Estimation des modèles FARIMA avec un bruit non corrélé mais non indépendant

Youssef Esstafa [1] , Yacouba Boubacar Maïnassara [2] & Bruno Saussereau [3]

[1] *youssef.esstafa@univ-fcomte.fr*

[2] *yacouba.boubacar_mainassara@univ-fcomte.fr*

[3] *bruno.saussereau@univ-fcomte.fr*

[1,2,3] *Université Bourgogne Franche-Comté,*
Laboratoire de mathématiques de Besançon,
UMR CNRS 6623,
16 route de Gray,
25030 Besançon, France.

Résumé. Dans ce travail, nous étudions les propriétés asymptotiques (convergence et normalité) de l'estimateur des moindres carrés des paramètres d'un modèle FARIMA (pour Fractionally AutoRegressive Integrated Moving-Average) avec un bruit non corrélé mais qui peut contenir des dépendances non linéaires. Les modèles FARIMA occupent une place centrale pour la modélisation des processus à mémoire longue, ils permettent d'identifier les phénomènes de persistance. Relâcher l'hypothèse standard d'indépendance sur le bruit permet à ces modèles de couvrir une large classe de processus à mémoire longue non linéaires. La convergence forte et la normalité asymptotique de l'estimateur sont démontrées sous certaines hypothèses d'ergodicité et de mélange.
Mots-clés. Processus à mémoire longue, les modèles FARIMA, l'estimateur des moindres carrés, la consistance, la normalité asymptotique ...

Abstract. This work considers the problem of estimating a fractionally integrated autoregressive moving average (FARIMA) model when only ergodic and mixing assumptions can be made concerning the noise process. Relaxing the independence assumption considerably extends the range of application of the FARIMA models, and allows to cover linear representations of general nonlinear processes. The method of least squares is used to estimate the parameters of the model. Strong consistency and asymptotic normality are shown to hold for the estimator.

Keywords. Long memory process, FARIMA models, least squares estimator, consistency, asymptotic normality ...

1 Introduction

La notion de mémoire longue est apparue au début des années 1950, historiquement pour l'étude du comportement inhabituel des niveaux du fleuve Nil en Égypte (voir Hurst [1951], Mandelbrot [1965]). Les processus à mémoire longue occupent une place de plus en plus importante dans la littérature des séries temporelles (voir Granger and Joyeux [1980], Fox and Taqqu [1986], Beran [1992]). En effet, les processus à mémoire longue s'avèrent plus adaptés à l'étude de certaines séries chronologiques issues par exemple de l'hydrologie, l'économie, la climatologie, l'économétrie financière et les trafics informatiques.

Afin de modéliser le comportement de mémoire longue, plusieurs modèles peuvent être utilisés. Les processus FARIMA sont parmi les modèles les plus connus et les plus utilisés dans ce contexte, ils sont une généralisation des modèles ARIMA, pour lesquels le paramètre de différenciation est un entier. Les processus FARIMA permettent au paramètre de différenciation de prendre des valeurs réelles.

Nous appelons FARIMA forts les modèles standards dans lesquels le terme d'erreur est supposé être une suite indépendante et identiquement distribuée (*i.e.* iid), et nous parlons de modèles FARIMA faibles quand les hypothèses sur le bruit sont moins restrictives. Le problème qui nous préoccupera sera l'analyse statistique de ces modèles.

Plusieurs méthodes d'estimation des paramètres des modèles FARIMA forts ont été proposées depuis les années 1950. Sous certaines conditions de régularités fortes sur la densité spectrale du processus, Dalhaus [1989] a montré la convergence et la normalité asymptotique de l'estimateur du maximum de vraisemblance exacte des paramètres d'un modèle FARIMA(p, d, q) et Fox et Taqqu [1986] ont établi les propriétés asymptotiques de l'estimateur du maximum de vraisemblance approximé de Whittle [1951].

