



**HAL**  
open science

# Statistique Paramétrique des processus Longue Mémoire

Dominique Guegan

► **To cite this version:**

Dominique Guegan. Statistique Paramétrique des processus Longue Mémoire. Annales de l'ISUP, 1991, XXXVI (1-2), pp.125-140. hal-03664801

**HAL Id: hal-03664801**

**<https://hal.science/hal-03664801>**

Submitted on 11 May 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Pub. Inst. Stat. Univ. Paris  
XXXVI, fasc. 1-2, 1991, 125 à 140

## STATISTIQUE PARAMÉTRIQUE DES PROCESSUS LONGUE MÉMOIRE

D. GUÉGAN

Université Paris XIII\* et L.S.T.A.

### Résumé

Dans cet article nous nous intéressons aux différents problèmes statistiques abordés dans les processus présentant un comportement longue mémoire. Les processus considérés sont les processus de régression avec bruits à longue mémoire et les processus FARMA  $(p,d,q)$ . Les propriétés statistiques étudiées concernent l'estimation, les tests et la prévision.

### Mots clefs

Processus longue mémoire - Brownien fractal - FARMA - Modèle de régression - Estimation - Tests - Prévisions.

Classification AMS : 60G11 - 62M10.

Université Paris XIII

Institut Galilée

Département de Mathématiques

Avenue J.B. Clément

93430 Villetaneuse.

## I - Introduction

Nous nous intéressons dans cet article aux propriétés statistiques des processus longue mémoire. Les processus présentant un tel comportement et pour lesquels certaines statistiques ont été étudiées peuvent être classés en deux catégories :

- Les modèles de régression avec bruit longue mémoire.
- Les processus pour lesquels la fonction de covariance  $\rho(k)$  est telle que :

$$(1) \quad \rho(k) \simeq L(\theta)|k|^{-\alpha(\theta)} \quad \text{quand } k \rightarrow \infty$$

où  $L(\theta)$  est une fonction à variation bornée, décroissant lentement. Cette catégorie de processus englobe en particulier les processus FARMA, GARMA, le mouvement brownien fractal, les modèles de cointégration.

Les principales propriétés statistiques de ces modèles, étudiées dans la littérature, concernent l'estimation, les tests et les problèmes de prévision. La plupart des travaux sur ces sujets datent d'une quinzaine d'années, beaucoup de résultats sont parcellaires et une théorie de l'inférence statistique sur ces modèles reste à faire. Actuellement le point de vue le moins développé est celui de l'identification. Le problème n'est pas simple et ce n'est que très récemment que des études essentiellement empiriques se sont développées, particulièrement dans le domaine de l'économie avec les travaux de Carlin, Dempster et Jonas (1985) et Carlin et Dempster (1989). Une des grandes difficultés pour identifier un comportement longue mémoire est d'être capable de "séparer" correctement ce qui revient au comportement mémoire courte de la série de ce qui appartient au comportement mémoire longue, en particulier en partant d'une modélisation par les processus ARMA. Le problème de la surdifférentiation n'est pas à négliger dans ce contexte.

La suite de notre article est organisée de la manière suivante : dans le paragraphe 2 nous considérons les problèmes d'estimation dans le cas d'une régression, le paragraphe 3 est consacré à l'estimation essentiellement pour les processus FARMA (p,d,q), mais aussi de manière plus générale pour les processus possédant une fonction de corrélation telle que (1). Le paragraphe 4 traite des problèmes de tests et le paragraphe 5 précise quelques résultats concernant la prévision.

## II - Estimation dans les modèles de régression

Dans cette section nous nous intéressons au problème de l'estimation, de la moyenne de la série observée et/ou du paramètre de régression  $\beta$ , de l'ordre de différentiation du bruit et de la variance du même bruit.

Appelons  $\{Y(t)\}$  le processus observé et supposons que le processus considéré suive le modèle de régression suivant :

$$(2) \quad Y(t) = X(t)' \cdot \beta + \varepsilon(t)$$

où  $X(t) = (X_1(t), \dots, X_k(t))$  est un vecteur de régresseurs de dimension  $k$ , où  $\{\varepsilon(t)\}$  est un processus stationnaire, centré, de densité spectrale  $f(\lambda)$ , de variance  $\sigma_\varepsilon^2$  et où  $\beta = (\beta_1, \dots, \beta_k)'$  est un vecteur de dimension  $k$  représentant les paramètres de régression inconnus. On suppose que la densité spectrale du bruit est de la forme suivante :

$$(3) \quad f(\lambda) = f^*(\lambda) / |1 - e^{i\lambda}|^{2d}, \quad 0 < d < \frac{1}{2}$$

