

CHAPITRE 5. MODELES AUTOREGRESSIFS.

Dans ce chapitre est étudiée l'estimation du modèle autorégressif ou de la partie autorégressive du modèle mixte ARMA. Les équations exprimant la condition d'optimalité du modèle autorégressif seront établies sous des hypothèses de moins en moins restrictives quant aux bases de fonctions et à l'innovation du modèle autorégressif. Le point essentiel est ici l'introduction d'un vecteur dont les composantes sont les produits du signal par chacune des fonctions de base. Ce vecteur se retrouvera dans toutes les équations d'optimalité, et tous les estimateurs traités dans cette partie, qu'ils concernent le modèle autorégressif ou le modèle à moyenne ajustée MA. Le second point traité est le passage des équations, conditions nécessaires d'optimalité aux estimateurs du modèle, passage qui s'articule autour d'une forme de pseudo-ergodicité du signal, observée en dépit de la non-stationnarité, passage aussi qui s'effectue au travers de plusieurs choix, d'abord entre un système d'équations régulier ou de Cramer, et un système sur-dimensionné, avec plus d'équations que de paramètres, ensuite entre un estimateur affecté d'"effets de bord" mais associé à une solution peu coûteuse, et un estimateur dégagé de ces effets, mais au prix d'un accroissement du coût du calcul.

1. Mise en équations.

1.1. Signal vectoriel associé au signal scalaire.

Le modèle autorégressif évolutif est décrit par une équation analogue à (2-1), mais les coefficients a_1 sont des fonctions du temps obtenues par

(2-20) comme combinaisons linéaires des fonctions $f_i(t)$ connues. Il serait alors logique d'écrire ce modèle sous la forme (2-23) suivante:

$$(2-23) \quad y_t + a_1(t)y_{t-1} + \dots + a_p(t)y_{t-p} = \epsilon_t$$

C'est sous cette forme qu'est écrit le modèle AR évolutif dans les travaux cités au chapitre précédent. Il s'avère cependant plus efficace de redéfinir l'origine des temps, afin de faire coïncider l'indice temporel des deux termes du produit prédicteur-signal retardé, donc multiplier y_{t-i} par $a_i(t-i)$, ce qui donne:

$$(2-24) \quad y_t + a_1(t-1)y_{t-1} + \dots + a_p(t-p)y_{t-p} = \epsilon_t$$

Quel est l'intérêt de cette écriture ? C'est de faire apparaître dans chaque produit les variantes décalées d'un même signal vectoriel, et des vecteurs constants:

$$(2-25) \quad a_i(t-i)y_{t-i} = \left[\sum_{j=0}^m a_{ij} f_j(t-i) \right] y_{t-j}$$

$$= \sum_{j=0}^m a_{ij} (f_j(t-i)y_{t-j})$$

$$= [a_{i0} \dots a_{im}] Y_t$$

Le vecteur Y_t est le vecteur des produits de y_t par chacune des fonctions de la base, ou en quelque sorte des projections de la fonction temporelle y_t sur la base:

$$(2-26) \quad Y_t = \begin{bmatrix} f_0(t) \\ \vdots \\ f_m(t) \end{bmatrix} y_t$$

Réécrire le modèle (1-24) à l'aide de (1-25) permet de faire apparaître ce modèle autorégressif non-stationnaire comme une régression, cette fois stationnaire, du signal y_t sur le passé du vecteur Y_t :

$$(2-27) \quad y_t + [a_{10} \dots a_{1m} \quad a_{20} \dots a_{pm}] \begin{bmatrix} Y_{t-1} \\ \vdots \\ Y_{t-p} \end{bmatrix} = \epsilon_t$$

Tout est dit. Toute la méthodologie des modèles évolutifs tient en ce tour de passe-passe: remplacer un signal scalaire non-stationnaire par un signal vectoriel, donc plus complexe, mais pour lequel le modèle deviendra invariant dans le temps et par conséquent identifiable avec un effort modéré. Il est nécessaire de préciser le sens à accorder ici au mot "identifiable" car ce sens dépend des données disponibles pour l'identification. Si on dispose d'un nombre élevé de signaux $y_t(i)$ qui peuvent être considérés comme autant de réalisations du signal aléatoire y_t , on pourra estimer par une moyenne d'ensemble la covariance $R_{t,s} = E(y_t y_s)$ du signal, pour tous les instants t, s . Il est alors possible d'identifier le modèle autorégressif non-stationnaire (2-24) (ou 2-23), la solution est donnée par (2-28) où $\sigma^2(t)$ représente la variance de ϵ_t .

$$(2-28) \quad \begin{bmatrix} R_{t,t} & R_{t,t-1} & \dots & R_{t,t-p} \\ R_{t-1,t} & R_{t-1,t-1} & & \\ \vdots & & \ddots & \\ R_{t-p,t} & & & R_{t-p,t-p} \end{bmatrix} \begin{bmatrix} 1 \\ a_1(t-1) \\ \vdots \\ a_p(t-p) \end{bmatrix} = \begin{bmatrix} \sigma^2(t) \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Lorsqu'une seule réalisation de y_t est connue, ce qui est la situation courante en traitement du signal, aucune estimation de $R_{t,s}$ n'est possible,

et le système (2-28) ne peut pas être résolu. C'est alors qu'il est nécessaire d'introduire la variable temps dans le modèle par exemple avec le modèle évolutif et ses fonctions $f_i(t)$, car il devient ensuite possible de remplacer les moyennes d'ensemble par des moyennes temporelles et on retrouve un modèle identifiable. C'est désormais dans ce cadre que se poursuit l'étude du modèle (2-24).

