi quand n tend vers infini vers une variable du χ^2 à $p' - p''$ degrés de liberté. D'autre part, un développement de Taylor donne

$$L_\infty(\theta_0) \approx L_\infty(\hat{\theta}') + \frac{1}{2} (\theta_0 - \hat{\theta}')^* \left(\frac{\partial^2}{\partial \theta} L_\infty \right) (\hat{\theta}') (\theta_0 - \hat{\theta}'),$$

Le signe * désignant le transposé. D'après les propriétés bien connues de $L_\infty(\theta) = E[L_n(\theta)]$ et de l'estimateur du maximum de vraisemblance, le dernier terme de l'expression précédente tend vers $p'/2$ quand n tend vers infini.

Donc $L_\infty(\hat{\theta}') - L_\infty(\hat{\theta}'')$ converge vers $(p' - p'')/2$ tandis que $L_n(\hat{\theta}') - L_n(\hat{\theta}'')$ converge en loi vers une variable de moyenne $(p' - p'')/2$.

Par suite, il faut prendre $\left[L_n(\hat{\theta}') - p' \right] - \left[L_n(\hat{\theta}'') - p'' \right]$ pour obtenir un estimateur asymptotiquement sans biais de $L_\infty(\hat{\theta}') - L_\infty(\hat{\theta}'')$. Ce résultat a été montré dans le cas général par Findley (1985), sous des conditions raisonnables. On est donc amené au critère

$$\text{AIC} = 2 (- \text{maximum de log vraisemblance} + \text{nombre de paramètres}).$$

Dans le cas du modèle AR d'ordre p , la fonction log-vraisemblance a pour expression approchée

$$-\frac{n}{2} \left[\log(2\pi\sigma^2) \right] + \sum_{t=p+1}^n \left(X_t + \sum_{j=1}^p a_j X_{t-j} \right)^2 / \sigma^2.$$

Cette fonction a pour maximum $-(n/2) \log(\hat{\sigma}_p^2) + \text{Const.}$ où $\hat{\sigma}_p^2$ est défini comme

précédemment. Le critère d'Akaike prend alors la forme

$$\text{AIC}(p) = n \log \hat{\sigma}_p^2 + 2p . \quad (4.1)$$

L'utilisation de la vraisemblance approchée permet de simplifier les calculs sans changer les propriétés asymptotiques de la procédure. Notons que la vraisemblance est obtenue sous l'hypothèse que le processus est gaussien, mais grâce au théorème de la limite centrale, les résultats asymptotiques ci-dessous restent valables sans cette hypothèse. En utilisant l'approximation $\log(1 + p/n) \approx p/n$ (p petit devant n), on voit que $\log \text{FPE}(p) \approx \text{AIC}(p)/n$. Les critères AIC et FPE sont donc asymptotiquement équivalents. Le lien entre l'AIC et le principe du maximum de l'entropie a été montré par Shimzu (1978) (voir aussi Bednar et Roberts, 1985).

La performance du critère AIC (ou FPE) pour le choix de l'ordre du modèle AR a été beaucoup étudié, aussi bien sur le plan empirique (Jones, 1975, Bhansali et Downham 1977) que sur le plan théorique (Bhansali et Downham 1977, Shibata, 1976, 1980, 1981). Il est montré que si la série observée suit réellement un modèle AR d'ordre p_0 , alors le critère AIC choisira un ordre $p < p_0$ avec une probabilité tendant vers 0 quand n tend vers infini. Toutefois, pour tout $p > p_0$, le critère choisira l'ordre p avec une probabilité tendant vers une limite positive. Cette limite est liée à certaines probabilités associées au processus de marche aléatoire (Bhansali et Downham, 1977). Par la suite, la procédure d'Akaike n'est pas convergente comme estimateur de l'ordre du modèle. Toutefois, Shibata (1980, 1981) a montré que cette procédure est asymptotiquement optimale dans le cas où le processus sous-jacent admet seulement une représentation AR d'ordre infini mais n'est pas un processus AR (d'ordre fini). Dans ce cas, comparer l'ordre choisi avec le vrai ordre (qui est infini) n'a pas de sens. Shibata a en fait montré l'optimalité de la procédure d'Akaike selon deux critères : l'erreur de prédiction à un pas et l'erreur de l'estimation de la densité spectrale, basée sur l'ajustement du modèle AR d'ordre choisi. Taniguchi (1980) a montré que le résultat de Shibata reste valable dans le cas du modèle ARMA, en ce qui concerne le deuxième critère d'optimalité.