où  $f^*(\lambda)$  est une fonction positive, continue. Comme  $f(\lambda)$  diverge à l'infini quand  $\lambda \rightarrow 0$ , le processus  $\{\varepsilon(t)\}$  est fortement dépendant et ses autocorrélations  $\gamma_k = E \varepsilon(t)\varepsilon(t+k)$ ,  $k = 0, 1, \dots$  ne sont pas absolument sommables. Il s'agit donc bien d'un processus longue mémoire. Pour le processus  $\{Y(t)\}$  défini par (2)-(3), les paramètres inconnus à estimer sont donc  $\beta$ ,  $d$  et  $\sigma_\varepsilon^2$ , ainsi que la moyenne  $\bar{Y}(T)$  pour un échantillon  $Y(1), \dots, Y(T)$ . En théorie de la régression linéaire classique, deux types d'estimateurs sont en général particulièrement étudiés pour  $\bar{Y}(T)$  et  $\beta$  : il s'agit de l'estimateur des moindres carrés que nous désignerons par LSE et de l'estimateur sans biais linéaire, classiquement désigné par BLUE. Ces deux estimateurs sont comparés en terme d'efficacité relative (rapport de la variance du BLUE et de la variance du LSE si l'on veut connaître l'efficacité du LSE par rapport au BLUE), ou asymptotique (limite du rapport précédent quand  $T \rightarrow \infty$ ), et l'on sait qu'en ce qui concerne la moyenne d'un échantillon, le BLUE peut ne pas être le meilleur estimateur dès lors que la série n'est pas gaussienne. Grenander a établi, en 1954, pour le modèle (2) si  $f(\lambda)$  est une fonction continue et positive et sous certaines conditions, que pour la moyenne des observations, le LSE était asymptotiquement plus efficace que le BLUE. Il s'agit ici d'établir des résultats du même type dans le cas où la densité  $f(\lambda)$  est définie par (3).

### 1. Estimation de la moyenne du processus observé et du paramètre $\beta$

- En 1974, Adenstedt étudie les propriétés de la moyenne empirique  $\bar{Y}(T) = \frac{1}{T} \sum_{t=1}^T Y(t)$ , pour le modèle (2)-(3) avec  $X_1(t) = 1$  et  $k = 1$ . Il considère deux estimateurs le BLUE =  $\bar{Y}_{\text{BLUE}}(T)$  dont la variance est :

$$\text{var } \bar{Y}_{\text{BLUE}}(T) = \frac{B(n+1, 1-2d)}{B(1-d, 1-d)}$$

et l'estimateur sans biais

$$\bar{Y}_S(T) = \sum_{t=1}^T C_t(T, d) Y(t)$$

où

$$C_t(T, d) = C_T^t \frac{B(t+1-d, T-d-t+1)}{B(1-d, 1-d)}$$

L'efficacité asymptotique de ce dernier par rapport au BLUE est établie :

$$\text{var } \bar{Y}_{\text{BLUE}}(T) / \text{var } \bar{Y}_S(T) \rightarrow 1 \quad \text{quand } T \rightarrow \infty.$$

- En 1985, Beran et Kunsch, pour le même modèle que Adenstedt, considèrent un  $M$ -estimateur  $M_T$  de la moyenne empirique défini par :

$$\sum_{t=1}^T \psi(Y(t) - M_T) = 0$$

(voir, Huber (1981), pour la définition de ces estimateurs).

L'efficacité du  $M$ -estimateur relativement à  $\bar{Y}(T)$ , pour une fonction  $\psi$  quelconque est égale à 1, et l'efficacité asymptotique du  $M$ -estimateur est nulle. Ce type d'estimateur est donc robuste si on choisit pour  $\psi$  une fonction bornée (et non linéaire) dont le comportement ne s'éloigne pas trop du processus longue mémoire.

- En 1988, Yagima considère le modèle (2)-(3) avec  $X_i(t) = t^{i-1}$  et  $k \geq 1$ . Il s'agit donc d'une régression polynomiale. Il estime le régresseur  $\beta$  à partir des deux estimateurs BLUE et LSE et calcule leurs matrices de covariance. Les résultats sont les suivants :

$$\lim_{T \rightarrow \infty} D_T E[(\hat{\beta}_{LSE} - \beta)(\hat{\beta}_{LSE} - \beta)'] D_T / T^{2d} = 2\pi f^*(0) M^{-1} H(d) M^{-1}$$

où  $M = (m_{ij}) = \{(2i-1)(2j-1)\}^{1/2} / (i+j-1)$

et  $H(d) = (h_{ij}(d)) = \{[(2i-1)(2j-1)]^{1/2} \Gamma(1-2d) / \{\Gamma(d)\Gamma(1-d)\}\}$ .

$$\int_0^1 \int_0^1 x^{i-1} y^{j-1} |x-y|^{2d-1} dx dy,$$

et

$$\lim_{T \rightarrow \infty} D_T E[(\hat{\beta}_{BLUE} - \beta)(\hat{\beta}_{BLUE} - \beta)'] D_T / T^{2d} = 2\pi f^*(0) W^{-1}(d)$$

où  $W(d) = (w_{ij}(d)) = \frac{\Gamma(i-d)\Gamma(j-d)\{(2i-1)(2j-1)\}^{1/2}}{\Gamma(i-2d)\Gamma(j-2d)(i+j-1-2d)}$

avec  $D_T = \text{diag}(\|X(1)\|_T, \|X(2)\|_T, \dots, \|X(k)\|_T)$ .

L'efficacité du LSE par rapport au BLUE est alors obtenue, mais on peut voir qu'elle est plus faible pour  $k=2$  que pour  $k=1$ . Par contre l'efficacité asymptotique n'est pas acquise, ce qui est complètement différent de ce qui se passe quand la densité  $f(\lambda)$  est positive et continue.

