REVUE DE STATISTIQUE APPLIQUÉE

J.-P. INDJEHAGOPIAN

Les processus stochastiques vectoriels ARMA : une procédure d'identification

Revue de statistique appliquée, tome 27, nº 3 (1979), p. 33-45

<http://www.numdam.org/item?id=RSA_1979__27_3_33_0>

LES PROCESSUS STOCHASTIQUES VECTORIELS ARMA : UNE PROCÉDURE D'IDENTIFICATION

J.-P. INDJEHAGOPIAN*

RESUME

Cet article considère une classe particulière de processus stochastiques vectoriels stationnaires : les modèles autorégressifs - moyennes mobiles ARMA multivariés et présente une procédure pour modéliser les séries temporelles multidimensionnelles à l'aide de cette classe de modèles. Cette procédure reprend le schéma itératif bien connu de Box et Jenkins : identifier (spécifier) - estimer - contrôler en proposant une solution au problème difficile de l'identification (spécification) du modèle.

1 – INTRODUCTION

Depuis quelques années, dans le domaine de la prévision économique (au sens le plus large), la méthodologie de Box et Jenkins [3] tend à être de plus en plus utilisée au détriment de l'approche économétrique. En effet, dans un certain nombre de domaines de l'économie, il n'existe pas encore de théories économiques satisfaisantes qui rendent parfaitement compte des faits observés (par exemple l'inflation, la monnaie, le chômage, le comportement des consommateurs, etc.). L'état actuel de ces théories conduit généralement à des modèles économétriques empiriques présentant des erreurs de spécification et/ou des autocorrélations dans les séries résiduelles etc. Les prévisions faites à partir de ces modèles ne sont plus optimales vis-à-vis du critère, généralement retenu, du minimum de l'espérance du carré de l'erreur de prévision. En ce sens, des études comparatives ont montré que les modèles non stationnaires ARIMA de Box et Jenkins donnaient de meilleures prédictions.

La procédure proposée par Box et Jenkins pour modéliser une série unidimensionnelle est cependant critiquable dans la mesure où elle n'intègre pas les effets croisés entre la série étudiée et d'autres séries économiques. On peut alors se demander si les prévisions sur chacune des séries pourraient être améliorées si on tenait compte des effets croisés (unidirectionnels ou avec rétroaction). Pour juger de l'amélioration en terme de prévision, il est alors nécessaire d'élaborer un modèle stochastique multivarié à partir de la connaissance de l'historique de la chronique multidimensionnelle.

(*) Ecole Supérieure des Sciences Economiques et Commerciales BP 105 - 95021 Cergy Pontoise Cedex.

Mots clefs : **Analyse des séries temporelles, Box-Jenkins, prédiction.**

L'objet de cet article est de présenter une méthode d'identification (spécification) de modèles ARMA multivariés qui furent introduits pour la première fois par Quenouille [18]. Cet article comporte trois parties. La première partie présente quelques concepts de base sur les processus faiblement stationnaires et des résultats récents sur la matrice de covariance et la factorisation canonique d'un processus ARMA. La partie suivante aborde les propriétés des composantes d'un vecteur aléatoire évoluant suivant un processus vectoriel ARMA et enfin la troisième partie présente une procédure de construction de modèles ARMA.

2 - PROCESSUS STOCHASTIQUES MULTIVARIES STATIONNAIRES

Soit $\{\underset{\sim}{z}_t \,;\, t \in Z\}$ un processus stochastique temporel discret tel que la variable aléatoire vectorielle $\underset{\sim}{z}_t$ comporte m composantes $z_{1t}, \ldots, z_{mt}$. Ce processus vectoriel est dit faiblement stationnaire (ou stationnaire au deuxième ordre) si chaque composante z_{it} est faiblement stationnaire (i.e les moments d'ordre un et deux existent et sont invariants au cours du temps) et si de plus la covariance pour tout couple de composantes est stationnaire. Ces définitions peuvent être résumées comme suit :

$$\text{(i)} \quad E(z_{it}) = \mu_i \qquad 1 \leqslant i \leqslant m$$

$$\text{(ii)} \quad E[(z_{it} - \mu_i)(z_{jt+k} - \mu_j)] = \gamma_{ij}(k) \qquad 1 \leqslant i, j \leqslant m \,;\, k \in Z$$

Pour la suite de l'exposé, on définit la matrice de covariance Γ_k au décalage k du processus faiblement stationnaire $\underset{\sim}{z}_t$ par

$$\Gamma_k = \{\gamma_{ij}(k)\} = E[(\underset{\sim}{z}_t - \underset{\sim}{\mu})(\underset{\sim}{z}_{t+k} - \underset{\sim}{\mu})']$$

où $\qquad \underset{\sim}{\mu}' = (\mu_1, \ldots, \mu_m)$

De même on définit la matrice des coefficients de corrélation croisée P_k et la fonction génératrice de covariance $\Gamma(B)$ du processus faiblement stationnaire, respectivement par :

$$P_k = \{\rho_{ij}(k)\} \quad \text{avec} \quad \rho_{ij}(k) = \frac{\gamma_{ij}(k)}{\{\gamma_{ii}(0)\,\gamma_{jj}(0)\}^{1/2}}$$

$$\Gamma(B) = \sum_{k=-\infty}^{+\infty} \Gamma_k B^k$$

où B représente l'opérateur de décalage $B\underset{\sim}{z}_t = \underset{\sim}{z}_{t-1}$, $B^k \underset{\sim}{z}_t = \underset{\sim}{z}_{t-k}$.