Pour corriger la tendance de surestimation de l'ordre de la procédure d'Akaike (quand le processus sous-jacent est un processus AR), on remplacera le terme $2p$ dans (4.1) par $\alpha_n p$, où α_n est une suite donnée. Bhansali et Downham (1977) a considéré le choix $\alpha_n = \alpha$ une constante donnée, ainsi que Atkinson (1980) dans le contexte

différent de la régression. Il s'avère que $\alpha > 2$ permet seulement d'atténuer la surestimation de l'ordre de la procédure. Pour obtenir une procédure convergente, il est nécessaire que α_n tend vers infini avec n , et Hannan et Quinn (1979) ont montré que la vitesse minimum de croissance de α_n est $2c \log \log n$, $c > 1$. Mais il existe des arguments suggérant l'utilisation d'une vitesse encore plus élevée. En effet, Akaike (1978, 1979) a obtenu à partir des arguments du type Bayésien, un nouveau critère appelé BIC (Bayesian Information Criterion). Celui-ci s'écrit, pour un modèle AR d'ordre p ,

$$\text{BIC}(p) = n \log \hat{\sigma}_p^2 - 2(n-p) \log \left(1 - \frac{p}{n} \right) + p \log n + p \log \left[\frac{1}{p} \left(\hat{\sigma}_X^2 / \hat{\sigma}_p^2 - 1 \right) \right]$$

où $\hat{\sigma}_p^2$ désigne la variance empirique du processus. Pour p fixé et n tendant vers infini, on a

$$\text{BIC}(p) = n \log \hat{\sigma}_p^2 + p \log n + o(1).$$

En négligeant le terme $o(1)$ on obtient une version simplifiée de BIC qui est précisément le critère de Schwarz (1978) : $\log \hat{\sigma}_p^2 + p \log n$.

En résumé, pour le modèle AR, on dispose d'un critère de choix automatique de l'ordre de la forme $n \log \hat{\sigma}_p^2 + \alpha_n p$. Le choix de α_n dépend de l'objectif de l'analyse. S'il s'agit d'une estimation spectrale ou d'une prédiction non paramétrique (où on n'est pas sûr que le processus sous-jacent soit autorégressif), il est recommandé de prendre $\alpha_n = 2$ ou $\alpha_n = \alpha > 2$. S'il s'agit d'estimer l'ordre d'un processus (réellement) autorégressif, il est préférable de prendre $\alpha_n = \log n$ ou $\alpha_n = 2c \log \log n$, $c > 1$. Une comparaison de ces différentes procédures appliquées aux séries réelles d'EEG (électroencéphalogramme) se trouvent dans Franke, Gasser et Steinberg (1985).

Sur le plan pratique, le calcul du critère peut être effectué rapidement à l'aide de l'algorithme de Levinson-Durbin. Cet algorithme permet d'ajuster de façon récursive des modèles AR d'ordre successif. On obtient ainsi les autocorrélations partielles empiriques, soient \hat{a}_{pp} , $p \leq 1$, et on calcule les $\hat{\sigma}_p^2$ récursivement par

$$\hat{\sigma}_p^2 = \hat{\sigma}_{p-1}^2 (1 - \hat{a}_{pp}^2).$$

Cette relation montre le lien entre la procédure automatique et l'approche interactive basée sur l'examen de la fonction d'autocorrélation partielle empirique. En effet, la différence des critères correspondant aux ordres p et p' est approximativement

$$n \left(\hat{a}_{pp}^2 - \alpha_n \right) + \dots + n \left(\hat{a}_{pp'}^2 - \alpha_n \right).$$