- En 1990, Beran, Kunsch et Hampel s'intéressent à un modèle de régression tel que (2)-(3) avec des contrastes. Le modèle considéré est le suivant, si l'on suppose  $p$  traitements possibles :

$$(4) \quad Y(t) = \sum_{j=1}^p \beta_j X(t, j) + \varepsilon(t)$$

où  $\beta_j$  représente la moyenne pour le  $j^{\text{ième}}$  traitement et où la quantité  $\sum_{j=1}^p \lambda_j \beta_j$

est appelé le contraste sous la contrainte  $\sum_{j=1}^p \lambda_j = 0$ . L'estimateur  $\hat{\beta}_{LSE}$  est calculé

ainsi que la variance du contraste  $\sum_{j=1}^p \lambda_j \hat{\beta}_{LSE}$ . Ils montrent qu'ils peuvent améliorer

l'efficacité relative de l'estimateur  $\hat{\beta}_{LSE}$  par rapport à l'estimateur  $\hat{\beta}_{BLUE}$  en imposant un effet de "randomisation" sur l'allocation de chaque traitement  $X(t, j)$ . Il semble alors qu'asymptotiquement la corrélation existant sur les erreurs du processus  $\{\varepsilon(t)\}$  (à cause de la forme de la densité donnée par (3)), n'affecte pas la variance du contraste.

- Enfin Yagima (1988, 1991) s'intéresse au modèle de régression général (2)-(3). Il obtient tout d'abord la consistance de l'estimateur  $\hat{\beta}_{LSE}$  sous certaines conditions, puis comme pour le modèle de régression polynomial, il donne le comportement asymptotique des matrices de covariance des estimateurs  $\hat{\beta}_{LSE}$  et  $\hat{\beta}_{BLUE}$  et il obtient alors une condition nécessaire et suffisante pour le  $\hat{\beta}_{LSE}$  d'être asymptotiquement efficace généralisant ainsi en un certain sens les travaux de Grenander et Rosenblatt (1957) et Anderson (1971). Il répond donc en partie au problème posé au début de ce paragraphe.

L'ensemble de ces travaux montre que pour un modèle de régression, l'estimation de la moyenne et du régresseur  $\beta$  est très importante, d'autant qu'intervient un phénomène longue mémoire sur les erreurs. En effet, ceci implique que la variance de l'erreur ne décroît plus en  $n^{-1}$  mais en  $n^{-2d-1}$ , ce qui a une conséquence non négligeable sur les intervalles de confiance que l'on peut construire pour décider de la validité de l'estimateur considéré.

La plupart des auteurs ont abordé ce problème en étudiant l'efficacité relative ou asymptotique de l'estimateur des moindres carrés par rapport à un estimateur de référence sans biais. On constate d'ailleurs qu'il peut y avoir perte d'efficacité du LSE quand la densité spectrale n'est pas positive et continue.

Il n'est pas possible pour l'instant d'établir une théorie de la régression avec erreur à longue mémoire dans la mesure où les techniques utilisées sont spécifiques à chaque modèle, et ne s'étendent donc pas facilement, et où un certain nombre de problèmes sont encore non résolus. Notons que Percival (1985) a étudié aussi la variance empirique d'un processus FARMA(p,d,q) et a établi que l'estimateur habituellement considéré pour la variance est fortement biaisé dans le cas d'un tel processus. Enfin il faut remarquer que le paramètre  $d$  intervient dans la vitesse de convergence de la variance de l'estimateur de la moyenne ou du régresseur  $\beta$  et que l'estimation de ce paramètre est donc fondamentale. Tous les précédents travaux ont abordé le problème de l'estimation en supposant  $d$  connu.

## 2. Estimation de $d$ et de $\sigma_\varepsilon^2$

- En 1988, Yagima, pour le modèle de régression (2)-(3), le régresseur  $\beta$  étant estimé de manière consistante, construit pour  $(d, \sigma_\varepsilon^2)$  un estimateur à partir de l'approximation de Whittle de la vraisemblance :  $\hat{d}_T$  est l'estimateur qui minimise la quantité suivante :

$$U_T(d, \hat{\beta}_T) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \sum_{t=1}^T (Y(t) - X'(t) \cdot \hat{\beta}) e^{i\lambda t} \right|^2 / |1 - e^{id}|^{2d} d\lambda$$

et  $\hat{\sigma}_{\varepsilon, T}^2 = \frac{1}{T} U_T(\hat{d}_T, \hat{\beta}_T)$ .

Les estimateurs  $\hat{d}_T$  et  $\hat{\sigma}_{\varepsilon, T}^2$  sont consistants, sous certaines conditions de régularité. Ce résultat généralise, en un certain sens, un résultat de Fox et Taquq (1986) qui pour le

cas particulier,  $X_1(t) = 1$  et  $k = 1$  dans l'équation (2)-(3) avaient obtenu la consistance et la normalité asymptotique des estimateurs  $\hat{d}_T$  et  $\hat{\sigma}_{\varepsilon,T}^2$  obtenus à partir de la même construction.

Pour le modèle de régression polynomial obtenu à partir de (2)-(3), où  $X_i(t) = t^{i-1}$  et  $k \geq 1$ , Yagima (1988) obtient la normalité asymptotique pour ses estimateurs ou bien une vitesse de convergence, à savoir :

• Si  $0 < d < \frac{1}{4}$ ,  $\sqrt{T}(\hat{d}_T - d, \hat{\sigma}_{\varepsilon,T}^2 - \sigma_\varepsilon^2) \sim N(0, A)$ , où

$$A = \begin{pmatrix} \frac{6}{\pi^2} & 0 \\ 0 & 2\sigma_\varepsilon^4 + K_4 \end{pmatrix}$$

où  $K_4 = E\varepsilon^4(t) - 3\sigma_\varepsilon^4$ .

