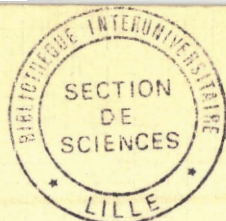


N° d'ordre : 187

55376
1988
5



THESE

55376
1988
5

présentée à

L'UNIVERSITE DES SCIENCES ET TECHNIQUES DE LILLE FLANDRES ARTOIS

pour obtenir

LE TITRE DE DOCTEUR DE L'UNIVERSITE
SPECIALITE : MATHEMATIQUES APPLIQUEES

par

GONCALVES Esmeralda

PROCESSUS FRACTIONNAIRES

Membres du Jury :

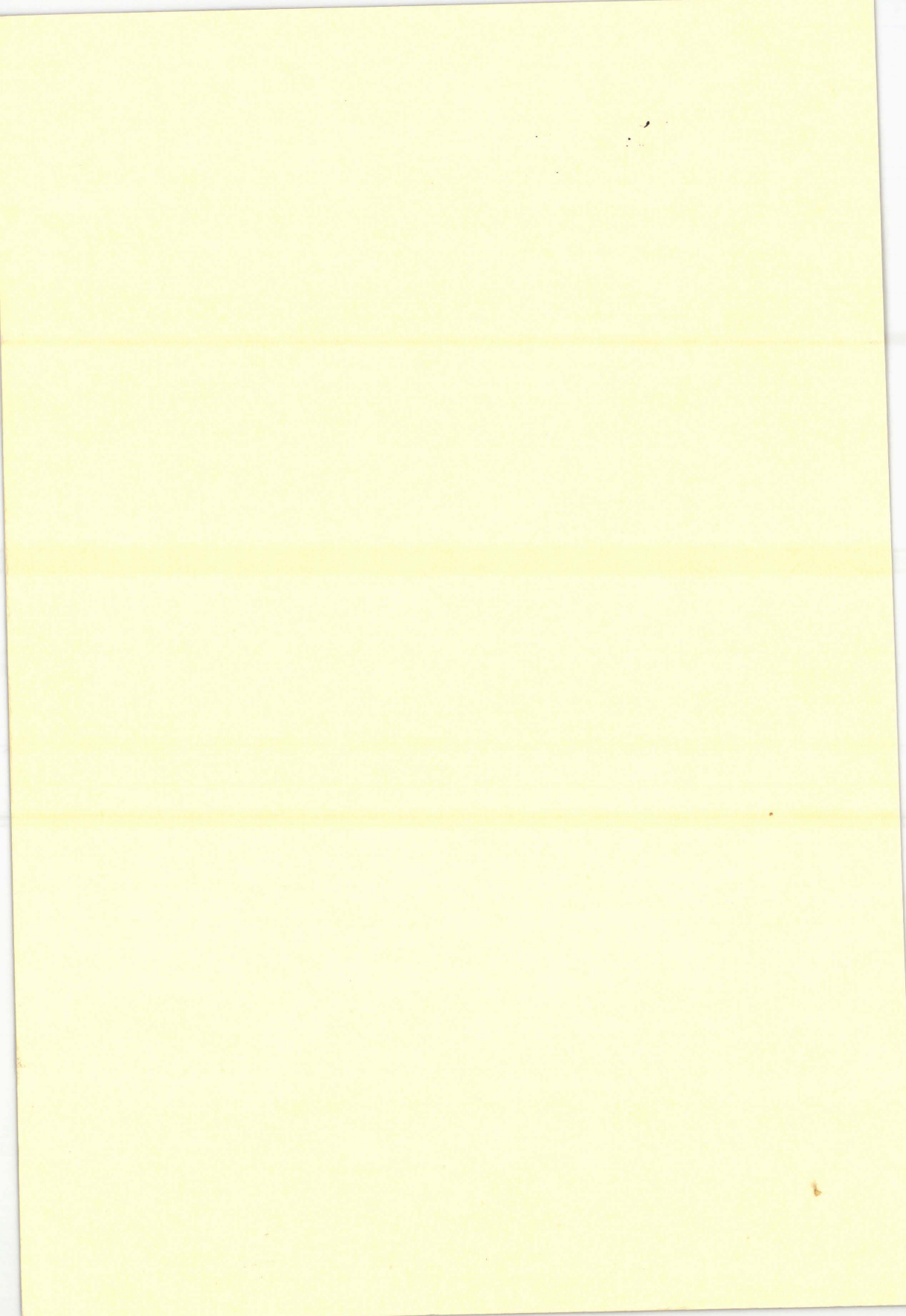
Président	:	Pierre JACOB, Université de Lille I
Rapporteurs	}	Alain TROGNON, Directeur-Adjoint à l'E.N.S.A.E.
		Marc HALLIN, Université Libre de Bruxelles
	}	Christian GOURIEROUX, Directeur de Thèse, Université de Lille I
Membres		Nazaré MENDES LOPES, Université de Coimbra (Portugal)
		Michel DELECROIX, Université de Lille I

SCD LILLE 1



D 030 254757 2

er 1988



N° d'ordre : 187

55376
1988
5

55376
1988
5

THESE

présentée à



L'UNIVERSITE DES SCIENCES ET TECHNIQUES DE LILLE FLANDRES ARTOIS

A RETOURNER LE

15 FEV. 1995

Président

Rapport

Membre

ra (Portugal)



0300158978



Je tiens à exprimer ma profonde gratitude à Monsieur le Professeur Christian GOURIEROUX. Il est clair que, sans son aide précieuse, ce travail n'aurait pu être entrepris ni terminé. Depuis le début et tout au long de ce travail, je ne sais si ce sont sa patience et sa confiance qu'il faut le plus admirer ou bien ses questions et suggestions.

Je remercie Monsieur le Professeur Pierre JACOB de l'honneur qu'il me fait en acceptant de présider le jury de cette thèse.

Madame le Professeur Nazaré MENDES LOPES et Messieurs les Professeurs Michel DELECROIX, Marc HALLIN et Alain TROGNON ont accepté de faire partie de ce jury. Qu'ils trouvent ici l'expression de ma sincère reconnaissance.

Je remercie vivement tous les membres, enseignants ou collègues, du Laboratoire de Statistique et Probabilités de Lille I de l'ambiance cordiale dont j'ai pu bénéficier pendant la réalisation de ce travail.

C'est avec beaucoup de gentillesse que Mesdames Arlette LENGAIGNE et Jacqueline JEAN ont dactylographié ce texte. Je tiens à les en remercier très cordialement.

Ce travail a bénéficié des soutiens du Gouvernement Français, du Secrétariat d'Etat au Plan et de l'Université de Coimbra.



P L A N

A - INTRODUCTION

ETUDE PROBABILISTE DES PROCESSUS MADE

B - UNE GENERALISATION DES PROCESSUS ARMA

1. Introduction.
2. Importance des relations de récurrence dans l'étude des processus ARMA.
3. Définition et principales propriétés des processus MADE.
4. Coefficients du développement moyenne mobile et forme des autocovariances.

C - PROPRIETES LIMITES DES MOMENTS EMPIRIQUES

1. Propriétés asymptotiques de la moyenne empirique (cas stationnaire).
2. Propriétés asymptotiques de la moyenne empirique (cas non stationnaire).
3. Propriétés asymptotiques des covariances empiriques (cas stationnaire).
4. Etude des moments empiriques d'ordre deux (cas non stationnaire).



QUELQUES APPLICATIONS STATISTIQUES

D - ESTIMATION DES PARAMETRES D'UN MODELE FRACTIONNAIRE

1. Estimateurs du (pseudo) maximum de vraisemblance des paramètres d'un modèle $MA(\infty)$.
2. Estimation par le maximum de vraisemblance d'un processus MADE(1).
3. Estimation par la méthode des moments.
4. Estimation par le maximum de vraisemblance d'un processus MADE quelconque.

E - TEST DU MULTIPLICATEUR DE LAGRANGE EN SERIES TEMPORELLES

1. Description du test.
2. Modèles linéaires approchés.

F - AGREGATION DE PROCESSUS AUTOREGRESSIFS D'ORDRE 1

1. Introduction.
2. Agrégation de processus autorégressifs d'ordre 1.
3. Un exemple d'agrégation.
4. Etude des corrélations.
5. Test de l'hypothèse d'homogénéité.

G - TESTS DE BRUIT BLANC

1. Introduction.
2. Construction de tests convergents.
3. Test de bruit blanc, lorsque l'alternative est fractionnaire.



Ⓐ

INTRODUCTION

I - TENDANCES

Disposant d'une suite d'observations indexées par le temps (X_t , $t = 1 \dots T$) et souhaitant analyser cette série chronologique, une première idée (après avoir analysé la nature des données) est peut être de faire son tracé graphique. A partir d'un tel tracé, on commence souvent à voir certaines caractéristiques de la série et donc à se poser diverses questions :

a) Y-a-t-il des variations lentes de la série ? dans des sens déterminés ? Ces variations se prolongent-elles durant une période de temps assez longue ?

b) Y-a-t-il des fluctuations autour du mouvement détecté en a) et quel est leur degré de régularité ?

c) La série paraît-elle ou non stationnaire, éventuellement après transformation des données ?

d) Y-a-t-il des points qui semblent ne pas respecter les "règles générales" vues en a), b), c) ?

.....

Dans ce travail nous nous sommes concentrés sur le premier type de régularité : évolution moyenne à long terme, encore appelée mouvement systématique, trend, tendance ...

L'intérêt d'une étude approfondie de la composante tendancielle est multiple :

i) Elle est importante pour séparer les phénomènes de long terme, des phénomènes d'ajustement autour de ce long terme.

ii) De part sa régularité, elle est une composante importante pour la détermination de bonnes prévisions.

iii) Sa détermination, puis son élimination peuvent aussi permettre une étude plus précise des autres composantes de la série : cycle, saisonnalité, points aberrants, ruptures ...

L'étude de la composante tendancielle passe nécessairement par une modélisation de celle-ci et c'est un exemple de telle modélisation qui est au centre de ce travail.

Il existe actuellement diverses façons plus ou moins précises, plus ou moins complexes de modéliser une tendance.

i) L'une des plus anciennes est directement liée à une méthode de détermination de la tendance, dite méthode d'ajustement. L'idée est de se donner a priori une forme fonctionnelle pour cette tendance, forme dépendant d'un petit nombre de paramètres $T(t;\beta)$. On retient généralement comme forme un polynôme de bas degré. La tendance est alors déterminée en choisissant une valeur du paramètre β par moindres carrés, c'est-à-dire en minimisant :

$$\text{Min}_{\beta} \sum_{t=1}^T (X_t - T(t,\beta))^2 .$$

Du point de vue de la modélisation, un modèle sous-jacent est :

$X_t = T(t,\beta) + \varepsilon_t$, où la suite $\varepsilon = (\varepsilon_t)$ est composée de variables centrées, non corrélées, de même variance.

Un tel modèle sous-jacent présente une certaine dissymétrie puisque la tendance est supposée représentable par une fonction déterministe,

alors que les autres composantes sont ici regroupées dans le terme aléatoire ϵ_t . (Signalons cependant que la saisonnalité est souvent introduite aussi sous forme déterministe, ce qui n'élimine pas le caractère dissymétrique de la formalisation).

ii) Cette approche présente ainsi l'inconvénient de devoir choisir une forme fonctionnelle valable pour l'ensemble de la période d'observation. Elle manque de ce fait de souplesse et notamment se comporte assez mal, lorsqu'il y a des ruptures de niveau ou de pente dans les séries. La méthode d'ajustement par polynômes locaux [voir Kendall-Stuart (1968) (The Advanced Theory of Statistics, Vol. 3, Griffin)] a pour but d'éviter cette difficulté. L'idée est d'ajuster la fonction $T(t, \beta)$ pour chaque date t à partir de données récentes ; plus précisément, notons $\hat{\beta}_t$ la solution de :

$$\text{Min}_{\beta} \sum_{j=t-M}^{t+M} [X_j - T(j, \beta)]^2 ,$$

la valeur de la tendance à la date t est prise égale à $T(t, \hat{\beta}_t)$. Ici encore la tendance est définie indirectement par l'intermédiaire de la méthode servant à la déterminer et il est difficile d'explicitier un modèle sous-jacent à la démarche précédente.

iii) Une troisième méthode de détermination de la tendance est celle implicitement utilisée dans certaines formes de prévision. Ainsi la prévision de X_t par lissage exponentiel est du type $\hat{X}_t = \lambda X_{t-1} + \lambda(1-\lambda) X_{t-2} + \dots + \lambda(1-\lambda)^{j-1} X_{t-j} + \dots$, $0 < \lambda < 1$ et consiste à faire une moyenne des valeurs passées avec des coefficients décroissants exponentiellement.

Ces coefficients sont de somme 1 de sorte que les constantes sont prévues sans erreur. Une démarche d'estimation de la tendance proche de cette idée de schéma exponentiel consiste à "lisser" la série en remplaçant la valeur X_t par une moyenne mobile pondérée :

$$T_t = \sum_{j=-M}^M \alpha_j X_{t-j} \quad , \quad \text{avec} \quad \sum_{j=-M}^M \alpha_j = 1 .$$

De telles moyennes mobiles sont par exemple utilisées pour estimer les tendances dans le programme de désaisonnalisation CENSUS X 11.

iv) Finalement le développement récent de la modélisation autorégressive-moyenne mobile a conduit à introduire des modèles pour prendre en compte certaines non stationnarités. Des non stationnarités de type polynomial sont par exemple introduites dans la formulation ARIMA (autorégressive-moyenne mobile intégrée), où la suite X_t , $t \in \mathbb{Z}$ est considérée comme une réalisation d'un processus satisfaisant une relation de récurrence du type :

$$(1-L)^d \phi(L) X_t = \Theta(L) \varepsilon_t \quad ,$$

où ε est un bruit blanc, ϕ et Θ des polynômes de degrés p et q , ayant leurs racines à l'extérieur du disque unité, d est un entier positif et L l'opérateur retard.

Cette formulation présente cependant certaines discontinuités, notamment dues au fait que l'exposant d est entier.

Celles-ci se répercutent alors au niveau des propriétés du modèle et entraînent de plus un manque de souplesse de celui-ci, expliquant qu'il ne s'ajuste pas toujours de façon adéquate aux données disponibles.

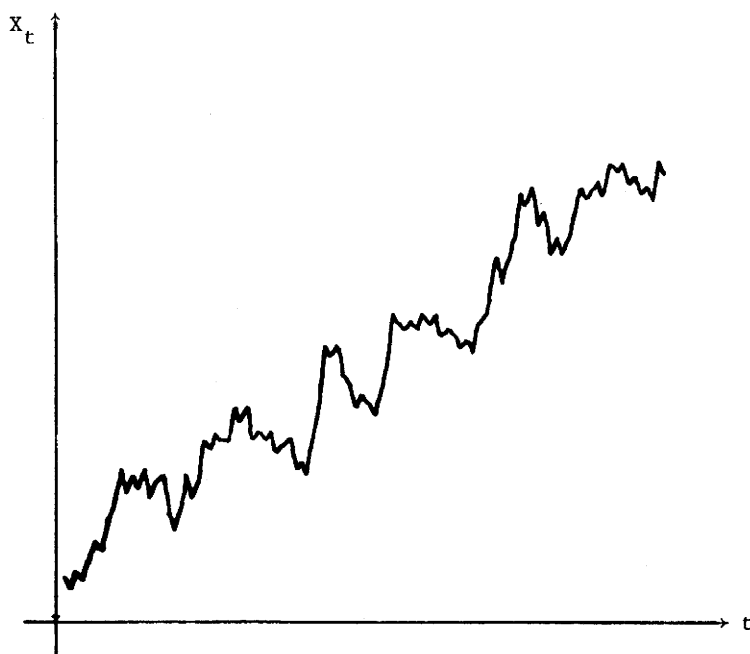
L'approche proposée ici pour modéliser la tendance passe par la considération d'un nouveau processus construit à partir de la série initiale (X_t) . Il est défini par :

$$\tilde{X}_T(r) = \frac{X_{[Tr]}}{T^\alpha}, \alpha > 0, 0 \leq r \leq 1$$

où $[\]$ désigne la partie entière.

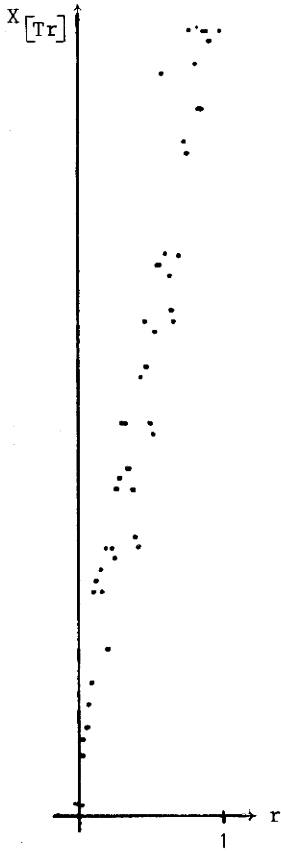
L'exemple suivant permet de visualiser l'effet de la transformation qui est proposée.

i) La série initiale



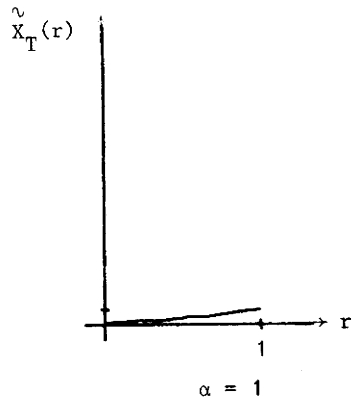
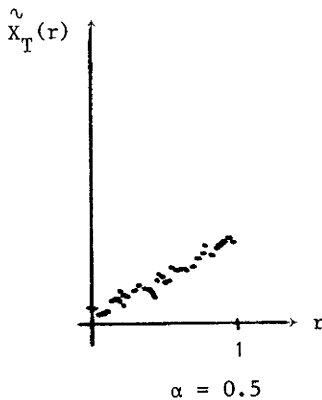
On peut parler de l'existence d'une tendance (en moyenne) croissante a peu près linéaire.

ii) La représentation graphique de la série $X_{[Tr]}$, $0 \leq r \leq 1$,



montre qu'en partant de X_t
on fait une modification
de l'échelle des temps,
de façon à ramener toute
la série dans $t \in [0,1]$

iii) La série $\tilde{X}_T(r) = \frac{X_{[Tr]}}{T^\alpha}$ ($\alpha = 0.5$, $\alpha = 1$)



On voit que la normalisation faite avec $\alpha = 1$ conduit à un tracé assez régulier à peu près linéaire en r ce qui est évidemment compatible avec la première impression de l'existence d'une tendance linéaire. Ce tracé présente l'avantage de diminuer beaucoup les fluctuations autour de la composante tendancielle et donc de mieux mettre en évidence cette dernière.

Dans la suite nous dirons que le processus (X_t) présente une composante tendancielle d'ordre α si le processus

$$\{\tilde{X}_T(r) = \frac{X_{[Tr]}}{T^\alpha}, r \in [0,1]\} \text{ converge en distribution vers un processus}$$

$\{X_\infty(r), r \in [0,1]\}$ non dégénéré et continu en moyenne quadratique.

Les processus \tilde{X}_T et X_∞ sont des "fonctions aléatoires" de $D[0,1]$, l'espace des fonctions réelles définies sur $[0,1]$ qui sont continues à droite en chaque point et qui ont des limites à gauche finies. La convergence en distribution signifie la convergence des lois de \tilde{X}_T vers celle de X_∞ .

Cette modélisation des tendances permet de retrouver la partie dominante du modèle sous-jacent à la modélisation ; comme

$$\left[\frac{X_{[Tr]}}{T^\alpha} \right] \longrightarrow [X_\infty(r)] \text{ et comme la limite est continue en moyenne quadratique}$$

on a localement (c'est-à-dire, pour $t \in [T(r-\epsilon), T(r+\epsilon)]$, ϵ petit) la valeur de X_t qui est peu différente de $t^\alpha X_\infty(r)$ au sens où la différence entre X_t et cette valeur est proche de zéro. Ainsi la série initiale est "localement" assimilable à un monôme At^α où A dépend à la fois de r et de l'état de la nature.

La formalisation précédente recouvre un certain nombre de modèles usuels.

i) Un premier exemple est celui où X_t est déterministe :
 $X_t = at^c$, $c > 0$. Le processus X_t présente une tendance en moyenne
 d'ordre c :

$$\tilde{X}_T(r) = a \left(\frac{[Tr]}{T} \right)^c \xrightarrow{D} X_\infty(r) = a r^c, \quad 0 \leq r \leq 1.$$

Le processus limite garde le caractère déterministe du processus de départ.

ii) Un deuxième exemple est fourni par le théorème de Donsker
 [Billingsley, P. (1968) Convergence of probability measures, Wiley, p. 137].

Si ε_t , $t \geq 1$ sont des variables indépendantes de même loi, de moyenne
 nulle et variance σ^2 , $0 < \sigma^2 < \infty$ et si $X_t = \sum_{j=1}^t \varepsilon_j$ on a :

$$\tilde{X}_T(r) = \frac{1}{\sigma\sqrt{T}} X_{[Tr]} = \sum_{j=1}^{[Tr]} \frac{\varepsilon_j}{\sigma\sqrt{T}}$$

et $\tilde{X}_T \xrightarrow{D} B$, où B est un processus du mouvement brownien sur $[0,1]$.

Le processus $X_t = \sum_{j=1}^t \varepsilon_j$ présente ici une composante tendancielle
 d'ordre $1/2$: $X_t \sim \sigma \left(\frac{t}{r} \right)^{1/2} B(r)$. Il s'agit d'une tendance en variance
 puisque $E X_t = 0$, $V X_t = \sigma^2 t$.

Une autre représentation du processus $X_t = \sum_{j=1}^t \varepsilon_j$ en termes
 d'équation aux différences est :

$$(1-L) X_t = \varepsilon_t, \quad t \geq 1, \quad X_0 = 0.$$

De façon plus générale nous verrons que tout processus satisfaisant une condition du type $(1-L)^d X_t = \varepsilon_t$, $t \geq 1$, $X_t = 0$ $t \leq 0$, $d > \frac{1}{2}$, admet une composante tendancielle au sens précédent. Le degré de différenciation d et l'ordre α de la composante tendancielle sont liés par la relation $\alpha = d - \frac{1}{2}$.



II - PLAN DE LA THESE

1) Une généralisation des processus ARMA.

Dans la première partie de ce travail nous nous intéressons aux processus stationnaires réguliers dont les coefficients de la représentation moyenne mobile infinie $(a_j, j \in \mathbb{N})$ satisfont une équation de récurrence linéaire à coefficients polynomiaux de degré un :

$$(A.1) \quad (-b_{-1} + c_0 j) a_j + [-b_0 + c_1(j-1)] a_{j-1} \\ + \dots + [-b_{k-1} + c_k(j-k)] a_{j-k} = 0, \quad \forall j \in \mathbb{Z},$$

avec $b_j, c_j \in \mathbb{R}$ et $(b_{-1}, c_0), (b_{k-1}, c_k)$ non nuls.

Cette approche a été suggérée par le fait que diverses équations de récurrence à coefficients constants jouent un rôle important pour la modélisation autorégressive-moyenne mobile [ARMA] : équations satisfaites par les coefficients moyennes mobiles, par les coefficients autorégressifs, par la fonction d'autocovariance ... Il était donc naturel d'étendre ce type de caractérisation en élargissant la classe d'équations aux différences considérées.

Sous les hypothèses :

- i) $c_k \neq 0$,
- ii) le polynôme $C(z) = \sum_{j=0}^k c_j z^j$ a toutes ses racines distinctes,

nous donnons des conditions nécessaires et suffisantes pour qu'il existe un processus stationnaire régulier compatible avec l'équation (A.1). De tels processus sont appelés MADE d'ordre k [Moving Average (satisfying) a

Difference Equation]. Nous présentons alors les principales propriétés de ces processus : forme de la densité spectrale et comportement de celle-ci pour les basses fréquences, expression des autocovariances, détermination des équations aux différences satisfaites par la suite des autocovariances, existence d'une représentation autorégressive infinie ...

La classe de processus ainsi définie comprend quelques cas particuliers importants, que nous étudions de façon plus détaillée : processus MADE(1), MADE(2) correspondant aux petits ordres, processus à mémoire longue obtenus lorsque les autocovariances décroissent lentement [Granger (1980), Granger-Joyeux (1980), Hosking (1981), Geweke-Porter-Hudak (1983) ...].

2) Propriétés limites des moments empiriques.

Le chapitre suivant est consacré à l'étude des propriétés limites des moyennes empiriques, autocovariances empiriques, périodogramme ... d'un processus MADE. De façon à simplifier la présentation, nous nous sommes restreint au cas de processus MADE(1).

Ceux-ci peuvent se mettre sous la forme :

$$(A.2) \quad (1-\alpha L)^d X_t = \varepsilon_t, \quad \text{où } L \text{ désigne l'opérateur}$$

retard, où α et d sont deux paramètres réels et $\varepsilon = (\varepsilon_t)$ un bruit blanc. Afin que le processus X soit stationnaire les paramètres doivent être contraints par :

$$\begin{aligned} & |\alpha| > 1, \quad d \text{ entier négatif,} \\ \text{ou} & \quad |\alpha| < 1, \quad d \text{ quelconque,} \\ \text{ou} & \quad |\alpha| = 1, \quad d < \frac{1}{2}. \end{aligned}$$

En fait on voit facilement que, lorsque ces contraintes ne sont pas satisfaites, on peut encore définir un processus X satisfaisant (A.2.), mais que ce processus est alors non stationnaire. L'étude asymptotique est faite en séparant systématiquement les cas stationnaire et non stationnaire.

a) Cas stationnaire

Nous commençons par rappeler quelques résultats classiques sur la convergence en moyenne quadratique (vitesse en $T^{+1/2}$, où T est le nombre d'observations) et sur la convergence en loi (normalité asymptotique) de la moyenne empirique. Les conditions d'application de ces résultats sont "grosso-modo" des conditions de sommabilité de la suite des autocovariances ou de continuité de la densité spectrale en zéro. On vérifie immédiatement qu'elles ne sont pas satisfaites pour toutes les valeurs admissibles des paramètres. En particulier ces propriétés classiques ne s'appliquent pas, lorsque la décroissance des autocovariances est trop lente. Dans notre exemple ceci se produit lorsque $\alpha = \pm 1$ et que l'exposant d est "assez" proche de $\frac{1}{2}$. Ces divers cas sont étudiés en détail dans ce chapitre. Par exemple l'étude de la convergence de la moyenne empirique lorsque $\alpha = 1$ et $d < \frac{1}{2}$ se révèle assez intéressante. En examinant directement l'expression de sa variance (ou en raisonnant dans le domaine des fréquences), on voit qu'il y a convergence en moyenne quadratique, que celle-ci a lieu à la vitesse $T^{-d+1/2}$ si $-\frac{1}{2} < d < \frac{1}{2}$ et T^{+1} , si $d < -\frac{1}{2}$, donc à une vitesse différente de la vitesse usuelle. Bien qu'il soit aussi possible d'étendre les résultats de convergence en moyenne quadratique, il n'en va pas de même concernant la normalité asymptotique. La moyenne empirique convenablement

normalisée ne converge généralement pas en loi vers une loi normale. Des résultats de ce type peuvent cependant être obtenus en supposant la normalité des erreurs. Nous présentons un tel résultat de convergence dans $D[0,1]$, résultat dû à Taqqu (1975).

Le même type d'approche peut être suivi pour l'étude des autocovariances empiriques. Cette étude est cependant rendue plus difficile vue que même sous l'hypothèse de normalité des erreurs, les autocovariances empiriques convenablement normalisées tendent en distribution vers une loi non normale. Cette loi est reliée aux processus dits de Rosenblatt [Rosenblatt (1956), (1961), Fox-Taqqu (1981)].

b) Cas non stationnaire

L'effet de la non stationnarité a été étudié initialement dans le cas de la marche aléatoire et a débouché sur le théorème de Donsker. Ce type de résultat est ici étendu au cas de processus fractionnaires de la forme :

$$X_t = (1-L)^{-d} u_t, \quad d > \frac{1}{2},$$

où le processus u admet la représentation $\phi(L) u_t = \theta(L) \tilde{\varepsilon}_t$,

$$\tilde{\varepsilon}_t = \begin{cases} \varepsilon_t, & t \geq 0 \\ 0 & \text{sinon} \end{cases} \quad \varepsilon = (\varepsilon_t, t \in \mathbb{Z}) \text{ étant un bruit blanc de variance } \sigma^2.$$

Un tel processus admet une écriture moyenne mobile infinie du type :

$$X_t = \sum_{\tau=1}^t \lambda(t-\tau) u_\tau ,$$

où la fonction réelle λ est équivalente asymptotiquement à : $\lambda(t) \sim t^\mu$,
si t tend vers l'infini, avec $\mu > -\frac{1}{2}$.

C'est cette écriture qui est utilisée pour obtenir le comportement asymptotique du processus X convenablement normalisé. Plus précisément un résultat récent d'invariance fonctionnelle de Akonom-Gouriéroux (1987) montre que le processus indexé par $r \in [0,1]$:

$$\frac{X_{\lfloor Tr \rfloor}}{T^{\mu+1/2}} , \quad r \in [0,1] ,$$

converge en loi [au sens de $\mathcal{D}[0,1]$] vers le processus

$$\frac{\Theta(1)}{\Phi(1)} \frac{\sigma}{\Gamma(d)} \int_0^r (r-s)^\mu dB(s) , \quad \text{où } B \text{ désigne un mouvement brownien.}$$

On déduit alors de cette propriété du processus initial les comportements asymptotiques de la moyenne empirique $\frac{1}{T^{d-1/2}} \bar{X}_T$,
des autocovariances empiriques $T^{-2d+1} \hat{\gamma}_T(0)$, $T^{-2d+1} \hat{\gamma}_T(1)$

On donne également des résultats sur la fonction périodogramme au voisinage de la fréquence nulle et en particulier on établit la convergence en loi de $\frac{1}{T^{2d}} I_T(\frac{\omega}{T})$, où I_T désigne le périodogramme.

3) Estimation des paramètres d'un modèle fractionnaire.

Le but de ce chapitre est d'étudier l'estimation par la méthode du maximum de vraisemblance des paramètres d'un modèle fractionnaire stationnaire :

$$X_t = \frac{K_0 \prod_{k=1}^{\infty} (1 - \alpha_k L)^{\delta_k}}{K_1 \prod_{k=1}^{\infty} (1 - \beta_k L)^{d_k}} \varepsilon_t, \text{ avec } |\alpha_k| < 1, |\beta_k| < 1$$

et $\varepsilon = (\varepsilon_t)$ un bruit blanc.

On peut dans ce type de modèle, utiliser les résultats généraux valables pour l'estimation de paramètres apparaissant dans un modèle de séries temporelles stationnaires. Nous commençons par rappeler brièvement ces divers résultats.

i) On donne les diverses écritures possibles de la log - vraisemblance et de ses approximations : écritures dans le domaine des temps, dans le domaine des fréquences, écritures conditionnelles.

ii) Si $L_T^*(x, \theta)$ désigne cette vraisemblance l'estimateur est alors défini comme solution de :

$$\text{Max}_{\theta} \text{Log } L_T^*(x, \theta),$$

ou de façon équivalente comme solution de :

$$\text{Min}_{\theta} K_T^*(P/P_0) = \frac{1}{T} \text{Log } L_T^*(x; \theta_0) - \frac{1}{T} \text{Log } L_T^*(x; \theta),$$

où θ_0 et P_0 désignent la vraie valeur du paramètre et la vraie loi des observations. On est alors amené pour étudier l'existence asymptotique de l'estimateur et sa convergence à étudier la fonction limite :

$$K(P/P_0) = \lim_T K_T^*(P/P_0) \quad (\text{limite presque sûre}).$$

iii) On donne diverses expressions de cette fonction limite qui s'interprète comme une mesure de proximité entre lois de processus stationnaires et est l'analogie de l'information de Kullback classique. En particulier

on décrit diverses décompositions de cette information fréquence par fréquence et aussi en fonction des valeurs des variables conditionnantes.

iv) Une fois établie la convergence de l'estimateur du maximum de vraisemblance $\hat{\theta}_T$, on voit que celui-ci satisfait asymptotiquement les conditions du premier ordre :

$$\frac{\partial \text{Log } L_T^*}{\partial \theta} (x, \hat{\theta}_T) = 0 .$$

On est donc conduit à étudier les propriétés asymptotiques du vecteur des scores $\frac{1}{\sqrt{T}} \frac{\partial \text{Log } L_T^*}{\partial \theta} (x; \theta_0)$. En particulier, on détermine diverses expressions de sa variance asymptotique, analogue de la matrice d'information de Fisher usuelle et on vérifie que cette matrice apparaît naturellement si on approche au voisinage de la vraie loi l'information de Kullback.

v) Finalement on montre la normalité asymptotique et l'efficacité asymptotique de l'estimateur du maximum de vraisemblance et on applique ces résultats au cas du processus fractionnaire.

4) Test du multiplicateur de Lagrange en séries temporelles.

Les développements asymptotiques effectués dans le paragraphe 3 servent aussi de base à diverses procédures de test. L'une des plus classiques est celle du multiplicateur de Lagrange que nous appliquerons à divers problèmes dans les chapitres suivants. On décrit cette procédure dans le cas général d'un modèle de série temporelle stationnaire.

On part d'un processus stationnaire admettant des représentations moyenne-mobile et autorégressive :

$$\phi_{\theta}(L) X_t = \varepsilon_t \iff X_t = A_{\theta}(L) \varepsilon_t$$

où $\theta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ et on cherche à établir des procédures de test de l'hypothèse $H_0 = \{\beta = 0\}$ contre $H_1 = \{\beta \neq 0\}$ à partir de l'idée du multiplicateur de Lagrange qui consiste à accepter H_0 lorsque le score estimé sous H_0 est proche de zéro.

En utilisant des résultats du paragraphe précédent, on déduit la loi du score estimé sous l'hypothèse nulle et on construit la statistique de test associée.

Cette statistique admet diverses écritures selon la forme du score retenue ; en particulier à partir de la forme conditionnelle on déduit une interprétation de la statistique en terme de régression. On peut dans certains cas obtenir la même statistique de test pour tester l'hypothèse nulle H_0 contre une alternative H_1 ou pour tester H_0 contre une autre alternative H_1^* . Ceci se produit, lorsque les deux alternatives sont localement équivalentes au voisinage de l'hypothèse nulle.

Considérant le cas d'un modèle décomposable :

$$\phi_{\theta}(L) = \phi_{1,\alpha}(L) \phi_{2,\beta}(L) ,$$

on utilise ce concept d'alternative locale pour obtenir d'autres interprétations du test du multiplicateur de Lagrange. Celles-ci consistent essentiellement à remplacer le modèle initial pour un autre modèle linéarisé par rapport au paramètre d'intérêt β . Ici encore ceci permet le calcul de la statistique de test à partir uniquement de régressions linéaires.

5) Agrégation de processus autorégressifs d'ordre 1.

Dans ce chapitre nous considérons une suite de processus $(X_{i,t}, i = 1, 2, \dots)$ admettant une écriture autorégressive d'ordre 1 pour $t \geq 0$.

Nous commençons par déterminer le modèle de série temporelle satisfait par la série agrégée ce qui permet de caractériser toutes les séries qui peuvent s'interpréter comme des agrégats de processus AR(1).

Nous étudions ensuite, de façon plus détaillée, le cas où la loi du coefficient de régression est une loi bêta : la série agrégée satisfait une écriture hypergéométrique contenant comme cas particulier certains processus fractionnaires ; elle peut être non stationnaire (même si toutes les séries au niveau désagrégé le sont) ; elle peut être stationnaire avec les coefficients du développement moyenne mobile décroissant lentement ...

L'étude des autocorrélations de la série agrégée et des corrélations temporelles des séries désagrégées montre que les premières sont toujours supérieures à la moyenne des secondes si la loi du coefficient de régression ne charge que les valeurs positives. Il y a donc un biais d'agrégation.

Un test de l'hypothèse d'homogénéité, fondé sur le principe du multiplicateur de Lagrange, est proposé dans le cadre du modèle obtenu avec loi bêta ; on montre enfin que l'oubli d'hétérogénéité conduit à des estimateurs biaisés.

6) Tests de bruit blanc

L'approche du multiplicateur de Lagrange est développée pour tester l'hypothèse qu'un processus $(X_t, t \in \mathbb{Z})$ est un bruit blanc indépendant. On suppose que sous l'hypothèse générale le processus admet une représentation moyenne-mobile infinie inversible $\phi_\theta(L) X_t = \varepsilon_t$ où θ est un paramètre scalaire et tel que l'hypothèse nulle s'écrit $H_0 = (\theta=0)$.

Des tests convergents sont proposés en exploitant le fait que, sous l'hypothèse nulle, il n'y a pas de corrélation entre X_t et une combinaison linéaire des valeurs passées X_{t-1}, X_{t-2}, \dots

Le comportement asymptotique de la statistique de test sous l'hypothèse de bruit blanc et sous des suites d'alternatives locales est étudié ; on déduit les formes de la statistique conduisant aux tests localement les plus puissants et on montre que le test fondé sur le score coïncide avec le test optimal.

Les puissances locales correspondantes à deux suites d'alternatives locales sont aussi comparées.

On s'intéresse ensuite au test de l'hypothèse de bruit blanc lorsque le modèle alternatif est fractionnaire : $(1-\alpha L)^d X_t = \varepsilon_t$. On peut espérer distinguer les alternatives autorégressives des alternatives moyenne mobile à travers le signe de d ($d > 0$ ou $d < 0$). On sera conduit à proposer une démarche de type minimax pour prendre en compte les diverses alternatives locales, dû à l'existence d'un problème d'identifiabilité des paramètres α et d sous l'hypothèse nulle.

B

UNE GENERALISATION DES PROCESSUS ARMA



Une généralisation des processus ARMA

Esmeralda GONÇALVES *

RÉSUMÉ. - Le but de cet article est d'introduire une classe de modèles généralisant les modèles ARMA usuels et permettant de résoudre au moins deux difficultés rencontrées avec cette dernière modélisation : la phase d'identification c'est-à-dire d'estimation des degrés autorégressif et moyenne mobile et les phénomènes de sur et sous-identification. Les processus étudiés dits processus MADE sont des processus stationnaires, dont le développement moyenne mobile infinie en fonction du bruit, a des coefficients satisfaisant une équation de récurrence linéaire à *coefficients polynomiaux de degré 1*.

Nous discutons l'existence et l'unicité de tels processus, vérifions qu'ils contiennent comme cas particuliers les processus ARMA ainsi que les processus à mémoire longue étudiés par Granger. Nous donnons ensuite divers résultats concernant la fonction d'autocovariance, la densité spectrale...

A Generalization of ARMA Processes

ABSTRACT. - The aim of this paper is the introduction of a class of models generalizing the usual ARMA models and giving possible solutions for two difficulties encountered in ARMA modeling, the identification step, i. e. the estimation of the autoregressive and moving average orders, and the phenomenas of under and over-differencing. The studied processes, called MADE processes (Moving Average satisfying a Difference Equation) are regular stationary processes, whose infinite moving average expansion has coefficients satisfying a difference equation with coefficients which are polynomials of degree 1.

We discuss the existence and the uniqueness of such processes, we verify that this class contains as particular cases the ARMA processes and the long memory models studied by Granger. Then we give a number of results concerning the autocovariance function, the spectral density function...



1 Introduction

Les aspects dynamiques des modèles macroéconomiques : anticipations, description des processus d'ajustement, liens de causalité entre variables, séparation court terme-long terme... ont donné lieu ces dernières années à de nombreuses études aussi bien théoriques que pratiques. L'étude de ces aspects dynamiques est habituellement menée en utilisant des modélisations autorégressives-moyennes mobiles soit sur les processus observables exogènes ou endogènes, soit sur les processus d'erreur. Cette modélisation consiste à supposer que la valeur présente d'une variable est fonction linéaire de ses valeurs passées et de termes d'erreurs; plus précisément on a une relation du type :

$$X_t - \varphi_1 X_{t-1} \dots - \varphi_p X_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} \dots - \theta_q \varepsilon_{t-q}$$

où (ε_t) est une suite de variables non observables centrées non corrélées et de même variance (bruit blanc). Ce type de modèle essentiellement descriptif s'est avéré très souple d'utilisation. Cependant on peut dans son emploi rencontrer certaines difficultés aussi bien au niveau de son estimation, qu'au niveau de l'adéquation avec la réalité. Deux exemples classiques de telles difficultés sont les suivants :

i) Un modèle autorégressif-moyenne mobile contient plusieurs paramètres : les ordres p et q , les coefficients autorégressifs et moyennes mobiles $\varphi_1 \dots \varphi_p$, $\theta_1 \dots \theta_q$ et la variance σ^2 du bruit blanc. Si l'estimation des paramètres à valeurs réelles φ_i , θ_j , σ^2 ne pose pas de difficultés particulières, il n'en va pas de même de l'estimation des ordres p et q . Ces deux paramètres p et q sont à valeurs entières et, bien que diverses procédures théoriques d'estimation de p et q aient été proposées dans la littérature, l'absence de connaissance des propriétés de ces méthodes conduit souvent le praticien à les estimer de façon très empirique.

ii) Une seconde difficulté souvent rencontrée dans l'utilisation des modèles ARMA concerne la description de séries non stationnaires. On considère généralement qu'il suffit de travailler en accroissement (différentiation) pour se ramener à un cas stationnaire, où le modèle ARMA sera applicable. Cette démarche peut cependant se révéler inadéquate si la série initiale bien que stationnaire présente des corrélations « assez fortes » entre passé et futur mêmes lointains, par exemple décroissant vers zéro à une vitesse $\frac{1}{h}$ où h est le décalage entre passé et futur. Travailler en accroissement conduit alors à un phénomène classique appelé surdifférentiation et qui tend à rendre inopérante la démarche habituelle.

Dans cet article nous nous proposons de décrire une classe de modèles, contenant comme cas particuliers les modèles ARMA, et dont l'utilisation peut donner certaines réponses aux difficultés précédentes. L'idée de base est simplement de rendre réels les deux ordres p et q , ce qui revient à introduire des puissances fractionnaires dans les écritures des polynômes autorégressif et moyenne mobile.

2 Importance des relations de récurrence dans l'étude des processus ARMA

L'analyse des séries temporelles a connu un grand développement depuis la parution du livre de BOX et JENKINS [1970], où étaient rassemblées les principales propriétés des processus stationnaires autorégressif-moyenne mobile (ARMA) et décrites des méthodes adaptées d'estimation des paramètres. La simplicité d'utilisation de la modélisation ARMA s'explique en grande partie par la forme particulière de la fonction de transfert.

Considérons en effet un processus stationnaire $X = (X_t, t \in \mathbb{Z})$ satisfaisant l'équation de récurrence :

$$(1) \quad \varphi(L) X_t = \theta(L) \varepsilon_t,$$

où : $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ est un bruit blanc du second-ordre, de variance σ^2 , où L désigne l'opérateur retard, et où φ et θ sont deux polynômes de degrés respectifs p et q , admettant des racines strictement à l'extérieur du disque unité et n'ayant pas de racine commune.

Le processus X défini par la relation (1) admet une représentation moyenne mobile infinie :

$$(2) \quad X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} = A(L) \varepsilon_t,$$

où la fonction de transfert : $A(z) = \sum_{j=0}^{\infty} a_j z^j$ est donnée par $A(z) = \frac{\theta(z)}{\varphi(z)}$. Les processus ARMA sont ainsi caractérisés par le fait que la fonction de transfert coïncide avec une fraction rationnelle. Cette propriété peut être réécrite de diverses façons. On peut par exemple remarquer que le processus X admet aussi une représentation autorégressive infinie :

$$(3) \quad \sum_{j=0}^{\infty} \pi_j X_{t-j} = \pi(L) X_t = \varepsilon_t,$$

où $\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{1}{A(z)} = \frac{\varphi(z)}{\theta(z)}$ est aussi égal à une fraction rationnelle on utilise le même type de propriété qui est satisfaite par la fonction génératrice d'autocovariance :

$$(4) \quad G(z) = \sum_{h=-\infty}^{+\infty} \gamma(h) z^h = \sigma^2 A(z) A\left(\frac{1}{z}\right) = \sigma^2 \frac{\theta(z)\theta(1/z)}{\varphi(z)\varphi(1/z)},$$

avec $\gamma(h) = \text{Cov}(X_t, X_{t+h})$.

Au lieu de caractériser la modélisation ARMA par l'intermédiaire de fractions rationnelles, on peut aussi introduire des définitions équivalentes

fondées sur des équations de récurrence linéaire. Ainsi on déduit directement de l'écriture : $A(z)\varphi(z) = \theta(z)$ et de l'identification terme à terme des coefficients que $A(z)$ est une fraction rationnelle si et seulement si la suite des coefficients $(a_j, j \in \mathbb{N})$ satisfait une équation de récurrence linéaire à coefficients constants (correspondant aux coefficients du polynôme φ), avec un second membre (correspondant aux coefficients du polynôme θ).

La même propriété est évidemment satisfaite pour les coefficients du développement autorégressif infini et pour la suite des covariances. Cette seconde caractérisation est à la base d'un grand nombre de méthodes d'estimation ou d'identification des processus ARMA.

Des approches similaires peuvent être développées à partir des transformées bien choisies de la fonction de transfert. A notre connaissance la remarque qui suit et qui est le cœur de cet article a été faite pour la première fois sous une forme différente par SPOLIA-CHANDER-O'CONNOR [1980]. Considérons la dérivée logarithmique de la fonction de transfert $A(z)$. Elle est donnée par :

$$(5) \quad \frac{1}{A(z)} \frac{dA(z)}{dz} = \frac{(d\theta(z)/dz)\varphi(z) - (d\varphi(z)/dz)\theta(z)}{\theta(z)\varphi(z)}$$

et coïncide donc aussi avec une fraction rationnelle irréductible :

$$(6) \quad \frac{1}{A(z)} \frac{dA(z)}{dz} = \frac{B(z)}{C(z)}$$

Comme φ et θ sont sans racines communes, des simplifications entre le numérateur $\frac{d\theta(z)}{dz}\varphi(z) - \frac{d\varphi(z)}{dz}\theta(z)$ et $\theta(z)\varphi(z)$ peuvent seulement être faites, lorsque θ ou φ admettent des racines multiples. Dans la forme irréductible (6) le polynôme C apparaissant au dénominateur admet donc des racines distinctes obtenues en prenant la réunion des racines de φ et des racines de θ .

Le polynôme B du numérateur admet un degré inférieur strictement à celui de C . Remarquons que cette écriture (5) fait jouer un rôle plus symétrique aux polynômes autorégressif et moyenne mobile.

L'équation (6) a une écriture équivalente en terme d'équation aux différences. Introduisons les écritures explicites des polynômes B et C :

$$B(z) = b_0 + b_1 z + \dots + b_{K-1} z^{K-1},$$

$$C(z) = c_0 + c_1 z + \dots + c_K z^K, \quad c_K \neq 0,$$

et de la série A :

$$A(z) = \sum_{j=-\infty}^{+\infty} a_j z^j \quad \text{avec } a_j = 0, \quad \forall j < 0;$$

on déduit de l'égalité : $C(z) \frac{dA(z)}{dz} - A(z)B(z) = 0$ que les coefficients $a_j, j \in \mathbb{Z}$ sont tels que :

$$(7) \quad (c_0 j) a_j + [-b_0 + c_1(j-1)] a_{j-1} + \dots + [-b_{K-1} + c_K(j-K)] a_{j-K} = 0, \quad \forall j.$$

Ils satisfont donc aussi une équation de récurrence linéaire sans second membre, dont les coefficients sont des polynômes de degré 1 en j . Il est clair qu'une telle équation (qui existe aussi pour les suites (π_j) et $(\gamma(h))$) peut servir à construire de nouvelles méthodes d'estimation et d'identification des processus ARMA.

La remarque précédente incite à étudier les processus moyennes mobiles infinies dont la fonction de transfert satisfait une équation du type (6) avec des polynômes B et C sans racines communes, tels que $d^\circ B \leq d^\circ C - 1$ et un polynôme C, dont toutes les racines sont distinctes. Dans la suite ces processus sont appelés MADE(K) (Moving Average (satisfying) a Difference Equation), où $K = d^\circ C$.

Il est facile de voir que l'ensemble de ces processus contient strictement la classe des processus ARMA.

Considérons pour simplifier le cas où $K = d^\circ C = 1$. Cette classe comprend les processus ARMA de la forme :

$$(1 - \alpha L)^d X_t = \varepsilon_t \quad \text{avec } d \in \mathbb{N}^*,$$

ou ceux de la forme :

$$X_t = (1 - \alpha L)^d \varepsilon_t \quad \text{avec } d \in \mathbb{N}^*.$$

Pour ces processus l'équation (6) s'écrit :

$$\frac{1}{A(z)} \frac{dA(z)}{dz} = \frac{-\alpha\delta}{1 - \alpha z} \quad \text{avec } \delta \in \mathbb{Z}^*.$$

Il est clair que des processus sans représentation ARMA peuvent être obtenus en prenant une valeur δ non entière. Considérer la classe MADE(1) permet ainsi de disposer d'un sur-modèle à la fois pour les AR(1), les MA(1) (et les processus autorégressifs purs ou moyennes mobiles pures admettant une seule racine, avec un ordre de multiplicité supérieur à 1).

Dans la section 3, nous étudions les principales propriétés des processus MADE. Nous commençons par établir des conditions nécessaires et suffisantes portant sur les coefficients des polynômes B et C pour qu'il existe un processus stationnaire dont les coefficients du développement moyenne mobile satisfassent la relation (7). Puis nous donnons les formes des représentations moyenne mobile infinie et autorégressive infinie, l'expression de la densité spectrale et explicitons les équations différentielles satisfaites par les diverses fonctions génératrices.

Dans la section 4, nous nous intéressons aux formes des coefficients du développement moyenne mobile et des autocovariances. Nous donnons leurs expressions complètes pour quelques modèles simples, en particulier pour un modèle correspondant à $B(z) = b_0$, $C(z) = 1 - \alpha z$. Ce cas limite permet de retrouver les modèles dits à mémoire longue (GRANGER [1980], GRANGER-JOYEUX [1980], HOSKING [1981], GEWEKE-PORTER-HUDAK [1983] ...). Lorsque les modèles sont d'ordre trop élevé, il est évidemment difficile d'obtenir des formes explicites et nous présentons diverses procédures numériques permettant leur détermination. Diverses démonstrations et quelques propriétés utiles des fonctions hypergéométriques sont données dans les annexes.

3 Définition et principales propriétés des processus MADE

3.a. Définition

Dans la suite de l'article un *processus MADE* est un processus stationnaire régulier, à valeurs réelles, dont la décomposition moyenne-mobile infinie :

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} \left(\text{avec } \varepsilon \text{ bruit blanc, de variance } \sigma^2, a_j \in \mathbb{R}, a_0 = 1 \text{ et } \sum_{j=0}^{\infty} a_j^2 < \infty \right)$$

admet des coefficients satisfaisant une équation de récurrence

linéaire du type :

$$(8) \quad \begin{aligned} &(-b_{-1} + c_0 j) a_j + [-b_0 + c_1(j-1)] a_{j-1} \\ &\quad + \dots + [-b_{K-1} + c_K(j-K)] a_{j-K} = 0 \end{aligned}$$

$\forall j \in \mathbb{Z}$ avec $b_j, c_j \in \mathbb{R}$ et $(b_{-1}, c_0), (b_{K-1}, c_K)$, non nuls.

Comme auparavant, nous posons $a_j = 0, \forall j < 0$. Cette équation est *a priori* plus générale que l'équation (7) vue auparavant; on montrera cependant que b_{-1} est nécessairement nul.

Tous les processus ainsi définis n'ont pas le même intérêt du point de vue pratique. Ainsi il paraît possible d'imposer les conditions supplémentaires suivantes, qui permettront dans la suite d'importantes simplifications et qui sont en particulier satisfaites pour les processus ARMA.

HYPOTHÈSE I : i. $c_K \neq 0$;

ii. le polynôme $C(z) = \sum_{j=0}^K c_j z^j$ a toutes ses racines distinctes.

Comme nous l'avons vu dans la partie introductive, l'hypothèse ii est automatiquement satisfaite dans le cas des processus ARMA à condition de retenir pour polynôme C le dénominateur de la forme irréductible de la fraction $\frac{1}{A(z)} \frac{dA(z)}{dz}$.

L'hypothèse i est, elle, identique à la condition $d^\circ B \leq d^\circ C - 1$.

3.b. Existence et unicité des processus MADE

L'existence d'un processus stationnaire régulier satisfaisant (8) implique évidemment diverses restrictions sur les coefficients b et c intervenant dans la relation.

Celles-ci sont conséquences des conditions $a_0 = 1$ et $\sum_{j=0}^{\infty} a_j^2 < \infty$ imposées au développement moyenne-mobile infinie.

i. **Restriction due à l'égalité $a_0 = 1$**

La relation (8) est évidemment satisfaite pour tout indice $j < 0$, du fait de la condition $a_j = 0, \forall j < 0$. Si nous écrivons maintenant cette relation pour $j = 0$, nous avons :

$$(-b_{-1} + c_0 \cdot 0) a_0 = 0 \Leftrightarrow b_{-1} a_0 = b_{-1} = 0.$$

Le coefficient b_{-1} est donc nécessairement nul et nous l'omettrons dans la suite. Remarquons que sous cette condition, il résulte de l'hypothèse $(b_{-1}, c_0) \neq 0$ que le coefficient c_0 de $a_j, j \geq 1$ est toujours non nul; on peut donc, pour $j \geq 1$, exprimer a_j en fonction de a_{j-1}, \dots, a_{j-k} et déduire de l'équation (8) des expressions uniques de (a_j) en fonction des coefficients b et c . Nous disposons en effet des valeurs initiales : $a_0 = 1, a_{-1} = a_{-2} = \dots = a_{-k+1} = 0$.

On a par exemple :

$$a_0 = 1,$$

$$c_0 a_1 + (-b_0 a_0) = 0 \Leftrightarrow a_1 = \frac{b_0}{c_0},$$

$$2c_0 a_2 + (-b_0 + c_1) a_1 + (-b_1 a_0) = 0 \Leftrightarrow a_2 = \frac{1}{2c_0} \left[(b_0 - c_1) \frac{b_0}{c_0} + b_1 \right]$$

... et ainsi de suite.

ii. **Restriction due à l'existence de la somme $\sum_{j=0}^{\infty} a_j^2$**

PROPRIÉTÉ 1 : Il existe un processus stationnaire régulier satisfaisant la relation (8) si et seulement si, on a :

$$b_{-1} = 0$$

et, si considérant la décomposition en éléments simples de

$$\frac{B(z)}{C(z)} = \sum_{k=1}^{K_0} \frac{-\alpha_k \delta_k}{1 - \alpha_k z} + \sum_{k=1}^{K_1} \frac{\beta_k d_k}{1 - \beta_k z} \quad \text{avec } \operatorname{Re}(\delta_k) \geq 0$$

et $\operatorname{Re}(d_k) > 0$, avec $K_0 + K_1 = K$, on a :

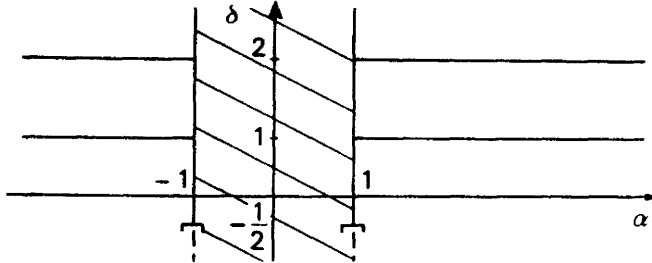
a) $|\beta_k| < 1$ ou $\left(|\beta_k| = 1 \text{ et } \operatorname{Re}(d_k) < \frac{1}{2} \right), \forall k = 1, \dots, K_1,$

b) δ_k réel entier ou $(|\alpha_k| \leq 1 \text{ et } \delta_k \text{ non entier}), \forall k = 1, \dots, K_0.$

Preuve : Voir appendice 1.

En fait lorsque les termes α_k et β_k sont de modules strictement inférieurs à 1, on a même l'absolue sommabilité $\sum_{j=0}^{\infty} |a_j| < \infty$.

Exemple 1 : Afin d'illustrer la propriété d'existence 1, on peut considérer le cas simple où $\frac{B(z)}{C(z)} = \frac{-\alpha\delta}{1-\alpha z}$, où δ est non contraint. La zone d'existence peut être représentée graphiquement dans un repère (α, δ) .



En plus de la zone hachurée, le domaine d'existence comprend un certain nombre de demi-droites; celles verticales correspondent à la zone frontière $\alpha = \pm 1$; celles horizontales à des valeurs entières de δ , c'est-à-dire à des représentations moyennes mobiles pures avec une racine à l'intérieur du disque unité.

Divers corollaires sont immédiatement déduits de la propriété précédente et de sa démonstration.

COROLLAIRE 2 : Un processus stationnaire régulier est MADE si et seulement si $\frac{1}{A(z)} \frac{dA(z)}{dz}$ a la forme d'une fraction rationnelle en z .

COROLLAIRE 3 : Un processus stationnaire est MADE si et seulement si

$$\text{la fonction de transfert s'écrit : } A(z) = \frac{\prod_{k=1}^{\kappa_0} (1 - \alpha_k z)^{\delta_k}}{\prod_{k=1}^{\kappa_1} (1 - \beta_k z)^{d_k}}$$

Ce corollaire s'obtient évidemment par intégration de la décomposition en éléments simples de $\frac{1}{A(z)} \frac{dA(z)}{dz} = \frac{B(z)}{C(z)}$ donnée dans la propriété 1.

Signalons simplement que l'écriture $(1 - \beta z)^d$ correspond à la détermination principale, c'est-à-dire celle prenant la valeur 1 pour $z=0$. Cette détermination principale est donnée par :

$$(9) \quad (1 - \beta z)^d = 1 + d(-\beta z) + \dots + \frac{d(d-1) \dots (d-j+1)}{j!} (-\beta z)^j + \dots$$

Ce développement en série existe sous les conditions a), b) de la propriété 1 ce qui donne ainsi un sens à l'expression de $A(z)$.

COROLLAIRE 4 : Un processus stationnaire régulier est MADE si et seulement si il satisfait une équation du type :

$$\prod_{k=1}^{\kappa_1} (1 - \beta_k L)^{d_k} X_t = \prod_{k=1}^{\kappa_0} (1 - \alpha_k L)^{\delta_k} \epsilon_t,$$

$$\text{Re } d_k > 0, \quad \text{Re } \delta_k \geq 0.$$

Ce corollaire montre qu'il est possible de distinguer les « séries autorégressives » des « séries moyennes mobiles » en étudiant les signes des numérateurs intervenant dans la décomposition en éléments simples.

COROLLAIRE 5 : Si les coefficients $\beta_k, d_k, \alpha_k, \delta_k$ satisfont les conditions de la propriété 1, il existe un unique processus stationnaire régulier vérifiant

$$\prod_{k=1}^{K_1} (1 - \beta_k L)^{d_k} X_t = \prod_{k=1}^{K_0} (1 - \alpha_k L)^{\delta_k} \varepsilon_t,$$

$$\operatorname{Re} d_k > 0, \quad \operatorname{Re} \delta_k \geq 0.$$

COROLLAIRE 6 : Un processus MADE admet plusieurs représentations possibles fondées sur le bruit blanc ε , c'est-à-dire correspond à plusieurs couples de polynômes B et C. Ces couples se déduisent tous de la forme irréductible de $\frac{1}{A(z)} \frac{dA(z)}{dz}$. En particulier, il existe un couple pour lequel

le degré K du polynôme C est le plus petit possible.

Dans la suite, nous notons K ce degré minimal et nous dirons que le processus admet une représentation MADE(K). Du fait de la bijection entre les définitions en terme de fractions rationnelles et en terme d'équations aux différences à coefficients polynomiaux de degré 1, on peut aussi interpréter K comme l'ordre minimal de telles équations aux différences satisfaites par les coefficients (a_j) .