La généralisation des procédures précédentes aux modèles ARMA est immédiate. On minimise le critère

$$\log \hat{\sigma}_{pq}^2 + \alpha_n (p + q) \quad (4.2)$$

où $\alpha_n = 2$ pour AIC, $= \log n$ pour BIC, $= c \log \log n$ pour le critère d'Hannan et Quinn. Ici, $\hat{\sigma}_{pq}^2$ est l'estimateur de la variance de l'innovation dans un modèle ARMA d'ordre q , obtenue comme la moyenne quadratique des résidus d'un ajustement standard (voir par ex. Box et Jenkins, 1970) du modèle. La forme (4.2) résulte en fait de l'utilisation d'une fonction de vraisemblance approchée, calculée sous l'hypothèse gaussienne. Cette forme peut aussi bien résulter de l'utilisation d'une fonction de vraisemblance exacte, toujours sous l'hypothèse gaussienne, à condition d'incorporer un terme jacobien (voir Ansley, 1979). Ici, l'hypothèse gaussienne est utilisée pour écrire la vraisemblance mais les propriétés asymptotiques de la procédure ne dépendent pas de cette hypothèse. Notons toutefois que même en travaillant avec la vraisemblance approchée, l'utilisation des critères précédents est très lourde, car l'ajustement de modèle ARMA n'est pas récursif en ordre comme dans le cas du modèle AR, et demande en plus un calcul itératif coûteux. Pour cette raison, Hannan et Rissanen (1982) ont introduit un autre critère de la forme

$$n \log \hat{\sigma}^2(p, q) + (p + q) \log n \quad (4.3)$$

où $\hat{\sigma}^2(p, q)$ est l'estimateur de $\sigma^2(p, q, \infty)$, c'est-à-dire $\hat{\sigma}^2(p, q, m)$ avec m tendant vers infini avec n , introduit vers la fin du paragraphe 3. Notons que Hannan et Rissanen ont choisi m par la procédure Akaike, appliquée au modèle autorégressif, ce qui conduit à une procédure à deux étapes.

Une autre approche pour le choix de modèle a été proposée par Rissanen (1978), basée sur un principe qu'il appelle "longueur de description minimum". Le modèle est vu comme une description des données observées, et on cherche naturellement à trouver la description la plus courte. Rissanen considère donc le codage de données, qui est optimal pour le modèle proposé. On sait que le code sera (en moyenne) le plus court si et seulement si les données sont réellement engendrées selon celui-ci. Dans le cas où une famille paramétrique de modèles, indexée par θ est donnée, on cherchera le modèle qui rendra le plus court le code correspondant, ce qui conduit à un estimateur de $\hat{\theta}$ de θ , qui est en fait asymptotiquement équivalent à l'estimateur du maximum de vraisemblance. Mais le code résultant des observations ne suffit pas, car il faut coder aussi le modèle choisi, c'est-à-dire $\hat{\theta}$. La longueur totale de ces deux codes est le critère de comparaison entre différentes classes de modèles. Dans le cas de modèles ARMA ce principe conduit au critère de Schwarz, à des termes négligeables près. Rissanen a appliqué son principe dans un cadre "non prédictif", c'est-à-dire les données sont codées seulement quand elles sont toutes observées. Dans un travail ultérieur, Rissanen (1986) a formulé le principe de "longueur de description minimum prédictif" dans lequel le codage d'une observation s'effectue, dès qu'elle est disponible, selon sa loi de probabilité conditionnelle (suivant le modèle adopté) sachant les observations passées. Il a aussi introduit la notion de complexité stochastique, définie uniquement à partir des observations, de la donnée d'une classe de modèles et éventuellement des connaissances a priori, mais ne faisant aucune référence à l'existence d'un "vrai modèle". Pour plus de détails voir Rissanen (1978, 1983, 1986, 1987) et aussi Wallace et Freeman (1987).