(5) • Si  $d = \frac{1}{4}$ ,  $(\hat{d}_T - d, \hat{\sigma}_{\varepsilon,T}^2 - \sigma_\varepsilon^2) = Op\left(\left(\frac{\log T}{T}\right)^{1/2}\right)$ .

(6) • Si  $\frac{1}{4} < d < \frac{1}{2}$ ,  $(\hat{d}_T - d, \hat{\sigma}_{\varepsilon,T}^2 - \sigma_\varepsilon^2) = op\left(\frac{1}{T^{1-2d}}\right)$ .

Dans le tableau suivant nous résumons l'ensemble des résultats obtenus jusqu'à présent à notre connaissance pour un modèle de régression avec longue mémoire.

statistiques		
modèles (2)-(4)	$\hat{Y}_{LSE}, \hat{\beta}_{LSE}$	$\hat{d}_T, \hat{\sigma}_{\varepsilon,T}^2$
régression simple $Y(t) = \beta + \varepsilon(t)$	Adenstedt (1974) efficacité $\bar{X}_{LSE}$ Beran-Kunsch (1985) M-estimateurs	Fox-Taqqu (1986) (Whittle) consistance - Norm. Asymp
régression polynomiale $Y(t) = \beta_1 + \beta_2 t + \dots, \beta_k t^k + \varepsilon(t)$	Yagima (1988) efficacité $\hat{\beta}_{LSE}$	Yagima (1988) (Whittle) consistance - Norm. Asymp ou vitesse de conv.
régression avec contraste : (4)	Kunsch, Beran, Hampel (1990) efficacité $\hat{\beta}_{LSE}$	
Modèle général (2)-(3)	Yagima (1991) efficacité $\hat{\beta}_{LSE}$ Yagima (1991) consistance $\hat{\beta}_{LSE}$	Yagima (1991) (Whittle) consistance

Tableau 2.1.

### III - Estimation pour des processus longue mémoire

Dans ce paragraphe nous nous intéressons aux processus  $\{X(t)\}$  stationnaires, de moyenne  $\mu$  (éventuellement nulle), de densité spectrale  $f_\theta(\lambda)$ ,  $\lambda \in [-\pi, \pi]$ , où  $\mu$  et  $\theta$  sont des paramètres à estimer. On suppose que

$$(7) \quad f_\theta(\lambda) \sim |\lambda|^{-\alpha(\theta)} L_\theta(\lambda) \quad \text{quand } \lambda \rightarrow 0,$$

où  $0 < \alpha(\theta) < 1$  et où  $L_\theta(\lambda)$  est une fonction qui varie lentement en 0.

La classe des processus  $\{X(t)\}$  dont la densité spectrale est définie par (7) et qui sont des processus self-similaires contient en particulier les processus FARMA  $(p, d, q)$  et le mouvement brownien fractal. Pour ces modèles les problèmes d'estimation concernent, essentiellement l'estimation des paramètres de différentiation, classiquement notés  $d$  ou  $H$ , les paramètres de la partie ARMA du modèle, la variance  $\sigma_\varepsilon^2$  et la moyenne de l'échantillon  $X(1), \dots, X(T)$ . Pour la clarté de l'exposé et en vue d'uniformiser les résultats existants, nous donnons les résultats pour des processus  $\{X(t)\}$  dont la densité spectrale est donnée par (3). Ceci recouvre essentiellement les processus FARMA  $(p, d, q)$  introduits par Granger et Joyeux (1980) et Hosking (1981). A chaque fois que cela sera possible nous préciserons les résultats dans le cas d'une densité spectrale plus générale donnée par (7). Supposons donc que le processus  $\{X(t)\}$  soit défini par une équation du type suivant :

$$\phi(B)(1-B)^d(X(t) - \mu) = \theta(B)\varepsilon(t), \quad 0 < d < \frac{1}{2}$$

$$\text{où} \quad \phi(B) = 1 - \phi_1 B \dots - \phi_p B^p$$

$$\theta(B) = 1 - \theta_1 B \dots - \theta_q B^q$$

$$(1-B)^d = \sum_{i=0}^{\infty} C_d^i (-1)^i B^i,$$

et où  $\{\varepsilon(t)\}$  est un bruit blanc gaussien de moyenne zéro et de variance  $\sigma_\varepsilon^2$ . On suppose que les polynômes  $\phi(\cdot)$  et  $\theta(\cdot)$  ont des racines de module  $> 1$ , et n'ont pas de zéros communs.

### 1. Estimation de la moyenne de l'échantillon.

En 1988, Samarov et Taqqu considèrent les estimateurs LSE et BLUE de la moyenne du processus  $\{X(t)\}$ . Ils obtiennent l'efficacité du LSE par rapport au BLUE :

$$e(T, f_d) = \frac{n^2 B(T, 1-2d)}{B(1-d, 1-d) \sum_{j=1}^T \sum_{k=1}^T R_d(j-k)}$$

$$\text{où} \quad R_d(k) = \int_{-\pi}^{\pi} e^{i\lambda k} f_d(\lambda) d\lambda$$

$$\text{avec} \quad f_d(\lambda) = \frac{1}{2\pi} |\lambda|^{-2d} \quad \text{quand} \quad \lambda \rightarrow 0.$$