3. c. Densité spectrale

Les processus MADE peuvent évidemment être caractérisés par la forme de leur densité spectrale; celle-ci est donnée par :

$$(10) \quad f(\omega) = \frac{\sigma^2}{2\pi} |A(\exp i\omega)|^2 = \frac{\sigma^2}{2\pi} \frac{\prod_{k=1}^{K_0} |(1 - \alpha_k \exp i\omega)^{\delta_k}|^2}{\prod_{k=1}^{K_1} |(1 - \beta_k \exp i\omega)^{d_k}|^2}.$$

L'examen de la démonstration de la propriété 1 montre que la condition $\operatorname{Re} d_k < \frac{1}{2}$, lorsque $|\beta_k| = 1$ est nécessaire et suffisante pour assurer l'intégrabilité de la fonction f au voisinage du point ω_0 pour lequel $\beta_k \exp i\omega_0 = 1$.

Remarque : Lorsque les exposants d_k et δ_k sont réels, la densité spectrale peut être facilement exprimée en terme de sinus. On a :

$$\begin{aligned} |(1 - \alpha \exp i\omega)^\delta|^2 &= (|1 - \alpha \exp i\omega|^2)^\delta = [1 + \alpha^2 - 2\alpha \cos \omega]^\delta \\ &= \left[(1 - \alpha)^2 + 4\alpha \sin^2 \frac{\omega}{2} \right]^\delta. \end{aligned}$$

D'où on déduit :

$$f(\omega) = \frac{\sigma^2}{2\pi} \frac{\prod_{k=1}^{K_0} \left[(1 - \alpha_k)^2 + 4\alpha_k \sin^2 \frac{\omega}{2} \right]^{\delta_k}}{\prod_{k=1}^{K_1} \left[(1 - \beta_k)^2 + 4\beta_k \sin^2 \frac{\omega}{2} \right]^{d_k}}.$$

Remarque : La densité spectrale peut être infinie en certains points. Ceci se produit, lorsque pour un indice k on a $|\beta_k| = 1$ et $\text{Re } d_k < \frac{1}{2}$. Considérons pour simplifier le cas où l'exposant d_k associé est réel et notons $\beta_k = \exp(-i\omega_k)$. Au voisinage de la valeur $\omega = \omega_k$, on a :

$$\begin{aligned} \text{Log } f(\omega) &\approx \text{cte} - d_k \log |1 - \exp - i\omega_k \exp i\omega|^2 \\ &\approx \text{cte} - d_k \log [2(1 - \cos(\omega - \omega_k))] \\ &\approx -d_k \log(\omega - \omega_k)^2 \approx -2d_k \log |\omega - \omega_k|. \end{aligned}$$

Ainsi la densité spectrale tend vers l'infini lorsque ω tend vers ω_k , ceci évidemment en restant intégrable. De plus l'exposant réel d_k associé à la racine β_k de module 1 permet de mesurer la vitesse avec laquelle a lieu cette divergence.

Dans le cas d'exposants complexes, la vitesse est fonction de $\text{Re } d_k$ et non de d_k , comme on le vérifie facilement à partir de l'écriture développée donnée ci-dessous.

Remarque : Certains exposants d_k ou δ_k peuvent être complexes. Les coefficients de A étant réels, ils figurent alors avec leur conjugué. Considérons ainsi un processus MADE (2) défini par :

$$(1 - \beta L)^d (1 - \bar{\beta} L)^{\bar{d}} X_t = \varepsilon_t,$$

où β et d sont tous deux complexes et satisfont les contraintes de la propriété 1. La densité spectrale s'écrit :

$$f(\omega) = \frac{\sigma^2}{2\pi} \frac{1}{(1 - \beta \exp i\omega)^d (1 - \bar{\beta} \exp -i\omega)^{\bar{d}} (1 - \bar{\beta} \exp i\omega)^{\bar{d}} (1 - \beta \exp -i\omega)^d}$$

Séparant les parties réelles et imaginaires de d , nous avons :

$$f(\omega) = \frac{\sigma^2}{2\pi} \frac{1}{|1 - \beta \exp i\omega|^{2 \text{Re } d} |1 - \bar{\beta} \exp i\omega|^{2 \text{Re } \bar{d}}} \frac{1}{[(1 - \beta \exp i\omega)/(1 - \bar{\beta} \exp -i\omega)]^{i \text{Im } d} [(1 - \bar{\beta} \exp i\omega)/(1 - \beta \exp -i\omega)]^{-i \text{Im } \bar{d}}}$$

Comme les rapports $\frac{1 - \beta \exp i\omega}{1 - \bar{\beta} \exp -i\omega}$ et $\frac{1 - \bar{\beta} \exp i\omega}{1 - \beta \exp -i\omega}$ sont de modules 1, ils peuvent être notés $\exp -i\theta(\omega)$ et $\exp +i\bar{\theta}(\omega)$ respectivement. Si de plus nous notons $\beta = |\beta| \exp(-i\omega_0)$ la décomposition du complexe β faisant apparaître son module $|\beta|$ et son argument ω_0 , nous obtenons :

$$f(\omega) = \frac{\sigma^2}{2\pi} \times \frac{1}{[1 + |\beta|^2 - 2|\beta| \cos(\omega - \omega_0)]^{\text{Re } d} [1 + |\beta|^2 - 2|\beta| \cos(\omega + \omega_0)]^{\text{Re } \bar{d}}} \frac{1}{\exp \text{Im } d [\theta(\omega) + \bar{\theta}(\omega)]}$$

Remarquons que, le dernier terme $1/\exp \text{Im } d [\theta(\omega) + \bar{\theta}(\omega)]$ étant continu, borné (puisque $\theta(\omega)$ et $\bar{\theta}(\omega)$ sont compris entre 0 et 2π), la densité spectrale

$f(\omega)$ peut uniquement tendre vers l'infini si et seulement si le premier terme tend vers l'infini. Ceci nous ramène donc au cas d'un exposant réel (remplacement de d par sa partie réelle).

Les résultats précédents peuvent évidemment être écrits en terme de fonction génératrice d'autocovariance. On a :

$$G(z) = \sum_{h=-\infty}^{+\infty} \gamma(h) z^h = \sigma^2 A(z) A\left(\frac{1}{z}\right),$$

ou de manière équivalente en écrivant les dérivées logarithmiques :

$$\begin{aligned} \frac{1}{G(z)} \frac{dG(z)}{dz} &= \frac{1}{A(z)} \frac{dA(z)}{dz} + \frac{1}{A(1/z)} \frac{dA(1/z)}{dz}, \\ \frac{1}{G(z)} \frac{dG(z)}{dz} &= \frac{B(z)}{C(z)} - \frac{1}{z^2} \frac{B(1/z)}{C(1/z)}, \\ \frac{1}{G(z)} \frac{dG(z)}{dz} &= \frac{z^{K+1} B(z) C(1/z) - z^{K-1} B(1/z) C(z)}{C(z) C(1/z) z^{K+1}} \end{aligned}$$

$$(11) \quad \frac{1}{G(z)} \frac{dG(z)}{dz} = \frac{B^*(z)}{C^*(z)}$$

où

$$B^*(z) = z^{K+1} B(z) C\left(\frac{1}{z}\right) - z^{K-1} B\left(\frac{1}{z}\right) C(z)$$

et

$$C^*(z) = z^{K+1} C(z) C\left(\frac{1}{z}\right).$$

Ainsi la fonction génératrice d'autocovariance satisfait une propriété similaire à celle vérifiée par la fonction de transfert : sa dérivée logarithmique coïncide avec une fraction rationnelle en z .

Le dénominateur de cette fraction est de degré $2K+1$ et, à un facteur z près, coïncide avec un polynôme symétrique $\left[z^K C(z) C\left(\frac{1}{z}\right) \right]$; le numérateur est de degré $2K$.

Comme le dénominateur comporte z en facteur et que celui-ci n'est pas facteur du numérateur, on en déduit immédiatement que la suite des autocovariances satisfait une relation de récurrence linéaire d'ordre $2K$, dont les coefficients sont des polynômes de degrés inférieurs ou égaux à 1. Notant :

$$B^*(z) = b_0^* + b_1^* z + \dots + b_{2K}^* z^{2K}$$

et

$$C^*(z) = c_1^* z + c_2^* z^2 + \dots + c_{2K+1}^* z^{2K+1},$$

on a :

$$(12) \quad \sum_{k=0}^{2K} (-b_k^* + c_{k+1}^*) \gamma(h-k) = 0, \quad \forall h \in \mathbb{Z}.$$

L'équation (11) permet de retrouver la fonction $G(z)$ à une constante multiplicative près. Elle caractérise donc la fonction génératrice d'autocorrélation :

$$R(z) = \frac{1}{\gamma(0)} G(z) = \sum_{h=-\infty}^{+\infty} \rho(h) z^h.$$

Comme l'équation de récurrence (12) est équivalente à l'équation différentielle (11), dès que la fonction G existe, on en déduit que la relation (12)

admet une seule suite solution telle que : $\rho(0) = 1$ et $\sum_{h=-\infty}^{+\infty} \rho(h)^2 < +\infty$; celle-ci n'est autre que la fonction d'autocorrélation du processus.

Exemple 2 : A titre d'illustration, considérons le cas d'un processus MADE (1). Nous savons que $\frac{1}{A(z)} \frac{dA(z)}{dz} = \frac{-\alpha\delta}{1-\alpha z}$ (où δ est un réel de signe quelconque) et nous en déduisons que :

$$\begin{aligned} \frac{1}{G(z)} \frac{dG(z)}{dz} &= \frac{-\alpha\delta}{1-\alpha z} - \frac{1}{z^2} \frac{-\alpha\delta}{1-(\alpha/z)} \\ &= \frac{\alpha\delta(1-z^2)}{z[-\alpha z^2 + z(1+\alpha^2) - \alpha]}, \\ \frac{1}{G(z)} \frac{dG(z)}{dz} &= (z - (1/z)) \frac{\delta}{z^2 - z((1+\alpha^2)/\alpha) + 1}. \end{aligned}$$

La fonction d'autocovariance satisfait donc la relation :

$$(\delta + h) \gamma(h) - \frac{(1+\alpha^2)}{\alpha} (h-1) \gamma(h-1) + (-\delta + h-2) \gamma(h-2) = 0$$

pour toute valeur h dans \mathbb{Z} , ou de manière plus symétrique dans l'indice :

$$(13) \quad (\delta + h + 1) \gamma(h+1) - \frac{1+\alpha^2}{\alpha} h \gamma(h) + (-\delta + h - 1) \gamma(h-1) = 0, \\ \forall h \in \mathbb{Z}.$$

Exemple 3 : Considérons pour simplifier le cas où $|\alpha| < 1$ et où $\delta = -1$, c'est-à-dire celui d'un processus AR (1) :

$$(1 - \alpha L) X_t = \varepsilon_t.$$

L'équation (13) se réduit à :

$$\begin{aligned} h \gamma(h+1) - \frac{1+\alpha^2}{\alpha} h \gamma(h) + h \gamma(h-1) &= 0, \quad \forall h \in \mathbb{Z} \\ \Rightarrow \gamma(h+1) - \frac{1+\alpha^2}{\alpha} \gamma(h) + \gamma(h-1) &= 0, \quad \forall h \geq 1. \end{aligned}$$

La suite $(\gamma(h), h > 0)$ satisfait ainsi une équation de récurrence linéaire d'ordre 2 à coefficients constants. Les racines de l'équation caractéristique associée étant α et $\frac{1}{\alpha}$, on a donc : $\gamma(h) = A \alpha^h + B \frac{1}{\alpha^h}$.

La condition d'existence de $f(\omega) = \frac{1}{2\pi} \sum_h \gamma(h) \exp(ih\omega)$ entraîne alors $B=0$, puisque $|\alpha| < 1$. On a donc nécessairement $\gamma(h) = \gamma(0) \alpha^h, h \geq 0$, ce qui est bien la fonction d'autocovariance du processus AR(1).

Exemple 4 : Dans le cas limite du bruit blanc avec $\delta=0$, l'équation (13) s'écrit :

$$(h+1)\gamma(h+1) - \frac{1+\alpha^2}{\alpha} h\gamma(h) + (h-1)\gamma(h-1) = 0, \quad \forall h \in \mathbb{Z}.$$

La suite $(h\gamma(h), h \in \mathbb{Z})$ satisfait donc une équation de récurrence linéaire à coefficients constants. Comme les racines de l'équation caractéristique sont α et $\frac{1}{\alpha}$, on a :

$$h\gamma(h) = A \alpha^h + \frac{B}{\alpha^h}, \quad \forall h \in \mathbb{Z}.$$

En particulier pour $h=0$, on obtient : $0 = A + B$ et donc $h\gamma(h) = A \left(\alpha^h - \frac{1}{\alpha^h} \right)$. Cette forme n'est compatible avec la « convergence » de la série que si $A=0$. On a alors : $\gamma(h) = 0, \forall h \neq 0$, ce qui correspond bien à un bruit blanc. Remarquons de plus que, comme on pouvait s'y attendre, le paramètre α intervenant dans l'équation de récurrence, n'apparaît plus dans la suite d'autocovariance solution.

Exemple 5 : La détermination de la solution « convergente » satisfaisant $\gamma(0) = 1$ n'est pas toujours aussi directe. Ainsi dans le cas d'un MA(1), on a : $\delta=1$ et la relation (13) se réduit à :

$$(h+2)\gamma(h+1) - \frac{1+\alpha^2}{\alpha} h\gamma(h) + (h-2)\gamma(h-1) = 0.$$

Pour $h=2$, on obtient : $4\gamma(3) - \frac{1+\alpha^2}{\alpha} 2\gamma(2) = 0$; de sorte que les valeurs $\gamma(h), h \geq 2$ sont caractérisées par la seule connaissance de $\gamma(2)$.

3. d. Représentation autoregressive

On peut se demander dans quelle mesure un processus MADE admet aussi une représentation autorégressive infinie unilatérale, dans le bruit ε :

$$(14) \quad X_t - \pi_1 X_{t-1} \dots - \pi_j X_{t-j} \dots = \varepsilon_t.$$

On dira alors que le processus est inversible.

On peut vérifier que dans le cas des processus MADE une condition nécessaire et suffisante pour pouvoir écrire $\varepsilon_t = \sum_{j=-\infty}^{+\infty} \pi_j X_{t-j}$ est que l'inverse

de la densité spectrale soit intégrable :

$$\int_{-\pi}^{\pi} \frac{1}{f(\omega)} d\omega < \infty.$$

On retrouve ainsi une condition habituellement utilisée lorsque la densité spectrale est bornée (ce qui n'est pas forcément le cas ici).

Utilisant la forme de la densité spectrale (10), on voit immédiatement que cette condition implique :

$$(15) \quad \operatorname{Re} \delta_k < \frac{1}{2}$$

pour tout indice k pour lequel le coefficient α_k est de module 1.

Sous cette condition supplémentaire et sous les conditions de la propriété 1 assurant l'existence du processus MADE, on peut inverser l'opérateur $A(L)$. Deux cas sont alors à distinguer : si parmi les α_k figure un élément de module strictement supérieur à 1, on obtient une écriture du type :

$$\varepsilon_t = \sum_{j=-\infty}^{+\infty} \pi_j X_{t-j}$$

faisant intervenir des valeurs postérieures de X .

Dans le cas contraire où tous les α sont de modules inférieurs ou égaux à 1, on a bien la représentation cherchée.

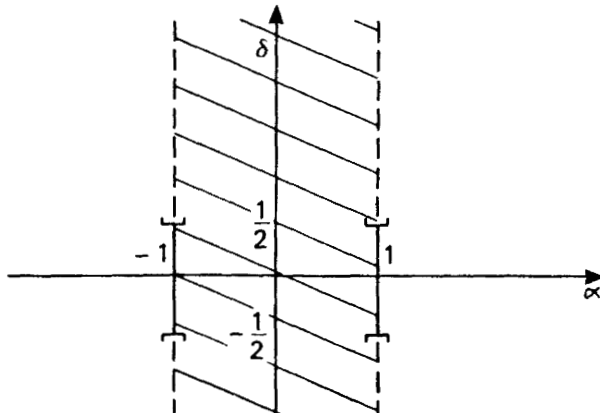
En résumé, nous avons la propriété suivante :

PROPRIÉTÉ 7 : Un processus MADE admet une représentation autorégressive unilatérale si et seulement si :

i. $\forall k = 1, \dots, K_0, |\alpha_k| \leq 1$;

ii. $\operatorname{Re} \delta_k < \frac{1}{2}$ pour tout indice k tel que $|\alpha_k| = 1$.

Exemple 6 : Si nous reprenons l'exemple 1 d'un processus pour lequel $\frac{B(z)}{C(z)} = \frac{-\alpha\delta}{1-\alpha z}$, la zone d'existence et d'inversibilité est :



Si nous nous restreignons aux processus inversibles, le résultat précédent nous conduit à séparer les processus MADE en deux classes :

i. celle des processus pour lesquels le polynôme C intervenant dans la forme irréductible a toutes ses racines à l'extérieur strictement du disque unité. Ces processus sont dans la suite dits *génériques* ;

ii. la classe complémentaire des processus pour lesquels il existe dans C une racine de module 1. Ces processus seront dits à *mémoire longue*.

4 Coefficients du développement moyenne mobile et forme des autocovariances

4. a. Processus MADE (1)

Il s'agit des processus moyennes mobiles infinies satisfaisant la relation :

$$(16) \quad (1 - \alpha L)^{-\delta} X_t = \varepsilon_t$$

$$\Leftrightarrow (17) \quad X_t = (1 - \alpha L)^\delta \varepsilon_t$$

où α est un réel de module inférieur ou égal à 1 et où δ est un réel tel que $\frac{1}{2} > \delta > -\frac{1}{2}$, si $\alpha = \pm 1$. Nous ne distinguons donc pas dans cette section deux cas selon le signe de l'exposant ; nous posons cependant $d = -\delta$ pour simplifier certaines formules.

4. a. 1. Coefficients du développement moyenne mobile infinie

On a déjà donné en (9) le développement de $(1 - \alpha L)^\delta$:

$$(18) \quad (1 - \alpha L)^\delta = 1 + \delta(-\alpha L) + \frac{\delta(\delta-1)}{2}(-\alpha L)^2 + \dots \\ + \frac{\delta(\delta-1)\dots(\delta-j+1)}{j!}(-\alpha L)^j + \dots$$

Cette expression peut être présentée sous une forme plus simple en introduisant $d = -\delta$. Nous avons :

$$(1 - \alpha L)^\delta = \sum_{j=0}^{\infty} \frac{(-d)(-d-1)\dots(-d-j+1)}{j!} (-1)^j \alpha^j L^j \\ = \sum_{j=0}^{\infty} (d+j-1)\dots(d+1) d \frac{\alpha^j}{j!} L^j$$

Notant :

$$(19) \quad \frac{\Gamma(d+j)}{\Gamma(d)} = (d+j-1) \dots (d+1) d,$$

nous obtenons l'écriture résumée :

$$(20) \quad (1-\alpha L)^\delta = \sum_{j=0}^{\infty} \frac{\Gamma(d+j)}{\Gamma(d)} \frac{\alpha^j}{j!} L^j \quad \text{avec } d = -\delta.$$

Remarque : L'écriture précédente correspond à l'expression de $(1-\alpha L)^\delta$ à partir de la série hypergéométrique de Gauss (voir par exemple ABRAMOWITZ-STEGUN [1965], GRADSHTEYN-RYZHIK [1965] et Appendice 2). Celle-ci est définie par :

$$(21) \quad F(a, b, c, z) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{j=0}^{\infty} \frac{\Gamma(a+j)\Gamma(b+j)}{\Gamma(c+j)} \frac{z^j}{j!}$$

et nous avons :

$$(22) \quad (1-\alpha L)^\delta = F(-\delta, b, b, \alpha L)$$

C'est généralement cette interprétation qui a été retenue dans la littérature (GRANGER [1980], HOSKING [1981]...) pour établir les principales propriétés des processus à mémoire longue (cas $\alpha=1$). Ainsi pour mettre en évidence les récurrences entre les coefficients du développement moyenne mobile, ces divers auteurs ont utilisés les relations de Gauss sur les fonctions hypergéométriques contiguës. Il paraît cependant plus éclairant d'utiliser comme nous l'avons fait dans la première partie les équations différentielles sous-jacentes pour obtenir ces mêmes résultats.

Connaissant la forme explicite des coefficients a_j du développement moyenne mobile infinie $X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$, il est évidemment facile d'étudier le comportement de la suite a_j et de voir comment celui-ci se modifie avec les paramètres α et δ .

Dans le cas d'un processus MADE(1) l'équation de récurrence sur les coefficients (a_j) est :

$$(23) \quad ja_j + [\alpha\delta - \alpha(j-1)]a_{j-1} = 0, \quad \forall j \in \mathbb{Z}.$$

Ceci implique :

$$(24) \quad \frac{a_j}{a_{j-1}} = \frac{\alpha(d+j-1)}{j},$$

pour $j \geq 1$ (sauf pour les cas limites des MA purs où les a_j sont nuls à partir d'un certain rang).

Exemple 7: Considérons à titre d'illustration le cas où α est un réel positif inférieur à 1.

Lorsque d est compris entre 0 et 1 (passage du bruit blanc à une représentation AR (1))

$$\forall j \geq 1: 0 \leq \frac{a_j}{a_{j-1}} \leq 1.$$

Tous les coefficients sont ainsi positifs et décroissent avec j . De plus on voit immédiatement que cette décroissance est plus forte lorsque la valeur de d est plus faible.

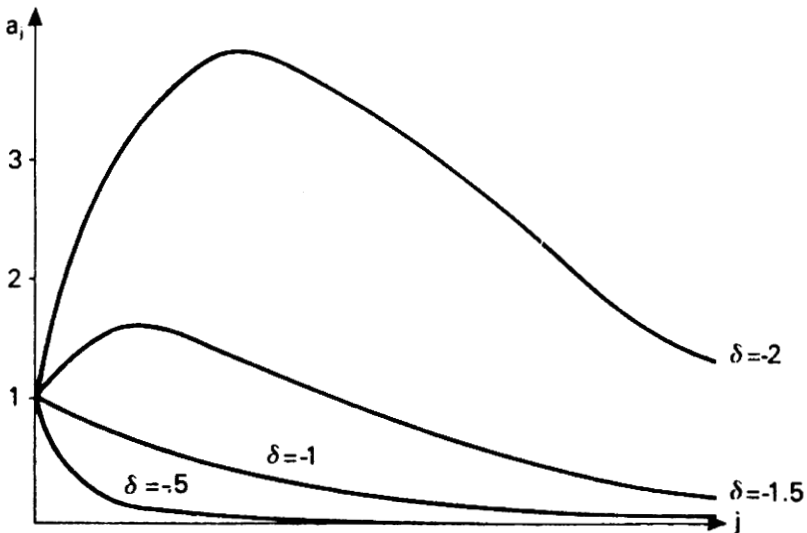
Lorsque d est compris entre -1 et 0 (passage d'une représentation MA (1) à un bruit blanc), on a : $\frac{a_j}{a_{j-1}} = \frac{\alpha(-\delta-1+j)}{j}$. Ce rapport est négatif pour $j=1$ et positif pour $j \geq 2$. On a donc $a_0 = 1 > 0$, $a_j \leq 0$, $\forall j \geq 1$. Il y a de plus décroissance du module de a_j et cette décroissance est plus accentuée, lorsque d se rapproche de zéro.

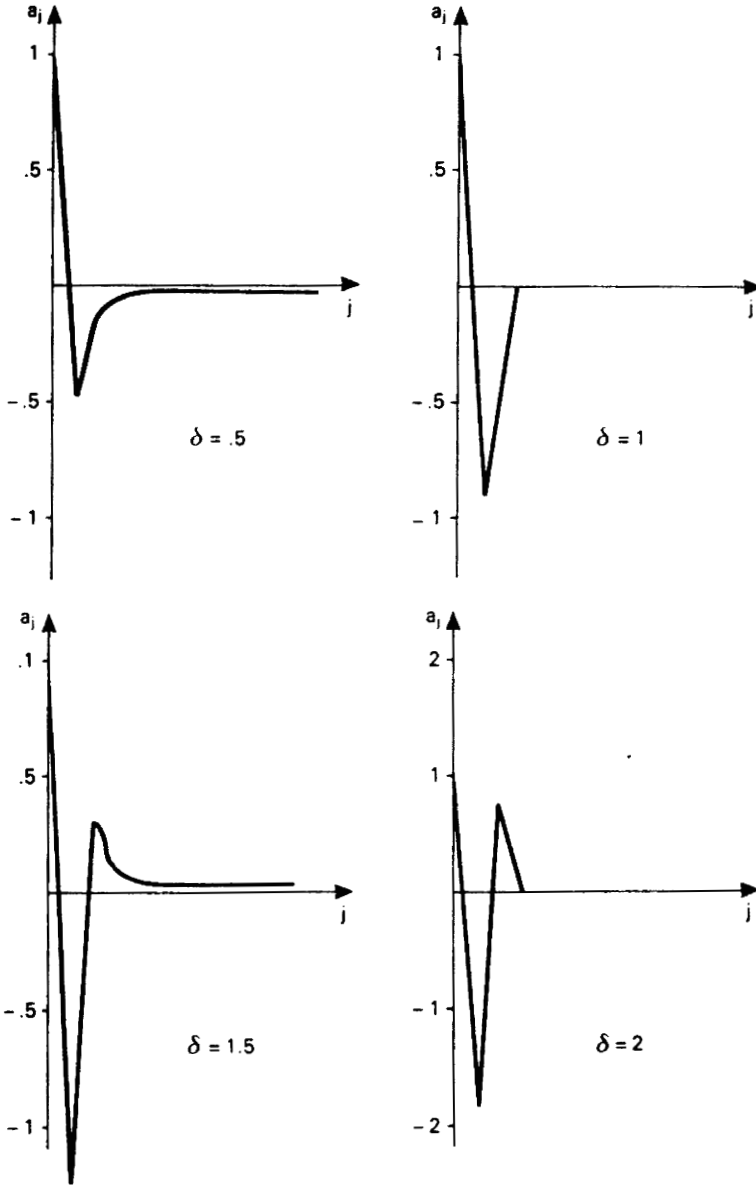
Il ne faudrait cependant pas croire que, dans le cas particulier d'un MADE(1), les coefficients a_j ont toujours une évolution monotone; ce résultat est du aux contraintes $-1 \leq d \leq 1$, que nous avons imposées jusqu'à maintenant. Si nous considérons maintenant un exposant tel que d soit supérieur à 1, nous voyons que :

$$\frac{a_j}{a_{j-1}} = \frac{\alpha(d+j-1)}{j} < 1.$$

$$\Leftrightarrow j > \frac{\alpha(d-1)}{1-\alpha} \geq 0.$$

Il y a donc croissance, puis décroissance des coefficients en fonction de j . Nous donnons dans la figure suivante pour $\alpha=0,9$, $\delta=-2, -1,5, -1, -0,5, 0, 0,5, 1, 1,5, 2$ les graphes des (a_j) .





Le comportement asymptotique des coefficients (a_j) se déduit directement de la formule de Sheppard :

$$\frac{\Gamma(j+a)}{\Gamma(j+b)} \approx j^{a-b}, \text{ si } j \text{ tend vers l'infini.}$$

Comme : $a_j = \frac{\Gamma(d+j)}{\Gamma(d)} \frac{\alpha^j}{j!} = \frac{\Gamma(d+j)}{\Gamma(1+j)} \frac{\alpha^j}{\Gamma(d)}$, on en déduit :

$$(25) \quad a_j \approx \frac{j^{d-1} \alpha^j}{\Gamma(d)}.$$

Remarque : Dans le cas de modèles à mémoire longue, où par exemple $\alpha=1$, on a : $a_j \approx \frac{j^{d-1}}{\Gamma(d)}$. On voit que $a_j^2 \approx \frac{j^{2d-2}}{\Gamma(d)^2}$ et la suite des a_j est de carré sommable si et seulement si : $2d-2 < -1 \Leftrightarrow d < \frac{1}{2}$; on retrouve ainsi directement dans ce cas très particulier le résultat de la propriété 1.

De plus on remarque que la décroissance n'est pas exponentielle, mais est de l'ordre d'une puissance négative de j . Bien que le processus soit stationnaire, l'influence de X_t sur X_{t+h} , $h \geq 0$ va être « importante » même pour de grandes valeurs de h . Ceci explique la terminologie « mémoire longue » retenue pour ce type de modèles. Cette décroissance lente correspond évidemment dans le domaine des fréquences à l'existence d'un point où la densité spectrale est infinie (ici pour $\omega=0$).

4. a.2. Forme des autocovariances

Nous avons vu dans la section 3 que la densité spectrale est donnée par :

$$(10) \quad f(\omega) = \frac{\sigma^2}{2\pi} |1 - \alpha \exp i\omega|^{2\delta} = \frac{\sigma^2}{2\pi} \left[(1-\alpha)^2 + 4\alpha \sin^2 \frac{\omega}{2} \right]^\delta,$$

et que les autocovariances $[\gamma(h)]$ satisfont :

$$(13) \quad (\delta+h+1)\gamma(h+1) - \frac{1+\alpha^2}{\alpha} h\gamma(h) + (-\delta+h-1)\gamma(h-1) = 0,$$

$$\forall h \in \mathbb{Z}.$$

La forme explicite des autocovariances se déduit directement de celle des coefficients a_j . On a, pour $h \geq 0$:

$$\begin{aligned} \gamma(h) &= \sigma^2 \sum_{j=0}^{\infty} a_j a_{j+h} \\ &= \sigma^2 \sum_{j=0}^{\infty} \frac{\Gamma(d+j)}{\Gamma(d)} \frac{\alpha^j}{j!} \frac{\Gamma(d+j+h)}{\Gamma(d)} \frac{\alpha^{j+h}}{(j+h)!} \\ &= \left[\frac{\sigma^2 \Gamma(h+d)}{\Gamma(h+1)\Gamma(d)} \alpha^h \right] \times \left[\frac{\Gamma(h+1)}{\Gamma(h+d)\Gamma(d)} \sum_{j=0}^{\infty} \frac{\Gamma(d+j)\Gamma(d+h+j)}{\Gamma(h+1+j)} \frac{\alpha^{2j}}{j!} \right]. \end{aligned}$$

On reconnaît dans le second terme entre crochet une valeur de la fonction hypergéométrique (voir (21)). D'où le résultat suivant :

$$(26) \quad \gamma(h) = \sigma^2 \frac{\Gamma(h+d)}{\Gamma(h+1)\Gamma(d)} \alpha^h F(d, d+h, h+1, \alpha^2), \quad \forall h \geq 0.$$

Remarque : L'équation de récurrence satisfaite par la suite des autocovariances peut être directement retrouvée à partir de la forme explicite (26)

de $\gamma(h)$. On déduit des relations de contiguïté de Gauss (voir Appendice 2), l'égalité suivante :

$$(c-1)F(a, b-1, c-1; z) + (az-bz-c+1)F(a, b, c; z) + \frac{b(c-a)}{c}zF(a, b+1, c+1; z) = 0.$$

Prenons alors $a=d$, $b=d+h$, $c=h+1$, $z=\alpha^2$ et remplaçons les fonctions hypergéométriques par leurs expressions en fonction des autocovariances, nous obtenons :

$$\begin{aligned} h \frac{\Gamma(h)\Gamma(d)}{\sigma^2\Gamma(h+d-1)} \frac{\gamma(h-1)}{\alpha^{h-1}} + (-h\alpha^2-h) \frac{\Gamma(h+1)\Gamma(d)}{\sigma^2\alpha^h\Gamma(h+d)} \gamma(h) \\ + \frac{(d+h)(h+1-d)}{h+1} \alpha^2 \frac{\Gamma(h+2)\Gamma(d)}{\sigma^2\alpha^{h+1}\Gamma(h+1+d)} \gamma(h+1) = 0, \quad \forall h \in \mathbb{Z} \\ \Leftrightarrow \frac{\gamma(h-1)}{\Gamma(h+d-1)} - \frac{(1+\alpha^2)}{\alpha} h \frac{\gamma(h)}{\Gamma(h+d)} + (h+1-d) \frac{\gamma(h+1)}{\Gamma(h+d)} = 0, \quad \forall h \in \mathbb{Z} \\ \Leftrightarrow (h-1+d)\gamma(h-1) - \frac{1+\alpha^2}{\alpha} h \gamma(h) + (h+1-d)\gamma(h+1) = 0, \quad \forall h \in \mathbb{Z}. \end{aligned}$$

Nous retrouvons bien la relation (13).

Remarque : L'équation de récurrence (13) satisfaite par la fonction d'autocovariance peut être assez différente de celle habituellement considérée pour les processus ARMA. Considérons un autoregressif d'ordre p défini par $(1-\alpha L)^p X_t = \varepsilon_t$. Un tel processus correspond à $d=p \in \mathbb{N}^*$ et l'équation de récurrence devient pour $p=1, 2, \dots$

$$\begin{aligned} p=1, \quad \alpha\gamma(h+1) - (1+\alpha^2)\gamma(h) + \alpha\gamma(h-1) = 0 \\ \Leftrightarrow (1-\alpha L)(\alpha-L)\gamma(h) = 0 \\ p=2, \quad (h-1)\gamma(h+1) - \frac{1+\alpha^2}{\alpha} h \gamma(h) + (h+1)\gamma(h-1) = 0 \end{aligned}$$

Si l'opérateur $1-\alpha L$ apparaît bien dans la relation associée à $p=1$, l'opérateur $(1-\alpha L)^2$ n'est pas en évidence dans celle associée à $p=2$.

Si nous nous restreignons au cas des processus à mémoire longue $\alpha = \pm 1$, les expressions des autocovariances se simplifient. On a :

$$\gamma(h) = \frac{\sigma^2\Gamma(h+d)}{\Gamma(h+1)\Gamma(d)} \alpha^h F(d, d+h, h+1, 1).$$

Or il résulte des propriétés classiques de la fonction hypergéométrique (Appendice 2) que :

$$F(a, b, c; 1) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}.$$

Nous en déduisons que :

$$\gamma(h) = \frac{\sigma^2\Gamma(h+d)}{\Gamma(h+1)\Gamma(d)} \alpha^h \frac{\Gamma(h+1)\Gamma(1-2d)}{\Gamma(h+1-d)\Gamma(1-d)}.$$

PROPRIÉTÉ 8 : Les autocovariances d'un processus MADE (1) à mémoire longue sont données par :

$$\gamma(h) = \sigma^2 \alpha^h \frac{\Gamma(h+d)}{\Gamma(h+1-d)} \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)}, \quad h \geq 0,$$

et les autocorrélations par :

$$\rho(h) = \alpha^h \frac{\Gamma(1-d)}{\Gamma(d)} \frac{\Gamma(h+d)}{\Gamma(h+1-d)}, \quad h \geq 0.$$

On peut donner des formes équivalentes pour l'expression de $\gamma(h)$. Ainsi remarquant que : $\frac{\Gamma(d+1)}{\Gamma(d-h+1)} = (-1)^h \frac{\Gamma(-d+h)}{\Gamma(-d)}$, on obtient :

$$\gamma(h) = \sigma^2 (-\alpha)^h \frac{\Gamma(-d+1)}{\Gamma(-d-h+1)} \frac{\Gamma(1-2d)}{\Gamma(h+1-d)\Gamma(1-d)}.$$

$$(27) \quad \gamma(h) = \sigma^2 (-\alpha)^h \frac{\Gamma(1-2d)}{\Gamma(-d-h+1)\Gamma(h+1-d)}.$$

Remarquons que la fonction du second membre étant paire en h l'égalité est vraie pour toutes les valeurs de h aussi bien positives que négatives.

Remarque : Les formules précédentes peuvent aussi être trouvées par inversion de la densité spectrale. Ainsi considérons le cas $\alpha = 1$, on a :

$$f(\omega) = \frac{\sigma^2}{2\pi} \left[4 \sin^2 \frac{\omega}{2} \right]^\delta = \frac{\sigma^2}{2\pi} 2^{2\delta} \left(\sin \frac{\omega}{2} \right)^{2\delta}$$

et

$$\begin{aligned} \gamma(h) &= \int_{-\pi}^{\pi} \cos(\omega h) f(\omega) d\omega = \int_0^{2\pi} \cos(\omega h) f(\omega) d\omega \\ &= \frac{\sigma^2}{2\pi} 2^{2\delta} \int_0^{2\pi} \cos(\omega h) \left(\sin \frac{\omega}{2} \right)^{2\delta} d\omega \\ &= \frac{\sigma^2}{2\pi} 2^{2\delta+1} \int_0^{\pi} \cos(2\lambda h) (\sin \lambda)^{2\delta} d\lambda. \end{aligned}$$

On peut alors utiliser une formule classique d'intégration de produit de sinus et de cosinus (appendice 2, formule (48))

$$\int_0^{\pi} (\sin x)^{b-1} \cos ax dx = \frac{\pi \cos(a\pi/2)}{2^{b-1} b \mathbf{B}[(b+a+1)/2, (b-a+1)/2]}, \quad \text{Re}(b) > 0,$$

pour en déduire que :

$$\begin{aligned} \gamma(h) &= \frac{\sigma^2}{\pi} 2^{2\delta} \frac{\pi \cos(h\pi)}{2^{2\delta} (2\delta+1) \mathbf{B}[\delta+h+1, \delta-h+1]} \\ &= \sigma^2 \frac{\cos(\pi h) \Gamma(2\delta+2)}{(2\delta+1) \Gamma(\delta+h+1) \Gamma(\delta-h+1)} \\ &= \sigma^2 (-1)^h \frac{\Gamma(2\delta+1)}{\Gamma(\delta+h+1) \Gamma(\delta-h+1)}. \end{aligned}$$

Une telle approche a par exemple été suivie par HOSKING [1981] et GRANGER-JOYEUX [1980].

Remarque : Utilisant l'équivalence

$$F(a, b+\lambda, c+\lambda; z) \approx \frac{1}{(1-z)^a}, \quad \text{si } |z| < 1 \text{ et si } \lambda \text{ tend vers } \pm \infty,$$

on voit que, lorsque h est grand, l'autocovariance est approximativement donnée par :

$$\begin{aligned} \gamma(h) &= \sigma^2 \frac{\Gamma(h+d)}{\Gamma(h+1)\Gamma(d)} \alpha^h F(d, d+h, h+1, \alpha^2) \\ &\approx \sigma^2 \frac{h^{d-1} \alpha^h}{\Gamma(d)(1-\alpha^2)^d} \quad \text{pour } |\alpha| < 1. \end{aligned}$$

Lorsque $\alpha = 1$, on a :

$$\gamma(h) = \sigma^2 \frac{\Gamma(h+d)}{\Gamma(h+1-d)} \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} \approx \sigma^2 h^{2d-1} \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)}.$$

Il y a, dans ce cas limite, décroissance des autocovariances à une vitesse* plus lente que la vitesse exponentielle usuelle.

4.b. Processus MADE (2)

Considérons maintenant un processus MADE d'ordre deux. Il satisfait une relation du type :

$$(28) \quad (1-\alpha_1 L)^{-\delta_1} (1-\alpha_2 L)^{-\delta_2} X_t = \varepsilon_t$$

$$\Leftrightarrow (29) \quad X_t = (1-\alpha_1 L)^{\delta_1} (1-\alpha_2 L)^{\delta_2} \varepsilon_t.$$

Comme dans le paragraphe précédent nous ne faisons pas la distinction selon le signe de l'exposant.

4.b.1. Coefficients du développement moyenne mobile infinie

Remplaçant dans (29), les opérateurs $(1-\alpha_1 L)^{\delta_1}$ et $(1-\alpha_2 L)^{\delta_2}$ par leurs développements en série nous obtenons :

$$A(L) = \left(\sum_{j=0}^{\infty} \frac{\Gamma(d_1+j)}{\Gamma(d_1)} \frac{\alpha_1^j L^j}{j!} \right) \left(\sum_{j=0}^{\infty} \frac{\Gamma(d_2+j)}{\Gamma(d_2)} \frac{\alpha_2^j L^j}{j!} \right),$$

de sorte que le terme général du développement moyenne mobile infinie est donné par :

$$a_k = \sum_{j=0}^k \frac{\Gamma(d_1+j)}{\Gamma(d_1)} \frac{\alpha_1^j}{j!} \frac{\Gamma(d_2+k-j)}{\Gamma(d_2)} \frac{\alpha_2^{k-j}}{(k-j)!}$$

$$\begin{aligned}
 &= \alpha_2^k \sum_{j=0}^k \frac{\Gamma(d_1+j)}{\Gamma(d_1)} \frac{\Gamma(d_2+k-j)}{\Gamma(d_2)} \frac{1}{\Gamma(k-j+1)} \frac{(\alpha_1/\alpha_2)^j}{j!} \\
 &= \alpha_2^k \frac{\Gamma(d_2+k)}{\Gamma(d_2)\Gamma(k+1)} \sum_{j=0}^k \frac{\Gamma(d_1+j)}{\Gamma(d_1)} \frac{\Gamma(d_2+k-j)}{\Gamma(d_1)} \frac{\Gamma(k+1)}{\Gamma(k+1-j)} \frac{(\alpha_1/\alpha_2)^j}{j!}.
 \end{aligned}$$

Remarquant que :

$$\frac{\Gamma(d_2+k-j)}{\Gamma(d_2+k)} = (-1)^j \frac{\Gamma(-d_2-k+1)}{\Gamma(-d_2-k+1+j)}$$

et que :

$$\frac{\Gamma(k+1)}{\Gamma(k+1-j)} = (-1)^j \frac{\Gamma(-k+j)}{\Gamma(-k)}$$

(d'après appendice 2, formule (45)) on obtient :

$$a_k = \alpha_2^k \frac{\Gamma(d_2+k)}{\Gamma(d_2)\Gamma(k+1)} \sum_{j=0}^k \frac{\Gamma(d_1+j)}{\Gamma(d_1)} \frac{\Gamma(-d_2-k+1)}{\Gamma(-d_2-k+1+j)} \frac{\Gamma(-k+j)}{\Gamma(-k)} \frac{(\alpha_1/\alpha_2)^j}{j!}$$

Comme $\frac{\Gamma(-k+j)}{\Gamma(-k)} = 0$ si $j \geq k+1$, on peut aussi écrire :

$$a_k = \alpha_2^k \frac{\Gamma(d_2+k)}{\Gamma(d_2)\Gamma(k+1)} \sum_{j=0}^{\infty} \frac{\Gamma(d_1+j)}{\Gamma(d_1)} \frac{\Gamma(-d_2-k+1)}{\Gamma(-d_2-k+1+j)} \frac{\Gamma(-k+j)}{\Gamma(-k)} \frac{(\alpha_1/\alpha_2)^j}{j!},$$

$$(30) \quad a_k = \alpha_2^k \frac{\Gamma(d_2+k)}{\Gamma(d_2)\Gamma(k+1)} F \left[d_1; -k; -d_2-k+1; \frac{\alpha_1}{\alpha_2} \right]$$

Remarque : Lorsque $\alpha_1 = \alpha_2$, on a :

$$F(d_1; -k; -d_2-k+1; 1) = \frac{\Gamma(-d_2-k+1)\Gamma(-d_2-d_1+1)}{\Gamma(-d_2+1)\Gamma(-d_2-d_1-k+1)}$$

(d'après la formule générale donnant $F(a, b, c, 1)$ voir Appendice 2 formule (44)).

Donc :

$$F(d_1; -k; -d_2-k+1; 1) = \frac{\Gamma(d_2)}{\Gamma(d_2+k)} \frac{\Gamma(d_2+d_1+k)}{\Gamma(d_2+d_1)}$$

(d'après Appendice 2 formule (45)) et l'expression de

$$a_k = \alpha_2^k \frac{\Gamma(d_2+d_1+k)}{\Gamma(d_2+d_1)\Gamma(k+1)}$$

coïncide bien avec celle correspondant au processus MADE (1) avec $\alpha = \alpha_2$, $d = d_1 + d_2$.

4.b.2. Formes des autocovariances

Une expression simple de l'autocovariance d'ordre h en terme de fonctions hypergéométriques ne peut être donnée que pour des processus MADE (2) particuliers, par exemple les processus satisfaisant :

$$(1 - \alpha_1 L)(1 - L)^{-\delta_2} X_t = \varepsilon_t$$

Posant : $Y_t = (1 - L)^{\delta_2} \varepsilon_t$, on a en effet :

$$X_t = \sum_{j=0}^{\infty} \alpha_1^j Y_{t-j}$$

et cette relation entraîne que les autocovariances de X et Y sont liées par :

$$\begin{aligned} \gamma_X(h) &= \sum_{j=0}^{\infty} \sum_{l=0}^{\infty} \alpha_1^{j+l} \gamma_Y(h+l-j) \\ &= \sum_{j=0}^{\infty} \sum_{l=-j}^{\infty} \alpha_1^{2j+l} \gamma_Y(h+l) \\ &= \sum_{j=0}^{\infty} \sum_{l=-j}^0 \alpha_1^{2j+l} \gamma_Y(h+l) + \sum_{j=0}^{\infty} \sum_{l=1}^{\infty} \alpha_1^{2j+l} \gamma_Y(h+l) \\ &= \sum_{l=-\infty}^0 \sum_{j=-l}^{\infty} \alpha_1^{2j+l} \gamma_Y(h+l) + \sum_{j=0}^{\infty} \sum_{l=1}^{\infty} \alpha_1^{2j+l} \gamma_Y(h+l) \end{aligned}$$

Dans ces sommations, on peut alors séparer celles où intervient l'indice j :

$$\begin{aligned} \gamma_X(h) &= \sum_{l=-\infty}^0 \alpha_1^l \gamma_Y(h+l) \left[\sum_{j=-l}^{\infty} \alpha_1^{2j} \right] + \left[\sum_{l=1}^{\infty} \alpha_1^l \gamma_Y(h+l) \right] \left[\sum_{j=0}^{\infty} \alpha_1^{2j} \right] \\ &= \sum_{l=-\infty}^0 \alpha_1^l \gamma_Y(h+l) \frac{\alpha_1^{-2l}}{1 - \alpha_1^2} + \sum_{l=1}^{\infty} \alpha_1^l \gamma_Y(h+l) \frac{1}{1 - \alpha_1^2} \\ &= \frac{1}{1 - \alpha_1^2} \left[\sum_{l=0}^{\infty} \alpha_1^l \gamma_Y(h-l) + \sum_{l=1}^{\infty} \alpha_1^l \gamma_Y(h+l) \right]. \end{aligned}$$

Nous pouvons donc donner une expression maniable de $\gamma_X(h)$, lorsque les sommes du second membre se simplifient.

Comme le processus Y est MADE (1) à mémoire longue, on a :

$$\gamma_Y(h) = \sigma^2 \frac{\Gamma(h+d_2)}{\Gamma(h+1-d_2)} \frac{\Gamma(1-2d_2)}{\Gamma(d_2)\Gamma(1-d_2)}, \quad \text{d'après la propriété (8).}$$

Remplaçant dans l'expression de $\gamma_X(h)$, on en déduit :

$$\begin{aligned} \gamma_X(h) &= \frac{\sigma^2}{1 - \alpha_1^2} \frac{\Gamma(1-2d_2)}{\Gamma(d_2)\Gamma(1-d_2)} \\ &\times \left[\sum_{l=0}^{\infty} \frac{\Gamma(h-l+d_2)}{\Gamma(h-l+1-d_2)} \alpha_1^l + \sum_{l=1}^{\infty} \frac{\Gamma(h+l+d_2)}{\Gamma(h+l+1-d_2)} \alpha_1^l \right] \end{aligned}$$

$$\begin{aligned}
 &= \frac{\gamma_Y(h)}{1-\alpha_1^2} \left[\sum_{l=0}^{\infty} \frac{\Gamma(h-l+d_2)}{\Gamma(h+d_2)} \frac{\Gamma(h+1-d_2)}{\Gamma(h-l+1-d_2)} \frac{\Gamma(l+1)}{\Gamma(1)} \frac{\alpha_1^l}{l!} \right. \\
 &\quad \left. + \sum_{l=1}^{\infty} \frac{\Gamma(h+l+d_2)}{\Gamma(h+d_2)} \frac{\Gamma(h+1-d_2)}{\Gamma(h+l+1-d_2)} \frac{\Gamma(l+1)}{\Gamma(1)} \frac{\alpha_1^l}{l!} \right] \\
 &= \frac{\gamma_Y(h)}{1-\alpha_1^2} \left[\sum_{l=0}^{\infty} \frac{\Gamma(-d_2-h+1)}{\Gamma(-d_2-h+l+1)} \frac{\Gamma(d_2-h+l)}{\Gamma(d_2-h)} \frac{\Gamma(l+1)}{\Gamma(1)} \frac{\alpha_1^l}{l!} \right. \\
 &\quad \left. + \sum_{l=0}^{\infty} \frac{\Gamma(h+l+d_2)}{\Gamma(h+d_2)} \frac{\Gamma(h+1-d_2)}{\Gamma(h+l+1-d_2)} \frac{\Gamma(l+1)}{\Gamma(1)} \frac{\alpha_1^l}{l!} - 1 \right] \\
 &= \frac{\gamma_Y(h)}{1-\alpha_1^2} \{F[1; d_2-h; -d_2-h+1; \alpha_1] + F[1; d_2+h; -d_2+h+1; \alpha_1] - 1\}.
 \end{aligned}$$

PROPRIÉTÉ 9 : Le processus MADE (2) défini par :

$$(1-\alpha_1 L)(1-L)^{-d_2} X_t = \varepsilon_t$$

admet pour autocovariance :

$$\begin{aligned}
 \gamma_X(h) &= \frac{\sigma^2}{1-\alpha_1^2} \frac{\Gamma(h+d_2)}{\Gamma(h+1-d_2)} \frac{\Gamma(1-2d_2)}{\Gamma(d_2)\Gamma(1-d_2)} \\
 &\quad \times \{F[1; d_2-h; -d_2-h+1; \alpha_1] \\
 &\quad + F[1; d_2+h; -d_2+h+1; \alpha_1] - 1\}.
 \end{aligned}$$

4. c. Résolution par la méthode d'André

Dans le cas d'un processus MADE quelconque les coefficients (a_j) du développement moyenne mobile peuvent être explicités par résolution directe de l'équation de récurrence :

$$(7) \quad c_0 a_j + \dots + [-b_{\kappa-1} + c_{\kappa}(j-\kappa)] a_{j-\kappa} = 0, \quad \forall j \in \mathbb{Z},$$

que nous écrirons de façon à alléger les notations :

$$(31) \quad p_0(j) a_j + \dots + p_{\kappa}(j-\kappa) a_{j-\kappa} = 0, \quad \forall j \in \mathbb{Z}.$$

Dans cette écriture $p_0(j) \dots p_{\kappa}(j)$ désignent des polynômes de degré un. Si nous normons le terme d'indice le plus élevé, nous pouvons réécrire (31) sous la forme :

$$(32) \quad a_j - q_1(j) a_{j-1} \dots - q_{\kappa}(j) a_{j-\kappa} = 0, \quad \forall j \neq 0,$$

$$\text{avec } q_{\kappa}(j) = -\frac{p_{\kappa}(j-\kappa)}{p_0(j)}.$$

La méthode d'ANDRÉ [1878], JORDAN [1965] consiste à interpréter une telle équation de récurrence comme un système linéaire dans les coefficients inconnus (a_j) .

Nous avons le système :

$$\begin{bmatrix} 1 - q_1(J) & -q_2(J) & \cdots & -q_K(J) & 0 & \cdots & 0 \\ 0 & 1 & -q_1(J-1) & -q_{K-1}(J-1) & -q_K(J-1) & \cdots & 0 \\ \vdots & & 1 & \cdots & & & \\ 0 & & & & & & 1 & -q_1(2) \\ & & & & & & & 1 \end{bmatrix} \times \begin{bmatrix} a_j \\ a_{j-1} \\ \vdots \\ a_1 \end{bmatrix} = \begin{bmatrix} y_j \\ y_{j-1} \\ \vdots \\ y_1 \end{bmatrix}$$

dans lequel le second membre se déduit immédiatement des conditions initiales. On a :

$$y_j = 0, \quad \forall j \geq K + 1,$$

$$y_j = q_j(j) a_0 = -\frac{p_j(0)}{p_0(j)} a_0 = \frac{b_{j-1}}{c_{0j}}, \quad \text{si } j \leq K.$$

Remarquant que la matrice du système précédent est triangulaire supérieure avec des 1 sur la diagonale, on voit immédiatement que le coefficient a_j est égal au déterminant de la matrice :

$$\begin{bmatrix} y_j & -q_1(J) & -q_2(J) & \cdots & 0 \\ y_{j-1} & 1 & -q_1(J-1) & \cdots & \vdots \\ \vdots & \vdots & 1 & & \vdots \\ \vdots & \vdots & \vdots & 1 & -q_1(2) \\ y_1 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

Il est donc de la forme : $a_j = \sum_{h=0}^{j-1} y_{j-h} \psi[J, h]$ en désignant par $\psi[J, h]$ le cofacteur de y_{j-h} .

Un calcul direct montre que :

$$\begin{aligned} \psi[J, 0] &= 1, \\ \psi[J, 1] &= q_1(J), \\ \psi[J, 2] &= q_1(J) q_1(J-1) + q_2(J), \\ \psi[J, 3] &= q_1(J) q_1(J-1) q_1(J-2) + q_3(J) + q_1(J) q_2(J-1) + q_2(J) q_1(J-2) \end{aligned}$$

et, poursuivant par récurrence ces développements, on vérifie facilement que :

$$(33) \quad \begin{aligned} \psi[J, h] &= \sum_{h_1 + h_2 + \dots + h_i = h} q_{h_1}(J) q_{h_2}(J-h_1) q_{h_3}(J-h_1-h_2) \times \dots \\ &\quad \times q_{h_i}(J-h_1 \dots -h_{i-1}), \end{aligned}$$

où la somme porte sur toutes les partitions possibles de l'entier h avec répétition et permutation.

Le nombre de termes apparaissant dans cette somme est donc 2^{h-1} .

Finalement, tenant compte de la nullité de certains y_j , on a :

$$(34) \quad a_j = \sum_{k=1}^{\kappa} y_k \psi[J, J-k]$$
$$= \sum_{k=1}^{\kappa} y_k \sum_{h_1 + \dots + h_i = J-k} q_{h_1}(J) \dots q_{h_i}(J-h_1 \dots -h_{i-1}).$$

Il est clair qu'une telle formule explicite n'a pas d'intérêt numérique, puisqu'il est plus simple de résoudre de manière itérative l'équation de récurrence linéaire. Elle peut en revanche se révéler utile pour trouver dans certains cas des formes analytiques des coefficients a_j .

APPENDICE 1

Condition nécessaire et suffisante

pour que $\sum_{j=0}^{\infty} a_j^2 < \infty$

i. Remarquons d'abord que la condition $\sum_{j=0}^{\infty} a_j^2 < \infty$ est équivalente au fait que la fonction $\omega \mapsto A(\exp i\omega)$ est de carré intégrable sur $[0, 2\pi[$ par rapport à la mesure de Lebesgue (et en particulier existe au sens de $L^2([0, 2\pi[, \lambda_{[0, 2\pi[})$). La série $A(\exp i\omega)$ existe donc $\lambda_{[0, 2\pi[}$, p. p. Ceci entraîne que le rayon de convergence de $A(z) = \sum_{j=0}^{\infty} a_j z^j$ est supérieur ou égal à 1.

ii. Remarquons de plus que par hypothèse $A(0) = a_0 = 1$ et donc que $A(z)$ est non nulle sur un disque ouvert non vide centré en zéro : $\dot{B}(0, R)$. Nous choisissons dans la suite le rayon R de ce disque inférieur au plus petit module des racines du polynôme C et au rayon de convergence de A . Ceci est possible, puisque C n'admet pas zéro comme racine (en effet $(b_{-1}, c_0) \neq (0, 0)$ entraîne $c_0 = C(0) \neq 0$ à cause de la première restriction $b_{-1} = 0$).

iii. Introduisant les polynômes : $B(z) = b_0 + b_1 z + \dots + b_{k-1} z^{k-1}$ et $C(z) = c_0 + c_1 z + \dots + c_k z^k$, on voit facilement que la relation de récurrence (8) est équivalente à l'égalité :

$$(35) \quad \frac{dA(z)}{dz} C(z) - A(z) B(z) = 0, \quad \forall z \in \dot{B}(0, R),$$

ou de manière équivalente à :

$$(36) \quad \frac{1}{A(z)} \frac{dA(z)}{dz} = \frac{B(z)}{C(z)}, \quad \forall z \in \dot{B}(0, R).$$

iv. Utilisant l'hypothèse 1 (§ 3. a), on peut décomposer la fraction rationnelle $\frac{B(z)}{C(z)}$ en éléments simples ne comprenant que des facteurs d'ordre 1.

Ce nombre K de facteurs est égal au nombre de racines de C , qui ne sont pas également racines de B . On a :

$$\frac{1}{A(z)} \frac{dA(z)}{dz} = \sum_{k=1}^K \frac{-\alpha_k \delta_k}{1 - \alpha_k z} \quad \text{avec } \alpha_k \neq 0, \delta_k \neq 0.$$

Il est à ce niveau intéressant de séparer les termes correspondant à une valeur positive ou nulle de la partie réelle de δ de ceux correspondant à

une valeur négative de celle-ci. Avec un changement de notation, nous avons :

$$\frac{1}{A(z)} \frac{dA(z)}{dz} = \sum_{k=1}^{k_0} \frac{-\alpha_k \delta_k}{1-\alpha_k z} + \sum_{k=1}^{k_1} \frac{\beta_k d_k}{1-\beta_k z}, \quad \forall z \in \hat{B}(0, R),$$

avec $\text{Re } \delta_k \geq 0$ et $\text{Re } d_k > 0$.

Les fonctions apparaissant dans les deux membres étant holomorphes sur $\hat{B}(0, R)$ admettent des primitives définies à une constante près (CARTAN [1964], p. 61). En terme de primitives l'équation devient :

$$A(z) = \text{Cte} \prod_{k=1}^{k_0} (1-\alpha_k z)^{\delta_k} \prod_{k=1}^{k_1} (1-\beta_k z)^{-d_k}.$$

Comme $A(0) = a_0 = 1$, on en déduit la forme obligée de :

$$A(z) = \prod_{k=1}^{k_0} (1-\alpha_k z)^{\delta_k} \prod_{k=1}^{k_1} (1-\beta_k z)^{-d_k} \quad \text{sur la boule } \hat{B}(0, R).$$

Les deux fonctions apparaissant dans l'égalité précédente sont analytiques sur $\hat{B}(0, R)$ et ont même développement en série. Les coefficients a_j sont donc égaux aux coefficients apparaissant dans le développement de :

$$\prod_{k=1}^{k_0} (1-\alpha_k z)^{\delta_k} \prod_{k=1}^{k_1} (1-\beta_k z)^{-d_k} = \sum_{j=0}^{\infty} a_j z^j.$$

Il reste donc uniquement à regarder dans quels cas ces coefficients satisfont la condition $\sum_{j=0}^{\infty} a_j^2 < \infty$.

v. Pour cela nous pouvons d'après i commencer par étudier le rayon de convergence de la série correspondant au développement de :

$$\prod_{k=1}^{k_0} (1-\alpha_k z)^{\delta_k} \prod_{k=1}^{k_1} (1-\beta_k z)^{-d_k}.$$

a) Remarquons d'abord que s'il existe un coefficient β_k de module strictement supérieur à 1 ou s'il existe un couple (α_k, δ_k) avec $|\alpha_k| > 1$ et δ_k non entier, le rayon de convergence de la série est strictement inférieur à 1. Ceci n'est pas compatible avec i et dans ce cas il n'existe pas de processus stationnaire, dont les coefficients de la décomposition moyenne mobile satisfont la relation (8).

b) Un autre cas simple à étudier est celui, où tous les coefficients β_k sont de modules strictement inférieurs à 1 et où tous les coefficients α_k associés à une valeur non entière de δ_k sont aussi de modules strictement inférieur à 1. Dans ce cas le rayon de convergence est strictement supérieur à 1. Ceci entraîne l'existence de $\sum_{j=0}^{\infty} |a_j|$, donc aussi celle de $\sum_{j=0}^{\infty} a_j^2$.

c) Finalement, il reste à étudier le cas limite, où le rayon de convergence est égal à 1. Ceci se produit, lorsque tous les coefficients β_k sont de modules

inférieurs ou égaux à 1, que les coefficients α_k associés aux δ_k non entiers sont de modules inférieurs ou égaux à 1 et que certains de ces coefficients β_k et/ou α_k sont de module 1.

