Estimation de la variance asymptotique de l'estimateur des moindres carrés des modèles FARIMA faibles

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

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

[2] *youssef.esstafa@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 considérons le problème de l'estimation de la matrice de variance asymptotique de l'estimateur des moindres carrés des modèles FARIMA (pour Fractionally AutoRegressive Integrated Moving-Average) dans le cas où les erreurs sont supposées non-corrélées mais non nécessairement indépendantes ni même des différences de martingales. Nous proposons un estimateur convergent de cette matrice de variance asymptotique et nous illustrons ensuite les résultats théoriques obtenus par des simulations.

Mots-clés. Processus non-linéaires, modèles FARIMA faibles, estimateur des moindres carrés, estimateur de la matrice de covariance, convergence.

Abstract. In this work, we propose a consistent estimator of the limiting variance matrix of the least-squares estimator of Fractionally AutoRegressive Integrated Moving-Average (FARIMA) models with uncorrelated but not necessarily independent nor martingale differences errors (i.e. weak FARIMA models). The theoretical results are illustrated by means Monte Carlo experiments.

Keywords. Non-linear processes, weak FARIMA models, least-squares estimator, covariance matrix estimate, consistency.

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 et Joyeux [1980], Fox et 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.

On dit qu'un processus $(X_t)_{t\in\mathbb{Z}}$ centré et stationnaire au second ordre, à valeurs réelles, admet une représentation FARIMA(p, d_0, q) faible si, 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, L est l'opérateur retard et $0 < d_0 < 1/2$.
Supposons, sans perte de généralité, que $a_p^2 + b_q^2 \neq 0$ (par convention $a_0 = b_0 = 1$).

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

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

où pour tout $j \in \mathbb{N}$ et pour tout $d \in \left]0, 1/2\right[$,

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

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

Par opposition au modèle FARIMA faible donné dans (1), nous appelons FARIMA forts les modèles standards dans lesquels le terme d'erreur est supposé être une suite indépendante et identiquement distribuée (iid). Dans la modélisation des séries temporelles, la validité des différentes étapes de la méthodologie traditionnelle de Box et Jenkins, à savoir les étapes d'identification, d'estimation et de validation, dépend des propriétés du bruit blanc. 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. L'objectif de ce travail est de proposer un estimateur convergent de la matrice de variance asymptotique des modèles FARIMA faibles en utilisant la méthode des noyaux.

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

$$\Theta := \{\theta = (\theta_1, \theta_2, ..., \theta_{p+q}, d)';\ a_\theta(z) = 1 + \sum_{j=1}^{p} \theta_j z^j \text{ et } b_\theta(z) = 1 + \sum_{j=1}^{q} \theta_{p+j} z^j$$
$$\text{ont leurs zéros en dehors du disque unité et n'ont aucune racine commune}$$
$$\text{et } d \in [d_1, d_2] \subset]0, 1/2[\}.$$

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

$$\epsilon_t(\theta) = \sum_{j\geq 0} \alpha_j(d) X_{t-j} + \sum_{i=1}^{p} \theta_i \sum_{j\geq 0} \alpha_j(d) X_{t-i-j} - \sum_{j=1}^{q} \theta_{p+j} \epsilon_{t-j}(\theta). \tag{2}$$

Notons que $\epsilon_t(\theta_0) = \epsilon_t$ p.s.. Étant donné une réalisation de longueur n, $X_1, X_2, ..., X_n$, les variables aléatoires $\epsilon_t(\theta)$ peuvent être approximées, pour $0 < t \leq n$, par $\tilde{\epsilon}_t(\theta)$ (avec $\tilde{\epsilon}_t(\theta) = X_t = 0$ si $t \leq 0$) définies récursivement comme solutions de

$$\tilde{\epsilon}_t(\theta) = \sum_{j=0}^{t-1} \alpha_j(d) X_{t-j} + \sum_{i=1}^{p} \theta_i \sum_{j=0}^{t-i-1} \alpha_j(d) X_{t-i-j} - \sum_{j=1}^{q} \theta_{p+j} \tilde{\epsilon}_{t-j}(\theta). \tag{3}$$

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

$$Q_n(\widehat{\theta}_n) = \min_{\theta\in\Theta} Q_n(\theta), \text{ où } Q_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \tilde{\epsilon}_t^2(\theta).$$

