Piecewise model selection for non-stationary long memory data

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]Laboratoire MAS, Ecole Centrale de Paris
Grande Voie des Vignes, 92295 Chatenay-Malabry, France

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

Résumé – Nous étudions le problème de la sélection d'un modèle paramétrique localement stationnaire pour un signal aléatoire à longue mémoire. Le nombre de ruptures et leurs localisations sont inconnus ainsi que les paramètres de chaque sous-série. Nous proposons un critère basé sur le principe de description de longueur minimale ("Minimum Description Length") pour sélectionner le meilleur modèle. L'optimisation de ce critère est réalisée au moyen d'un algorithme génétique. Des simulations de Monte Carlo montrent que ce critère est plus performant que le critère BIC et le critère proposé par Davis et al. (2006). L'application de la méthode aux données réelles du Nil entre 622 et 1284 av. J-C confirme d'autres études concluant à un changement structurel des données autour de 722 av. J-C.

Abstract – We study the model selection problem for a locally stationary long memory signal by dividing the signal into stationary blocks. In this piecewise model, the number and the locations of the break points are unknown as well as the parameters of each regime. We propose a model selection criterion based on the minimum description length (MDL) principle. Monte Carlo simulations show that our criterion performs better than BIC and the criterion proposed by Davis et al. (2006). The application of our method to the Nile river data for the years 622-1284 AD confirms previous studies which conclude that a structural break exists around the year 722 AD.

1 Introduction

Fitting a piecewise stationary model to data consists in identifying the different blocks and in selecting and estimating an appropriate model for each stationary piece. This problem is addressed by Kitagawa and Akaike (1978) and Davis et al. (2006) when the pieces are autoregressive processes, and by Davis et al. (2008) when the blocks are a type of nonlinear time series including piecewise generalized autoregressive conditionally heteroscedastic processes. Here, we are interested in long range dependent (LRD) processes. These type of series appear in many areas, including hydrology, meteorology, economics, finance and telecommunications; see for instance, Beran (1994) and Taqqu and Teverovsky (1997). A commonly used model for LRD processes is the FARIMA model, introduced by Granger and Joyeux (1980) and Hosking (1981). The main feature of a stationary FARIMA process is that its covariance function decays hyperbolically, while the covariance function of an ARMA process at least decays exponentially.

In practice, estimating a LRD model accurately requires more data than estimating a short-memory model, which in turn, increases the chance of structural changes over time. Much of real data exhibit both structural changes and LRD; see e.g. Beran and Terrin (1994). Then, it may be unrealistic to assume that the data can be modeled by a stationary process with constant parameters. Previous studies discussing structural changes in LRD processes include Gil-Alana (2008) who considers a linear regression with a fractional noise disturbance where the sub-series have different regression coefficients and fractional orders. Ray and Tsay (2002) use a Bayesian method for detecting the changes in the mean and the LRD parameter of a FARIMA process with a fixed ARMA part.

This work proposes a piecewise FARIMA process to model a local stationary long-memory time series. It is a pure structural change model in the sense that all parameters including the ARMA orders are allowed to change between two regimes. Moreover, the number of structural BPs is assumed to be unknown. Fitting data to this model can be treated as a statistical model selection problem which can be solved by the minimum description length (MDL) principle by Rissanen (1978). MDL principle is used by Davis et al. (2006) for a piecewise AR process, and by Davis et al. (2008) for some nonlinear piecewise stationary time series. Empirical results show good performance results for estimating the BP number and their locations for these models. Here, we adapt the MDL principle to the piecewise FARIMA process. The implementation of this principle leads to a criterion which performs well in practice.

The rest of this article is organized as follows. In Section 2, we introduce the piecewise stationary FARIMA model and in Section 3, we present the criterion based on the MDL principle. In Section 4, Monte Carlo simulation results are presented and in Section 5, the yearly minima of the Nile river data is considered. Finally, concluding remarks are given in Section 6.

2 Piecewise FARIMA model

We address the multiple structural change problem for a non-stationary time series in which the segments are modelled by stationary zero-mean FARIMA processes. More precisely, let m denote the unknown BP number and n the length of the time series. For $j = 1, \dots, m$, let τ_j be the BP between the jth and $(j+1)$th FARIMA regime, and set $\tau_0 = 1$ and $\tau_{m+1} = n + 1$. For $j = 1, \dots, m + 1$, the jth piece of the observed time series $\{Y_t\}$ is modeled by

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

where $\{X_{t,j}\}$, $t \in \mathbb{Z}$, is the FARIMA(p_j, d_j, q_j) process defined by the difference equation