Nous allons simultanément étudier ces deux types de conditions en distinguant les parties nécessaire et suffisante de la proposition.

vi. Supposons que les coefficients α_k, β_k satisfont les conditions précédentes et que pour les indices tels que $|\beta_k|=1$, on ait $\operatorname{Re} d_k < \frac{1}{2}$. On sait alors que les développements en série des divers termes intervenant dans le produit, par exemple :

$$(1 - \beta_k z)^{-d_k} = 1 + (-d_k)(-\beta_k z) + \dots + \frac{(-d_k)(-d_k-1)\dots(-d_k-j+1)}{j!} (-\beta_k z)^j + \dots$$

ont des coefficients de carré sommable (voir ABRAMOWITZ-STEGUN [1965] 15. 1). On peut donc donner un sens $\lambda_{[0, 2\pi[}$ p. p. au produit :

$$A(\exp i\omega) = \frac{\prod_{k=1}^{k_0} (1 - \alpha_k \exp i\omega)^{\delta_k}}{\prod_{k=1}^{k_1} (1 - \beta_k \exp i\omega)^{d_k}}$$

Pour savoir si les coefficients du développement en série de $A(\exp i\omega)$ sont de carrés sommables, il suffit alors d'examiner l'intégrabilité de $|A(\exp i\omega)|^2$ sur $[0, 2\pi[$.

Cette intégration pose uniquement un problème aux pôles du dénominateur. Ces pôles ω_k correspondent aux indices k pour lesquels $|\beta_k|=1$ et sont tels que $\beta_k \exp i\omega_k = 1$. Comme les termes β_k sont par hypothèse distincts, le développement limité de $A(\exp i\omega)$ au voisinage d'un pôle ω_{k_0} particulier montre immédiatement que $|A(\exp i\omega)|^2$ est équivalent à

Cte $\times \frac{1}{|(1 - \beta_{k_0} \exp i\omega)^{d_{k_0}}|^2}$. La condition $\operatorname{Re} d_{k_0} < \frac{1}{2}$ impliquant l'intégrabi-

lité de $\frac{1}{|(1 - \beta_{k_0} \exp i\omega)^{d_{k_0}}|^2}$, on voit que $|A(\exp i\omega)|^2$ est intégrable au voisinage de chacun de ses pôles. Comme le nombre de pôles est fini, que le domaine d'intégration est borné, on en déduit que $|A(\exp i\omega)|^2$ est intégrable sur l'ensemble du domaine $[0, 2\pi[$.

vii. Il nous reste maintenant à montrer que la condition $\operatorname{Re} d_k < \frac{1}{2}$ satisfaite pour les indices correspondant à $|\beta_k|=1$ est nécessaire.

a) Pour cela, nous avons besoin du lemme suivant :

LEMME 10 : Si $\sum_{j=0}^{\infty} |a_j|^2 < \infty$ la suite $\sum_{j=0}^j a_j \left(\frac{\exp i\omega}{\rho}\right)^j$ converge dans $L^2([0, 2\pi[, \lambda_{[0, 2\pi[})$ uniformément en ρ pour $\rho \geq 1$.

Preuve : En effet pour toute valeur de ρ , $\rho \geq 1$, on a : $\sum_{j=0}^{\infty} \frac{|a_j|^2}{\rho^{2j}} < \infty$. Donc la série $\sum_{j=0}^{\infty} a_j \left(\frac{\exp i\omega}{\rho} \right)^j$ existe au sens de L^2 . Nous notons $F_\rho(\omega)$ sa somme. On a alors :

$$\left\| F_\rho(\omega) - \sum_{j=0}^J a_j \left(\frac{\exp i\omega}{\rho} \right)^j \right\|_2^2 = \sum_{j=J+1}^{\infty} \frac{|a_j|^2}{\rho^{2j}} \leq \sum_{j=J+1}^{\infty} |a_j|^2.$$

La majoration étant indépendante de ρ , $\rho \geq 1$, on en déduit le résultat cherché. \square

Remarquant que, d'après l'inégalité triangulaire, on a :

$$\begin{aligned} \|F_1(\omega) - F_\rho(\omega)\|_2 &\leq \left\| F_1(\omega) - \sum_{j=0}^J a_j \exp ij\omega \right\|_2 \\ &\quad + \left\| \sum_{j=0}^J a_j \exp ij\omega - \sum_{j=0}^J a_j \frac{\exp ij\omega}{\rho^j} \right\|_2 \\ &\quad + \left\| F_\rho(\omega) - \sum_{j=0}^J a_j \frac{\exp ij\omega}{\rho^j} \right\|_2, \end{aligned}$$

on déduit du lemme que :

$$\forall \varepsilon > 0, \exists J_0, \forall J \geq J_0, \forall \rho \geq 1, \left\| F_\rho(\omega) - \sum_{j=0}^J a_j \frac{\exp ij\omega}{\rho^j} \right\|_2 < \varepsilon$$

et que :

$$\forall \varepsilon, \exists J_0, \forall \rho \geq 1, \|F_1(\omega) - F_\rho(\omega)\|_2 \leq 2\varepsilon + \left\| \sum_{j=1}^{J_0} a_j \exp ij\omega - \sum_{j=1}^{J_0} a_j \frac{\exp ij\omega}{\rho^j} \right\|_2.$$

Faisant tendre ρ vers 1, on a :

$$\forall \varepsilon > 0, \limsup_{\rho \rightarrow 1} \|F_1(\omega) - F_\rho(\omega)\|_2 \leq 2\varepsilon.$$

Cette inégalité étant satisfaite pour toute valeur $\varepsilon > 0$, on a le corollaire suivant :

COROLLAIRE 11 : $F_\rho(\omega)$ converge au sens de $L^2([0, 2\pi[, \lambda_{10}, 2\pi])$ vers $F_1(\omega)$, lorsque ρ tend vers 1 en décroissant.

b) Comme $\sum_{j=0}^{\infty} |a_j|^2 < \infty$, nous pouvons donner un sens à la somme $\sum_{j=0}^{\infty} a_j \frac{\exp ij\omega}{\rho^j} = A \left(\frac{\exp i\omega}{\rho} \right)$ pour $\rho > 1$ et savons qu'elle coïncide nécessaire-

ment avec la fonction :

$$A\left(\frac{\exp i \omega}{\rho}\right) = \frac{\prod_{k=1}^{\kappa_0} (1 - (\alpha_k/\rho) \exp i \omega)^{\delta_k}}{\prod_{k=1}^{\kappa_1} (1 - (\beta_k/\rho) \exp i \omega)^{d_k}}$$

pour $\rho > 1$, ceci quelles que soient les valeurs des exposants d_k, δ_k .

Cependant comme les exposants ne sont pas *a priori* contraints, nous ne

savons pas si la fonction $\frac{\prod_{k=1}^{\kappa_0} (1 - \alpha_k \exp i \omega)^{\delta_k}}{\prod_{k=1}^{\kappa_1} (1 - \beta_k \exp i \omega)^{d_k}}$ a ou non un sens.

On peut malgré tout, d'après le corollaire précédent, dire que :

$$A(\exp i \omega) = \lim_{\rho \rightarrow 1} \frac{\prod_{k=1}^{\kappa_0} (1 - (\alpha_k/\rho) \exp i \omega)^{\delta_k}}{\prod_{k=1}^{\kappa_1} (1 - (\beta_k/\rho) \exp i \omega)^{d_k}},$$

où la limite est au sens de $L^2([0, 2\pi[, \lambda_{(0, 2\pi[})$.

Donc :

$$+\infty > \int_0^{2\pi} |A(\exp i \omega)|^2 d\omega = \lim_{\rho \rightarrow 1} \int_0^{2\pi} \left| \frac{\prod_{k=1}^{\kappa_0} (1 - (\alpha_k/\rho) \exp i \omega)^{\delta_k}}{\prod_{k=1}^{\kappa_1} (1 - (\beta_k/\rho) \exp i \omega)^{d_k}} \right|^2 d\omega.$$

On vérifie alors facilement par un calcul direct qu'une condition nécessaire pour que la suite :

$$\int_0^{2\pi} \left| \frac{\prod_{k=1}^{\kappa_0} (1 - (\alpha_k/\rho) \exp i \omega)^{\delta_k}}{\prod_{k=1}^{\kappa_1} (1 - (\beta_k/\rho) \exp i \omega)^{d_k}} \right|^2 d\omega$$

reste bornée, lorsque ρ tend vers 1, est que $\operatorname{Re} d_k < \frac{1}{2}$. Ceci montre donc la condition nécessaire.

Remarquons d'ailleurs qu'utilisant maintenant la partie suffisante on voit qu'il était effectivement possible d'écrire :

$$A(\exp i \omega) = \frac{\prod_{k=1}^{\kappa_0} (1 - \alpha_k \exp i \omega)^{\delta_k}}{\prod_{k=1}^{\kappa_1} (1 - \beta_k \exp i \omega)^{d_k}}$$

avec une égalité au sens de L^2 .

APPENDICE 2

Quelques formules concernant la fonction hypergéométrique

Références : ABRAMOWITZ-STEGUN [1965] [A. S.], ERDELYI [1953] [E.],
GRADSHTEYN-RYZHIK [1965] [G. R.].

1. Définition

$$(37) \quad F(a, b, c, z) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{j=0}^{\infty} \frac{\Gamma(a+j)\Gamma(b+j)}{\Gamma(c+j)} \frac{z^j}{j!}$$

[(A. S.) 15. 1. 1].

2. Relations de contiguïté de Gauss

$$(38) \quad (c-b-1)F(a, b, c, z) + bF(a, b+1, c, z) - (c-1)F(a, b, c-1, z) = 0$$

[(A. S.) 15. 2. 24].

$$(39) \quad c(1-z)F(a, b, c, z) - cF(a, b-1, c, z) + (c-a)zF(a, b, c+1, z) = 0$$

[(A. S.) 15. 2. 25].

On déduit de ces deux relations, la nouvelle égalité :

$$(40) \quad (c-1)F(a, b-1, c-1, z) + (az-bz-c+1)F(a, b, c, z) \\ + \frac{b(c-a)}{c} zF(a, b+1, c+1, z) = 0.$$

Preuve : Remplaçant dans l'équation (38), $F(a, b+1, c, z)$ par son expression déduite de (39), c'est-à-dire par :

$$F(a, b+1, c, z) = \frac{1}{c(1-z)} [cF(a, b, c, z) - (c-a)zF(a, b+1, c+1, z)],$$

et $F(a, b, c-1, z)$ par son expression déduite de (39) c'est-à-dire par :

$$(c-1)F(a, b, c-1, z) = \frac{1}{1-z} [(c-1)F(a, b-1, c-1, z) \\ - (c-1-a)zF(a, b, c, z)],$$

nous obtenons :

$$\left[(c-b-1) + \frac{b}{(1-z)} + \frac{(c-1-a)z}{1-z} \right] F(a, b, c, z) - \frac{b(c-a)}{c(1-z)} z F(a, b+1, c+1, z) - \frac{c-1}{1-z} F(a, b-1, c-1, z) = 0$$

$$\Leftrightarrow (c-1)F(a, b-1, c-1, z) - (bz-az+c-1)F(a, b, c, z) + \frac{b(c-a)}{c} z F(a, b+1, c+1, z) = 0 \quad \square$$

3. Fonctions s'exprimant en terme de fonctions hyperg om etriques

- (41) $(1-\alpha z)^\delta = F(-\delta, b, b, \alpha z)$ [(A.S.) 15. 1. 8].
- (42) $\cos(2az) = F\left(-a, a, \frac{1}{2}, \sin^2 z\right)$ [(A.S.) 15. 1. 17].
- (43) $\pi \sec(\pi z) = \Gamma\left(\frac{1}{2}-z\right)\Gamma\left(\frac{1}{2}+z\right)$ [(E.) p. 3].

4. Propri t s diverses de la fonction hyperg om etrique

- (44) $F(a, b, c, 1) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}$ [(A.S.) 15. 1. 20].
- (45) $\frac{\Gamma(d+1)}{\Gamma(d-h+1)} = (-1)^h \frac{\Gamma(-d+h)}{\Gamma(-d)}$ (par calcul direct).
- (46) $\Gamma(2z) = 2^{2z-1} \frac{\Gamma(z)\Gamma(z+(1/2))}{\sqrt{\pi}}$ [(A.S.) 6. 1. 18].
- (47) $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ [(A.S.) 6. 1. 8].

5. Formules d'int gration

(48) $\int_0^\pi (\sin x)^{b-1} \cos ax \, dx = \frac{\pi \cos(a\pi/2)}{2^{b-1} b B[(b+a+1)/2, (b-a+1)/2]}$

pour $\text{Re}(b) > 0$

[(G.R.) 3. 631. 8].

6. Formule de Saalschutz

$$(49) \quad \sum_{j=0}^n \frac{(a)_j (b)_j (-n)_j}{(c)_j (1+a+b-c-n)_j j!} = \frac{(c-a)_n (c-b)_n}{(c)_n (c-a-b)_n}$$

$$\text{où } (a)_j = \frac{\Gamma(a+j)}{\Gamma(a)}$$

[(E.) p. 66]

7. Formule de Sheppard

$$(50) \quad \frac{\Gamma(j+a)}{\Gamma(j+b)} \approx j^{a-b}, \quad \text{si } j \text{ tend vers l'infini,}$$

(déduite de la formule de Stirling [(A.S.) 6.1.37]).

● Références bibliographiques

- ABRAMOWITZ, M. et STEGUN, I. (1965). — *Handbook of Mathematical Functions*, Dover.
- ANDRÉ, D. (1878). — « Terme général d'une série quelconque », *Annales de l'École Normale Supérieure*, p. 375-408.
- AZENCOTT, R. et DACUNHA-CASTELLE, D. (1984). — *Séries d'observations irrégulières*, Masson.
- BOX, G. et JENKINS, G. (1970). — *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.
- CARTAN, H. (1961). — *Théorie élémentaire des fonctions analytiques d'une ou plusieurs variables complexes*, Hermann, Paris.
- ERDŐLYI, MAGNUS, OBERHETINGER et TRICOMI (1953). — *Higher Transcendental Functions*, Vol. 1, McGraw-Hill.
- GEWEKE, J. et PORTER-HUDAK, S. (1983). — « The Estimation and Application of Long Memory Time Series Models », *discussant paper* 8302, University of Wisconsin-Madison.
- GRADSHTEYN, I. et RYZHIK, I. (1965). — *Tables of Integrals, Series and Products*, Academic Press.
- GRANGER, C. W. (1978). — « New classes of time series models », *Statistician*, 3 et 4, 27, p. 237-253.
- GRANGER, C. W. (1980). — « Long Memory Relationships and the Aggregation of Dynamic Models », *Journal of Econometrics*, 14, p. 227-238.
- GRANGER, C. W. et JOYEUX, R. (1980). — « An Introduction to Long Memory Time Series Models and Fractional Differencing », *Journal of Time Series Analysis*, 1, p. 15-29.
- HIPEL et McLEOD (1978). — « Preservation of the rescaled adjusted range, parts 1, 2 and 3 », *Water Resources Research*, 14, 3, p. 491-518.
- HOSKING, J. (1981). — « Fractional differencing », *Biometrika*, 68, p. 165-176.
- JORDAN, C. (1965). — *Calculus of Finite Differences*, Chelsea Publishing Company, New York.
- MANDELBROT, B. et VAN NESS, J. (1968). — « Fractional Brownian Motions, Fractional Noises and Applications », *S.I.A.M. Review*, 10, p. 422-437.
- MANDELBROT, B. (1971). — « A Fast Fractional Gaussian Noise Generator », *Water Resources Research*, 7, p. 543-553.

- PORTER-HUDAK, S. (1982). — « Long-Term Memory Modelling — A Simplified Spectral Approach », *Ph. D. Thesis*, University of Wisconsin.
- ROSENBLATT, M. (1976). — « Fractional Integrals of Stationary Processes and the Central Limit Theorem », *Journal of Applied Probability*, 13, p. 723-732.
- ROZANOV, Y. (1967). — *Stationary Random Processes*, Holden-Day, San Francisco.
- SPOLIA, S., CHANDER, S. et O'CONNOR, K. (1980). — « An Autocorrelation Approach for Parameter Estimation of Fractional Order Equal-Root Autoregressive Models Using Hypergeometric Functions », *Journal of Hydrology*, 47, p. 1-17.

©

PROPRIETES LIMITES DES MOMENTS EMPIRIQUES



PROPRIETES LIMITES DES MOMENTS EMPIRIQUES

ou

(de la stationnarité à la non-stationnarité)

Dans le précédent chapitre, nous avons introduit la notion de processus stationnaire MADE, en imposant aux coefficients du développement moyenne mobile de satisfaire une équation aux différences à coefficients polynômiaux. Nous avons vu que cette classe de processus contenait comme cas particuliers les processus ARMA stationnaires ainsi que les processus, dits à mémoire longue, pour lesquels les autocovariances tendent lentement vers zéro.

Nous allons maintenant examiner les propriétés asymptotiques des moyennes empiriques et variances-covariances empiriques calculées à partir d'un tel processus.

Nous nous restreindrons pour cette étude aux processus MADE(1), définis par :

$$(1-\alpha L)^d X_t = \varepsilon_t \quad , \quad \text{où } \varepsilon \text{ est un bruit blanc .}$$

Cette restriction a essentiellement pour but d'alléger les notations et les résultats pourraient être étendus à des processus MADE généraux. Nous signalerons certaines de ces extensions, lorsqu'elles découlent directement des propriétés qui seront montrées. Dans les autres cas, on pourrait le vérifier en faisant des décompositions de fractions rationnelles en éléments simples.

Nous allons voir que les résultats limites classiques utilisés dans la théorie des processus ARMA stationnaires ne sont plus nécessairement valables. Ainsi la vitesse de convergence de la moyenne empirique n'est pas forcément en \sqrt{T} , la distribution limite des autocovariances empiriques peut ne pas être gaussienne ...

Dans le premier paragraphe, nous nous intéressons à la convergence en moyenne quadratique vers zéro de $\bar{X}_T = \frac{1}{T} \sum_{t=1}^T X_t$. Nous commençons par rappeler les résultats usuels et leurs conditions d'application. Nous regardons alors dans quelle zone de valeurs des paramètres (α, d) ils s'appliquent. A l'extérieur de cette zone, nous calculons l'expression exacte de la variance de \bar{X}_T , ce qui nous permet d'en déduire les vitesses de convergence en moyenne quadratique vers zéro. Divers résultats concernant la convergence en loi sont également donnés.

Dans le second paragraphe, nous regardons ce que deviennent les propriétés limites, lorsqu'on passe aux processus MADE non stationnaires. Divers résultats sur les vitesses de convergence et les lois limites sont directement déduits de versions du théorème de Donsker.

Le troisième paragraphe est consacré à l'étude des autocovariances empiriques. Ici aussi nous rappelons les conditions classiques assurant la convergence des autocovariances à la vitesse \sqrt{T} et leur normalité asymptotique et décrivons les zones de paramètres correspondants.

A l'extérieur de cette zone, nous vérifions pour le cas stationnaire, en suivant l'approche de Rosenblatt [1961], que la convergence en loi a lieu vers une loi non normale.

Finalement diverses propriétés asymptotiques relatives aux autocovariances empiriques et au périodogramme sont montrées pour le cas non stationnaire (paragraphe 4).

L'intérêt de ces divers résultats théoriques est essentiellement de permettre de développer des procédures d'estimation et de tests sur les processus MADE, ce qui fera l'objet des chapitres suivants.

Signalons finalement que pour alléger la présentation, nous avons mis en annexe diverses démonstrations.

I - PROPRIETES ASYMPTOTIQUES DE LA MOYENNE EMPIRIQUE

(CAS STATIONNAIRE)

I.A. - LES RESULTATS CLASSIQUES

Soit $X = (X_t, t \in \mathbb{Z})$ un processus stationnaire admettant la représentation moyenne mobile infinie :

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}, \quad \text{avec} \quad \sum_{j=0}^{\infty} a_j^2 < \infty$$

et $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ un bruit blanc, de variance $\sigma^2 > 0$. Nous notons : $(\gamma(h), h \geq 0)$ sa fonction d'autocovariance et f sa densité spectrale.

On peut distinguer trois types de résultats concernant : la variance de \bar{X}_T , la convergence en moyenne quadratique de \bar{X}_T vers zéro et la convergence en loi de $\sqrt{T} \bar{X}_T$.

Forme de la variance :

Propriété 1.1. - On a :

$$\begin{aligned} \text{V} \bar{X}_T &= \frac{1}{T^2} \sum_{t,s=1}^T \gamma(t-s) \\ &= \frac{1}{T} \sum_{h=-(T-1)}^{T-1} \left(1 - \frac{|h|}{T}\right) \gamma(h) \\ &= \int_{-\pi}^{\pi} \left(\frac{\sin \frac{1}{2} \omega T}{T \sin \frac{\omega}{2}} \right)^2 f(\omega) d\omega \end{aligned}$$

voir Anderson [1971] p. 444

Convergence en moyenne quadratique :

Propriété 1.2.-

i) Si $\sum_{h=-\infty}^{+\infty} \gamma(h) < +\infty$,

alors :
$$\lim_{T \rightarrow \infty} T V \bar{X}_T = \sum_{h=-\infty}^{+\infty} \gamma(h) .$$

ii) Si la densité spectrale f est continue en $\omega = 0$ alors :

$$\lim_{T \rightarrow \infty} T V \bar{X}_T = 2\pi f(0)$$

voir Anderson [1971] p. 459 .

Remarquons que, comme \bar{X}_T est de moyenne nulle, ces résultats entraînent que, pour tout $\delta > 0$, $T^{\frac{1}{2}-\delta} \bar{X}_T$ converge en moyenne quadratique vers zéro. On notera d'autre part que cette convergence a lieu, sans qu'il ait été nécessaire d'imposer l'absolue convergence de la suite des autocovariances. Evidemment si $\sum_{h=-\infty}^{+\infty} |\gamma(h)| < +\infty$, on a simultanément la convergence de la suite des autocovariances et la continuité de f en zéro, de sorte que les deux résultats i) et ii) sont tous deux valables.

Convergence en loi :

Propriété 1.3.- Si le bruit blanc ε est indépendant (*) et
si $\sum_{h=-\infty}^{+\infty} |\gamma(h)| < +\infty$, alors $\sqrt{T} \bar{X}_T$ converge en loi vers la
loi normale centrée, de variance : $\sum_{h=-\infty}^{+\infty} \gamma(h) = 2\pi f(0)$.

voir Anderson [1971] p. 478 .

Remarquons que l'application des trois propriétés précédentes peut poser problème dans deux cas assez différents :

i) Lorsque la densité spectrale n'est pas bornée les propriétés 1.2 ii) et 1.3. ne s'appliquent évidemment pas. Cette difficulté existe pour les modèles à mémoire lorsque :

$$(1-L)^d X_t = \varepsilon_t, \text{ avec } 0 < d < \frac{1}{2} .$$

Comme dans de tels cas, on se "rapproche" de la non-stationnarité, on peut raisonnablement penser que la vitesse de convergence sera plus lente.

ii) Lorsque la densité spectrale est continue nulle en zéro, les propriétés 1.2. ii) et 1.3. s'appliquent. Cependant, comme $\lim_{T \rightarrow \infty} V(\sqrt{T} \bar{X}_T) = 0$, on peut penser que le facteur retenu \sqrt{T} n'est pas assez important et que la convergence a lieu plus rapidement que \sqrt{T} . Nous rencontrerons cette difficulté pour les modèles : $(1-L)^d X_t = \varepsilon_t$, avec $d < 0$.

(*) Dans la suite "bruit blanc indépendant" signifie que les composantes ε_t sont indépendantes de même loi, centrée de variance σ^2 .

$$\lim_{T \rightarrow \infty} V(\sqrt{T} \bar{X}_T) = 2\pi f(o) \quad (\neq 0)$$

ii) Si ε est un bruit blanc indépendant et si $|\alpha| < 1$, d quelconque ou $\alpha = -1$, $d \leq 0$, ou $|\alpha| > 1$, d entier négatif, la série des autocovariances est absolument convergente et donc :

$$\sqrt{T} \bar{X}_T \xrightarrow{d} N [0, 2\pi f(o)] \quad .$$

Nous aurions également pu appliquer 1.2. ii) par exemple au cas $\alpha = 1$, $d < 0$. Cependant comme la densité spectrale est nulle en zéro, nous aurions trouvé une loi dégénérée. Il nous reste maintenant à étudier les cas limites :

$$\alpha = -1, \quad 0 < d < \frac{1}{2}$$

et surtout le cas : $\alpha = 1, \quad d < \frac{1}{2}$.

I.C. - ETUDE DU CAS $\alpha = -1, \quad 0 < d < \frac{1}{2}$.

Propriété 1.5.- Le résultat 1.4. ii) de normalité asymptotique s'applique également au cas : $\alpha = -1, \quad 0 < d < \frac{1}{2}$:

$$\sqrt{T} \bar{X}_T \xrightarrow{d} N [0, 2\pi f(o)] \quad .$$

Preuve :

i) Introduisons le processus : $Y_t = (1+L) X_t = X_t + X_{t-1}$.

On a : $Y_t = (1+L)^{1-d} \varepsilon_t$ de sorte que le processus Y satisfait la propriété 1.4. ii).

On a donc :

$$\sqrt{T} \bar{Y}_T \xrightarrow{d} N [0, 2\pi f_Y(o)] = N [0, \sigma^2 2^{2-2d}] .$$

ii) Remarquons pour terminer que :

$$\bar{Y}_T = \frac{1}{T} \sum_{t=1}^T Y_t = \frac{1}{T} \sum_{t=1}^T (X_t + X_{t-1}) = 2 \bar{X}_T + \frac{X_0 - X_T}{T} .$$

On a donc :

$$\sqrt{T} \bar{X}_T = \frac{1}{2} \sqrt{T} \bar{Y}_T - \frac{X_0 - X_T}{2 \sqrt{T}} .$$

Comme X est un processus stationnaire, $\frac{X_0 - X_T}{\sqrt{T}}$ tend en probabilité vers zéro lorsque T tend vers l'infini.

$\sqrt{T} \bar{X}_T$ a donc asymptotiquement même loi que $\frac{1}{2} \sqrt{T} \bar{Y}_T$,
c'est-à-dire : $\sqrt{T} \bar{X}_T \xrightarrow{d} N [0, \sigma^2 2^{-2d}] = N [0, 2\pi f_X(o)]$. ■

I.D. - ETUDE DU CAS $\alpha = 1$, $d < \frac{1}{2}$.

I.D.1. - Calcul de $V \bar{X}_T$.

Le processus X est défini par : $(1-L)^d X_t = \varepsilon_t$, $d < \frac{1}{2}$.

Il est possible dans ce cas d'obtenir la forme explicite de la variance de \bar{X}_T .

Propriété 1.6.- On a :

$$V \bar{X}_T = \frac{\sigma^2}{T^2} \frac{2^{2\delta-1}}{\sqrt{\pi}} \frac{\Gamma(\delta - \frac{1}{2})}{\Gamma(\delta)} \left\{ 1 - \frac{\Gamma(\delta)}{\Gamma(\delta+T)} \frac{\Gamma(-\delta+1+T)}{\Gamma(-\delta+1)} \right\}$$

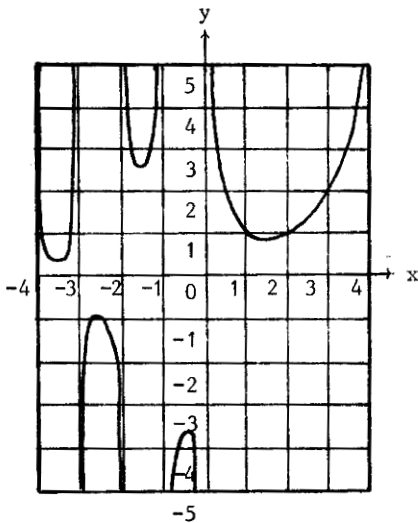
avec $\delta = -d$ et $d \neq 0$, $d \neq -\frac{1}{2}$.

Preuve : Voir Appendice 1.

La nécessité de distinguer les cas $d = 0$ et $d = -\frac{1}{2}$ provient de la définition de la fonction gamma. Si les rapports : $\frac{\Gamma(\delta+1)}{\Gamma(\delta)}$ et $\frac{\Gamma(-\delta+1)}{\Gamma(-\delta+1)}$ ont toujours un sens, il n'en est pas de même du terme

$\frac{\Gamma(\delta - \frac{1}{2})}{\Gamma(\delta)}$ figurant en facteur.

Rappelons simplement à ce niveau que la fonction gamma n'est pas définie pour les valeurs entières négatives ou nulles.



Fonction Gamma

Comme $\delta > -\frac{1}{2}$, de telles difficultés ne peuvent se produire que pour $\delta = 0$ et $\delta = \frac{1}{2}$.

i) Considérons d'abord le cas $d = 0$. Nous savons que le processus satisfait $X_t = \epsilon_t$ et est un bruit blanc. On a donc :

$$V \bar{X}_T = \frac{\sigma^2}{T} .$$

D'autre part, prenant la limite de la formule 1.6, lorsque δ tend vers zéro, on a :

$$\begin{aligned} \lim_{\delta \rightarrow 0} V \bar{X}_T &= \frac{\sigma^2}{T^2} \frac{1}{2\Gamma(\frac{1}{2})} \Gamma(-\frac{1}{2}) \left[\lim_{\delta \rightarrow 0} \left\{ \frac{1}{\Gamma(\delta)} \right\} - \frac{\Gamma(1+T)}{\Gamma(T)} \right] \\ &= -\frac{\sigma^2}{T^2} \frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{1}{2})} [0 - T] \\ &= \frac{\sigma^2}{T} . \end{aligned}$$

Corollaire 1.7.- La formule 1.6. peut être utilisée par continuité pour le cas limite $d = 0$.

ii) En revanche, si nous considérons l'autre cas limite $d = -\frac{1}{2}$ et cherchons : $\lim_{\delta \rightarrow \frac{1}{2}} V \bar{X}_T$, nous obtenons : $\lim_{\delta \rightarrow \frac{1}{2}} V \bar{X}_T = \infty$. Ceci montre qu'il n'est pas possible pour ce cas de passer la formule 1.6. par continuité. Il faut donc refaire un calcul direct.

Propriété 1.8.- Lorsque $d = -\frac{1}{2}$, on a :

$$V \bar{X}_T = \frac{2 \sigma^2}{\pi} \left\{ \frac{1}{\Gamma(T - \frac{1}{2})} + \frac{1}{T^2} \{-1 + \gamma + 2 \log 2 + \Psi[T - \frac{1}{2}] + \frac{1}{2T-1}\} \right\}$$

où γ désigne la constante d'Euler, et

$$\Psi(z) = \frac{1}{\Gamma(z)} \frac{d\Gamma(z)}{dz} .$$

Preuve : voir appendice 2.

Remarque 1.9.- Lorsque d est un entier négatif, le processus

X est un processus moyenne mobile pur avec racine multiple 1 :

$$X_t = (1-L)^\delta \varepsilon_t, \quad \delta \in \mathbb{N}.$$

Dans ce cas particulier, la formule 1.6 se réduit à :

$$V \bar{X}_T = \frac{\sigma^2}{T^2} \frac{2^{2\delta-1}}{\sqrt{\pi}} \frac{\Gamma(\delta - \frac{1}{2})}{\Gamma(\delta)}, \quad \text{si } T \geq \delta$$

(puisque le terme $\frac{\Gamma(1+T-\delta)}{\Gamma(1-\delta)} = 0$).

Développant les fonctions Γ et remarquant que $\sqrt{\pi} = \Gamma(\frac{1}{2})$,

on obtient :

$$V \bar{X}_T = \frac{\sigma^2}{T^2} 2^{2\delta-1} \frac{\frac{1}{2} \frac{3}{2} \cdots \frac{2\delta-3}{2}}{(\delta-1)(\delta-2)\dots\dots 1}$$

$$V \bar{X}_T = \frac{2\sigma^2}{T^2} \frac{(2\delta-2)!}{[(\delta-1)!]^2}.$$

I.D.2. - Convergence en moyenne quadratique.

On peut alors directement étudier la vitesse de convergence en moyenne quadratique de \bar{X}_T vers zéro. Il suffit pour cela de chercher des équivalents de $V \bar{X}_T$, lorsque T tend vers l'infini. Disposant de deux formes distinctes de cette variance pour $d \neq -\frac{1}{2}$ et $d = -\frac{1}{2}$, il nous faut considérer séparément ces deux cas.

i) Cas $d \neq -\frac{1}{2}$.

Utilisant l'équivalence : $\frac{\Gamma(-\delta+1+T)}{\Gamma(\delta+T)} \sim T^{-2\delta+1}$ au voisinage de l'infini, nous avons :

$$V \bar{X}_T \sim \frac{\sigma^2}{T^2} \frac{2^{2\delta-1}}{\sqrt{\pi}} \frac{\Gamma(\delta - \frac{1}{2})}{\Gamma(\delta)} \left\{ 1 - \frac{\Gamma(\delta)}{\Gamma(-\delta+1)} T^{-2\delta+1} \right\} .$$

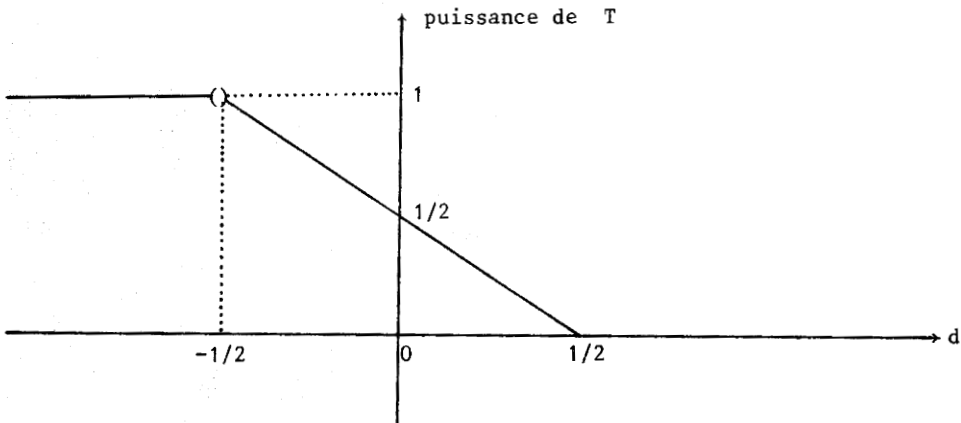
Nous en déduisons immédiatement que la variance est asymptotiquement du même ordre que $\frac{1}{T^{\text{Min}(2, 2\delta+1)}}$. De façon plus précise on a la propriété suivante :

Propriété 1.10.-

i) Si $d < -\frac{1}{2}$, $\lim_{T \rightarrow \infty} (T^{2d+1} V \bar{X}_T) = \frac{\sigma^2}{\sqrt{\pi}} \frac{2^{2\delta-1}}{\Gamma(\delta)} \frac{\Gamma(\delta - \frac{1}{2})}{\Gamma(\delta)}$

ii) Si $-\frac{1}{2} < d < \frac{1}{2}$, $\lim_{T \rightarrow \infty} (T^{-2d+1} V \bar{X}_T) = -\frac{\sigma^2}{\sqrt{\pi}} \frac{2^{2\delta-1}}{\Gamma(-\delta+1)} \frac{\Gamma(\delta - \frac{1}{2})}{\Gamma(\delta)}$

L'évolution de la vitesse avec la valeur de d est la suivante :



On retrouve la vitesse usuelle (puissance $\frac{1}{2}$) pour le bruit blanc ($d = 0$).
On observe également le résultat attendu : vitesse plus lente pour $d > 0$
et plus rapide pour $d < 0$. En fait cette vitesse diminue, lorsque d
augmente.

Moins intuitive est la borne 1 obtenue pour cette vitesse, lorsque
 $d < -\frac{1}{2}$. Elle peut cependant facilement être comprise, si on considère les
deux processus MA purs définis par :

$$X_t^1 = \varepsilon_t - \varepsilon_{t-1} = (1-L) \varepsilon_t \quad (d = -1) ,$$

$$X_t^2 = \varepsilon_t - 2 \varepsilon_{t-1} + \varepsilon_{t-2} = (1-L)^2 \varepsilon_t \quad (d = -2) .$$

Un calcul direct donne en effet :

$$\sum_{t=1}^T X_t^1 = \varepsilon_T - \varepsilon_0 , \quad \sum_{t=1}^T X_t^2 = \varepsilon_T - \varepsilon_{T-1} - \varepsilon_0 + \varepsilon_{-1} .$$

On voit alors que :

$$V(T \bar{X}_T^1) = T^2 V(\bar{X}_T^1) = 2 \sigma^2 , \quad \sqrt{T} ,$$

$$V(T \bar{X}_T^2) = T^2 V(\bar{X}_T^2) = 4 \sigma^2 , \quad \sqrt{T} .$$

Les deux vitesses sont bien identiques, correspondant à la puissance 1 .

ii) Cas $d = -\frac{1}{2}$.

Propriété 1.11.- Si $d = -\frac{1}{2}$, on a :

$$\lim_{T \rightarrow \infty} \left[\frac{T^2}{\text{Log } T} V \bar{X}_T \right] = \frac{2 \sigma^2}{\pi} .$$

Preuve : voir appendice 2.

L'équivalent de $V \bar{X}_T$ n'est donc pas une puissance de T ce qui explique les difficultés que nous avons rencontrées.

I.D.3. - Approche spectrale.

La densité spectrale du processus X est proportionnelle à $f(\omega) \propto [\sin \frac{\omega}{2}]^{-2d}$. Au voisinage de $\omega = 0$, elle est donc équivalente à $f(\omega) \propto \omega^{-2d}$. La puissance apparaissant dans ce développement limité de la densité spectrale au voisinage de zéro est manifestement très liée à celle donnant la vitesse de convergence de \bar{X}_T , au moins pour $d \neq -\frac{1}{2}$. On peut donc se demander, s'il ne serait pas possible de montrer les résultats précédents en utilisant une approche spectrale. Pour cela nous pouvons utiliser certaines propriétés des densités spectrales satisfaisant des conditions de Lipschitz.

Définition 1.12.-

i) Soit f une fonction continue réelle définie sur un intervalle I fermé de \mathbb{R} . Le module de continuité de f est :

$$\omega(\delta; f) = \sup \{ |f(\omega_2) - f(\omega_1)|, \omega_1, \omega_2 \in I, |\omega_1 - \omega_2| < \delta \}.$$

ii) La fonction f vérifie une condition de Lipschitz d'ordre β , $\beta > 0$ si et seulement si il existe un réel C indépendant de δ tel que :

$$\omega(\delta, f) \leq C \delta^\beta.$$

On peut alors montrer le résultat suivant relatif au noyau de

Fejer :
$$L_T(\omega) = \frac{1}{T} \left[\frac{\sin \frac{\omega T}{2}}{\sin \frac{\omega}{2}} \right]^2 .$$

Propriété 1.13. - Introduisons la fonction :

$$\sigma_T(\omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) L_T(\omega-\theta) d\theta .$$

Alors si f vérifie une condition de Lipschitz d'ordre β , $\beta > 0$, on a le résultat suivant sur la convergence de $\sigma_T(\omega)$ vers $f(\omega)$.

i) Si $0 < \beta < 1$:

$$\sup_{\omega \in I} |f(\omega) - \sigma_T(\omega)| = O(T^{-\beta}) ;$$

ii) Si $\beta = 1$:

$$\sup_{\omega \in I} |f(\omega) - \sigma_T(\omega)| = O\left(\frac{\text{Log } T}{T}\right) .$$

Preuve : voir Hannan [1970] p. 506-513, Zygmund [1959].

Cette propriété s'applique directement au cas des processus MADE avec

$\alpha = 1$. Comme $f(\omega) \propto (\sin \frac{\omega}{2})^{-2d}$, on a :

$$\begin{aligned} |f(\omega_1) - f(\omega_2)| &= K \left| \left(\sin \frac{\omega_1}{2}\right)^{-2d} - \left(\sin \frac{\omega_2}{2}\right)^{-2d} \right| \\ &= K \left| \frac{\omega_1}{2} - \frac{\omega_2}{2} \right| |2d| \left| \cos \frac{\tilde{\omega}}{2} \right| \left| \sin \frac{\tilde{\omega}}{2} \right|^{-2d-1} \end{aligned}$$

où $\tilde{\omega}$ est un point intermédiaire de $[\omega_1, \omega_2]$.

On en déduit que pour $d \leq -\frac{1}{2}$: $|f(\omega_1) - f(\omega_2)| \leq C |\omega_1 - \omega_2|$;
c'est-à-dire f vérifie une condition de Lipschitz d'ordre 1.

D'autre part les autocovariances, coefficients du développement en série de Fourier de $f(\omega) = \sum_h \gamma(h) \cos(\omega h)$, sont telles que

$$\gamma(h) = O\left(\frac{1}{h^{1-2d}}\right) ; \text{ lorsque } 0 < -2d < 1 \text{ on a } f \in \text{Lip}(-2d)$$

(Bary [1964] p. 217).

$$\text{Finalement pour } \omega = 0 \text{ et } d < 0 : f(0) = 0 \text{ et } \sigma_T(0) = \frac{T}{2\pi} V \bar{X}_T .$$

Appliquant la propriété 1.13 on déduit le corollaire suivant :

Corollaire 1.14.-

i) Si $-\frac{1}{2} < d < 0$:

$$V \bar{X}_T = O(T^{2d-1})$$

ii) Si $d \leq -\frac{1}{2}$:

$$V \bar{X}_T = O\left(\frac{\text{Log } T}{T^2}\right)$$

Ces résultats sont évidemment compatibles avec ceux obtenus en 1.10 et 1.11

Ils sont cependant moins précis, puisque :

i) On sait seulement pour $-\frac{1}{2} < d < 0$ que $V \bar{X}_T$ est d'ordre inférieur ou égal à T^{2d-1} .

ii) Les constantes $\lim_{T \rightarrow \infty} (T^{-2d+1} V \bar{X}_T)$ ne sont pas données.

iii) La propriété ne s'applique pas pour $0 < d < \frac{1}{2}$.

En revanche, il est clair que la propriété de Lipschitz est aussi satisfaite pour un processus du type :

$X_t = (1-L)^{-d} u_t$ où $d < 0$ et où le processus u admet une représentation ARMA avec des polynômes autorégressif et moyenne mobile ayant leurs racines à l'extérieur du disque unité. Le corollaire 1.14 s'étend donc directement à ce cas.

I.D.4. - Convergence en loi.

i) Non invariance fonctionnelle

Concernant la convergence en loi, nous pouvons d'entrée remarquer qu'il ne sera pas possible de montrer une telle convergence avec une loi limite indépendante de la loi commune des erreurs ε_t .

En effet prenant par exemple le cas $d = -1$ (mais la même remarque est valable pour toutes les valeurs de d entières négatives ou nulles), nous avons :

$$T \bar{X}_T^1 = \sum_{t=1}^T X_t^1 = \varepsilon_T - \varepsilon_0 ,$$

dont la loi est la convolée de la loi de ε et de sa symétrisée, en supposant les $(\varepsilon_t, t \in \mathbb{Z})$ indépendants, de même loi. Elle dépend donc de façon évidente de la loi de ε . Ainsi bien qu'il y ait convergence en loi de $T \bar{X}_T^1$, la limite ne possède pas les propriétés d'invariance habituellement souhaitées.

On peut cependant obtenir certains résultats de convergence faible dans le cas où le processus $(X_t, t \in \mathbb{Z})$ est gaussien. Ces résultats sont dus à Taqqu [1975] et sont fondés sur la notion de mouvement brownien fractionnaire, notion qui servira également dans l'étude du cas non stationnaire.

ii) Mouvement brownien

Un mouvement brownien indexé par $t \in \mathbb{R}^+$ est un processus $B(t;.)$, $t \geq 0$ défini par les trois conditions :

- (*) $B(0;.) = 0$ (p.s.) ,
- (**) le processus $B(t;.)$ est à accroissements indépendants, c'est-à-dire pour toute famille $0 \leq t_1 \leq t_2 \leq \dots \leq t_k$ les variables $B(t_1;.)$, $B(t_2;.) - B(t_1;.)$, ..., $B(t_k;.) - B(t_{k-1};.)$ sont indépendantes ,
- (***) les accroissements suivent des lois normales $\forall t \geq s \geq 0$, $B(t;.) - B(s;.) \rightsquigarrow N(0, t-s)$.

On peut prolonger un tel mouvement brownien à un ensemble d'indices composé des réels positifs et négatifs. Il suffit pour cela de construire un mouvement brownien $\tilde{B}(t;.)$, $t \geq 0$ indépendant du mouvement $B(t;.)$, puis de poser :

$$B(t;.) = \begin{cases} B(t;.) & , \quad \text{si } t \geq 0 \quad , \\ \tilde{B}(-t;.) & , \quad \text{si } t \leq 0 \quad . \end{cases}$$

On vérifie facilement que le processus ainsi défini par prolongement possède bien encore les propriétés d'accroissements indépendants et de normalité des accroissements. Il est tel que :

$$B(t;.) \rightsquigarrow N(0, |t|) \quad , \quad \forall t \in \mathbb{R} \quad .$$

Les autocovariances sont données par :

$$\begin{aligned} \text{Cov} [\bar{B}(t;.), B(s;.)] &= \begin{cases} 0 & , \text{ si } ts \leq 0 , \\ \text{Min}(|t|, |s|) & , \text{ si } ts \geq 0 . \end{cases} \\ &= \frac{1}{2} (|s| + |t| - |s-t|) . \end{aligned}$$

Notons finalement qu'il satisfait la propriété de changement d'échelles :

$$B(at;.) \stackrel{\Delta}{=} \sqrt{a} B(t;.) , \quad a > 0 ,$$

où $\stackrel{\Delta}{=}$ désigne l'égalité des distributions.

iii) Mouvement brownien fractionnaire

La propriété relative au changement d'échelle peut être généralisée et on peut notamment rechercher des processus $(Z(t;\omega) , t \in \mathbb{R})$ satisfaisant la condition plus large :

$$(1.15) \quad Z(at;\omega) \stackrel{\Delta}{=} a^H Z(t;\omega) , \quad a \geq 0 ,$$

où H est un paramètre positif. Le processus est alors dit auto-semblable ("self similar").

Mandelbrot et Van Ness [1968] ont utilisé cette condition pour construire un tel processus à partir du mouvement brownien.

Définition 1.16.- Soit H un réel compris entre 0 et 1.

On appelle mouvement brownien fractionnaire de paramètre H , un processus $(B_H(t;\omega) , t \in \mathbb{R})$ défini par :

$$B_H(t; \omega) = \frac{1}{\Gamma(H + \frac{1}{2})} \left\{ \int_{-\infty}^0 [(t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}] dB(s; \omega) + \int_0^t (t-s)^{H-\frac{1}{2}} dB(s; \omega) \right\}, \quad \text{si } t \geq 0$$

$$B_H(t; \omega) = + \frac{1}{\Gamma(H + \frac{1}{2})} \left\{ \int_{-\infty}^t [(t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}] dB(s; \omega) - \int_t^0 (-s)^{H-\frac{1}{2}} dB(s; \omega) \right\}, \quad \text{si } t \leq 0.$$

où $(B(t; \omega), t \in \mathbb{R})$ désigne un mouvement brownien et $\int \cdot dB(s; \omega)$ une intégrale stochastique par rapport à B .

Comme pour le mouvement brownien linéaire B , la définition aux indices positifs et négatifs de B_H est obtenue en imposant la condition initiale $B_H(0; \omega) = 0$ et en retournant le temps. En particulier, on a :

$$\forall t \geq 0, \quad B_H(-t; \cdot) \stackrel{\Delta}{=} B_H(t; \cdot).$$

Mandelbrot et Van Ness ont proposé une formulation unique pour ce mouvement brownien en utilisant une différence convergente de deux intégrales divergentes :

$$B_H(t_2, \cdot) - B_H(t_1, \cdot) = \frac{1}{\Gamma(H + \frac{1}{2})} \left\{ \int_{-\infty}^{t_2} (t_2-s)^{H-\frac{1}{2}} dB(s; \omega) - \int_{-\infty}^{t_1} (t_1-s)^{H-\frac{1}{2}} dB(s; \omega) \right\}.$$

Utilisant la formule d'intégration par parties :

$$\int_a^b f(s) dB(s; \omega) = [B(s; \omega) f(s)]_a^b - \int_a^b B(s; \omega) df(s),$$

on peut donner une autre expression du mouvement brownien fractionnaire ; nous l'explicitons pour $t \geq 0$. On a :

$$(1.17) \quad B_H(t; \omega) = - \frac{1}{\Gamma(H + \frac{1}{2})} \left\{ \int_{-\infty}^0 B(s; \omega) d[(t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}] \right. \\ \left. + \int_0^t B(s; \omega) d(t-s)^{H-\frac{1}{2}} \right\} .$$

A partir de cette forme, on vérifie facilement l'auto-similitude du processus B_H [voir appendice 3] :

$$B_H(at; \cdot) \stackrel{\Delta}{=} a^H B_H(t; \cdot) \quad , \quad a \geq 0 .$$

La définition de B_H à partir d'intégrales stochastiques permet immédiatement de connaître la forme de la loi du processus B_H . Cette loi est gaussienne centrée. D'autre part les accroissements sont stationnaires et cette propriété utilisée de manière jointe avec l'auto-similitude fournit la forme des moments d'ordre deux. On a, par exemple pour $t_2 \geq t_1 \geq 0$:

$$\begin{aligned} & \text{Cov} [B_H(t_1; \cdot) , B_H(t_2; \cdot)] \\ &= E [B_H(t_1; \cdot) B_H(t_2; \cdot)] \\ &= \frac{1}{2} \{ E[B_H^2(t_1; \cdot)] + E[B_H^2(t_2; \cdot)] - E[B_H(t_2; \cdot) - B_H(t_1; \cdot)]^2 \} \\ &= \frac{1}{2} \{ E[t_1^H B_H(1; \cdot)]^2 + E[t_2^H B_H(1; \cdot)]^2 \\ & \quad - E [B_H(t_2 - t_1 + t_1; \cdot) - B_H(t_1; \cdot)]^2 \} . \end{aligned}$$

Comme $B_H(\tau + t_1; \cdot) - B_H(t_1; \cdot) \stackrel{\Delta}{=} B_H(\tau; \cdot)$, on en déduit :

$$\begin{aligned} & \text{Cov} [B_H(t_1;.), B_H(t_2;.)] \\ &= \frac{1}{2} \{ t_1^{2H} + t_2^{2H} - (t_2 - t_1)^{2H} \} v_{B_H}(1,.) \end{aligned}$$

Plus généralement, on a :

$$(1.18) \quad \begin{aligned} & \text{Cov} [B_H(t_1;.), B_H(t_2;.)] \\ &= \frac{v_{B_H}(1,.)}{2} \{ |t_1|^{2H} + |t_2|^{2H} - |t_1 - t_2|^{2H} \} . \end{aligned}$$

Les corrélations sérielles sont donc égales à :

$$\text{Corr} [B_H(t_1;.), B_H(t_2;.)] = \frac{1}{2} \frac{|t_1|^{2H} + |t_2|^{2H} - |t_1 - t_2|^{2H}}{|t_1|^H |t_2|^H} ,$$

et les covariances ont une expression développée immédiatement déduite de (1.18) en remarquant que :

$$\begin{aligned} v_{B_H}(1,.) &= \frac{1}{\Gamma(H + \frac{1}{2})^2} \left\{ \int_{-\infty}^0 \left[(1-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}} \right]^2 ds \right. \\ &\quad \left. + \int_0^1 (1-s)^{2H-1} ds \right\} \\ &= \frac{1}{\Gamma(H + \frac{1}{2})^2} \left\{ \int_{-\infty}^0 \left[(1-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}} \right]^2 ds + \frac{1}{2H} \right\} . \end{aligned}$$

iv) Convergence faible

Taqqu [1975] établit le résultat de convergence faible ci-dessous applicable à des processus stationnaires dont les autocovariances ont des équivalents polynomiaux. La définition de tels équivalents utilise la notion de fonction à variation lente.

Définition 1.19.- Une fonction L de \mathbb{R} à valeurs dans \mathbb{R}^+ est dite à variation lente à l'infini si :

$$\forall t > 0 \quad \lim_{x \rightarrow \infty} \frac{L(xt)}{L(x)} = 1 .$$

La propriété de base est la suivante :

Théorème 1.20.- Soit $X = (X_t, t \in \mathbb{Z})$ un processus stationnaire gaussien centré. Supposons que ses autocovariances soient telles que :

$$\sum_{i=1}^T \sum_{j=1}^T \gamma(i-j) \sim K T^{2H} L(T) \quad ,$$

lorsque T tend vers l'infini, où $0 < H < 1$, et $K > 0$ et L est à variation lente à l'infini, alors :

$$Z_T(r) = \frac{1}{d_T} \sum_{i=1}^{[Tr]} X_i \xrightarrow{d} \sqrt{\frac{K}{V B_H(1,.)}} B_H(r,.) \quad ,$$

où d_T est tel que $d_T^2 \sim T^{2H} L(T)$ et $r \in [0,1]$.

Preuve : voir Taqqu (1975).

Ce résultat est évidemment beaucoup plus puissant que celui dont nous avons réellement besoin dans l'étude des processus MADE. En particulier il donne un résultat de convergence faible pour le processus $(Z_T(r), r \in [0,1])$. Ces processus sont à valeurs dans l'espace $D[0,1]$ des fonctions définies sur $[0,1]$, continues à droite et ayant des limites à gauche finies ; cet espace est muni habituellement de la topologie de Skorokhod [voir Billingsley [1968] (p. 109 et suivantes)].

Le théorème 1.20 s'applique directement aux processus MADE(1) correspondant à $\alpha = 1$ et $-\frac{1}{2} < d < \frac{1}{2}$. En effet, on a vu dans la propriété 1.10 que :

$$\begin{aligned} \lim_{T \rightarrow \infty} (T^{-2d+1} \vee \bar{X}_T) &= \lim_{T \rightarrow \infty} \left[T^{-2d-1} \sum_{i=1}^T \sum_{j=1}^T \gamma(i-j) \right] \\ &= - \frac{\sigma^2 2^{2\delta-1}}{\sqrt{\pi}} \frac{\Gamma(\delta - \frac{1}{2})}{\Gamma(-\delta + 1)} . \end{aligned}$$

On a dans ce cas : $2H = 2d+1$ et la condition $0 < H < 1$ équivaut à $-\frac{1}{2} < d < \frac{1}{2}$; d'autre part la constante K est égale à

$$K = - \frac{\sigma^2 2^{2\delta-1}}{\sqrt{\pi}} \frac{\Gamma(\delta - \frac{1}{2})}{\Gamma(-\delta + 1)} .$$

Appliquant le théorème 1.20 à ce processus et pour la valeur $r = 1$, on en déduit le corollaire suivant.

Corollaire 1.21. - Soit $X = (X_t, t \in \mathbb{Z})$ un processus gaussien stationnaire satisfaisant :

$$(1-L)^d X_t = \varepsilon_t \quad , \quad -\frac{1}{2} < d < \frac{1}{2} \quad ,$$

alors $\frac{1}{T^{d - \frac{1}{2}}} \bar{X}_T$ converge en loi vers une variable normale

$$N \left[0 , - \frac{\sigma^2 2^{2\delta-1}}{\sqrt{\pi}} \frac{\Gamma(\delta - \frac{1}{2})}{\Gamma(-\delta + 1)} \right]$$

Le corollaire aurait évidemment pu être obtenu directement en remarquant que $\frac{1}{T^{d-1/2}} \bar{X}_T$ suit une loi normale et qu'il y a convergence de la moyenne et de la variance.

II - PROPRIETES ASYMPTOTIQUES DE LA MOYENNE EMPIRIQUE

[CAS NON STATIONNAIRE]

II.A. - DEFINITION DE PROCESSUS FRACTIONNAIRES

NON STATIONNAIRES

Jusqu'à maintenant nous avons essentiellement étudié des processus fractionnaires satisfaisant la condition de stationnarité. Si par exemple, nous considérons la relation $(1-\alpha L)^d X_t = \varepsilon_t$, où ε est un bruit blanc, nous savons qu'il existe une solution stationnaire, lorsque les paramètres α , d satisfont certaines contraintes, essentiellement si $|\alpha| < 1$ ou si $\alpha = \pm 1$ et $d < \frac{1}{2}$. Dans les autres cas une telle équation ne peut admettre que des solutions non stationnaires. On sait alors que, dans le cas usuel des processus ARIMA, il faut pour définir une solution non stationnaire fixer de façon adéquate une ou des conditions initiales et que selon le choix effectué le processus peut dans certains cas satisfaire l'équation de définition depuis la date initiale ou bien seulement la satisfaire asymptotiquement. Il existe cependant de nombreux problèmes en particulier ceux liés à la prévision pour lesquels le choix de l'une ou l'autre des conditions initiales a peu d'influence. Par analogie avec ce cas classique, nous allons introduire des processus fractionnaires non stationnaires en choisissant une convention maniable sur les conditions initiales. Nous nous intéresserons essentiellement à une non stationnarité provenant de la présence d'une racine égale à 1.

Définition 2.1.- Un processus $X = (X_t, t \in \mathbb{N})$ est dit non stationnaire fractionnaire intégré, s'il satisfait une relation de type :

$$X_t = (1-L)^{-d} u_t, \quad t \in \mathbb{N}$$

où $d > \frac{1}{2}$

et où $u_t = A(L) \tilde{\varepsilon}_t$ avec $A(z) = \sum_{j=0}^{+\infty} \alpha_j z^j$ et

$$\tilde{\varepsilon}_t = 0 \quad \text{si } t \leq 0, \quad \tilde{\varepsilon}_t = \varepsilon_t \quad \text{si } t > 0, \quad \varepsilon = (\varepsilon_t, t \in \mathbb{Z})$$

étant un bruit blanc de variance σ^2 .

Les conditions initiales sont donc introduites dans cette définition en mettant à zéro les valeurs de u_t antérieures à la date initiale. Une autre convention naturelle aurait consisté à mettre à zéro les valeurs antérieures du processus lui-même et à imposer $(1-L)^d X_t = A(L) \varepsilon_t, t \geq 0$.

Un avantage de la première est évidemment sa simplicité de maniement puisque les valeurs du processus X sont données sous forme explicite.

Remarquons également que la généralisation du paragraphe I aurait consisté à prendre $A(L) = 1$.

II.B. - LE THEOREME D'INVARIANCE FONCTIONNELLE

Théorème d'invariance fonctionnelle 2.2. - Soit $(u_\tau, \tau \geq 1)$

un processus et $(S_t = \sum_{\tau=1}^t u_\tau, t \geq 1)$ le processus cumulé correspondant :

- a) Si $E u_\tau = 0, \forall \tau$;
- b) Si $\sup_{\tau} E |u_\tau|^{\beta+\epsilon} < \infty$ pour $\beta > 2$ et $\epsilon > 0$;
- c) Si $\eta^2 = \lim_{T \rightarrow \infty} E \left[\frac{1}{T} S_T^2 \right]$ existe et est strictement positif ;
- d) Si $(u_\tau, \tau \geq 1)$ est un processus fortement mélangeant de coefficients de mélangeance α_k satisfaisant $\sum_{k=1}^{\infty} \alpha_k^{1-2/\beta} < \infty$,

alors le processus $(\overset{\vee}{S}_T(r), r \in [0,1])$ défini par :

$$\overset{\vee}{S}_T(r) = \frac{1}{\eta\sqrt{T}} S_{[Tr]} \quad \text{où } [\] \text{ désigne la partie entière converge}$$

en loi vers un mouvement brownien $(B(r), r \in [0,1])$.