Il est important de souligner une difficulté propre au problème du choix de l'ordre du modèle ARMA lié au surajustement. En effet, si $p > p_0$ et $q > q_0$ alors le modèle n'est pas identifiable et l'estimateur du maximum de vraisemblance sera très instable et le maximum de la vraisemblance n'aura pas le comportement standard (la matrice d'information de Fisher est en fait singulière). Hannan (1982) a considéré le cas où on ajuste un modèle ARMA d'ordre (1, 1) au bruit blanc et a montré que le maximum de $L_n(\theta)$ tend vers infini avec n (voir aussi Veres, 1987). Par suite, AIC dans ce cas choisira l'ordre (1, 1) avec une probabilité tendant vers 1. Dans le cas de BIC, à cause de la sévère pénalité associée au choix du modèle d'ordre élevé, ce critère choisira encore correctement l'ordre avec une probabilité tendant vers 1 (ce résultat de convergence a été établi de façon générale par Hannan, 1980). Toutefois on n'a pas de résultat de convergence pour le critère d'Hannan et Quinn ($\alpha_n = c \log \log n$). Dans tous les cas, il est clair qu'une procédure du choix de l'ordre basée sur le maximum de vraisemblance, favorise les modèles strictement surajustés. Or, on veut précisément éviter de les choisir

car ils ne sont pas identifiables et donc les paramètres estimés n'ont pas de signification. Notons aussi que l'ajustement par maximum de vraisemblance d'un tel modèle est aussi très difficile car les estimateurs ont tendance à aller vers le bord du domaine admissible des paramètres (de plus ils ne sont pas convergents). Le surajustement est inévitable car la procédure de choix consiste à ajuster des modèles de tout ordre, jusqu'à un maximum fixé à l'avance.

Une solution à ce problème est la procédure à plusieurs étapes proposée par Hannan et Kavalieris (1984, voir aussi Hannan, Kavalieris et MacKinsack, 1986, Hannan et Deistler, 1988). Cette procédure débute par la procédure à deux étapes décrite plus haut. Celle-ci peut aussi engendrer la surestimation de l'ordre, au sens que l'ordre choisi (p, q) converge vers $p_0 + v, q_0 + v$ où $v > 0$ (voir Hannan et Kavalieris, 1984). Toutefois, les paramètres estimés $\hat{a}_j(p, q, m)$ possèdent la propriété remarquable d'être convergents même si $p > p_0, q > q_0$. Plus précisément, dans le dernier cas, si m tend vers l'infini avec n avec une vitesse adéquate,

$1 + \sum_{j=1}^p \hat{a}_j(p, q, m) z^j$ converge vers $\Phi(z) (1 + \sum a_j z^j)$ où $\Phi(z)$ est un polynôme de degré $v = \max(p, q) - p_0$, de coefficients Φ_j avec $\Phi_0 = 1$ et $-\Phi_1, \dots, \Phi_v$ les

coefficients de régression de Y_t sur Y_{t-1}, \dots, Y_{t-v} , Y_t étant le processus $X_t + \sum a_j X_{t-j} = e_t + \sum b_j e_{t-j}$ (voir Hannan et Kavalieris, 1984, Pham, 1988).

Hannan et Kavalieris (1984) ont proposé différentes méthodes pour corriger la surestimation de la procédure à deux étapes dont l'une consiste à en ajouter un troisième.

Dans cette étape, on commence par calculer les estimations \tilde{e}_t de e_t , par la récurrence

$$\tilde{e}_t = - \sum_{j=1}^{\tilde{q}} \tilde{b}_j \tilde{e}_{t-j} + X_t + \sum_{j=1}^{\tilde{p}} \tilde{a}_j X_{t-j}$$

où (\tilde{p}, \tilde{q}) est l'ordre et $\tilde{a}_j, j = 1, \dots, \tilde{p}$ et $\tilde{b}_j, j = 1, \dots, \tilde{q}$ sont des estimateurs des paramètres, choisis à l'étape 2 (notons que cette étape comporte déjà une correction du choix de l'ordre obtenu par minimisation de (4.3)). On calcule ensuite de façon récursive