ROZANOV [19] et HANNAN [8] ont montré à partir de l'analyse spectrale qu'il était possible de généraliser le théorème de décomposition de WOLD [23] pour un processus $\underset{\sim}{z}_t$ centré stationnaire (du deuxième ordre). Cette décomposition appelée factorisation canonique se présente sous la forme :

$$(2.1.) \qquad \underset{\sim}{z}_t = \Psi(B)\underset{\sim}{a}_t$$

où :

(i) $\Psi(B) = \sum_{k=0}^{\infty} \Psi_k B^k$ est une matrice dont les éléments sont des fonctions rationnelles de B et où $\{\Psi_k\}$ est une suite de matrices de dimension (m,m)

(ii) $\{\underset{\sim}{a}_t\}$ est un processus vectoriel bruit blanc satisfaisant à $E(\underset{\sim}{a}_t \underset{\sim}{a}'_{t+k}) = \delta_0^k I$ avec I matrice unité de dimension (m,m)

(iii) les dénominateurs, polynomiaux en B, de chaque élément de $\Psi(B)$ ont tous leurs poles situés à l'extérieur du disque unité et les racines de det $\Psi(B) = 0$ ne sont pas situées à l'intérieur du disque unité.

La décomposition (2.1.) n'est cependant pas unique. Il est en effet possible de trouver d'autres factorisations canoniques du processus faiblement stationnaire $\underset{\sim}{z}_t$ qui conduisent à la même fonction génératrice de covariance Γ. Ces représentations se distinguent par la structure probabiliste des innovations $\underset{\sim}{a}_t$. (Cf. HAUGH et BOX [10], HANNAN [8], INDJEHAGOPIAN [12] et [13], LEDOLTER [15].

HANNAN [8] et ROZANOV [19] ont montré que si les éléments d'une représentation $\Psi(B)$ sont des fonctions rationnelles de B, il est alors possible de factoriser Ψ sous la forme (principe de parcimonie) :

$$\Psi(B) = [\Phi(B)]^{-1}\,\Theta(B)$$

où $\Phi(B)$ et $\Theta(B)$ sont respectivement des polynomes de degrés p et q :

$$\Phi(B) = I - \sum_{k=1}^{p} \Phi_k B^k$$

$$\Theta(B) = I - \sum_{k=1}^{q} \Theta_k B^k$$

Si de plus, on suppose que les racines de det $\Phi(B) = 0$ sont extérieures au disque unité (condition de stationnarité du processus linéaire) et que les racines de det $\Theta(B) = 0$ sont à l'extérieur du disque unité (condition d'inversibilité) et que de plus ces deux équations n'ont aucune racine en commun, alors le processus $\underset{\sim}{z}_t$ peut s'écrire :

$$\Phi(B)\underset{\sim}{z}_t = \Theta(B)\underset{\sim}{a}_t$$

soit

(2.2) $$\underset{\sim}{z}_t - \Phi_1 \underset{\sim}{z}_{t-1} - \dots - \Phi_p \underset{\sim}{z}_{t-p} = \underset{\sim}{a}_t - \Theta_1 \underset{\sim}{a}_{t-1} - \dots - \Theta_{t-q} \underset{\sim}{a}_{t-q}$$

avec : $$E(\underset{\sim}{a}_t \underset{\sim}{a}'_{t+k}) = \delta_0^k\,\Omega$$

où Ω est une matrice (m,m) symétrique définie positive et δ_0^k le symbole de KRONECKER.

La représentation (2.2.) est appelée modèle ARMA multivarié d'ordre p et q avec les composantes du bruit blanc (vectoriel) $\underset{\sim}{a}_t$ corrélées deux à deux aux mêmes époques i.e. :

$$E(a_{it}\, a_{jt+k}) = \delta_0^k\, \omega_{ij} \qquad 1 \leqslant i,j \leqslant m \;;\; k \in Z$$

Cette représentation (2.2.) fut introduite pour la première fois par QUENOUILLE [18]. On peut montrer qu'il est toujours possible de trouver une version de (2.2) telle que la matrice Ω de covariance du processus bruit blanc soit diagonale ou réduite à la matrice unité (cf. HAUGH et BOX [10] et INDJEHAGOPIAN [12]).