L'étude de la fonction  $e(T, f_d)$  est assez "intrigante". En effet on peut constater que l'estimateur LSE a une vitesse de convergence plus lente que le BLUE quand la spectre a un zéro à l'origine d'ordre 1. D'autre part l'efficacité du LSE décroît avec la longueur de l'échantillon. Il semble donc que dans ce cas, l'estimateur BLUE soit meilleur parce qu'il prend en compte toute la structure longue mémoire du processus, ce qui ne serait pas le cas du LSE. On peut remarquer aussi que

$$\lim_{T \rightarrow \infty} e(T, f_d) \rightarrow 1$$



pour  $0 < d < \frac{1}{2}$ . En cela ce résultat confirme l'étude faite de Beran et Kunsch (1985). Si l'on note  $e(\infty, f_\alpha)$  l'efficacité asymptotique, on a aussi :

$$\lim_{d \rightarrow 0} e(\infty, f_d) = 1.$$

## 2. Estimation des paramètres $d$ et $\sigma_\varepsilon^2$ .

Un grand nombre d'approches ont été envisagées pour estimer ces paramètres, la moyenne du processus  $\{X(t)\}$  étant supposée connue.

- Il y a eu tout d'abord les travaux de Hurst (1951), Mandelbrot et Wallis (1969) et Mandelbrot et Taquq (1979) qui à partir de la statistique R/S ont obtenu un estimateur de  $d$  ayant d'assez bonnes propriétés de robustesse.

- D'un point de vue spectral, l'estimation de  $d$  à partir du périodogramme a été étudiée par Janacek (1982) qui considère le log-spectrum, par Geweke et Porter-Hudak (1983) qui utilisent une régression du log-spectrum sur une constante. Ceux-ci obtiennent la distribution asymptotique de  $d$  sous certaines conditions de régularité. Graf, Hampel, Tacier (1984) font une étude expérimentale de  $d$  à partir de simulations.

Kashyap et Eom (1988) estiment  $d$  par un algorithme des moindres carrés dans le domaine fréquentiel basé sur une représentation du logarithme du périodogramme. Ils obtiennent la normalité asymptotique de l'estimateur de  $d$ . Une approche bayésienne est discutée dans Carlin, Dempster et Jonas (1986).

- En 1985, Yagima étudie à la fois l'estimateur des moindres carrés et l'estimateur du maximum de vraisemblance pour  $(d, \sigma_\varepsilon^2)$ . Il prouve la consistance des deux estimateurs pour  $0 < d < \frac{1}{2}$ . Il obtient de plus la normalité asymptotique ou la vitesse de convergence pour l'estimateur du maximum de vraisemblance  $(\hat{d}_T, \hat{\sigma}_{\varepsilon,T}^2)$  :

- pour  $0 < d < \frac{1}{4}$ ,

$$\sqrt{T}(d_T - d, \hat{\sigma}_{\varepsilon,T}^2 - \sigma^2) \rightarrow N(0, A) \quad \text{quand } T \rightarrow \infty$$

où

$$A = \begin{pmatrix} \frac{6}{\pi^2} & 0 \\ 0 & 2\sigma_\varepsilon^2 \end{pmatrix}$$

- Pour  $\frac{1}{2} < d < \frac{1}{4}$ , les vitesses de convergence sont celles données en (5) et (6).

Dans le cas de l'estimateur du maximum de vraisemblance, il étend les travaux de Hannan (1979), et dans le cas de l'estimateur des moindres carrés, il suit les méthodes développées par Rosenblatt (1979). Dans son article, Yagima considère deux types d'extension :

- d'une part quand la moyenne  $\bar{X}_T$  du paramètre est estimée : les résultats obtenus sont les mêmes que précédemment.

- d'autre part quand le bruit  $\{\varepsilon(t)\}$  n'est plus gaussien. Il suffit alors que le processus  $\{X(t)\}$  soit stationnaire, ergodique avec des moments finis d'ordre 4 pour que les résultats précédents soient maintenus.



modèles statistiques	FARMA(0, d, 0)	FARMA(p, d, q)	(7)
$\bar{X}_{LSE}$	Beran, Kunsch (1985) Samarov, Taqqu (1988) efficacité		
d	<ul style="list-style-type: none"> <li>• Approche R/S Hurst (1951), Mandelbrot, Wallis (1969), Mandelbrot, Taqqu (1979).</li> <li>• Approche fréquentielle Janacek (1982), Graf, Hampel, Tacier (1984) Geweke, Porter Hudak (1983) norm.asympt. Kashyap, Eom (1988) norm.asympt.</li> </ul>	<ul style="list-style-type: none"> <li>• Approche bayésienne Carlin, Dempster, Jonas. (1986),</li> </ul>	
$d, \sigma^2$ $\phi_1, \dots, \phi_p$	<ul style="list-style-type: none"> <li>• Max. de vrais. exact Yagima (1985) cons. et norm. asymp.</li> <li>• Appr. Whittle Fox, Taqqu (1986) et Dalhaus (1989) cons. et norm. asympt.</li> </ul>	<ul style="list-style-type: none"> <li>• Max de vrais. appr. (Box Jenkins) Li et McLeod (1986) norm. asympt.</li> </ul>	<ul style="list-style-type: none"> <li>• Appr. Whittle Fox, Taqqu (1986) et Dalhaus (1989) cons. et norm. asymp.</li> <li>• Max. de vrais. exact. Dalhaus (1989) cons. et norm. asympt.</li> </ul>

Tableau 3.1.

#### IV - Les Tests pour les processus longue mémoire

En ce qui concerne les modèles présentant un comportement longue mémoire, peu de tests ont été pour l'instant développés. Ceux que l'on rencontre peuvent se ranger en quatre classes : des tests paramétriques basés sur la construction d'un intervalle de confiance, des tests de discrimination, des tests d'adéquation et des tests "longue mémoire", tentant de prendre en compte le comportement longue mémoire. Il s'agit en général d'étude plutôt expérimentale que théorique. En matière de tests pour les processus longue mémoire, la théorie reste à faire.