$$\eta_t = - \sum_{j=1}^{\tilde{q}} \tilde{b}_j \eta_{t-j} + y_t, \quad \xi_t = \sum_{j=1}^{\tilde{p}} \tilde{a}_j \xi_{t-j} + \tilde{e}_t,$$

et on fait, pour tout couple (p, q) , une régression de $\ddot{e}_t + \eta_t - \xi_t$ sur $-\eta_{t-1}, \dots, -\eta_{t-p}, \xi_{t-1}, \dots, \xi_{t-q}$. La moyenne quadratique des résidus de cette régression sera utilisée comme $\hat{\sigma}_{pq}^2$ dans (4.2) pour le choix de l'ordre du modèle.

Une méthode analogue, proposée dans Pham (1986) procède de la façon suivante. L'idée est de modifier la quantité $\hat{\sigma}_{pq}^2$ dans (4.3) de sorte que $-n/2$ fois son logarithme comporte comme un maximum de vraisemblance (i.e. diffère de la vraisemblance évaluée en le vrai modèle par une variable du χ^2), même quand $p > p_0, q > q_0$. Pour (p, q) donné, notons $\theta = (a_1, \dots, a_p, b_1, \dots, b_q)$ et $\hat{\theta}$ son estimateur de maximum de vraisemblance alors $\hat{\sigma}_{pq}^2$ apparaît comme moyenne quadratique des $\hat{e}_t(\hat{\theta}), t = 1, \dots, n$, où les $\hat{e}_t(\hat{\theta})$ sont définis par la récurrence

$$\hat{e}_t(\theta) = - \sum_{j=1}^q b_j \hat{e}_{t-j}(\theta) + X_t + \sum_{j=1}^p a_j X_{t-j}$$

avec les données initiales nulles. Notons Θ_0 l'ensemble des paramètres correspondant au vrai modèle et soit θ^* un estimateur $n^{1/2}$ convergent de θ quand $p \leq p_0$ ou $q \geq q_0$, au sens que $n^{1/2} \sup_{\theta \in \Theta_0} \|\theta^* - \theta\|$ est borné en probabilité quand n tend vers infini, Θ_0 étant l'ensemble des paramètres correspondant au vrai modèle. La

modification consiste alors à prendre comme $\hat{\sigma}_{pq}^2$ le minimum de $\sum_{t=1}^n e_t^*(\theta)^2$ où

$$e_t^*(\theta) = \hat{e}_t(\theta^*) + \left(\frac{\delta}{\delta \theta} \hat{e}_t \right) (\theta^*) (\theta - \theta^*) .$$

On montre qu'on peut prendre comme θ^* les $\hat{a}_j(p, q, m)$ précédents et les estimateurs de b_j qui en découlent. Cette méthode diffère de la méthode d'Hannan et Kavalieris par le fait qu'il n'y a pas de choix de l'ordre à l'étape 2. On se sert uniquement d'un estimateur fourni par cette étape pour construire le critère.

5. CONCLUSION

Beaucoup de méthodes pour le choix de l'ordre du modèle ARMA ont été proposées, ce qui témoigne l'intérêt du problème. La littérature dans ce domaine est très vaste et continue de croître. Cet article ne prétend pas d'être exhaustif. Une comparaison