Preuve : voir Herrndorf [1984] corollaire 1. ■

Comme dans la section précédente les divers processus sont à valeurs dans l'espace $D[0,1]$, muni de la topologie de Skorokhod.

Les coefficients de mélangeance apparaissant dans le théorème précédent sont définis de la façon suivante : introduisons les σ -algèbres $B_m = \sigma(u_\tau, 1 \leq \tau \leq m)$ et $B_{m+k}^n = \sigma(u_\tau, m+k \leq \tau \leq n)$ engendrées respectivement par les composantes d'indices inférieurs à m et par celles d'indices compris entre $m+k$ et n . On commence par définir un coefficient de mélangeance relatif à de tels couples de σ -algèbres :

$$\alpha_k(n) = \begin{cases} \sup \{ |P(A \cap B) - P(A)P(B)| & : A \in \mathcal{B}_m, B \in \mathcal{B}_{m+k}^n, \\ & 1 \leq m \leq n-k \} & , \quad k \leq n-1, \\ 0 & , \quad \text{si } k \geq n. \end{cases}$$

Les coefficients de mélangeance du théorème sont alors obtenus par

$$\alpha_k = \sup_n \alpha_k(n) \quad [\text{voir Rosenblatt (1956)}].$$

Nous arrêterons ici la discussion technique de ce résultat. L'important pour ce qui suit est de remarquer qu'il s'applique notamment aux processus u satisfaisant une représentation ARMA stationnaire :

$$\phi(L) u_t = \psi(L) \varepsilon_t,$$

où ε est un bruit blanc indépendant dont les composantes admettent une loi commune avec des moments d'ordre strictement supérieur à deux. Dans ce cas, on a :

$$\eta^2 = \lim_{T \rightarrow \infty} E \left[\frac{1}{T} S_T^2 \right] = 2\pi f_u(0) = \sigma^2 \frac{|\psi(1)|^2}{\phi(1)^2},$$

en désignant par σ^2 la variance du bruit ε .

On voit qu'il s'agit effectivement d'un théorème d'invariance fonctionnelle au sens où le résultat est indépendant de la forme de la loi commune des ε_t et ne dépend en fait de cette loi que par l'intermédiaire de la variance σ^2 .

Ce résultat peut être vu comme une généralisation naturelle du théorème de Donsker [voir Billingsley (1968)], qui correspond au cas particulier $u_t = \varepsilon_t$.

II.C. - UNE CONSEQUENCE DU THEOREME D'INVARIANCE

FONCTIONNELLE

Nous allons maintenant utiliser le théorème d'Herrndorf pour en déduire le comportement limite de certaines sommes pondérées des composantes du processus u .

Etant donnée une fonction réelle λ , on introduit le processus non stationnaire $(Y_t, t \geq 1)$ défini par :

$$(2.3) \quad Y_t = \sum_{\tau=1}^t \lambda(t-\tau) u_{\tau},$$

où $(u_{\tau}, \tau \geq 1)$ satisfait les hypothèses du théorème 2.2.

Ce processus peut donc être considéré comme admettant une écriture moyenne mobile infinie : $Y_t = \sum_{\tau=-\infty}^t \lambda(t-\tau) \hat{u}_{\tau}$ fondé sur le processus \hat{u} déduit du processus u en mettant à zéro les composantes d'indices négatifs.

Nous allons nous intéresser au comportement asymptotique du processus :

$$(2.4) \quad \tilde{Y}_T(r) = \frac{Y_{[Tr]}}{\eta T^{\mu + \frac{1}{2}}},$$

en supposant que la fonction λ est asymptotiquement telle que

$$(2.5) \quad \lambda(t) \sim t^{\mu}$$

lorsque t tend vers l'infini avec : $\mu > -\frac{1}{2}$.

Nous donnons ci-dessous une présentation heuristique du résultat renvoyant à l'appendice 4 pour la preuve complète, c'est-à-dire essentiellement pour les justifications des interversions des limites et des intégrales.

Nous avons :

$$\begin{aligned} \tilde{Y}_T(r) &= \frac{Y_{[Tr]}}{\eta T^{\mu + \frac{1}{2}}} = \frac{1}{\eta T^{\mu + \frac{1}{2}}} \sum_{\tau=1}^{[Tr]} \lambda([Tr] - \tau) u_\tau \\ &= \sum_{\tau=1}^{[Tr]} \frac{1}{T^\mu} \lambda([Tr] - \tau) \left(\frac{1}{\eta\sqrt{T}} u_\tau \right) \\ &= \sum_{\tau=1}^{[Tr]} \frac{1}{T^\mu} \lambda([Tr] - \tau) \int_{\tau-1/T}^{\tau/T} d \tilde{S}_T(s) \\ &= \int_0^1 \left(\sum_{\tau=1}^{[Tr]} \frac{\lambda([Tr] - \tau)}{T^\mu} \mathbb{1}_{\frac{\tau-1}{T} \leq s < \frac{\tau}{T}} \right) d \tilde{S}_T(s) . \end{aligned}$$

Lorsque T tend vers l'infini, nous savons d'après le théorème 2.2. que le processus \tilde{S}_T converge en loi vers un mouvement brownien B ; d'autre part la fonction :

$$f_T(s) = \sum_{\tau=1}^{[Tr]} \frac{\lambda([Tr] - \tau)}{T^\mu} \mathbb{1}_{\frac{\tau-1}{T} \leq s < \frac{\tau}{T}} ,$$

converge vers :

$$f_\infty(s) = (r-s)^\mu \mathbb{1}_{0 \leq s < r} .$$

Passant à la limite sous l'intégrale [voir Appendice 4], on en déduit la convergence en loi du processus \tilde{Y}_T vers le processus \tilde{Y}_∞ , tel que

$$\tilde{Y}_\infty(r) = \int_0^r (r-s)^\mu dB(s) .$$

Propriété 2.6.- Considérons un processus défini par :

$$Y_t = \sum_{\tau=1}^t \lambda(t-\tau) u_\tau ,$$

où le processus u satisfait les conditions du théorème 2.2. et où la fonction λ est asymptotiquement équivalente à t^μ avec $\mu > -\frac{1}{2}$.

Le processus défini par : $\tilde{Y}_T(r) = \frac{Y[T_r]}{n T^{\mu + \frac{1}{2}}}$
 converge en loi vers : $Y_\infty(r) = \int_0^r (r-s)^\mu dB(s) .$

Remarquons que la condition $\mu > -\frac{1}{2}$ est utile à deux niveaux différents : d'une part le processus étant non stationnaire explosif il est nécessaire de le raccourcir pour obtenir la convergence, ce qui demande une puissance $\mu + \frac{1}{2} > 0$ pour T ; d'autre part il faut que l'intégrale stochastique définissant \tilde{Y}_∞ existe.

Or :
$$V\left(\int_0^r (r-s)^\mu dB(s)\right) = \int_0^r (r-s)^{2\mu} ds = \int_0^r u^{2\mu} du$$

existe lorsque $\mu > -\frac{1}{2}$.

II.D. - APPLICATION AUX PROCESSUS FRACTIONNAIRES INTEGRES

Revenons maintenant aux processus fractionnaires intégrés. D'un point de vue pratique, nous nous restreignons aux processus :

$$X_t = (1-L)^{-d} \tilde{u}_t ,$$

où le processus u associé admet une représentation ARMA stationnaire fondée sur un bruit blanc indépendant admettant un moment d'ordre strictement supérieur à 2 :

$$\phi(L) u_t = \theta(L) \varepsilon_t , \quad V \varepsilon_t = \sigma^2 .$$

La fonction $(1-L)^{-d}$ de l'opérateur retard admet un développement en série entière, de sorte que le processus X s'écrit sous la forme :

$$X_t = \sum_{\tau=1}^t \lambda(t-\tau) u_\tau = \sum_{\tau=1}^t \frac{\Gamma(t-\tau+d)}{\Gamma(d) (t-\tau)!} u_\tau .$$

Le terme général de ce développement est tel que : $\lambda(t) \sim t^{d-1} / \Gamma(d)$.

Nous sommes donc dans les conditions d'application de la propriété 2.6 en prenant $\mu = d-1 > -\frac{1}{2}$, c'est-à-dire $d > \frac{1}{2}$.

Corollaire 2.7.- Un processus non stationnaire fractionnaire intégré :

$$X_t = (1-L)^{-d} \tilde{u}_t , \quad \text{avec } d > \frac{1}{2} \quad \text{et}$$

$$\phi(L) u_t = \theta(L) \varepsilon_t , \quad \varepsilon \text{ bruit blanc indépendant est tel que :}$$

$$\tilde{X}_T(r) = \frac{X[Tr]}{\eta T^{d-\frac{1}{2}}} \quad \text{converge en loi vers}$$

$$\tilde{X}_\infty(r) = \frac{1}{\Gamma(d)} \int_0^r (r-s)^{d-1} dB(s), \quad r \in [0,1].$$

Ce corollaire fournit donc une idée du "terme dominant" de la série X initiale. L'explosion du processus non stationnaire a lieu à la vitesse $T^{d-\frac{1}{2}}$; d'autre part le coefficient de cette puissance $T^{d-\frac{1}{2}}$ peut être considérée comme asymptotiquement normal. En particulier prenant $r = 1$, on voit que :

$$\frac{1}{T^{d-\frac{1}{2}}} X_T \xrightarrow{d} \eta \tilde{X}_\infty(1),$$

où :

$$\eta \tilde{X}_\infty(1) \rightsquigarrow N \left[0 ; \frac{\eta^2}{\Gamma^2(d)} \int_0^1 (1-s)^{2d-2} ds \right],$$

$$\eta \tilde{X}_\infty(1) \rightsquigarrow N \left[0 ; \frac{\sigma^2}{\Gamma^2(d)} \frac{\Theta(1)^2}{\Phi(1)^2} \frac{1}{2d-1} \right].$$

On remarquera finalement que le processus limite $\tilde{X}_\infty(r)$ est directement déduit d'un mouvement brownien fractionnaire en mettant par convention à zéro les composantes de B associées aux indices négatifs.

II.E. - CONVERGENCE DE LA MOYENNE EMPIRIQUE

Propriété 2.8. - Sous les hypothèses du corollaire 2.7 ,

$$\frac{1}{T^{d+\frac{1}{2}}} \sum_{t=1}^T X_t = \frac{1}{T^{d-\frac{1}{2}}} \bar{X}_T \quad \text{converge en loi vers :}$$

$$\frac{\eta}{\Gamma(d)} \int_0^1 (1-s)^{d-1} B(s) ds \quad .$$

Preuve : Pour obtenir ce résultat, il suffit d'écrire :

$$\begin{aligned} \frac{1}{T^{d+\frac{1}{2}}} \sum_{t=1}^T X_t &= \frac{1}{T} \sum_{t=1}^T \frac{1}{T^{d-\frac{1}{2}}} X_t \\ &= \frac{1}{T} \sum_{t=1}^T \tilde{X}_T\left(\frac{t}{T}\right) \\ &= \sum_{t=1}^T \int_{t-1/T}^{t/T} \tilde{X}_T\left(\frac{t}{T}\right) dr \\ &= \int_0^1 \left[\sum_{t=1}^T \mathbb{1}_{\frac{t-1}{T} \leq r < \frac{t}{T}} \tilde{X}_T\left(\frac{t}{T}\right) \right] dr \quad . \end{aligned}$$

Cette variable converge en loi vers : $\int_0^1 \tilde{X}_\infty(r) dr$.

Remplaçant $\tilde{X}_\infty(r)$ par son expression, nous avons :

$$\begin{aligned} \int_0^1 \tilde{X}_\infty(r) dr &= \frac{\eta}{\Gamma(d)} \int_0^1 \left[\int_0^r (r-s)^{d-1} dB(s) \right] dr \\ &= \frac{\eta}{\Gamma(d)} \int_0^1 \left[\int_s^1 (r-s)^{d-1} dr \right] dB(s) \\ &= \frac{\eta}{\Gamma(d)d} \int_0^1 (1-s)^d dB(s) = \frac{\eta}{\Gamma(d)} \int_0^1 B(s) (1-s)^{d-1} ds \quad . \quad \text{Q.E.D.} \end{aligned}$$

La convergence de la moyenne empirique a donc lieu à la "vitesse" : $T^{-d+\frac{1}{2}}$; plus le degré de non stationnarité est élevé plus la convergence est lente.

Cette convergence en loi a lieu vers une loi normale et ceci sans hypothèse de normalité sur la loi du bruit blanc ϵ . Le cas limite où ce résultat d'invariance fonctionnelle est perdu est $d = \frac{1}{2}$; nous avons en effet vu en I corollaire 1.21. que la normalité asymptotique était satisfaite dans le cas où ϵ est gaussien.

La propriété 2.8. donne aussi une formule pour la variance asymptotique de la moyenne empirique.

Corollaire 2.9. - Pour un processus X non stationnaire fractionnaire intégré, la variance de \bar{X}_T peut être approchée par :

$$V_{as} \bar{X}_T = \frac{\sigma^2 \Theta(1)^2}{\phi(1)^2} \frac{T^{2d-1}}{d^2(2d+1)} \frac{1}{\Gamma^2(d)}$$

Preuve : On a :

$$\begin{aligned} V \left[\int_0^1 (1-s)^{d-1} B(s) ds \right] &= V \left[\frac{\eta}{d} \int_0^1 (1-s)^d dB(s) \right] \\ &= \frac{\eta^2}{d^2} \int_0^1 (1-s)^{2d} ds = \frac{\eta^2}{d^2} \frac{1}{2d+1} \end{aligned}$$

formule dont on déduit immédiatement le résultat.

Q.E.D.

Remarque 2.10.- Cette formule correspond bien à ce que l'on pouvait raisonnablement attendre en examinant le cas particulier d'une marche aléatoire. Dans ce cas, on a :

$$X_t = (1-L)^{-1} \varepsilon_t = \sum_{\tau=1}^t \varepsilon_{\tau} .$$

Donc en sommant, on obtient :

$$\sum_{t=1}^T X_t = \sum_{t=1}^T \sum_{\tau=1}^t \varepsilon_{\tau} = \varepsilon_T + 2 \varepsilon_{T-1} + 3 \varepsilon_{T-2} + \dots + T \varepsilon_1 .$$

On en déduit que :

$$\begin{aligned} V \bar{X}_T &= \frac{1}{T^2} \{ \sigma^2 [1 + 2^2 + \dots + (T-1)^2 + T^2] \} \\ &= \frac{\sigma^2}{T^2} \left[\frac{T(T+1)(2T+1)}{6} \right] \sim \sigma^2 \frac{T}{3} . \end{aligned}$$

Dans ce cas on vérifie bien que : $\eta^2 = \sigma^2$, $2d-1 = 1$ et $d^2(2d+1) = 3$.

III - PROPRIETES ASYMPTOTIQUES DES COVARIANCES EMPIRIQUES
(CAS STATIONNAIRE)

Nous allons maintenant nous intéresser aux moments d'ordre deux, en séparant comme auparavant les cas stationnaire et non stationnaire. Nous nous limiterons à l'étude de la convergence en loi. D'autre part nous présentons les résultats dans le cas de covariances empiriques du type :

$$\tilde{\gamma}(h) = \frac{1}{T-h} \sum_{t=1}^{T-h} X_t X_{t+h} .$$

Des résultats de même type pourraient être déduits pour les covariances empiriques préalablement centrées, c'est-à-dire pour :

$$\tilde{\gamma}(h) = \frac{1}{T-h} \sum_{t=1}^{T-h} (X_t - \bar{X}_T) (X_{t+h} - \bar{X}_T) ;$$

il faudrait alors tenir compte des diverses propriétés de convergence faible de la moyenne empirique discutées dans le paragraphe I.

III.A. - LES RESULTATS CLASSIQUES

Soit $X = (X_t, t \in \mathbb{Z})$ un processus stationnaire admettant la représentation moyenne mobile infinie :

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} , \quad \text{avec} \quad \sum_{j=0}^{\infty} a_j^2 < \infty ,$$

où $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ est un bruit blanc, de variance $\sigma^2 > 0$.

La covariance empirique :

$$(3.1) \quad \tilde{\gamma}_T(h) = \frac{1}{T-h} \sum_{t=1}^{T-h} X_t X_{t+h} , \quad h = 0, 1, \dots, T-1$$

constitue un estimateur sans biais de la covariance théorique du processus $\gamma(h) = E(X_t X_{t+h})$, $h = 0, 1, \dots, T-1$.

Afin d'obtenir des renseignements sur le comportement asymptotique de cet estimateur et en particulier sur sa précision, nous serons amenés à considérer les moments d'ordre 4 du processus X . Si $E \varepsilon_t^4 < \infty$,

on définit le cumulant d'ordre 4 de X_t par :

$$(3.2) \quad K[h, g, \ell] = E(X_t X_{t+h} X_{t+g} X_{t+\ell}) \\ - [\gamma(h) \gamma(g-\ell) + \gamma(g) \gamma(h-\ell) + \gamma(\ell) \gamma(h-g)] .$$

Dans le cas particulier, où X est gaussien, ce cumulants est identiquement nul.

La première propriété ci-dessous concerne la convergence en moyenne quadratique des $\hat{\gamma}_T(h)$ vers $\gamma(h)$ et la seconde la convergence en loi.

Propriété 3.3.- Si $\sum_{\ell=-\infty}^{+\infty} \gamma^2(\ell) < \infty$ et si

$$\left| \sum_{\ell=-\infty}^{+\infty} K[h, -\ell, g-\ell] \right| < \infty, \text{ alors :}$$

$$\lim_{T \rightarrow \infty} T \text{Cov} [\hat{\gamma}_T(h), \hat{\gamma}_T(g)] \\ = \sum_{\ell=-\infty}^{+\infty} [\gamma(\ell) \gamma(\ell+h-g) + \gamma(\ell-g) \gamma(\ell+h) + K(h, -\ell, g-\ell)]$$

Preuve : voir Anderson [1971] p. 465 .

La vitesse de convergence est donc de l'ordre de $T^{1/2}$.

De plus, on peut simplifier l'expression des covariances asymptotiques,

lorsque ε est un bruit blanc indépendant. Notant : $K_4 = E \varepsilon_t^4 - 3 \sigma^4$

la kurtosis de ε_t , on a alors :

$$\sum_{\ell=-\infty}^{+\infty} K[h, -\ell, g-\ell] = K_4 \frac{\gamma(h) \gamma(g)}{\sigma^4} .$$

Remplaçant dans la formule donnant les covariances asymptotiques, on obtient :

$$(3.4) \quad \left\{ \begin{aligned} \Gamma_{hg} &= \lim_{T \rightarrow \infty} T \operatorname{Cov} [\tilde{Y}_T(h), \tilde{Y}_T(g)] \\ &= \sum_{\ell=-\infty}^{+\infty} [\gamma(\ell+h) \gamma(\ell+g) + \gamma(\ell-g) \gamma(\ell+h)] + \frac{K_4}{\sigma^4} \gamma(h) \gamma(g) \\ &= 4\pi \int_{-\pi}^{\pi} \cos(\omega g) \cos(\omega h) f^2(\omega) d\omega + \frac{K_4}{\sigma^4} \gamma(h) \gamma(g) \end{aligned} \right.$$

où f est la densité spectrale de X .

Propriété 3.5.- Convergence en loi

Si $X = (X_t, t \in \mathbb{Z})$ admet une représentation moyenne mobile infinie fondée sur un bruit blanc indépendant, si $E \varepsilon_t^4 < \infty$ et si $\sum_{j=0}^{\infty} |a_j| < \infty$, alors la loi limite de :

$$\sqrt{T} [\tilde{Y}_T(o) - \gamma(o), \dots, \tilde{Y}_T(K) - \gamma(K)]'$$

est la loi normale $N [0, [\Gamma_{h,g}]_{0 \leq h, g \leq K}]$, où les Γ_{hg} sont donnés par la formule (3.4).

Preuve : Voir Anderson (1971) p. 481.

Remarque 3.6.- On voit évidemment de façon immédiate que le même résultat est valable si on remplace les autocovariances empiriques

$$\tilde{Y}_T(h) \quad \text{par} \quad \hat{Y}_T(h) = \frac{1}{T} \sum_{t=1}^{T-h} X_t X_{t+h} .$$

Remarque 3.7.- Le résultat de convergence en loi est dans d'autres ouvrages donné sous des conditions autres que $\sum_{j=0}^{\infty} |a_j| < \infty$.

Ces conditions sont :

$$(*) \quad \sum_{j=-\infty}^{+\infty} |j| |\gamma(j)| < \infty$$

dans Azencott - Dacunha-Castelle [1984],

$$(**) \quad \sum_{j=0}^{\infty} j a_j^2 < \infty \quad \text{et} \quad \sum_{j=-\infty}^{+\infty} \gamma^2(j) < \infty ,$$

dans Rosenblatt [1985].

III.B. - CAS OU LES RESULTATS CLASSIQUES S'APPLIQUENT

Considérons maintenant le cas particulier d'un processus MADE(1) stationnaire défini par :

$$(1-\alpha L)^d X_t = \varepsilon_t \quad , \quad t \in \mathbb{Z} \quad .$$

Il résulte immédiatement des équivalents obtenus pour les autocovariances $\gamma(h)$ et pour les coefficients a_j du développement moyenne mobile que $\gamma(h)$ et a_j décroissent de façon exponentielle si $|\alpha| < 1$, qu'ils s'annulent à partir d'un certain rang si $|\alpha| > 1$ et d est un entier négatif et que lorsque $\alpha = \pm 1$, ils sont tels que :

$$|\gamma(h)| \sim h^{2d-1} \quad , \quad |a_j| \sim j^{d-1} \quad .$$

On en déduit que la condition $\sum_{h=-\infty}^{+\infty} \gamma^2(h) < \infty$ est vérifiée dans tous les cas de stationnarité, sauf si $\alpha = \pm 1$ et $d > \frac{1}{4}$; la condition $\sum_{j=0}^{\infty} |a_j| < +\infty$ l'est, sauf si $\alpha = \pm 1$ et $d > 0$.

Regroupant les propriétés 3.3 et 3.5 en nous plaçant dans le cas d'un bruit blanc indépendant, on obtient :

Propriété 3.8.- Soit X un processus stationnaire satisfaisant $(1-L)^d X_t = \varepsilon_t$, $t \in \mathbb{Z}$, où ε est un bruit blanc indépendant, admettant des moments jusqu'à l'ordre 4 ;

i) Si $|\alpha| < 1$, ou si $|\alpha| > 1$ et d est entier négatif, ou si $\alpha = \pm 1$ et $d < \frac{1}{4}$, alors :

$$\lim_{T \rightarrow \infty} T \text{Cov} [\tilde{Y}_T(h), \tilde{Y}_T(g)] = \Gamma_{hg} .$$

ii) Si $|\alpha| < 1$, ou si $|\alpha| > 1$ et d est entier négatif, ou si $\alpha = \pm 1$ et $d \leq 0$, alors :

$$\sqrt{T} [\tilde{Y}_T(o) - \gamma(o), \dots, \tilde{Y}_T(K) - \gamma(K)] \xrightarrow{d} N[0, [\Gamma_{hg}]] , \quad \forall k .$$

Remarque 3.9.- Sur cet exemple des processus MADE, on vérifie que si $\alpha = \pm 1$ la condition de Azencott - Dacunha-Castelle conduit à $d < -\frac{1}{2}$, alors que celle de Rosenblatt donne $d \leq 0$. Les théorèmes fournis dans Anderson ou dans Rosenblatt apparaissent donc beaucoup plus fin que celui figurant dans Azencott - Dacunha - Castelle.

Remarque 3.10.- Les éléments de la matrice de variance covariance asymptotique peuvent en pratique être facilement calculés. Introduisons en effet le processus Y défini par :

$$(1-\alpha L)^{2d} Y_t = \varepsilon_t .$$

On peut remarquer que si $\alpha = \pm 1$ le degré d étant inférieur à $\frac{1}{4}$ dans la propriété 3.8, Y est bien stationnaire (ceci explique l'apparition de cette valeur particulière $1/4$).

On a :

$$\begin{aligned} \Gamma_{hg} &= 4\pi \int_{-\pi}^{\pi} \cos(\omega g) \cos(\omega h) f_X^2(\omega) d\omega + \frac{K_4}{\sigma^4} \gamma_X(h) \gamma_X(g) \\ &= 4\pi \int_{-\pi}^{\pi} \frac{1}{2} \{ \cos [\omega(h+g)] + \cos [\omega(h-g)] \} f_X^2(\omega) d\omega + \frac{K_4}{\sigma^4} \gamma_X(h) \gamma_X(g) \\ &= \sigma^2 [\gamma_Y(h+g) + \gamma_Y(h-g)] + \frac{K_4}{\sigma^4} \gamma_X(h) \gamma_X(g) . \end{aligned}$$

Dans le cas particulier où $\alpha = \pm 1$, on a : (voir Gonçalves (1987)) :

$$\begin{aligned} \gamma_Y(h+g) + \gamma_Y(h-g) &= \frac{\sigma^2 \alpha^{h+g}}{\Gamma(2d)} \left[\frac{\Gamma(h+g+2d)}{\Gamma(h+g+1)} F[2d ; 2d + h+g ; h+g+1 ; 1] \right. \\ &\quad \left. + \frac{\Gamma(h-g+2d)}{\Gamma(h-g+1)} F[2d ; 2d + h-g ; h-g+1 ; 1] \right] \\ &= \sigma^2 \alpha^{h+g} \frac{\Gamma(1-4d)}{\Gamma(2d) \Gamma(1-2d)} \left[\frac{\Gamma(h+g+2d)}{\Gamma(h+g+1-2d)} + \frac{\Gamma(h-g+2d)}{\Gamma(h-g+1-2d)} \right] \end{aligned}$$

III.C. - ETUDE DU CAS $\alpha = 1$, $0 < d < \frac{1}{2}$.

Il est clair d'après les résultats du paragraphe précédent que les théorèmes classiques ne permettent pas de déterminer la loi asymptotique des autocovariances empiriques si $\alpha = \pm 1$ et $0 < d < \frac{1}{2}$. Les cas $\alpha = 1$ et $\alpha = -1$ étant symétriques l'un de l'autre, nous pouvons nous restreindre dans la présentation à $\alpha = 1$.

Des résultats sur ce cas ont été obtenus par Rosenblatt (1961) et Fox-Taqqu (1985). De façon à alléger la présentation, il a paru préférable de commencer par donner les propriétés de convergence faible qui en résultent pour les processus MADE, puis d'expliquer ensuite comment sont obtenus les résultats généraux.

III.C.1. - Convergence en loi des autocovariances empiriques
dans le cas des processus MADE avec $\alpha = 1$, $0 < d < \frac{1}{2}$.

Propriété 3.11. - Soit X un processus stationnaire défini par
 $(1-L)^d X_t = \varepsilon_t$, où ε est un bruit blanc indépendant, gaussien
et où $0 < d < \frac{1}{4}$, alors :

$$\sqrt{T} [\tilde{\gamma}_T(h) - \gamma(h)] \xrightarrow{d} \left[\sum_{K=-\infty}^{+\infty} [\gamma^2(K) + \gamma(K-h)\gamma(K+h)] \right]^{1/2} B(1, \cdot) .$$

Sous l'hypothèse supplémentaire de normalité du bruit, on conserve donc la vitesse de convergence en $T^{1/2}$ et la normalité asymptotique.

Propriété 3.12.- Soit X un processus stationnaire défini par : $(1-L)^d X_t = \varepsilon_t$, où ε est un bruit blanc indépendant gaussien et où $\frac{1}{4} < d < \frac{1}{2}$, alors

$$\frac{1}{T^{2d-1}} [\tilde{Y}_T(h) - \gamma(h)] \xrightarrow{d} \gamma(o) R_{\frac{1}{2}-d} (1) ,$$

où $R_{\frac{1}{2}-d} (1)$ est une variable dont la loi n'est pas normale.

Il y a donc modification de la vitesse de convergence et perte de la normalité asymptotique. On remarque aussi une différence importante avec le cas où $d < \frac{1}{4}$ puisque la variance de la loi limite est indépendante de l'indice h . La loi limite est liée aux processus dits de Rosenblatt et aux distributions dites exotiques.

III.C.2. - L'approche de Rosenblatt.

Le processus dit de Rosenblatt apparaît pour les premières fois dans deux articles : Rosenblatt [1956] et [1961]. L'auteur cherche à obtenir la loi limite de la variance empirique d'un processus stationnaire, gaussien, centré $X = (X_t, t \in \mathbb{Z})$, dont la suite des autocovariances est : $\gamma(h) = (1+h^2)^{-\gamma}$, avec $0 < \gamma < \frac{1}{4}$. La démarche suivie est tout à fait naturelle. Elle consiste à :

a) Déterminer la fonction caractéristique de $\sum_{t=1}^T X_t^2$,

b) A montrer ensuite qu'il existe une fonction $H(T)$ tendant vers l'infini avec T et telle que la fonction caractéristique de

$$\frac{1}{H(T)} \left[\sum_{t=1}^T X_t^2 - T \gamma(o) \right] \text{ converge,}$$

c) puis à étudier la loi associée à la fonction limite et à voir notamment qu'elle ne correspond pas à une loi normale.

a) Fonction caractéristique d'une forme quadratique de vecteur gaussien.

L'approche repose sur la propriété ci-dessous qui a d'ailleurs son intérêt propre et sera notamment utilisée dans l'étude statistique des processus fractionnaires.

Propriété 3.13.- Soit X un vecteur gaussien de dimension finie T , centré et de matrice de variance-covariance Ω . Alors :

$$\begin{aligned}\psi(z) &= E \left[\exp iz \sum_{t=1}^T X_t^2 \right] = E \left[\exp iz X'X \right] \\ &= \exp \frac{1}{2} \sum_{k=1}^{\infty} \frac{z^k 2^k i^k}{k} \text{Tr}(\Omega^k),\end{aligned}$$

si z est choisi assez petit.

Preuve : Voir appendice 4.

Remarquons à ce niveau que les réels positifs $\text{Tr}(\Omega^k)$ ont les expressions développées :

$$\text{Tr } \Omega = \sum_j \omega_{jj},$$

$$\text{Tr } \Omega^2 = \sum_j \sum_i \omega_{ij} \omega_{ji},$$

$$\text{Tr } \Omega^k = \sum_{i_1} \sum_{i_2} \dots \sum_{i_k} \omega_{i_1 i_2} \omega_{i_2 i_3} \dots \omega_{i_k i_1}.$$

Dans le cas où les composantes de $X : X_1 \dots X_T$, correspondent aux valeurs successives d'un processus stationnaire, on a donc :

$$\text{Tr } \Omega^k = \sum_{i_1=1}^T \sum_{i_2=1}^T \dots \sum_{i_k=1}^T \gamma(i_1-i_2) \gamma(i_2-i_3) \dots \gamma(i_k-i_1)$$

b) Application à l'étude de la variance empirique d'un processus stationnaire gaussien.

Considérons alors le processus stationnaire $X = (X_t, t \in \mathbb{Z})$, d'autovariance $\gamma(h) = (1+h^2)^{-\gamma}$, et intéressons nous à la convergence faible de :

$$(3.14) \quad Z_T = T^{2\gamma} [\hat{\gamma}_T(o) - \gamma(o)] = T^{2\gamma} \left[\frac{1}{T} \sum_{t=1}^T X_t^2 - \gamma(o) \right].$$

La fonction caractéristique de Z_T est :

$$\begin{aligned} \psi_T(z) &= E \exp (iz Z_T) \\ &= \exp [- iz T^{2\gamma} \gamma(o)] E \left[\exp i \frac{z}{T^{1-2\gamma}} X'X \right] \\ &= \exp [- iz T^{2\gamma} \gamma(o)] \exp \left\{ \frac{1}{2} \sum_{k=1}^{\infty} \frac{z^k 2^k i^k}{k(T^{1-2\gamma})^k} \text{Tr}(\Omega_T^k) \right\}, \end{aligned}$$

en notant Ω_T la matrice de variance-covariance de $(X_1, \dots, X_T)'$.

Comme $\text{Tr } \Omega_T = T\gamma(o)$, on voit qu'il y a une simplification par le terme correspondant à $k = 1$, de sorte que :

$$\psi_T(z) = \exp \frac{1}{2} \sum_{k=2}^{\infty} \frac{z^k 2^k i^k}{k} \frac{\text{Tr}(\Omega_T^k)}{(T^{1-2\gamma})^k}.$$

D'autre part, on a :

$$\begin{aligned} \frac{\text{Tr}(\Omega_T^k)}{[T^{1-2\gamma}]^k} &= \frac{1}{(T^{1-2\gamma})^k} \sum_{i_1=1}^T \cdots \sum_{i_k=1}^T \gamma(i_1-i_2) \cdots \gamma(i_k-i_1) \\ &= \frac{1}{T^k} \sum_{i_1=1}^T \cdots \sum_{i_k=1}^T \left[\frac{1}{T^2} + \left(\frac{i_1-i_2}{T} \right)^2 \right]^{-\gamma} \cdots \left[\frac{1}{T^2} + \left(\frac{i_k-i_1}{T} \right)^2 \right]^{-\gamma} \end{aligned}$$

quantité qui, lorsque T tend vers l'infini, peut être approchée par :

$$\frac{1}{T^k} \sum_{i_1=1}^T \cdots \sum_{i_k=1}^T \left(\frac{i_1-i_2}{T} \right)^{-2\gamma} \cdots \left(\frac{i_k-i_1}{T} \right)^{-2\gamma}$$

Comme il s'agit d'une somme de Riemann, elle tend vers l'intégrale correspondante :

$$(3.15) \quad c_k = \int_0^1 \cdots \int_0^1 [x_1-x_2]^{-2\gamma} [x_2-x_3]^{-2\gamma} \cdots [x_k-x_1]^{-2\gamma} dx_1 \cdots dx_k$$

Propriété 3.16.- Si X est stationnaire, gaussien centré, de covariances

$$\gamma(h) = (1+h^2)^{-\gamma}, \quad 0 < \gamma < \frac{1}{4},$$

la fonction caractéristique de :

$$Z_T = T^{2\gamma-1} \left[\sum_{t=1}^T X_t^2 - T \gamma(0) \right]$$

converge, pour z assez petit, vers :

$$\psi_\infty(z) = \exp \frac{1}{2} \sum_{k=2}^{\infty} \frac{z^k}{k} \frac{2^k}{k} c_k,$$

où les constantes c_k sont données par la formule (3.15).

Remarque 3.17.- [voir Taqqu (1975)] . Dans la démonstration précédente, nous avons évidemment sauté une étape. Le rayon de convergence de la série apparaissant sous l'exponentielle dans l'expression de ϕ_T dépend de l'indice T et il nous aurait fallu vérifier que la limite inférieure de ces rayons n'est pas nulle. En fait il suffit de regarder, s'il existe une constante $C > 0$ telle que $c_k \leq C^k$.

Effectuant le changement de variable : $y_j = x_j - x_{j+1}$
 $j = 1, \dots, k-1$ on obtient :

$$c_k = \int \dots \int y_1^{-2\gamma} \dots y_{k-1}^{-2\gamma} (y_1 + \dots + y_{k-1})^{-2\gamma} dy_1 \dots dy_{k-1}$$

et appliquant l'inégalité de Schwarz :

$$c_k \leq \left[\int_{-1}^1 |y|^{-4\gamma} dy \right]^{1/2} \left[\int_{-1}^1 \dots \int_{-1}^1 |y_1|^{-4\gamma} \dots |y_{k-1}|^{-4\gamma} dy_1 \dots dy_{k-1} \right]^{1/2}$$
$$= \left[\frac{2}{1-4\gamma} \right]^{k/2} .$$

Remarque 3.18.- Il est clair d'autre part que dans la démonstration l'important n'est pas la forme précise des autocovariances $\gamma(h) = (1+h^2)^{-\gamma}$, mais essentiellement le fait que $\gamma(h)$ soit asymptotiquement équivalent à $h^{-2\gamma}$ (voir Taqqu (1975)). Cette condition étant réalisée pour les processus MADE puisque $\gamma(h) \sim h^{2d-1}$, on pourra leur appliquer le résultat précédent. Comme $\gamma = \frac{1}{2} - d$, la condition $0 < \gamma < \frac{1}{4}$ devient $\frac{1}{4} < d < \frac{1}{2}$. Ceci explique le résultat de convergence en loi donné en (3.12).

c) Le processus de Rosenblatt.

La fonction caractéristique limite, que nous avons obtenue, contient sous l'exponentielle des termes de tous degrés, puisque $c_k > 0, \forall k$. Elle correspond donc à une loi non normale. Cette fonction caractéristique a été généralisée au cas de processus indexés par $[0,1]$ [voir Fox-Taquq (1985)].

Définition 3.19.- On appelle processus de Rosenblatt de paramètre γ un processus $(R_\gamma(t), t \in [0,1])$, dont les lois fini-dimensionnelles sont telles que :

$$\begin{aligned} E \exp \left[i \sum_{j=1}^p z_j R_\gamma(t_j) \right] \\ = \exp \left\{ \frac{1}{2} \sum_{k=2}^{\infty} \frac{2^k i^k}{k} \sum_{s_1, \dots, s_k \in \{1, 2, \dots, p\}} z_{s_1} z_{s_2} \dots z_{s_k} \right. \\ \left. \int_0^1 \dots \int_0^1 \mathbb{1}_{x_1 \leq t_{s_1}} \dots \mathbb{1}_{x_k \leq t_{s_k}} (x_1 - x_2)^{-2\gamma} \dots (x_k - x_1)^{-2\gamma} dx_1 \dots dx_k \right\} \end{aligned}$$

Dans le cas où on ne s'intéresse qu'à la loi de $R_\gamma(1)$, on obtient :

$$E \left[\exp iz R_\gamma(1) \right] = \exp \frac{1}{2} \sum_{k=2}^{\infty} \frac{2^k i^k z^k}{k} c_k,$$

c'est-à-dire la fonction caractéristique limite vue en b.

III.C.3. - L'approche par les polynômes d'Hermite.

Une étude complète de la détermination de la loi asymptotique des autocovariances empiriques a été faite par Fox-Taquq (1985) et est fondée sur les transformés du processus par les polynômes d'Hermite.

a) Rappelons que les polynômes d'Hermite sont donnés par :

$$(3.20) \quad H_q(x) = (-1)^q \exp \frac{x^2}{2} \frac{d^q}{dx^q} \left(\exp - \frac{x^2}{2} \right), \quad q \in \mathbb{N} .$$

En particulier les premiers de ces polynômes sont :

$H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = x^2 - 1$. Ces polynômes constituent un système orthogonal dans $L^2\left[\mathbb{R}, \frac{1}{\sqrt{2\pi}} \exp - \frac{x^2}{2} dx\right]$ et plus précisément satisfont :

$$E[H_q(x) H_\ell(x)] = \int_{-\infty}^{+\infty} H_q(x) H_\ell(x) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = q! \delta_{\ell q},$$

où $\delta_{\ell q}$ est le symbole de Kronecker.

b) Les résultats obtenus par Fox et Taqqu concernent des processus X stationnaires, gaussiens, tels que :

$$E X_t = 0, \quad E X_t^2 = 1 \quad \text{et} \quad \gamma(h) = E(X_t X_{t+h}) \sim h^{-2\gamma} L(h),$$

lorsque h tend vers l'infini. γ est contraint par $0 < \gamma < \frac{1}{2}$ et

L est une fonction à variation lente à l'infini [voir définition (1.19)].

Les résultats établis sont les suivants :

Propriété 3.21. - Soient X un processus satisfaisant les hypothèses du b) et $(a_s, -\infty < s < +\infty)$ une suite de réels vérifiant $a_{-s} = a_s$ et $\sum_{s=-\infty}^{+\infty} |a_s| < +\infty$.

i) Si $\frac{1}{4} < \gamma < \frac{1}{2}$, le processus :

$$Z_T(r) = \frac{1}{\sqrt{T}} \left\{ \sum_{j=1}^{[Tr]} \sum_{k=1}^{[Tr]} a_{j-k} H_m(X_j) H_m(X_k) \right. \\ \left. - E \left(\sum_{j=1}^{[Tr]} \sum_{k=1}^{[Tr]} a_{j-k} H_m(X_j) H_m(X_k) \right) \right\}$$

converge en loi dans $D[0,1]$ vers le processus $\sigma_m B(r)$, où B est le mouvement brownien et où :

$$\sigma_m^2 = (m!)^2 \sum_{n=1}^m \left\{ \binom{m}{n}^2 \sum_{s_1=-\infty}^{+\infty} \sum_{s_2=-\infty}^{+\infty} a_{s_1} a_{s_2} [\gamma(s_1) \gamma(s_2)]^{m-n} \right. \\ \left. \sum_{q=0}^n \binom{n}{q}^2 \sum_{k=-\infty}^{+\infty} [\gamma(k) \gamma(k+s_1-s_2)]^q [\gamma(k+s_1) \gamma(k-s_2)]^{n-q} \right\} .$$

ii) Si $0 < \gamma < \frac{1}{4}$, le processus :

$$Z_T(r) = \frac{1}{T^{1-2\gamma} \gamma_L(T)} \left\{ \sum_{j=1}^{[Tr]} \sum_{k=1}^{[Tr]} a_{j-k} H_m(X_j) H_m(X_k) \right. \\ \left. - E \left[\sum_{j=1}^{[Tr]} \sum_{k=1}^{[Tr]} a_{j-k} H_m(X_j) H_m(X_k) \right] \right\}$$

converge faiblement dans $D[0,1]$ vers le processus :

$$m! \cdot m \left[\sum_{s=-\infty}^{+\infty} a_s \gamma(s)^{m-1} \right] R_\gamma(r) .$$

On voit directement apparaître sur ce théorème les deux types de lois limites possibles.

III.C.4. - Application aux processus MADE.

De façon à obtenir les covariances empiriques il suffit de prendre le polynôme d'Hermite de degré $m = 1$ et la suite $(a_s, -\infty < s < +\infty)$ telle que : $a_h = a_{-h} = 1$ et $a_j = 0$, si $j \neq h$ ou $j \neq -h$. Prenant de plus $r = 1$, on a :

$$\begin{aligned} & [\text{Tr}] \quad [\text{Tr}] \\ & \sum_{j=1}^T \sum_{k=1}^T a_{j-k} H_m(X_j) H_m(X_k) \\ & = \sum_{j=1}^T \sum_{k=1}^T a_{j-k} X_j X_k \\ & = \begin{cases} 2 \sum_{j=1}^{T-h} X_j X_{j+h} & , \quad \text{si } h \neq 0 \quad , \\ \sum_{j=1}^T X_j^2 & , \quad \text{si } h = 0 \quad . \end{cases} \end{aligned}$$

D'autre part, on a :

$$\begin{aligned} \sigma_1^2 &= \sum_{s_1=-\infty}^{+\infty} \sum_{s_2=-\infty}^{+\infty} a_{s_1} a_{s_2} \sum_{K=-\infty}^{+\infty} [\gamma(K+s_1) \gamma(K-s_2) + \gamma(K) \gamma(K+s_1-s_2)] \\ &= \begin{cases} 2 \sum_{K=-\infty}^{+\infty} \gamma^2(K) & , \quad \text{si } h = 0 \quad , \\ 4 \sum_{K=-\infty}^{+\infty} [\gamma^2(K) + \gamma(K-h) \gamma(K+h)] & , \quad \text{si } h \neq 0 \quad . \end{cases} \end{aligned}$$

On en déduit directement la propriété (3.11) :

$$\sqrt{T} \left| \tilde{\gamma}_T(h) - \gamma(h) \right| \xrightarrow{d} \left[\sum_{K=-\infty}^{+\infty} \gamma^2(K) + \gamma(K-h) \gamma(K+h) \right]^{1/2} B(1, \cdot) ,$$

comme corollaire de la propriété (3.21) ; on déduit aussi de la propriété 3.21 la propriété 3.12 :

$$\frac{1}{T^{2d-1}} \left[\tilde{\gamma}_T(h) - \gamma(h) \right] \xrightarrow{d} \gamma(0) R_{\frac{1}{2}-d} (1)$$

(le terme $\gamma(0)$ étant du au fait que la propriété 3.21 est déduite pour des processus de variance 1).

IV - ETUDE DES MOMENTS EMPIRIQUES D'ORDRE DEUX

[CAS NON STATIONNAIRE] .

Dans le corollaire 2.7, nous avons vu que, pour un processus fractionnaire intégré :

$$X_t = (1-L)^{-d} u_t \quad , \quad \text{avec } d > \frac{1}{2} \quad \text{et}$$

$$\Phi(L) u_t = \textcircled{+}(L) \tilde{\varepsilon}_t \quad , \quad \varepsilon \text{ bruit blanc indépendant ,}$$

le processus raccourci correspondant :

$$\tilde{X}_T(r) = \frac{X_{[Tr]}}{T^{d - \frac{1}{2}}} \quad ,$$

converge en loi vers $\tilde{X}_\infty(r) = \frac{\eta}{\Gamma(d)} \int_0^r (r-s)^{d-1} dB(s)$, $r \in [0,1]$. Ce résultat

va maintenant être utilisé pour déterminer le comportement asymptotique

des variance et covariance empiriques dans le cas non stationnaire.

IV.A. - VARIANCE EMPIRIQUE

Nous avons :

$$\begin{aligned} \frac{1}{T^{2d}} \sum_{t=1}^T X_t^2 &= \frac{1}{T} \sum_{t=1}^T \left(\frac{X_t}{T^{d - \frac{1}{2}}} \right)^2 \\ &= \frac{1}{T} \sum_{t=1}^T \tilde{X}_T\left(\frac{t}{T}\right)^2 \\ &= \sum_{t=1}^T \int_{t-1/T}^{t/T} \tilde{X}_T\left(\frac{t}{T}\right)^2 dr \\ &= \int_0^1 \left[\sum_{t=1}^T \tilde{X}_T\left(\frac{t}{T}\right)^2 \mathbb{1}_{\frac{t-1}{T} \leq r < \frac{t}{T}} \right] dr \end{aligned}$$

Utilisant la convergence en loi de $\hat{X}_T^v(r)$ vers $\hat{X}_\infty^v(r)$ et le théorème de continuité [Billingsley [1968]], on en déduit la convergence en loi de

$$\begin{aligned} & \frac{1}{T^{2d}} \sum_{t=1}^T X_t^2 \text{ vers :} \\ & \int_0^1 \hat{X}_\infty^2(r) \, dr \\ & = \frac{n^2}{\Gamma^2(d)} \int_0^1 \left[\int_0^r (r-u)^{d-1} \, dB(u) \right]^2 \, dr \\ & = \frac{n^2}{\Gamma^2(d)} \int_0^1 \int_0^1 \int_0^1 (r-u)^{d-1} (r-v)^{d-1} \mathbb{1}_{u \leq r} \mathbb{1}_{v \leq r} \, dr \, dB(u) \, dB(v) \\ & = \frac{n^2}{\Gamma^2(d)} \int_0^1 \int_0^1 \left[\int_{\text{Min}(u,v)}^1 (r-u)^{d-1} (r-v)^{d-1} \, dr \right] \, dB(u) \, dB(v) . \end{aligned}$$

Propriété 4.1.- Dans le cas d'un processus non stationnaire fractionnaire intégré on a la convergence en loi de

$$\begin{aligned} & \frac{1}{T^{2d}} \sum_{t=1}^T X_t^2 = \frac{1}{T^{2d-1}} \hat{\gamma}_T(o) \text{ vers :} \\ & \frac{n^2}{\Gamma^2(d)} \int_0^1 \int_0^1 \left[\int_{\text{Min}(u,v)}^1 (r-u)^{d-1} (r-v)^{d-1} \, dr \right] \, dB(u) \, dB(v) . \end{aligned}$$

Il y a donc convergence en loi et ceci est valable sans aucune hypothèse de normalité sur la loi du processus. La limite obtenue n'a évidemment aucune interprétation simple en terme de covariance théorique du processus X . Dans ce cas non stationnaire les variances théoriques $\gamma_t(o) = E X_t^2$ dépendent en effet de l'indice t , de sorte qu'on ne peut espérer montrer une convergence de $\hat{\gamma}_T(o)$ vers une valeur stationnaire $\gamma(o)$.

IV.B. - COVARIANCE EMPIRIQUE

Le résultat (4.1) peut être étendu aux autocovariances empiriques. Nous donnons cette extension pour l'autocovariance d'ordre 1. L'idée est de commencer par étudier le produit croisé :

$$(4.2) \quad S_T = \sum_{t=1}^T X_{t-1} [\bar{X}_t - X_{t-1}] .$$

On a en remarquant que : $X_t = X_{t-1} + (X_t - X_{t-1})$:

$$\begin{aligned} S_T &= \frac{1}{2} \left[\sum_{t=1}^T X_t^2 - \sum_{t=1}^T X_{t-1}^2 - \sum_{t=1}^T (X_t - X_{t-1})^2 \right] \\ &= \frac{1}{2} X_T^2 - \frac{1}{2} \sum_{t=1}^T (X_t - X_{t-1})^2 . \end{aligned}$$

Le premier terme $\frac{1}{2} X_T^2$ est d'ordre T^{2d-1} .

Le second terme $\frac{1}{2} \sum_{t=1}^T (X_t - X_{t-1})^2$ est directement lié à la variance empirique de la série différenciée. Cette dernière est stationnaire, si

$d < \frac{3}{2}$ et est non stationnaire, si $d > \frac{3}{2}$. Si $d < \frac{3}{2}$, on sait alors que $\frac{1}{T} \sum_{t=1}^T (X_t - X_{t-1})^2$ tend vers une limite et donc le second terme

est d'ordre T ; si $d > \frac{3}{2}$, il résulte de la propriété 4.1. que

$\frac{1}{T^{2d-2}} \sum_{t=1}^T (X_t - X_{t-1})^2$ tend en loi vers une limite. On en déduit que,

si $d < 1$, le terme dominant est le second et si $d > 1$, le premier.

Propriété 4.3. - Etant donné un processus non stationnaire

fractionnaire intégré, on a :

$$\text{si } d < 1 : \frac{1}{T} S_T \xrightarrow{d} -\frac{1}{2} \text{plim}_T \frac{1}{T} \sum_{t=1}^T (X_t - X_{t-1})^2 ,$$

$$\left| \begin{array}{l} \text{si } d = 1 : \frac{1}{T} S_T \xrightarrow{d} \frac{1}{2} \quad \hat{X}_\infty^2(1) - \frac{1}{2} \text{plim}_T \frac{1}{T} \sum_{t=1}^T (X_t - X_{t-1})^2, \\ \text{si } d > 1 : \frac{1}{T^{2d-1}} S_T \xrightarrow{d} \frac{1}{2} \quad \hat{X}_\infty^2(1) = \frac{1}{2} \frac{\eta^2}{\Gamma^2(d)} \left[\int_0^1 (1-s)^{d-1} dB(s) \right]^2. \end{array} \right.$$

Remarquons que sauf dans le cas limite $d = 1$, la covariance empirique $\frac{1}{T} S_T$ entre la série initiale et la série différenciée a asymptotiquement un signe fixe : négatif si $d < 1$, positif si $d > 1$.

On déduit directement des propriétés (4.1) et (4.3) le comportement de l'autocovariance empirique d'ordre 1 : $\hat{\gamma}_T(1) = \frac{1}{T} \sum_{t=1}^T X_t X_{t-1}$.
On a en effet :

$$S_T = \sum_{t=1}^T X_{t-1} X_t - \sum_{t=1}^T X_{t-1}^2 = T \hat{\gamma}_T(1) - T \hat{\gamma}_{T-1}(0),$$

et donc :
$$\hat{\gamma}_T(1) = \frac{1}{T} S_T + \hat{\gamma}_{T-1}(0).$$

Lorsque $d \leq 1$, le premier terme $\frac{1}{T} S_T$ est d'ordre T^0 et pour $d > 1$, il est d'ordre T^{2d-2} . D'après 4.1, le second terme est d'ordre T^{2d-1} .
On en déduit immédiatement le corollaire suivant :

Propriété 4.4.- Etant donné un processus non stationnaire fractionnaire intégré, on a, puisque $d > \frac{1}{2}$, l'équivalence asymptotique en probabilité entre $\frac{\hat{\gamma}_T(1)}{T^{2d-1}}$ et $\frac{\hat{\gamma}_T(0)}{T^{2d-1}}$ et en particulier le fait que ces deux suites ont même loi limite.

Ceci rejoint la proposition 1 p. 62 de Akonom (1987) montrant que pour $\frac{1}{2} < d < \frac{3}{2}$, $\frac{1}{T^{2d-1}} \hat{\gamma}_T(h)$ converge en loi vers une variable aléatoire indépendante de l'indice h . L'étude qui vient d'être faite

montre que ce résultat passe aux valeurs de $d > \frac{3}{2}$.

IV.C. - PERIODOGRAMME

Les propriétés d'un processus à l'ordre deux peuvent aussi dans le cas stationnaire être analysées dans le domaine des fréquences, c'est-à-dire à partir de la densité spectrale ou de son estimateur usuel le périodogramme. Bien que nous soyons ici dans un contexte non stationnaire nous pouvons cependant introduire par analogie le pseudo-spectre du processus :

$$f_X(\omega) = \frac{\sigma^2}{2\pi} \frac{|\Theta[\exp i\omega]|^2}{|1-\exp i\omega|^{2d} |\Phi(\exp i\omega)|^2} .$$

Au voisinage de $\omega = 0$, cette fonction est équivalente à :

$$f_X(\omega) \sim \frac{\sigma^2}{2\pi} \frac{\Theta(1)^2}{\Phi(1)^2} \frac{1}{\omega^{2d}} = \frac{\eta^2}{2\pi} \frac{1}{\omega^{2d}} ,$$

de sorte que l'étude du pseudo-spectre au voisinage de l'origine fournit des renseignements intéressants sur la valeur du degré de différentiation d .

Le pseudo-spectre étant en pratique inconnu, nous nous proposons de remplacer cette étude par celle du périodogramme :

$$(4.5) \quad I_T(\omega) = \frac{1}{2\pi T} \left| \sum_{t=1}^T X_t \exp it\omega \right|^2 .$$

De façon à mener cette étude au voisinage de $\omega = 0$, nous allons faire tendre ω vers zéro lorsque T tend vers l'infini. De façon plus précise nous avons le résultat suivant :

Propriété 4.6.- Si X est un processus non stationnaire fractionnaire intégré, on a :

$$\frac{1}{T^{2d}} I_T(\frac{\omega}{T}) \xrightarrow{d} \frac{\eta^2}{\Gamma^2(d)2\pi} \left| \int_0^1 \left[\int_s^1 \exp(i\omega r) (r-s)^{d-1} dr \right] dB(s) \right|^2 .$$

Preuve : La démarche est la même que dans les paragraphes précédents. On a :

$$\begin{aligned} \frac{1}{T^{2d}} I_T(\frac{\omega}{T}) &= \frac{1}{2\pi T^{2d+1}} \left| \sum_{t=1}^T X_t \exp i \frac{t}{T} \omega \right|^2 \\ &= \frac{1}{2\pi} \left| \frac{1}{T} \sum_{t=1}^T \frac{X_t}{T^{d-1/2}} \exp i \frac{t}{T} \omega \right|^2 \\ &= \frac{1}{2\pi} \left| \frac{1}{T} \sum_{t=1}^T \exp(i \frac{t}{T} \omega) \tilde{X}_T(\frac{t}{T}) \right|^2 \\ &= \frac{1}{2\pi} \left| \int_0^1 \sum_{t=1}^T \exp(i \frac{t}{T} \omega) \tilde{X}_T(\frac{t}{T}) \mathbb{1}_{\frac{t-1}{T} \leq r < \frac{t}{T}} dr \right|^2 , \end{aligned}$$

qui converge en loi vers :

$$\begin{aligned} &\frac{1}{2\pi} \left| \int_0^1 \exp i\omega r \tilde{X}_\infty(r) dr \right|^2 \\ &= \frac{\eta^2}{\Gamma^2(d)2\pi} \left| \int_0^1 \exp i\omega r \left[\int_0^r (r-s)^{d-1} dB(s) \right] dr \right|^2 . \end{aligned}$$

Le résultat s'en déduit par interversion des signes d'intégration. ■

On voit au niveau de cette démonstration qu'il serait peut être plus intéressant de calculer uniquement :

$$\tilde{I}_T(\omega) = \sum_{t=1}^T X_t \exp it\omega .$$

On aurait alors : $\frac{1}{T} \frac{1}{d + \frac{1}{2}} \tilde{I}_T(\frac{\omega}{T}) \xrightarrow{d} \frac{\eta}{\Gamma(d)} \int_0^1 \left[\int_s^1 \exp i\omega r (r-s)^{d-1} dr \right] dB(s)$

et on pourrait notamment étudier très faiblement les liaisons asymptotiques entre les valeurs associées à deux fréquences différentes ω_1 et ω_2 :

$$\begin{aligned} & \text{Cov}_{as} \left[\frac{1}{T} \frac{1}{d + \frac{1}{2}} \tilde{I}_T(\frac{\omega_1}{T}), \frac{1}{T} \frac{1}{d + \frac{1}{2}} \tilde{I}_T(\frac{\omega_2}{T}) \right] \\ &= \frac{\eta^2}{\Gamma^2(d)} \int_0^1 \left[\int_s^1 \exp(i\omega_1 r_1) (r_1-s)^{d-1} dr_1 \right] \left[\int_s^1 \exp(i\omega_2 r_2) (r_2-s)^{d-1} dr_2 \right] ds \\ &= \frac{\eta^2}{\Gamma^2(d)} \int_0^1 \int_0^1 \int_0^1 \exp i(\omega_1 r_1 - \omega_2 r_2) (r_1-s)^{d-1} (r_2-s)^{d-1} \mathbb{1}_{s \leq \text{Min}(r_1, r_2)} dr_1 dr_2 ds . \end{aligned}$$

VITESSE DE CONVERGENCE DE LA MOYENNE EMPIRIQUE

$\alpha \backslash d$	$d < -\frac{1}{2}$	$d = -\frac{1}{2}$	$-\frac{1}{2} < d < \frac{1}{2}$	$\frac{1}{2} < d$
$ \alpha > 1$	avec d entier négatif : $T^{1/2}$			
$\alpha = 1$	T^1	$\frac{T}{\sqrt{\log T}}$	$T^{1/2-d}$	$T^{1/2-d}$
$ \alpha < 1$	$T^{1/2}$	$T^{1/2}$	$T^{1/2}$	$T^{1/2}$
$\alpha = -1$	$T^{1/2}$	$T^{1/2}$	$T^{1/2}$	

LOI LIMITE DE LA MOYENNE EMPIRIQUE

((N) sous hypothèse de normalité du bruit,

(S) sans hypothèse de normalité).

α \ d	$d < -\frac{1}{2}$	$d = -\frac{1}{2}$	$-\frac{1}{2} < d < \frac{1}{2}$	$\frac{1}{2} < d$
$\alpha = 1$			normale (N)	normale (S)
$ \alpha < 1$	normale (S)	normale (S)	normale (S)	normale (S)
$\alpha = -1$	normale (S)	normale (S)	normale (S)	

VITESSE DE CONVERGENCE DES AUTOCOVARIANCES EMPIRIQUES

((N) avec hypothèse de normalité,

(S) sans hypothèse de normalité).

α \ d	$d \leq 0$	$0 < d < \frac{1}{4}$	$\frac{1}{4} < d < \frac{1}{2}$	$\frac{1}{2} < d$
$\alpha = 1$	$T^{1/2}$ (S)	$T^{1/2}$ (S)	$T^{1/2-d}$ (N)	T^{1-2d} (*) (S)
$ \alpha < 1$	$T^{1/2}$ (S)	$T^{1/2}$ (S)	$T^{1/2}$ (S)	$T^{1/2}$ (S)

(*) Calculé par écart à zéro et non par écart à $\gamma(0)$.