1.2. Equations d'optimalité du modèle AR.

L'identification du modèle requiert le choix d'un critère d'optimalité, puis la maximisation du critère. Réécrire le modèle (2-24) met en lumière la triple signification de ϵ_t qui est à la fois l'entrée du modèle autorégressif vu comme un système linéaire, et l'innovation du signal y_t , c'est-à-dire la différence entre y_t et \hat{y}_t conditionnellement à son passé y_{t-1}, y_{t-2}, \dots , mais aussi une erreur de prédiction car \hat{y}_t est la prédiction linéaire de y_t , le signal étant du second ordre. Le choix du critère est immédiat, ϵ_t étant une erreur, il faut minimiser la variance de cette erreur. Soit θ le vecteur des paramètres a_{ij} :

$$(2-29) \quad \theta = [a_{10} \dots a_{1m} \quad a_{20} \quad \dots \quad a_{pm}]^T$$

la variance de ϵ_t s'écrit:

$$(2-30) \quad E(\epsilon_t^2) = [1 \theta^T] E \begin{pmatrix} y_t \\ y_{t-1} \\ \cdot \\ \cdot \\ y_{t-p} \end{pmatrix} [y_t \quad y_{t-1}^T \quad \dots \quad y_{t-p}^T] \begin{bmatrix} 1 \\ \theta \end{bmatrix}$$

La minimisation de $E(\epsilon_t^2)$ par rapport à θ conduit au système d'équations:

$$(2-31) \quad E \left(\begin{bmatrix} Y_{t-1} \\ \cdot \\ \cdot \\ Y_{t-p} \end{bmatrix} [Y_{t-1}^T \dots Y_{t-p}^T] \right) \theta = E \left(\begin{bmatrix} Y_{t-1} \\ \cdot \\ \cdot \\ Y_{t-p} \end{bmatrix} Y_t \right)$$

Ce système s'écrit encore:

$$(2-32) \quad E \left(\begin{bmatrix} Y_t \\ Y_{t-1} \\ \cdot \\ \cdot \\ Y_{t-p} \end{bmatrix} [Y_t \ Y_{t-1}^T \dots Y_{t-p}^T] \begin{bmatrix} 1 \\ \theta \end{bmatrix} \right) = \begin{bmatrix} \sigma^2(t) \\ 0 \\ \cdot \\ \cdot \\ 0 \end{bmatrix}$$

On reconnaît dans (2-31) ou (2-32) la forme habituelle des équations de Yule-Walker pour la prédiction linéaire, mais les éléments du passé sont remplacés par les vecteurs Y_t , ce qui accroît la taille du système à résoudre par rapport aux équations (2-28). Par contre le vecteur inconnu ne dépend plus du temps. C'est cette remarque, déjà faite, qui expliquera qu'on puisse utiliser des sommes temporelles pour faire l'estimation.

1.3. Interprétation du signal vectoriel.

Il est utile de considérer un cas particulier qui va éclairer le rôle du vecteur Y_t . Ce sera le cas où la fonction $f_0(t)$ est égale à la fonction constante $f_0(t)=1$. La première composante du vecteur Y_t est alors égale à y_t , le signal non-stationnaire lui-même. Il en découle un lien entre le vecteur θ solution du modèle non-stationnaire et le modèle autorégressif vectoriel de Y_t . Remarquons d'abord que si y_t est autorégressif d'ordre p , Y_t est aussi autorégressif du même ordre p , et également non-stationnaire. En effet si:

$$(2-33) \quad E(Y_t/Y_{t-1}, Y_{t-2}, \dots) = E(Y_t/Y_{t-1}, \dots, Y_{t-p})$$

alors, puisque les fonctions $f_i(t)$ sont certaines,

$$(2-34) \quad E(Y_t/Y_{t-1}, Y_{t-2}, \dots) = E(Y_t/Y_{t-1}, \dots, Y_{t-p})$$

Soient alors A_0, A_1, \dots, A_p les matrices des coefficients du modèle vectoriel de Y_t , et ξ_t l'innovation de Y_t :

$$(2-35) \quad A_0 Y_t + A_1 Y_{t-1} + \dots + A_p Y_{t-p} = \xi_t$$

Les matrices A_i dépendent du temps. Plusieurs normalisations du modèle autorégressif vectoriel existent suivant qu'on choisisse d'imposer à la variance de ξ_t d'être la matrice identité, auquel cas A_0 n'est défini qu'à une transformation unitaire près, ou d'imposer à A_0 d'être la matrice identité. Retenant ce dernier choix, la solution optimale au problème d'identification des A_i est obtenue en minimisant la variance Ξ de l'erreur ξ_t . Comme cette variance est une matrice positive, le choix des normes de Ξ et donc des critères d'identification est large. Heureusement, la solution $[A_1, \dots, A_p]$ obtenue est la même, que l'on choisisse de minimiser la trace de Ξ , son déterminant ... Si R_Y est la matrice de covariance de Y_t et de son passé jusque $t-p$, la variance de l'erreur ξ_t s'écrit à l'instant t :

$$(2-36) \quad \Xi = [I \ A_1 \ \dots \ A_p] R_Y \begin{bmatrix} I \\ A_1^T \\ \cdot \\ \cdot \\ A_p^T \end{bmatrix}$$

avec

$$(2-37) \quad R_Y = E \left(\begin{bmatrix} Y_t \\ Y_{t-1} \\ \cdot \\ \cdot \\ Y_{t-p} \end{bmatrix} \begin{bmatrix} Y_t^T & Y_{t-1}^T & \dots & Y_{t-p}^T \end{bmatrix} \right)$$