Cas particulier de modèles ARMA multivariés

1) Modèle autorégressif multivarié (centré) d'ordre p :

$$(I - \Phi_1 B - \cdots - \Phi_p B^p) \underset{\sim}{z}_t = \underset{\sim}{a}_t \tag{2.3}$$

où le processus $\{\underset{\sim}{a}_t\}$ est tel que :

$$E(\underset{\sim}{a}_t \underset{\sim}{a}'_{t+k}) = \delta_0^k \, \Omega$$

A partir de l'équation (2.3.), on démontre facilement que les matrices de covariance satisfont à la relation :

$$\Gamma_k - \Gamma_{k-1} \Phi'_1 - \cdots - \Gamma_{k-p} \Phi'_p = 0 \quad k \geqslant 1 \tag{2.4}$$

et

$$\Gamma_0 - \Phi_1 \Gamma_1 - \cdots - \Phi_p \Gamma_p = \Omega \tag{2.5}$$

En remplaçant dans (2.4) le décalage k par $1, 2, \ldots p$ et en tenant compte de la relation $\Gamma_k = \Gamma'_{-k}$, on peut exprimer les coefficients matriciels Φ_j à partir de combinaisons linéaires des covariances $\Gamma_0, \ldots, \Gamma_p$. Ces relations linéaires généralisent les équations de Yules-Walker pour un modèle autorégressif univarié (cf. BOX et JENKINS [3]).

2) Modèle moyenne mobile multivarié (centré) d'ordre q :

$$\underset{\sim}{z}_t = (I - \Theta_1 B - \cdots - \Theta_q B^q) \underset{\sim}{a}_t \tag{2.6}$$

où le processus innovation $\underset{\sim}{a}_t$ satisfait à :

$$E(\underset{\sim}{a}_t \underset{\sim}{a}'_{t+k}) = \delta_0^k \, \Omega$$

En utilisant (2.6), la matrice de covariance d'un MA(q) s'écrit :

$$\Gamma_k = E(\underset{\sim}{z}_t \underset{\sim}{z}'_{t+k}) = E[\Theta(B) \underset{\sim}{a}_t \underset{\sim}{a}_{t+k} \Theta'(B)]$$

$$\Gamma_k = \begin{cases} \sum_{i=0}^{q} \Theta_i \Omega \Theta'_i & \Theta_0 = I \; ; \; k = 0 \\ -\Omega \Theta'_k + \sum_{i=1}^{q-k} \Theta_i \Omega \Theta'_{i+k} & 1 \leqslant k \leqslant q \\ 0 & k \geqslant q + 1 \end{cases} \tag{2.7}$$

Cette relation généralise celle trouvée par Box et Jenkins [3] dans le cas d'un MA(q) univarié.

On peut montrer qu'il y a 2^q modèles MA(q) possibles qui satisfont à (2.7) (cf. HANNAN [7]). Aussi dans une perspective de prédiction vers le futur on sélectionne le modèle MA(q) qui vérifie la condition d'inversibilité (i.e le polynome $\det \Theta(B) = 0$ n'a pas de racines à l'intérieur du disque unité).

3) Modèle autorégressif - moyenne mobile (centré) d'ordre p et q

On démontre aussi qu'un processus $\underset{\sim}{z}_t$ solution de l'équation :

$$I - \Phi_1 z_{t-1} - \cdots - \Phi_p z_{t-p} = I - \Theta_1 a_{t-1} - \ldots - \Theta_q a_{t-q}$$

avec

$$E(\underset{\sim}{a}_t \underset{\sim}{a}'_{t+k}) = \delta_0^k \Omega$$

a ses matrices de covariance qui, à partir du décalage q + 1, vérifie la propriété

$$\Gamma_k - \Gamma_{k-1}\Phi_1' - \cdots - \Gamma_{k-p}\Phi_p' = 0 \qquad k \geqslant q+1 \tag{2.8}$$

Cette relation est identique à celle du modèle ARMA (p , q) univarié.

4) Modèle multivarié saisonnier

Les séries économiques multidimensionnelles que l'on est amené à analyser présentent généralement un comportement saisonnier. Il est donc nécessaire d'étendre la classe des modèles ARMA multivariés. afin de pouvoir décrire ce type de séries. HILLMER [11] propose de généraliser le modèle saisonnier multiplicatif de BOX et JENKINS [3] au cas d'un processus m-vectoriel $\underset{\sim}{z}_t$ centré et faiblement stationnaire, de périodicité s, i.e.,

$$\Phi(B)\,\Phi_s(B^s)\underset{\sim}{z}_t = \Theta(B)\,\Theta_s(B^s)\underset{\sim}{a}_t$$

où les polynomes $\Phi(B)$ et $\Theta(B)$ sont définis comme en (2.2) et où :

$$\Phi_s(B^s) = I - \Phi_{1,s} B^s - \cdots - \Phi_{P,s} B^{Ps}$$

$$\Theta_s(B^s) = I - \Theta_{1,s} B^s - \cdots - \Theta_{Q,s} B^{Qs}$$

avec $\Phi_{j,s}$ $(1 \leqslant j \leqslant P)$ et $\Theta_{\ell,s}$ $(1 \leqslant \ell \leqslant Q)$ des matrices de dimension (m,m).