LOI LIMITE DES AUTOCOVARIANCES EMPIRIQUES

((N) avec hypothèse de normalité,

(S) sans hypothèse de normalité).

α \ d	$d \leq 0$	$0 < d < \frac{1}{4}$	$\frac{1}{4} < d < \frac{1}{2}$	$\frac{1}{2} < d$
$\alpha = 1$	Normale (S)	Normale (N)	Rosenblatt (N)	loi d'une forme quadra- tique de Brownien (S)
$ \alpha < 1$	Normale (S)	Normale (S)	Normale (S)	Normale (S)

A P P E N D I C E 1

PREUVE DE LA PROPRIETE 1.6 :

La variance de la moyenne empirique est donnée par :

$$V \bar{X}_T = \int_{-\pi}^{\pi} \frac{(\sin \frac{\omega T}{2})^2}{T^2 \sin^2 \frac{\omega}{2}} f(\omega) d\omega$$

[voir propriété 1.1.]

Comme $f(\omega) = \frac{\sigma^2}{2\pi} (4 \sin^2 \frac{\omega}{2})^\delta = \frac{\sigma^2}{2\pi} 2^{2\delta} (\sin \frac{\omega}{2})^{2\delta}$, on a après remplacement :

$$\begin{aligned} V \bar{X}_T &= \int_{-\pi}^{\pi} (\sin \frac{\omega T}{2})^2 \frac{\sigma^2}{2\pi T^2} 2^{2\delta} (\sin \frac{\omega}{2})^{2\delta-2} d\omega \\ &= \frac{\sigma^2}{\pi T^2} 2^{2\delta-2} \int_{-\pi}^{\pi} (1-\cos \omega T) (\sin \frac{\omega}{2})^{2\delta-2} d\omega . \end{aligned}$$

Utilisant le développement hypergéométrique de $\cos \omega T$ [voir Abramowitz - Stegun [1965] 15.1.17] :

$$\begin{aligned} \cos \omega T &= F(-T, T, \frac{1}{2}, \sin^2 \frac{\omega}{2}) \\ &= 1 + \sum_{j=1}^T \frac{\Gamma(-T+j) \Gamma[T+j]}{\Gamma[\frac{1}{2} + j]} \frac{\Gamma(\frac{1}{2})}{\Gamma(-T) \Gamma(T)} \frac{(\sin \frac{\omega}{2})^{2j}}{j!} , \end{aligned}$$

on obtient :

$$\begin{aligned}
 V \bar{X}_T &= - \frac{\sigma^2}{\pi T^2} 2^{2\delta-2} \sum_{j=1}^T \frac{\Gamma(-T+j) \Gamma[T+j]}{\Gamma[\frac{1}{2} + j]} \frac{\Gamma(\frac{1}{2})}{\Gamma(-T) \Gamma(T)} \frac{1}{j!} \int_{-\pi}^{\pi} (\sin \frac{\omega}{2})^{2j+2\delta-2} d\omega \\
 &= - \frac{\sigma^2}{\pi T^2} 2^{2\delta-1} \sum_{j=1}^T \frac{\Gamma(-T+j) \Gamma[T+j]}{\Gamma[\frac{1}{2} + j]} \frac{\Gamma(\frac{1}{2})}{\Gamma[-T] \Gamma[T]} \frac{1}{j!} \int_0^{\pi} (\sin \omega)^{2j+2\delta-2} d\omega
 \end{aligned}$$

Comme $\text{Re } \delta > -\frac{1}{2}$, on a $\text{Re}(2j + 2\delta - 2) > -1$ et on peut donc appliquer la formule [voir Gradshteyn - Ryzhik (1965) 3.631.8]

$$\int_0^{\pi} (\sin x)^{b-1} dx = \frac{\pi}{2^{b-1} b B[\frac{b+1}{2}, \frac{b+1}{2}]} \text{ pour } \text{Re}(b) > 0.$$

On a donc :

$$\begin{aligned}
 V \bar{X}_T &= - \frac{\sigma^2}{\pi T^2} 2^{2\delta-1} \sum_{j=1}^T \frac{\Gamma[-T+j] \Gamma[T+j]}{\Gamma[\frac{1}{2} + j]} \frac{\Gamma(\frac{1}{2})}{\Gamma(-T) \Gamma(T)} \\
 &\quad \cdot \frac{\pi}{j!} \frac{1}{2^{2j+2\delta-2}} \frac{1}{2^{2j+2\delta-1}} \frac{\Gamma(2j+2\delta)}{\Gamma(j+\delta)^2}.
 \end{aligned}$$

Utilisant le fait que [voir Abramowitz - Stegun [1965] 6.1.18] :

$$\begin{aligned}
 \Gamma(2z) &= 2^{2z-1} \frac{\Gamma(z) \Gamma(z + \frac{1}{2})}{\sqrt{\pi}}, \text{ on voit que :} \\
 \frac{\Gamma(2j+2\delta)}{(2j+2\delta-1) \Gamma(j+\delta)} &= \frac{2^{2j+2\delta-1} \Gamma(j + \delta + \frac{1}{2})}{2 [j + \delta - \frac{1}{2}] \sqrt{\pi}} = \frac{2^{2j+2\delta-2}}{\sqrt{\pi}} \Gamma(j + \delta - \frac{1}{2})
 \end{aligned}$$

Cette égalité permet de simplifier l'expression de la variance :

$$\begin{aligned}
 v \bar{X}_T &= - \frac{\sigma^2}{T^2} \frac{2^{2\delta-1}}{\sqrt{\pi}} \sum_{j=1}^T \frac{\Gamma[-T+j] \Gamma[T+j]}{\Gamma[\frac{1}{2}+j]} \frac{\Gamma(\frac{1}{2})}{\Gamma(-T) \Gamma(T)} \frac{1}{j!} \frac{\Gamma(j + \delta - \frac{1}{2})}{\Gamma(j + \delta)} \\
 &= - \frac{\sigma^2}{T^2} \frac{2^{2\delta-1}}{\sqrt{\pi}} \frac{\Gamma(\delta - \frac{1}{2})}{\Gamma(\delta)} \sum_{j=1}^T \frac{(-T)_j}{(\frac{1}{2})_j} \frac{(T)_j}{j!} \frac{(\delta - \frac{1}{2})_j}{(\delta)_j}
 \end{aligned}$$

en notant $(T)_j = \frac{\Gamma(T+j)}{\Gamma(T)}$.

On peut alors appliquer la formule de Saalschütz [cf. Erdelyi (1953)] :

$$\sum_{j=0}^T \frac{(a)_j (b)_j (-T)_j}{(c)_j (1+a+b-c-T)_j j!} = \frac{(c-a)_T (c-b)_T}{(c)_T (c-a-b)_T}$$

pour en déduire que :

$$\begin{aligned}
 v \bar{X}_T &= - \frac{\sigma^2}{T^2} \frac{2^{2\delta-1}}{\sqrt{\pi}} \frac{\Gamma(\delta - \frac{1}{2})}{\Gamma(\delta)} \left\{ \frac{(\frac{1}{2} - T)_T (1-\delta)_T}{(\frac{1}{2})_T (1-T-\delta)_T} - 1 \right\} \\
 &= - \frac{\sigma^2}{T^2} \frac{2^{2\delta-1}}{\sqrt{\pi}} \frac{\Gamma(\delta - \frac{1}{2})}{\Gamma(\delta)} \left\{ \frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{1}{2} - T)} \frac{\Gamma(1+T-\delta)}{\Gamma(1-\delta)} \right. \\
 &\quad \left. \frac{\Gamma[T + \frac{1}{2}]}{\Gamma(\frac{1}{2})} \frac{\Gamma(1-\delta)}{\Gamma(1-\delta-T)} - 1 \right\} \\
 &= - \frac{\sigma^2}{T^2} \frac{2^{2\delta-1}}{\sqrt{\pi}} \frac{\Gamma(\delta - \frac{1}{2})}{\Gamma(\delta)} \left\{ \frac{\Gamma(\frac{1}{2})^2 \Gamma(1+T-\delta) \Gamma(1-\delta-T)}{\Gamma(\frac{1}{2}+T) \Gamma(\frac{1}{2} - T) \Gamma(1-\delta)^2} - 1 \right\}
 \end{aligned}$$

Comme $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ et que $\Gamma(\frac{1}{2} + T) \Gamma(\frac{1}{2} - T) = \pi \sec(\pi T)$, on a :

$$\begin{aligned} V \bar{X}_T &= -\frac{\sigma^2}{T^2} \frac{2^{2\delta-1} \Gamma(\delta - \frac{1}{2})}{\sqrt{\pi} \Gamma(\delta)} \left\{ (-1)^T \frac{\Gamma(T-\delta+1)}{\Gamma(-\delta+1)} \frac{\Gamma(-\delta-T+1)}{\Gamma(-\delta+1)} - 1 \right\} \\ &= -\frac{\sigma^2}{T^2} \frac{2^{2\delta-1} \Gamma(\delta - \frac{1}{2})}{\sqrt{\pi} \Gamma(\delta)} \left\{ \frac{\Gamma(T-\delta+1)}{\Gamma(-\delta+1)} \frac{\Gamma(\delta)}{\Gamma(\delta+T)} - 1 \right\} \blacksquare \end{aligned}$$

A P P E N D I C E 2

DETERMINATION DE $V \bar{X}_T$ ET DE SON EQUIVALENT ASYMPTOTIQUE,

LORSQUE : $\alpha = 1$, $d = -\frac{1}{2}$.

EXPRESSION DE $V \bar{X}_T$

i) Appliquant la propriété 8 de [Gonçalves [1987]], on sait que pour un tel processus, la fonction d'autocovariance est donnée par :

$$\begin{aligned} \gamma(h) &= \sigma^2 \frac{\Gamma(h - \frac{1}{2})}{\Gamma(h + \frac{3}{2})} \frac{\Gamma(2)}{\Gamma(-\frac{1}{2}) \Gamma(\frac{3}{2})} , \quad h \geq 0 \\ &= \sigma^2 \frac{1}{(h + \frac{1}{2})(h - \frac{1}{2})} \frac{1}{-2 \sqrt{\pi} \frac{1}{2} \sqrt{\pi}} \\ &= -\frac{\sigma^2}{\pi} \frac{1}{(h - \frac{1}{2})(h + \frac{1}{2})} \end{aligned}$$

Cette série est absolument convergente et on sait que :

$$\sum_{h=-\infty}^{+\infty} \gamma(h) = 2\pi f(0) = 0 .$$

ii) D'après la propriété 1.1, on a :

$$\begin{aligned}
 v \bar{X}_T &= \frac{1}{T} \sum_{h=-(T-1)}^{T-1} \left(1 - \frac{|h|}{T}\right) \gamma(h) \\
 &= \frac{\gamma(0)}{T} + \frac{2}{T} \sum_{h=1}^{T-1} \gamma(h) - \frac{2}{T^2} \sum_{h=1}^{T-1} h \gamma(h) .
 \end{aligned}$$

Comme : $\gamma(0) + 2 \sum_{h=1}^{T-1} \gamma(h) = -2 \sum_{h=T}^{\infty} \gamma(h)$, on peut écrire :

$$v \bar{X}_T = -\frac{2}{T} \sum_{h=T}^{+\infty} \gamma(h) - \frac{2}{T^2} \sum_{h=1}^{T-1} h \gamma(h) .$$

iii) Calcul du terme $\sum_{h=T}^{\infty} \gamma(h)$.

On a :

$$\begin{aligned}
 \sum_{h=T}^{\infty} \gamma(h) &= -\frac{\sigma^2}{\pi} \sum_{h=T}^{\infty} \frac{\Gamma(h - \frac{1}{2})}{\Gamma(h + \frac{3}{2})} \\
 &= -\frac{\sigma^2}{\pi} \sum_{\ell=0}^{\infty} \frac{\Gamma(\ell + T - \frac{1}{2})}{\Gamma(\ell + T + \frac{3}{2})} \\
 &= -\frac{\sigma^2}{\pi} \frac{\Gamma(T - \frac{1}{2}) \Gamma(1)}{\Gamma(T + \frac{3}{2})} F(1, T - \frac{1}{2}, T + \frac{3}{2}, 1) \\
 &= -\frac{\sigma^2}{\pi} \frac{\Gamma(T - \frac{1}{2}) \Gamma(1)}{\Gamma(T + \frac{3}{2})} \frac{\Gamma(T + \frac{3}{2}) \Gamma[T + \frac{3}{2} - (T - \frac{1}{2}) - 1]}{\Gamma(T + \frac{3}{2} - 1) \Gamma[T + \frac{3}{2} - (T - \frac{1}{2})]} ,
 \end{aligned}$$

car : $F(a, b, c, 1) = \frac{\Gamma(c) \Gamma(c-a-b)}{\Gamma(c-a) \Gamma(c-b)}$.

D'où :

$$\sum_{h=T}^{\infty} \gamma(h) = -\frac{\sigma^2}{\pi} \frac{\Gamma(T - \frac{1}{2})}{\Gamma(T + \frac{1}{2})} \frac{\Gamma(1)}{\Gamma(2)} = -\frac{\sigma^2}{\pi} \frac{\Gamma(T - \frac{1}{2})}{\Gamma(T + \frac{1}{2})} = -\frac{\sigma^2}{\pi} \frac{1}{T - \frac{1}{2}}$$

iv) Calcul du terme $\sum_{h=1}^{T-1} h \gamma(h)$.

$$\begin{aligned} \text{On a :} \quad h \gamma(h) &= -\frac{\sigma^2}{\pi} \frac{h}{(h - \frac{1}{2})(h + \frac{1}{2})} \\ &= -\frac{\sigma^2}{2\pi} \left[\frac{1}{h - \frac{1}{2}} + \frac{1}{h + \frac{1}{2}} \right] . \end{aligned}$$

Or :

$$\begin{aligned} &\sum_{h=1}^{T-1} \left[\frac{1}{h - \frac{1}{2}} + \frac{1}{h + \frac{1}{2}} \right] \\ &= 2 + 2 \sum_{h=1}^{T-2} \frac{1}{h + \frac{1}{2}} + \frac{1}{T - \frac{1}{2}} \\ &= 2 \left[1 + 2 \sum_{h=1}^{T-2} \frac{1}{2h + 1} + \frac{1}{2T - 1} \right] \\ &= 2 \left[-1 + 2 \sum_{\ell=1}^{T-1} \frac{1}{2\ell - 1} + \frac{1}{2T - 1} \right] . \end{aligned}$$

On déduit alors de la définition de la fonction $\Psi(z) = \frac{d}{dz} \text{Log } \Gamma(z)$ [voir Abramowitz - Stegun (1965) 6.3.4.] que :

$$2 \sum_{\ell=1}^{T-1} \frac{1}{2\ell - 1} = \gamma + 2 \text{Log } 2 + \Psi \left[T - \frac{1}{2} \right] ,$$

où γ est la constante d'Euler.

Remplaçant dans la somme à calculer, nous obtenons :

$$\sum_{h=1}^{T-1} h \gamma(h) = -\frac{\sigma^2}{\pi} \left[-1 + \gamma + 2 \text{Log } 2 + \Psi \left[T - \frac{1}{2} \right] + \frac{1}{2T - 1} \right] .$$

v) Expression de $V \bar{X}_T$

Finalement, on a par addition :

$$\begin{aligned}
 V \bar{X}_T &= \frac{2 \sigma^2}{\pi T} \frac{1}{T - \frac{1}{2}} + \frac{2 \sigma^2}{\pi T^2} \left[-1 + \gamma + 2 \operatorname{Log} 2 + \Psi \left[T - \frac{1}{2} \right] + \frac{1}{2T-1} \right] \\
 &= \frac{2 \sigma^2}{\pi} \left\{ \frac{1}{T(T - \frac{1}{2})} + \frac{1}{T^2} \left[-1 + \gamma + 2 \operatorname{Log} 2 + \Psi \left[T - \frac{1}{2} \right] + \frac{1}{2T-1} \right] \right\}
 \end{aligned}$$

DEVELOPPEMENT ASYMPTOTIQUE DE $V \bar{X}_T$

Le développement asymptotique de la fonction Ψ est [voir Erdelyi [1953] p. 47 formule 4] :

$$\Psi(z) = \operatorname{Log} z - \frac{1}{2z} - \sum_{n=1}^m B_{2n} \frac{z^{-2n}}{2n} + O(z^{-2m-2}) ,$$

où les coefficients B_{2n} sont les nombres de Bernoulli :

$$B_{2n} = 2 (-1)^{n+1} (2n)! \sum_{r=1}^{\infty} (2\pi r)^{-2n}$$

[voir Erdelyi [1953] p. 38 formule 16].

Le développement à l'ordre 1 de la fonction Ψ est donc :

$$\Psi(z) = \operatorname{Log} z - \frac{1}{2z} + O\left(\frac{1}{z^2}\right) ,$$

et

$$\Psi\left(T - \frac{1}{2}\right) = \operatorname{Log}\left(T - \frac{1}{2}\right) - \frac{1}{2T-1} + O\left(\frac{1}{T^2}\right)$$

Donc :

$$V \bar{X}_T = \frac{2 \sigma^2}{\pi} \left\{ \frac{1}{T(T - \frac{1}{2})} + \frac{1}{T^2} \left[-1 + \gamma + 2 \text{Log } 2 + \text{Log}(T - \frac{1}{2}) + o\left(\frac{1}{T}\right) \right] \right\},$$

$$\lim_{T \rightarrow \infty} \frac{T^2}{\text{Log}(T - \frac{1}{2})} V \bar{X}_T = \frac{2 \sigma^2}{\pi},$$

$$\lim_{T \rightarrow \infty} \frac{T^2}{\text{Log } T} V \bar{X}_T = \frac{2 \sigma^2}{\pi}.$$

A P P E N D I C E 3

AUTO-SIMILITUDE DU PROCESSUS BROWNIEN FRACTIONNAIRE.

Partant de la formule (1.17), on a :

$$\frac{1}{a^H} B_H(at, \omega) = - \frac{1}{a^H \Gamma(H + \frac{1}{2})} \left\{ \int_{-\infty}^0 B(s, \omega) d[(at-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}] \right. \\ \left. + \int_0^{at} B(s, \omega) d[at-s]^{H-\frac{1}{2}} \right\} .$$

Effectuant le changement de variable $s = au$, on obtient :

$$\frac{1}{a^H} B_H(at, \omega) = - \frac{1}{\Gamma(H + \frac{1}{2})} \frac{1}{\sqrt{a}} \left\{ \int_{-\infty}^0 B(au; \omega) d[(t-u)^{H-\frac{1}{2}} - (-u)^{H-\frac{1}{2}}] \right. \\ \left. + \int_0^t B(au; \omega) d(t-u)^{H-\frac{1}{2}} \right\} .$$

Comme $\frac{1}{\sqrt{a}} B(au; \omega) \stackrel{\Delta}{=} B(u; \omega)$, on en déduit que :

$$\frac{1}{a^H} B_H(at, \omega) \stackrel{\Delta}{=} - \frac{1}{\Gamma(H + \frac{1}{2})} \left\{ \int_{-\infty}^0 B(u; \omega) d[(t-u)^{H-\frac{1}{2}} - (-u)^{H-\frac{1}{2}}] \right. \\ \left. + \int_0^t B(u; \omega) d(t-u)^{H-\frac{1}{2}} \right\} .$$

$$= B_H(t, \omega) .$$

A P P E N D I C E 4

FONCTION CARACTERISTIQUE D'UNE FORME QUADRATIQUE
DE VECTEUR GAUSSIEN.

Soit X un vecteur gaussien : $X \rightsquigarrow N[0, \Omega]$, on sait qu'il existe une matrice orthogonale Q permettant de rendre diagonale la matrice de variance Ω :

$$Q \Omega Q' = \begin{bmatrix} \lambda_1 & & & & \\ & \ddots & & & \\ & & \circ & & \\ & & & \ddots & \\ \circ & & & & \lambda_T \end{bmatrix} , \text{ avec } \lambda_t \geq 0 , \forall t .$$

On a alors : $Y = QX \rightsquigarrow N[0, \text{diag } \lambda]$

et $X'X = Y'Y$.

On en déduit que :

$$\begin{aligned}
 E [\exp iz X'X] &= E \exp [iz Y'Y] \\
 &= E \left[\exp iz \sum_{t=1}^T Y_t^2 \right] \\
 &= \prod_{t=1}^T [\exp iz Y_t^2] && \text{(car les composantes } Y_t \\
 &&& \text{non corrélées sont indépendantes)} \\
 &= \prod_{t=1}^T \frac{1}{(1-2iz\lambda_t)^{1/2}} && \text{(en utilisant la forme de la} \\
 &&& \text{fonction caractéristique d'une} \\
 &&& \text{loi du } \chi^2) \\
 &= \exp \sum_{t=1}^T -\frac{1}{2} \text{Log}(1-2iz\lambda_t) && \text{(en prenant la détermination} \\
 &&& \text{principale du logarithme).}
 \end{aligned}$$

Comme : $\text{Tr } \Omega^k = \sum_{t=1}^T \lambda_t^k \leq (\text{Tr } \Omega)^k = \left(\sum_{t=1}^T \lambda_t \right)^k$, ce qui entraîne :

$\forall t : \lambda_t^k \leq (\text{Tr } \Omega)^k$, on peut pour z assez petit remplacer

$\text{Log}(1-2iz\lambda_t)$ par son développement en série entière :

$$\begin{aligned}
 E [\exp iz X'X] &= \exp \left[\sum_{t=1}^T \frac{1}{2} \sum_{k=1}^{\infty} \frac{(2iz\lambda_t)^k}{k} \right] \\
 &= \exp \frac{1}{2} \left[\sum_{k=1}^{\infty} \frac{2^k z^k i^k}{k} \sum_{t=1}^T \lambda_t^k \right] \\
 &= \exp \frac{1}{2} \sum_{k=1}^{\infty} \frac{2^k z^k i^k}{k} \text{Tr } \Omega^k .
 \end{aligned}$$

R E F E R E N C E S.

ABRAMOWITZ, M. et I.A. STEGUN (1965) : Handbook of mathematical functions,
Dover.

AKONOM, J. (1987) : Sur les processus transformés d'un ARIMA ou d'un
processus de Wiener, problème d'estimation.
D.P.

AKONOM, J., GOURIÉROUX C. (1987) : A functional limit theorem for
fractional processes.
CEPREMAP, D.P. 8801

ANDERSON, T.W. (1971) : The statistical analysis of time series,
Wiley.

AZENCOTT, R. et D. DACUNHA-CASTELLE (1984) : Séries d'observations
irrégulières.
Masson, Paris.

BARY, N.K. (1964) : A treatise on trigonometric series,
Vol. 1, Pergamon Press.

BILLINGSLEY, P. (1968) : Convergence of probability measures.
Wiley, New-York.

ERDELYI, MAGNUS, OBERHETTINGER et TRICOMI (1953) : Higher transcendental
functions.
Vol. 1, McGraw-Hill.

FOX, R. et TAQUU, M. (1985) : Non central limit theorem for quadratic
forms in random variables having long range dependence,
Ann. of Proba., 13, 428-446.

- FOX, R. et TAQQU, M. (1987) : Central limit theorems for quadratic forms in random variables having long range dependence, Prob. Th. Rel. Fields 74, 213-240.
- GRADSHTEYN, I. et I. RYZHIK (1965) : Tables of integrals, series and products, Academic Press.
- HANNAN E.J. (1970) : Multiple time series, Wiley.
- HERRNDORF, N. (1984) : A functional central limit theorem for weakly dependent sequences of random variables, Annals of Probability 12, 141-153.
- MANDELBROT, B. et M. TAQQU (1979) : Robust R/S analysis of long run serial correlation, Proceedings of the 42nd session of the International Statistical Institute.
- MANDELBROT, B. et J.W. VAN NESS (1968) : Fractional brownian motions, fractional noises and applications, SIAM Rev. 10, 422-437.
- ROSENBLATT, M. (1956) : A central limit theorem and a strong mixing condition, Proc. Nat. Acad. Sci. USA 42, 43-47.
- ROSENBLATT, M. (1961) : Independence and dependence, Proc. 4th Berkeley Symposium in Math. Stat. and Prob., University of California, Berkeley Univ. Press 431-443.
- ROSENBLATT, M. (1985) : Stationary sequences and random fields, Birkhäuser.

TAQQU, M. (1975) : Weak convergence to fractional brownian motion and
to the Rosenblatt'process,
Z. Wahrsch. verw. Gebiete 31, 287-302.

TAQQU, M. (1979) : Convergence of integrated processes of arbitrary
Hermite rank,
Z. Wahrsch. verw. Gebiete, 50, 53-83.

ZYGMUND, A. (1959) : Trigonometric series,
Vol. I., Cambridge Univ. Press.



①

ESTIMATION DES PARAMETRES D'UN MODELE FRACTIONNAIRE

ESTIMATION DES PARAMETRES D'UN MODELE FRACTIONNAIRE

Dans ce chapitre, on souhaite étudier diverses méthodes d'estimation des paramètres d'un modèle fractionnaire. On dispose de T observations successives d'un processus satisfaisant :

$$X_t = \frac{\prod_{k=1}^K (1 - \alpha_k L)^{\delta_k}}{\prod_{k=1}^K (1 - \beta_k L)^{d_k}} \varepsilon_t ,$$

où $\varepsilon = (\varepsilon_t, t \in Z)$ est un bruit blanc indépendant de variance σ^2 admettant des moments jusqu'à l'ordre 4 et où les paramètres α_k et β_k sont tels que $|\alpha_k| < 1$ et $|\beta_k| < 1$. Sous ces conditions le processus admet des représentations moyenne mobile infinie et autorégressive infinie, dont les coefficients décroissent asymptotiquement de façon exponentielle.

Sous ces hypothèses les paramètres $\alpha_k, \beta_k, \delta_k, d_k, \sigma^2$ peuvent être estimés par une méthode du type pseudo-maximum de vraisemblance fondée sur la loi normale.

Nous commençons dans le paragraphe 1 par rappeler comment s'applique cette méthode dans le cas de processus stationnaires, donnons les diverses manières d'approcher la vraisemblance et explicitons les formes de la matrice d'information de Fisher.

Ces résultats sont dans le second paragraphe appliqués aux processus fractionnaires.

Finalement dans le paragraphe 3 nous discutons des procédures d'estimation convergentes fondées sur les équations de récurrence linéaires satisfaites par les autocovariances du processus. Les méthodes d'estimation obtenues s'apparentent à la résolution du système d'équations de Yule-Walker utilisé de façon classique dans l'étude des processus ARMA.

Dans tout ce chapitre, nous analyserons de façon privilégiée les processus MADE(1) : $X_t = (1 - \alpha L)^\delta \varepsilon_t$, et les propriétés des estimateurs des paramètres au voisinage de l'hypothèse de bruit blanc. Cette étude locale constitue en effet une introduction aux problèmes de test analysés dans le chapitre suivant.

I - ESTIMATEURS DU PSEUDO-MAXIMUM DE VRAISEMBLANCE DES PARAMETRES D'UN MODELE MA(∞)

I.A - LE MODELE PARAMETRIQUE

Nous considérons un processus stationnaire univarié $(X_t, t \in Z)$ admettant une écriture moyenne mobile infinie :

$$(1.1) \quad X_t = A(L) \varepsilon_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j},$$

où ε est un bruit blanc indépendant, de variance σ^2 et où $a_0 = 1$.

Ce processus est supposé admettre aussi une représentation autorégressive infinie :

$$(1.2) \quad X_t - \sum_{j=1}^{\infty} \phi_j X_{t-j} = \psi(L) X_t = \varepsilon_t, \quad t \in \mathbb{Z}.$$

Ces deux développements sont supposés converger à une vitesse suffisamment rapide ; de façon plus précise, nous nous restreignons au cas où :

$$\exists c : 0 < c < 1 \text{ telle que } |a_j| < c^j, \quad |\phi_j| < c^j, \quad \forall j.$$

Cette condition technique est essentiellement utile pour appliquer les théorèmes asymptotiques usuels sur les autocovariances empiriques [voir Hannan (1973)] et pour mener à bien l'approximation de Toeplitz [voir I.B].

Nous notons $\{\gamma(h) = E(X_t X_{t+h})\}$, $h \in \mathbb{Z}$ la fonction d'autocovariance

$$\text{du processus et } f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) \exp i\omega h \text{ sa densité spectrale.}$$

La densité spectrale est la transformée de Fourier de la suite des autocovariances. Celles-ci peuvent donc s'écrire par la transformation de Fourier inverse :

$$(1.3) \quad \gamma(h) = \int_{-\pi}^{\pi} \exp(i\omega h) f(\omega) d\omega.$$

De plus, on a l'identité de Kolmogorov :

$$(1.4) \quad \text{Log } \sigma^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{Log } [2\pi f(\omega)] d\omega = E_{\mu} \text{Log } [2\pi f(\omega)],$$

en notant μ la loi uniforme sur $]-\pi, \pi]$ (en fait sur $\mathbb{R} / 2\pi \mathbb{Z}$).

Le processus X admet une représentation spectrale :

$$(1.5) \quad X_t = \int_{-\pi}^{\pi} \exp(-it\omega) d\xi(\omega),$$

où $(\xi(\omega), \omega \in]-\pi, \pi])$ est un processus à accroissements orthogonaux.

Lorsque le processus X est gaussien, les accroissements de ξ le sont également.

Les "accroissements infinitésimaux" $d\xi(\omega)$ sont centrés, de variance

$V[d\xi(\omega)] = f(\omega) d\omega$. La variable $d\xi(\omega)$ s'interprète comme le poids

aléatoire de la composante de fréquence ω dans la décomposition spectrale

de X . Cette interprétation se retrouve au niveau des décompositions des

moments d'ordre deux : de la relation (1.3), on déduit : $\gamma(o) = \int_{-\pi}^{\pi} f(\omega) d\omega$,

c'est-à-dire que la variance de X_t est la moyenne des variances des $d\xi(\omega)$;

de la relation de Kolmogorov (1.4), on déduit :

$$\sigma^2 = \exp \int_{\mu} E \text{Log} [2\pi f(\omega)] ,$$

c'est à dire que la variance de l'innovation est (à une constante près) la

moyenne géométrique des variances des $d\xi(\omega)$. Dans la suite, nous définissons

un modèle semi-paramétrique en spécifiant les résumés d'ordre deux comme

fonctions de paramètres (θ, σ^2) . [Nous traitons dès le début de manière

séparée la variance du bruit]. Les divers résumés à l'ordre deux du processus

sont alors indexés et on écrit :

$$(1.6) \quad A_{\theta}^2(L), \phi_{\theta}^2(L), \gamma_{\theta, \sigma^2}^2(h) = \sigma_{\theta}^2 c_{\theta}^2(h), f_{\theta, \sigma^2}^2(\omega) = \sigma_{\theta}^2 g_{\theta}^2(\omega), \theta \in \Theta .$$

Nous repérons les vraies valeurs de ces résumés, c'est-à-dire celles

correspondant aux vraies valeurs θ_0, σ_0^2 des paramètres par un indice 0

et nous notons donc : $A_{\theta_0}^2(L), \phi_{\theta_0}^2(L), \gamma_{\theta_0, \sigma_0^2}^2(h), f_{\theta_0, \sigma_0^2}^2(\omega)$.

Ces vraies valeurs sont supposées identifiables au second ordre, au sens où

la connaissance de $(\gamma_{\theta, \sigma^2}^{(h)}, h \in Z)$ ou de $(f_{\theta, \sigma^2}^{(\omega)}, \omega \in]-\pi, \pi])$ [resp de $A_{\theta}^{(L)}$ ou $\phi_{\theta}^{(L)}$] permet de retrouver sans ambiguïté la vraie valeur (θ, σ^2) du paramètre [resp : θ du paramètre].

I.B - LA (PSEUDO) LOG-VRAISEMBLANCE ET SES APPROXIMATIONS

I.B.a - Log-vraisemblance exacte

Dans la suite, on construira des vraisemblances en faisant comme si les erreurs $(\varepsilon_t, t \in Z)$ suivaient une loi normale. Si nous disposons d'observations X_1, \dots, X_T du processus X , la log-vraisemblance correspondante est alors donnée par :

$$(1.7) \quad \text{Log } L_T(x; \theta, \sigma^2) = -\frac{T}{2} \text{Log } 2\pi - \frac{1}{2} \log \det \Gamma_T(\theta, \sigma^2) - \frac{1}{2} x_T' \Gamma_T(\theta, \sigma^2)^{-1} x_T,$$

où $x_T = (X_1, \dots, X_T)'$ et où $\Gamma_T(\theta, \sigma^2)$ est la matrice de variance-covariance de $(X_1, \dots, X_T)'$ et a pour terme général :

$$\gamma_{\theta, \sigma^2}^{(i-j)} = \sigma^2 c_{\theta}^{(i-j)}, \quad 1 \leq i, j \leq T.$$

Cette log-vraisemblance peut être réécrite en séparant les deux types de paramètres θ et σ^2 . Pour cela introduisons $C_T(\theta) = \frac{1}{\sigma^2} \Gamma_T(\theta, \sigma^2)$, on a :

$$(1.8) \quad \text{Log } L_T(x; \theta, \sigma^2) = -\frac{T}{2} \text{Log } 2\pi - \frac{T}{2} \text{Log } \sigma^2 - \frac{1}{2} \log \det C_T(\theta)$$

$$- \frac{1}{2\sigma^2} x_T' C_T(\theta)^{-1} x_T .$$

Remarquons que si la forme quadratique $x_T' C_T(\theta) x_T$ a une forme simple, vu que la matrice C_T est de Toeplitz, c'est-à-dire a mêmes éléments sur les paradiagonales parallèles à la diagonale principale, il n'en est pas de même de la forme : $x_T' C_T(\theta)^{-1} x_T$; l'inverse d'une matrice de Toeplitz n'est en effet généralement pas de Toeplitz.

On peut proposer diverses approximations intuitives de la log-vraisemblance exacte. Nous donnons ci-dessous les deux les plus classiques : celle par les lois conditionnelles et l'approximation dans le domaine des fréquences.

I.B.b - Approximation par les lois conditionnelles

La loi du vecteur $x_T = (X_1, \dots, X_T)'$ peut être explicitée en faisant apparaître les diverses lois conditionnelles successives. Si nous notons : $l(X_t/X_{t-1}, \dots, X_1; \theta, \sigma^2)$ la densité de X_t conditionnelle à X_{t-1}, \dots, X_1 , nous avons :

$$L_T(x; \theta, \sigma^2) = \prod_{t=1}^T l(X_t/X_{t-1}, \dots, X_1; \theta, \sigma^2) .$$

Chacune des (pseudo) lois conditionnelles est normale, de moyenne

$$E(X_t/X_{t-1}, \dots, X_1) , \text{ de variance } V(X_t/X_{t-1}, \dots, X_1) .$$

Lorsque t tend vers l'infini, la prévision $E(X_t/X_{t-1}, \dots, X_1)$ converge en moyenne quadratique vers la prévision sur le passé infini :

$$\begin{aligned} E(X_t/X_{t-1}, \dots, X_1) &\neq E(X_t/X_{t-1}, \dots, X_1, \dots) \\ &= \sum_{j=1}^{\infty} \psi_j(\theta) X_{t-j} = - [\phi(\theta)(L) - 1] X_t . \end{aligned}$$

De même la variance : $V(X_t/X_{t-1}, \dots, X_1)$ peut être approchée par l'erreur de prévision σ^2 .

Une approximation naturelle de la log-vraisemblance exacte est donc :

$$(1.9) \quad \text{Log } L_T^c(x; \theta, \sigma^2) = -\frac{T}{2} \text{Log } 2\pi - \frac{T}{2} \text{Log } \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^T [\phi_\theta(L) X_t]^2.$$

Cette fonction sera appelée log-vraisemblance conditionnelle théorique. Elle présente le défaut de ne pas dépendre des composantes du processus X au travers des seules observations X_1, \dots, X_T . Elle peut être modifiée pour tenir compte de cette difficulté. Par exemple, on peut introduire la log-vraisemblance conditionnelle :

$$(1.10) \quad \text{Log } L_T^{ac}(x; \theta, \sigma^2) = -\frac{T}{2} \text{Log } 2\pi - \frac{T}{2} \text{Log } \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^T [X_t - \phi_1(\theta) X_{t-1} \dots - \phi_{t-1}(\theta) X_1]^2.$$

I.B.c - Approximation dans le domaine des fréquences

Finalement une autre approximation introduite par Whittle (1951) se révèle très utile. Elle est obtenue en écrivant la log-vraisemblance dans le domaine des fréquences.

Remarquons d'abord que l'approximation (1.9) par les lois conditionnelles conduit à remplacer $\text{log det } \Gamma_T(\theta; \sigma^2)$ par $T \text{log } \sigma^2$.

D'autre part, si nous considérons la matrice de Toeplitz infinie $\Gamma(\theta; \sigma^2)$ nous pouvons l'assimiler à la suite des coefficients $(\gamma_{\theta, \sigma}^2(h))$, $h \in \mathbb{Z}$, de transformée de Fourier $f_{\theta, \sigma}^2(\omega)$.

La fonction spectrale inverse $\tilde{f}_{\theta, \sigma}^2(\omega) = \frac{1}{(2\pi)^2} \frac{1}{f_{\theta, \sigma}^2(\omega)}$ est alors

associée à la suite des autocovariances inverses du processus :

$(\gamma_{\theta, \sigma}^2(h))$, $h \in \mathbb{Z}$. Cette suite est reliée à celle des autocovariances par

l'égalité $\gamma_{\theta, \sigma}^2 * \tilde{\gamma}_{\theta, \sigma}^2 = e$, où e est l'élément neutre du produit

de convolution. Introduisant la matrice de Toeplitz infinie $\tilde{\Gamma}_{\theta, \sigma}^2$ associée

à la suite $\tilde{\gamma}_{\theta, \sigma}^2$, la relation sur les suites, en terme de produit de

convolution, se traduit matriciellement par : $\Gamma_{\theta, \sigma}^2 \tilde{\Gamma}_{\theta, \sigma}^2 = \text{Id}_{\infty}$.

$$\Leftrightarrow \tilde{\Gamma}_{\theta, \sigma}^2 = \Gamma_{\theta, \sigma}^{-1}$$

Il est donc naturel d'approcher l'inverse $\Gamma_{\theta, \sigma}^{-1}$ par une sous matrice principale de $\tilde{\Gamma}_{\theta, \sigma}^2$ de taille T .

On aboutit ainsi à l'approximation :

$$x_T' \Gamma_T(\theta, \sigma)^{-1} x_T \approx x_T' \tilde{\Gamma}_T(\theta, \sigma) x_T$$

On peut alors utiliser les résultats classiques sur les formes quadratiques fondées sur des matrices de Toeplitz pour en déduire que :

$$\begin{aligned} x_T' \tilde{\Gamma}_T(\theta, \sigma) x_T &= \int_{-\pi}^{\pi} \left| \sum_{t=1}^T X_t \exp - it\omega \right|^2 \tilde{f}_{\theta, \sigma}^2(\omega) d\omega \\ &= \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \left| \sum_{t=1}^T X_t \exp - it\omega \right|^2 \frac{1}{f_{\theta, \sigma}^2(\omega)} d\omega \end{aligned}$$

Appelant $\hat{f}_T(\omega) = \frac{1}{2\pi T} \left| \sum_{t=1}^T X_t \exp - it\omega \right|^2$ le périodogramme d'ordre T

du processus X , on a donc :

$$x_T' \hat{\Gamma}_T(\theta, \sigma^2) x_T = \frac{T}{2\pi} \int_{-\pi}^{\pi} \frac{\hat{\Gamma}_T(\omega)}{f_{\theta, \sigma^2}(\omega)} d\omega = T E_{\mu} \left(\frac{\hat{\Gamma}_T}{f_{\theta, \sigma^2}} \right),$$

où μ est la loi uniforme sur $]-\pi, \pi]$.

Remplaçant dans l'expression de la (pseudo) log-vraisemblance exacte, nous obtenons :

$$(1.11) \text{Log } \hat{L}_T(x; \theta, \sigma^2) = -\frac{T}{2} \text{Log } 2\pi - \frac{T}{2} \text{Log } \sigma^2 - \frac{T}{2\sigma^2} E_{\mu} \left(\frac{\hat{\Gamma}_T}{g_{\theta}} \right).$$

Finalement l'égalité de Kolmogorov conduit à :

$$\text{Log } \sigma^2 = E_{\mu} [\log 2\pi f_{\theta, \sigma^2}(\omega)] = E_{\mu} [\log (2\pi \sigma^2 g_{\theta}(\omega))]$$

$$\Leftrightarrow \text{Log } 2\pi + E_{\mu} \text{Log } g_{\theta}(\omega) = 0.$$

On a donc :

$$\text{Log } \hat{L}_T(x; \theta, \sigma^2) = -\frac{T}{2} E_{\mu} \text{Log } g_{\theta} - \frac{T}{2} \text{Log } \sigma^2 - \frac{T}{2\sigma^2} E_{\mu} \left(\frac{\hat{\Gamma}_T}{g_{\theta}} \right).$$

I.C - DEFINITION ET CONVERGENCE DES ESTIMATEURS DU MAXIMUM DE VRAISEMBLANCE

I.C.1 - Les divers estimateurs et la condition de convergence

Disposant de diverses formes exacte ou approchées de la log-vraisemblance, on peut définir divers estimateurs du (pseudo)maximum de vraisemblance en maximisant l'une ou l'autre de ces formes. Un tel estimateur est donc solution du problème :

$$(1.12) \quad \text{Max}_{\theta, \sigma} \text{Log } L_T^*(x; \theta, \sigma^2) \quad ,$$

où $L_T^* = L_T$ ou $L_T^{c.c}$ ou L_T^* . Il est dans la suite noté : $(\hat{\theta}_T, \hat{\sigma}_T^2)$.

Pour savoir s'il existe (asymptotiquement) une telle solution et si elle est convergente vers les vraies valeurs (θ_0, σ_0^2) , l'approche usuelle consiste à étudier le comportement asymptotique de $\frac{1}{T} \text{Log } L_T^*(x; \theta, \sigma^2)$ ou de façon équivalente celui de :

$$K_T^*(P/P_0) = \frac{1}{T} \text{Log } L_T^*(x; \theta, \sigma_0^2) - \frac{1}{T} \text{Log } L_T^*(x; \theta, \sigma^2) \quad .$$

Le problème de recherche de l'estimateur du maximum de vraisemblance équivaut à :

$$(1.13) \quad \text{Min}_{\theta, \sigma} K_T^*(P/P_0) \quad .$$

Si la fonction objectif $K_T^*(P/P_0)$ converge p.s uniformément en (θ, σ^2) [ou uniformément sur tout compact] vers une limite $K^*(P/P_0)$, si de plus cette limite admet un minimum unique en (θ_0, σ_0^2) atteint pour (θ_0, σ_0^2) , alors il existe une solution au problème à distance finie (1.13) à partir d'un certain rang aléatoire T et cette solution converge vers la solution (θ_0, σ_0^2) du problème limite :

$$(1.14) \quad \text{Min}_{\theta, \sigma} K^*(P/P_0) \quad .$$

Cette approche due à Jennrich [1969] s'applique en particulier au cas des séries temporelles stationnaires, qui nous intéresse. Sous les hypothèses concernant la décroissance exponentielle des coefficients des développements moyenne-mobile ou autorégressif, les diverses approximations de la log-vrai-

semblance apparaissent asymptotiquement correctes au sens où la limite de $K_T^*(P/P_0)$ ne dépend pas de la forme exacte ou approchée retenue. Dans la suite, nous notons $K(P/P_0)$ cette limite commune.

Nous allons maintenant donner diverses formes équivalentes de cette limite commune, qui n'est autre que le contraste de Kullback entre la pseudo-loi P utilisée pour construire la log-vraisemblance et la vraie loi P_0 .

I.C.2 - Expressions du contraste de Kullback

D'après la discussion précédente, il est clair qu'une définition possible de $K(P/P_0)$ est :

$$(1.15) \quad K(P/P_0) = \underset{T \rightarrow \infty}{p} \lim \left[\frac{1}{T} \text{Log } L_T^*(x; \theta_0, \sigma_0^2) - \frac{1}{T} \text{Log } L_T^*(x; \theta, \sigma^2) \right]$$

D'autres expressions sont déduites des divers formes de L_T^* .

a) Utilisation de l'approximation dans l'espace des fréquences

A partir de la forme (1.11) de la log-vraisemblance, on obtient :

$$\begin{aligned} \tilde{K}_T(P/P_0) &= \frac{1}{T} \left[-\frac{T}{2} \text{Log } 2\pi - \frac{T}{2} \text{Log } \sigma_0^2 - \frac{T}{2\sigma_0^2} E_{\mu} \left(\frac{\hat{T}}{g_{\theta}} \right) \right] \\ &\quad - \frac{1}{T} \left[-\frac{T}{2} \text{Log } 2\pi - \frac{T}{2} \text{Log } \sigma^2 - \frac{T}{2\sigma^2} E_{\mu} \left(\frac{\hat{T}}{g_{\theta}} \right) \right] \\ &= \frac{1}{2} \text{Log } \frac{\sigma^2}{\sigma_0^2} - \frac{1}{2\sigma_0^2} E_{\mu} \left(\frac{\hat{T}}{g_{\theta}} \right) + \frac{1}{2\sigma^2} E_{\mu} \left(\frac{\hat{T}}{g_{\theta}} \right) . \end{aligned}$$

Bien que le périodogramme \hat{I}_T ne converge pas vers la vraie densité spectrale, on a cependant [voir Azencott-Dacunha-Castelle (1984)] la convergence d'une intégrale $E_\mu(g \hat{I}_T)$ vers $E_\mu \left(g f_{\theta, \sigma}^2 \right)$ pour toute fonction g continue sur $]-\pi, \pi[$.

Ce résultat peut être directement appliqué, ce qui donne :

$$\underset{T \rightarrow \infty}{\text{plim}} \tilde{K}_T(P/P_0) = \frac{1}{2} \text{Log} \frac{\sigma^2}{2\sigma_0^2} - \frac{1}{2\sigma_0^2} E_\mu \sigma_0^2 + \frac{1}{2\sigma_0^2} E_\mu \left(\sigma_0^2 \frac{g_\theta}{g_\theta} \right),$$

c'est-à-dire :

$$(1.16) \quad K(P/P_0) = \frac{1}{2} E_\mu \left\{ \text{Log} \frac{\sigma^2}{\sigma_0^2} - 1 + \frac{\sigma_0^2}{\sigma^2} \frac{g_\theta}{g_\theta} \right\}.$$

Remarque 1.17 : Comme pour toute valeur θ , on a :

$E_\mu \text{Log} g_\theta = -\text{Log} 2\pi$, une autre écriture équivalente de ce contraste est

$$K(P/P_0) = \frac{1}{2} E_\mu \left\{ \text{Log} \frac{f_{\theta, \sigma}^2}{f_{\theta, \sigma_0}^2} - 1 + \frac{f_{\theta, \sigma_0}^2}{f_{\theta, \sigma}^2} \right\}.$$

Remarque 1.18 : Dans le cas de deux lois normales centrées $\tilde{P} = N(0, \sigma^2)$

et $\tilde{P}_0 = N(0, \sigma_0^2)$ le contraste de Kullback vaut :

$$K(\tilde{P}/\tilde{P}_0) = \frac{1}{2} E_0 \text{Log} \frac{1(Y)}{1(Y)}$$

$$= \underset{0}{\overset{\approx}{E}} \left\{ \frac{1}{2} \text{Log} \frac{\sigma^2}{\sigma_0^2} - \frac{1}{2\sigma_0^2} Y^2 + \frac{1}{2\sigma^2} Y^2 \right\}$$

$$= \frac{1}{2} \left\{ \text{Log} \frac{\sigma^2}{\sigma_0^2} - 1 + \frac{\sigma_0^2}{\sigma^2} \right\} .$$

Sous l'hypothèse de normalité la mesure stochastique ξ associée au processus est à accroissements normaux ; on a en fait pour toute valeur h :

$$\frac{\xi(\omega+h) - \xi(\omega)}{h} \approx N \left[0, \frac{1}{h} \int_{\omega}^{\omega+h} f(u) du \right] , \text{ où } f \text{ désigne la densité spectrale.}$$

La quantité : $K(\omega) = \frac{1}{2} \left\{ \text{Log} \frac{f_{\theta, \sigma^2}}{f_{\theta_0, \sigma_0^2}} - 1 + \frac{f_{\theta_0, \sigma_0^2}}{f_{\theta, \sigma^2}} \right\}$ s'interprète

donc comme la proximité entre les lois du processus associées aux valeurs (θ, σ^2) et (θ_0, σ_0^2) et correspondant à la fréquence ω . La proximité entre les deux lois du processus apparaît alors d'après (1.17) comme la moyenne des proximités fréquence par fréquence : $K(P/P_0) = \underset{\mu}{E} K(\omega)$.

b) Utilisation des lois conditionnelles

Une autre expression du contraste de Kullback s'obtient à partir de l'écriture en terme de lois conditionnelles. On a :

$$K(P/P_0) = \lim_{T \rightarrow \infty} \left\{ \frac{1}{T} \text{Log} L_T^c(x; \theta, \sigma^2) - \frac{1}{T} \text{Log} L_T^c(x; \theta_0, \sigma_0^2) \right\}$$

$$= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \left[\text{Log} l_{\theta, \sigma^2}(X_t/X_{t-1}, X_{t-2}, \dots) - \text{Log} l_{\theta_0, \sigma_0^2}(X_t/X_{t-1}, \dots) \right] ,$$

en notant 1_o et 1 les densités conditionnelles associées aux valeurs (θ_o, σ_o^2) et (θ, σ^2) du paramètre. Utilisant la stationnarité du processus et les théorèmes ergodiques, on en déduit que :

$$K(P/P_o) = E_o \left[\text{Log} \frac{1_o(X_t/X_{t-1}, X_{t-2}, \dots)}{1(X_t/X_{t-1}, X_{t-2}, \dots)} \right]$$

(1.19)
$$K(P/P_o) = E_o E_o \left[\text{Log} \frac{1_o(X_t/X_{t-1}, X_{t-2}, \dots)}{1(X_t/X_{t-1}, X_{t-2}, \dots)} / X_{t-1}, X_{t-2}, \dots \right]$$

$= E_o K(P/P_o) \quad \text{[par définition]}$

Le contraste apparaît ainsi comme la moyenne des contrastes conditionnels au passé. La formule (1.19) est donc une formule de décomposition du contraste dans le domaine des temps, analogue à celle établie en (1.18) dans le domaine des fréquences.

Finalement remplaçant les densités conditionnelles par leurs expressions, nous avons :

$$K(P/P_o) = E_o \frac{1}{2} \left\{ \text{Log} \frac{\sigma_o^2}{\sigma_o^2} - \frac{1}{2} \left[\frac{1}{\sigma_o} \text{E} \left[(L) X_t \right]^2 \right] + \frac{1}{2} \left[\frac{1}{\sigma_o} \text{E} \left[(L) X_t \right]^2 \right] \right\}$$

(1.20)
$$K(P/P_o) = \frac{1}{2} \left\{ \text{Log} \frac{\sigma_o^2}{\sigma_o^2} - \frac{1}{2} V_o \left[\frac{1}{\sigma_o} \text{E} \left[(L) X_t \right] \right] + \frac{1}{2} V_o \left[\frac{1}{\sigma_o} \text{E} \left[(L) X_t \right] \right] \right\} .$$

Remarque 1.21 : Cette dernière formule peut d'ailleurs servir à retrouver

directement l'écriture dans le domaine des fréquences. La densité spectrale de $\phi_{\theta}(L) X_t$, lorsque les valeurs des paramètres sont (θ_o, σ_o^2) , est en

$$\text{effet donnée par : } f_{\phi_{\theta}(L)X}(\omega) = |\phi_{\theta}(\exp i\omega)|^2 f_{\theta_o, \sigma_o^2}(\omega) = \frac{\sigma_o^2}{2\pi} \frac{f_{\theta_o, \sigma_o^2}(\omega)}{f_{\theta_o, \sigma_o^2}(\omega)}$$

De plus on a :

$$V_{\theta_o}[\phi_{\theta}(L) X_t] = \int_{-\pi}^{\pi} f_{\phi_{\theta}(L)X}(\omega) d\omega = \sigma_o^2 E_{\mu} \left[\frac{f_{\theta_o, \sigma_o^2}}{f_{\theta_o, \sigma_o^2}} \right]$$

Il suffirait alors de remplacer les variances par ce type d'expression pour retrouver l'écriture fréquentielle.

I.C.3 - Existence d'une solution unique au problème limite

Pour vérifier la convergence d'un estimateur du maximum de vraisemblance, il faut notamment regarder si le problème limite $\text{Min}_{\theta, \sigma^2} K(P/P_o)$ admet bien la solution $\theta = \theta_o, \sigma^2 = \sigma_o^2$.

L'application : $x \mapsto -\log x - 1 + x = h(x)$ est continue à valeurs positives et ne s'annule que pour $x = 1$.

$$\text{Comme : } K(P/P_o) = \frac{1}{2} E_{\mu} \left(h \left[\frac{f_{\theta_o, \sigma_o^2}}{f_{\theta, \sigma^2}} \right] \right) \text{ le contraste est toujours positif.}$$

D'autre part et remarquant que le rapport $\frac{f_{\theta_o, \sigma_o^2}}{f_{\theta, \sigma^2}}$ est une fonction continue,

il ne s'annule que si :

$$f_{\theta_0, \sigma_0^2} = f_{\theta, \sigma^2} .$$

Le problème limite admet donc la vraie valeur du paramètre comme unique solution si et seulement si :

$$f_{\theta, \sigma^2} = f_{\theta_0, \sigma_0^2} \iff \theta = \theta_0, \sigma^2 = \sigma_0^2 ,$$

c'est-à-dire si le paramètre est identifiable. Cette condition d'identifiabilité avait été introduite dans les hypothèses.

I.D - VECTEURS DES SCORES

Si la vraie valeur (θ_0, σ_0^2) appartient à l'intérieur de l'espace des valeurs possibles du paramètre, si la log-vraisemblance $\text{Log } L_T^*(x; \theta, \sigma^2)$ est dérivable, il résulte de la convergence de l'estimateur du maximum de vraisemblance que celui-ci satisfait asymptotiquement la condition du premier ordre :

$$(1.22) \frac{\partial \text{Log } L_T^*(x; \hat{\theta}_T, \hat{\sigma}_T^2)}{\partial (\theta', \sigma^2)'} = 0 .$$

Ceci nous conduit à expliciter le vecteur des scores $\frac{\partial \text{Log } L_T^*}{\partial (\theta', \sigma^2)'}$. Il peut

l'être pour les divers choix de la vraisemblance exacte ou approchée.

I.D.1 - Score exact

Commençons par considérer la forme exacte donnée en (1.8). On a :

$$\frac{\partial \text{Log } L_T}{\partial \sigma^2} (x; \theta, \sigma^2) = - \frac{T}{2\sigma^2} + \frac{1}{2\sigma^4} x_T' C_T(\theta)^{-1} x_T ,$$

et

$$(1.23) \quad \frac{\partial \text{Log} L_T}{\partial \theta_k} (x; \theta, \sigma^2) = - \frac{1}{2} \frac{\partial}{\partial \theta_k} [\text{Log det } C_T(\theta)] - \frac{1}{2\sigma^2} \frac{\partial}{\partial \theta_k} (x_T' C_T(\theta)^{-1} x_T)$$

$$= - \frac{1}{2} \text{Tr} \left[C_T(\theta)^{-1} \frac{\partial C_T(\theta)}{\partial \theta_k} \right] + \frac{1}{2\sigma^2} x_T' C_T(\theta)^{-1} \frac{\partial C_T(\theta)}{\partial \theta_k} C_T(\theta)^{-1} x_T,$$

(en utilisant quelques formules classiques de dérivation matricielle [voir Rao (1973)]).

I.D.2 - Score approché déduit de la forme conditionnelle

De la forme conditionnelle (1.9) [qui pourrait être remplacée par (1.10)], on déduit :

$$\frac{\partial \text{Log} L_T^c}{\partial \sigma^2} (x; \theta, \sigma^2) = - \frac{T}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{t=1}^T [\phi_\theta(L) X_t]^2$$

et

$$(1.24) \quad \frac{\partial \text{Log} L_T^c}{\partial \theta} (x; \theta, \sigma^2) = - \frac{1}{\sigma^2} \sum_{t=1}^T \frac{\partial \phi_\theta(L) X_t}{\partial \theta} + \phi_\theta(L) X_t$$

$$= + \frac{1}{\sigma^2} \sum_{t=1}^T \left(\sum_{j=1}^{\infty} \frac{\partial \phi_j(\theta)}{\partial \theta} X_{t-j} \right) \left(X_t - \sum_{j=1}^{\infty} \phi_j(\theta) X_{t-j} \right).$$

Sous cette forme autorégressive infinie, on retrouve les interprétations usuelles en terme de moindres carrés non linéaires.

L'équation de vraisemblance $\frac{\partial \text{Log } L_T^c}{\partial \theta} (x; \hat{\theta}_T, \hat{\sigma}_T^2) = 0$ apparaît comme une condition d'orthogonalité entre les résidus :

$$\hat{\varepsilon}_t = X_t - \sum_{j=1}^{\infty} \varphi_j(\hat{\theta}_T) X_{t-j},$$

et les "variables explicatives" :

$$\sum_{j=1}^{\infty} \frac{\partial \varphi_j(\hat{\theta}_T)}{\partial \theta_k} X_{t-j}, \quad k \text{ variant.}$$

L'autre équation de vraisemblance fournit $\hat{\sigma}_T^2$ à partir de la somme des carrés des résidus :

$$\hat{\sigma}_T^2 = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t^2.$$

I.D.3 - Score approché dans le domaine des fréquences

Finalement, on déduit de (1.11) que :

$$\frac{\partial \text{Log } \tilde{L}_T}{\partial \sigma^2} (x; \theta, \sigma^2) = -\frac{T}{2\sigma^2} + \frac{T}{2\sigma^4} E_{\mu} \left(\frac{\hat{T}_T}{g_{\theta}} \right),$$

$$\text{et } \frac{\partial \text{Log } \tilde{L}_T}{\partial \theta} (x; \theta, \sigma^2) = \frac{T}{2\sigma^2} E_{\mu} \left[\hat{T}_T \frac{\partial g_{\theta}}{\partial \theta} \frac{1}{g_{\theta}^2} \right].$$

Dérivant par rapport à θ , l'égalité de Kolmogorov :

$$\text{Log } 2\pi + E_{\mu} \text{Log } g_{\theta} = 0,$$

on voit immédiatement que cette seconde dérivée peut aussi s'écrire :

$$(1.25) \quad \frac{\partial \text{Log } \hat{L}_T^*}{\partial \theta} (x; \theta, \sigma^2) = \frac{T}{2\sigma^2} E_{\mu} \left[\left(\hat{\Gamma}_T - \sigma^2 g_{\theta} \right) \frac{\partial g_{\theta}}{\partial \theta} \frac{1}{g_{\theta}^2} \right]$$

Ce vecteur de score peut être exprimé en faisant apparaître les covariances empiriques : $\hat{\gamma}_T(h) = \frac{1}{T} \sum_{t=1}^{T-|h|} X_t X_{t+h}$. Celles-ci sont liées au périodogramme par :

$$\hat{\Gamma}_T(\omega) = \frac{1}{2\pi} \sum_{h=-T}^{T-1} \hat{\gamma}_T(h) \exp(ih\omega)$$

On a donc :

$$(1.26) \quad \frac{\partial \text{Log } \hat{L}_T^*}{\partial \theta} (x; \theta, \sigma^2) = - \frac{T}{8\pi^2 \sigma^2} \sum_{h=-T}^{T-1} \hat{\gamma}_T(h) \int_{-\pi}^{\pi} \exp(ih\omega) \frac{\frac{\partial}{\partial \theta} g_{\theta}(\omega)}{g_{\theta}^2(\omega)} d\omega$$

I.E - LOIS ASYMPTOTIQUES ET MATRICE D'INFORMATION DE FISHER

I.E.1 - Loi asymptotique des scores

Dans ce sous paragraphe nous résumons les propriétés asymptotiques des vecteurs des scores évalués en la vraie valeur.

