Structural Break Estimation for piecewise FARIMA Models

Li Song[1,2], Pascal Bondon[1]

[1]CNRS UMR 8506, Université Paris-Sud
3 rue Joliot-Curie, 91192 Gif-sur-Yvette Cedex, France

[2]School of Electronic Information, Wuhan University,
Wuhan 430079, Hubei, China

song@lss.supelec.fr, bondon@lss.supelec.fr

Résumé − Nous étudions le problème de l'estimation de ruptures structurelles dans un signal aléatoire à longue mémoire du type FARIMA(p, d, q). Le nombre de ruptures et leurs localisations sont inconnus ainsi que les ordres (p, d, q) et les paramètres de chaque sous-série. Nous proposons une procédure basée sur des estimateurs du maximum de vraisemblance et un critère d'information permettant d'estimer les points de rupture et les paramètres des sous-séries à longue dépendance. On montre par des simulations de Monte Carlo que la méthode fonctionne bien sur des modèles FARIMA par morceaux quand chaque sous-série contient un nombre raisonnable de données de l'ordre de 800.

Abstract − We consider the problem of estimating the structural changes in a long memory piecewise FARIMA processes. The number of break points as well as their locations, the orders and the parameters of each sub-segments are assumed to be unknown. A 4-step procedure based on maximum likelihood estimate and distance criteria is proposed to find out the break points and to estimate the parameters of each segment. Its effectiveness is shown by Monte Carlo simulations even when the length of piecewise stationary sub-segment is above 800.

1 Introduction

An increasing number of studies have been witnessed in two issues. On the one hand, issue related to estimation, testing and computation for piecewise models involving structural changes receives increasing interest in literature, e.g Bai (1997), Andrews (2003) and Bai and Perron (2006). On the other hand, issue related to stationary processes exhibiting long range dependence (LRD) has been observed in many areas, including hydrology, meteorology, economics and finance, and telecommunications; see, for example, Beran (1994), and Park and Willinger (2000). Commonly used model for such LRD processes is the fractional autoregressive integrated moving-average (FARIMA) model, introduced by Granger and Joyeux (1980) and Hosking (1981). The main feature of the stationary FARIMA(p, d, q) process is that its autocovariance function declines at a hyperbolic rate, slower than geometric rate of stationary ARMA(p, q) processes.

But less attention has been paid to the problem of modeling a process in the presence of both structural changes and LRD because they are easy to confused. The possibility of confusing long memory and structural change has of course arisen occasionally, in literatures of mathematical statistics like Bhattacharya et al. (1983) and Künsch (1986). Kuan and Hsu (1998) points out that even the structural test statistics of Hidalgo and Robinson (1996) designs specifically for the famous FARIMA data with known potential structural change date may still have large size distortions in small samples. Nunes et al. (1995) shows that the stationary LRD processes is likely to incur a spurious structural change, because many well-known structural change tests may suggest a structural change has occurred, even though there is no such change at all in the LRD processes. So people prefer to study the partial structural change model in which only some coefficients are allowed to change. For example, Ray and Tsay (2002) assumes that the ARMA parameters are constant. This partial structural change model can not meet the needs both in theory and in practice. In theory, accurate estimation of a FARIMA(p, d, q) model requires a large sample of data, which in turn increases the chance of structural changes over time. In practice, as said in Stoev et al. (2006) and Song et al. (2008), the assumption that real data can be modeled by a stationary process with constant parameters may be unrealistic.

Therefore, a piecewise stationary model with more flexibility in modeling structural changes and long memory, as well as the corresponding estimation procedure seems to be needed. This paper intends to fill this gap. The remainder of the paper is organized as follows. Section 2 outlines the piecewise stationary FARIMA model description, Section 3, a 4-step fitting procedure is described. Section 4 applies the procedure to simulated data. Section 5 concludes.