Les premiers principaux résultats de ce travail portent sur les propriétés asymptotiques de l'estimateur des moindres carrés et sont obtenus par Boubacar Maïnassara, Y. Esstafa, Y. et Saussereau, B. [2018]. Ces propriétés ont été prouvées sous certaines conditions de régularités sur le processus $(\epsilon_t)_{t\in\mathbb{Z}}$, dont celles essentielles sont :
H1 : Le processus $(\epsilon_t)_{t\in\mathbb{Z}}$ est strictement stationnaire et ergodique.
H2 : Nous avons $\mathbb{E}|\epsilon_t|^{4+2\nu} < \infty$ et $\sum_{h=0}^{\infty} \{\alpha_\epsilon(h)\}^{\nu/2+\nu} < \infty$ pour un certain $\nu > 0$.
H3 : Supposons que $\sum_{i,j,k\in\mathbb{Z}} |\text{cum}(\epsilon_0, \epsilon_i, \epsilon_j, \epsilon_k)| < \infty$.

Pour tout délai h, $\alpha_\epsilon(h)$ définit le coefficient de mélange à l'ordre h du processus $(\epsilon_t)_t$. Ce coefficient permet de contrôler la dépendance probabiliste entre les tribus engendrées par

ϵ_u pour $u \leq t$ et ϵ_u pour $u \geq t+h$. Dans l'hypothèse **H3** nous supposons la sommabilité des cumulants joints d'ordre 4 de $(\epsilon_t)_t$, ce qui permet de contrôler la dépendance statistique de ce processus. Le théorème suivant donne le comportement asymptotique de $(\hat{\theta}_n)_n$.

Théorème 1 *Supposons que* $(\epsilon_t)_{t\in\mathbb{Z}}$ *vérifie* (1). *Sous les hypothèses* **H1**, **H2** *et* **H3**, *nous avons*

$$\widehat{\theta}_n \xrightarrow[n\to+\infty]{p.s.} \theta_0 \text{ et } \sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \Sigma := J^{-1}IJ^{-1}\right),$$

où

$$J = 2\mathbb{E}\left[\frac{\partial \epsilon_t}{\partial \theta}\frac{\partial \epsilon_t}{\partial \theta'}\right] \; p.s. \quad et \; I = \sum_{k=-\infty}^{\infty} \mathrm{Cov}(H_t(\theta_0), H_{t-k}(\theta_0))$$

avec

$$H_t(\theta) = 2\epsilon_t(\theta)\frac{\partial \epsilon_t(\theta)}{\partial \theta}, \quad \forall \theta \in \Theta.$$

2 Estimation de la matrice de variance asymptotique

La matrice de variance asymptotique de l'estimateur des moindres carrés des modèles FARIMA faibles dépend de deux matrices d'informations différentes. Il est donc nécessaires de fournir un estimateur convergent de chaque matrice. La matrice J peut facilement être estimée empiriquement par :

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

Remarquons que

$$I = \lim_{n\to\infty} Var\left\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n} H_t(\theta_0)\right\}.$$

En utilisant l'absolue sommabilité de la fonction d'autocovariance du processus $(H_t)_t$ et la stationnarité de ce dernier, nous obtenons que

$$I = \sum_{k=-\infty}^{\infty} \Delta_k(\theta_0),$$

où, pour tout $k \in \mathbb{Z}$,

$$\Delta_k(\theta_0) = \mathrm{Cov}(H_t(\theta_0), H_{t-k}(\theta_0)).$$

Remarque : Dans le cas fort, la matrice I vaut $2\sigma_\epsilon^2 J$, et donc la matrice de variance asymptotique se réduit à $\Sigma_s := 2\sigma_\epsilon^2 J^{-1}$. Dans le cas général des FARIMA faibles, $I \neq 2\sigma_\epsilon^2 J$ alors $\Sigma \neq \Sigma_s$.

