# Un Modèle FARIMA Localement Stationnaire

Li Song, Pascal Bondon

# Un Modèle FARIMA Localement Stationnaire

Li Song & Pascal Bondon

*CNRS UMR 8506, Université Paris-Sud, France.*

Résumé

Nous étudions le problème de la modélisation d'une série chronologique non stationnaire à longue mémoire au moyen d'un processus localement autorégressif à moyenne mobile fractionnairement intégrée. Le nombre de points de ruptures et leurs localisations sont inconnus ainsi que les paramètres de chaque sous-série. Nous présentons une méthode d'estimation des points de ruptures et des paramètres des sous-séries dont nous montrons les bonnes performances au moyen de simulations de Monte Carlo.

Abstract

We consider the problem of modeling a non-stationary long-memory time series using a piecewise fractional autoregressive integrated moving average process. The number as well as the locations of structural break points and the parameters of each segment are assumed to be unknown. A four-step procedure is proposed to find out the break points and to estimate the parameters of each segment. Its effectiveness is shown by Monte Carlo simulations.

## 1 Modèle

Le modèle autorégressif à moyenne mobile fractionnairement intégrée (FARIMA) est un processus stationnaire à longue mémoire introduit il y a une vingtaine d'années et qui est utilisé dans de nombreux domaines de l'hydrologie à l'économie, voir Beran (1994) et Doukhan et al. (2003). Cependant, certaines données ne sont pas stationnaires même après élimination de la tendance. En particulier, le paramètre de longue mémoire dans le modèle FARIMA peut dépendre du temps. On rencontre ce cas par exemple en géophysique, en océanographie et en météorologie, voir Ray & Tsay (2002), Wang & Wang (2006) et Stoev et al. (2006). Pour modéliser ces données, nous proposons un processus FARIMA localement stationnaire. Plus précisément, nous supposons que la série non stationnaire $\{Y_t\}$, $t = 1, \ldots, n$, peut être divisée en $m+1$ blocs stationnaires de types FARIMA. Pour $j = 1, \ldots, m$, $\tau_j$ désigne le point de rupture (PR) entre le $j$-ème et le $(j+1)$-ème processus FARIMA, et on pose $\tau_0 = 1$ et $\tau_{m+1} = n+1$. Le $j$-ème bloc de $\{Y_t\}$ est donc modélisé par

$$Y_t = X_{t,j}, \qquad \tau_{j-1} \leq t < \tau_j, \tag{1}$$

où $X_{t,j}$ est le modèle FARIMA $(p_j, d_j, q_j)$ défini par

$$\Phi_j(B)X_{t,j} = \Theta_j(B)(1-B)^{-d_j}\epsilon_t \tag{2}$$

avec $\epsilon_t \sim$ iid N$(0,1)$, $\Phi_j(z) = 1 - \phi_{j,1}z - \cdots - \phi_{j,p_j}z^p$, $\Theta_j(z) = 1 + \theta_{j,1}z + \cdots + \theta_{j,q_j}z^q$ et $d_j \in (0, 1/2)$. Soit $p \geq \max(p_j)$, $q \geq \max(q_j)$, $\alpha_j = (d_j, \phi_{j,1}, \ldots, \phi_{j,p}, \theta_{j,1}, \ldots, \theta_{j,q})$ où $\phi_{j,k} = 0$ si $k > p_j$ et $\theta_{j,k} = 0$ si $k > q_j$, et $\beta_j = (p_j, q_j, \alpha_j)$. Le vecteur $\beta_j$ contient les paramètres du $j$-ème modèle et $\beta_j$ est constant sur chaque intervalle $[\tau_{j-1}, \tau_j)$.