On suppose en outre que les racines de $\det \Phi_s(B^s) = 0$ sont à l'extérieur du disque unité (stationnarité de l'autorégressif saisonnier), que les racines de $\det \Theta_s(B^s) = 0$ ne sont pas à l'intérieur du disque unité (condition d'inversibilité du moyenne mobile saisonnier) et que de plus, ces deux équations n'ont aucune racine en commun.

Ces modèles saisonniers peuvent être traités comme des modèles ARMA multivariés non saisonniers d'ordres plus élevés dont les paramètres matriciels, compte tenu des conditions énoncées ci-dessus, satisfont un certain nombre de contraintes.

3 – MODELES SUIVIS PAR LES COMPOSANTES D'UN PROCESSUS VECTORIEL APPARTENANT A LA CLASSE DES ARMA

Dans l'analyse des séries temporelles, un problème important est celui de l'identification (spécification) du modèle dans la classe des modèles linéaires ARMA. Pour tenter de résoudre ce problème difficile, BOX et JENKINS [3] ont proposé, dans le cas des séries temporelles univariées, d'utiliser la fonction d'autocorrélation estimée et la fonction d'autocorrélation partielle estimée. L'extension d'une telle procédure au cas des séries temporelles multidimensionnelles est actuellement difficile car on ne sait pas très bien analyser la structure des matrices de corrélation d'un processus ARMA multivarié.

Pour palier à cet inconvénient nous proposerons au paragraphe suivant une méthode d'identification (spécification) qui nécessite l'analyse de chacune des composantes de la chronique multidimensionnelle. A cet effet, il est nécessaire de trouver les modèles suivis par les composantes d'une variable aléatoire vectorielle $\underset{\sim}{z}_t$ satisfaisant à l'équation d'un ARMA multivarié.

Pour cela considérons le cas d'un modèle autorégressif AR(p) dont le processus m-vectoriel $\underset{\sim}{z}_t$ soit solution de :

$$\phi(B)\underset{\sim}{z}_t = (I - \phi_1 B - \cdots - \phi_1 B^p)\underset{\sim}{z}_t = \underset{\sim}{a}_t$$

ou encore, en décomposant la matrice $\phi(B)$ en bloc de matrices :

$$\begin{bmatrix} H_1(B) & H_2'(B) \\ & \\ H_3(B) & \Phi_{m-1}(B) \end{bmatrix} \begin{bmatrix} z_{1t} \\ \cdot \\ \cdot \\ \cdot \\ z_{mt} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ \cdot \\ \cdot \\ \cdot \\ a_{mt} \end{bmatrix} \tag{3.1}$$

avec :

$H_1(B)$ un opérateur scalaire, $H_2(B)$ et $H_3(B)$ des matrices de dimension $(m-1, 1)$ et $\Phi_{m-1}(B)$ l'opérateur matriciel $(m-1, m-1)$.

A l'aide de ces blocs et si $\Phi_{m-1}(B)$ est inversible, on obtient :

$$[H_1(B) - H_2'(B)\,\Phi_{m-1}^{-1}(B)\,H_3(B)]\,z_{1t} = a_{1t} - H_2'(B)\,\Phi_{m-1}^{-1}(B)\,\underset{\sim}{a}_t^{1'} \tag{3.2}$$

avec :

$$\underset{\sim}{a}_t^{1'} = (a_{2t}, \ldots, a_{mt})$$

Pour exprimer (3.2) sous forme de polynomes en B appliqués à la composante z_{1t} et aux composantes du bruit a_t, multiplions (3.2) par le déterminant de $\Phi_{m-1}(B)$, soit :

$$\begin{aligned} &[\det \Phi_{m-1}(B) . H_1(B) - H_2'(B)\,\Phi_{m-1}^{*}(B)\,H_3(B)]\,z_{1t} = \\ &\qquad = \det \Phi_{m-1}(B) . a_{1t} - H_2'(B)\,\Phi_{m-1}(B)\,\underset{\sim}{a}_t^{1'} \end{aligned} \tag{3.3}$$

où $\Phi_{m-1}^{*}(B)$ est la matrice adjointe de $\Phi_{m-1}(B)$

A partir de (3.3) on peut conclure que la composante z_{1t}, issue d'un processus multivarié AR(p), suit en général un processus autorégressif avec bruit corrélé (il en serait de même des autres composantes). Si les matrices $H_2(B)$ et

$H_1(B)$ n'ont pas de facteurs communs avec la matrice $\Phi_{m-1}(B)$, il s'ensuit que la composante z_{1t} suit un processus ARMA dont l'ordre de l'autorégressif est mp (il en serait de même des autres composantes du processus).