Francq et Zakoïan [1998] ont montré la convergence forte et la normalité asymptotique de l'estimateur des moindres carrés des paramètres d'un modèles ARMA (AutoRegressive Moving-Average) faible. Dans ce travail, nous proposons une extension de leur procédure aux cas des modèles FARIMA faibles. Nous établissons la convergence forte et la normalité asymptotique des paramètres des modèles FARIMA faibles. Une attention particulière a été consacrée à l'estimation de la matrice de variance asymptotique.

2 Modèles et hypothèses

Soit $(X_t)_{t\in\mathbb{Z}}$ un processus stationnaire au second ordre, à valeurs réelles, tel que pour tout $t \in \mathbb{Z}$,

$$a(L)\,(1-L)^{d_0}\,X_t = b(L)\epsilon_t, \tag{1}$$

où $(\epsilon_t)_{t\in\mathbb{Z}}$ est une suite de variables aléatoires centrées ($\mathbb{E}\left[\epsilon_t\right] = 0, \forall t$), non corrélées, et de même variance $\sigma_\epsilon^2 > 0$ sur un certain espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. Les polynômes $a(z) = 1 + \sum_{i=1}^p a_i z^i$ et $b(z) = 1 + \sum_{j=1}^q b_j z^j$ ($p, q \in \mathbb{N}$ et $a_1, ..., a_p, b_1, ..., b_q \in \mathbb{R}$) ont leurs racines en dehors du disque unité et n'ont aucun zéro en commun et $0 < d_0 < 1/2$. Supposons, sans perte de généralité, que $a_p^2 + b_q^2 \neq 0$ (par convention $a_0 = b_0 = 1$).

$(1-L)^{d_0}$ est l'opérateur de différenciation fractionnaire, il est défini en utilisant la formule du binôme généralisée par :

$$(1-L)^{d_0} = \sum_{j=0}^{+\infty} \alpha_j L^j,$$

où pour tout $j \in \mathbb{N}$,

$$\alpha_j = \frac{\Gamma(j-d_0)}{\Gamma(j+1)\Gamma(-d_0)},$$

avec $\Gamma(.)$ est la fonction Gamma.

Le vecteur des vraies valeurs des paramètres

$$\theta_0 := (a_1, a_2, ..., a_p, b_1, b_2, ..., b_q, d_0)'$$

appartient à l'espace compact des paramètres

$$\begin{aligned} \Theta := \{\theta = (\theta_1, \theta_2, ..., \theta_{p+q}, d)';\ a_\theta(z) = 1 + \theta_1 z + ... + \theta_p z^p \\ \text{et } b_\theta(z) = 1 + \theta_{p+1} z + ... + \theta_{p+q} z^q \text{ ont leurs zéros en dehors du disque unité}\}. \end{aligned}$$

Pour tout $\theta \in \Theta$, soit $(\epsilon_t(\theta))_{t\in\mathbb{Z}}$ le processus stationnaire au second ordre solution de

$$\epsilon_t(\theta) = a_\theta(L)(1-L)^d X_t - \sum_{j=1}^q \theta_{p+j}\epsilon_{t-j}(\theta),\ \forall t \in \mathbb{Z}.$$

Remarquons, en utilisant ces notations, que $\epsilon_t(\theta_0) = \epsilon_t$ p.s. $\forall t \in \mathbb{Z}$.
Étant donné une observation de longueur n, $X_1, X_2, ..., X_n$, les variables aléatoires $\epsilon_t(\theta)$ peuvent être approximées par $\tilde{\epsilon}_t(\theta)$ définies récursivement comme solutions de

$$\tilde{\epsilon}_t(\theta) = a_\theta(L)(1-L)^d X_t - \sum_{j=1}^q \theta_{p+j}\tilde{\epsilon}_{t-j}(\theta), \quad \forall 0 < t \leq n.$$