systematique de la performance des différentes méthodes est souhaitable mais une telle étude ne semble pas avoir été faite. En ce qui concerne les méthodes interactives, la comparaison est difficile vu le fait qu'il y a un élément subjectif dans le choix retenu. Une autre difficulté est l'absence d'un bon critère pour cette comparaison. La considération de la fréquence où le procédé a choisi l'ordre correct est insuffisante, car le procédé peut choisir un ordre incorrect mais le modèle correspondant reste très proche du vrai modèle, ce qui est moins grave que si ce choix aboutit à un modèle complètement différent. En fait, si le vrai modèle est bien distinct des autres, la plupart des procédés de choix permettent de l'identifier. C'est seulement dans le cas où il existe des modèles alternatifs assez proches de celui-ci que l'identification est difficile, mais dans ce cas, choisir un modèle alternatif en question n'est pas très grave. En pratique, le procédé d'identification de modèles est adopté souvent selon le goût du praticien ou par simplicité (l'existence d'un logiciel qui implante ce procédé, par ex.). La méthode empirique simple suivante est aussi couramment employée et donne des résultats satisfaisants dans beaucoup de cas. Elle consiste à ajuster des modèles AR d'ordre 1, 2, 3 et 4 pour les données trimestrielles ou 12 pour les données mensuelles (pour tenir compte de la saisonnalité), puis examiner les autocorrélations empiriques des résidus. Si pour un modèle AR (p), il y a "troncation" de celle-ci au rang q , (c'est-à-dire la série des résidus semble provenir d'un modèle MA (q), on adoptera le modèle ARMA (p, q). En ce qui concerne les méthodes automatiques, il s'avère que le critère d'Akaike a tendance à surestimer l'ordre et de plus l'ordre choisi a une plus grande dispersion que celui obtenu par les critères d'Hannan ou BIC. Cependant ceci ne s'applique qu'au cas où le modèle sous-jacent est bien un modèle ARMA d'un certain ordre. Dans le cas où le modèle ARMA n'est qu'une approximation du vrai modèle et il s'agit de déterminer l'ordre optimal pour la meilleure approximation, le critère d'Akaike est plus adapté que les autres. L'adoption d'un critère dépend donc du contexte de son utilisation.

REMERCIEMENTS

Nous remercions le rapporteur anonyme pour ses critiques constructives et pour avoir attiré notre attention sur certaines références.

REFERENCES

- Akaike, H.** (1969) Fitting autoregressive model for prediction. *Ann. Inst. Statist. Math.*, 21, pp. 243-247.
- Akaike, H.** (1970) Statistical predictor identification. *Ann. Inst. Statist. Math.*, 22, pp. 203-217.
- Akaike, H.** (1973) Information theory and an extension of the maximum likelihood principle. Dans *2nd Int. Symposium on Information Theory*, pp. 267-281. Eds B.N. Pretop et F. Csaki, Académia Budapest Kiado, Budapest.
- Akaike, H.** (1978) A Bayesian analysis for the minimum AIC procedure. *Ann. Inst. Statist. Math.*, 30, pp. 9-14.
- Akaike, H.** (1979) A Bayesian extension extension of the minimum AIC procedure for autoregressive model fitting. *Biometrika*, 66, pp. 237-242.
- Ansley, C. S.** (1979) An algorithm for the exact likelihood of a mixed autoregressive-moving average process, *Biometrika*, 66, pp. 59-65.
- Atkinson, A.C.** (1980) A note on the generalized information criterion for choice of a model. *Biometrika*, 67, pp. 413-418.
- Béguin, J.M., Gourieroux, C., Montfort, A.** (1980) Identification of a mixed autoregressive moving average process : the corner method. *Time Series*, pp. 423-426, ed. O.D. Andersen, Amsterdam : North-Holland.
- Bednard, J.B., Roberts, B.** (1985) On the relationship between Levinson recursion and the R and S array for ARMA model identification. *Comm. Statist. Theory and Methods*, 14, pp.1217-1248.
- Berlinet, A.** (1984) Sur quelques problèmes d'estimation fonctionnelle et de statistique de processus. Thèse doctorat, Université de Lille.
- Berlinet, A.** (1985) Sequence transformations as a statistical tools. *Applied Numerical Math.*, 1, pp. 531-544. North-Holland.

Bhansali, R.J. (1980) Autoregressive and windowed estimates of the inverse autocorrelation. *Biometrika*, 67, pp. 551-566.

Bhansali, R.J. (1983) The inverse partial autocorrelation function of a time series and its applications. *J. Mult. Anal.*, 13, pp. 310-327.

Bhansali, R.J., Downham, D.Y. (1977) Some properties of the order of an autoregressive model selected by a generalization of Akaike's FPE criterion. *Biometrika*, 64, pp. 547-551.