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

$\{\epsilon_{t,j}\}$, $t \in \mathbb{Z}$, $j = 1, \dots, m+1$, is a sequence of iid zero-mean Gaussian random variables with unit variance, $\sigma_j > 0$, B is the backward operator $BX_t = X_{t-1}$, $d_j \in (0, 1/2)$, and the polynomials $\Phi_j(z) = 1 - \phi_{j,1}z - \cdots - \phi_{j,p_j}z^{p_j}$ and $\Theta_j(z) = 1 + \theta_{j,1}z + \cdots + \theta_{j,q_j}z^{q_j}$ with real coefficients have no common zeros and neither $\Phi_j(z)$ nor $\Theta_j(z)$ has zeros in the closed unit disk $\{z \in \mathbb{C} : |z| \le 1\}$. The process $(1-B)^{-d_j}\epsilon_{t,j}$ is defined by

$$(1-B)^{-d_j}\epsilon_{t,j} = \sum_{k=0}^{\infty} \varphi_k(d_j)\epsilon_{t-k,j}, \tag{3}$$

where $\varphi_0(d_j) = 1$ and $\varphi_k(d_j) = \prod_{s=1}^{k} \frac{d_j+s-1}{s}$ for $k \ge 1$. Since $d_j < 1/2$, $\sum_{k=0}^{\infty} \varphi_k(d_j)^2 < \infty$.

The parameters of the jth regime are $\alpha_j = (d_j, \phi_{j,1}, \dots, \phi_{j,p_j}, \theta_{j,1}, \dots, \theta_{j,q_j}, \sigma_j)$ and α_j is constant for each interval $[\tau_{j-1}, \tau_j)$. The piecewise FARIMA process $\{Y_t\}$ is characterized by the BP number m, the BP locations $\tau_1, \dots, \tau_m$ and the parameters $\alpha_1, \dots, \alpha_{m+1}$.

3 Model selection using MDL

Fitting model (1)–(2) to the data $y = (y_i)_{1 \le i \le n}$ consists in finding the "best" vector $\gamma = (m, \tau_1, \dots, \tau_m, \alpha_1, \dots, \alpha_{m+1})$. This can be treated as a statistical model selection problem in which candidate models may have different number of parameters. One efficient strategy to solve this problem is to use the MDL principle. By viewing statistical modeling as a way of generating descriptions of observed data, the central idea of the MDL principle is to represent an entire class of candidate probability distributions as models, and to select the model which allows the shortest coding of the data and of the model itself.

We adopt the two-part description length method used by Rissanen; see e.g. Lee (2001). Let $L(\cdot)$ denote the code length of an object. Then using model (1)–(2) to encode y, $L(y)$ can be decomposed into

$$L(y) = L(\widehat{\gamma}) + L(y|\widehat{\gamma}),$$

where $\widehat{\gamma}$ is vector γ in which parameters $\alpha_1, \dots, \alpha_{m+1}$ are replaced by the maximum likelihood estimates (MLEs) $\widehat{\alpha}_1, \dots, \widehat{\alpha}_{m+1}$ and $L(y|\widehat{\gamma})$ is the code length for encoding y with model (1)–(2) defined by $\widehat{\gamma}$. The "best" model is the one minimizing $L(y)$.

Let us first derive an expression for $L(\widehat{\gamma})$. Let $n_j = \tau_j - \tau_{j-1}$ be the number of observations in the jth FARIMA regime. Since the τ_j's contain the same information as the n_j's, we have

$$L(\widehat{\gamma}) = L(m) + \sum_{j=1}^{m+1} \{L(n_j) + L(p_j) + L(q_j) + L(\widehat{\alpha}_j)\}. \tag{4}$$

According to Rissanen (1983), for any nonnegative integer x, we have

$$L(x) = \begin{cases} \log_2 c + \log_2 x + \log_2 \log_2 x + \cdots & \text{if } x > 0, \\ 0 & \text{if } x = 0, \end{cases} \tag{5}$$

where c is a constant approximately equal to 2.865 and the sum involves only the nonnegative terms, whose number is clearly finite. To determine $L(\widehat{\alpha}_j)$, we use the following result of Rissanen (1989) : a MLE of a real-valued parameter computed from N data can be effectively encoded with $\frac{1}{2}\log_2 N$ bits. Each of the $p_j + q_j + 2$ parameters in $\widehat{\alpha}_j$ is computed from n_j data. Therefore, we have $L(\widehat{\alpha}_j) = \frac{p_j+q_j+2}{2}\log_2 n_j$.