Les valeurs initiales inconnues sont remplacées par zéro :

$$\forall t \leq 0,\ X_t = 0 \text{ et } \tilde{\epsilon}_t(\theta) = 0.$$

La variable aléatoire $\widehat{\theta}_n$ est dite estimateur des moindres carrés de θ_0, si, presque sûrement, elle satisfait

$$Q_n(\widehat{\theta}_n) = \min_{\theta \in \Theta} Q_n(\theta),$$

où

$$Q_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \tilde{\epsilon}_t^2(\theta).$$

Pour prouver la convergence forte et la normalité asymptotique de l'estimateur des moindres carrés, il sera commode de considérer la fonction

$$O_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \epsilon_t^2(\theta).$$

2.1 Propriétés asymptotiques de l'estimateur des moindres carrés

Le premier résultat principal est la convergence forte de l'estimateur des moindres carrés, cette propriété asymptotique est prouvée sous certaines conditions de régularités sur le processus $(\epsilon_t)_{t\in\mathbb{Z}}$.

Soit l'hypothèse :

H1 : Le processus $(\epsilon_t)_{t\in\mathbb{Z}}$ est strictement stationnaire et ergodique.
Le premier résultat est donné par le théorème suivant :

Théorème 1 (Convergence forte) Supposons que $(X_t)_{t\in\mathbb{Z}}$ vérifie (1) et appartient à L^2. Sous les hypothèses précédentes et **H1**, nous avons

$$\widehat{\theta}_n \xrightarrow[n\to+\infty]{p.s.} \theta_0.$$

Le deuxième résultat porte sur la normalité asymptotique de l'estimateur des moindres carrés, il exige deux hypothèses supplémentaires, une sur les moments d'ordres 4 de $(\epsilon_t)_{t\in\mathbb{Z}}$ et les coefficients de mélanges et l'autre sur l'appartenance du vrai paramètre à l'intérieur de l'espace compact des paramètres.

Notons $\mathcal{F}_{-\infty}^{t}$ la $\sigma-$ algèbre engendrée par $\{\epsilon_u : u \leq t\}$ et $\mathcal{F}_{t+k}^{\infty}$ la tribu générée par $\{\epsilon_u : u \geq t+k\}$. Alors, les coefficients de mélange forts $(\alpha_\epsilon(h))_{h\in\mathbb{N}^*}$ du processus stationnaire $(\epsilon_t)_{t\in\mathbb{Z}}$ sont définis pour tout h par :

$$\alpha_\epsilon(h) = \sup_{A\in\mathcal{F}_{-\infty}^{t},\ B\in\mathcal{F}_{t+k}^{\infty}} \mid P(A\cap B) - P(A)P(B) \mid .$$

Soient les hypothèses suivantes :

H2 : Il existe un réel $\nu > 0$ tel que $\mathbb{E}\left[\mid \epsilon_t \mid^{4+2\nu}\right] < \infty$ et les coefficients de mélange du processus $(\epsilon_t)_{t\in\mathbb{Z}}$ vérifient $\sum_{h=1}^{\infty} h^2 \left\{\alpha_\epsilon(h)\right\}^{(\nu-1)/\nu} < \infty$.

H3 : Nous avons $\theta_0 \in \overset{\circ}{\Theta}$, où $\overset{\circ}{\Theta}$ est l'intérieur de Θ.

La deuxième partie de l'hypothèse **H2** sur les coefficients de mélanges est donnée dans Donald W. K. Andrews [1991].

Théorème 2 (Normalité asymptotique) Supposons que $(X_t)_{t\in\mathbb{Z}}$ vérifie (1) et appartient à L^2. Sous les hypothèses du théorème 1 et **H2–H3**, nous avons

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \underset{n\to+\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}\left(0, J^{-1}IJ^{-1}\right),$$

où $J = J(\theta_0)$, et $I = I(\theta_0)$, avec

$$J(\theta) = \lim_{n\to\infty}\left\{\frac{\partial^2}{\partial\theta\partial\theta'}Q_n(\theta)\right\}\ p.s. \text{ et } I(\theta) = \lim_{n\to\infty} Var\left\{\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta)\right\}.$$

Remarque : Dans le cas fort, la matrice I vaut $2\sigma_\epsilon^2 J$, et donc la matrice de variance asymptotique se réduit à $2\sigma_\epsilon^2 J^{-1}$.