# 2 Estimation

Le problème d'ajustement du modèle (1)–(2) consiste à estimer $(\tau_1, \ldots, \tau_m, \beta_1, \ldots, \beta_{m+1})$. On divise la série $\{Y_t\}$ en blocs de longueur $E$ et on considère la série (éventuellement tronquée) formée des $K = [n/E]$ blocs définis sur les intervalles $I_k = ((k-1)E, kE]$ pour $k = 1, \ldots, K$. On suppose que $E$ est suffisamment petit pour qu'il y ait au plus un PR par intervalle $I_k$ et qu'au moins un intervalle sépare deux PR consécutifs. Par ailleurs, on suppose pour l'instant que $m$ est connu (voir remarque 1). Notre méthode d'estimation est constituée des quatre étapes suivantes :

**Étape 1 : estimation locale.** Pour chaque intervalle $I_k$, on estime le vecteur $\alpha_k$ par l'estimateur du maximum de vraisemblance gaussien (EMV) $\hat{\alpha}_k$ et on sélectionne une paire $(\hat{p}_k, \hat{q}_k)$ au moyen du critère BIC.

**Étape 2 : sélection des intervalles contenant un PR.** Si $\{Y_t\}$ satisfait (1)–(2), on s'attend à ce que $\hat{\beta}_k$ soit proche de $\beta_k$ s'il n'y pas de PR dans $I_k$ et si $E$ est suffisamment grand. À l'inverse, s'il y a un PR dans $I_k$ et pas de PR dans $I_{k-1}$ et $I_{k+1}$, $\hat{\beta}_k$ doit être assez différent de $\hat{\beta}_{k-1}$ et $\hat{\beta}_{k+1}$. Posons $k_0 = 0$, $k_{m+1} = K$, et

$$(\hat{k}_1, \ldots, \hat{k}_m) = \underset{1 \leq k_1 < \cdots < k_m < K}{\operatorname{argmin}} \sum_{j=1}^{m+1} \sum_{k=k_{j-1}+1}^{k_j} \left( \|\hat{\alpha}_k - \bar{\alpha}_j\|^2 + \psi(|\hat{p}_k - \bar{p}_j|) + \psi(|\hat{q}_k - \bar{q}_j|) \right), \tag{3}$$

où $\bar{\alpha}_j = \dfrac{1}{k_j - k_{j-1}} \displaystyle\sum_{k=k_{j-1}+1}^{k_j} \hat{\alpha}_k$, $\bar{p}_j$ (resp. $\bar{q}_j$) est l'ordre le plus fréquemment sélectionné parmi les ordres $\hat{p}_k$ (resp. $\hat{q}_k$) pour $k = k_{j-1}+1, \ldots, k_j$. Dans le cas où $\bar{p}_j$ (resp. $\bar{q}_j$) n'est pas unique, on choisit l'ordre le plus petit. La fonction $\psi(\cdot)$ est positive et croissante. Soit $J_k = ((k-0.5)E, (k+0.5)E]$ pour $k = 1, \ldots, K-1$. On choisit $(J_{\hat{k}_1}, \ldots, J_{\hat{k}_m})$ comme étant les intervalles contenant un PR.

**Étape 3 : estimation des PR.** Supposons que tous les intervalles $J_{\hat{k}_j}$ sont les bons, i.e., $\tau_j \in J_{\hat{k}_j}$ pour $j = 1, \ldots, m$. Soit $j$ fixé. Alors il n'y a pas de PR dans l'intervalle "précédant", $((\hat{k}_{j-1} + 0.5)E, (\hat{k}_j - 0.5)E]$ où $\hat{k}_0 + 0.5 = 0$, et on définit $\hat{\beta}_p$ comme l'EMV de $\beta_j$ utilisant toutes les données de cet intervalle. De la même manière, soit $\hat{\beta}_s$ l'EMV de $\beta_{j+1}$ basé sur les données dans l'intervalle "suivant", $((\hat{k}_j + 0.5)E, (\hat{k}_{j+1} - 0.5)E]$ où