Ces propriétés sont :

i) La convergence P₀ p.s du vecteur $\frac{1}{T} \frac{\partial \text{Log } \hat{L}_T^*}{\partial (\theta', \sigma^2)} (x; \theta_0, \sigma_0^2)$ vers zéro ;

ii) la normalité asymptotique de ce vecteur convenablement normalisé :

$$(1.27) \quad \frac{1}{\sqrt{T}} \frac{\partial \text{Log } \hat{L}_T^*}{\partial (\theta', \sigma^2)} (x; \theta_0, \sigma_0^2) \xrightarrow{d} N[0; J(\theta_0, \sigma_0^2)] ;$$

iii) l'équivalence asymptotique des approches fondées sur les diverses formes

de la vraisemblance. Si L_T^* et L_T^{**} sont deux de ces formes, l'écart :

$$\frac{1}{\sqrt{T}} \frac{\partial \text{Log} L_T^*}{\partial (\theta', \sigma^2)} (x; \theta_0, \sigma_0^2) - \frac{1}{\sqrt{T}} \frac{\partial \text{Log} L_T^{**}}{\partial (\theta', \sigma^2)} (x; \theta_0, \sigma_0^2)$$

est infiniment petit en probabilité.

Comme auparavant les résultats d'équivalence asymptotique résultent essentiellement des hypothèses de décroissance exponentielle des coefficients des développements moyennes mobiles ou autorégressifs. [Voir Whittle (1951), Hannan (1973)]. Dans la suite nous admettons cette équivalence.

Nous l'utiliserons pour obtenir selon le choix de L_T^* des expressions différentes de la matrice limite $J(\theta_0, \sigma_0^2)$, qui s'interprète comme une matrice d'information de Fisher.

a) Utilisation du score dans le domaine des fréquences

Les propriétés asymptotiques du score peuvent être directement déduites de celles du périodogramme. On a en effet sous nos hypothèses :

PROPRIÉTÉ 1.28 :

Pour toute fonction φ complexe multivariée continue bornée sur

$]-\pi, \pi[$:

i)
$$\int_{-\pi}^{\pi} \hat{\Gamma}_T(\omega) \varphi(\omega) d\omega \xrightarrow{P_{\theta_0, \sigma_0^2}} \int_{-\pi}^{\pi} f_{\theta_0, \sigma_0^2}^2(\omega) \varphi(\omega) d\omega ;$$

ii) De plus si ε est gaussien :

$$\sqrt{T} E_{\mu} \left\{ \left[\hat{\Gamma}_T - f_{\theta_0, \sigma_0^2}^2 \right] \varphi \right\} \xrightarrow{d} N \left[0, 2 E_{\mu} \left[\varphi \bar{\varphi}' f_{\theta_0, \sigma_0^2}^2 \right] \right] .$$

Voir [Anderson (1971)].

De l'expression du score dans le domaine des fréquences (voir I.D.3), on déduit :

$$\begin{bmatrix} \frac{\partial \text{Log} \tilde{L}_T}{\partial \sigma^2} (x; \theta, \sigma^2) \\ \frac{\partial \text{Log} \tilde{L}_T}{\partial \theta} (x; \theta, \sigma^2) \end{bmatrix} = \frac{T}{2\sigma^2} \begin{bmatrix} E_{\mu} \left[\frac{\hat{I}_T - f}{\theta, \sigma^2} \right] \\ E_{\mu} \left[\left(\hat{I}_T - f \right) \frac{\partial g_{\theta}}{\partial \theta} \frac{1}{g_{\theta}} \right] \end{bmatrix}$$

Il s'agit d'une forme, où apparaissent des intégrales de fonctions continues ; ceci permet l'application directe de la propriété (1.28).

On a :

$$(1.29) \quad \frac{1}{T} \begin{bmatrix} \frac{\partial \text{Log} \tilde{L}_T}{\partial \sigma^2} (x; \theta, \sigma^2) \\ \frac{\partial \text{Log} \tilde{L}_T}{\partial \theta} (x; \theta, \sigma^2) \end{bmatrix} \xrightarrow{P_{\theta} \text{ ps}} 0$$

et

$$(1.30) \quad \frac{1}{\sqrt{T}} \begin{bmatrix} \frac{\partial \text{Log} L_T}{\partial \sigma^2} (x; \theta_0, \sigma_0^2) \\ \frac{\partial \text{Log} L_T}{\partial \theta} (x; \theta_0, \sigma_0^2) \end{bmatrix} \xrightarrow{d} N \left[0, J(\theta_0, \sigma_0^2) \right]$$

$$\text{avec : } J(\theta_0, \sigma_0^2) = \frac{2}{4\sigma_0^4} \begin{bmatrix} E_{\mu} \left(\frac{f^2}{\theta_0, \sigma_0^2} \right) & E_{\mu} \left(\frac{\partial \text{Log } g_{\theta_0}}{\partial \theta'} \right) \\ \sigma_0^2 E_{\mu} \left(\frac{\partial \text{Log } g_{\theta_0}}{\partial \theta} \right) & \sigma_0^4 E_{\mu} \left(\frac{\partial \text{Log } g_{\theta_0}}{\partial \theta} \frac{\partial \text{Log } g_{\theta_0}}{\partial \theta'} \right) \end{bmatrix}$$

Utilisant l'égalité de Kolmogorov, qui entraîne $E_{\mu} \frac{\partial \text{Log } g_{\theta_0}}{\partial \theta} = 0$, on voit que l'information de Fisher est :

$$(1.31) \quad J(\theta_0, \sigma_0^2) = \begin{bmatrix} \frac{1}{4\sigma_0^4} & 0 \\ 0 & \frac{1}{2} E_{\mu} \left(\frac{\partial \text{Log } g_{\theta_0}}{\partial \theta} \frac{\partial \text{Log } g_{\theta_0}}{\partial \theta'} \right) \end{bmatrix}$$

b) Utilisation du score conditionnel

Utilisons maintenant la forme conditionnelle :

$$\frac{\partial \text{Log} L_T^c}{\partial (\theta', \sigma^2)'} (x; \theta, \sigma^2) = \sum_{t=1}^T \frac{\partial}{\partial (\theta', \sigma^2)'} \text{Log} 1 (X_t / X_{t-1}, \dots; \theta, \sigma^2) .$$

Les variables $Z_t = \frac{\partial}{\partial (\theta', \sigma^2)'} \text{Log} 1 (X_t / X_{t-1}, \dots; \theta, \sigma^2)$ intervenant dans

la sommation sont telles que :

$$E_0 [Z_t / X_{t-1}, X_{t-2}, \dots] = E_0 \left[\frac{\partial}{\partial (\theta', \sigma^2)'} \text{Log} 1 (X_t / X_{t-1}, \dots; \theta, \sigma^2) / X_{t-1}, X_{t-2}, \dots \right] = 0 ,$$

et donc que $E_0 [Z_t / Z_{t-1}, Z_{t-2}, \dots] = 0 .$

La suite $(Z_t, t = 1 \dots T \dots)$ constitue donc une différence de martingale centrée. Utilisant en plus la stationnarité de cette suite $(Z_t, t = 1 \dots T \dots)$, on déduit directement que :

$$\frac{1}{T} \frac{\partial \text{Log} L_T^c}{\partial (\theta', \sigma^2)'} (x; \theta, \sigma^2) \xrightarrow{P, \text{ps}} E_0 \left[\frac{\partial}{\partial (\theta', \sigma^2)'} \text{Log} 1 (X_t / X_{t-1}, \dots; \theta, \sigma^2) \right] = 0 ,$$

et :

$$\frac{1}{\sqrt{T}} \frac{\partial \text{Log} L_T^c}{\partial (\theta', \sigma^2)'} (x; \theta, \sigma^2) \xrightarrow{d} N [0 ; J(\theta, \sigma^2)] ,$$

$$\text{avec : } J(\theta, \sigma^2) = V_0 \left[\frac{\partial}{\partial (\theta', \sigma^2)'} \text{Log} 1 (X_t / X_{t-1}, \dots; \theta, \sigma^2) \right] .$$

Développant cette dernière formule en utilisant I.D.2, on obtient :

$$J(\theta, \sigma^2) = V \begin{bmatrix} -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4} \epsilon_t^2 \\ \frac{1}{2\sigma^2} \epsilon_t \cdot \sum_{j=1}^{\infty} \frac{\partial \varphi_j(\theta)}{\partial \theta} X_{t-j} \end{bmatrix}$$

$$J(\theta, \sigma^2) = \begin{bmatrix} \frac{1}{2\sigma^4} & 0 \\ 0 & \frac{1}{2\sigma^2} V \left[\frac{\partial \varphi(\theta)}{\partial \theta} X \right] \end{bmatrix}$$

On voit clairement dans cette approche où est intervenue l'hypothèse ϵ gaussien. Elle sert essentiellement à écrire $E \epsilon_t^4 = 3 \sigma^4$ et $E \epsilon_t^3 = 0$.
 Sous ces deux seules conditions les résultats sont encore valables.

Remarque 1.32 : Une autre forme de la matrice d'information peut facilement être déduite de cette approche conditionnelle :

$$\begin{aligned} J(\theta, \sigma^2) &= V \left[\frac{\partial}{\partial(\theta', \sigma^2)} \text{Log} l(X_t / X_{t-1}, \dots; \theta, \sigma^2) \right] \\ &= E \left[\frac{\partial}{\partial(\theta', \sigma^2)} \text{Log} l(X_t / X_{t-1}, \dots; \theta, \sigma^2) \frac{\partial}{\partial(\theta', \sigma^2)} \text{Log} l(X_t / X_{t-1}, \dots; \theta, \sigma^2) \right] \\ &= E E \left[\frac{\partial}{\partial(\theta', \sigma^2)} \text{Log} l(X_t / X_{t-1}, \dots; \theta, \sigma^2) \right. \\ &\quad \left. \frac{\partial}{\partial(\theta', \sigma^2)} \text{Log} l(X_t / X_{t-1}, \dots; \theta, \sigma^2) / X_{t-1}, X_{t-2}, \dots \right] \end{aligned}$$

$$\begin{aligned}
 &= E_0 \left[- \frac{\partial^2}{\partial(\theta', \sigma^2)' \partial(\theta', \sigma^2)} \text{Log} l(X_t / X_{t-i}; \theta_0, \sigma_0^2) \right] \quad (\text{d'après une} \\
 & \hspace{15em} \text{propriété classique}) \\
 &= p \lim_T \left[- \frac{1}{T} \sum_{t=1}^T \frac{\partial^2 \text{Log} l(X_t / X_{t-i}; \theta_0, \sigma_0^2)}{\partial(\theta', \sigma^2)' \partial(\theta', \sigma^2)} \right] \\
 &= p \lim_T - \frac{1}{T} \frac{\partial^2 \text{Log} L_T^c(x; \theta_0, \sigma_0^2)}{\partial(\theta', \sigma^2)' \partial(\theta', \sigma^2)} .
 \end{aligned}$$

I.E.2 - Normalité asymptotique d'un estimateur du maximum de vraisemblance

La loi asymptotique de $(\hat{\theta}_T, \hat{\sigma}_T^2)$ se déduit de la manière usuelle de celle du score. Effectuant un développement limité en probabilité de la condition du premier ordre, on obtient :

$$\begin{aligned}
 &\frac{1}{\sqrt{T}} \frac{\partial \text{Log} L_T^*(x; \hat{\theta}_T, \hat{\sigma}_T^2)}{\partial(\theta', \sigma^2)'} = 0 \\
 \Leftrightarrow &\frac{1}{\sqrt{T}} \frac{\partial \text{Log} L_T^*(x; \theta_0, \sigma_0^2)}{\partial(\theta', \sigma^2)'} + \frac{1}{T} \frac{\partial^2 \text{Log} L_T^*(x; \theta_0, \sigma_0^2)}{\partial(\theta', \sigma^2)' \partial(\theta', \sigma^2)} \sqrt{T} \begin{pmatrix} \hat{\theta}_T - \theta_0 \\ \hat{\sigma}_T^2 - \sigma_0^2 \end{pmatrix} = o_P(1) ,
 \end{aligned}$$

$$\sqrt{T} \begin{pmatrix} \hat{\theta}_T - \theta_0 \\ \hat{\sigma}_T^2 - \sigma_0^2 \end{pmatrix} = \left[- p \lim_T \frac{1}{T} \frac{\partial^2 \text{Log} L_T^*(x; \theta_0, \sigma_0^2)}{\partial(\theta', \sigma^2)' \partial(\theta', \sigma^2)} \right]^{-1} \frac{1}{\sqrt{T}} \frac{\partial \text{Log} L_T^*(x; \theta_0, \sigma_0^2)}{\partial(\theta', \sigma^2)'} + o_P(1)$$

$$\sqrt{T} \begin{pmatrix} \hat{\theta}_T - \theta_0 \\ \hat{\sigma}_T^2 - \sigma_0^2 \end{pmatrix} = J(\theta_0, \sigma_0^2)^{-1} \frac{1}{\sqrt{T}} \frac{\partial \text{Log } L_T^*(x; \theta_0, \sigma_0^2)}{\partial (\theta', \sigma^2)'} + o_p(1) .$$

On en déduit alors la propriété classique :

PROPRIÉTÉ 1.33 :

L'estimateur du maximum de vraisemblance est asymptotiquement normal :

$$\sqrt{T} \begin{pmatrix} \hat{\theta}_T - \theta_0 \\ \hat{\sigma}_T^2 - \sigma_0^2 \end{pmatrix} \xrightarrow{d} Z ,$$

avec $Z \approx N \left[0, J(\theta_0, \sigma_0^2)^{-1} \right] .$

On déduit de la forme particulière (1.31) de la matrice d'information de Fisher le corollaire :

Corollaire 1.34 :

- i) Les estimateurs $\hat{\theta}_T$ et $\hat{\sigma}_T^2$ sont asymptotiquement indépendants.
- ii) La précision asymptotique de $\hat{\theta}_T$ est indépendante de la valeur de la variance σ_0^2 .

Il résulte des formules établies pour la matrice d'information que :

$$\begin{aligned}
 (1.35) \quad V_{as} \left[\sqrt{T} (\hat{\theta}_T - \theta_0) \right] &= 2 \left(E_{\mu} \left[\begin{array}{cc} \frac{\partial \text{Log } g_{\theta}}{\partial \theta} & \frac{\partial \text{Log } g_{\theta}}{\partial \theta'} \end{array} \right] \right)^{-1} \\
 &= \sigma_0^2 \left[V_0 \left(\begin{array}{c} \frac{\partial \phi_{\theta_0}^{(L)}}{\partial \theta} \\ X_t \end{array} \right) \right]^{-1} \\
 &= \sigma_0^2 \left(V_0 \left[\begin{array}{c} \frac{\partial \text{Log } \phi_{\theta_0}^{(L)}}{\partial \theta} \\ \epsilon_t \end{array} \right] \right)^{-1}
 \end{aligned}$$

II - ESTIMATION PAR LE MAXIMUM DE VRAISEMBLANCE D'UN PROCESSUS MADE(1)

II.A - IDENTIFIABILITE

Nous allons appliquer les résultats du chapitre précédent au cas d'un processus MADE (1) :

$$(2.1) \quad (1 - \alpha L)^d X_t = \epsilon_t,$$

où $\epsilon = (\epsilon_t, t \in Z)$ est un bruit blanc de variance $\sigma^2 > 0$, et où $|\alpha| < 1$ et $d \in R$.

Nous nous restreignons à ce cas qui sera utilisé ultérieurement au niveau des tests d'hypothèses. Les résultats se généralisent cependant facilement au cas d'un processus MADE quelconque avec $|\alpha_k| < 1$; $|\beta_k| < 1$ (voir § IV).

Pour étudier l'identifiabilité du couple (α_0, d_0) , il faut regarder si :

$$(1 - \alpha_0 L)^{d_0} = (1 - \alpha L)^d \implies \alpha = \alpha_0 \quad \text{et} \quad d = d_0.$$

On montre aisément le résultat ci-dessous.

PROPRIETE 2.2 :

(α, d) est identifiable si et seulement si $\alpha d \neq 0$.

Toutes les valeurs $(\alpha, 0)$ et $(0, d)$ non identifiables correspondent au cas limite du bruit blanc.

II.B - CALCUL DU SCORE ET MATRICE D'INFORMATION

II.B.1 - Expression du score dans le domaine des temps

Evaluons les composantes du vecteur des scores correspondant aux paramètres α, d , à partir de la forme conditionnelle (1.24). On a :

$$\begin{aligned} \frac{\partial \text{Log } L_T^c}{\partial \alpha} (x; \alpha, d, \sigma^2) &= - \frac{1}{\sigma^2} \sum_{t=1}^T \frac{\partial (1-\alpha L)^d}{\partial \alpha} X_t (1-\alpha L)^d X_t \\ &= - \frac{1}{\sigma^2} \sum_{t=1}^T d(1-\alpha L)^{d-1} L X_t (1-\alpha L)^d X_t \\ &= - \frac{d}{\sigma^2} \sum_{t=1}^T \epsilon_t (1-\alpha L)^{-1} \epsilon_{t-1} . \end{aligned}$$

$$\begin{aligned} \frac{\partial \text{Log } L_T^c}{\partial d} (x; \alpha, d, \sigma^2) &= - \frac{1}{\sigma^2} \sum_{t=1}^T \frac{\partial (1-\alpha L)^d}{\partial d} X_t (1-\alpha L)^d X_t \\ &= - \frac{1}{\sigma^2} \sum_{t=1}^T \left[(1-\alpha L)^d \text{Log } (1-\alpha L) \right] X_t (1-\alpha L)^d X_t \end{aligned}$$

$$= - \frac{1}{\sigma^2} \sum_{t=1}^T \epsilon_t \text{Log} (1-\alpha L) \epsilon_t$$

Les deux composantes du vecteur des scores apparaissent donc comme des produits scalaires entre l'innovation présente ϵ_t et des

combinaisons linéaires : $\frac{d}{\sigma^2} (1-\alpha L)^{-1} \epsilon_{t-1}$ et $-\frac{1}{\sigma^2} \text{Log} (1-\alpha L) \epsilon_t$ des innovations passées.

II.B.2 - Eléments de la matrice d'information

Nous notons : $\tilde{J} = \begin{bmatrix} J_{\alpha\alpha} & J_{\alpha d} \\ J_{d\alpha} & J_{dd} \end{bmatrix}$ les éléments de la matrice d'information

associés aux paramètres α , d . Ils sont fournis par :

$$\begin{aligned} J_{\alpha\alpha} &= E \left\{ \frac{d^2}{\sigma^4} \epsilon_t^2 \left[(1 - \alpha L)^{-1} \epsilon_{t-1} \right]^2 \right\} \\ &= \frac{d^2}{\sigma^4} E \epsilon_t^2 E \left[(1 - \alpha L)^{-1} \epsilon_{t-1} \right]^2 \quad (\text{d'après l'indépendance du bruit}) \\ &= \frac{d^2}{\sigma^2} V \left[(1 - \alpha L)^{-1} \epsilon_{t-1} \right] = \frac{d^2}{\sigma^2} V \left(\sum_{j=0}^{\infty} \alpha^j \epsilon_{t-1-j} \right) \\ &= \frac{d^2}{\sigma^2} \sigma^2 \sum_{j=0}^{\infty} \alpha^{2j} = \frac{d^2}{1 - \alpha^2} \end{aligned}$$

$$\begin{aligned} J_{\alpha d} &= E \left\{ \frac{d}{\sigma^2} \epsilon_t \left[(1 - \alpha L)^{-1} \epsilon_{t-1} \right] \left(-\frac{1}{\sigma^2} \right) \epsilon_t \left[\text{Log} (1-\alpha L) \epsilon_t \right] \right\} \\ &= + \frac{d}{\sigma^4} E \epsilon_t^2 E \left[(1 - \alpha L)^{-1} \epsilon_{t-1} \left[-\text{Log} (1-\alpha L) \epsilon_t \right] \right] \end{aligned}$$

$$= \frac{d}{\sigma^2} E \left(\sum_{i=0}^{\infty} \alpha^i \epsilon_{t-i-1} \sum_{j=1}^{\infty} \frac{\alpha^j}{j} \epsilon_{t-j} \right)$$

$$= d \sum_{j=1}^{\infty} \frac{\alpha^{2j-1}}{j} = -\frac{d}{\alpha} \text{Log} (1 - \alpha^2) .$$

$$J_{dd} = E \left\{ \frac{1}{\sigma^2} \epsilon_t^2 \left[\text{Log} (1 - \alpha L) \epsilon_t \right]^2 \right\}$$

$$= \frac{1}{\sigma^2} V \left(\sum_{j=1}^{\infty} \frac{\alpha^j}{j} \epsilon_{t-j} \right) = \sum_{j=1}^{\infty} \frac{\alpha^{2j}}{j^2} = c(\alpha) , \text{ par définition.}$$

Une expression de la matrice d'information est donc :

$$(2.3) \quad J = \begin{bmatrix} \frac{d^2}{1 - \alpha^2} & -\frac{d}{\alpha} \text{Log} (1 - \alpha^2) \\ -\frac{d}{\alpha} \text{Log} (1 - \alpha^2) & \sum_{j=1}^{\infty} \frac{\alpha^{2j}}{j^2} \end{bmatrix}$$

II.B.3 - Expression de la matrice d'information dans le domaine des fréquences

Partons maintenant de la formule (1.31) . On a l'expression de la densité

$$\text{spectrale : } f_{\alpha, d, \sigma}^2(\omega) = \frac{\sigma^2}{2\pi} [1 + \alpha^2 - 2 \alpha \cos \omega]^{-d} = \sigma^2 g_{\alpha, d}(\omega) .$$

On en déduit :

$$\text{Log } g_{\alpha, d}(\omega) = -\text{Log } 2 \pi - d \text{Log} [1 + \alpha^2 - 2 \alpha \cos \omega] ,$$

et les dérivées correspondantes sont données par :

$$\left\{ \begin{array}{l} \frac{\partial \text{Log } g_{\alpha, d}(\omega)}{\partial \alpha} = -d \frac{2\alpha - 2\cos\omega}{1 + \alpha^2 - 2\alpha\cos\omega} , \\ \frac{\partial \text{Log } g_{\alpha, d}(\omega)}{\partial d} = -\text{Log} [1 + \alpha^2 - 2\alpha\cos\omega] , \end{array} \right.$$

Dans le domaine des fréquences la matrice J s'écrit donc :

$$(2.4) \quad \tilde{J} = \frac{1}{2} \left[\begin{array}{cc} d^2 E_{\mu} \left\{ \frac{(2\alpha - 2\cos\omega)^2}{(1 + \alpha^2 - 2\alpha\cos\omega)^2} \right\} & d E_{\mu} \left\{ \frac{(2\alpha - 2\cos\omega)}{(1 + \alpha^2 - 2\alpha\cos\omega)} \text{Log}(1 + \alpha^2 - 2\alpha\cos\omega) \right\} \\ d E_{\mu} \left\{ \frac{(2\alpha - 2\cos\omega)}{(1 + \alpha^2 - 2\alpha\cos\omega)} \text{Log}(1 + \alpha^2 - 2\alpha\cos\omega) \right\} & E_{\mu} \left\{ \text{Log}^2(1 + \alpha^2 - 2\alpha\cos\omega) \right\} \end{array} \right]$$

On peut facilement comparer de façon directe les deux expressions de \tilde{J} (voir Annexe 1 J).

II.C - ESTIMATEURS DU MAXIMUM DE VRAISEMBLANCE

II.C.1 - Equations de vraisemblance

Les estimateurs du maximum de vraisemblance $\hat{\alpha}_T$ et \hat{d}_T sont obtenus comme solution de :

$$\left\{ \begin{array}{l} \sum_{t=1}^T \hat{\epsilon}_t (1 - \hat{\alpha}_T L)^{-1} \hat{\epsilon}_{t-1} = 0 , \\ \sum_{t=1}^T \hat{\epsilon}_t \text{Log}(1 - \hat{\alpha}_T L) \hat{\epsilon}_t = 0 , \end{array} \right.$$

$$\text{avec : } \hat{e}_t = (1 - \hat{\alpha}_T L)^{\hat{d}_T} \tilde{X}_t ,$$

$$\text{où } \tilde{X}_t = \begin{cases} X_t & , \quad \text{si } 1 \leq t \leq T , \\ 0 & , \quad \text{sinon .} \end{cases}$$

Il s'agit évidemment d'un système non linéaire en $(\hat{\alpha}_T, \hat{d}_T)$ qui ne peut être résolu que numériquement.

II.C.2 - Identifiabilité

On peut retrouver facilement la condition d'identifiabilité par l'étude de l'inversibilité de la matrice d'information \tilde{J} . On a :

$$(2.5) \det \tilde{J} = d^2 \sum_{K=2}^{\infty} \alpha^{2K} \left[\sum_{n=0}^{K-1} \left[\frac{1}{(n+1)^2} - \frac{2}{(K+1)(n+1)} \right] \right] ,$$

(Voir Annexe 2).

On vérifie alors que les termes $\sum_{n=0}^{K-1} \left[\frac{1}{(n+1)^2} - \frac{2}{(K+1)(n+1)} \right]$, $K = 2, 3, \dots$

sont tous strictement positifs (voir Annexe 2).

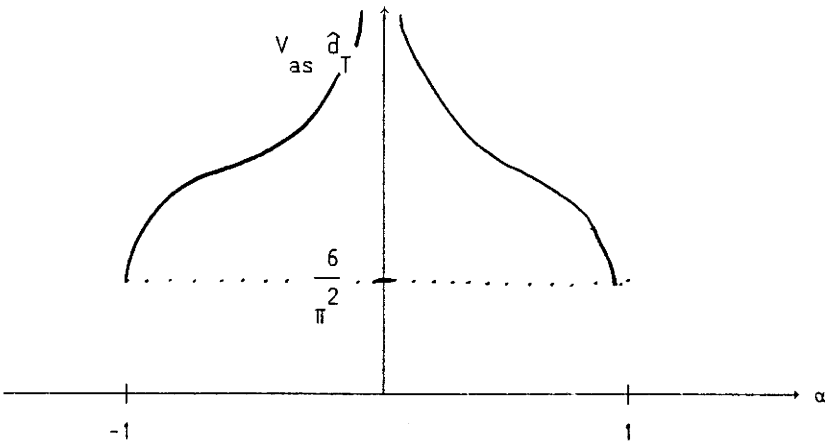
Il en résulte que le déterminant ne peut être nul, que si l'un des deux paramètres d ou α est nul, ce qui correspond bien à la non identifiabilité.

II.C.3 - Evolution des variances asymptotiques de d et de α

La variance asymptotique de l'estimateur \hat{d}_T est donnée par :

$$(2.6) \quad V_{as} \sqrt{T} (\hat{\alpha}_T - d) = \frac{\alpha^2}{\alpha^2 \sum_{j=1}^{\infty} \frac{\alpha^{2j}}{j^2} - (1-\alpha^2) \text{Log}^2(1-\alpha^2)} \quad (\text{voir Annexe 2}).$$

Elle ne dépend que du paramètre α et on voit facilement que son graphe à la forme ci-dessous :



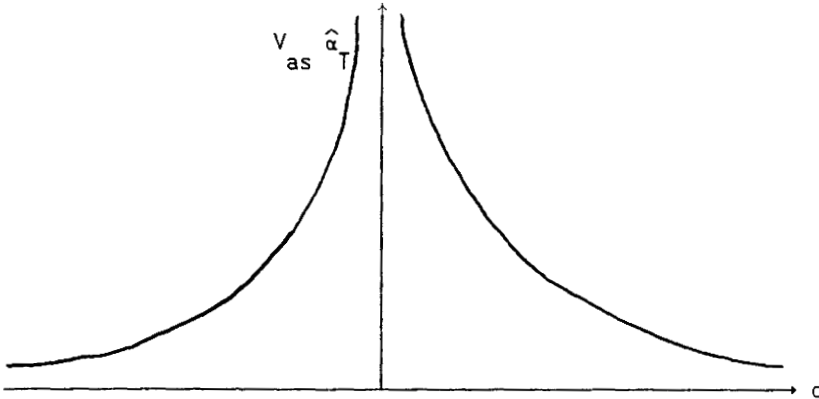
Ainsi plus le paramètre α est proche de 1 en module, plus l'estimateur de d est précis, ce qui est bien compatible avec les résultats discutés dans le chapitre sur les propriétés des moments empiriques de processus MADE.

La variance asymptotique de l'estimateur $\hat{\alpha}_T$ est :

$$(2.7) \quad V_{as} \left[\sqrt{T} (\hat{\alpha}_T - \alpha) \right] = \frac{1}{d^2} \frac{\sum_{j=1}^{\infty} \frac{\alpha^{2j}}{j^2}}{\frac{1}{1-\alpha^2} \sum_{j=1}^{\infty} \frac{\alpha^{2j}}{j^2} - \frac{1}{\alpha^2} \text{Log}^2(1-\alpha^2)}$$

Elle dépend des deux paramètres α et d .

Sa variation par rapport à d à α fixé est :



L'étude de sa variation en fonction de α^2 n'a pas été possible sous forme analytique.

II.C.4 - Corrélation asymptotique entre les deux estimateurs

Cette corrélation est :

$$\begin{aligned} \text{Corr}_{\text{as}} \left[\sqrt{T} (\hat{d}_T - d), \sqrt{T} (\hat{\alpha}_T - \alpha) \right] \\ = \frac{\frac{d}{\alpha} \text{Log} (1-\alpha^2)}{\sqrt{\frac{d^2}{1-\alpha^2}} \sqrt{\sum_{j=1}^{\infty} \frac{\alpha^{2j}}{j^2}}} = \text{Sng}(d) \frac{\sqrt{1-\alpha^2} \text{Log} (1-\alpha^2)}{\alpha \sqrt{\sum_{j=1}^{\infty} \frac{\alpha^{2j}}{j^2}}} \end{aligned}$$

Elle ne dépend de d qu'au travers de son signe. Ceci montre de façon claire l'impact différent d'un modèle MADE "autoregressif" ($d > 0$) ou "moyenne mobile" ($d < 0$).

On remarque en plus que les deux estimateurs deviennent asymptotiquement non

corrélés, lorsque α^2 tend vers 1 .

III - ESTIMATION PAR LA METHODE DES MOMENTS

III.A - PRINCIPE DE LA METHODE

Considérons un processus MADE(1) défini par :

$$(1 - \alpha L)^d X_t = \varepsilon_t \quad , \quad \text{où les } \varepsilon \text{ sont gaussiens, de variance } \sigma^2 .$$

Nous savons [voir Gonçalves (1987)] que les autocovariances du processus satisfont la relation de récurrence :

$$(3.1) \quad (-d+h+1) \gamma(h+1) - \frac{1+\alpha^2}{\alpha} h \gamma(h) + (d+h-1) \gamma(h-1) = 0 \quad , \quad \forall h \in \mathbb{Z} .$$

On peut alors exprimer α et d en fonction des autocovariances en écrivant cette relation pour deux valeurs successives de h . On a :

$$\begin{cases} (-d+h+1) \gamma(h+1) - u h \gamma(h) + (d+h-1) \gamma(h-1) = 0 \quad , \\ (-d+h+2) \gamma(h+2) - u (h+1) \gamma(h+1) + (d+h) \gamma(h) = 0 \quad , \end{cases}$$

$$\text{avec } u = \frac{1+\alpha^2}{\alpha} .$$

Ce système d'équations est linéaire en u et d :

$$\begin{bmatrix} -\gamma(h-1) + \gamma(h+1) & h \gamma(h) \\ -\gamma(h) + \gamma(h+2) & (h+1) \gamma(h+1) \end{bmatrix} \begin{bmatrix} d \\ u \end{bmatrix} = \begin{bmatrix} (h+1) \gamma(h+1) + (h-1) \gamma(h-1) \\ (h+2) \gamma(h+2) + h \gamma(h) \end{bmatrix} .$$

ou bien si la matrice du membre de gauche est inversible :

$$\begin{bmatrix} d \\ u \end{bmatrix} = \begin{bmatrix} \gamma(h+1) - \gamma(h-1) & h \gamma(h) \\ \gamma(h+2) - \gamma(h) & (h+1) \gamma(h+1) \end{bmatrix}^{-1} \begin{bmatrix} (h+1) \gamma(h+1) + (h-1) \gamma(h-1) \\ (h+2) \gamma(h+2) + h \gamma(h) \end{bmatrix}$$

Comme on dispose d'estimateurs convergents : $\hat{\gamma}_T(h) = \frac{1}{T} \sum_{t=1}^{T-h} X_t X_{t+h}$, des autocovariances, on en déduit directement des estimateurs convergents des paramètres d'intérêt d et u . Ceux-ci sont définis par :

$$(3.2) \begin{bmatrix} \tilde{d}_T(h) \\ \tilde{u}_T(h) \end{bmatrix} = \begin{bmatrix} \hat{\gamma}_T(h+1) - \hat{\gamma}_T(h-1) & h \hat{\gamma}_T(h) \\ \hat{\gamma}_T(h+2) - \hat{\gamma}_T(h) & (h+1) \hat{\gamma}_T(h+1) \end{bmatrix}^{-1} \begin{bmatrix} (h+1) \hat{\gamma}_T(h+1) + (h-1) \hat{\gamma}_T(h-1) \\ (h+2) \hat{\gamma}_T(h+2) + h \hat{\gamma}_T(h) \end{bmatrix}$$

Les estimateurs $(\hat{\gamma}_T(h), h \in \mathbb{N}^*)$ étant sous les conditions de régularité classiques asymptotiquement normaux, il en est de même des estimateurs $[\tilde{d}_T(h), \tilde{u}_T(h)]$. La variance asymptotique de ceux-ci peut être déterminée en utilisant la formule de Slutsky. En effet les estimateurs $\tilde{d}_T(h)$ et $\tilde{u}_T(h)$ sont des fonctions différentiables de $\hat{\gamma}_T(h-1), \hat{\gamma}_T(h), \hat{\gamma}_T(h+1), \hat{\gamma}_T(h+2)$:

$$\begin{bmatrix} \tilde{d}_T(h) \\ \tilde{u}_T(h) \end{bmatrix} = k[\hat{\gamma}_T(h-1), \hat{\gamma}_T(h), \hat{\gamma}_T(h+1), \hat{\gamma}_T(h+2)]$$

Notant Δ la matrice jacobienne de la fonction k évaluée en $\gamma(h-1), \gamma(h), \gamma(h+1), \gamma(h+2)$, on a :

$$V_{as} \begin{bmatrix} \sqrt{T} \begin{bmatrix} \tilde{d}_T(h) - d \\ \tilde{u}_T(h) - u \end{bmatrix} \end{bmatrix} = \Delta V_{as} \begin{bmatrix} \sqrt{T} \begin{bmatrix} \hat{\gamma}_T(h-1) - \gamma(h-1) \\ \hat{\gamma}_T(h) - \gamma(h) \\ \hat{\gamma}_T(h+1) - \gamma(h+1) \\ \hat{\gamma}_T(h+2) - \gamma(h+2) \end{bmatrix} \end{bmatrix} \Delta'$$

Le calcul explicite de la matrice jacobienne et son produit avec la matrice de variance asymptotique des autovariances empiriques ne conduit cependant à aucune simplification analytique.

Des estimateurs de d et α sont facilement déduits des précédents. Il suffit de choisir pour estimateur de α la seule solution $\tilde{\alpha}_T(h)$ de module strictement inférieur à 1 de l'équation :

$$\tilde{u}_T(h) = \tilde{\alpha}_T(h) + \frac{1}{\tilde{\alpha}_T(h)} .$$

Le couple $\left[\tilde{d}_T(h), \tilde{\alpha}_T(h) \right]'$ est asymptotiquement normal, et sa précision asymptotique se déduit facilement de la formule de Slutsky.

IV - ESTIMATION PAR LE MAXIMUM DE VRAISEMBLANCE D'UN PROCESSUS MADE

QUELCONQUE

Les résultats concernant les estimateurs du maximum de vraisemblance peuvent être étendus à un processus MADE quelconque défini par :

$$X_t = \frac{\prod_{k=1}^{K_0} (1 - \alpha_k L)^{\delta_k}}{\prod_{k=1}^{K_1} (1 - \beta_k L)^{d_k}} \varepsilon_t .$$

Nous les décrivons sans entrer dans les détails des calculs.

IV.A - CONDITIONS D'IDENTIFIABILITE

Le vecteur $\theta = (\alpha_1, \dots, \alpha_{K_0}, \delta_1, \dots, \delta_{K_0}, \beta_1, \dots, \beta_{K_1}, d_1, \dots, d_{K_1})'$ est identifiable si et seulement si :

- a) $\alpha_i \delta_i \neq 0$, $i = 1, \dots, K_0$,
- b) $\beta_j d_j \neq 0$, $j = 1, \dots, K_1$,
- c) $\alpha_1, \dots, \alpha_{K_0}$, $\beta_1, \dots, \beta_{K_1}$ sont tous différents.

IV.B - LA LOG VRAISEMBLANCE CONDITIONNELLE

Elle est donnée par :

$$\text{Log} L_T^c(x; \theta, \sigma^2) = -\frac{T}{2} \text{Log } 2\pi - \frac{T}{2} \text{Log } \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^T \left[\frac{\prod_{k=1}^{K_1} (1 - \beta_k L)^{d_k}}{\prod_{k=1}^{K_0} (1 - \alpha_k L)^{\delta_k}} X_t \right]^2$$

Les composantes du vecteur des scores sont :

$$\frac{\partial \text{Log } L_T^c}{\partial \beta_j} (x; \theta, \sigma^2) = \frac{d_j}{\sigma^2} \sum_{t=1}^T \epsilon_t (1 - \beta_j L)^{-1} \epsilon_{t-1}$$

$$\frac{\partial \text{Log } L_T^c}{\partial \alpha_j} (x; \theta, \sigma^2) = -\frac{\delta_j}{\sigma^2} \sum_{t=1}^T \epsilon_t (1 - \alpha_j L)^{-1} \epsilon_{t-1}$$

$$\frac{\partial \text{Log } L_T^c}{\partial d_j} (x; \theta, \sigma^2) = -\frac{1}{\sigma^2} \sum_{t=1}^T \epsilon_t \left[\text{Log} (1 - \beta_j L) \epsilon_t \right]$$

$$\frac{\partial \text{Log } L_T^c}{\partial \delta_j} (x; \theta, \sigma^2) = \frac{1}{\sigma^2} \sum_{t=1}^T \epsilon_t \left[\text{Log} (1 - \alpha_j L) \epsilon_t \right]$$

IV.C - LA MATRICE D'INFORMATION

Nous décomposons celle-ci en blocs :

$\alpha_1, \dots, \alpha_{K_0}, \delta_1, \dots, \delta_{K_0}, \beta_1, \dots, \beta_{K_1}, d_1, \dots, d_{K_1}$

$$J = \begin{array}{|c|c|c|c|} \hline J_{\alpha\alpha} & J_{\alpha\delta} & J_{\alpha\beta} & J_{\alpha d} \\ \hline & J_{\delta\delta} & J_{\delta\beta} & J_{\delta d} \\ \hline & & J_{\beta\beta} & J_{\beta d} \\ \hline & & & J_{dd} \\ \hline \end{array}$$

$\alpha \qquad \delta \qquad \beta \qquad d$

$$J = \left[\begin{array}{cccc} \left[\frac{\delta_i \delta_j}{1 - \alpha_i \alpha_j} \right] & \left[-\frac{\delta_i}{\alpha_i} \text{Log}(1 - \alpha_i \alpha_j) \right] & \left[\frac{-\delta_i d_j}{1 - \alpha_i \beta_j} \right] & \left[\frac{\delta_i}{\alpha_i} \text{Log}(1 - \alpha_i \beta_j) \right] \\ (K_0, K_0) & (K_0, K_0) & (K_0, K_1) & (K_0, K_1) \\ \left[\sum_{k=1}^{\infty} \frac{(\alpha_i \alpha_j)^k}{k^2} \right] & \left[\frac{d_j}{\beta_j} \text{Log}(1 - \alpha_i \beta_j) \right] & \left[-\sum_{k=1}^{\infty} \frac{(\alpha_i \beta_j)^k}{k^2} \right] & \\ (K_0, K_0) & (K_0, K_1) & (K_0, K_1) & \\ & \left[\frac{d_i d_j}{1 - \beta_i \beta_j} \right] & \left[-\frac{d_i}{\beta_i} \text{Log}(1 - \beta_i \beta_j) \right] & \\ & (K_1, K_1) & (K_1, K_1) & \\ & & \left[\sum_{k=1}^{\infty} \frac{(\beta_i \beta_j)^k}{k^2} \right] & \\ & & (K_1, K_1) & \end{array} \right]$$

IV.D - UN EXEMPLE

Ainsi pour un processus défini par : $X_t = \frac{(1 - \alpha L)^\delta}{(1 - \beta L)^d} \epsilon_t$, la matrice

d'information est donnée par :

$$J = \begin{bmatrix} \frac{\delta^2}{1 - \alpha^2} & -\frac{\delta}{\alpha} \text{Log}(1 - \alpha^2) & -\frac{\delta d}{1 - \alpha\beta} & \frac{\delta}{\alpha} \text{Log}(1 - \alpha\beta) \\ -\frac{\delta}{\alpha} \text{Log}(1 - \alpha^2) & \sum_{k=1}^{\infty} \frac{\alpha^{2k}}{k^2} & \frac{d}{\beta} \text{Log}(1 - \alpha\beta) & -\sum_{k=1}^{\infty} \frac{(\alpha\beta)^k}{k^2} \\ -\frac{\delta d}{1 - \alpha\beta} & \frac{d}{\beta} \text{Log}(1 - \alpha\beta) & \frac{d^2}{1 - \beta^2} & -\frac{d}{\beta} \text{Log}(1 - \beta^2) \\ \frac{\delta}{\alpha} \text{Log}(1 - \alpha\beta) & -\sum_{k=1}^{\infty} \frac{(\alpha\beta)^k}{k^2} & -\frac{d}{\beta} \text{Log}(1 - \beta^2) & \sum_{k=1}^{\infty} \frac{\beta^{2k}}{k^2} \end{bmatrix}$$

A N N E X E 1

COMPARAISON DIRECTE DES DEUX EXPRESSIONS DE \tilde{J}

$$\begin{aligned}
 \text{On a : } J_{\alpha\alpha} &= \frac{d^2}{4\pi} \int_{-\pi}^{\pi} \left[\frac{2\alpha - 2\cos w}{1 + \alpha^2 - 2\alpha\cos w} \right]^2 dw \\
 &= \frac{d^2}{4\pi\alpha^2} \int_{-\pi}^{\pi} \left[\frac{2\alpha^2 - 2\alpha\cos w}{1 + \alpha^2 - 2\alpha\cos w} \right]^2 dw \\
 &= \frac{d^2}{4\pi\alpha^2} \int_{-\pi}^{\pi} \left[1 + \frac{-1 + \alpha^2}{1 + \alpha^2 - 2\alpha\cos w} \right]^2 dw \\
 &= \frac{d^2}{4\pi\alpha^2} \left\{ 2\pi + (-1 + \alpha^2)^2 \int_{-\pi}^{\pi} \frac{dw}{(1 + \alpha^2 - 2\alpha\cos w)^2} \right. \\
 &\quad \left. + 2(-1 + \alpha^2) \int_{-\pi}^{\pi} \frac{dw}{1 + \alpha^2 - 2\alpha\cos w} \right\} .
 \end{aligned}$$

Si Z est un processus tel que : $(1 - \alpha L)^2 Z_t = \epsilon_t$, on a :

$$f_Z(w) = \frac{\sigma^2}{2\pi} \frac{1}{(1 + \alpha^2 - 2\alpha\cos w)^2} .$$

Si Y est un processus tel que : $(1 - \alpha L) Y_t = \epsilon_t$, on a :

$$f_Y(w) = \frac{\sigma^2}{2\pi} \frac{1}{1 + \alpha^2 - 2\alpha\cos w} .$$

On en déduit que :

$$J_{\alpha\alpha} = \frac{d^2}{4\pi\alpha^2} \left\{ 2\pi + (-1+\alpha^2)^2 \frac{2\pi}{\sigma^2} \gamma_Z(0) + 2(-1+\alpha^2) \frac{2\pi}{\sigma^2} \gamma_Y(0) \right\} .$$

Y étant autorégressif d'ordre 1, on a : $\gamma_Y(0) = \frac{\sigma^2}{1-\alpha^2}$.

Le processus Z a lui pour variance :

$$\begin{aligned} \gamma_Z(0) &= \frac{\sigma^2}{(1-\alpha^2)^2} \sum_{i=0}^1 \begin{pmatrix} 1 \\ 1-i \end{pmatrix} \begin{pmatrix} 1+i \\ i \end{pmatrix} \left(\frac{\alpha^2}{1-\alpha^2} \right)^i \\ &= \frac{\sigma^2}{(1-\alpha^2)^2} \left[1 + 2 \frac{\alpha^2}{1-\alpha^2} \right] . \end{aligned}$$

On en déduit :

$$\begin{aligned} J_{\alpha\alpha} &= \frac{d^2}{4\pi\alpha^2} \left\{ 2\pi + (-1+\alpha^2)^2 \frac{2\pi}{(1-\alpha^2)^2} \left[1 + 2 \frac{\alpha^2}{1-\alpha^2} \right] + 2(-1+\alpha^2) \frac{2\pi}{1-\alpha^2} \right\} \\ &= \frac{d^2}{1-\alpha^2} . \end{aligned}$$

Les autres vérifications s'effectuent de façon similaire.

A N N E X E 2

ETUDE DE LA MATRICE D'INFORMATION DE FISHER D'UN MADE(1)

a) Expression du déterminant

On a :

$$\begin{aligned} \det \tilde{J} &= \frac{d^2}{1 - \alpha^2} \sum_{j=1}^{\infty} \frac{\alpha^{2j}}{j^2} - \frac{d^2}{\alpha} \left[\text{Log} (1 - \alpha^2) \right]^2 \\ &= d^2 \left\{ \sum_{n=0}^{\infty} \alpha^{2n} \sum_{j=1}^{\infty} \frac{\alpha^{2j}}{j^2} - \frac{1}{\alpha} \sum_{n=1}^{\infty} \frac{\alpha^{2n}}{n} \sum_{j=1}^{\infty} \frac{\alpha^{2j}}{j} \right\} \\ &= d^2 \sum_{K=1}^{\infty} \alpha^{2K} \left\{ \sum_{n=0}^{K-1} \frac{1}{(K-n)^2} - \sum_{n=0}^{K-1} \frac{1}{(n+1)(K-n)} \right\} \end{aligned}$$

Comme : $\frac{1}{(n+1)(K-n)} = \frac{1}{K+1} \left[\frac{1}{(n+1)} + \frac{1}{(K-n)} \right]$, on a :

$$\sum_{n=0}^{K-1} \frac{1}{(n+1)(K-n)} = \frac{1}{K+1} \sum_{n=0}^{K-1} \left[\frac{1}{(n+1)} + \frac{1}{(K-n)} \right] = \frac{2}{K+1} \sum_{n=0}^{K-1} \frac{1}{n+1}$$

On en déduit que :

$$\det \tilde{J} = d^2 \sum_{K=1}^{\infty} \alpha^{2K} \sum_{n=0}^{K-1} \left[\frac{1}{(n+1)^2} - \frac{2}{(K+1)(n+1)} \right]$$

b) Signe des coefficients

Le premier coefficient correspondant à $K = 1$ est nul. Les autres coefficients sont donnés par :

$$\sum_{n=0}^{K-1} \left[\frac{1}{(n+1)^2} - \frac{2}{(K+1)(n+1)} \right] = c_{K-1} \quad (\text{par définition}) .$$

$$\text{On a : } c_1 = \left[1 - \frac{2}{3} \right] + \left[\frac{1}{4} - \frac{1}{3} \right] = \frac{1}{4} > 0 .$$

De plus :

$$\begin{aligned} c_K - c_{K-1} &= 2 \sum_{n=0}^{K-1} \left[\frac{1}{(K+1)(n+1)} - \frac{1}{(K+2)(n+1)} \right] + \frac{1}{(K+1)^2} - \frac{2}{(K+2)(K+1)} \\ &= 2 \sum_{n=0}^{K-1} \frac{1}{(K+1)(K+2)(n+1)} + \frac{1}{(K+1)^2} - \frac{2}{(K+2)(K+1)} \\ &= 2 \sum_{n=1}^{K-1} \frac{1}{(K+1)(K+2)(n+1)} + \frac{1}{(K+1)^2} > 0 . \end{aligned}$$

c) Expression de la variance asymptotique de \hat{d}_T et évolution

$$\begin{aligned} \text{On a : } V_{\text{as}} \left[\sqrt{T} (\hat{d}_T - d) \right] &= \frac{1}{\det J} \frac{d^2}{1 - \alpha^2} \\ &= \frac{\alpha^2}{\alpha^2 \sum_{j=1}^{\infty} \frac{\alpha^{2j}}{j^2} - (1-\alpha^2) \text{Log}^2(1-\alpha^2)} = \frac{\alpha^2}{a(\alpha^2)} . \end{aligned}$$

Il s'agit d'une fonction paire. Elle prend la valeur :

$$\frac{1}{\sum_{j=1}^{\infty} \frac{1}{j^2}} = \frac{6}{\pi^2} , \text{ si } \alpha^2 = 1 \text{ et la valeur } +\infty , \text{ si } \alpha^2 \text{ est nul.}$$

Sa dérivée par rapport à α^2 est :

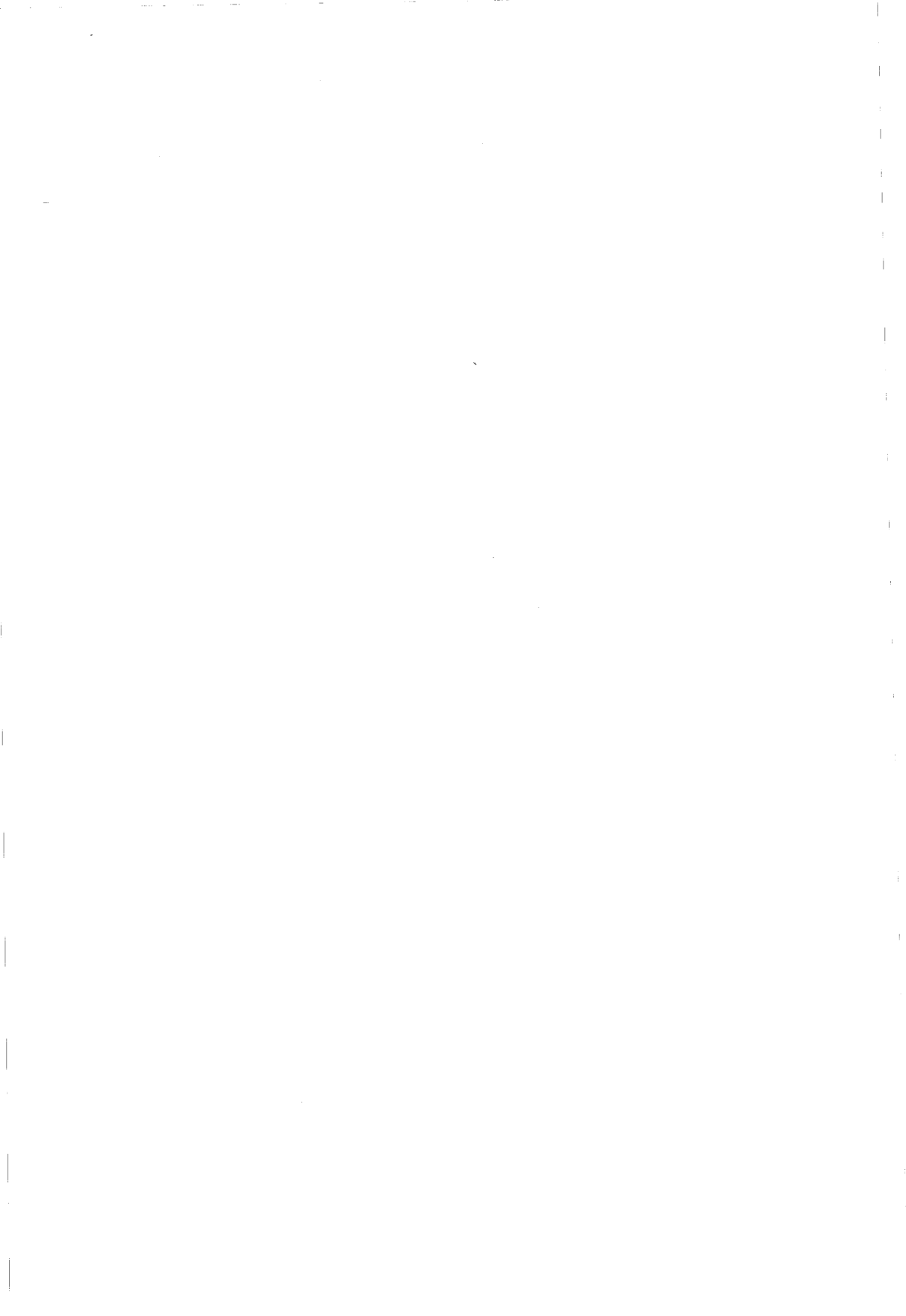
$$\begin{aligned} \frac{\partial V}{\partial(\alpha^2)} &= \left[\sqrt{T} (\hat{\alpha}_T - d) \right] = \\ &= \frac{1}{a^2(\alpha^2)} \left[\alpha^2 \sum_{j=1}^{\infty} \frac{\alpha^{2j}}{j^2} - (1-\alpha^2) \text{Log}^2(1-\alpha^2) - \alpha^2 \sum_{j=1}^{\infty} \frac{\alpha^{2j}}{j} - \sum_{j=1}^{\infty} \frac{\alpha^{2j+2}}{j} \right. \\ &\quad \left. - \alpha^2 \text{Log}^2(1-\alpha^2) - 2 \alpha^2 \text{Log}(1-\alpha^2) \right] \\ &= - \frac{1}{a^2(\alpha^2)} \text{Log}(1-\alpha^2) \left[\text{Log}(1-\alpha^2) + \alpha^2 \right] < 0 . \end{aligned}$$

On remarque aussi que cette dérivée vaut $-\infty$ pour $\alpha^2 = 1$.



R E F E R E N C E S

- Anderson, T. (1971) : The statistical analysis of time series, Wiley.
- Azencott, R. et D., Dacunha Castelle (1984) : Séries d'observations irrégulières, modélisation et prévision, Masson, Paris.
- Basawa, I. et Prakasa-Rao, B. (1980) : Statistical inference for stochastic processes, Academic , London.
- Fox, R. et M. Taqqu (1986) : Large properties of parameter estimates for strongly dependent stationary gaussian time series, The Annals of Stat. 14, 517-532.
- Gonçalves, E. (1987) : Une généralisation des processus ARMA, Annales d'Economie et de Statistique, 5, 109-146.
- Hannan, E. (1973) : The asymptotic theory of linear time series models, J. Appl. Probab. 10, 130-145.
- Jennrich, R. (1969) : Asymptotic properties of nonlinear least squares estimators, The Annals of Math. Stat. 40, 633-643.
- Rao, R. (1973) : Linear statistical inference and its applications, Wiley.
- Walker, A. (1964) : Asymptotic properties of least squares estimators of parameters of the spectrum of a stationary non deterministic time series, J. Austral. Math. Soc. 4, 363-384.
- Whittle, P. (1951) : Hypothesis testing in time series analysis, Hafaer, New-York.



Ⓔ

**TEST DU MULTIPLICATEUR
DE LAGRANGE EN SERIES TEMPORELLES**



TEST DU MULTIPLICATEUR
DE LAGRANGE EN SERIES TEMPORELLES.

Dans ce chapitre nous considérons un processus stationnaire admettant des représentations moyenne mobile et autorégressive :

$$\phi_{\theta}(L) X_t = \varepsilon_t \iff X_t = A_{\theta}(L) \varepsilon_t ,$$

où $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ est un bruit blanc indépendant gaussien de variance σ^2 . Nous supposons que le paramètre θ peut être décomposé en deux sous-paramètres $\theta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ et nous nous proposons d'introduire des procédures de test de l'hypothèse $H_0 = \{\beta = 0\}$ contre $H_1 = \{\beta \neq 0\}$.

Dans le premier paragraphe, nous décrivons l'approche par le multiplicateur de Lagrange. Si $\hat{\alpha}_T, \hat{\beta}_T, \hat{\sigma}_T^2$ désignent les estimateurs du maximum de vraisemblance des paramètres et $\hat{\alpha}_T^0, \hat{\beta}_T^0 = 0, \hat{\sigma}_T^{02}$ ces estimateurs évalués sous l'hypothèse nulle $H_0 = \{\beta = 0\}$, l'idée du test consiste à accepter l'hypothèse H_0 , lorsque le score estimé sous H_0 : $\frac{\partial \text{Log } L_T^*}{\partial \beta}(x; \hat{\alpha}_T^0, 0, \hat{\sigma}_T^{02})$ est proche de zéro. Nous commençons par donner la loi asymptotique du score estimé sous H_0 et l'utilisons ensuite pour construire la statistique de test.

Dans le second paragraphe nous interprétons ce test d'un point de vue local. Ceci permet l'introduction d'un modèle linéaire constituant une bonne approximation du modèle initial au sens où les deux tests du multiplicateur de Lagrange de H_0 effectués sur ces deux modèles coïncident.

I - DESCRIPTION DU TEST

I.A. - Les estimateurs du maximum de vraisemblance

Nous supposons satisfaites, au niveau des modèles non contraint et contraint par l'hypothèse H_0 , les conditions assurant l'existence asymptotique des estimateurs du maximum de vraisemblance, leur convergence, la possibilité de les exprimer en fonction du vecteur des scores ... Reprenant les résultats du chapitre précédent, nous avons vu que :

$$\sqrt{T} \begin{bmatrix} \hat{\theta}_T - \theta_0 \\ \hat{\sigma}_T^2 - \sigma_0^2 \end{bmatrix} = J(\theta_0, \sigma_0^2)^{-1} \frac{1}{\sqrt{T}} \frac{\partial \text{Log } L_T^*}{\partial (\theta', \sigma^2)} (x; \theta_0, \sigma_0^2) + o_p(1).$$

Ceci entraîne en développant :

$$\sqrt{T} \begin{pmatrix} \hat{\alpha}_T - \alpha_0 \\ \hat{\beta}_T - \beta_0 \end{pmatrix} = \begin{bmatrix} J_{\alpha\alpha} & J_{\alpha\beta} \\ J_{\beta\alpha} & J_{\beta\beta} \end{bmatrix}^{-1} \frac{1}{\sqrt{T}} \begin{bmatrix} \frac{\partial \text{Log } L_T^*}{\partial \alpha} (x; \theta_0, \sigma_0^2) \\ \frac{\partial \text{Log } L_T^*}{\partial \beta} (x; \theta_0, \sigma_0^2) \end{bmatrix} + o_p(1),$$

développement valable sous l'hypothèse générale.

Si nous nous plaçons sous l'hypothèse nulle $H_0 = \{\beta = 0\}$, nous avons le même résultat avec les éléments de la matrice d'information évalués sous H_0 ; si ceux-ci sont notés $J_{\alpha\alpha}^0$, $J_{\alpha\beta}^0$, $J_{\beta\beta}^0$, nous avons donc :

$$(1.1) \quad \sqrt{T} \begin{bmatrix} \hat{\alpha}_T - \alpha_0 \\ \hat{\beta}_T \end{bmatrix} = \begin{bmatrix} J_{\alpha\alpha}^0 & J_{\alpha\beta}^0 \\ J_{\beta\alpha}^0 & J_{\beta\beta}^0 \end{bmatrix}^{-1} \frac{1}{\sqrt{T}} \begin{bmatrix} \frac{\partial \text{Log } L_T^*}{\partial \alpha} [x ; \alpha_0, 0 ; \sigma_0^2] \\ \frac{\partial \text{Log } L_T^*}{\partial \beta} (x ; \alpha_0, 0, \sigma_0^2) \end{bmatrix}$$

sous l'hypothèse H_0 .