2 Model description

Broadly speaking, the original data in the piecewise model may have several stationary sub-segments in where models' coefficients keep unchanged. A node between the two sub-segments is the break point (BP) where the structure changes. We suppose that the non-stationary process Y_t, $t = 1, \ldots, n$, can be segmented into $m + 1$ blocks of stationary FARIMA processes. For $j = 1, \ldots, m$, denote the BP between the jth and $(j + 1)$th FARIMA processes as τ_j, and set $\tau_0 = 1$ and $\tau_{m+1} = n + 1$. Then the jth block of $\{Y_t\}$ is modeled by

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

where $X_{t,j}$ is the FARIMA (p_j, d_j, q_j) process defined like (2)

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

B is the backward operator $BX_t = X_{t-1}$, ε_t is a sequence of zero-mean iid random variables with finite variance σ_ε^2, $d_j \in (0, 1/2)$, and the polynomials $\phi_j(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ and $\psi(z) = 1 + \psi_1 z + \cdots + \psi_q z^q$ with real coefficients have no common zeros and neither $\phi(z)$ nor $\psi(z)$ has zeros in the closed unit disk $\{z \in \mathbb{C} : |z| \le 1\}$. The process $(1 - B)^{-d}\epsilon_t$ is defined by

$$(1 - B)^{-d}\epsilon_t = \sum_{k=0}^{\infty} \varphi_k(d)\epsilon_{t-k}$$

where $\varphi_0(d) = 1$ and $\varphi_k(d) = \prod_{s=1}^{k} \frac{d+s-1}{s}$ for $k \ge 1$. Since $d < 1/2$, $\sum_{k=0}^{\infty} \varphi_k(d)^2 < \infty$ and the series in $(1 - B)^{-d}\epsilon_t$ converges in the mean square sense. Y_t is characterized by slowly decaying non-summable autocovariances

$$\gamma(k) = \mathrm{Cov}(X_t, X_{t+k}) \sim c_\gamma |k|^{2d-1}.(|k| \to \infty)$$

Let $\beta_j = (p_j; q_j; \alpha_j)$ the parameters of the jth model, where P and Q are the highest order of all (p_j, q_j), $\alpha_j = (d_j; \phi_{j,1}, \ldots, \phi_{j,p_j}, \ldots, \phi_{j,P}; \theta_{j,1}, \ldots, \theta_{j,q_j} \ldots, \theta_{j,Q})$, $\phi_{j,g} = 0$ when $g > p_j$, and $\theta_{j,g} = 0$ when $g > q_j$. β_j is constant on each interval $[\tau_{j-1}, \tau_j)$.

3 Estimation procedure

The problem of fitting model (1)–(2) to data consists in finding $(\tau_1, \ldots, \tau_m, \beta_1, \ldots, \beta_{m+1})$. The first problem is to estimate the BPs accurately, which can be realized by detecting the structural parameters changes. It is shown in Stoev et al. (2006) that some of the best available techniques to estimate the parameters may be misled by non-stationary characters of the observed time series, and some of these non-stationarity effects can often be alleviated by estimating the parameters using data locally. That is to say, it is better to divide the original time series into a set of elementary sub-series of length E and use the data in the same sub-series to get a local parameter estimation. Predefining a suitable length E for the elementary sub-series is not always an easy task: on the one hand, due to LRD properties, a reasonable number of observations are needed to obtain adequate parameter estimates, and then E can't be too short; on the other, the probability of meeting a BP increases as E grows. E should vary with the hidden true model order and the estimate method. Hence some restriction should be put on E, and E is chosen by empirical experience.

In the following, we consider the truncated series formed by the $K = [n/E]$ elementary sub-series defined on the intervals $I_k = ((k - 1)E, kE]$ for $k = 1, \ldots, K$. We make the following assumptions:

(i) the number m of BPs of the truncated series is known and $m \le K/2$ (see Remark 1 when m is unknown);

(ii) there is at most one BP in each interval;

(iii) at least one interval separates two consecutive BPs.

Then, the m BPs τ_j, $j = 1, \ldots, m$, are dispersed into a few elementary intervals.

The following four steps procedure is proposed to fitting model (1)–(2) to a local stationary time series.