GAMBOTTO, 1979, fait remarquer que si dans le modèle autorégressif vectoriel (stationnaire, mais sa remarque vaut aussi pour le modèle (2-36) non-stationnaire à l'instant t), on minimise la trace de la variance Ξ , relativement aux éléments des matrices A_i , on obtient autant d'équations que de lignes dans les matrices A_i , et ces $m+1$ équations sont indépendantes les unes des autres. En effet le terme ξ_{ii} en i -ème position sur la diagonale de Ξ est le produit de la i -ème ligne de la matrice $[I \ A_1 \dots A_p]$ par la matrice R_Y et par la i -ème colonne de la transposée de $[I \ A_1 \dots A_p]$, c'est-à-dire la transposée de la i -ème ligne de $[I \ A_1 \dots A_p]$. Chacune de ces lignes intervient isolément dans un des termes ξ_{ii} de la trace de Ξ , et une seule fois.

Si nous appliquons la remarque précédente à notre modèle non-stationnaire, en comparant les équations (2-30) et (2-36), le vecteur θ , ou plutôt son transposé θ^T apparaît comme la première ligne de la matrice $[A_1 \dots A_p]$. En effet le produit de la première ligne de I , soit $(1 \ 0 \dots 0)$ par Y_t donne la première composante de Y_t qui n'est autre que y_t quand $f_0(t)=1$. Ceci montre que la première ligne de la matrice $[A_1 \dots A_p]$ est indépendante du temps. Il n'en sera pas de même pour les lignes suivantes. Ainsi le vecteur Y_t acquiert une deuxième signification. Il était auparavant un vecteur auxiliaire qui permettait d'exprimer le modèle autorégressif non-stationnaire comme une régression de y_t sur le passé du vecteur Y_t , et

qui donc assumait toute la non-stationnarité du modèle. Il devient maintenant un prolongement non seulement du signal y_t mais aussi et surtout de son modèle AR. Il suffira donc d'estimer le modèle vectoriel AR de Y_t pour obtenir par simple sélection de la première ligne, le vecteur θ des coefficients du modèle scalaire de y_t . L'intérêt est évident: la solution d'un système linéaire (2-31) dont le second membre est quelconque relativement à la matrice à inverser, ou la solution du système (2-32) où la matrice n'a pas de structure simple à cause du mélange entre scalaires et vecteurs, seront remplacés par la solution des équations de Yule-Walker d'un modèle vectoriel. Gain en facilité d'écriture de l'algorithme et du programme, gain en vitesse de calcul, comme on le verra au chapitre suivant.

Ce prolongement présente une difficulté qui se surmonte facilement: si la première ligne des A_i est invariante temporellement, il n'en est pas ainsi des autres lignes, ce qui devrait empêcher l'emploi d'estimateurs ergodiques pour les espérances mathématiques, avec comme conséquence l'impossibilité d'estimer le modèle. Mais ceci n'est pas gênant, car seule la première ligne du modèle doit être estimée correctement. Si on fait l'hypothèse que les autres lignes du modèle autorégressif vectoriel sont invariantes, on pourra estimer le modèle, mais ce ne sera pas le modèle correct pour Y_t , les lignes seront une moyenne sur l'intervalle de temps où est faite l'estimation, entre les valeurs prises à chaque instant par ces lignes. La première ligne sera alors correcte, et le modèle scalaire sera bien estimé. La structure de la matrice R_Y , constituée à partir de décalages temporels de Y_t donne par cette sommation sur le temps, une structure de matrice de Toëplitz par blocs, et la résolution du modèle AR se fait par l'algorithme de LEVINSON, 1946, tel qu'il a été étendu au cas vectoriel par WHITTLE, 1963, puis WIGGINS, ROBINSON, 1965.

1.4. Equations pour la partie AR du modèle ARMA.

Pour passer d'un estimateur du modèle autorégressif à un estimateur de la partie autorégressive du modèle ARMA, il faut revenir sur la minimisation de la variance de ϵ_t . En écrivant l'équation (2-30) comme:

$$(2-38) \quad 0 = \text{Grad}_{\theta} E(\epsilon_t^2) = 2 \cdot E(\epsilon_t \text{Grad}_{\theta} \epsilon_t)$$

avec:

$$(2-39) \quad \text{Grad}_{\theta} \epsilon_t = \begin{bmatrix} Y_{t-1} \\ \cdot \\ \cdot \\ Y_{t-p} \end{bmatrix}$$

on fait apparaître l'équation (2-38) comme une équation d'orthogonalité entre ϵ_t et les p derniers échantillons du passé de Y_t . Une telle relation résulte directement de l'équation du modèle AR (2-24) et de la blancheur de ϵ_t . Le modèle mixte ARMA sera donné par (2-40), où ϵ_t sera toujours un bruit blanc centré, mais ici sa variance sera 1.

$$(2-40) \quad y_t + a_1(t-1)y_{t-1} + \dots + a_p(t-p)y_{t-p} = b_0(t)\epsilon_t + b_1(t-1)\epsilon_{t-1} + \dots + b_q(t-q)\epsilon_{t-q}$$