A titre d'exemple, considérons tout d'abord le modèle autorégressif bivarié :

$$(I - \phi B)\underset{\sim}{z}_t = \underset{\sim}{a}_t$$

(3.4) $$\begin{bmatrix} 1-\phi_{11}B & -\phi_{12}B \\ -\phi_{21}B & 1-\phi_{22}B \end{bmatrix} \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

La résolution du système matriciel conduit à :

(3.5) $$[1-(\phi_{11}+\phi_{22})B-(\phi_{12}\phi_{21}-\phi_{11}\phi_{22})B^2]\,z_{1t} = (1-\phi_{22}B)\,a_{1t}+\phi_{12}\,a_{2t-1}$$

Il apparait ainsi que la composante z_{1t} suit un processus autorégressif en général d'ordre 2 avec des erreurs corrélées (si a_{1t} et a_{2t} sont corrélés).

Si l'on se place dans l'hypothèse où le coefficient ϕ_{12} est nul (pas de rétroaction ou "feedback" de z_2 sur z_1), alors le modèle (3.5) après simplification par $1-\phi_{22}B$ s'écrit sous la forme d'un AR(1) :

(3.6) $$(1-\phi_{11}B)\,z_{1t} = a_{1t}$$

En développant des calculs similaires sur la composante z_{2t}, on montre aisément en conservant l'hypothèse de *non rétroaction* de z_2 sur z_1, que l'on retombe sur le modèle de BOX et JENKINS avec fonction de transfert v(B) :

$$z_{2t} = v(B)\,z_{1t} + \Psi(B)\epsilon_t$$

où v(B) s'exprime comme le rapport de deux polynomes de degré un et où la variable exogène z_{1t} est non corrélée au bruit blanc ϵ_{t+k} pour tout k.

Considérons maintenant un modèle moyenne mobile MA(q) multivarié. La relation (2.7) montre que la matrice de covariance Γ_k est nulle dès que le décalage k est supérieur à l'ordre q du MA multivarié. Cette propriété implique que l'autocovariance de chaque composante de $\underset{\sim}{z}_t$ est nulle dès que le décalage k est supérieur à q. Chaque composante suit donc un modèle MA univarié d'ordre au plus égal à l'ordre du modèle MA multivarié.

4 – PROCEDURES DE CONSTRUCTION D'UN MODELE ARMA

Une approche similaire à la procédure itérative proposée par BOX et JENKINS [3] peut être utilisée pour la construction d'un modèle stochastique multivarié. Cette méthodologie comprend les phases suivantes :

1) Considérer une classe de modèles stochastiques linéaires suffisamment générale pour couvrir les situations variées et identifier (spécifier) une sous classe de modèles compatibles avec les données multiples ;

2) Estimer les paramètres inconnus du modèle multivarié dans la sous classe retenue ;

3) Contrôler la validité du modèle, itérer la procédure si les tests statistiques sont négatifs.

a) Spécification

Cette première phase consiste à sélectionner les ordres p et q du processus ARMA multivarié en supposant que la série multidimensionnelle observée est une réalisation de ce processus. Au préalable, il convient de "stationnariser" chacune des séries composantes observées. A cet effet, on utilise les opérateurs différences $(1 - B)^d$ et/ou les différences saisonnières $(1 - B^s)^D$ si s est la périodicité cf. BOX et JENKINS [3] et plus généralement, les transformations de BOX et COX [2] ou de ANSLEY [1].

Les outils statistiques utilisés pour cette phase d'identification sous principalement la fonction matricielle de covariance estimée $k \to C_k$ et la fonction matricielle de corrélation croisée estimée $k \to R_k$, données respectivement par :

$$C_k = \{C_{ij}(k)\} \quad \text{avec} \quad C_{ij}(k) = \begin{cases} \dfrac{1}{n} \displaystyle\sum_{t=1}^{n-k} (z_{it} - \bar{z}_i)(z_{jt+k} - \bar{z}_j) & ; \; k \geqslant 0 \\ \dfrac{1}{n} \displaystyle\sum_{t=1-k}^{n} (z_{it} - \bar{z}_i)(z_{jt+k} - \bar{z}_j) & ; \; k < 0 \end{cases}$$

où $\bar{z}_i$ et $\bar{z}_j$ sont respectivement les moyennes des séries composantes stationnaires z_i et z_j $(1 \leqslant i, j \leqslant m)$

et $$R_k = \{r_{ij}(k)\} \quad \text{avec} \quad r_{ij}(k) = \frac{c_{ij}(k)}{\{c_{ii}(0)\, c_{jj}(0)\}^{1/2}}$$

Ces outils sont actuellement d'un emploi peu commode car il est difficile de déceler la structure théorique de la fonction matricielle des corrélations croisées à partir de son estimation.