Step 1 : Local estimation. For each interval I_k, a pair $(\hat{p}_k, \hat{q}_k)$ is selected by employing the Bayes Information Criterion (BIC) as suggested in Torre et al. (2007), and the model's parameters α_k are estimated by the Gaussian maximum-likelihood estimates (MLE) based on autoregressive approximations $\hat{\alpha}_k$, see e.g. Granger and Joyeux (1980). Therefore, Step 1 gives the local estimates $\hat{\beta}_k = (\hat{p}_k, \hat{q}_k, \hat{\alpha}_k)$ for $k = 1, \ldots, K$.

Step 2 : Selection of intervals with a BP. If model (1)–(2) is suitable for the data, one expects that $\hat{\beta}_k$ is close to the true values of the parameters when there is no BP in the interval I_k. Now, if there is a BP in I_k and no BP in I_{k-1} and I_{k+1}, $\hat{\beta}_k$ should be significantly different from both $\hat{\beta}_{k-1}$ and $\hat{\beta}_{k+1}$. Then, let $k_0 = 0$, $k_{m+1} = K$, and

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

where $\bar{\alpha}_j = \dfrac{1}{k_j - k_{j-1}} \displaystyle\sum_{k=k_{j-1}+1}^{k_j} \hat{\alpha}_k$, $\bar{p}_j$ (resp. $\bar{q}_j$) is the order which is the most frequently selected among the orders $\hat{p}_k$ (resp. $\hat{q}_k$) for $k = k_{j-1} + 1, \ldots, k_j$. In the case where $\bar{p}_j$ (resp. $\bar{q}_j$) is not unique, the lowest order is chosen. Function $\psi(\cdot)$ is positive and strictly increasing. Let $J_k = ((k - 0.5)E, (k + 0.5)E]$ for $k = 1, \ldots, K - 1$. We select the intervals $(J_{\hat{k}_1}, \ldots, J_{\hat{k}_m})$ as being those containing a BP.

Step 3 : Estimation of the BPs. Suppose that all the intervals $J_{\hat{k}_j}$ are selected properly, i.e., $\tau_j \in J_{\hat{k}_j}$. Therefore, for any fixed j, there is no BP in the "previous"

block between $J_{\hat{k}_{j-1}}$ and $J_{\hat{k}_j}$, viz. $((\hat{k}_{j-1}+0.5)E, (\hat{k}_j - 0.5)E]$ where we set $\hat{k}_0 + 0.5 = 0$, and we define $\hat{\beta}_p$ as the MLE of β_j based on the data in this block. In the same way, let $\hat{\beta}_n$ be the MLE of β_{j+1} based on the data in the "next" block between $J_{\hat{k}_j}$ and $J_{\hat{k}_{j+1}}$, viz. $((\hat{k}_j + 0.5)E, (\hat{k}_{j+1} - 0.5)E]$ where we set $\hat{k}_{m+1} - 0.5 = K$. We treat $\hat{\beta}_p$ and $\hat{\beta}_n$ as two benchmarks. These estimates are more precise than any local estimate calculated in Step 1 since they involve more data. Suppose that $l \in J_{\hat{k}_j}$ is the BP τ_j. Then we can calculate the MLE $\hat{\beta}_{l_p}$ of β_j and $\hat{\beta}_{l_n}$ of β_{j+1} based respectively on $((\hat{k}_{j-1}+0.5)E, l]$ and $(l, (\hat{k}_{j+1} - 0.5)E]$. These estimates should be close to benchmarks $\hat{\beta}_p$ and $\hat{\beta}_n$, respectively. Hence, our choice of the BP estimate $\hat{\tau}_j$ is based on the following criterion

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

To reduce the complexity, $\hat{\beta}_{l_p}$ and $\hat{\beta}_{l_n}$ are calculated using the data in $(l-E, l)$ and $(l, l+E)$, respectively, and this gives good results in practice as shown in Section 4.

Step 4 : Estimation of the parameters of each stationary block. Once $(\hat{\tau}_1, \ldots, \hat{\tau}_m)$ are obtained, the parameters β_j of the stationary sequence $X_{t,j}$ for $j = 1, \ldots, m+1$, can be estimated on the basis on the data in $(\hat{\tau}_{j-1}, \hat{\tau}_j]$, where $\hat{\tau}_0 = 1$ and $\hat{\tau}_{m+1} = KE$.