Les résultats concernant le comportement asymptotique des estimateurs du maximum de vraisemblance peuvent aussi être appliqués au cas contraint.

On obtient sous H_0 :

$$\sqrt{T} \begin{bmatrix} \hat{\alpha}_T^0 - \alpha_0 \\ \hat{\sigma}_T^{02} - \sigma_0^2 \end{bmatrix} = \begin{bmatrix} J_{\alpha\alpha}^0 & 0 \\ 0 & \frac{1}{2\sigma_0^4} \end{bmatrix}^{-1} \frac{1}{\sqrt{T}} \begin{bmatrix} \frac{\partial \text{Log } L_T^*}{\partial \alpha} [x ; \alpha_0, 0, \sigma_0^2] \\ \frac{\partial \text{Log } L_T^*}{\partial \sigma^2} [x ; \alpha_0, 0, \sigma_0^2] \end{bmatrix} + o_p(1)$$

Ceci entraîne en particulier :

$$(1.2) \quad \sqrt{T}(\hat{\alpha}_T^0 - \alpha_0) = [J_{\alpha\alpha}^0]^{-1} \frac{1}{\sqrt{T}} \frac{\partial \text{Log } L_T^*}{\partial \alpha} [x ; \alpha_0, 0, \sigma_0^2] + o_p(1)$$

I.B. - Loi asymptotique du score estimé

Considérons alors le score estimé $\frac{1}{\sqrt{T}} \frac{\partial \text{Log } L_T^*}{\partial \beta} (x ; \hat{\alpha}_T^0, 0, \hat{\sigma}_T^{02})$.

Il résulte de la forme explicite du score que celui-ci s'écrit :

$$\frac{1}{\sqrt{T}} \frac{\partial \text{Log } L_T^*}{\partial \beta} (x ; \hat{\alpha}_T^o, 0, \hat{\sigma}_T^{o2}) = \frac{1}{\hat{\sigma}_T^{o2}} h_T(x ; \hat{\alpha}_T^o) ;$$

on peut donc écrire :

$$\frac{1}{\sqrt{T}} \frac{\partial \text{Log } L_T^*}{\partial \beta} (x ; \hat{\alpha}_T^o, 0, \hat{\sigma}_T^{o2}) = \frac{\sigma_o^2}{\hat{\sigma}_T^{o2}} \frac{1}{\sqrt{T}} \frac{\partial \text{Log } L_T^*}{\partial \beta} (x ; \hat{\alpha}_T^o, 0, \sigma_o^2).$$

Comme $\hat{\alpha}_T^o$ tend vers α_o à la vitesse $\frac{1}{\sqrt{T}}$, nous pouvons faire un développement limité en probabilité, sous H_o :

$$\begin{aligned} & \frac{1}{\sqrt{T}} \frac{\partial \text{Log } L_T^*}{\partial \beta} [x ; \hat{\alpha}_T^o, 0, \hat{\sigma}_T^{o2}] \\ &= \frac{\sigma_o^2}{\hat{\sigma}_T^{o2}} \left[\frac{1}{\sqrt{T}} \frac{\partial \text{Log } L_T^*}{\partial \beta} (x ; \alpha_o, 0, \sigma_o^2) + \frac{1}{T} \frac{\partial^2 \text{Log } L_T^*}{\partial \beta \partial \alpha'} (x ; \alpha_o, 0, \sigma_o^2) \right. \\ & \quad \left. \sqrt{T} (\hat{\alpha}_T^o - \alpha_o) + o_p(1) \right] \\ &= \frac{\sigma_o^2}{\hat{\sigma}_T^{o2}} \left[\frac{1}{\sqrt{T}} \frac{\partial \text{Log } L_T^*}{\partial \beta} (x ; \alpha_o, 0, \sigma_o^2) - J_{\beta\alpha}^o \sqrt{T} (\hat{\alpha}_T^o - \alpha_o) + o_p(1) \right] \\ &= \frac{1}{\sqrt{T}} \frac{\partial \text{Log } L_T^*}{\partial \beta} (x ; \alpha_o, 0, \sigma_o^2) - J_{\beta\alpha}^o \sqrt{T} (\hat{\alpha}_T^o - \alpha_o) + o_p(1) \\ & \quad \text{(car } \sigma_o^2 = p \lim \hat{\sigma}_T^{o2}) \\ &= \frac{1}{\sqrt{T}} \frac{\partial \text{Log } L_T^*}{\partial \beta} (x ; \alpha_o, 0, \sigma_o^2) - J_{\beta\alpha}^o (J_{\alpha\alpha}^o)^{-1} \frac{1}{\sqrt{T}} \frac{\partial \text{Log } L_T^*}{\partial \alpha} (x ; \alpha_o, 0, \sigma_o^2) \\ & \quad + o_p(1) . \end{aligned}$$

Comme $J_{\beta\alpha}^0$, $J_{\alpha\alpha}^0$ sont des covariance et variance de vecteurs des scores et que celui-ci est asymptotiquement normal, on a :

$$\begin{aligned} & \frac{1}{\sqrt{T}} \frac{\partial \text{Log } L_T^*}{\partial \beta} [x ; \hat{\alpha}_T^0, 0, \hat{\sigma}_T^{02}] \\ &= \frac{1}{\sqrt{T}} \frac{\partial \text{Log } L_T^*}{\partial \beta} - \text{Cov}_{\text{as}} \left[\frac{1}{\sqrt{T}} \frac{\partial \text{Log } L_T^*}{\partial \beta}, \frac{1}{\sqrt{T}} \frac{\partial \text{Log } L_T^*}{\partial \alpha} \right] \\ & \quad \left[\text{V}_{\text{as}} \left(\frac{1}{\sqrt{T}} \frac{\partial \text{Log } L_T^*}{\partial \alpha} \right) \right]^{-1} \frac{1}{\sqrt{T}} \frac{\partial \text{Log } L_T^*}{\partial \alpha} + o_p(1) \end{aligned}$$

Propriété 1.3.- Le vecteur des scores estimés sous H_0 est asymptotiquement équivalent au résidu de la régression théorique de

$$\frac{1}{\sqrt{T}} \frac{\partial \text{Log } L_T^*}{\partial \beta} \text{ sur } \frac{1}{\sqrt{T}} \frac{\partial \text{Log } L_T^*}{\partial \alpha} .$$

En particulier, on en déduit immédiatement sa loi

Propriété 1.4.- Sous l'hypothèse $H_0 = \{\beta = 0\}$, on a :

$$\frac{1}{\sqrt{T}} \frac{\partial \text{Log } L_T^*}{\partial \beta} (x ; \hat{\alpha}_T^0, 0, \hat{\sigma}_T^{02}) \xrightarrow{d} Z ,$$

avec : $Z \rightsquigarrow N[0, J_{\beta\beta}^0 - J_{\beta\alpha}^0 (J_{\alpha\alpha}^0)^{-1} J_{\alpha\beta}^0]$

I.C. - Statistique de test

Lorsque la matrice $J_{\beta\beta}^0 - J_{\beta\alpha}^0 (J_{\alpha\alpha}^0)^{-1} J_{\alpha\beta}^0$ est inversible, on peut passer à la forme quadratique associée :

$$\frac{1}{\sqrt{T}} \frac{\partial \text{Log } L_T^*}{\partial \beta'} (x ; \hat{\alpha}_T^0, 0, \hat{\sigma}_T^{02}) [J_{\beta\beta}^0 - J_{\beta\alpha}^0 (J_{\alpha\alpha}^0)^{-1} J_{\alpha\beta}^0]^{-1}$$

$$\frac{1}{\sqrt{T}} \frac{\partial \text{Log } L_T^*}{\partial \beta} (x ; \hat{\alpha}_T^0, 0, \hat{\sigma}_T^{02})$$

dont la loi est une loi du khi-deux avec un nombre de degrés de liberté égal à la taille du sous-paramètre β . Le même résultat reste vrai si les éléments de la matrice d'information sont remplacés par des estimateurs convergents sous H_0 .

Propriété 1.5. - Si β est un sous-paramètre de taille r , si $\chi_{95\%}^2(r)$ désigne le quantile à 95 % de la loi du khi-deux à r degrés de liberté, si \hat{J}^0 est un estimateur convergent de J^0 sous l'hypothèse nulle $H_0 = (\beta = 0)$, on appelle statistique du multiplicateur de Lagrange :

$$\xi_{LM} = \frac{1}{\sqrt{T}} \frac{\partial \text{Log } L_T^*}{\partial \beta'} (x ; \hat{\alpha}_T^0, 0, \hat{\sigma}_T^{02}) [\hat{J}_{\beta\beta}^0 - \hat{J}_{\beta\alpha}^0 \hat{J}_{\alpha\alpha}^{0-1} \hat{J}_{\alpha\beta}^0]^{-1}$$

$$\frac{1}{\sqrt{T}} \frac{\partial \text{Log } L_T^*}{\partial \beta} (x ; \hat{\alpha}_T^0, 0, \hat{\sigma}_T^{02}) .$$

Le test du multiplicateur de Lagrange associé consiste à :

accepter l'hypothèse H_0 , si : $\xi_{LM} \leq \chi_{95\%}^2(r)$,
 refuser cette hypothèse sinon .

Ce test est asymptotiquement de niveau 95 % .

L'approche par le multiplicateur de Lagrange a été introduite par [Aitchison et Silvey (1958)] et [Silvey (1959)] dans le cas i.i.d. Sa généralisation aux séries temporelles est due à [Hosking (1980)].

II - MODELES LINEAIRES APPROCHES

II.A. - Une forme de la statistique $\xi_{L.M}$ facilement interprétable

Il existe divers choix pour les statistiques du multiplicateur de Lagrange. Ces choix dépendent du type de vraisemblance exacte ou approchée retenue et des estimateurs de J utilisés. Retenant la vraisemblance conditionnelle, on a :

$$\frac{\partial \text{Log } \hat{L}_T^c}{\partial \theta} (x ; \theta, \sigma^2) = -\frac{1}{\sigma^2} \cdot \sum_{t=1}^T \frac{\partial \phi_{\theta}(L)}{\partial \theta} \tilde{X}_t \quad \phi_{\theta}(L) \tilde{X}_t ,$$

où

$$\tilde{X}_t = \begin{cases} X_t , & \text{si } 1 \leq t \leq T , \\ 0 , & \text{sinon .} \end{cases}$$

Posons :

$$Z_{1t} = \frac{\partial \phi_{\theta_0}(L)}{\partial \alpha} \tilde{X}_t , \quad Z_{2t} = \frac{\partial \phi_{\theta_0}(L)}{\partial \beta} \tilde{X}_t ,$$

$$\tilde{\varepsilon}_t^0 = \phi_{\theta_0}(L) \tilde{X}_t .$$

L'estimateur contraint de la variance est : $\hat{\sigma}_T^2 = \frac{1}{T} \sum_{t=1}^T \tilde{\varepsilon}_t^2$.

D'autre part des estimateurs convergents de $J_{\beta\alpha}^0$, $J_{\beta\beta}^0$, $J_{\alpha\alpha}^0$ sont :

$$\hat{J}_{\beta\alpha}^0 = \frac{1}{\hat{\sigma}_T^2} \frac{1}{T} \sum_{t=1}^T Z_{2t} Z'_{1t} , \quad \hat{J}_{\beta\beta}^0 = \frac{1}{\hat{\sigma}_T^2} \frac{1}{T} \sum_{t=1}^T Z_{2t} Z'_{2t} ,$$

$$\hat{J}_{\alpha\alpha}^0 = \frac{1}{\hat{\sigma}_T^2} \frac{1}{T} \sum_{t=1}^T Z_{1t} Z'_{1t} .$$

Remplaçant dans l'expression de la statistique du multiplicateur de Lagrange, nous obtenons le résultat suivant.

Propriété 2.1. - Un choix possible de la statistique du multiplicateur de Lagrange pour tester $H_0 = \{\beta = 0\}$ est :

$$\xi_{L.M} = \frac{1}{\hat{\sigma}_T^2} \left(\sum_{t=1}^T Z_{2t} \tilde{\varepsilon}_t^0 \right)' \left\{ \sum_{t=1}^T Z_{2t} Z_{2t}' - \sum_{t=1}^T Z_{2t} Z_{1t}' \left[\sum_{t=1}^T Z_{1t} Z_{1t}' \right]^{-1} \sum_{t=1}^T Z_{1t} Z_{2t}' \right\}^{-1} \left(\sum_{t=1}^T Z_{2t} \tilde{\varepsilon}_t^0 \right),$$

avec :

$$Z_{1t} = \frac{\partial \hat{\theta}_T^0(L)}{\partial \alpha} \tilde{X}_t, \quad Z_{2t} = \frac{\partial \hat{\theta}_T^0(L)}{\partial \beta} \tilde{X}_t,$$

$$\tilde{\varepsilon}_t^0 = \hat{\phi}_{\hat{\theta}_T^0(L)} \tilde{X}_t.$$

Cette forme de la statistique a une interprétation directe en terme de régression. Introduisons le modèle de régression fictif :

$$(2.2) \quad \tilde{\varepsilon}_t^0 = Z_{1t}' a + Z_{2t}' b + e_t, \quad t = 1, \dots, T,$$

et l'hypothèse nulle fictive $H_0 = \{b = 0\}$.

Si nous notons $M = Z(Z'Z)^{-1}Z'$ et $M_1 = Z_1(Z_1'Z_1)^{-1}Z_1'$ les projecteurs respectifs sur les espaces engendrés par les colonnes de Z et par les colonnes de Z_1 , il résulte des équations de vraisemblance que : $M_1 \tilde{\varepsilon}^0 = 0 \iff Z_1' \tilde{\varepsilon}^0 = 0$.

On peut alors facilement vérifier que la statistique de test s'exprime à partir des sommes de carrés des résidus correspondant à ce modèle fictif sous l'hypothèse générale $[SCR_H]$ et sous l'hypothèse nulle (SCR_{H_0}) :

$$(2.3) \quad \xi_{LM} = \frac{SCR_{H_0} - SCR_H}{\hat{\sigma}_T^2}$$

En effet, on a :

$$\begin{aligned} SCR_{H_0} - SCR_H &= \tilde{\varepsilon}^{o'} (Id - M_1) \tilde{\varepsilon}^o - \tilde{\varepsilon}^{o'} (Id - M) \tilde{\varepsilon}^o \\ &= \tilde{\varepsilon}^{o'} (M - M_1) \tilde{\varepsilon}^o = \tilde{\varepsilon}^{o'} M \tilde{\varepsilon}^o \\ &= \tilde{\varepsilon}^{o'} Z (Z'Z)^{-1} Z' \tilde{\varepsilon}^o \\ &= [0 \quad \tilde{\varepsilon}_o' \quad Z_2'] \begin{bmatrix} Z_1'Z_1 & Z_1'Z_2 \\ Z_2'Z_1 & Z_2'Z_2 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ Z_2' \tilde{\varepsilon}^o \end{bmatrix} \\ &= \tilde{\varepsilon}^{o'} Z_2 \{Z_2'Z_2 - Z_2'Z_1 [Z_1'Z_1]^{-1} Z_1'Z_2\}^{-1} Z_2' \tilde{\varepsilon}^o \\ &= \hat{\sigma}_T^2 \xi_{LM} . \end{aligned}$$

II.B. - Alternatives localement équivalentes

Considérons deux modèles possibles pour décrire la série stationnaire X :

$$(2.4) \quad \phi_{\alpha, \beta}(L) X_t = \varepsilon_t \quad , \quad v_{\varepsilon_t} = \sigma^2 \quad ,$$

$$(2.5) \quad \phi_{\alpha, \beta}^*(L) X_t = \eta_t \quad , \quad v_{\eta_t} = \sigma^2 \quad .$$

(Nous conservons des notations identiques pour les paramètres afin de faciliter les écritures).

Le principe d'un test du multiplicateur de Lagrange consistant à regarder le comportement de la dérivée de la log vraisemblance au voisinage de l'hypothèse nulle, il est naturel d'étudier les modèles dans ce même voisinage.

Définition 2.6. - Les deux modèles seront dits fournir des alternatives localement équivalentes si les deux conditions suivantes sont satisfaites :

- i) $\phi_{\alpha,0}(L) = \phi_{\alpha,0}^*(L)$,
- ii) $\frac{\partial \phi_{\alpha,0}(L)}{\partial \beta} = A \frac{\partial \phi_{\alpha,0}^*(L)}{\partial \beta}$, où A est une matrice inversible.

La première égalité signifie que les hypothèses $H_0 = (\beta = 0)$ définie sur le modèle (2.4) et $H_0^* = (\beta = 0)$ définie sur le modèle (2.5) correspondent aux mêmes familles de lois : $H_0 = H_0^*$. La seconde égalité concerne les alternatives $H_1 = (\beta \neq 0)$ et $H_1^* = (\beta \neq 0)$ associées aux deux modèles. Si on ne retient de ces hypothèses que les lois correspondant aux valeurs petites de β , les deux hypothèses sont "équivalentes" au premier ordre.

La condition (i) entraîne évidemment que les estimateurs contraints de α et σ^2 évalués en utilisant (2.4) ou (2.5) sont confondus. Elle

entraîne ainsi que $\frac{\partial \phi_{\alpha,0}(L)}{\partial \alpha} = \frac{\partial \phi_{\alpha,0}^*(L)}{\partial \alpha}$ et donc les vecteurs

$$Z_{1t} = \frac{\partial \phi_{\alpha T, 0}^o(L)}{\partial \alpha} \tilde{X}_t \quad \text{et} \quad Z_{1t}^* = \frac{\partial \phi_{\alpha T, 0}^{*o}(L)}{\partial \alpha} \tilde{X}_t$$

sont confondus.

De même il résulte de la condition (ii) que les vecteurs Z_{2t} et Z_{2t}^* sont en liaison linéaire bijective. On en déduit la propriété ci-dessous :

Propriété 2.7. - Si deux modèles fournissent des alternatives localement équivalentes, les statistiques du multiplicateur de Lagrange calculées à partir de l'approche conditionnelle (2.1) sont confondues : $\xi_{LM} = \xi_{LM}^*$.

Ce résultat peut être utilisé de deux manières distinctes.

i) Lorsque le test refuse l'hypothèse nulle, il est clair qu'il n'y a aucune raison a priori de se diriger plutôt vers H_1 ou vers H_1^* .

ii) La propriété peut servir pour déterminer plus facilement la statistique ξ_{LM} en remplaçant le modèle initial par un autre modèle plus maniable fournissant une alternative localement équivalente.

II.C. - Modèles décomposables

Afin d'illustrer la démarche des alternatives localement équivalentes, nous considérons le cas usuel où l'opérateur autorégressif se décompose en :

$$(2.8) \quad \phi_{\alpha,\beta}(L) = \phi_{1,\alpha}(L) \phi_{2,\beta}(L) \quad , \quad \text{avec} \quad \phi_{1,\alpha}(0) = 1$$

(et donc aussi $\phi_{2,\beta}(0) = 1$).

Développant au premier ordre cet opérateur au voisinage de $\beta = 0$, nous obtenons :

$$\phi_{\alpha,\beta}(L) \# \phi_{1,\alpha}(L) \left[\phi_{2,0}(L) + \frac{\partial \phi_{2,0}(L)}{\partial \beta} \beta \right] .$$

Comme il est toujours possible de choisir par convention la décomposition de façon que :

$$(2.9) \quad \phi_{2,0}(L) = 1 \quad ,$$

nous avons :

$$\phi_{\alpha,\beta}(L) \# \phi_{1,\alpha}(L) + \phi_{1,\alpha}(L) \frac{\partial \phi_{2,0}(L)}{\partial \beta} \beta .$$

Il est donc à ce stade naturel de considérer les deux modèles :

$$(2.10) \quad \phi_{1,\alpha}(L) \phi_{2,\beta}(L) X_t = \epsilon_t \quad , \quad \forall \epsilon_t = \sigma^2 \quad \text{et}$$

$$(2.11) \quad \phi_{\alpha,\beta}^*(L) X_t = \left[\phi_{1,\alpha}(L) + \phi_{1,\alpha}(L) \frac{\partial \phi_{2,0}(L)}{\partial \beta} \beta \right] X_t = \eta_t \quad , \quad \forall \eta_t = \sigma^2 .$$

On vérifie immédiatement que :

$$\phi_{\alpha,0}(L) = \phi_{1,\alpha}(L) = \phi_{\alpha,0}^*(L) \quad ,$$

et que :

$$\frac{\partial \phi_{\alpha,0}(L)}{\partial \beta} = \phi_{1,\alpha}(L) \frac{\partial \phi_{2,0}(L)}{\partial \beta} = \frac{\partial \phi_{\alpha,0}^*(L)}{\partial \beta} .$$

Propriété 2.12.- Le modèle décomposable (2.8) peut pour le test du multiplicateur de Lagrange de $H_0 = \{\beta = 0\}$ être remplacé par

$$\phi_{1,\alpha}(L) X_t + \phi_{1,\alpha}(L) \frac{\partial \phi_{2,0}(L)}{\partial \beta} X_t \beta = \eta_t, \quad V\eta_t = \sigma^2,$$

qui fournit une alternative localement équivalente.

L'avantage est évidemment de remplacer le modèle initial généralement non linéaire en β par un autre modèle qui lui possède cette propriété de linéarité.

Un autre "modèle linéaire approché" peut facilement être déduit de celui que nous avons obtenu. Posons :

$$Z_t = \phi_{1,\alpha_T}(L) \frac{\partial \phi_{2,0}(L)}{\partial \beta} X_t$$

et introduisons le modèle fictif :

$$(2.13) \quad \phi_{1,\alpha}(L) X_t + Z_t' \beta = \xi_t, \quad V\xi_t = \mu^2.$$

Effectuant le test du multiplicateur de Lagrange de $H_0 = (\beta = 0)$ en faisant comme si, dans Z_t , $\hat{\alpha}_T$ était déterministe, on voit que la statistique du multiplicateur de Lagrange coïncide avec celle du modèle (2.8).

Propriété 2.14.- La statistique ξ_{LM} peut être calculé à partir du modèle linéaire approché :

$$\phi_{1,\alpha}(L) X_t + Z_t' \beta = \xi_t, \quad V\xi_t = \mu^2,$$

avec $Z_t = \phi_{1, \hat{\alpha}_T^o(L)} \frac{\partial \phi_{2, o}^{(L)}}{\partial \beta} X_t$ et en faisant comme si $\hat{\alpha}_T^o$ était déterministe.

Cette dernière approximation est très simple d'utilisation :

i) on commence généralement par estimer le modèle sous l'hypothèse nulle, ce qui fournit les estimateurs $\hat{\alpha}_T^o$, $\hat{\sigma}_T^{o2}$ et les résidus contraints ε_t^{o} , $t = 1, \dots, T$;

ii) on obtient alors d'après (2.14) la statistique ξ_{LM} en testant $\{\beta = 0\}$ dans le modèle fictif :

$$\phi_{1, \alpha}^{(L)} X_t + Z_t' \beta = \xi_t \quad , \quad \text{avec} \quad Z_t = \frac{\partial \phi_{2, o}^{(L)}}{\partial \beta} \varepsilon_t^{o} .$$

Le test se ramène donc à un test d'oubli de variables ; celles-ci Z_t se calculant très facilement à partir des résidus ε_t^{o} .

Une autre interprétation du test du multiplicateur de Lagrange se déduit de l'expression du score estimé. On a en effet :

$$\begin{aligned} \frac{\partial \text{Log } L_T^C}{\partial \beta} (x ; \hat{\alpha}_T^o, 0, \hat{\sigma}_T^{o2}) &= - \frac{1}{\hat{\sigma}_T^{o2}} \sum_{t=1}^T \left[\frac{\partial \phi_{2, o}^{(L)}}{\partial \beta'} \phi_{1, \hat{\alpha}_T^o(L)} X_t \right] \\ &\quad \left[\phi_{2, o}^{(L)} \phi_{1, \hat{\alpha}_T^o(L)} X_t \right] \\ &= - \frac{1}{\hat{\sigma}_T^{o2}} \sum_{t=1}^T \frac{\partial \phi_{2, o}^{(L)}}{\partial \beta'} \varepsilon_t^{o} \cdot \varepsilon_t^{o} . \end{aligned}$$

Propriété 2.15. - Tester l'hypothèse $H_0 = \{\beta = 0\}$ par la

méthode du multiplicateur de Lagrange revient à tester l'orthogonalité entre les résidus contraints ε_t^{o} et les variables $Z_t = \frac{\partial \phi_{2, o}^{(L)}}{\partial \beta} \varepsilon_t^{o}$.

II.D. - Modèles inversés

Etant donné un modèle décomposable

$$(2.16) \quad \phi_{1,\alpha}(L) \phi_{2,\beta}(L) X_t = \varepsilon_t, \quad V\varepsilon_t = \sigma^2, \quad \text{avec } \phi_{2,0}(L) = 1,$$

on est souvent conduit à considérer simultanément le modèle partiellement inversé :

$$(2.17) \quad \phi_{1,\alpha}(L) \frac{1}{\phi_{2,\beta}(L)} X_t = \eta_t, \quad V\eta_t = \sigma^2.$$



Propriété 2.18.- Le modèle décomposable (2.8) et le modèle partiellement inversé (2.17) fournissent des alternatives localement équivalentes pour l'hypothèse $H_0 = \{\beta = 0\}$.

Preuve : On a : $\phi_{\alpha,\beta}^*(L) = \phi_{1,\alpha}(L) \frac{1}{\phi_{2,\beta}(L)}$. On peut alors vérifier les deux conditions de la propriété (2.6) :

$$i) \quad \phi_{\alpha,0}^*(L) = \phi_{1,\alpha}(L) \frac{1}{\phi_{2,0}(L)} = \phi_{1,\alpha}(L) = \phi_{\alpha,0}(L) :$$

$$ii) \quad \frac{\partial \phi_{\alpha,0}^*(L)}{\partial \beta} = - \frac{\phi_{1,\alpha}(L)}{\phi_{2,0}^2(L)} \frac{\partial \phi_{2,0}(L)}{\partial \beta} = - \text{Id} \left[\phi_{1,\alpha}(L) \frac{\partial \phi_{2,0}(L)}{\partial \beta} \right] = - \text{Id} \frac{\partial \phi_{\alpha,0}(L)}{\partial \beta}.$$

Q.E.D.

Exemple 2.19.- L'application la plus classique de cette propriété générale est donnée dans Godfrey-Wickens (1982). L'hypothèse nulle H_0 correspond à l'existence d'une représentation ARMA(p,q) pour le processus. Les alternatives directe et inversée sont obtenues en prenant pour $\phi_{2,\beta}(L)$ un polynôme de retards de degré r . Les deux alternatives sont :

$H_1 = \{\text{le processus admet une représentation ARMA}(p+r, q)\}$,

$H_1^* = \{\text{le processus admet une représentation ARMA}(p, q+r)\}$.

Ces deux alternatives sont localement équivalentes au voisinage de l'hypothèse nulle.

R E F E R E N C E S

- AITCHISON, J et S.D. SILVEY (1958) : Maximum likelihood estimation of parameters subject to restraints,
A.M.S., Vol. 29 p. 813-828.
- DAVIES, R. (1977) : Hypothesis testing when a nuisance parameter is present only under the alternative,
Biometrika 64, 247-254.
- GODFREY, L. (1979) : Testing the adequacy of a time series model,
Biometrika 64, 67-72.
- GODFREY, L. (1981) : On the invariance of Lagrange multiplier test with respect to certain changes in the alternative hypothesis,
Econometrica 49, 1443-1455.
- GODFREY, L. et WICKENS, M.R. (1982) : Tests for misspecification using locally equivalent alternative models
dans Chow-Corsi "Evaluating the reliability of macro-economic models". Wiley.
- HOSKING, J.R. (1980) : Lagrange multiplier tests of time series models,
JRSS B 42, 170-181.
- McLEOD A.I. et W.K. LI (1983) : Diagnostic checking ARMA time series models using squared residual autocorrelations,
Journal of Time Series Analysis 4, 269-273.
- POSKITT, D.S. et A.R. TREMAYNE (1981) : An approach to testing linear time series models,
Annals of Stat. 9, 974-986.
- POSKITT, D.S. et A.R. TREMAYNE (1984) : Testing misspecification in vector time series models with exogenous variables,
JRSSB 46, 304-315.

SAIKKONEN, P. et R. LUUKKONEN (1986) : Lagrange multiplier tests for testing non linearities in time series models, Univ. Helsinki D.P.

SILVEY, S. (1959) : The Lagrangian multiplier test, Annals of Mathematical Statistics 30, 389-407.



**AGREGATION DE PROCESSUS AUTOREGRESSIFS
D'ORDRE 1**

I - INTRODUCTION.

Les modèles usuels de séries temporelles comme les représentations autorégressives (ou autorégressives moyennes mobiles) sont en pratique utilisées aussi bien dans le cadre d'études microéconomiques (par exemple dans la formalisation des erreurs d'un modèle panel) que macroéconomiques (détermination des multiplicateurs dynamiques). Cependant une telle pratique n'est mathématiquement justifiée que si la classe de séries temporelles retenue possède certaines propriétés d'invariances vis à vis de l'agrégation.

Dans cet article, nous nous intéressons essentiellement à l'agrégation de processus autorégressifs d'ordre 1, suivant ainsi l'exemple étudié par Granger (1980).

Dans le paragraphe 2, nous déterminons le modèle de série temporelle satisfait par la série agrégée. Ceci permet de caractériser toutes les séries temporelles linéaires pouvant s'interpréter comme agrégats de séries autorégressives d'ordre 1.

Dans le paragraphe 3, nous étudions plus en détail le cas où la distribution d'hétérogénéité du coefficient de régression est une loi bêta. On voit alors que les séries agrégées satisfont des écritures hypergéométriques et sortent donc du cadre ARMA traditionnel. En particulier il apparaît que des agrégats de séries autorégressives stationnaires peuvent être non stationnaires ou être stationnaires avec des coefficients moyennes mobiles décroissant lentement (mémoire longue).

Dans le paragraphe 4, nous comparons les corrélations temporelles de la série agrégée avec la moyenne des corrélations temporelles des séries désagrégées et vérifions que les premières sont toujours plus grandes que les secondes,

lorsque les liaisons temporelles sont positives. Les écarts entre ces deux types de corrélations fournissent alors des mesures du degré d'hétérogénéité adaptées au cas des séries temporelles.

Finalement nous examinons dans le paragraphe 5 les implications statistiques de ces résultats, proposons un test d'homogénéité et discutons le sens du biais dû à l'oubli de la composante individuelle.

II - AGREGATION DE PROCESSUS AUTOREGRESSIFS D'ORDRE 1

2.A - LE PRINCIPE

Nous suivrons la procédure usuelle d'agrégation telle qu'elle est présentée par exemple par Granger (1980). Nous considérons une suite de processus autorégressifs d'ordre 1, avec des coefficients de régression différents et des erreurs pouvant être corrélées entre elles. Cette corrélation éventuelle est prise en compte par l'intermédiaire d'une décomposition de l'erreur en effet temporel et effet croisé. De façon plus précise, on pose :

$$(1) \quad X_{it} = \varphi_i X_{i,t-1} + c_i \varepsilon_t + \eta_{it} \quad , \quad t > 0 \quad , \quad i = 1, 2, \dots$$

$$\text{avec } X_{i,-1} = 0 \quad .$$

Les bruits $(\varepsilon_t), (\eta_{1t}), (\eta_{2t}) \dots$ sont indépendants entre eux, centrés, de variances respectives $\sigma^2 = V \varepsilon_t$ et $\mu^2 = V \eta_{it}$, $\forall i$.

Remarquons que nous avons retenu une écriture autorégressive pour les seuls indices positifs. Ceci permet de considérer simultanément le cas de processus asymptotiquement stationnaires, lorsque $|\varphi_i| < 1$ et celui de processus non stationnaires, lorsque $|\varphi_i| > 1$.

Afin de mener à bien la procédure d'agrégation, nous ferons également des hypothèses sur les coefficients structurels φ_i, c_i . On suppose que ces valeurs peuvent être considérées comme résultant de tirages indépendants dans une même loi, que ces tirages sont indépendants des valeurs prises par les bruits et que les variables φ et c sont indépendantes entre elles. Le paramètre c qui permet d'introduire une hétéroscédasticité est contraint à être strictement positif.

La moyenne calculée sur les n premières séries est :

$$\begin{aligned} \bar{X}_{n,t} &= \frac{1}{n} \sum_{i=1}^n X_{i,t} \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{k=0}^t \varphi_i^k [c_i \varepsilon_{i,t-k} + \eta_{i,t-k}] \\ &= \sum_{k=0}^t \left[\frac{1}{n} \sum_{i=1}^n \varphi_i^k c_i \right] \varepsilon_{t-k} + \sum_{k=0}^t \left[\frac{1}{n} \sum_{i=1}^n \varphi_i^k \eta_{i,t-k} \right] . \end{aligned}$$

Faisant tendre n vers l'infini, on obtient :

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \varphi_i^k c_i &= E(\varphi^k c) = E(\varphi^k) (E c) , \\ \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \varphi_i^k \eta_{i,t-k} &= E(\varphi^k \eta_{i,t-k}) = E(\varphi^k) E(\eta_{i,t-k}) = 0 . \end{aligned}$$

On voit alors qu'à la limite le processus agrégé : $\bar{X}_t = \lim_{n \rightarrow \infty} \bar{X}_{n,t}$ satisfait une relation du type :

$$(2) \quad \bar{X}_t = \sum_{k=0}^t E(\varphi^k) E c \varepsilon_{t-k} = \sum_{k=0}^t E(\varphi^k) \bar{\varepsilon}_{t-k} ,$$

$$\text{avec } \bar{\varepsilon}_t = E c \varepsilon_t .$$

Nous avons mené cette agrégation de façon heuristique, sans justifier les divers passages à la limite. Il faut évidemment que certaines conditions de régularité soient satisfaites, en particulier que les divers moments $E \varphi^k$ existent.

2.B - CLASSIFICATION DES SERIES AGREGÉES

On peut distinguer plusieurs cas selon que la série agrégée est ou non (asymptotiquement) stationnaire et, lorsque la série est stationnaire, en fonction de la vitesse avec laquelle les coefficients moyennes mobiles tendent vers zéro. Ceci conduit aux propriétés suivantes :

PROPRIÉTÉ 3 :

i) La série agrégée est non (asymptotiquement) stationnaire si

$$\sum_{k=0}^{\infty} (E \varphi^k)^2 = +\infty .$$

ii) Elle est (asymptotiquement) stationnaire dans le cas contraire.

PROPRIÉTÉ 4 :

Si la série agrégée est (asymptotiquement) stationnaire :

i) elle est à mémoire longue, si $\sum_{k=0}^{\infty} |E \varphi^k| = +\infty$;

ii) elle est à mémoire courte, si $\sum_{k=0}^{\infty} |E \varphi^k| < +\infty$.

Lorsque les séries désagrégées ne sont pas toutes (p.s) asymptotiquement stationnaires, on a $P(|\varphi| > 1) > 0$. On en déduit que $E \varphi^{2k} > P(|\varphi| > 1)$ et en particulier que $\sum_{k=0}^{\infty} (E \varphi^k)^2 = +\infty$.

PROPRIETE 5 :

Une condition nécessaire pour que la série agrégée soit asymptotiquement stationnaire est que presque toutes les séries désagrégées le soient.

Comme nous le verrons dans le chapitre suivant cette condition n'est pas suffisante. Il est donc possible que la propriété de stationnarité soit perdue par agrégation.

2.C - CARACTERISATION DES SERIES AGREGÉES

Etant donnée une série $[X_t = \sum_{k=0}^t a_k \bar{\varepsilon}_{t-k}, t \in \mathbb{N}]$, on peut se demander dans quels cas elle peut être considérée comme provenant d'une agrégation de processus autorégressifs d'ordre 1. Il faut pour cela chercher des conditions pour que a_k puisse s'écrire $a_k = E \varphi^k$. De telles conditions sont classiques.

PROPRIETE 6 :

La série $[X_t = \sum_{k=0}^t a_k \bar{\varepsilon}_{t-k}, t \in \mathbb{N}]$, est une série agrégée de processus autorégressifs d'ordre 1, si et seulement si on a :

$$\left| \forall n \forall \alpha_0, \alpha_1 \dots \alpha_n \in \mathbb{R}, \sum_{i=0}^n \sum_{j=0}^n \alpha_i \alpha_j a_{i+j} \geq 0 \right.$$

Preuve : Voir [Choquet (1969), p. 279].

Cette condition correspond simplement à la positivité de la matrice de Hankel infinie :

$$\begin{pmatrix} a_0 & a_1 & a_2 & \dots & \dots \\ a_1 & a_2 & & & \\ a_2 & & & & \\ \vdots & & & & \\ \vdots & & & & \\ \vdots & & & & \end{pmatrix}$$

Elle est donc facile à vérifier en pratique à partir des mineurs principaux, puisqu'elle équivaut à :

$$\det \begin{pmatrix} a_0 & a_1 & \dots & a_i \\ a_1 & & & \vdots \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ a_i & \dots & \dots & a_{2i} \end{pmatrix} \geq 0, \quad \forall i \geq 0$$

Ainsi la première contrainte effective est :

$$\det \begin{bmatrix} a_0 & a_1 \\ a_1 & a_2 \end{bmatrix} = a_2 - a_1^2 = E \psi^2 - (E \psi)^2 \geq 0$$

Il est clair que l'ensemble des sous-déterminants peut-être considéré comme une mesure vectorielle d'hétérogénéité des coefficients de régression au niveau désagrégé. La première composante étant identique à $V \psi$.

De plus il est parfois possible de retrouver sans ambiguïté la loi du coefficient de regression φ .

PROPRIETE 7 :

Lorsque la condition de la propriété 6 est satisfaite, il existe une loi unique pour φ telle que $a_k = E \varphi^k$, $\forall k$ si et seulement si la série $\sum_{k=0}^{\infty} \frac{a_k s^k}{k!}$ est absolument convergente pour une valeur strictement positive s .

Preuve : Voir [Cramer (1945)].

On peut alors se demander dans quel cas cette loi ne charge que les valeurs φ avec $|\varphi| < 1$. Nous avons vu que si $P(|\varphi| > 1) > 0$, les coefficients $a_k = E \varphi^k$ ne tendent pas vers zero.

Inversement si $P(|\varphi| < 1) = 1$, on a :

$$\forall \epsilon > 0, \exists \eta : 1 > \eta > 0 : P(|\varphi| > 1 - \eta) < \epsilon .$$

On a alors : $E|\varphi|^k < (1 - \eta)^k (1 - \epsilon) + \epsilon \quad \forall k$.

On en déduit : $\overline{\lim}_k E|\varphi|^k < \epsilon$ et ceci étant valable pour toute valeur ϵ positive $\overline{\lim}_k E|\varphi|^k = 0$.

PROPRIETE 8 :

Lorsque la condition de la propriété 6 est satisfaite, les lois possibles de φ ne chargent que les valeurs $|\varphi| < 1$ si et seule-

ment si $\lim_{k \rightarrow \infty} a_k = 0$.

Exemple 9 : A titre d'exemple considérons un processus autoregressif pur

$$X_t = \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + \bar{\varepsilon}_t, \quad t > 0, \quad \text{avec } X_{-1} = \dots = X_{-p} = 0.$$

Dans quels cas peut-il s'interpréter comme un agrégat d'autorégressifs d'ordre 1 ? L'écriture moyenne mobile de ce processus est :

$$X_t = \sum_{k=0}^t \left(\sum_{l=1}^p \mu_l z_l^k \right) \bar{\varepsilon}_{t-k}, \quad \text{en appelant } z_l, \quad l = 1, \dots, p \text{ les}$$

inverses des racines (supposées distinctes) du polynôme autorégressif

$$\phi(z) = 1 - \varphi_1 z - \dots - \varphi_p z^p. \quad \text{La forme des coefficients moyennes-mobiles montre}$$

immédiatement que s'il s'agit d'un agrégat d'autorégressifs d'ordre 1, la

loi de φ est nécessairement discrète chargeant les valeurs z_l ,

$l = 1, \dots, p$ avec les poids μ_l , $l = 1, \dots, p$. L'interprétation en terme

d'agrégat est ainsi possible lorsque :

- i) toutes les racines du polynôme autorégressif sont réelles,
- ii) les solutions μ_l , $l = 1, \dots, p$ du système :

$$\begin{cases} \sum_{l=1}^p \mu_l = 1 \\ \sum_{l=1}^p \mu_l z_l^k - \sum_{j=1}^k \varphi_j \sum_{l=1}^p \mu_l z_l^{k-j} = 0 \quad k = 1, \dots, p-1, \end{cases}$$

qui détermine sans ambiguïté les μ_l , sont toutes positives.

Ainsi pour un processus AR(2) ce système se réduit à :

$$\left\{ \begin{array}{l} \mu_1 + \mu_2 = 1 \\ \mu_1 z_1 + \mu_2 z_2 - \varphi_1 (\mu_1 + \mu_2) = 0 \end{array} \right. \iff \left\{ \begin{array}{l} \mu_1 = \frac{z_2 - \varphi_1}{z_2 - z_1} \\ \mu_2 = \frac{\varphi_1 - z_1}{z_2 - z_1} \end{array} \right.$$

Désignant par z_2 la plus grande des inverses des deux racines, la condition est simplement : z_1, z_2 réelles telles que $z_2 > \varphi_1 > z_1$, avec au moins une inégalité stricte.

III - UN EXEMPLE D'AGREGATION

3.A - AGREGATION AVEC LOI BETA

Nous allons détailler les résultats du chapitre précédent lorsque la loi du coefficient de régression φ est une loi bêta transformée par homothétie : $B(p, q; \alpha)$, $p > 0$, $q > 0$, $\alpha > 0$. Cette loi est continue sur $[0, \alpha]$, de densité :

$$g(\varphi) = \frac{\varphi^{p-1} (1 - \varphi/\alpha)^{q-1}}{\alpha^p B(p, q)} \mathbb{1}_{(\varphi)} \quad , \quad \text{avec} \quad B(p, q) = \frac{\Gamma(p) \Gamma(q)}{\Gamma(p+q)} .$$

Les moments de cette loi sont :

$$E(\varphi^k) = \alpha^k \frac{\Gamma(p+k) \Gamma(p+q)}{\Gamma(p) \Gamma(p+q+k)} .$$

La représentation moyenne mobile du processus agrégé est donc :

$$X_t = \sum_{k=0}^t \frac{\Gamma(p+k) \Gamma(p+q)}{\Gamma(p) \Gamma(p+q+k)} \alpha^k L^k \bar{\varepsilon}_t .$$

Introduisons le processus $\varepsilon_t \approx \begin{cases} \bar{\varepsilon}_t & , \text{ si } t > 0 \\ 0 & , \text{ sinon} \end{cases}$, on a :

$$X_t = \left[\sum_{k=0}^{\infty} \frac{\Gamma(p+k)}{\Gamma(p)} \frac{\Gamma(p+q)}{\Gamma(p+q+k)} \alpha^k L^k \right] \varepsilon_t ,$$

$$X_t = \left[\sum_{k=0}^{\infty} \frac{\Gamma(p+k)}{\Gamma(p)} \frac{\Gamma(p+q)}{\Gamma(p+q+k)} \frac{\Gamma(1+k)}{\Gamma(1)} \frac{\alpha^k L^k}{k!} \right] \varepsilon_t ,$$

$$(9) \quad X_t = F[p; 1; p+q; \alpha L] \varepsilon_t ,$$

où F désigne la fonction hypergéométrique [Abramowitz-Stegun (1965)].

On obtient ainsi par agrégation de processus autorégressifs d'ordre 1, une classe de séries temporelles sur laquelle nous allons maintenant donner quelques éléments.

3.8 - CLASSIFICATION DES SERIES AGREGEES

i) Si $\alpha > 1$, nous savons que les séries désagrégées ne sont pas presque toutes stationnaires et la série agrégée est donc non stationnaire.

ii) Il nous reste à considérer le cas $\alpha < 1$.

D'après la formule de Stirling, on a : $\frac{\Gamma(p+k)}{\Gamma(p+q+k)} \approx k^{-q}$, lorsque k tend vers l'infini. Ceci permet d'obtenir un équivalent du coefficient moyenne mobile :

$$a_k = E \psi^k \approx \frac{\Gamma(p+q)}{\Gamma(p)} \alpha^k k^{-q} .$$

On en déduit alors la classification suivante pour les séries agrégées :

VALEURS DES PARAMETRES		TYPE DE SERIE AGREGEE
$\alpha > 1$		non stationnaire
$\alpha = 1$	$q < \frac{1}{2}$	non stationnaire
	$\frac{1}{2} < q < 1$	stationnaire, mémoire longue
	$1 < q$	stationnaire, mémoire courte
$\alpha < 1$		stationnaire, mémoire courte

3.C - QUELQUES AGREGATS PARTICULIERS

Comme la moyenne d'une loi $B(p,q,1)$ est égale à : $\frac{p}{p+q} = d$ et sa

variance à : $\frac{p q}{(p+q)^2 (p+q+1)}$, il est intéressant d'effectuer un changement

de paramètres pour ne faire apparaître que des paramètres facilement interprétables. Ainsi on pourrait retenir :

α donnant la valeur maximale des coefficients,

$d = \frac{p}{p+q}$ donnant la moyenne de la loi au paramètre d'homothétie près,

$\gamma = \frac{1}{p+q}$, qui permet de faire varier la variance de la loi à moyenne et coefficient d'homothétie donnés.

On voit alors que certains modèles classiques sont retrouvés comme cas particulier du modèle hypergéométrique (9) :

$$X_t = F [d/\gamma ; 1 ; 1/\gamma ; \alpha L] \varepsilon_t$$

i) Si $\gamma = 0$, c'est-à-dire si $p+q = +\infty$, la variance de φ est nulle et le modèle se réduit à un modèle autorégressif d'ordre 1 :

$$\bar{X}_t = [1 - d \alpha L]^{-1} \varepsilon_t$$

Lorsque $\gamma = 0$, on ne peut évidemment identifier séparément les deux autres paramètres.

ii) Si $\gamma = p+q = 1$, ce qui implique $0 < p < 1$, on a :

$$\begin{aligned} \bar{X}_t &= F[p; 1; \alpha L] \varepsilon_t \\ &= \sum_{k=0}^{\infty} \frac{\Gamma(p+k)}{\Gamma(p)} \frac{\alpha^k L^k}{k!} \varepsilon_t \\ &= \sum_{k=0}^{\infty} \frac{-p(-p-1)\dots(-p-k+1)}{k!} (-1)^k \alpha^k L^k \varepsilon_t, \end{aligned}$$

$$\langle == \rangle \bar{X}_t = (1 - \alpha L)^{-p} \varepsilon_t$$

Le modèle obtenu est un autorégressif fractionnaire [voir Granger-Joyeux (1980), Mandelbrot-Van Ness (1968), Gonçalves (1987)...] :

$$(1 - \alpha L)^p \bar{X}_t = \varepsilon_t, \text{ avec } 0 < p < 1.$$

Si $\alpha = 1$, on retrouve en particulier le modèle ARIMA $(\alpha, p, 0)$.

Ce modèle est non stationnaire si $\alpha = 1$ et $p = 1 - q > \frac{1}{2}$.

iii) Finalement remarquant que : $E \varphi = \alpha d$

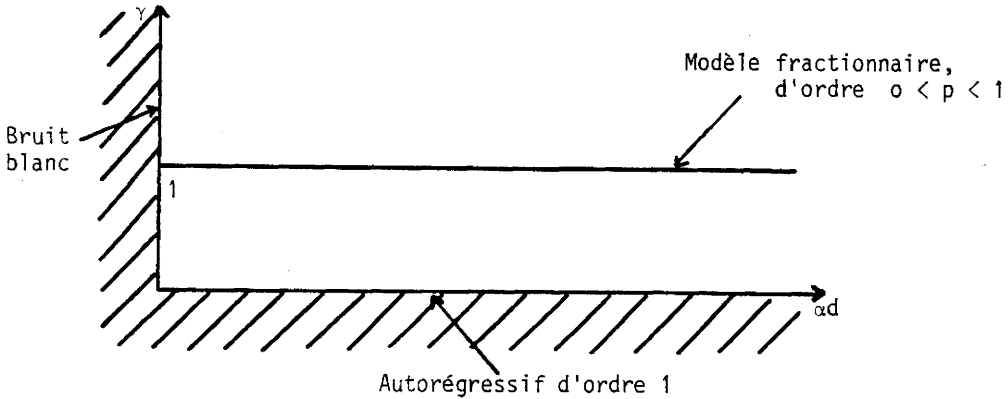
$$V \varphi = \alpha^2 d(1-d) \frac{\gamma}{1+\gamma},$$

on voit que $E \varphi = V \varphi = 0 \langle == \rangle \alpha d = 0$.

Le cas limite du bruit blanc correspond donc à la contrainte : $\alpha d = 0$.

Dans l'espace $(\gamma, \alpha d)$ ces divers cas particuliers apparaissent de la façon

suivante :



3.D - AGREGATIONS SUCCESSIVES

Nous avons vu que par agrégation de processus autorégressifs d'ordre 1 , on obtenait un processus tel que :

$$\bar{X}_t = \sum_{k=0}^{\infty} E(\varphi^k) \varepsilon_{t-k} .$$

En particulier, dès que la loi de φ est non dégénérée, on a $E \varphi^2 \neq (E \varphi)^2$ et le processus \bar{X} n'a pas de représentation autorégressive d'ordre 1 . La classe des processus AR(1) ne possède donc aucune propriété de stabilité par agrégation.

On peut maintenant regarder si de telles propriétés peuvent dans certains cas exister pour les processus à représentation hypergéométrique.

La question s'écrit mathématiquement : existe-t-il une loi de probabilité pour p, q, α telle que :

$$E (F[p; 1; p+q; \alpha L])$$

soit de la forme $F[\bar{p}; 1; \bar{p}+\bar{q}; \bar{\alpha} L]$?

Si nous notons :

$$G(p,q,\alpha;x) = \frac{1}{B(p,q)} \int_0^x \frac{\varphi^{p-1} (1 - \varphi/\alpha)^{q-1}}{\alpha^p} d\varphi, \quad 0 \leq x \leq \alpha.$$

on obtient en intégrant par parties :

$$G(p,q,\alpha;x) = \frac{p}{p+q} G(p+1,q,\alpha;x) + \frac{q}{p+q} G(p,q+1,\alpha;x).$$

Cette relation montre qu'étant donnés des processus hypergéométriques :

$X_{jt} = F(p_j; 1; p_j + q_j; \alpha L) \varepsilon_t^*$, $j = 1, \dots, J$ avec $p_j + q_j = \delta$ indépendant de j et $p_{j+1} = p_j + 1$, on peut trouver des poids β_j , $j = 1, \dots, J$ de

façon que le processus agrégé $\sum_{j=1}^J \beta_j X_{jt}$ admette aussi une représentation hypergéométrique de paramètres $\bar{p} = \min_j p_j$, $\bar{p} + \bar{q} = \delta$, $\bar{\alpha} = \alpha$.

IV - ETUDE DES CORRELATIONS

4.A - ACCROISSEMENT DES CORRELATIONS

Nous avons vu que la procédure d'agrégation avait pour effet d'augmenter l'ordre des retards et de faire se "rapprocher" la représentation de la non stationnarité. De tels effets devraient également se retrouver au niveau de la fonction d'autocorrélation.

Au niveau désagrégé la fonction d'autocorrélation est :

$$\rho_i^h = \varphi_i^h, \quad i = 1, 2, \dots, \quad h = 0, 1, \dots$$

et donc la moyenne de ces fonctions d'autocorrélation est :

$$(10) \quad \bar{\rho}^h = E \varphi^h, \quad h = 0, 1, \dots$$

Calculons maintenant la fonction d'autocorrélation associée à la série agrégée.

Ce calcul est possible lorsque cette série est stationnaire ce que nous supposons. On obtient :

$$(11) \quad \tilde{\rho}(h) = \frac{\sum_{k=0}^{\infty} E(\varphi^k) E(\varphi^{k+h})}{\sum_{k=0}^{\infty} [E(\varphi^k)]^2}, \quad \forall h \in \mathbb{N}.$$

Si la loi du coefficient de régression charge uniquement les réels positifs les fonctions $\varphi \rightarrow \varphi^k$, $k > 0$ et $\varphi \rightarrow \varphi^h$, $h > 0$ sont croissantes et donc :

$$\text{Cov}(\varphi^k, \varphi^h) = E(\varphi^{k+h}) - E(\varphi^k) E(\varphi^h) > 0.$$

On en déduit que :

$$\tilde{\rho}(h) > E \varphi^h = \bar{\rho}(h).$$

PROPRIETE 12 :

Si la loi de φ charge les valeurs positives, les corrélations au niveau agrégé sont supérieures aux moyennes des corrélations désagrégées.

4.8 - CAS DES PROCESSUS FRACTIONNAIRES

Considérons par exemple une distribution d'hétérogénéité de type $B(p,q;\alpha)$ avec $p+q = 1$, qui correspond à un processus agrégé satisfaisant :

$$(1 - \alpha L)^p X_t = \tilde{\varepsilon}_t.$$

$$\text{Nous avons : } E \varphi^k = \frac{\Gamma(p+k)}{\Gamma(p)} \frac{\alpha^k}{k!} = \frac{p(p+1)\dots(p+k-1)}{k!} \alpha^k.$$

D'autre part, les moments croisés du processus agrégé se calculent à partir

des sommes :

$$\begin{aligned} & \sum_{k=0}^{\infty} E(\varphi^k) E(\varphi^{k+h}) \\ &= \sum_{k=0}^{\infty} \frac{\Gamma(p+k)}{\Gamma(p)} \frac{\Gamma(p+k+h)}{\Gamma(p)} \frac{\alpha^{2k+h}}{k! (k+h)!} \\ &= \alpha^h \frac{\Gamma(p+h)}{\Gamma(p) \Gamma(h+1)} \sum_{k=0}^{\infty} \frac{\Gamma(p+k)}{\Gamma(p)} \frac{\Gamma(p+k+h)}{\Gamma(p+h)} \frac{\Gamma(h+1)}{\Gamma(k+h+1)} \frac{\alpha^{2k}}{k!} \\ &= \alpha^h \frac{\Gamma(p+h)}{\Gamma(p) \Gamma(h+1)} F[\rho, \rho+h, h+1; \alpha^2] \end{aligned}$$

On en déduit :

$$\bar{q}(h) = \frac{\alpha^h \Gamma(p+h)}{\Gamma(p) \Gamma(h+1)} \frac{F[\rho, \rho+h, h+1; \alpha^2]}{F[\rho, \rho, 1; \alpha^2]}$$

c'est-à-dire :

$$(13) \quad \frac{\bar{q}(h)}{\bar{q}(h)} = \frac{F[\rho, \rho+h, h+1; \alpha^2]}{F[\rho, \rho, 1; \alpha^2]}$$

Ce rapport est comme nous l'avons vu auparavant supérieur à 1 .

On peut d'ailleurs remarquer que :

$$\frac{\bar{q}(h)}{\bar{q}(h)} = \sum_{k=0}^{\infty} \beta_k^{(h)} \alpha^{2k}$$

$$\text{et que : } \frac{\beta_k^{(h+1)}}{\beta_k^{(h)}} = \frac{\rho+k+h}{\rho+h} \frac{h+1}{k+h+1} > 1, \text{ car } \rho < 1 .$$

On en déduit que le rapport $\frac{\bar{q}(h)}{\bar{q}(h)}$ est fonction croissante de h .

D'autre part, utilisant l'équivalence asymptotique :

$$F[a, b+\lambda, c+\lambda, z] \approx \frac{1}{(1-z)^a}, \text{ si } |z| < 1 \text{ et si } \lambda \text{ tend vers l'infini,}$$

on voit que ce rapport est asymptotiquement donné par :

$$(14) \quad \lim_{h \rightarrow \infty} \frac{\bar{q}(h)}{\underline{q}(h)} = \frac{1}{(1-\alpha)^2 F(p, p, 1; \alpha^2)}$$

V - TEST DE L'HYPOTHESE D'HOMOGENEITE

5.A - DERIVATION DE LA STATISTIQUE DU MULTIPLICATEUR DE LAGRANGE

Plaçons nous de nouveau dans le cas d'une agrégation de processus autorégressifs d'ordre 1 effectuée selon une loi bêta. Nous avons vu que la série agrégée satisfait une représentation hypergéométrique :

$$X_t = F\left(\frac{d}{\gamma}; 1; \frac{1}{\gamma}; \alpha L\right) \varepsilon_t,$$

où γ est un paramètre positif directement lié à la variance du coefficient d'autorégression. Dans ce paragraphe, on se propose de développer un test d'homogénéité, c'est-à-dire un test de l'hypothèse nulle $H_0 = \langle \gamma=0 \rangle$ contre l'alternative $H_1 = \langle \gamma > 0 \rangle$. Nous avons noté précédemment que sous l'hypothèse d'homogénéité le paramètre α n'est pas identifiable ; afin de résoudre cette difficulté, nous développons le test sous la contrainte $\alpha = 1$, c'est-à-dire pour le modèle : $X_t = F\left(\frac{d}{\gamma}; 1; \frac{1}{\gamma}; L\right) \varepsilon_t$. On suppose de plus les composantes du bruit indépendantes de même loi.

Nous fondons le test sur le principe du multiplicateur de Lagrange. Notant $\diamond_{d,\gamma} (L) X_t = \varepsilon_t$ la forme autorégressive du processus, la statistique du multiplicateur de Lagrange fait intervenir l'estimateur du maximum de vraisemblance des paramètres calculé sous l'hypothèse nulle :

$$\hat{\alpha}_T^0 = \frac{\sum_{t=1}^T X_t X_{t-1}}{\sum_{t=1}^T X_{t-1}^2} \quad (\text{puisque sous } H_0, d \text{ est le coefficient d'un autorégressif d'ordre 1}).$$

$$\hat{\gamma}_T^0 = 0, \quad \hat{\sigma}_T^{o2} = \frac{1}{T} \sum_{t=1}^T (X_t - \hat{\alpha}_T^0 X_{t-1})^2 = \frac{1}{T} \sum_{t=1}^T (\tilde{\varepsilon}_t^0)^2,$$

où $\tilde{\varepsilon}_t^0$ désigne le résidu sous l'hypothèse nulle.

Elle s'exprime alors en fonction des scores estimés sous contraintes :

$$Z_{1t} = \frac{\partial \phi_{T,0}^{(L)}(\tilde{X}_t)}{\partial d} \quad \text{et} \quad Z_{2t} = \frac{\partial \phi_{T,0}^{(L)}(\tilde{X}_t)}{\partial \gamma},$$

où $\tilde{X}_t = X_t$, si $t > 0$, $\tilde{X}_t = 0$ sinon. Elle est donnée par [voir page 192]

$$(15) \quad \xi_{L.M} = \frac{1}{\hat{\sigma}_T^{o2}} \left[\sum_{t=1}^T Z_{2t} \tilde{\varepsilon}_t^0 \right]^2 / \left[\sum_{t=1}^T Z_{2t}^2 - \frac{\left[\sum_{t=1}^T Z_{2t} Z_{1t} \right]^2}{\sum_{t=1}^T Z_{1t}^2} \right]$$

et suit asymptotiquement sous l'hypothèse d'homogénéité une loi du khi-deux à un degré de liberté.