According to Rissanen (1989), $L(y|\widehat{\gamma})$ is the negative of the $\log_2$-likelihood function at the MLEs $\widehat{\alpha}_1, \dots, \widehat{\alpha}_{m+1}$. Since the segments in model (1)–(2) are independent and Gaussian, we have

$$L(y|\widehat{\gamma}) = \sum_{j=1}^{m+1} \mathcal{L}_j(y_j; \widehat{\alpha}_j), \tag{6}$$

where

$$\mathcal{L}_j(y_j; \widehat{\alpha}_j) = \frac{n_j}{2}\log_2(2\pi) + \frac{1}{2}\log_2(\det \widehat{V}_j) + \frac{\log_2 e}{2}\, \underline{y}_j' \widehat{V}_j^{-1} \underline{y}_j,$$

$\widehat{V}_j$ is the covariance matrix with size n_j of the FARIMA process $\{X_{t,j}\}$ in (2) where the vector of parameters α_j is replaced by $\widehat{\alpha}_j$, and $\underline{y}_j = (y_{\tau_{j-1}}, \dots, y_{\tau_j - 1})'$ is the vector of observations in the jth piece in (1). Combining (4) and

(6), we propose to select the best model (1)–(2) for y as the one that minimizes with respect to $(m, \tau_1, \ldots, \tau_m, p_1, \ldots, p_{m+1}, q_1, \ldots, q_{m+1})$ criterion C defined by

$$\mathrm{C} = L(m) + \sum_{j=1}^{m+1} \{L(n_j) + L(p_j) + L(q_j) + \frac{p_j + q_j + 2}{2} \log_2 n_j + \mathcal{L}_j(y_j; \widehat{\alpha}_j)\}, \quad (7)$$

where functions L and $\mathcal{L}_j$ are defined by (5) and (3), respectively.

Applying the criterion proposed by Davis et al. (2008) to model (1)–(2) amounts to minimize with respect to $(m, \tau_1, \ldots, \tau_m, p_1, \ldots, p_{m+1}, q_1, \ldots, q_{m+1})$ function D defined by

$$\mathrm{D} = \log_2^+ m + (m+1)\log_2 n + \sum_{j=1}^{m+1} \{\log_2^+ p_j + \log_2^+ q_j + \frac{p_j + q_j + 2}{2} \log_2 n_j + \mathcal{L}_j(y_j; \widehat{\alpha}_j)\}, \quad (8)$$

where for any nonnegative integer x, $\log_2^+ x = \log_2 x$ if $x \geq 1$ and $\log_2^+ 0 = 0$. On the other hand, optimizing the Bayesian information criterion for model (1)–(2) is equivalent to minimize with respect to $(m, \tau_1, \ldots, \tau_m, p_1, \ldots, p_{m+1}, q_1, \ldots, q_{m+1})$ function BIC defined by

$$\mathrm{BIC} = \sum_{j=1}^{m+1} \left\{ \frac{p_j + q_j + 2}{2} \log_2 n_j + \mathcal{L}_j(y_j; \widehat{\alpha}_j) \right\}. \quad (9)$$

The differences between criteria C, D and BIC lie in the penalty term. The expression of the code length of an integer is different in C and D. Observe that $L(0) = 0$ and $L(1) = \log_2 c$ in (7), while $\log_2^+ 0 = \log_2^+ 1 = 0$ in (8). Moreover, $L(x)$ is significantly different from $\log_2 x$ when x is not large, which is the case when x is the BP number m or the orders (p_j, q_j) of the FARIMA models. We compare the performances of these three criteria using a piecewise FARIMA model with multiple BPs in the simulation section.

4 Simulation

Since the search space is huge, the practical optimization of C, D or BIC is a complicated task and we use an automatic methodology based on a genetic algorithm first proposed by Holland (1975). In this simulation, we consider a piecewise FARIMA model of length $n = 4000$ with three BPs at $\tau_1 = 600$, $\tau_2 = 1600$ and $\tau_3 = 2400$. In (2), we take $\sigma_j = 1$ for $j = 1, \ldots, 4$; the parameters for four regimes are $(d_j, \phi_j, \theta_j) = (0.3, 0, 0)$; $(0.1, 0.6, -0.7)$; $(0.4, 0, -0.5)$; $(0.2, 0.8, 0)$. In the following, we use the standardized break fraction $\lambda_j = \tau_j / n$, and then the true break fractions are $\lambda_1 = 0.15$, $\lambda_2 = 0.40$, $\lambda_3 = 0.60$. All results are based on 1000 replications. We compare criteria C, D and BIC for the selection of the piecewise model (1)–(2). The mean values and standard errors of the estimated break fractions are given in table 1. We see that BIC overestimates the BP number in 33% of the cases while C and D underestimate it in 20% and 28% of the cases, respectively. When the right BP number is selected, all criteria estimate the break fractions with good precision, C performing the best, in terms of bias and mean-square error.