La covariance $\Delta_k(\theta_0)$ sera estimée empiriquement, pour $0 \leq k < n$, par $\hat{\Delta}_k(\hat{\theta}_n)$ où, pour tout $\theta \in \Theta$,

$$\hat{\Delta}_k(\theta) = \frac{4}{n} \sum_{t=1}^{n-k} \tilde{\epsilon}_t(\theta) \frac{\partial \tilde{\epsilon}_t(\theta)}{\partial \theta} \left(\tilde{\epsilon}_{t+k}(\theta) \frac{\partial \tilde{\epsilon}_{t+k}(\theta)}{\partial \theta} \right)'$$

et

$$\hat{\Delta}_{-k}(\theta) = \hat{\Delta}_k(\theta)'.$$

Sous les hypothèses du Théorème 1, la statistique $\hat{\Delta}_k(\hat{\theta}_n)$ est un estimateur convergent de $\Delta_k(\theta_0)$. Par conséquent, on peut se demander si l'estimateur simple

$$\tilde{I}_n(\hat{\theta}_n) = \sum_{k=-n+1}^{n-1} \hat{\Delta}_k(\hat{\theta}_n)$$

est un estimateur convergent de I. La réponse est négative puisque, pour tout n,

$$\tilde{I}_n(\hat{\theta}_n) = \frac{4}{n} \left(\sum_{t=1}^{n} \tilde{\epsilon}_t(\hat{\theta}_n) \frac{\partial \tilde{\epsilon}_t(\hat{\theta}_n)}{\partial \theta} \right)^2 = n \left(\frac{\partial}{\partial \theta} Q_n(\hat{\theta}_n) \right)^2 = 0.$$

La matrice I est une somme infinie de moments d'ordre 4 de $(\epsilon_t)_t$, donc l'estimation de certains moments repose que sur quelques observations. Cela implique que certains moments sont probablement mal estimés (voir Francq, C. et Zakoïan, J.-M. [2000] pour une référence dans le cas des modèles ARMA faibles et voir Boubacar Maïnassara, Y. [2014] pour une extension au cas des modèles VARMA faibles). La solution classique utilisée pour résoudre ce problème consiste à pondérer les moments empiriques $\hat{\Delta}_k(\hat{\theta}_n)$. Le poids attribué à $\hat{\Delta}_k(\hat{\theta}_n)$ est de la forme $\omega(kh_n)$, où $(h_n)_{n\in\mathbb{N}}$ est une suite de nombres réels et la fonction $\omega : \mathbb{R} \to \mathbb{R}$ est supposée bornée, à support compact $[-c, c]$ et continue à l'origine avec $\omega(0) = 1$.

Un estimateur convergent de I est obtenu sans hypothèses supplémentaires sur le processus $(\epsilon_t)_t$.

Théorème 2 *Supposons que les hypothèses du Théorème 1 sont vérifiées et considérons la matrice*

$$\hat{I}_n(\theta) = \sum_{k=-T_n}^{T_n} \omega(kh_n) \hat{\Delta}_k(\theta),$$

où $T_n = [c/h_n]$ ($[x]$ est la partie entière de x). Si $(h_n)_n$ est choisie telle que

$$h_n \to 0 \text{ et } nh_n^2 \to \infty \text{ quand } n \to \infty,$$

alors

$$\hat{I}_n(\hat{\theta}_n) \to I \text{ en probabilité quand } n \to \infty.$$

Bibliographie

[1] Jan Beran. (1992) A goodness-of-fit test for time series with long range dependence. J. Roy. Statist. Soc. Ser. B, 54(3) :749–760.

[2] Yacouba Boubacar Maïnassara. (2014) Estimation of the variance of the quasi-maximum likelihood estimator of weak VARMA models. Electron. J. Stat., 8(2) :2701–2740.

[3] Boubacar Maïnassara, Y., Esstafa, Y. and Saussereau, B. (2018). Estimating FARIMA models with uncorrelated but non-independent error terms, preprint.

[4] Robert Fox and Murad S. Taqqu. (1986) Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. Ann. Statist., 14(2) :517–532.

[5] Christian Francq and Jean-Michel Zakoïan. (2000) Covariance matrix estimation for estimators of mixing weak ARMA models. J. Statist. Plann. Inference, 83(2) :369–394.

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

[7] H.E. HURST. (1951) Long-term storage capacity of reservoirs. pages 770–799.

[8] Benoit B. Mandelbrot. (1965) Une classe de processus stochastiques homothétiques à soi ; application à la loi climatologique de H. E. Hurst. 260 :3274–3277.