$\hat{k}_{m+1} - 0.5 = K$. Les estimateurs $\hat{\beta}_p$ et $\hat{\beta}_s$ sont plus précis que des estimateurs locaux calculés à l'étape 1 car ils sont basés sur plus de données. Supposons que $l \in J_{\hat{k}_j}$ est le PR $\tau_j$ et soit $\hat{\beta}_{l_p}$ et $\hat{\beta}_{l_s}$ les EMV de $\beta_j$ et $\beta_{j+1}$ basés respectivement sur les données dans les intervalles $((\hat{k}_{j-1} + 0.5)E, l]$ et $(l, (\hat{k}_{j+1} - 0.5)E]$. Ces estimateurs doivent être proches respectivement de $\hat{\beta}_p$ et $\hat{\beta}_s$. Ceci justifie le critère suivant pour estimer le PR $\tau_j$,

$$\hat{\tau}_j = \underset{l \in J_{\hat{k}_j}}{\operatorname{argmin}} \big( \|\hat{\alpha}_{l_p} - \hat{\alpha}_p\|^2 + \psi(|\hat{p}_{l_p} - \hat{p}_p|) + \psi(|\hat{q}_{l_p} - \hat{q}_p|) + \|\hat{\alpha}_{l_s} - \hat{\alpha}_s\|^2 + \psi(|\hat{p}_{l_s} - \hat{p}_s|) + \psi(|\hat{q}_{l_s} - \hat{q}_s|) \big) . \quad (4)$$

Pour réduire la complexité, $\hat{\beta}_{l_p}$ et $\hat{\beta}_{l_s}$ sont calculés respectivement en utilisant les données dans $(l-E, l)$ et $(l, l+E)$, et ceci donne de bons résultats en pratique (voir la simulation).
**Étape 4 : estimation des paramètres de chaque bloc stationnaire.** Les PR $(\hat{\tau}_1, \ldots, \hat{\tau}_m)$ étant sélectionnés, on estime les paramètres $\beta_j$ des séries stationnaires $X_{t,j}$ pour $j = 1, \ldots, m+1$ par MV au moyen des données dans les intervalles $(\hat{\tau}_{j-1}, \hat{\tau}_j]$ où $\hat{\tau}_0 = 1$ et $\hat{\tau}_{m+1} = KE$.

**Remarque 1.** La méthode suppose que le nombre $m$ de PR est connu, ce qui n'est pas le cas en pratique. Une procédure pour estimer $m$ consiste à augmenter un à un le nombre de PR dans la méthode. Lorsque ce nombre est égal à $m+1$, on trouve deux intervalles très proches qui correspondent en fait au même PR. Ceci permet de déterminer $m$ quand les PR ne sont pas trop proches les uns des autres. En pratique, cette méthode fonctionne bien quand au moins $2E$ données séparent chaque PR.

# 3 Simulation

Nous générons 100 réalisations de tailles $n = 40000$ du modèle (1)–(2) défini dans le tableau 1 avec les quatre PR $\tau_1 = 7829$, $\tau_2 = 16341$, $\tau_3 = 23623$ et $\tau_4 = 32176$. On prend $E = 2000$, on a donc $K = 20$ et les PR sont dans les intervalles $J_4$, $J_8$, $J_{12}$ et $J_{16}$. On prend $\psi(x) = \ln(x+1)$ dans (3)–(4).
Le tableau 2 contient les ordres les plus fréquemment sélectionnés à l'étape 1 pour chaque intervalle $I_k$ dans les 100 simulations. Notons que les PR appartiennent aux intervalles $I_4$, $I_9$, $I_{12}$ et $I_{17}$. On voit que les bons ordres sont choisis en majorité, sauf lorsqu'il y a un PR dans l'intervalle (voir les colonnes en gras).
On effectue l'étape 2 pour $1 \leq m \leq 6$. Pour $m \leq 4$, tous les intervalles sélectionnés sont bien séparés, tandis que pour $m = 5, 6$, nous trouvons des intervalles proches les uns des autres ce qui conduit à prendre $m = 4$. La figure 1 montre les intervalles trouvés dans l'étape 2 où $m = 4$. Les intervalles les plus fréquemment choisis sont bien les bons.
Soit $\lambda_j = \tau_j / n$. Le tableau 3 contient les moyennes $\hat{\mu}(\hat{\lambda}_j)$ et les erreurs standards $\hat{\sigma}(\hat{\lambda}_j)$ sur les 100 expériences des estimations des PR obtenues dans l'étape 3. On voit que les estimations sont très précises.