A cause de la moyenne mobile de $\epsilon_t \dots \epsilon_{t-q}$, on n'a plus l'orthogonalité de la partie autorégressive avec le signal $y_{t-1} y_{t-2} \dots$. La quantité w_t définie par (2-41) contenant les échantillons du passé de ϵ_t jusque ϵ_{t-q} , n'est orthogonale aux échantillons y_{t-i} et donc aux échantillons Y_{t-i} que pour $i > q$.

$$(2-41) \quad w_t = y_t + a_1(t-1)y_{t-1} + \dots + a_p(t-p)y_{t-p}$$

Ainsi:

$$(2-42) \quad E\left(\left(Y_t + \theta^T \begin{bmatrix} Y_{t-1} \\ \cdot \\ Y_{t-p} \end{bmatrix}\right) \cdot Y_{t-i}^T\right) = 0 \quad \text{pour } i > q$$

La relation (2-30) se trouvera donc transformée en la relation (2-43) exprimant une condition nécessaire pour que θ soit le vecteur des coefficients a_{ij} de la partie autorégressive.

$$(2-43) \quad E\left(\begin{bmatrix} Y_{t-q-1} \\ \cdot \\ Y_{t-q-p} \end{bmatrix} [Y_{t-1}^T \dots Y_{t-p}^T]\right) \theta = -E\left(\begin{bmatrix} Y_{t-q-1} \\ \cdot \\ Y_{t-q-p} \end{bmatrix} Y_t\right)$$

Cette dernière relation est l'équivalent non-stationnaire de l'équation utilisée par HSIA, LANDGREBE, 1967, puis GERSCH, 1970, pour estimer la partie autorégressive du modèle ARMA, relation que l'on retrouve aussi dans la méthode de MASSEY, 1969, où l'accent est mis sur une structure de Hankel au lieu d'une structure de Toeplitz pour la matrice multipliant θ dans (2-43). On pourrait aussi la retrouver par le biais du modèle ARMA vectoriel de Y_t dans le cas qui vient d'être étudié où $f_0(t)=1$. Il faut cependant remarquer que les équations (2-31), (2-32) donnant les conditions d'optimalité du modèle AR, ou (2-43) donnant celles de la partie AR du modèle ARMA ne constituent pas des conditions suffisantes pour que θ les vérifiant soit le vecteur des paramètres du modèle. En effet, aucune des matrices figurant dans ces trois systèmes linéaires n'est de rang complet car le vecteur aléatoire Y_t est le produit d'un scalaire aléatoire y_t par le vecteur des fonctions $f_0(t) \dots f_m(t)$ qui sont certaines. Les matrices multipliant θ dans les trois équations sont de rang au plus égal à p dans (2-31) et (2-43)

alors que leur dimension est $p.(m+1)$, et de rang égal à $(p+1)$ dans (2-32), pour une dimension de $1+p(m+1)$. C'est le passage de l'espérance mathématique à la somme temporelle de l'estimateur ergodique qui rend ces matrices inversibles et de rang complet, les conditions deviennent des conditions suffisantes sur θ .

2. Estimation.

2.1. Somme temporelle pour estimer l'espérance.

L'emploi du terme "estimateur" à propos des équations (2-31), (2-32) et (2-43) n'était pas justifié, car rien n'indiquait comment étaient estimées les espérances mathématiques figurant dans ces équations. C'est ce point qui est étudié maintenant, en supposant désormais que le signal y_t est connu par son observation sur l'intervalle $[0, T]$. Dans le cas stationnaire, l'estimateur retenu est toujours l'estimateur ergodique où la somme temporelle remplace la moyenne d'ensemble. Il est clair, du fait de l'invariance temporelle de θ que le remplacement dans les trois équations de l'espérance par une somme sur le temps, entraîne trois équations qui expriment encore des conditions nécessaires sur θ , et si la somme se fait sur un intervalle assez long, on obtient des conditions suffisantes pour que θ soit solution au problème posé. Pourtant ce dernier point a des interprétations fort différentes suivant la stationnarité de ϵ_t .

Supposons d'abord que l'innovation ϵ_t du modèle autorégressif est stationnaire, sa variance est invariante $E(\epsilon_t^2) = \sigma^2$, il est alors équivalent de minimiser cette variance à un instant t , ou la somme sur un intervalle $[0, T]$ de ϵ_t . L'équation (2-31) devient alors l'équation (2-44) et une transformation analogue s'effectue sur (2-32). L'estimateur ergodique est ainsi

pleinement justifié, et l'équation (2-44) a la signification d'une solution des moindres carrés.

$$(2-44) \quad \sum_{t \in T} \begin{bmatrix} Y_{t-1} \\ \cdot \\ \cdot \\ Y_{t-p} \end{bmatrix} [Y_{t-1}^T \dots Y_{t-p}^T] \cdot \theta = - \sum_{t \in T} \begin{bmatrix} Y_{t-1} \\ \cdot \\ \cdot \\ Y_{t-p} \end{bmatrix} Y_t$$