### 1. Test paramétrique.

On s'intéresse ici aux accroissements du bruit gaussien fractal défini pour  $0 < H < 1$ , notés  $\{X(t)\}$  tel que  $T^{2-2H} \text{var} \bar{X}_T$  tend vers

$$C(H) = 4H^{-1} \Gamma(-2H) \Gamma(+2H - 1) \sin\left(\frac{1}{2} \pi - \pi H\right) \sin(\pi - \pi H).$$

On cherche à construire un test de la moyenne empirique  $\bar{X}_T$  pour un tel modèle qui soit une amélioration du classique test de Student (basé sur l'indépendance des erreurs). L'idée est de multiplier la statistique de Student  $t$  par un facteur approprié qui prenne en compte le comportement longue mémoire du processus. Le test considéré est basé sur la statistique  $t^* = (\bar{X}_T - \mu) c^{-1/2}(\hat{H}) T^{1-\hat{H}}$  si  $\hat{H} \in (0, 1)$  et  $t = 0$ , sinon. On a au préalable estimé  $H$  par l'une des méthodes exposées au paragraphe précédent par exemple, et l'on suppose que la vitesse de convergence de l'estimateur  $\hat{H}$  est plus petite que  $n^{-1/2}$ . La distribution asymptotique de  $t^*$  est obtenue. Elle permet de construire des intervalles de confiance pour  $\bar{X}_T$ , voir Beran (1989).

### 2. Test de discrimination.

Sowell (1990) cherche à discriminer entre un modèle (1) présentant une tendance déterministe et un modèle (2) avec "unit root", à savoir :

$$(1) \quad X(t) = \mu t + \varepsilon(t)$$

et

$$(2) \quad X(t) = \mu + X(t-1) + \varepsilon(t)$$

Il étudie les différences  $\Delta X(t)$  et il se sert de la détermination de l'ordre d'un modèle fractionnaire ajusté pour discriminer (1) et (2). Si  $d$  est proche de  $-1$ , on décide le modèle (1), si  $d$  est proche de  $0$ , on décide le modèle (2).

Robinson (1990) propose un test, à partir des multiplicateurs de Lagrange permettant de discriminer entre un FARMA  $(0, d, 0)$  sous  $H_0$  et un modèle ARCH sous  $H_1$ .

### 3. Test d'adéquation.

Un processus fractionnaire FARMA  $(p, d, q)$  étant ajusté et  $\mu$  étant supposé connu, Li et McLeod (1986) propose de tester l'adéquation du modèle à partir d'une généralisation du test de Portmanteau. La démarche est la même que celle de Ljung et Box (1978).

### 4. Test de "Longue mémoire".

Etant donné une série vectorielle  $Y(t) = (Y_1(t), \dots, Y_k(t))$ , on cherche à prendre en compte un comportement "à long terme" contre un comportement "à court terme". Pour cela on définit un vecteur d'équilibre :

$$\alpha'_0 Y(t) = \alpha_{10} Y_1(t) + \dots + \alpha_k Y_n(t).$$

Si  $m_i$  représente la moyenne de chaque série  $Y_i(t)$ , on considère la condition suivante :

$$H_0 : \alpha' m = \alpha_{10} m_1 + \alpha_{20} m_2 + \dots + \alpha_{k0} m_k = 0.$$

La condition  $H_0$  implique la nullité de la série  $\alpha'_0 Y(t)$  à chaque  $t$ . C'est une condition d'équilibre à court terme. On teste donc  $H_0 : \alpha' m = 0$  contre  $H_1 : \alpha' m \neq 0$ . La

statistique considérée est :

$$t(\alpha_0) = \frac{\sqrt{T} \alpha_0' \bar{Y}_T}{\sqrt{\alpha_0' \sum_{h=-H}^H \hat{\Gamma}(h) \cdot \alpha_0}}$$

où  $\Gamma(h) = \text{cov}(Y(t), Y(t+h))$ . Cette statistique est asymptotiquement normale sous  $H_0$ . On accepte  $H_0$  si  $|t(\alpha_0)| < 1,96$ , sinon on rejette. Ce test permet donc de détecter un comportement en moyenne de court terme.

Une autre approche est basée sur les corrélations du processus  $Y(t)$ . On suppose au préalable que le processus admet une représentation sous forme de moyenne mobile :

$$Y(t) = m + \varepsilon(t) + \theta_1 \varepsilon(t-1) + \dots + \theta_q \varepsilon(t-q)$$

où  $\{\varepsilon(t)\}$  est un bruit blanc ayant une matrice de covariance  $\Gamma$  inversible. On cherche alors à établir une notion de codépendance et on dit que  $\alpha_0$  est un vecteur de codépendance de degré  $b$  si et seulement si  $\alpha_0' Y(t)$  est non corrélé avec  $\varepsilon(t-q+b-1)$ ,  $\varepsilon(t-q+b-2)$ , ..., soit si l'on a la condition

$$H_0 : \alpha_0' \Gamma(q-b+1) = \alpha_0' \Gamma(q-b+2) = \dots = \alpha_0' \Gamma(q) = 0.$$

Posons  $\Lambda(q-b+1) = \alpha_0' \hat{\Gamma}(q-b+1)$  où  $\hat{\Gamma}(k)$  représente les covariances empiriques des  $(Y_j(t))_{j=1, \dots, T}$ , et  $\hat{Q}$  la matrice d'élément générique :

$$q_{hh'} = \text{cov}[\sqrt{T} - \Lambda(q-h), \sqrt{T} \Lambda(q-h')].$$

On considère alors la statistique de Wald :

$$W_T = T[\Lambda'(q), \dots, \Lambda'(q-b+1)] \hat{Q}^{-1}[\Lambda'(q), \dots, \Lambda'(q-b+1)]$$

qui asymptotiquement suit un  $\chi_{Tb}^2$ . Le test est le suivant : on accepte l'hypothèse de codépendance si  $W_T < \chi_{1-\alpha}^2(Tb)$ , sinon on rejette.