| Paramètres | Série $X_{t,j}$ | | | | |
|---|---|---|---|---|---|
| $\beta_j$ | 1 | 2 | 3 | 4 | 5 |
| $p_j$ | 1 | 0 | 1 | 1 | 0 |
| $q_j$ | 2 | 0 | 0 | 1 | 1 |
| $d_j$ | 0.20 | 0.45 | 0.15 | 0.35 | 0.10 |
| $\phi_j$ | 0.50 | 0 | -0.80 | 0.30 | 0 |
| $\theta_j$ | (-0.60, -0.20) | 0 | 0 | -0.70 | 0.40 |

TAB. 1 – Ordres et paramètres du modèle.

| Intervalle | 1 | 2 | 3 | **4** | 5 | 6 | 7 | 8 | **9** | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Ordres $(\hat{p}_j, \hat{q}_j)$ | (1,2) | (1,2) | (1,2) | **(2,0)** | (0,0) | (0,0) | (0,0) | (0,0) | **(1,0)** | (1,0) |
| Fréquence | 48 | 50 | 57 | **37** | 79 | 78 | 86 | 82 | **37** | 67 |
| Intervalle | 11 | **12** | 13 | 14 | 15 | 16 | **17** | 18 | 19 | 20 |
| Ordres $(\hat{p}_j, \hat{q}_j)$ | (1,0) | **(2,0)** | (1,1) | (1,1) | (1,1) | (1,1) | **(4,3)** | (0,1) | (0,1) | (0,1) |
| Fréquence | 66 | **59** | 66 | 68 | 66 | 66 | **9** | 70 | 67 | 71 |

TAB. 2 – Ordres sélectionnés dans l'étape 1 pour chaque intervalle.

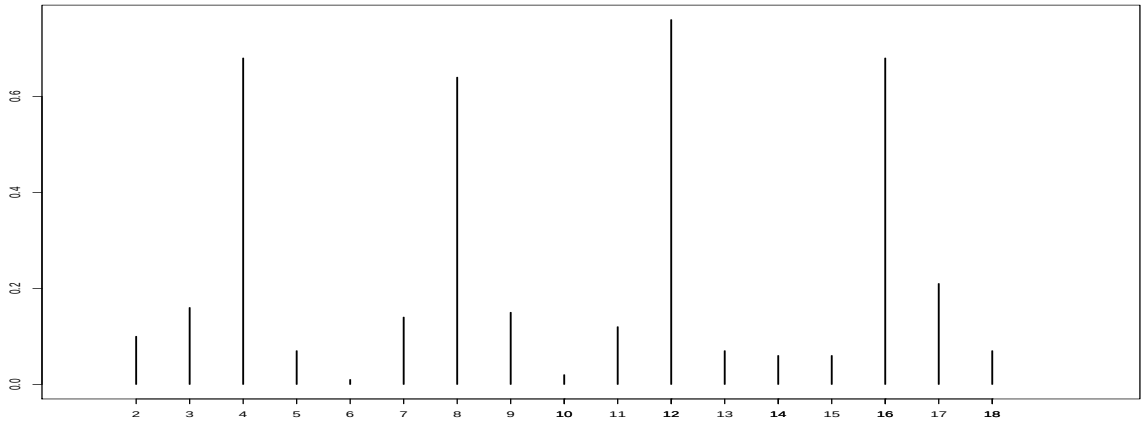


FIG. 1 – Intervalles sélectionnés dans l'étape 2.

| $\lambda_j$ | 0.1957 | 0.4085 | 0.5906 | 0.8044 |
|---|---|---|---|---|
| $\hat{\mu}(\hat{\lambda}_j)$ | 0.1946 | 0.4099 | 0.5959 | 0.8095 |
| $\hat{\sigma}(\hat{\lambda}_j)$ | 0.038 | 0.0146 | 0.0295 | 0.0209 |

TAB. 3 – Estimations des PR dans l'étape 3.