Si maintenant la variance de ϵ_t n'est plus constante, il n'est plus licite de minimiser la somme temporelle de ϵ_t^2 , et (2-44) n'a plus d'interprétation en terme de critère. L'équation (2-44) n'exprime qu'une relation d'orthogonalité entre ϵ_t et le passé de Y_t , et toute pondération différente sur l'intervalle $[0, T]$ des équations (2-31) conviendrait aussi pour constituer un estimateur semblable à (2-44). Par exemple si la variance $\sigma^2(t)$ de ϵ_t pouvait s'écrire $\sigma^2(t) = (G(t))^2 \sigma^2$ où σ^2 est constant, inconnu et $G(t)$ est connu, alors il devient légitime de minimiser σ^2 , et de prendre comme critère:

$$(2-45) \quad J = \sum_{t \in T} \frac{1}{G(t)^2} E(\epsilon_t^2)$$

L'équation (2-44) se transforme ainsi en (2-46):

$$(2-46) \quad \sum_{t \in T} \frac{1}{G(t)^2} \begin{bmatrix} Y_{t-1} \\ \cdot \\ \cdot \\ Y_{t-p} \end{bmatrix} [Y_{t-1}^T \dots Y_{t-p}^T] \cdot \theta = - \sum_{t \in T} \frac{1}{G(t)^2} \begin{bmatrix} Y_{t-1} \\ \cdot \\ \cdot \\ Y_{t-p} \end{bmatrix} Y_t$$

Ne disposant pas en général d'information sur la variance de ϵ_t , à priori, on retiendra (2-44) plutôt que (2-46), mais l'estimation de θ risquera d'être biaisée, et ce d'autant plus que $\sigma^2(t)$ variera beaucoup entre 0 et T.

Pour l'estimation de la partie autorégressive du modèle ARMA, on

retiendra l'équation nécessaire (2-43), et l'estimateur ergodique, ce qui donne l'équation (2-47):

$$(2-47) \quad \sum_{t \in \Gamma} \begin{bmatrix} Y_{t-q-1} \\ \cdot \\ \cdot \\ Y_{t-q-p} \end{bmatrix} [Y_{t-1}^T \dots Y_{t-p}^T] \cdot \theta = \sum_{t \in \Gamma} \begin{bmatrix} Y_{t-q-1} \\ \cdot \\ \cdot \\ Y_{t-q-p} \end{bmatrix} Y_t$$

Ici aussi, aucun critère n'est associé à cette équation, mais la non-stationnarité n'est pas en cause, car la même situation existe dans le cas stationnaire, où GERSCH, 1970, a montré que l'estimateur obtenu était non-biaisé. Ici, la non-stationnarité ne permet pas de faire un calcul explicite du biais de θ , mais il me semble que ce biais, par analogie avec ce qui se produit en (2-44) doit être nul si la variance de la variable w_t définie en (2-41) est constante, et croître avec la dynamique de $E(w_t)$, c'est-à-dire avec le rapport entre les amplitudes maximales et minimales de la variance de w_t , ainsi qu'avec la vitesse de variation ou le taux de non-stationnarité, pris en un sens intuitif, de la partie AR. Cette intrication des deux phénomènes rend très difficile l'analyse du biais de θ dans (2-47).

Ainsi les modèles AR et ARMA évolutifs permettent pour un signal non-stationnaire de tourner la non-ergodicité inhérente à un tel signal, et de fabriquer un estimateur à partir de sommes temporelles. Il est visible que ceci est permis par la conjugaison de deux propriétés: l'existence d'une transformation du modèle dans laquelle les paramètres inconnus deviennent tous invariants vis-à-vis du temps, et la linéarité des équations d'optimalité, qui est liée au fait que ces équations expriment des conditions d'orthogonalité. L'estimation qui en résulte, biaisée par les variations de l'énergie résiduelle pourra être rendue sans-biais par une estima-

tion à postériori des variations incriminées, et leur incorporation dans les équations.

2.2. Méthodes de "corrélation" et de "covariance".

Tel est le choix de l'estimateur des espérances mathématiques: une somme temporelle, mais ceci ne fixe pas totalement les équations (2-44) ou (2-47). Il est en effet possible de distinguer comme pour le modèle autorégressif stationnaire deux cas, connus sous les dénominations malencontreuses mais fermement établies de "méthode de corrélation" et "méthode de covariance". Leur opposition peut être interprétée de plusieurs façons. Tout d'abord, il est possible de considérer dans (2-44) ou (2-47) que chaque élément des matrices du terme de gauche, ou des vecteurs du terme de droite est estimé par une somme temporelle, utilisant au mieux tous les échantillons disponibles. Les matrices deviennent alors des matrices de Toeplitz par blocs. Si R_i est l'estimateur de $E(Y_{t-i} Y_t^T)$:

$$(2-48) \quad R_i = \sum_{t=i}^T Y_{t-i} Y_t^T \quad R_{-i} = \sum_{t=i}^T Y_t Y_{t-i}^T$$

et si V_i est l'estimateur de $E(Y_{t-i} y_t)$

$$(2-49) \quad V_i = \sum_{t=i}^T Y_{t-i} y_t$$

les systèmes (2-44) et (2-47) deviennent:

$$(2-50) \quad \begin{bmatrix} R_0 & R_{-1} & & & R_{-p+1} \\ R_1 & R_0 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & R_0 \\ R_{p-1} & & & & R_0 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \cdot \\ \cdot \\ \cdot \\ \theta_p \end{bmatrix} = \begin{bmatrix} V_1 \\ \cdot \\ \cdot \\ \cdot \\ V_p \end{bmatrix}$$

et

$$(2-51) \quad \begin{bmatrix} R_q & R_{q-1} & & R_{q-p+1} \\ R_{q+1} & R_q & \cdot & \\ & \cdot & \cdot & \\ & & & R_q \\ R_{q+p-1} & & & \end{bmatrix} \begin{bmatrix} \theta_1 \\ \cdot \\ \cdot \\ \cdot \\ \theta_p \end{bmatrix} = \begin{bmatrix} V_{q+1} \\ \cdot \\ \cdot \\ \cdot \\ V_{q+p} \end{bmatrix}$$