Box, G.P.E., Jenkins, G.M. (1970) *Time Series Analysis, Forecasting and Control*. San Francisco : Holden Day.

Chaverie, P., Szpiro, D., Topol, R. (1990) Identification de modèle à fonction de transfert : la méthode Padé transformée en z. *Ann. d'Economie et Statistique*, 17, pp.145-161.

Chatfield, C. (1979) Inverse autocorrelations. *J. Roy. Statist. Soc.*, A, 142, 3, pp. 363-377.

Cleveland, W.S. (1972) The inverse autocorrelations of a time series and their application. *Technometrics*, 14, pp. 277-293.

Chow, J.C. (1972) On estimating the order of an autoregressive moving average process with uncertain observation. *IEEE Trans. Automat. Control*, AC-17, pp. 707-709.

De Gooijer, J.G., Abraham, B., Gould, A., Robinson, L. (1985) Methods for determining the order of an autoregressive moving average process : a survey. *Inst. Stat. Rev.*, 53, 3, pp. 301-329.

Findley, D.F. (1985) On the unbiasedness property of AIC for exact or approximating linear stochastic time series models. *J. Time Series Anal.*, §, pp. 229-252.

Franke, J. (1985) A Levinson-Durbin recursion for autoregressive moving average process. *Biometrika*, 72, pp. 573-581.

Franke, J., Gasser, Th., Steinberg, H. (1985) Fitting autoregressive processes to EEG time series : an empirical comparison of estimates of the order. *IEEE Trans., ASSP-33*, pp. 1115-1193.

Glasbey, C.A. (1982) A generalization of partial autocorrelation useful in identification of ARMA models. *Technometrics*, 24, pp. 223-228.

Graupe, D., Krause, D.J., Moore, J.B. (1975) Identification of autoregressive moving average parameters of time series. *IEEE Trans. Automat. Control*, AC-20, pp. 104-107.

Gray, H.L., Kelley, G.D., McIntyre, D.D. (1978) A new approach to ARMA modelling. *Comm. Statist. Simul. Comp.*, B7, pp. 1-77.

Gray, H.L., Houston, A.G., Morgan, F.W. (1978) On G-spectral estimation. *Applied Time Series Analysis*, pp. 39-138, D.F. Findley ed. New York : Academic Press.

Hannan, E.J. (1980) The estimation of the order of an ARMA process, *Ann. Statist.*, 8, pp. 1071-1081.

Hannan, E.J. (1982) Testing for autocorrelation and Akaike's criterion. *Essays in Statistical Science. Papers in Honor of P.A. Moran*, *Appl. Prob. Trust*, pp. 403-412.

Hannan, E.J. (1987) Rational transfer function approximation (with comments). *Statistical Science*, 2, pp. 135-161.

Hannan, E.J., Deistler, M. (1988) *The statistical theory of linear system*. New York : Wiley.

Hannan, E.J., Kavalieris, L. (1984) A method for autoregressive moving average estimation. *Biometrika*, 71, 2, pp. 273-280.

Hannan, E.J., Kavalieris, L. (1984) Multivariate linear time series models. *Adv. Appl. Prob.*, 16, pp. 492-561.