Le tableau 4 contient les ordres les plus fréquemment sélectionnés ainsi que les moyennes et les erreurs standards des paramètres estimés dans l'étape 4 sur les 100 expériences. Nous omettons les paramètres $\hat{\phi}_{j,k}$ et $\hat{\theta}_{j,k}$ pour $k > 2$ car ceux-ci sont très proches de zéro. On voit que les valeurs estimées des paramètres sont proches des vraies valeurs données dans le tableau 1.

| Estimée | Série $X_{t,j}$ | | | | |
|---|---|---|---|---|---|
| $\hat{\beta}_j$ | 1 | 2 | 3 | 4 | 5 |
| $(\hat{p}_j,\ \hat{q}_j)$ | (1,2) | (0,0) | (1,0) | (1,1) | (0,1) |
| Fréquence | 59 | 91 | 68 | 77 | 76 |
| $\hat{\mu}(\hat{d}_j)$ | 0.22 | 0.44 | 0.14 | 0.36 | 0.10 |
| $\hat{\sigma}(\hat{d}_j)$ | 0.02 | 0.07 | 0.02 | 0.05 | 0.07 |
| $(\hat{\mu}(\hat{\phi}_{j,1}), \hat{\mu}(\hat{\phi}_{j,2}))$ | (0.56,-0.05) | (0.01,0.00) | (-0.65,0.05) | (0.32,0.00) | (-0.05,-0.02) |
| $(\hat{\sigma}(\hat{\phi}_{j,1}), \hat{\sigma}(\hat{\phi}_{j,2}))$ | (0.07,0.02) | (0.01,0.01) | (0.14,0.02) | (0.02,0.07) | (0.03,0.02) |
| $(\hat{\mu}(\hat{\theta}_{j,1}), \hat{\mu}(\hat{\theta}_{j,2}))$ | (-0.57,-0.24) | (-0.01,0.00) | (0.02,0.00) | (-0.69,-0.01) | (0.33,0.00) |
| $(\hat{\sigma}(\hat{\theta}_{j,1}), \hat{\sigma}(\hat{\theta}_{j,2}))$ | (0.13,0.10) | (0.05,0.02) | (0.03,0.01) | (0.14,0.01) | (0.07,0.01) |

TAB. 4 – Ordres et paramètres estimés dans l'étape 4.

En conclusion, quand chaque sous-série stationnaire contient suffisamment de données, les performances pratiques de notre méthode d'estimation pour le modèle (1)–(2) sont bonnes, non seulement pour estimer le nombre et les emplacements des PR mais aussi les paramètres des sous-séries.

# Références


J. Beran. *Statistics for long-memory processes*, volume 61 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York, 1994.

P. Doukhan, G. Oppenheim, and M. S. Taqqu, editors. *Theory and Applications of Long-Range Dependence*. Birkhäuser Boston Inc., Boston, MA, 2003.

B. K. Ray and R. S. Tsay. Bayesian methods for change-point detection in long-range dependent processes. *Journal of Time Series Analysis*, 23(6), 687–705, 2002.

S. Stoev, M. S. Taqqu, C. Park, G. Michailidis, and J. S. Marron. LASS : a tool for the local analysis of self-similarity. *Comput. Statist. Data Anal.*, 50(9), 2447–2471, 2006.

L. Wang and J. Wang. Testing and estimating for change in long memory parameter. *J. Stat. Comput. Simulation*, 76(4), 317–329, 2006.