Ces deux tests ont été développés par Gouriéroux et Peaucelle (1988).

Comme on peut le voir la théorie des tests concernant les processus "longue mémoire" reste à faire tant dans la construction de tests spécifiques que de tests de discrimination, leurs propriétés théoriques devant être étudiées.

## V - La prédiction dans les processus longue mémoire.

Lorsque Granger et Joyeux (1980) ont introduit le processus FARMA( $p, d, q$ ) leur but était de disposer d'un modèle qui prendrait en compte de manière plus pertinente que ne le fait le modèle ARIMA de Box-Jenkins (1976) par exemple, les prévisions à long terme, quand la fonction d'autocorrélation du processus a tendance à décroître lentement vers zéro. Or malgré l'intérêt fondamental des prévisions, tant pour des séries économiques qu'hydrologiques, peu de travaux sur cette approche ont vu le jour.

Peiris et Perera (1988) donnent la relation permettant de calculer pour un modèle FARMA( $p, d, q$ ) à la fois

- Le prédicteur optimal linéaire  $\hat{X}(t + \ell)$ ,  $\ell > 0$  défini par

$$\hat{X}(t + \ell) = \mathbf{P}_{\mathcal{L}_t^2(X)} X(t + \ell), \quad \ell > 0$$

où  $\mathbf{P}_{\mathcal{L}_t^2(X)}$  représente la projection orthogonale de  $\mathcal{L}^2(X)$  sur  $\mathcal{L}_t^2(X)$ , ( $\mathcal{L}^2(X)$  représentant le sous espace fermé de  $\mathcal{L}^2(\cdot)$  engendré par tous les  $X(t)$  et  $\mathcal{L}^2(\cdot)$  l'espace de Hilbert des v.a. de moyenne zéro et de variance finie), et

- le prédicteur non linéaire optimal  $\tilde{X}(t + \ell)$ ,  $\ell > 0$  défini par

$$\tilde{X}(t + \ell) = E[X(t + \ell)/X(t), X(t - 1) \dots]$$

Quelques études expérimentales sont basées sur l'étude des prévisions de processus à mémoire longue. L'idée de base étant surtout de se servir des prévisions d'un processus pour choisir entre un modèle à mémoire courte (type autorégressif) et un processus à mémoire longue. L'intérêt est loin d'être négligeable d'ailleurs, car on retrouve ainsi posé le principe de parcimonie dont Whittle (1954) et Hannan (1970) avaient déjà parlé dans ce même contexte.

Une première approche est celle de Granger et Joyeux (1980) : on se donne des observations obtenues à partir de leurs autocorrélations dont on suppose qu'elles présentent un comportement caractéristique d'un processus longue mémoire. Le modèle sous jacent considéré est le modèle FARMA(0,  $d$ , 0). Pour chaque valeur de  $d$ ,  $-0,9 < d < 0,4$ , on construit un processus ARMA long. On calcule alors les prédicteurs pour chaque modèle à l'horizon 10 et 20 ainsi que la variance de l'erreur de prédiction. On constate que l'erreur de prédiction décroît pour  $d$  variant de  $-0,9$  à  $0,1$ , et que l'erreur croît pour  $d$  variant de  $0,1$  à  $0,4$ . D'autre part il semble que pour une valeur de  $d$  proche de  $0,5$  la méthode Box-Jenkins ait tendance à ajuster un modèle autorégressif d'ordre faible et que comparé au modèle FARMA, l'erreur de prévision soit de 30% plus faible pour le FARMA à l'horizon 10 et 27% à l'horizon 20.

Une deuxième approche est celle de Geweke et Porter-Hudak (1983) : ils ajustent 4 types de modèles et choisissent comme critère de sélection celui dont l'erreur de prédiction à l'horizon 10 ou 20 est la plus faible. Les 4 modèles retenus sont les suivants :

$$(a) \quad (1 - B)^d X(t) = \varepsilon(t) \quad -\frac{1}{2} \leq d \leq \frac{1}{2}$$

$$(b) \quad \phi(B)(1 - B)^d X(t) = \theta(B)\varepsilon(t) \quad -\frac{1}{2} \leq d \leq \frac{1}{2}$$

$$(c) \quad ARIMA(p, d, q) \quad d \in \mathbb{N}$$

$$(d) \quad AR(50)$$

Les modèles (a) et (b) sont estimés à l'aide du périodogramme. Ils fournissent de bonnes prédictions à long terme. Le mode (c) est un modèle qui classiquement fournit de bonnes prédictions à court terme, tandis que le modèle (d) de part le grand nombre de ses termes doit normalement prendre en compte un comportement longue mémoire. Pour l'exemple choisi, les résultats donnent :