Le vecteur θ a été partitionné en vecteurs à $m+1$ composantes en accord avec la taille des blocs. Ceci donne la méthode dite "de corrélation". Si c'est la matrice entière qui est estimée par une somme temporelle, la matrice perd cette structure de Toeplitz, et l'intervalle Γ où s'effectue la somme doit être $\Gamma=[p,T]$ pour l'équation (2-44) et $\Gamma=[p+q,T]$ pour (2-47). Cela peut être la seconde façon d'interpréter l'opposition "corrélation"/"covariance". Si on veut réécrire les équations "de corrélation" (2-50) ou (2-51) sous leur forme primitive, (2-44) ou (2-47), on voit que l'intervalle Γ doit être $\Gamma=-\infty,+\infty[$ ou au moins $\Gamma=[0,T+p]$ pour (2-44)/(2-50) et $\Gamma=[0,T+p+q]$ pour (2-47)/(2-51). Les échantillons non observés ($t < 0$ ou $t > T$) sont remplacés par des zéros. Dans cette optique, la "méthode de covariance" est plus conforme au principe d'entropie maximale car elle ne fait pas d'hypothèse, n'introduit pas d'information relative aux échantillons non-observés, tandis que la "méthode de corrélation" impose la nullité de ces échantillons. Une troisième opposition entre les deux méthodes en découle. L'une sera affectée d'effets dits "de bord", l'autre pas. Le fait de prolonger par des zéros le signal disponible introduit une discontinuité dans ce signal qui est vue comme une non-stationnarité temporelle, ou une convolution dans le domaine spectral. Le remède utilisé en modélisation stationnaire est l'emploi d'une fenêtre de pondération ou d'apodisation qui amène progressivement le signal vers zéro aux bornes de l'intervalle. Il est clair qu'un tel emploi est à proscrire pour la modélisation non-stationnaire car la fenêtre interférerait

avec les évolutions du signal et de son modèle.

Ainsi dans le modèle non-stationnaire, autorégressif, ou partie autorégressive d'un ARMA, se trouvent en concurrence deux méthodes. La "méthode de corrélation" a l'avantage de donner au système à résoudre une structure simple, avec une matrice de Toeplitz. L'inconvénient est que le modèle est biaisé par les effets de bord dus à l'ajout d'échantillons nuls. Le biais ne peut pas être compensé par une fenêtre de pondération comme dans le cas stationnaire, mais ceci n'est pas très gênant car la longueur T de la fenêtre d'observation est beaucoup plus longue que celle utilisée dans les modèles AR stationnaires. Ainsi sur la parole, la durée des fenêtres est en général de 20 à 25 millisecondes pour le modèle stationnaire, et de 100 à 600 ms pour le modèle évolutif soit 4 à 30 fois plus. Le biais est inversement proportionnel à la durée, ce qui le rend négligeable dans les modèles évolutifs. La "méthode de covariance" ne présente pas ce biais, mais elle le paie par un système plus difficile à résoudre. Heureusement, si la matrice du système n'est plus de Toeplitz, elle reste un produit de deux matrices de Toeplitz et on sait que ceci est une condition qui assure l'existence d'un algorithme rapide (chapitre 6). Peut-on départager les deux méthodes? HALL, OPPENHEIM, WILLISKY, 1983, à cause du biais de la "méthode de corrélation" penchent nettement pour l'autre méthode. Je ne partage pas cette opinion. Les simulations que j'ai réalisées montrent un avantage certes pour la "méthode de covariance", mais le gain est si faible qu'il ne me paraît pas justifier le coût de cette méthode: le code Fortran de la "covariance" est environ huit fois plus encombrant que celui de la "corrélation".

Il reste un détail à expliciter dans ces méthodes. L'estimateur de l'espérance mathématique ne devrait pas être une somme mais une moyenne tem-

porelle. On se convaincra aisément que ceci n'a pas d'importance dans les équations (2-44) ou (2-47), mais pour les équations de "corrélation", l'intervalle sur lequel est estimé chaque bloc R_i a une longueur variable. Doit-on alors diviser R_i par T ou par $T-i$ (en réalité $T+1$ ou $T+1-i$ si on n'a pas soustrait au signal sa moyenne, ce qui retire un degré de liberté) ? Il est connu par l'étude du cas stationnaire que l'emploi de $T-i$ donne un estimateur des covariances non biaisé, alors que l'emploi de T induit un biais. Mais comme l'estimateur biaisé (avec T) garantit la positivité de la matrice de covariance dans (2-50) ou (2-51), contrairement à l'estimateur non-biaisé ($T-i$), c'est l'estimateur biaisé qui est retenu, et il est légitime de ne pas diviser par T qui figurerait aux deux côtés de l'équation. Ceci achève la discussion des équations (2-50) et (2-51).

2.3. Méthode sur-dimensionnée.

A côté des deux estimateurs précédents, la "méthode de corrélation" et la "méthode de covariance" adaptées au contexte non-stationnaire, on peut faire figurer un troisième type d'estimateur qui sera dit "sur-dimensionné". Nous allons le considérer ici dans le cas le plus général de l'estimation de la partie autorégressive d'un modèle ARMA. Revenons sur la relation d'orthogonalité (2-42) entre le résidu à l'instant t et le passé du signal Y_t pour tout i supérieur à l'ordre q de la partie MA. Les deux méthodes vues précédemment consistent à choisir p exemplaires des conditions d'orthogonalité, de façon à obtenir un système d'équations linéaires avec autant d'équations que d'inconnues, ce qui donne (2-31) ou (2-43), puis avec le remplacement des espérances mathématiques par des sommes temporelles (2-44) et (2-47). Ce sont bien sûr les p premières conditions qui sont retenus pour obtenir la meilleure précision possible car les premiers termes des

covariances sont mieux estimés que ceux correspondant à de plus lointains retards. Rien n'empêche cependant de considérer un nombre N d'équations, supérieur à p . C'est ce qu'a fait CADZOW, 1980-a et b, dans le contexte stationnaire. Voyons comme ceci se généralise dans le cadre non-stationnaire évolutif.