HAUGH [9], HAUGH et BOX [10], PIERCE [17] utilisent, dans le cas de deux séries, une procédure à deux étapes. La première étape consiste à identifier (spécifier) pour chacune des séries, préalablement "stationnarisée", un modèle ARMA univarié, la deuxième, vise à identifier le modèle ARMA bivarié reliant les deux bruits blancs estimés (ou résidus) issus des modèles univariés estimés. Cette identification s'effectue à partir de la fonction de corrélation croisée estimée des résidus. En combinant les modèles issus des deux étapes, on en déduit la forme du modèle bivarié sur la série bidimensionnelle initiale.

GRANGER et NEWBOLD [6] adoptent une procédure similaire. ZELLNER et PALM [24] et WALLIS [20] utilisent le fait que les modèles ARMA présentés en (2.2) peuvent s'écrire :

$$\det \Phi(B) \underset{\sim}{z}_t = \Phi^*(B)\, \Theta(B) \underset{\sim}{a}_t$$

où $\Phi^*(B)$ est la matrice adjointe de $\Phi(B)$ et où les racines de $\det \Phi(B) = 0$ sont supposées être à l'extérieur du disque unité. Il s'ensuit que chacune des séries suit un modèle ARMA. Ces modèles ARMA présentent généralement le même ordre et les mêmes paramètres pour leur partie autorégressive.

Dans le cas particulier d'un effet unidirectionnel entre deux variables (pas de rétroaction) BOX et JENKINS [3] proposent d'utiliser un modèle linéaire dynamique (fonction de transfert).

L'approche qui est discutée ici utilise les informations issues de l'analyse univariée de chaque série composante. De plus, il est tenu compte des propriétés suivantes, énoncées dans la partie 3 :

(i) Si le modèle multivarie est un AR(p) alors chaque série composante suit un modèle ARMA $(\bar{p}, \bar{q})$ où $\bar{p} \geqslant p$;

(ii) Si le modèle multivarié est un MA(q) alors chaque série composante suit aussi un MA$(\underline{q})$ ou $\underline{q} \leqslant q$.

La spécification du modèle multivarié revient à prendre un modèle de même nature que les modèles univariés i.e. si chaque série composante suit un MA dont l'ordre maximum est q, le modèle multivarié aura pour structure un MA(q), si chaque série composante suit un ARMA et si p (respectivement q) représente l'ordre minimum (respectivement maximum) des parties autorégressives (respectivement moyennes mobiles), le modèle multivarié sera un ARMA (p, q).

b) Estimation

Après avoir spécifié la structure du modèle i.e. les opérations de "stationnarisation" et les ordres p, q du modèle ARMA (éventuellement les ordres P et Q du modèle saisonnier multiplicatif) on doit estimer les paramètres matriciels.

Les procédures d'estimation des paramètres matriciels du modèle ARMA multivarié lorsque les ordres p et q sont spécifiés, s'appuient généralement sur la méthode du maximum de vraisemblance dans le cas où le processus vectoriel bruit blanc est normal. WILSON [22] propose une procédure itérative qui permet d'approximer la fonction de vraisemblance des paramètres en prenant les valeurs de démarrage des $\underset{\sim}{a}_t$ nuls. Ceci suppose que les valeurs de démarrage ont un effet négligeable sur la fonction de vraisemblance lorsque le nombre d'observations sur la série multidimensionnelle $\underset{\sim}{z}_t$ est grand. Cette procédure généralise l'algorithme de Marquardt (cf. DRAPER et SMITH [5]). HILLMER [11] et OSBORN [16] utilisent une méthode similaire en relaxant la condition sur les valeurs de démarrage des $\underset{\sim}{a}_t$ et en calculant la fonction de vraisemblance exacte.

c) Contrôle-validation

Cette phase permet de tester l'adéquation du modèle estimé à la série multiple observés. Deux types de contrôle sont mis en place sur les résidus.

1) Tests à l'aide de la fonction d'autocorrélation et de la fonction de corrélation croisée

Si le modèle est correctement spécifié et si on connaît les vraies valeurs des paramètres du modéle, les bruits vectoriels $\underset{\sim}{a}_t$ sont indépendants, d'espérance nulle et de matrice de covariance Ω i.e., $E(\underset{\sim}{a}_t \underset{\sim}{a}'_{t+k}) = \delta_0^k \, \Omega$. On montre, cf. BOX et JENKINS [3] que les coefficients d'autocorrélation estimés $r_{ii}(k)$ de a_{it} $(1 \leqslant i \leqslant m)$ sont non corrélés et asymptotiquement indépendants et distribués suivant une loi normale de moyen zéro et de variance n^{-1} (où n est le nombre de données disponibles pour chaque série). Dans la réalité on estime le modèle et on observe les résidus $\hat{a}_t$, on est ainsi amené à calculer $\hat{r}_{ii}(k)$ à partir des résidus $\hat{a}_{it}$. On montre cependant que les $\{\hat{r}_{ii}(k)\}$ ont la même distribution asymptotique que les $\{r_{ii}(k)\}$. De plus, on montre que cette dernière propriété reste valable lorsque l'on utilise les coefficients de corrélation croisés estimés $r_{ij}(k)$ entre les deux séries indépendantes a_{it} et a_{jt} (cf. HAUGH [9], HAUGH et BOX [10]).