## Erreur de prédiction

	$h = 10$	$h = 20$
(10)	8,359	17,837
(11)	7,804	16,996
(12)	8,969	18,647
(13)	8,046	16,757

## Références.

- Adenstedt R.K. (1974) : *On large sample estimation for the mean of a stationary random sequences*. Annals of Stat. Vol.2.6. 1095-1107.
- Anderson T.W. (1971) : *The statistical analysis of Time Series*. Wiley. N.Y.
- Beran J. (1989) : *A test of location for data with slowly decaying serial correlations*. Biometrika. 76. 261-269.
- Beran J., Kunsch H.R. (1985) : *Location estimators for processus with long range dependence*. Research report n° 50. Zurich.
- Box G.E.P., Jenkins G.H. (1976) : *Time Series Analysis : Forecasting and control*. 2<sup>nd</sup> ed. San Francisco. Holden Day.
- Carlin J.B., Dempster A.P. (1989) : *Sensitivity analysis of seasonal adjustments : empirical case studies*. J.A.S.A. vol. 84. 405. 6-32.
- Carlin J.B., Dempster A.P., Jonas A.B. (1985) : *On methods and models for bayesian time series analysis*. J. Econometrics. 30. 67-70.
- Dalhaus R. (1989) : *Efficient parameter estimation for self-similar processes*. Annals of stat. Vol. 17.4. 1749-1766.
- Fox R., Taqqu M. (1989) : *Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series*. Annals. of stat. Vol. 14.2. 517-532.
- Geweke J., Porter-Hudak S. (1983) : *The estimation and application of long memory time series models*. J.T.S.A. Vol. 4. 4. 221-238.
- Gouriéroux C., Peaucelle J. (1988) : *Detecting a long-run relationship*. CEPREMAP.

n° 8909. Paris.

Graf H., Hampel F.R., Tacier J.D. (1984) : *The problems of unsuspected serial correlation*. Lecture notes in statistics. 127-145.

Granger C.W.J., Joyeux R. (1980) : *An introduction to long-memory Time Series models and fractional differencing*. J.T.S.A. Vol. 1. 1. 15-29.

Grenander U. (1954) : *On estimation of regression coefficients in the case of an uncorrelated disturbance*. Ann. Math. Stat. 25. 252-272.

Grenander U., Rosenblatt M. (1957) : *Statistical Analysis of Stationary Time Series*. Wiley N.Y.

Hannan E.J. (1970) : *Multiple Time Series*. New York. Wiley.

Hannan E.J. (1979) : *The central Limit theorem for time series regression*. Stoch. Proc. and their Appl. 9. 281-289.

Hosking J.R.M. (1981) : *Fractional differencing*. Biometrika. 68.1. 165-174.

Huber P.J. (1981) : *Robust statistics*. Wiley. N.Y.

Hurst H.E. (1954) : *Long terme storage capacity reservoirs*. Trans. Amer. Soc. Civil Engineers. 116. 770-799.

Janacek G.J. (1982) : *Determining the degree of differencing for time Series via log-spectrum*. J.T.S.A. Vol. 3.3. 177-183.

Kashyap R.L., Eom K.B. (1988) : *Estimating in long memory Time Series models*. J.T.S.A. Vol. 9. 1. 35-41.

Kunsch H.R., Beran J., Hampel F.R. (1990) : *Robustness against dependence contrast versus constants*. Techn. Report A & M University and Zurich University.

Li W.K., McLeod A.I. (1986) : *Fractional Time Series modelling*. Biometrika. 73. 1. 217-221.

Ljung G.M., Box G.E.P. (1978) : *On a mesure of lack of fit in time series models*. Biometrika. 65. 297-303.

Mandelbrot B., Taquq M.S. (1979) : *Robust R/S analysis of long-run serial correlation*. 42<sup>nd</sup> session of the ISI. Manile. 4-14.

Mandelbrot B., Wallis J.R. (1969) : *Robustness of the rescaled range R/S in the measurement of noncyclic long-run statistical dependence*. Water Ressources Research. Vol. 5. 5. 967-988.

Peiris M.S., Perera B.J.C. (1988) : *On prediction with fractionally differenced ARIMA*



processes. J.T.S.A. Vol. 9. 3. 215-220.

**Robinson P.H.** (1990) : *Testing for strong serial correlation and dynamics conditional heteroskedasticity in multiple regression*. Journ. of Econometrics. Forthcoming.

**Rosenblatt M.** (1979) : *Some limit theorems for partial sums of quadratic forms in stationary Gaussian variables*. Z. Wahrsch. Verw. Gebiete. 49. 125-132.

**Samarov A., Taqqu M.S.** (1988) : *On the efficiency of the sample mean in long-memory noise*. J.T.S.A. Vol. 9. 2. 191-200.

**Sowell F.** (1990) : *Maximum likelihood estimation of stationary univariate fractionally integrated time series models*. Techn. Report. Carnegie Mellon University.

**Yagima Y.** (1985) : *On estimation of long-memory time series models*. Austr. Journ. Stat. 27. 3. 303-320.

**Yagima Y.** (1988) : *On estimation of the regression model with long memory stationary errors*. Annals of Stat. Vol. 16. 2. 791-807.

**Yagima Y.** (1990) : *Asymptotic properties of the LSE in a regression model with long-memory stationary processes*. Techn. Report. Univ. of Tokyo.

**Whittle P.** (1954) : *The statistical analysis of a sciche record*. J. Marine. Res. 13. 76-100.