Retenir les N premières occurrences de la condition (2-42) conduit au système suivant:

$$(2-52) \quad E \begin{pmatrix} Y_{t-q-1} \\ \cdot \\ \cdot \\ Y_{t-q-N} \end{pmatrix} (y_t + [Y_{t-1}^T \dots Y_{t-p}^T] \theta) = \begin{pmatrix} 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix}$$

Lorsque $N > p$, ce système sur déterminé ne possède pas de solution exacte dès que y_{t-p} n'est pas exactement un signal ARMA (p, q) , ce qui survient toujours en pratique. Le second membre ne peut être uniformément nul. Soit E_i le i -ème bloc de ce second membre:

$$(2-53) \quad E_i = E(Y_{t-q-i} (y_t + [Y_{t-1}^T \dots Y_{t-p}^T] \theta))$$

Le système (2-52) s'écrit avec E_i :

$$(2-54) \quad E \begin{pmatrix} Y_{t-q-1} \\ \cdot \\ \cdot \\ Y_{t-q-N} \end{pmatrix} (y_t + [Y_{t-1}^T \dots Y_{t-p}^T] \theta) = \begin{pmatrix} E_1 \\ \cdot \\ \cdot \\ E_N \end{pmatrix}$$

On résoudra ce système de façon approximative en minimisant la norme du vecteur au second membre. C'est l'extension non-stationnaire de la méthode connue sous le nom barbare de "covariance sur la corrélation". CADZOW, 1980-a et b, propose d'améliorer cette méthode en pondérant différemment la

contribution de chaque E_i au critère. Dans notre contexte, ceci revient à minimiser un critère J dépendant de poids γ_i .

$$(2-55) \quad J = \sum_{i=1}^N \gamma_i E_i^T E_i$$

Il est clair que si tous les poids γ_i sont égaux, minimiser (2-55), c'est-à-dire résoudre (2-54) par une méthode de moindres carrés est équivalent à utiliser la pseudo-inverse de la matrice multipliant θ dans (2-52) ou (2-54) pour obtenir la solution θ . Quand les poids γ_i diffèrent entre eux, on a seulement une méthode de moindres carrés pondérés sur les corrélations des Y_t . La minimisation du critère est facile à obtenir:

$$(2-56) \quad J = \sum_{i=1}^N \gamma_i E((Y_t + \theta^T \Phi_t) Y_{t-q-i}^T) E(Y_{t-q-i} (Y_t + \Phi_t^T \theta))$$

avec $\Phi_t^T = [Y_{t-1}^T \dots Y_{t-p}^T]$ d'où pour $\text{Grad}_{\theta} J = 0$:

$$(2-57) \quad \sum_{i=1}^N \gamma_i E(\Phi_t Y_{t-q-i}^T) E(Y_{t-q-i} (Y_t + \Phi_t^T \theta)) = 0$$

D'où le système:

$$(2-58) \quad \left[\sum_{i=1}^N \gamma_i E(\Phi_t Y_{t-q-i}^T) E(Y_{t-q-i} \Phi_t^T) \right] \theta = - \sum_{i=1}^N \gamma_i E(\Phi_t Y_{t-q-i}^T) E(Y_{t-q-i} Y_t)$$

Il ne reste plus pour obtenir l'estimateur sur-dimensionné non-stationnaire qu'à remplacer les espérances mathématiques par des sommes. On ne peut le faire qu'après un choix entre deux variantes: doit-on travailler sur (2-58), en faisant chaque somme sur la totalité des échantillons disponibles, c'est-à-dire en sommant $\Phi_t Y_{t-q-i}^T$ de $t=q+i$ à $t=T+1$ et $Y_{t-q-i} Y_t$ de $t=q+i$ à $t=T$, ou bien doit-on revenir au système initial (2-52) en faisant les sommes dans (2-58) de $t=q+n$ à $t=T$, quel que soit i . Dans le premier cas, on a une analogie avec la "méthode de corrélation" ce qui suppose le signal prolongé

par des zéros avec ce que cela implique d'effets de bord. Dans le second cas, qui est analogue à une "méthode de covariance", l'effet de bord disparaît, mais les sommes sont calculées sur moins de points, ce qui diminue la qualité de l'estimateur, par une augmentation de la variance. Dans les simulations décrites plus loin (chapitre 9), c'est le premier de ces estimateurs qui a été implanté.