Remarque 1. The procedure assumes that the number m of BPs is known, but of course m is unknown in practice. One way to estimate m consists in increasing sequentially one by one the number of BPs in the procedure. Indeed, when estimating a single BP model in the presence of multiple BPs, the estimate of the interval which contains the BP will be typically one of the true intervals with a BP, namely the one which is dominant in the sense that selecting this interval allows to minimize the sum of squared (3) where $m = 1$. Next, we minimize (3) with $m = 2$ which gives two dominating intervals with a BP. When iterating this process beyond the true number of BPs, two intervals are founded which are very close to each other and correspond to the same BP. This allows to determine the true number of BPs when the BPs are not too close to each others. We show in Section 4 that this method for finding m works well when at least $2E$ data separate each BP.

4 Simulation

We illustrate the estimation procedure and show its effectiveness by a simple Monte Carlo simulation. In this example, the target piecewise time series FARIMA(p, d, q) process $\{Y_t\}$, $t = 1, \ldots, n$ is generated by

$$Y_t = \begin{cases} 0.3Y_{t-1} + (1 - 0.7B)(1-B)^{-0.2}\epsilon_t, & \text{if } 1 \le t < \tau_1, \\ (1-B)^{-0.4}\epsilon_t, & \text{if } \tau_1 \le t < \tau_2, \\ (1 + 0.4B)(1-B)^{-0.3}\sigma_3\epsilon_t, & \text{if } \tau_2 \le t \le n, \end{cases} \quad (5)$$

where $n = 20000$ and $\epsilon_t \sim$ iid N$(0,1)$. Two BPs are at $\tau_1 = 8200$ and $\tau_2 = 13800$. We set $E = 2000$, therefore $K = 10$ and two BPs fall within the elementary interval I_5 and I_7; the intervals with a BP respectively are J_4 and J_7. All results are based on 100 realizations of process (5).

Table 1 exhibits the most frequently selected orders for each elementary intervals in Step 1. For most I_k's, these orders are the true ones. Observe that BIC performs well for detecting the low orders zero and one. Of course, when there is a BP and the order changes in the elementary intervals I_5 and I_7, the order given by BIC is less reliable (see the columns in bold in table 1).

Sub-series	1	2	3	4	**5**
Order $(\hat{p}_k, \hat{q}_k)$	(1,1)	(1,1)	(1,1)	(1,1)	**(0,0)**
Frequency	79	80	77	79	**45**
Sub-series	6	**7**	8	9	10
Order $(\hat{p}_k, \hat{q}_k)$	(0,0)	**(0,0)**	(0,1)	(0,1)	(0,1)
Frequency	93	**69**	83	85	88

TAB. 1: Selected orders in Step 1.

After the local estimation (Step 1), we calculate the intervals with a BP (Step 2) corresponding to the BPs numbers $m = 1, \ldots, 4$. For $m = 1, 2$, all the selected intervals in 100 realizations are well separated. For $m = 3, 4$, we find intervals close to each others in the same 100 realizations, which hints that the number of BPs should be 2. Table 2 gives the selected intervals in the 100 simulations for $m = 2, 3$ (the last column for each case indicates the number of frequencies the $(\hat{k}_1, \hat{k}_2)$ or $(\hat{k}_1, \hat{k}_2, \hat{k}_3)$ are selected). We see that when the BPs number used in Step 2 is the true BPs number plus one, the additional interval chosen by the procedure is close to an interval containing a BP. When $m = 2$, some selected intervals spread to a few intervals near the right ones. The estimation of the first interval diffuses from J_2 to J_5, that of the second interval from J_5 to J_8. This is partly caused by the impreciseness of the order selection in Step 1.

$m=2$			$m=3$			
$\hat{k}_1$	$\hat{k}_2$	%	$\hat{k}_1$	$\hat{k}_2$	$\hat{k}_3$	%
4	7	**71**	4	7	8	**38**
4	8	14	4	5	7	23
3	7	5	3	4	7	18
2	5	5	4	6	7	12
2	7	3	5	7	8	3
5	8	2	Others cases			6

TAB. 2: Selected intervals in Step 2 for $m = 2, 3$.