- Hannan, E.J., Kavalieris, L., MacKinsack, M.** (1986) Recursive estimation of linear systems. *Biometrika*, 73, pp. 119-134.
- Hannan, E.J., Quinn, B.G.** (1979) The determination of the order of an autoregression, *J. Roy. Statist. Soc., B*, 41, pp. 190-195.
- Hannan, E.J., Rissanen, J.** (1982) Recursive estimation of mixed autoregressive moving average order. *Biometrika*, 69, pp. 81-94.
- Jones, R.H.** (1975) Fitting autoregressions. *J. Amer. Statist. Assoc.*, 70, pp. 590-592.
- Mareschal, B., Mélard, G.** (1988) The corner method for identifying autoregressive moving average model. *Appl. Statist.*, 37, pp. 301-316.
- McClave, J.T.** (1978) Estimating the order of autoregressive model, the max χ^2 method. *J. Amer. Statist. Ass.*, 73, pp. 122-128.
- Pham, D.T.** (1983) A survey of time series analysis through parametric models. *Math. Operationsforschung. Statistik, series Statistics*, 14, 4, pp. 603-631.
- Pham, D.T.** (1984) A note on some statistics useful in identifying the order of autoregressive moving average model. *J. Time Series Anal.*, 5, pp. 273-279.
- Pham, D.T.** (1986) Parameter estimation and order selection for autoregressive moving average models. Rapport Recherche n° 614 M, Laboratoire TIM3, Institut IMAG, Grenoble.
- Pham, D.T.** (1988) Estimation of autoregressive parameters and order selection for ARMA models. *J. Time Series Anal.*, 9, 3, pp. 265-269.
- Parzen, E.** (1974) Some recent advances in time series modelling. *IEEE Trans. Automatic Control*, AC-19, pp. 723-729.
- Parzen, E.** (1977) Multiple time series modelling : determining the order of approximating autoregressive schemes. *Multivariate Analysis, IV*, pp. 283-295, ed. Krishnaiah. Amsterdam : North-Holland.

Pötscher, B.M. (1983) Order estimation in ARMA models by Lagrangian multiplier tests. *Annal. Statist.*, 11, pp. 872-885.

Rissanen, J. (1978) Modelling by shortest data description. *Automatica*, 14, pp. 465-471.

Rissanen, J. (1983) A universal prior for integer and estimation by minimum description length. *Ann. Statist.*, 11, pp. 416-431.

Rissanen, J. (1986) Stochastic complexity and modelling. *Ann. Statist.*, 14, pp. 1080-1100.

Rissanen, J. (1987) Stochastic complexity (with discussion). *J. Roy. Statist.*, B 49, pp. 223-239.

Schwarz, G. (1978) Estimating the dimension of a model. *Ann. Statist.*, 6, pp. 461-464.

Shibata, R. (1976) Selection of the order of autoregressive model by Akaike's information criterion. *Biometrika*, 63, pp. 117-126.

Shibata, R. (1980) Asymptotically efficient selection of the order of the for estimation parameters of a linear process. *Ann. Statist.*, 8, pp. 147-244.

Shibata, R. (1981) An optimal autoregressive spectral estimate. *Ann. Statist.*, 9, pp. 300-306.

Shibata, R. (1985) Various models selections techniques in time series analysis. *Handbook of Statistics*, 5, pp. 179-187. E.J. Hannan, P.R. Krishnaiah et M.M. Rao eds. Amsterdam : North-Holland.

Tsay, R.S., Tiao, G.C. (1984) Consistent estimate of autoregressive parameters and extended sample autocorrelation function for stationary and non stationary ARMA models. *J. Amer. Statist. Ass.*, 79, pp. 84-86.

Shanks, D. (1955) Nonlinear transformation of divergent or slowly convergent series. *J. Math. Phys.*, 34, pp. 1-42.

Shimzu, R. (1978) Entropy maximization principle and selection of the order of an autoregressive gaussian process. *Ann. Inst. Statist. Math.*, 30A, pp. 263-270.

Taniquichi, M. (1980) On selection of the spectral density model for a stationary process. *Ann. Inst. Statist. Math.*, 32A, pp. 401-419.

Veres, S. (1987) Asymptotic distribution of likelihood ratio for overparametrised ARMA process. *J. Time Series Anal.*, 8, 3, pp. 145-157.

Woodside, C.M. (1971) Estimation of the order of linear systems. *Automatica*, 7, pp. 727-733.

Woodward, W.A., Gray, H.L. (1981) On the relationship between the S array and the Box-Jenkins methods of ARMA model identification. *J. Amer. Statist. Assoc.*, 76, pp. 579-587.

Wallace, C.S., Freeman, P.R. (1987) Estimation and inference by compact coding (with discussion). *J. Roy. Statist. Soc.*, B 49, pp. 240-265.

Wynn, P. (1956) On a device to compute the e_m (S_m) transformation. *Math. Tables and Aids to Com.*, 10, pp. 91-96.