On pourrait obtenir un autre estimateur du type sur-dimensionné en retenant au lieu des relations d'orthogonalité (2-42) entre le résidu w_t et le passé de Y_t , les mêmes relations entre w_t et le passé de y_t , en retenant donc le signal au lieu de ses projections sur la base de fonctions. Il faut dans ce cas prendre N supérieur à $p(m+1)$ pour avoir plus d'équations que d'inconnues. La solution s'obtient de la même façon que (2-58).

$$(2-59) \quad \left[\begin{array}{c} N \\ \sum_{i=1} \gamma_i E(\phi_t y_{t-q-i}) E(y_{t-q-i} \phi_t^T) \end{array} \right] \theta = \sum_{i=1}^N \gamma_i E(\phi_t y_{t-q-i}) E(y_{t-q-i} y_t)$$

Les deux estimateurs possibles pour le remplacement des espérances des espérances par des sommes temporelles, se retrouvent aussi dans ce cas. Les quatre cas obtenus sont encore multipliés par la diversité des pondérations γ_i . On peut prendre le choix uniforme: $\gamma_i=1$. CADZOW, 1980-a et b, suggère le choix d'une fonction monotone décroissante vis à vis de i , et en particulier les puissances de $T-i$: $\gamma_i=(T-i)^k$ avec k compris entre 1 et 4. Il retient aussi pour N la valeur $N=T-1$, ce qui semble excessif, les dernières équations étant beaucoup plus biaisées, et pour compenser cela, pondérées plus sévèrement. Je préfère utiliser un nombre d'équations N compris entre trois et huit fois l'ordre du modèle, avec une pondération en cosinus: $\gamma_i=0.5+0.5\cos\pi(\frac{i-q}{n})$. CADZOW, MOSES, 1981, proposent un calcul optimal de la pondération γ_i sous l'hypothèse d'une répartition statistique uniforme des

zéros du modèle ARMA, les γ_i se relient alors aux corrélations du signal. Le calcul d'une pondération optimale dans le cas non-stationnaire semble beaucoup plus délicat, et peut-être impossible, et c'est ce qui fera la difficulté d'emploi des méthodes sur-dimensionnées car les choix de N et des γ_i restent assez arbitraires.

2.4. Comparaison des trois méthodes.

Dans ce chapitre ont été étudiés trois méthodes d'estimation du modèle autorégressif évolutif ou de la partie AR d'un modèle ARMA évolutif; ce sont des extensions non-stationnaires des méthodes dites "de corrélation", "de covariance" et "sur-dimensionnée". Leurs qualités en tant qu'estimateurs seront étudiées par la suite, mais il est utile de les comparer d'ores et déjà sur le plan de leur coût évalué en nombre d'opérations (additions ou multiplications). Le coût se répartit en deux phases de calcul, d'abord l'évaluation des deux membres du système d'équation, puis la résolution de ce système.

La première phase nécessite le calcul d'un certain nombre de covariances du vecteur Y_t . Le nombre d'opérations nécessaire au calcul de la matrice et du second membre du système linéaire a été évalué par HALL, OPPENHEIM, WILLSKY, 1983, pour les méthodes "de corrélation" et "de covariance". Leur calcul est en réalité erroné, pour deux raisons. La première est qu'ils ne comptent que le nombre d'éléments à calculer, alors que chacun est une somme sur un intervalle de temps variable entre T et T-p-q, ce qui change le nombre d'opérations. La seconde est que pour la méthode "de covariance", on ne peut pas séparer les deux phases, ou plutôt la façon dont se fait la seconde influe sur celle dont se fait la première. Ils supposent donc qu'il faut calculer la totalité de la matrice de covariance, or

l'algorithme de résolution utilise, comme ce sera montré au chapitre suivant les blocs de la matrice de la méthode "de corrélation" qui est Toeplitz par blocs. Les corrections pour obtenir la matrice de covariance sont incluses dans la résolution du système linéaire. Les deux méthodes ont donc presque le même coût dans la première phase, et c'est même la "méthode de covariance" qui nécessitera légèrement moins de calcul, du fait que les sommes, dans la matrice s'arrêteront à T-1.

Le calcul du nombre d'opérations conforme à (2-48), (2-49) et (2-50) donne:

$$(2-60) \quad M=(T+1)(p(m+1)(m+2)-\frac{1}{2}m(m+1))-\frac{1}{2}p(m+1)(p(m+2)-m)$$

Pour la méthode "de covariance" ceci deviendrait:

$$(2-61) \quad M=(T+1)(p(m+1)(m+2)-\frac{1}{2}m(m+1))-p(p(m+1)(m+2)-\frac{1}{2}m(m+1))$$

Ici M représente le nombre de multiplications nécessaire pour former le système à résoudre. Ce nombre est valable pour le modèle purement autorégressif et quand $f_0(t) \neq 1$. Si $f_0(t) = 1$, le vecteur au second membre se retrouve dans la matrice, et on économise $(T+1-p)(p-1)(m+1)$ opérations. Si on a affaire à un modèle ARMA d'ordre MA supérieur à l'ordre AR, le nombre M présente une valeur maximale de:

$$(2-61) \quad M=(T+1-p-q)(p(m+1)(2m+3)-(m+1)^2)$$

Au-delà de ces décomptes laborieux apparaît un coût de calcul de l'ordre de Tpm^2 dans le cas AR à $2Tpm^2$ dans le cas ARMA. Pour la méthode "surdimensionnée", le coût sera:

$$(2-61) \quad M=N \left[(T+1-q-N)p(m+1)(m+2)+p^2(m+1)^5+p(m+1)^3 \right]$$

L'ordre d'un tel calcul est de $N^2 p m^2$ (avec des termes résiduels en $N p^2 m^5$ (!)), soit beaucoup plus que dans les autres méthodes.

La seconde phase, de résolution du système, sera analysée au chapitre suivant. On y verra qu'existent des algorithmes rapides pour la résolution des systèmes associés aux deux premières méthodes, d'ordre $p^2 m^3$, tandis que la troisième méthode ne connaît pas de tels algorithmes, et son coût est de l'ordre de $p^3 m^3$.