Table 3 presents the sample means $\hat{\mu}(\hat{\lambda}_j)$ and standard deviations $\hat{\sigma}(\hat{\lambda}_j)$ of the BPs estimation in Step 3 over the 100 realizations. Following Davis et al. (2006), we use the standardized parameter $\lambda_j = \tau_j/n$. We see that the estimation are close to the true values and the standard deviations are quite small.

λ_j	0.410	0.690
$\hat{\mu}(\hat{\lambda}_j)$	0.390	0.690
$\hat{\sigma}(\hat{\lambda}_j)$	0.005	0.006

TAB. 3: Locations of BPs estimation in Step 3.

Table 4 gives the most frequently selected orders and the corresponding sample means and standard deviations of the parameters estimates in Step 4 for each stationary segment identified in Step 3. We see that the true orders are well identified and the estimated parameters are quite near to the true values given in (5).

Estimate	Segment $X_{t,j}$		
$\hat{\beta}_j$	1	2	3
$(\hat{p}_j, \hat{q}_j)$	(1,1)	(0,0)	(0,1)
Frequency	90	92	96
$\hat{\mu}(\hat{d}_j)$	0.20	0.39	0.30
$\hat{\sigma}(\hat{d}_j)$	0.04	0.05	0.04
$\hat{\mu}(\hat{\phi}_j)$	0.31	-	-
$\hat{\sigma}(\hat{\phi}_j)$	0.01	-	-
$\hat{\mu}(\hat{\theta}_j)$	-0.69	-	0.40
$\hat{\sigma}(\hat{\theta}_j)$	0.02	-	0.04

TAB. 4: Orders and parameters estimation in Step 4.

5 Conclusion

In this paper, we have proposed a piecewise FARIMA model and the methodology for fitting it to a piecewise stationary long memory signal. This model is able to capture the structural change properties of the signal, it is flexible and allows to model simultaneously long and short range dependence. The model fitting consists in a 4-step procedure designed to estimate both the BPs and the parameters. Simulations have shown that the practical performance of our method for the above piecewise FARIMA process is very good.

References

D. Andrews. "Tests for parameter instability and structural change with unknown change point". *Econometrica*, 71(1):395–397, 2003.

J. Bai. Estimating multiple breaks one at a time. *Econometric Theory*, 13(3):315–352, 1997.

J. Bai and P. Perron. Multiple structural change models: a simulation analysis. In *Econometric theory and practice*, pages 212–237. Cambridge Univ. Press, Cambridge, 2006.

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

R. N. Bhattacharya, V. K. Gupta, and E. Waymire. The Hurst effect under trends. *J. Appl. Probab.*, 20(3):649–662, 1983.

R. A. Davis, Thomas C. M. Lee, and G. A. Rodriguez-Yam. Structural break estimation for nonstationary time series models. *J. Amer. Statist. Assoc.*, 101(473): 223–239, 2006.

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

J. Hidalgo and P. M. Robinson. Testing for structural change in a long-memory environment. *J. Econometrics*, 70(1):159–174, 1996.

R. J. Hosking. Fractional differencing. *Biometrika*, 68(1): 165–176, 1981.

C. M. Kuan and C. C. Hsu. Change-Point Estimation of Fractionally Integrated Processes. *Journal of Time Series Analysis*, 19(6):693–708, 1998.

H. Künsch. Discrimination between monotonic trends and long-range dependence. *J. Appl. Probab.*, 23(4):1025–1030, 1986.

L. C. Nunes, C. M. Kuan, and P. Newbold. Spurious break. *Econometric Theory*, 11(4):736–749, 1995.

K. Park and W. Willinger. *Self-Similar Network Traffic and Performance Evaluation*. John Wiley & Sons, Inc., New York, NY, USA, 2000.

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.

L. Song, P. Bondon, Y. Cao, and Q. Cheng. A Time-Varying FARIMA Model for Internet Traffic. In *Image and Signal Processing, 2008. CISP'08*, volume 5, pages 83–87, 2008.

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.

K. Torre, D. Delignières, and L. Lemoine. Detection of long-range dependence and estimation of fractal exponents through ARFIMA modelling. *British Journal of Mathematical and Statistical Psychology*, 60(1):85–106, 2007.