Il nous reste à expliciter la forme des scores.

i) On a :

$$\frac{\partial \phi_{d,\gamma}^{(L)}}{\partial d} = - \frac{\frac{\partial}{\partial d} F\left(\frac{d}{\gamma}; 1; \frac{1}{\gamma}; L\right)}{F^2\left(\frac{d}{\gamma}; 1; \frac{1}{\gamma}; L\right)}$$

Le terme général du développement en série de la fonction hypergéométrique

$F\left(\frac{d}{\gamma}; 1; \frac{1}{\gamma}; L\right)$ est :

$$\frac{\Gamma\left(\frac{d}{\gamma} + i\right)}{\Gamma\left(\frac{d}{\gamma}\right)} \frac{\Gamma\left(\frac{1}{\gamma}\right)}{\Gamma\left(\frac{1}{\gamma} + i\right)} L^i .$$

On a :

$$\frac{\partial}{\partial d} \text{Log} \frac{\Gamma\left(\frac{d}{\gamma} + i\right)}{\Gamma\left(\frac{d}{\gamma}\right)} \frac{\Gamma\left(\frac{1}{\gamma}\right)}{\Gamma\left(\frac{1}{\gamma} + i\right)} = \frac{1}{\gamma} \left[\Psi\left(\frac{d}{\gamma} + i\right) - \Psi\left(\frac{d}{\gamma}\right) \right] ,$$

où $\Psi(z) = \frac{\Gamma'(z)}{\Gamma(z)}$ est la fonction digamma. Utilisant les propriétés de la fonction digamma [Abramowitz - Stegun (1965) 6.3.6], on a :

$$\frac{\partial}{\partial d} \left[\text{Log} \frac{\Gamma\left(\frac{d}{\gamma} + i\right)}{\Gamma\left(\frac{d}{\gamma}\right)} \frac{\Gamma\left(\frac{1}{\gamma}\right)}{\Gamma\left(\frac{1}{\gamma} + i\right)} \right] = \frac{1}{\gamma} \left[\sum_{j=0}^{i-1} \frac{1}{\frac{d}{\gamma} + j} \right] = \sum_{j=0}^{i-1} \frac{1}{d+j\gamma}$$

et :

$$\frac{\partial}{\partial d} \left[\frac{\Gamma\left(\frac{d}{\gamma} + i\right)}{\Gamma\left(\frac{d}{\gamma}\right)} \frac{\Gamma\left(\frac{1}{\gamma}\right)}{\Gamma\left(\frac{1}{\gamma} + i\right)} L^i \right] = \frac{\Gamma\left(\frac{d}{\gamma} + i\right)}{\Gamma\left(\frac{d}{\gamma}\right)} \frac{\Gamma\left(\frac{1}{\gamma}\right)}{\Gamma\left(\frac{1}{\gamma} + i\right)} \sum_{j=0}^{i-1} \frac{1}{d+j\gamma} L^i .$$

On en déduit la forme du score relatif à d , évalué en $\gamma = 0$:

$$\begin{aligned} \frac{\partial \phi_{d,0}(L)}{\partial d} &= - (1 - dL)^2 \sum_{i=1}^{\infty} i d^{i-1} L^i \\ &= - (1 - dL)^2 \frac{L}{(1 - dL)^2} = - L . \end{aligned}$$

Finalement, on a :

$$(16) \quad Z_{1t} = \frac{\frac{\partial \phi}{\partial d} (L)}{\hat{\alpha}_T^0} \tilde{X}_t = - \tilde{X}_{t-1} .$$

ii) Calculons de même l'autre composante du vecteur des scores. Après calcul (voir annexe), on trouve :

$$\frac{\partial \phi_{d,o} (L)}{\partial \gamma} = d(d-1) \frac{L^2}{1-dL} .$$

La variable Z_{2t} est donc :

$$Z_{2t} = \hat{\alpha}_T^0 (\hat{\alpha}_T^0 - 1) \frac{L^2}{1 - \hat{\alpha}_T^0 L} \tilde{X}_t ,$$

$$Z_{2t} = \hat{\alpha}_T^0 (\hat{\alpha}_T^0 - 1) (1 - \hat{\alpha}_T^0 L)^{-1} \tilde{X}_{t-2} .$$

iii) La statistique du multiplicateur de Lagrange s'écrit finalement :

$$(17) \quad \epsilon_{L.M} = \frac{1}{\hat{\sigma}_T^2} \frac{\left[\sum_{t=1}^T (1 - \hat{\alpha}_T^0 L)^{-1} \tilde{X}_{t-2} \cdot \tilde{\epsilon}_t \right]^2}{\sum_{t=1}^T \left[(1 - \hat{\alpha}_T^0 L)^{-1} \tilde{X}_{t-2} \right]^2 - \left[\sum_{t=1}^T (1 - \hat{\alpha}_T^0 L)^{-1} \tilde{X}_{t-2} \cdot \tilde{X}_{t-1} \right]^2 / \sum_{t=1}^T \tilde{X}_{t-1}^2}$$

5.8 - UNE STATISTIQUE DE TEST ASYMPTOTIQUEMENT EQUIVALENTE

Une autre forme de cette statistique équivalente sous l'hypothèse nulle s'en déduit en approchant le dénominateur. On a après calcul :

$$\sum_{t=1}^T \left[(1 - \hat{d}_T^0)^{-1} \tilde{X}_{t-2} \right]^2 \approx T \frac{\hat{\sigma}_T^{02}}{(1 - \hat{d}_T^{02})^3}$$

$$\sum_{t=1}^T \tilde{X}_{t-1} (1 - \hat{d}_T^0)^{-1} \tilde{X}_{t-2} \approx T \frac{\hat{d}_T^0}{(1 - \hat{d}_T^{02})^2}$$

$$\sum_{t=1}^T \tilde{X}_{t-1}^2 \approx T \frac{\hat{\sigma}_T^{02}}{1 - \hat{d}_T^{02}}$$

Le dénominateur peut donc être approché par :

$$T \frac{\hat{\sigma}_T^{02}}{(1 - \hat{d}_T^{02})^3} - \frac{T^2 \hat{\sigma}_T^{04} \hat{d}_T^{02} (1 - \hat{d}_T^{02})}{(1 - \hat{d}_T^{02})^4 T \hat{\sigma}_T^{02}}$$

$$= T \frac{\hat{\sigma}_T^{02}}{(1 - \hat{d}_T^{02})^3} [1 + \hat{d}_T^{02} - \hat{d}_T^{02}] = \frac{T \hat{\sigma}_T^{02}}{(1 - \hat{d}_T^{02})^3}$$

Une statistique équivalente à $\xi_{L.M}$ sous l'hypothèse nulle est donc :

$$(18) \quad \xi_{L.M}^* = \frac{(1 - \hat{d}_T^{02})^3}{\hat{\sigma}_T^{04}} \frac{1}{T} \left[\sum_{t=1}^T (1 - \hat{d}_T^0)^{-1} \tilde{X}_{t-2} \cdot \tilde{\varepsilon}_t \right]^2$$

5.C - L'ASPECT UNILATERAL DU TEST

Les statistiques que nous venons de calculer sont celles habituellement utilisées dans le cas d'un test bilatéral de $H_0^* = (\gamma = 0)$ contre

$H_1 = \langle \gamma \neq 0 \rangle$. Dans notre application il faut tenir compte de l'aspect unilatéral du test. Ceci conduit à modifier les statistiques de test en prenant :

$$(19) \quad T_{LM} = -s \sqrt{\hat{\epsilon}_{LM}} \quad \text{et} \quad T_{LM}^* = s \sqrt{\hat{\epsilon}_{LM}^*}, \quad \text{où } s \text{ désigne le}$$

signe de $\sum_{t=1}^T (1 - \hat{\alpha}_T^0 L)^{-1} \hat{X}_{t-2} \cdot \hat{\epsilon}_t^{\approx 0}$. Ainsi on a :

$$T_{LM}^* = \frac{(1 - \hat{\alpha}_T^{o2})^{3/2}}{\hat{\sigma}_T^{o2}} \frac{1}{\sqrt{T}} \sum_{t=1}^T (1 - \hat{\alpha}_T^0 L)^{-1} \hat{X}_{t-2} \cdot \hat{\epsilon}_t^{\approx 0}.$$

Ces deux statistiques ont asymptotiquement une loi normale centrée réduite sous l'hypothèse nulle $H_0 = \langle \gamma = 0 \rangle$.

Nous allons maintenant étudier leur comportement sous une suite d'alternatives locales du type $H_{1T} = \left\{ \gamma_T = \frac{\gamma_0}{\sqrt{T}} \right\}$, avec $\gamma_0 > 0$.

Pour cette suite le modèle initial :

$$\frac{1}{F \left(\frac{d}{\gamma_T}; 1; \frac{1}{\gamma_T}; L \right)} \bar{X}_t = \phi_{d, \gamma_T}^{(L)} \bar{X}_t = \hat{\epsilon}_t^{\approx 0},$$

peut être approché par :

$$\left(\phi_{d,0}^{(L)} + \frac{\partial \phi_{d,0}^{(L)}}{\partial \gamma} \frac{\gamma_0}{\sqrt{T}} \right) \bar{X}_t = \hat{\epsilon}_t^{\approx 0},$$

ou par :

$$X_t = \left(\frac{1}{\phi_{d,o}^{(L)}} - \frac{\frac{\partial \phi_{d,o}^{(L)}}{\partial \gamma}}{\phi_{d,o}^{(L)2}} \frac{\gamma_0}{\sqrt{T}} \right) \varepsilon_t$$

$$\Leftrightarrow X_t = \left[(1-dL)^{-1} + d(1-d)L^2(1-dL)^{-3} \frac{\gamma_0}{\sqrt{T}} \right] \varepsilon_t$$

$$\Leftrightarrow X_t - dX_{t-1} = \varepsilon_t + \frac{d(1-d)L^2}{(1-dL)^2} \frac{\gamma_0}{\sqrt{T}} \varepsilon_t$$

Sous cette suite d'alternatives locales, l'estimateur contraint du paramètre d peut être approché par :

$$\begin{aligned} \hat{d}_T^o &= \frac{\sum_t X_t X_{t-1}}{\sum_t X_{t-1}^2} \\ &= d + \frac{\frac{1}{T} \sum_t \left(\varepsilon_t + \frac{d(1-d)L^2}{(1-dL)^2} \frac{\gamma_0}{\sqrt{T}} \varepsilon_t \right) X_{t-1}}{\frac{1}{T} \sum_t X_{t-1}^2} + o_p(1) \\ &= d + o_p(1) \end{aligned}$$

Le numérateur de la statistique T_{LM} ou T_{LM}^* peut être approché par :

$$\begin{aligned} &\frac{1}{\sqrt{T}} \sum_{t=1}^T (1-d_T^o)^{-1} X_{t-2} \varepsilon_t \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^T (1-dL)^{-1} X_{t-2} (X_t - dX_{t-1}) + o_p(1) \end{aligned}$$

$$= \frac{1}{\sqrt{T}} \sum_{t=1}^T (1-dL)^{-1} X_{t-2} \left(\tilde{\varepsilon}_t + \frac{d(1-d)}{(1-dL)^2} \frac{\gamma_0}{\sqrt{T}} \tilde{\varepsilon}_{t-2} \right) + o_p(1)$$

$$= \frac{1}{T} \sum_{t=1}^T (1-dL)^{-1} X_{t-2} \gamma_0 \frac{d(1-d)}{(1-dL)^2} \tilde{\varepsilon}_{t-2} + o_p(1),$$

(car $\tilde{\varepsilon}_t$ est indépendant des valeurs X_{t-2}, X_{t-3}, \dots).

Ne conservant que le terme d'ordre supérieur dans l'écriture de X_{t-2} , on voit que le numérateur est pour la suite d'alternatives locales équivalent à :

$$\frac{1}{T} \sum_{t=1}^T (1-dL)^{-2} \tilde{\varepsilon}_{t-2} \gamma_0 \frac{d(1-d)}{(1-dL)^2} \tilde{\varepsilon}_{t-2} + o_p(1)$$

$$= \gamma_0 d(1-d) V((1-dL)^{-2} \tilde{\varepsilon}_{t-2}).$$

Comme $\gamma_0 > 0$ et que $0 < d < 1$ le numérateur converge donc vers une constante positive.

La région critique du test unilatéral est donc de la forme :

$$W_T = \left\{ T_{LM} > c \right\} \quad \text{ou} \quad W_T^* = \left\{ T_{LM}^* > c \right\},$$

où c est déterminée par la condition asymptotique sur le risque de première espèce. Si ce risque est fixé à α_0 , on doit avoir : $\lim_T P(W_T) = \alpha_0$
 $\Leftrightarrow 1 - u(c) = \alpha_0 \Leftrightarrow c = u^{-1}(1 - \alpha_0)$, où u désigne la fonction de répartition de la loi normale centrée réduite.

PROPRIÉTÉ 20 :

Le test du multiplicateur de Lagrange au seuil α_0 de l'hypothèse d'homogénéité dans le cadre d'un modèle hypergéométrique consiste à :
 accepter l'hypothèse d'homogénéité si :

$$T_{LM} < u^{-1}(1-\alpha_0) \quad \left[\text{ou si } T_{LM}^* < u^{-1}(1-\alpha_0) \right] ,$$

à la refuser dans le cas contraire.

5.D - BIAIS D'HETEROGENEITE

Finalement, on peut se demander quelles seraient les propriétés des estimateurs contraints par l'hypothèse d'homogénéité, lorsque celle-ci n'est pas satisfaite.

L'estimateur contraint du paramètre d est comme nous l'avons vu donné par :

$$\hat{d}_T^o = \frac{\sum_t X_t X_{t-1}}{\sum_t X_{t-1}^2} ;$$

et il converge donc vers la corrélation théorique à l'ordre 1 du processus X , c'est-à-dire vers $\tilde{\rho}(1)$.

On a vu d'autre part dans la propriété 12 que $\tilde{\rho}(1) > E \varphi$.

Comme d s'interprète justement comme la moyenne de φ , on en déduit l'existence d'un biais positif d'hétérogénéité.

PROPRIETE 21 :

L'estimateur du maximum de vraisemblance de d calculé sous l'hypothèse d'homogénéité est biaisé, lorsqu'il y a hétérogénéité, ce biais est toujours positif.

Ainsi l'oubli de l'hétérogénéité conduit à une surestimation systématique de ce coefficient.

Ce biais peut être explicité en fonction des paramètres. On a :

$$\hat{\rho}(1) - d = \frac{\sum_{i=0}^{\infty} a_i a_{i+1} - d \sum_{i=0}^{\infty} a_i^2}{\sum_{i=0}^{\infty} a_i^2}, \text{ où } a_i \text{ désigne le } i^{\text{e}} \text{ coefficient}$$

du développement moyenne mobile de \bar{X} , c'est-à-dire :

$$a_i = \frac{\Gamma\left(\frac{d}{\gamma} + i\right) \Gamma\left(\frac{1}{\gamma}\right)}{\Gamma\left(\frac{d}{\gamma}\right) \Gamma\left(\frac{1}{\gamma} + i\right)}$$

Comme : $a_{i+1} = \frac{\frac{d}{\gamma} + i}{\frac{1}{\gamma} + i} a_i$, on a :

$$\begin{aligned} \hat{\rho}(1) - d &= \frac{\sum_{i=0}^{\infty} a_i (a_{i+1} - d a_i)}{\sum_{i=0}^{\infty} a_i^2} \\ &= \frac{(1-d) \sum_{i=0}^{\infty} a_i^2 \frac{i}{\frac{1}{\gamma} + i}}{\sum_{i=0}^{\infty} a_i^2} \end{aligned}$$

On voit que ce biais est une fonction croissante de $1 - d$ et donc le biais tend à s'annuler, lorsqu'on se rapproche de la non stationnarité.

ANNEXE

DERIVEE DE $\phi_{d,\gamma}^{(L)}$ PAR RAPPORT A γ EN $\gamma = 0$

On a :

$$\phi_{d,\gamma}^{(L)} = \frac{1}{F\left(\frac{d}{\gamma}; 1; \frac{1}{\gamma}; L\right)} \quad \text{et} \quad \phi_{d,0}^{(L)} = (1 - dL)$$

On en déduit :

$$\begin{aligned} \frac{\phi_{d,\gamma}^{(L)} - \phi_{d,0}^{(L)}}{\gamma} &= \frac{\frac{1}{F\left(\frac{d}{\gamma}; 1; \frac{1}{\gamma}; L\right)} - (1 - dL)}{\gamma} \\ &= \frac{1 - F\left(\frac{d}{\gamma}; 1; \frac{1}{\gamma}; L\right) + dL F\left(\frac{d}{\gamma}; 1; \frac{1}{\gamma}; L\right)}{\gamma F\left(\frac{d}{\gamma}; 1; \frac{1}{\gamma}; L\right)} \\ &= \frac{-\sum_{i=1}^{\infty} \frac{r\left(\frac{d}{\gamma} + i\right)}{r\left(\frac{d}{\gamma}\right)} \frac{r\left(\frac{1}{\gamma}\right)}{r\left(\frac{1}{\gamma} + i\right)} L^i + dL \sum_{i=0}^{\infty} \frac{r\left(\frac{d}{\gamma} + i\right)}{r\left(\frac{d}{\gamma}\right)} \frac{r\left(\frac{1}{\gamma}\right)}{r\left(\frac{1}{\gamma} + i\right)} L^i}{\gamma F\left(\frac{d}{\gamma}; 1; \frac{1}{\gamma}; L\right)} \\ &= \frac{\sum_{i=1}^{\infty} \left[d \frac{r\left(\frac{d}{\gamma} + i - 1\right)}{r\left(\frac{d}{\gamma}\right)} \frac{r\left(\frac{1}{\gamma}\right)}{r\left(\frac{1}{\gamma} + i - 1\right)} - \frac{r\left(\frac{d}{\gamma} + i\right)}{r\left(\frac{d}{\gamma}\right)} \frac{r\left(\frac{1}{\gamma}\right)}{r\left(\frac{1}{\gamma} + i\right)} \right] L^i}{\gamma F\left(\frac{d}{\gamma}; 1; \frac{1}{\gamma}; L\right)} \end{aligned}$$

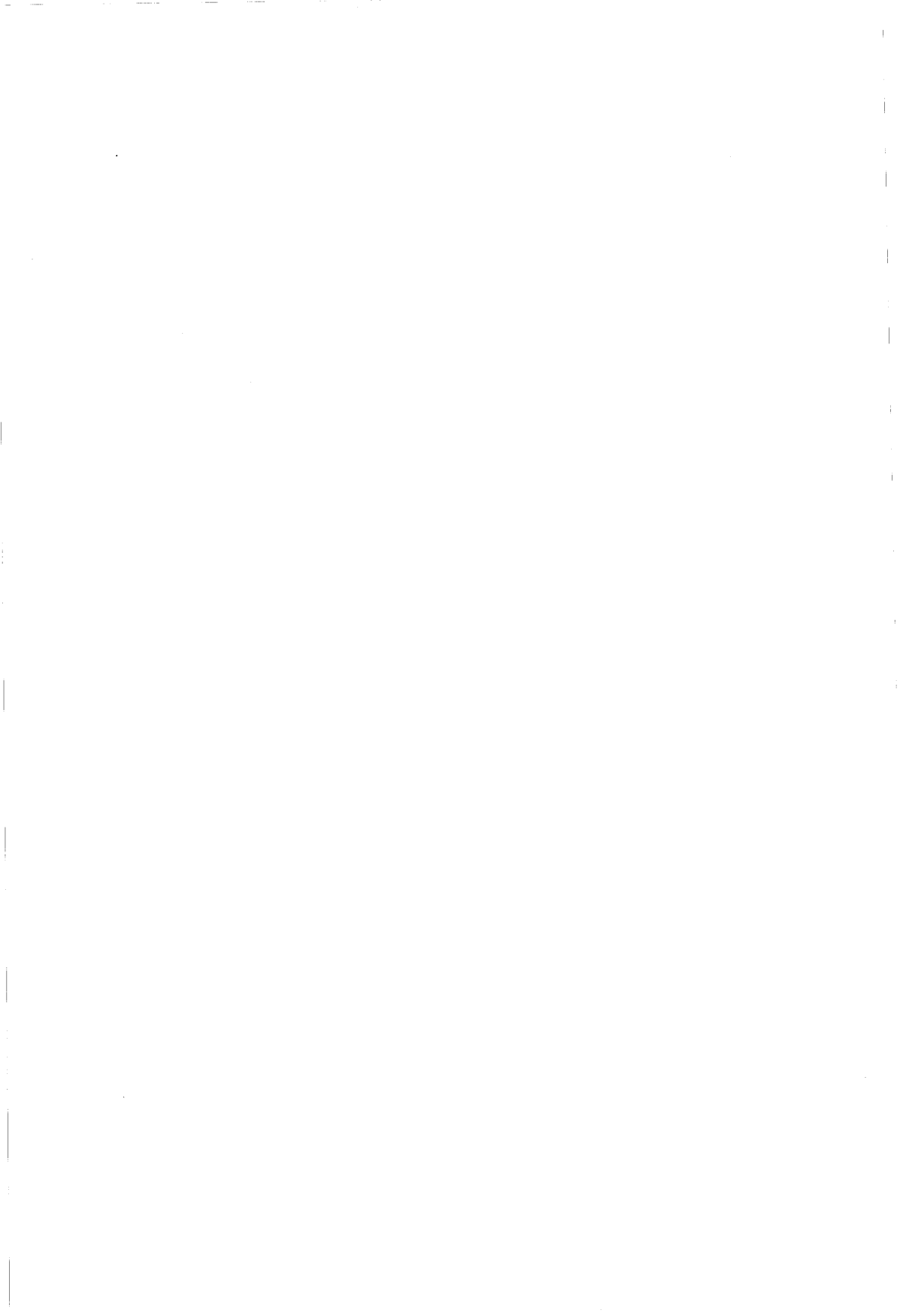
$$\begin{aligned}
 & \sum_{i=1}^{\infty} \frac{\Gamma\left(\frac{d}{\gamma} + i - 1\right)}{\Gamma\left(\frac{d}{\gamma}\right)} \frac{\Gamma\left(\frac{1}{\gamma}\right)}{\Gamma\left(\frac{1}{\gamma} + i - 1\right)} \left[d - \frac{\frac{d}{\gamma} + i - 1}{\frac{1}{\gamma} + i - 1} \right] L^i \\
 = & \frac{\quad}{\gamma F\left(\frac{d}{\gamma}; 1; \frac{1}{\gamma}; L\right)} \\
 & \sum_{i=1}^{\infty} \frac{\Gamma\left(\frac{d}{\gamma} + i - 1\right)}{\Gamma\left(\frac{d}{\gamma}\right)} \frac{\Gamma\left(\frac{1}{\gamma}\right)}{\Gamma\left(\frac{1}{\gamma} + i - 1\right)} \frac{(i-1)(d-1)}{\frac{1}{\gamma} + i - 1} L^i \\
 = & \frac{\quad}{\gamma F\left(\frac{d}{\gamma}; 1; \frac{1}{\gamma}; L\right)} \\
 & \sum_{i=1}^{\infty} \frac{\Gamma\left(\frac{d}{\gamma} + i - 1\right)}{\Gamma\left(\frac{d}{\gamma}\right)} \frac{\Gamma\left(\frac{1}{\gamma}\right)}{\Gamma\left(\frac{1}{\gamma} + i - 1\right)} \frac{i-1}{1 + \gamma(i-1)} L^i \\
 = & (d-1) \frac{\quad}{F\left(\frac{d}{\gamma}; 1; \frac{1}{\gamma}; L\right)} .
 \end{aligned}$$

Faisant tendre γ vers zéro, on obtient :

$$\begin{aligned}
 \frac{\partial \phi_{d,0} (L)}{\partial \gamma} &= (d-1) (1-dL) \sum_{i=1}^{\infty} d^{i-1} (i-1) L^i \\
 &= (d-1) (1-dL) \frac{d L^2}{(1-dL)^2} \\
 &= \frac{d(d-1) L^2}{1-dL} .
 \end{aligned}$$

REFERENCES

- Abramowitz, M. et I., Stegun (1965) : Handbook of Mathematical functions, Dover.
- Choquet, G. (1969) : Lectures on Analysis, Vol. II, Representation theory, W.A. Benjamin, Inc., New-York.
- Cramer, H. (1945) : Mathematical methods of statistics, Princeton University Press.
- Gonçalves, E. (1987) : Une généralisation des processus ARMA, Annales d'Economie et de Statistique, 5, 109-146.
- Granger, C. (1980) : Long memory relationships and the aggregation of dynamic models, Journal of Econometrics, 14, 227-238.
- Granger, C. et R., Joyeux (1980) : An introduction to long memory time series and fractional differencing, Journal of Time Series Analysis, 1, 15-29.
- Mandelbrot, B.B. et J.W. Van Ness (1968) : Fractional brownian motions, fractional noises and applications, SIAM Review, 10, 422-437.



G

TESTS DE BRUIT BLANC



I - INTRODUCTION.

Depuis leur présentation par Aitchison et Silvey (1958), Silvey (1959), les tests du multiplicateur de Lagrange ont été largement utilisés en statistique. Ils se révèlent en effet simples d'emploi, tout en possédant certaines propriétés d'optimalité locale.

Leur introduction en séries temporelles est cependant assez récente (Hosking (1980), Godfrey (1981), Godfrey-Wickens (1982)) ; il a en effet été nécessaire de généraliser les résultats initiaux d'Aitchison et Silvey, qui avaient été établis dans un cadre d'échantillonnage. D'autre part les premières applications de cette procédure ont essentiellement permis de retrouver des tests qui avaient été construits d'une autre manière, par exemple le test de Durbin-Watson ou celui dit test "portmanteau".

Dans les deux paragraphes qui suivent nous nous proposons de développer cette approche du multiplicateur de Lagrange pour tester l'hypothèse qu'un processus $X = (X_t, t \in Z)$ est un bruit blanc indépendant : $X_t = \epsilon_t$. Sous l'hypothèse générale le processus est supposé admettre une représentation moyenne mobile infinie inversible dépendant d'un paramètre scalaire θ . Donnée sous forme autorégressive cette représentation est :

$$\phi_{\theta}(L) X_t = \epsilon_t, \quad t \in Z, \quad \text{où } \epsilon = (\epsilon_t, t \in Z) \text{ est un bruit}$$

blanc indépendant de variance σ^2 et où $\phi_{\theta}(L)$ est un opérateur de retard paramétré par θ .

Par convention le paramètre est défini de façon que l'hypothèse nulle s'écrive $H_0 = (\theta = 0)$, ce qui signifie que $\phi_0(L) = 1$.

Dans le paragraphe II, nous considérons les tests fondés sur la valeur de la corrélation empirique entre la variable présente et une combinaison linéaire des valeurs retardées. Nous étudions le comportement asymptotique de cette statistique sous l'hypothèse de bruit blanc et sous des suites d'alternatives locales. Ceci nous permet de mettre en évidence les tests localement les plus puissants parmi les tests de cette forme et de vérifier que la procédure du multiplicateur de Lagrange satisfait une condition d'optimalité.

Dans le paragraphe III, nous étudions le problème du test de l'hypothèse de bruit blanc en supposant sous l'hypothèse maintenue que le processus est fractionnaire. L'intérêt d'une telle étude vient du fait que la procédure du multiplicateur de Lagrange pour tester l'hypothèse de bruit blanc contre l'hypothèse d'autorégressif d'ordre 1 et celle pour tester l'hypothèse de bruit blanc contre l'hypothèse de moyenne mobile d'ordre 1 sont identiques. Incluant dans l'hypothèse générale tous les processus fractionnaires du type :

$(1 - \alpha L)^d X_t = \varepsilon_t$, on peut espérer distinguer les alternatives autorégressives ($d > 0$) des alternatives moyennes mobiles ($d < 0$). On verra cependant que l'existence d'un problème d'identifiabilité des paramètres α , d sous l'hypothèse nulle rend impossible une application directe des procédures classiques de test. On propose alors une démarche de type minimax afin de prendre en compte les diverses alternatives locales.

II - CONSTRUCTION DE TESTS CONVERGENTS

II.A - FORME DES TESTS

Sous l'hypothèse de bruit blanc, les composantes X_t du processus sont non corrélées entre elles. Il y a donc non corrélation entre la valeur présente

du processus à la date t : X_t , et une combinaison linéaire des valeurs passées :

$$A(L) X_t = (a_1 L + a_2 L^2 + \dots + a_k L^k + \dots) X_t .$$

Il est donc naturel de considérer la statistique donnant la corrélation empirique entre X_t et $A(L) X_t$:

$$(2.1) \quad E_T(A) = \frac{\sum_{t=1}^T X_t [A(L) X_t]}{\sqrt{\sum_{t=1}^T X_t^2} \sqrt{\sum_{t=1}^T [A(L) X_t]^2}}$$

Sous l'hypothèse nulle, X est un bruit blanc indépendant et on peut appliquer les théorèmes limites classiques ; on a :

$$p \lim_T \frac{1}{T} \sum_{t=1}^T X_t^2 = V X_t = \sigma^2 ,$$

$$p \lim_T \frac{1}{T} \sum_{t=1}^T [A(L) X_t]^2 = V [A(L) X_t] = \left[\sum_{k=1}^{\infty} a_k^2 \right] \sigma^2 .$$

De plus il y a normalité asymptotique de la variable :

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T X_t [A(L) X_t] .$$

Remarquant que :

$$E [X_t A(L) X_t] = E [X_t (a_1 X_{t-1} + a_2 X_{t-2} + \dots)] = 0 ,$$

$$E [X_t A(L) X_t] X_{t+h} A(L) X_{t+h}$$

$$= E X_t E [(A(L) X_t) X_{t+h} A(L) X_{t+h}]$$

$$= 0, \quad \text{si } h < 0,$$

$$\begin{aligned} E [X_t A(L) X_t]^2 &= E X_t^2 E [A(L) X_t]^2 \\ &= V X_t V [A(L) X_t], \end{aligned}$$

on en déduit la moyenne et la variance de cette loi normale limite :

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T X_t [A(L) X_t] \xrightarrow[H_0]{d} N \left[0, V X_t V [A(L) X_t] \right].$$

On a donc :

$$(2.2) \quad \boxed{\sqrt{T} \xi_T(A) \xrightarrow[H_0]{d} N(0,1)}.$$

A chacune de ces statistiques, on peut associer un test de risque asymptotique α donné de la manière habituelle. Notons $u_{1-\alpha/2}$ le quantile à $1 - \frac{\alpha}{2}$ de la loi normale centrée réduite, le test consiste à :

$$(2.3) \quad \boxed{\begin{cases} \text{accepter l'hypothèse de bruit blanc, si } |\sqrt{T} \xi_T(A)| < u_{1-\alpha/2}, \\ \text{refuser l'hypothèse de bruit blanc, si } |\sqrt{T} \xi_T(A)| > u_{1-\alpha/2}, \end{cases}}$$

Remarquons que pour l'instant, il ne s'agit pas réellement d'un test, puisque la variable $\xi_T(A)$ peut dépendre de toutes les valeurs retardées de X , en particulier de celles non observables correspondant aux indices négatifs. En fait on peut montrer qu'en mettant à zéro ces valeurs la statistique obtenue possède les mêmes propriétés limites. Nous ferons pour cette raison cet abus de langage dans toute la suite du texte.

II.B - LOI DE LA STATISTIQUE DE TEST POUR DES ALTERNATIVES LOCALES

Afin d'obtenir la forme du test nous avons déterminé la loi de $\sqrt{T} \xi_T(A)$ sous l'hypothèse nulle. Nous allons maintenant la chercher sous des suites d'alternatives locales. On suppose que le modèle alternatif est :

$$(2.4) \quad \phi_{\theta} (L) X_t = \varepsilon_t ,$$

où (ε_t) est un bruit blanc indépendant, où $\phi_{\theta} (L)$ est un opérateur autorégressif fonction d'un paramètre scalaire θ . Il s'agit d'une suite d'alternatives locales au sens où le paramètre n'est pas choisi fixe, mais est pris fonction du nombre T d'observations. Plus précisément, nous posons :

$$(2.5) \quad \theta_T = \frac{\theta_0}{\sqrt{T}} .$$

On a alors les approximations :

$$\left[1 + \frac{\partial \phi_0 (L)}{\partial \theta} \frac{\theta_0}{\sqrt{T}} \right] X_t \approx \varepsilon_t$$

$$\Leftrightarrow X_t \approx \left[1 - \frac{\partial \phi_0 (L)}{\partial \theta} \frac{\theta_0}{\sqrt{T}} \right] \varepsilon_t ,$$

où le signe \approx signifie que les termes d'ordre supérieur sont négligés (en probabilité). On remarque que comme $\phi_{\theta} (0) = 1$, on a toujours

$$\frac{\partial \phi_0 (0)}{\partial \theta} = 0 .$$

Les diverses statistiques intervenant dans l'expression de $\sqrt{T} \xi_T(A)$ sont alors asymptotiquement telles que :

$$p \lim_T \frac{1}{T} \sum_{t=1}^T X_t^2 = \lim_T V \left[\left[1 - \frac{\frac{\partial \phi_o(L)}{\partial \theta}}{\frac{\theta_o}{\sqrt{T}}} \right] \varepsilon_t \right] = V \varepsilon_t = \sigma^2 ,$$

$$p \lim_T \frac{1}{T} \sum_{t=1}^T [A(L) X_t]^2 = \lim_T V \left[A(L) \left[1 - \frac{\frac{\partial \phi_o(L)}{\partial \theta}}{\frac{\theta_o}{\sqrt{T}}} \right] \varepsilon_t \right] = V (A(L) \varepsilon_t)$$

$$= \left[\sum_{k=1}^{\infty} a_k^2 \right] \sigma^2 .$$

Le numérateur est lui équivalent à :

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T X_t [A(L) X_t]$$

$$** \frac{1}{\sqrt{T}} \sum_{t=1}^T \left[\left[1 - \frac{\frac{\partial \phi_o(L)}{\partial \theta}}{\frac{\theta_o}{\sqrt{T}}} \right] \varepsilon_t \right] \left[\left[A(L) - A(L) \frac{\frac{\partial \phi_o(L)}{\partial \theta}}{\frac{\theta_o}{\sqrt{T}}} \right] \varepsilon_t \right]$$

$$** \frac{1}{\sqrt{T}} \sum_{t=1}^T \varepsilon_t [A(L) \varepsilon_t] - \frac{\theta_o}{T} \sum_{t=1}^T \frac{\frac{\partial \phi_o(L)}{\partial \theta}}{\frac{\theta_o}{\sqrt{T}}} \varepsilon_t A(L) \varepsilon_t .$$

$$\text{Comme : } \frac{1}{\sqrt{T}} \sum_{t=1}^T \varepsilon_t [A(L) \varepsilon_t] \xrightarrow{d} N [0, V \varepsilon_t V (A(L) \varepsilon_t)] ,$$

$$\text{et que : } \frac{1}{T} \sum_{t=1}^T \frac{\frac{\partial \phi_o(L)}{\partial \theta}}{\frac{\theta_o}{\sqrt{T}}} \varepsilon_t A(L) \varepsilon_t \xrightarrow{P} \text{Cov} \left[\frac{\frac{\partial \phi_o(L)}{\partial \theta}}{\frac{\theta_o}{\sqrt{T}}} \varepsilon_t , A(L) \varepsilon_t \right] ,$$

On a :

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T X_t [A(L) X_t] \xrightarrow{d} N \left[- \theta_o \text{Cov} \left[\frac{\frac{\partial \phi_o(L)}{\partial \theta}}{\frac{\theta_o}{\sqrt{T}}} \varepsilon_t , A(L) \varepsilon_t \right] , V \varepsilon_t V (A(L) \varepsilon_t) \right] .$$

PROPRIETE 2.6 :

Sous la suite d'alternatives locales : $\theta_T = \frac{\theta_0}{\sqrt{T}}$, on a :

$$\sqrt{T} \epsilon_T(A) \xrightarrow{d} N \left[\frac{-\theta_0 \text{Cov} \left[\frac{\partial \phi(L)}{\partial \theta} \epsilon_t, A(L) \epsilon_t \right]}{\sqrt{V \epsilon_t} \sqrt{V(A(L) \epsilon_t)}}, 1 \right].$$

Le test apparaît localement convergent, si le paramètre de décentrage est

non nul, c'est-à-dire si $\text{Cov} \left[\frac{\partial \phi(L)}{\partial \theta} \epsilon_t, A(L) \epsilon_t \right] \neq 0$.

II.C - RECHERCHE D'UN TEST LOCALEMENT OPTIMAL

La puissance locale asymptotique du test associé à $\epsilon_T(A)$ est une fonction croissante du paramètre de non centralité :

$$|\lambda| = \left| \frac{-\theta_0 \text{Cov} \left[\frac{\partial \phi(L)}{\partial \theta} \epsilon_t, A(L) \epsilon_t \right]}{\sqrt{V \epsilon_t} \sqrt{V(A(L) \epsilon_t)}} \right|.$$

Pour maximiser la puissance locale, il faut donc maximiser le module de la

corrélation entre $\frac{\partial \phi(L)}{\partial \theta} \epsilon_t$ et $A(L) \epsilon_t$.

Notons $\frac{\partial \phi(L)}{\partial \theta} = \psi_1 L + \psi_2 L^2 + \dots + \psi_k L^k + \dots$, ceci revient à maximiser

$$\left| \frac{\sum_{k=1}^K \psi_k a_k}{\sqrt{\sum_{k=1}^K \psi_k^2} \sqrt{\sum_{k=1}^K a_k^2}} \right| = \left| \frac{\langle \Psi, a \rangle}{\|\Psi\| \|a\|} \right| . \text{ Les solutions sont}$$

donc obtenues en choisissant le vecteur a en liaison affine avec le vecteur Ψ :

$$\exists \beta_0, \beta_1 : a_k = \beta_0 + \beta_1 \psi_k .$$

Comme de plus pour l'existence des processus, il faut que $\sum_{k=1}^{\infty} a_k^2 < \infty$,
 ceci entraine $\beta_0 = 0$.

PROPRIETE 2.7 :

Les tests localement les plus puissants parmi ceux fondés sur

$$\xi_T(A) , \text{ sont obtenus pour } A(L) \text{ proportionnel à } \frac{\partial \phi_o(L)}{\partial \theta} .$$

Le test optimal correspond donc à :

$$\xi_T^* = \frac{\sum_{t=1}^T X_t \frac{\partial \phi_o(L)}{\partial \theta} X_t}{\sqrt{\sum_{k=1}^T X_t^2} \sqrt{\sum_{k=1}^T \left[\frac{\partial \phi_o(L)}{\partial \theta} X_t \right]^2}}$$

II.D - TEST DU MULTIPLICATEUR DE LAGRANGE

Sous l'hypothèse de normalité des erreurs, il existe plusieurs expressions exacte ou approchées de la log vraisemblance. l'une de celle-ci est obtenue facilement en remplaçant les composantes non observables du bruit par leur

expression en fonction du processus X .

On a :

$$\begin{aligned} \text{Log } L_T(\theta) &= -\frac{T}{2} \text{Log } \sigma^2 - \frac{T}{2} \text{Log } 2\pi - \frac{1}{2\sigma^2} \sum_{t=1}^T \varepsilon_t^2 \\ &= -\frac{T}{2} \text{Log } \sigma^2 - \frac{T}{2} \text{Log } 2\pi - \frac{1}{2\sigma^2} \sum_{t=1}^T [\phi_\theta(L) X_t]^2. \end{aligned}$$

Le score correspondant est :

$$\frac{\partial \text{Log } L_T(\theta)}{\partial \theta} = -\frac{1}{\sigma^2} \sum_{t=1}^T [\phi_\theta(L) X_t] \left[\frac{\partial \phi_\theta(L)}{\partial \theta} X_t \right];$$

Évalué sous l'hypothèse nulle, c'est-à-dire pour $\theta = 0$, il est égal à :

$$\frac{\partial \text{Log } L_T(0)}{\partial \theta} = -\frac{1}{\sigma^2} \sum_{t=1}^T X_t \frac{\partial \phi_0(L)}{\partial \theta} X_t.$$

On retrouve le numérateur de la statistique optimale, de sorte qu'après réduction le test fondé sur le score coïncide avec le test localement le plus puissant donné en 2.7.

PROPRIÉTÉ 2.8 :

Le test fondé sur le score est localement le plus puissant parmi ceux fondés sur les statistiques $\xi_T(A)$, A quelconque.

II.E - COMPARAISON DES PUISSANCES LOCALES CORRESPONDANT A DEUX SUITES D'ALTERNATIVES LOCALES

Étant donné l'un des tests précédents associé à une statistique $\xi_T(A)$, on peut se demander s'il est plus puissant pour tester l'hypothèse nulle de

bruit blanc contre une alternative $\phi_{\theta}(L)X_t = \epsilon_t$, ou pour tester cette même

hypothèse contre une autre alternative : $\phi_{\tilde{\theta}}(L)X_t = \epsilon_t$. On peut évidemment

pour chacune de ces alternatives définir le modèle au voisinage de l'hypothèse

$H_0 : \langle \theta = 0 \rangle = \langle \tilde{\theta} = 0 \rangle$. Cependant les paramètres θ et $\tilde{\theta}$ n'ayant a

priori aucun rapport entre eux, nous devons pour pouvoir comparer les

puissances locales choisir une convention naturelle reliant entre eux ces deux paramètres.

Posons donc : $\theta_T = \theta_0 / \sqrt{T}$ et $\tilde{\theta}_T = \tilde{\theta}_0 / \sqrt{T}$, comment fixer de façon

jointe θ_0 et $\tilde{\theta}_0$ pour donner un sens à la comparaison ? On pourrait par

exemple les choisir de façon qu'avec les deux modèles les variances soient

localement les mêmes. On a :

$$V X_t \neq V \left[\epsilon_t - \frac{\partial \phi_{\theta_0}(L)}{\partial \theta} \frac{\theta_0}{\sqrt{T}} \epsilon_t \right] = \sigma^2 + \frac{\theta_0^2}{T} V \left[\frac{\partial \phi_{\theta_0}(L)}{\partial \theta} \epsilon_t \right],$$

avec le premier modèle,

$$\text{et : } V X_t \neq V \left[\epsilon_t - \frac{\partial \phi_{\tilde{\theta}_0}(L)}{\partial \tilde{\theta}} \frac{\tilde{\theta}_0}{\sqrt{T}} \epsilon_t \right] = \sigma^2 + \frac{\tilde{\theta}_0^2}{T} V \left[\frac{\partial \phi_{\tilde{\theta}_0}(L)}{\partial \tilde{\theta}} \epsilon_t \right]$$

avec le second modèle.

De façon à avoir le même développement des variances, il est donc naturel

$$\text{de prendre : } \theta_0 = \eta_0 \sqrt{V \left[\frac{\partial \phi_{\theta_0}(L)}{\partial \theta} \epsilon_t \right]},$$

et

$$\hat{\theta}_0 = \eta_0 / \sqrt{V \left[\frac{\partial \phi_0(L)}{\partial \theta} \epsilon_t \right]}$$

Le paramètre de décentrage est alors de la forme :

$$\lambda = \frac{\eta_0}{\sqrt{V \epsilon_t}} \frac{\text{Cov} \left[\frac{\partial \phi_0(L)}{\partial \theta} \epsilon_t, A(L) \epsilon_t \right]}{\sqrt{V \left[\frac{\partial \phi_0(L)}{\partial \theta} \epsilon_t \right]} \sqrt{V (A(L) \epsilon_t)}}$$

Ainsi le test paraît localement plus puissant dès que le module du cosinus de l'angle entre Ψ et a s'accroît.

Remarquons à ce niveau que deux alternatives pour lesquels ces cosinus sont égaux apparaissent équivalentes au niveau du test. En particulier toutes celles pour lesquelles ce cosinus est nul sont assimilables au cas du bruit blanc et leur réunion constitue donc l'hypothèse nulle implicite locale associée à la procédure de test (voir Mizon-Richard (19)).

III - TEST DE BRUIT BLANC LORSQUE L'ALTERNATIVE EST FRACTIONNAIRE

III.A - LE TEST USUEL

Lorsqu'on se restreint à un test de bruit blanc faisant intervenir un seul paramètre, les alternatives choisies sont généralement soit une représentation autorégressive d'ordre 1 :

$$X_t - \theta X_{t-1} = \epsilon_t, \quad |\theta| < 1,$$

soit une représentation moyenne mobile d'ordre 1,

$$X_t = \epsilon_t - \theta \epsilon_{t-1}, \quad |\theta| < 1.$$

Dans le premier cas on a $\phi_\theta(L) = 1 - \theta L$ et dans le second $\phi_\theta(L) = 1 / (1 - \theta L)$.

On vérifie alors que
$$\frac{\partial \phi_\theta(L)}{\partial \theta} = -L = -\frac{\partial \phi_\theta(L)}{\partial \theta}.$$

Ces deux alternatives sont donc localement équivalentes pour les tests associés aux statistiques $\epsilon_T(A)$. En particulier, il leur correspond une même statistique optimale donnée par :

$$(3.1) \quad \epsilon_T = \sqrt{T} \frac{\sum_{t=2}^T X_t X_{t-1}}{\sqrt{\sum_{t=1}^T X_t^2} \sqrt{\sum_{t=2}^T X_{t-1}^2}}.$$

(voir Godfrey-Wickens (1982)).

L'hypothèse nulle implicite locale associée à ce test est la réunion de toutes les représentations $\phi_\theta(L) X_t = \epsilon_t$ pour lesquelles :

$$\frac{\partial \phi_\theta(L)}{\partial \theta} = \psi_2 L^2 + \psi_3 L^3 + \dots. \text{ Elle comprend par exemple les processus autorégressifs du type } X_t - \theta X_{t-p} = \epsilon_t, \quad p \geq 2.$$

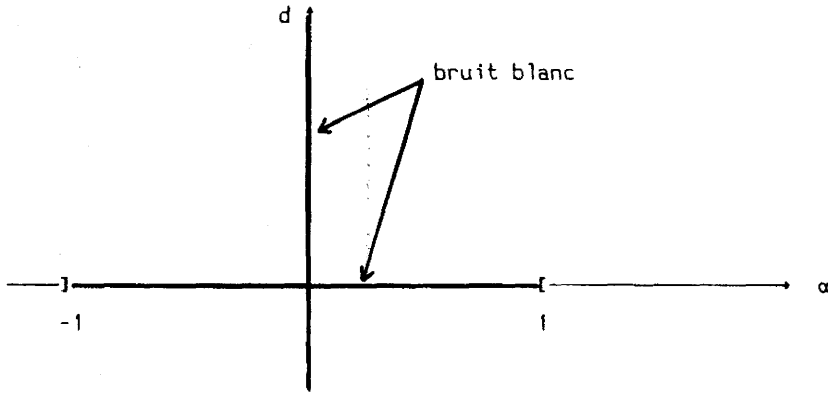
III.B - LES PROCESSUS FRACTIONNAIRES

Les deux modèles autorégressifs d'ordre 1 et moyenne mobile d'ordre 1 peuvent être considérés comme des cas particuliers d'un même modèle donné par :

$$(1 - \alpha L)^d X_t = \epsilon_t \quad , \quad |\alpha| < 1 \quad , \quad d \in \mathbb{R} \quad ,$$

et correspondent respectivement à $d = 1$ et $d = -1$. Les processus fractionnaires ont été étudiés par (Granger (1980), Granger-Joyeux (1980), Hosking (1981)...).

L'hypothèse nulle de bruit blanc peut être trouvée de plusieurs façons en fixant des contraintes sur les paramètres. On l'obtient ainsi en faisant $d = 0$ ou en faisant $\alpha = 0$.



Il apparaît clairement que la convergence vers le bruit blanc peut se produire de diverses façons : en faisant tendre d vers zéro et α vers une valeur constante non nulle α_0 , en faisant tendre α vers zéro et d vers une valeur constante non nulle d_0 , en faisant simultanément tendre α et d vers zéro. Ce sont ces divers cas que nous allons maintenant étudier.

III.C - ALTERNATIVE LOCALE A d FIXE

Prenons $d = d_0 \neq 0$ et paramétrons le modèle alternatif par $\theta = \alpha$.

On a : $\Phi_\theta(L) = (1 - \theta L)^{d_0}$. La dérivée prise en $\theta = 0$ est donc :

$$\frac{\partial \phi_o(L)}{\partial \theta} = \left[-d_o (1 - \theta L)^{d_o-1} L \right]_{\theta=0} = -d_o L .$$

Cette dérivée étant proportionnelle à $-L$, on en déduit que toutes ces alternatives sont localement équivalentes pour $d_o \neq 0$ et sont en particulier équivalentes localement à la représentation AR(1) [resp. MA(1)] . La statistique de test optimale, qui leur est associée est celle donnée en (3.1).

III.D - ALTERNATIVE LOCALE, LORSQUE α ET d TENDENT SIMULTANEMENT VERS ZERO

Lorsque les deux paramètres α et d tendent simultanément vers zéro, le développement limité de $\phi_{\alpha,d}(L) = (1 - \alpha L)^d$ montre que le terme d'ordre inférieur est :

$$\phi_{\alpha,d}(L) = -\alpha d L + o(\alpha d) .$$

On voit que ce terme est d'ordre deux en α, d et fait essentiellement intervenir le terme croisé αd . Ceci est une conséquence de la non identifiabilité du couple α, d sous l'hypothèse de bruit blanc.

On voit d'autre part que l'alternative locale est ici encore équivalente à celle associée aux représentations AR(1) ou MA(1). De sorte que la statistique optimale est toujours celle donnée en (3.1).

Il est clair que la non-identifiabilité du couple (α, d) empêche ici l'utilisation des résultats usuels sur le test du multiplicateur de Lagrange. Ainsi alors que l'hypothèse $H_o = (\alpha = 0, d = 0)$ est définie par deux contraintes sur les paramètres, l'alternative locale ne fait intervenir qu'une seule direction et il lui correspond une statistique de test dont la loi sous

l'hypothèse nulle est une loi du khi-deux avec un degré de liberté, inférieur au nombre de contraintes.

III.E - ALTERNATIVES LOCALES A α FIXE

a) Forme des alternatives locales.

Prenons maintenant $\alpha = \alpha_0 \neq 0$ et paramétrons le modèle par $\theta = d$.

$$\text{On a : } \phi_{\theta}(L) = (1 - \alpha_0 L)^{\theta}$$

$$\text{et } \frac{\partial \phi_{\theta}(L)}{\partial \theta} = \left[(1 - \alpha_0 L)^{\theta} \text{Log} (1 - \alpha_0 L) \right]_{\theta=0} = \text{Log} (1 - \alpha_0 L) = - \sum_{k=1}^{\infty} \frac{\alpha_0^k L^k}{k} .$$

Pour diverses valeurs de α_0 , ces opérateurs sont non proportionnels de sorte que les alternatives correspondantes ne sont pas localement équivalentes.

b) Quelques propriétés du modèle approché.

L'approche locale conduit ainsi à remplacer le modèle initial :

$$(1 - \alpha_0 L)^{\theta} X_t = \varepsilon_t ,$$

par le modèle approché :

$$\left[1 + \theta \text{Log} (1 - \alpha_0 L) \right] X_t = \varepsilon_t ,$$

ou de façon localement équivalente par :

$$(3.2) X_t = \left[1 - \theta \text{Log} (1 - \alpha_0 L) \right] \varepsilon_t .$$

Cette forme moyenne mobile infinie :

$$X_t = \left[1 - \theta \text{Log} (1 - \alpha_0 L) \right] \varepsilon_t = \varepsilon_t + \theta \sum_{k=1}^{\infty} \frac{\alpha_0^k}{k} \varepsilon_{t-k} ,$$

bien que n'étant pas un cas particulier de modèle fractionnaire possède

certaines propriétés de même type. Ainsi le développement moyenne mobile du processus fractionnaire

$$X_t = \sum_{k=0}^{\infty} a_k(\theta) \varepsilon_{t-k} ,$$

admet des coefficients égaux à : $a_k(\theta) = \frac{\Gamma(\theta+k)}{\Gamma(\theta)} \frac{\alpha_0^k}{k!}$.

Ceux du processus (3.2) sont donnés par : $\tilde{a}_k(\theta) = \theta \frac{\alpha_0^k}{k}$.

Bien que ces deux ensembles de coefficients ne coïncident pas, on voit qu'ils satisfont des relations de récurrence de même type :

$$k a_k(\theta) - \alpha_0(\theta + k - 1) a_{k-1}(\theta) = 0$$

et $k \tilde{a}_k(\theta) - \alpha_0(k - 1) \tilde{a}_{k-1}(\theta) = 0$.

Celle satisfaite par le "processus approché" se déduit de la précédente en mettant à zéro le paramètre θ .

c) Test du multiplicateur de Lagrange correspondant à $\alpha = \alpha_0$.

Le test localement optimal est ici fondé sur la statistique :

$$(3.3) \quad \xi_T(\alpha_0) = - \sqrt{T} \frac{\sum_{t=1}^T X_t \left[\sum_{k=1}^{\infty} \frac{\alpha_0^k X_{t-k}}{k} \right]}{\sqrt{\sum_{t=1}^T X_t^2} \sqrt{\sum_{t=1}^T \left[\sum_{k=1}^{\infty} \frac{\alpha_0^k X_{t-k}}{k} \right]^2}}$$

Remarquons à ce niveau que, si α_0 est connu, on peut distinguer à partir de cette statistique le cas autorégressif $d = 1$, du cas moyenne mobile

$d = -1$. En effet tester l'hypothèse de bruit blanc contre celle de représentation AR(1) est naturellement obtenu, en testant : $H_0 = (d = 0)$ contre $H_1 = (d > 0)$.

De façon similaire pour le cas moyenne mobile, on peut considérer

$H_0 : (d = 0)$ contre $\tilde{H}_1 : (d < 0)$.

Introduire des tests unilatéraux fondés sur $\xi_T(\alpha_0)$, c'est-à-dire prendre en compte le signe de cette statistique peut alors permettre de distinguer entre représentation autorégressive et représentation moyenne mobile.

Malheureusement en pratique la valeur α_0 est inconnue et du fait du problème de non identifiabilité de α_0 sous l'hypothèse nulle, on ne peut remplacer α_0 par son estimateur non contraint en conservant de bonnes propriétés. Dans le paragraphe suivant, nous allons essayer de proposer une procédure de type minimax permettant de se garder dans toutes les directions possibles des alternatives locales.

III.F - CONSTRUCTION DE PROCEDURES MINIMAX

D'un point de vue pratique, il est souhaitable de construire une procédure de test possédant de bonnes propriétés de puissance vis à vis des diverses alternatives possibles. Si nous nous restreignons à une étude locale, il résulte du paragraphe précédent que l'ensemble des alternatives locales associées aux diverses valeurs de α_0 engendre un espace de dimension infinie (alors que dans des conditions d'applications classiques de la procédure du multiplicateur de Lagrange cet espace devrait être de dimension deux). Ceci entraîne en particulier qu'on ne peut espérer trouver une procédure (de type

multiplieur de Lagrange) qui soit localement uniformément la plus puissante.

En revanche on peut rechercher s'il existe une procédure minimax.

Considérons pour l'instant la classe des statistiques de test introduites en

(2.1) :

$$\sqrt{T} \epsilon_T(A) = \frac{\sum_{t=1}^T X_t [A(L) X_t]}{\sqrt{\sum_{t=1}^T X_t^2} \sqrt{\sum_{t=1}^T [A(L) X_t]^2}}$$

Il nous faut examiner sa loi limite pour les diverses alternatives locales

possibles. Celles-ci sont du type $M\left[\alpha_0, \frac{d(\alpha_0)}{\sqrt{T}}\right]$ et il faut choisir la

vitesse $d(\alpha_0)$ de façon à rendre ces alternatives comparables les unes avec

les autres. Suivant la même démarche que dans le paragraphe II.E, on obtient

des choix de $d(\alpha_0)$ du type : $d(\alpha_0) = \frac{d_0}{\sqrt{V \text{Log}(1 - \alpha_0 L) \epsilon_t}}$

Sous l'alternative locale : $M\left[\alpha_0, \frac{d_0}{\sqrt{T} \sqrt{V \text{Log}(1 - \alpha_0 L) \epsilon_t}}\right]$, la

statistique de test a pour loi :

$$\sqrt{T} \epsilon_T(A) \xrightarrow{\alpha_0, d_0} N\left[\frac{-d_0}{\sqrt{V \epsilon_t}} \frac{\text{Cov}(\text{Log}(1 - \alpha_0 L) \epsilon_t, A(L) \epsilon_t)}{\sqrt{V \text{Log}(1 - \alpha_0 L) \epsilon_t} \sqrt{V A(L) \epsilon_t}}, 1\right]$$

La puissance locale du test est fonction croissante du paramètre de non-

centralité : $\left| \frac{d_0}{\sqrt{V \epsilon_t}} \frac{\text{Cov}(\text{Log}(1 - \alpha_0 L) \epsilon_t, A(L) \epsilon_t)}{\sqrt{V \text{Log}(1 - \alpha_0 L) \epsilon_t} \sqrt{V A(L) \epsilon_t}} \right|$

Ceci conduit à rechercher un test localement minimax défini de la façon

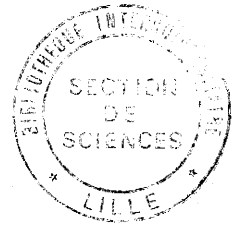
suivante :

Définition 3.4 :

Le test localement minimax parmi les tests fondés sur des statistiques du type $\xi_T(A)$ est obtenu en cherchant une solution en A de :

$$\text{Max}_A \text{ Min}_{\alpha_0} \frac{|\text{Cov}(\text{Log}(1 - \alpha_0 L) \varepsilon_t, A(L) \varepsilon_t)|}{\sqrt{V(\text{Log}(1 - \alpha_0 L) \varepsilon_t)} \sqrt{V A(L) \varepsilon_t}} .$$

L'idée est donc de se placer dans le cas le plus défavorable en retenant la valeur α_0 conduisant à la puissance minimale, puis de choisir A de façon à rendre le plus grand possible ce minimum.



REFERENCES

- Aitchison, J. et S.D. Silvey (1958) : Maximum likelihood estimation of parameters subject to restraints, A.M.S., Vol. 29, 813-828.
- Godfrey, L. (1981) : On the invariance of Lagrange multiplier test with respect to certain changes in the alternative hypothesis, *Econometrica* 49, 1443-1455.
- Godfrey, L. et Wickens, M.R. (1982) : Tests for misspecification using locally equivalent alternatives models, dans Chow-Corsi, "Evaluating the reliability of macro-economic models", Wiley.
- Granger, C.W. (1980) : Long memory relationships and the aggregation of dynamic models, *Journal of Econometrics* 14, 227-238.
- Granger, C.W. et R. Joyeux (1980) : An introduction to long memory time series models and fractional differencing, *Journal of Time Series Analysis* 1, 15-29.
- Hosking, J.R. (1980) : Lagrange multiplier tests of time series models, *JRSS B* 42, 170-181.
- Hosking, J.R. (1981) : Fractional differencing, *Biometrika* 68, 165-176.
- Mizon-Richard
- Silvey, S. (1959) : The Lagrangian multiplier test, *AMS* 30, 389-407.



R E S U M E

Nous proposons une généralisation de la notion de processus ARMA en permettant aux opérateurs autorégressifs et moyennes mobiles d'admettre des racines avec ordre de multiplicité fractionnaire. Nous commençons par étudier les propriétés probabilistes de tels processus : forme des autocovariances, du spectre, des développements autorégressifs et moyennes mobiles. Nous examinons ensuite les propriétés ergodiques concernant les moyennes empiriques covariances empiriques et périodogramme. Il apparaît nécessaire de distinguer divers cas selon que le processus est ou non stationnaire et lorsqu'il est stationnaire, selon que les conditions de mélangeance sont ou non satisfaites. Ceci conduit à des lois limites qui ne sont pas toujours gaussiennes et peuvent faire intervenir des processus de Rosenblatt.

La seconde partie du travail concerne l'aspect statistique. Nous commençons par examiner les diverses expressions approchées de la vraisemblance pour un modèle de série temporelle stationnaire, ce qui permet de dériver diverses expressions des informations, puis d'étudier les propriétés des estimateurs du maximum de vraisemblance et des tests de type score.

Divers problèmes sont ensuite examinés :

Agrégation de processus autorégressifs et test d'homogénéité.

Construction d'un test de bruit blanc et étude de son optimalité.

M O T S C L E S

- PROCESSUS
- MEMOIRE LONGUE
- MAXIMUM DE VRAISEMBLANCE
- AGREGATION
- ERGODICITE
- TEST DU SCORE
- TEST DE BRUIT BLANC