Ces propriétés statistiques permettent de tester pour chaque coefficient d'autocorrélation (de corrélation croisée) l'hypothèse nulle $\rho_{ii}(k) = 0$ (ou $\rho_{ij}(k) = 0$).

2) Test dit du "portmanteau"

Un autre test peut être utilisé pour contrôler l'indépendance de deux séries a_{it} et a_{jt} $(1 \leqslant i, j \leqslant m \; ; \; i \neq j)$. Ce test utilise la statistique $Q_K = n \sum_{k \in K} \hat{r}_{ij}^2(k)$ où K est un ensemble de décalages. Cette statistique, sous l'hypothèse nulle d'indépendance des deux séries, est asymptotiquement distribuée comme un χ^2 de degrés de liberté card K. En particulier, si $Q_M = n \sum_{k=-M}^{M} r_{ij}(k) \geqslant \chi_{2M+1}(\alpha)$ on rejettera l'hypothèse d'indépendance des deux séries, au seuil α où

$$\alpha = \text{Prob } \{\chi^2_{2M+1} \geqslant \chi^2_{2M+1}(\alpha)\}$$

A partir de la statistique Q_K HAUGH et BOX [10], PIERCE [17] montrent qu'il est possible de vérifier des schémas de causalité entre deux séries z_i et z_j.

d) Prévision à partir des modèles ARMA multivariés

Dans la classe des modèles univariés ARIMA (p, d, q), on peut montrer en utilisant la théorie générale de la prédiction développée principalement par KOLMOGOROV [14] et WIENER [21] que le prédicteur linéaire de l'observation future $z_{n+\ell}$ qui minimise l'espérance du carré de l'erreur de prévision, connaissant les valeurs passées $z_s, s \leqslant n$, est donnée par l'espérance conditionnelle

$$\hat{z}_n(\ell) = E(z_{n+\ell} \,|\, z_s, s \leqslant n)$$

Cette fonction linéaire des valeurs passées se calcule facilement à partir de

$$(1 - B)^d \, \phi(B) \, z_t = \theta(B) \, a_t \tag{4.1}$$

En effet, en supposant que le processus (4.1) est inversible, on peut développer en série entière $\pi(B) = (1 - B)^d \, \phi(B)/\theta(B) = 1 - \pi_1 B - \pi_2 B^2 - \ldots$ et

$$z_{n+\ell} = \sum_{j=1}^{\infty} \pi_j \, z_{n+\ell-j} + a_{n+\ell} \quad (\ell \geqslant 1) \tag{4.2}$$

A partir de (4.2) BOX et JENKINS [3] montrent que

$$\hat{z}_n(\ell) = \sum_{j=0}^{\infty} \pi_{j+1}^{(\ell)} \, z_{n-j} \quad \text{avec} \quad \pi_j^{(\ell)} = \pi_{j+\ell-1} + \sum_{k=1}^{\ell-1} \pi_k \, \pi_j^{(\ell-k)} \tag{4.3}$$

Dans le cas des processus m-vectoriels, la formule (4.3) peut être généralisée en remplaçant la suite des $\pi_j^{(\ell)}$ par la suite des matrices $\Pi_i^{(\ell)}$ de dimension (m, m). Cette généralisation correspond au prédicteur linéaire $\underset{\sim}{z}_n(\ell)$ optimal dans le sens où la matrice de covariance $V_n(\ell)$ du vecteur erreur de prévision $\underset{\sim}{e}_n(\ell) = \underset{\sim}{z}_{n+\ell} - \hat{\underset{\sim}{z}}_n(\ell)$ est la plus petite (1) cf. GRANGER et NEWBOLD [6]).

5 – CONCLUSION

Dans cet article, nous avons tout d'abord présenté la classe des processus stochastiques vectoriels stationnaires ARMA en insistant sur le fait que la factorisation canonique d'un processus, à partir de sa fonction génératrice de covariance, n'était pas unique. Nous avons donné aussi quelques exemples de processus ARMA en indiquent que l'équation aux différences sur la fonction d'autocovariance (équation de Yules WALKER) se généralisait au cas multidimensionnel. Malheureusement, l'état actuel des recherches dans ce domaine ne permet pas d'utiliser les relations de Yules WALKER généralisées pour identifier le processus vectoriel comme dans le cas unidimensionnel. Enfin, nous avons proposé une procédure d'identification (spécification) de modèles ARMA multivariés à partir, d'une part, de l'analyse de chacune des composantes de la série à l'aide de la démarche BOX et JENKINS et d'autre part, des résultats théoriques démontrés à la partie 3.