3 Estimation de la matrice de variance

Afin d'obtenir des intervalles de confiance ou de tester la significativité des coefficients FARIMA faibles, il sera nécessaire de disposer d'un estimateur au moins faiblement convergent de la matrice de variance asymptotique $J^{-1}IJ^{-1}$. La matrice J peut facilement être estimée empiriquement par :

$$\widehat{J} = \frac{2}{n}\sum_{t=1}^{n}\left\{\frac{\partial\tilde{\epsilon}_t(\theta)}{\partial\theta}\frac{\partial\tilde{\epsilon}_t(\theta)}{\partial\theta'}\right\}_{\theta=\widehat{\theta}_n}.$$

Notons

$$I = \lim_{n\to\infty} Var\left\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Upsilon_t\right\},\ \text{où } \Upsilon_t = 2\left\{\epsilon_t(\theta)\frac{\partial\epsilon_t(\theta)}{\partial\theta}\right\}_{\theta=\theta_0}.$$

Nous utiliserons la méthode d'estimation paramétrique de la densité spectrale introduite par Berk [1974]. Notons $\widehat{\mathbf{\Phi}}_r(z) = I_{p+q+1} + \sum_{i=1}^{r} \widehat{\Phi}_{r,i} z^i$, où $\widehat{\Phi}_{r,1}, ..., \widehat{\Phi}_{r,r}$ sont les coefficients de la régression des moindres carrés de $\widehat{\Upsilon}_t$ sur $\widehat{\Upsilon}_{t-1}, ..., \widehat{\Upsilon}_{t-r}$ et par $\widehat{\Sigma}_{\widehat{u}_r}$ la variance empirique de ces résidus.

Théorème 3 Supposons qu'il existe $\delta > 0$ tel que $\mathbb{E}\left[\mid \epsilon_t \mid^{8+4\delta}\right] < \infty$. Sous les hypothèses du théorème 2 et sous certaines hypothèses de régularités, nous avons

$$\widehat{I}^{SP} := \widehat{\mathbf{\Phi}}_r^{-1}(1)\widehat{\Sigma}_{\widehat{u}_r}\widehat{\mathbf{\Phi}}_r'^{-1}(1) \longrightarrow I$$

en probabilité quand $r = r(n) \to \infty$ et $r^3/n \to 0$ quand $n \to \infty$.

Bibliographie

[1] D. W. K. Andrews (1991), Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation, *Econometrica, Vol. 59, No. 3, 817-858.*

[2] J. Beran (1992), Statistical methods for data with long-range dependence. *Statistical science.,Vol, 7(4) :404–427.*

[3] Berk, K.N. (1974), Consistent Autoregressive Spectral Estimates. *Annals of Statistics,* **2***, 489–502.*

[4] R. Dahlhaus (1989), Efficient parameter estimation for self-similar processes. *Ann. Statist., 17(4) :1749–1766. ISSN 0090-5364.*

[5] R. Fox and M. S. Taqqu (1986), Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist., 14(2) :517–532. ISSN0090-5364.*

[6] Francq, C. and Zakoïan, J-M. (1998), Estimating linear representations of nonlinear processes, *Journal of Statistical Planing and Inference 68,145-165.*

[7] C. W. J. Granger and R. Joyeux (1980), An introduction to long-memory time series models and fractional differencing. *J. Time Ser. Anal., 1(1) :15–29.*

[8] H. E. Hurst (1951), Long term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers, 116 :770–779.*

[9] B. Mandelbrot (1965), Une classe processus stochastiques homothétiques à soi ; application à la loi climatologique H. E. Hurst. *C. R. Acad. Sci. Paris, 260 :3274–3277.*