Quelques exemples d'application de l'analyse des séries temporelles multidimensionnelles (macroéconomie, macroéconomie financière, etc.) utilisant les procédures soit de HAUGH [9] soit de ZELLNER et PALM [24] ont été récemment publiés (cf. CRAMER et MILLER [4], HAUGH et BOX [10], WALLIS [20], ZELLNER et PALM [24]). Plusieurs applications pratiques, dans le domaine de la macroéconomie, sont en cours d'élaboration en utilisant la procédure de l'auteur.

REFERENCES

[1] ANSLEY C.F., SPIVEY W.A. and WROBLESKY W.J. – A class of transformations for BOX and JENKINS seasonal models. *Appl. Statist.* (1977), 26, n° 2, 173-177.

[2] BOX G.E.P. and COX D.R. – An analysis of transformation, *Journal of Royal, Stat. Soc., Series B*, 26 (1964), 211-252.

(1) Soient deux matrices V_1 et V_2 de dimension (m, m) définies positives, V_1 est dite plus petite que V_2 si $u'V_1u < u'V_2u$ pour tout vecteur non nul u de dimension (m, 1).

[3] BOX G.E.P. and JENKINS G.M. – *Time Series Analysis, Forecasting and Control*, Holden Day, San Francisco, 1976.

[4] CRAMER R. and MILLER R. – Dynamic modelling of multivariate time series for use in bank analysis, *Jour. of Money, Credit and Banking*, Vol. VIII, n° 1, Feb. 1976.

[5] DRAPER N.R. and SMITH H. – *Applied Regression Analysis*, Wiley, New York, 1966.

[6] GRANGER C.W.J., and NEWBOLD P. – *Forecasting Economic Time Series*, Academic Press, New York, 1977.

[7] HANNAN E.J. – *Time Series Analysis*, Methuen, Londres, 1960.

[8] HANNAN E.J. – *Multiple Time Series*, Wiley, New York 1970.

[9] HAUGH L.D. – *The Identification of Time Series Interrelationships with Special Reference to Dynamic Regression Models*, Ph. D. Thesis, University of Wisconsin, Madison, 1972.

[10] HAUGH L.D. and BOX G.E.P. – Identification of Dynamic Regression (Distributed Lag) Models Connecting Two Time Series, *JASA*, 72, 1977, 121-130.

[11] HILLMER S.C. – *Time Series : Estimation, Smoothing and Seasonal Adjusting*, Ph. D. Thesis, University of Wisconsin, Madison, 1976.

[12] INDJEHAGOPIAN J.P. – *Modeles ARMA multidimensionnels, une procédure de construction,* Document de Recherche 78.001, CERESSEC, Centre de Recherche de l'ESSEC, Mai 1978.

[13] INDJEHAGOPIAN J.P. – Une procédure de construction de modèles ARIMA multivariés. Application. *Actes du Colloque Structures Economiques et Econométrie, Départ, de Mathématiques, Université Claude BERNARD,* Lyon, Mai 1978, p. II-13 à II-28.

[14] KOLMOGOROV A.N. – Interpolation and Extrapolation of Stationarity Random Sequence, 1941, Translation *Rand Corp.* Santa Monica, Calif., RM - 2090 - PR.

[15] LEDOLTER J. – *A Multivariate Time Series Approach to Modelling Macro-Economic Sequences*, R.M. 77-33, International Institut for Applied Systems Analysis, Laxenburg, Asutria, 1977.

[16] OSBORN D.R. – Exact and approximate maximum likehood estimators for moving average processes, *Journal of Royal Stat. Soc., Serie B*, 38, (1977), 114-118.

[17] PIERCE D.A. – Relationships - and the Lack Thereof - Between Economic Time Series, with Special Reference to Money and Interest Rates, (with discussion), *JASA*, 72 (1977), 11-26.

[18] QUENOUILLE M.H. – *The Analysis of Multiple Time Series*, Griffin, London, 1957.

[19] ROZANOV Y.A. – *Stationary Random Processes,* Holden-Day, San Francisco, 1967.

[20] WALLIS K.F. – Multiple Time Series and the Final form of Econometric Models, *Econometrica*, vol. 45, n° 6 (september 1977).

[21] WIENER N. – *The Extrapolation, Interpolation and Smoothing of Stationary Time Series*, Wiley, New York, 1949.

[22] WILSON G. – The Estimation of Parameters in Multivariate Time Series Models, *Journal of Royal Stat. Soc., Serie B*, 35 (1973), 76-85.

[23] WOLD H.O. – *A study in the analysis of stationary time series,* Almquist and Wicksell, Upsala, 2nd Edition, 1954.

[24] ZELLNER A. and F. PALM. – Time Series Analysis and Simultaneous Equation Econometric Models, *Journal of Econometrics*, 2, (1974), 17-54.