Criterion	m	Number	Mean and MSE (in parenthesis) of $\widehat{\lambda}_j$, $j = 1, \ldots, m$
C	2	167	0.1502, 0.4004 (3.089e-6),(2.343e-5)
	2	32	0.1503, 0.6005 (2.734e-6), (2.949e-5)
	3	801	0.1502, 0.4001, 0.6002 (2.202e-6), (1.232e-5), (3.062e-5)
D	2	133	0.1483, 0.4024 (2.328e-5), (9.453e-5)
	2	146	0.1507, 0.5974 (1.201e-5), (1.001e-4)
	3	721	0.1509, 0.4026, 0.6021 (1.027e-5), (7.857e-5), (8.081e-5)
BIC	3	669	0.1533, 0.4040, 0.5929 (3.865e-5), (1.178e-4), (4.568e-4)
	4	331	0.1448, 0.3928, 0.6107, 0.8019 (5.261e-5), (1.034e-4), (7.765e-4), (–)

TAB. 1: Estimated BPs.

In table 2, we present the percentage of right orders selection for the four regimes when C, D and BIC detect three BPs. All criteria tend to overestimate the orders, but C behaves better than both D and BIC since it finds the true orders more often and selects smaller orders in the other cases (the detailed results are omitted here).

(p_j, q_j) / Criterion	(0,0)	(1,1)	(0,1)	(1,0)
C	65.66	67.79	83.27	85.02
D	58.67	67.26	78.09	80.03
BIC	50.21	58.51	68.10	69.11

TAB. 2: Percentage of right model orders (p_j, q_j) selection for $j = 1, \ldots, 4$.

5 The River Nile data case

The time series of yearly minimal water levels of the Nile river for the years 622–1284 AD ($n = 663$) is one of the prime examples of LRD processes. The data are displayed in figure 1. A FARIMA$(0, d, 0)$ process with $d = 0.40$ fits these data well, as shown by Beran (1992) and Beran (1994). However, Beran and Terrin (1996) reveal that the series might not be completely homogeneous. Observations 1 to about 100 seem to be more independent than the others, implying a smaller value of d for the first 100 data than for the subsequent data. Palma et al. (2008) also find that there are some possible highly influential observations around the year 720 AD. Whitcher et al. (2002) observe that the structure change around this year coincides closely to the construction of a new device in 715 AD for measuring Nile river water levels. Beran and Terrin (1996) fit a FARIMA$(0, d, 0)$ process with $d = 0.04$ for the years 622–722 AD, and a FARIMA$(0, d, 0)$ process with $d = 0.38$ for the years 723–1284 AD.

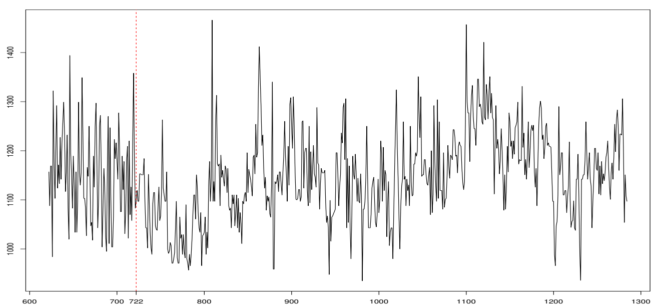

FIG. 1: Nile river data (622–1284 AD). The vertical dashed line indicates the estimated BP location by criterion C.

We apply criterion C to fit a piecewise FARIMA model to the Nile river data. Criterion C chooses a two-regimes model with one BP $\widehat{\tau}_1$ at 722 AD. For the first regime, a FARIMA$(2, d, 2)$ process with $d = 0.03$ is selected, while a FARIMA$(0, d, 0)$ process with $d = 0.45$ is chosen for the second regime. These results are very close to the results of Beran and Terrin (1996) and Whitcher et al. (2002). Since the first regime has only 100 data, estimation of orders (p_1, q_1) needs to be interpreted carefully.

6 Conclusions

In this article, we have proposed a method for modeling a non-stationary time series as a piecewise FARIMA process. The problem is in estimating the BP number and the locations, and in fitting an appropriate FARIMA model for each stationary regime. This is achieved by minimizing a MDL criterion. Numerical experiments have demonstrated good performance results of the proposed method. When applying our methodology to the Nile River data a two regime piecewise FARIMA model with a BP at 722 AD was selected.

References

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

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

J. Beran and N. Terrin. Estimation of the long-memory parameter, based on a multivariate central limit theorem. *J. Time Ser. Anal.*, 15(3):269–278, 1994.

J. Beran and N. Terrin. Testing for a change of the long-memory parameter. *Biometrika*, 83(3):627–638, 1996.

R. A. Davis, T. 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.

R. A. Davis, T. C. M. Lee, and G. A. Rodriguez-Yam. Break detection for a class of nonlinear time series models. *J. Time Ser. Anal.*, 29(5):834–867, 2008.

L. A. Gil-Alana. Fractional integration and structural breaks at unknown periods of time. *J. Time Ser. Anal.*, 29(1):163–185, 2008.

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. H. Holland. *Adaptation in natural and artificial systems.* University of Michigan Press, Ann Arbor, Mich., 1975. An introductory analysis with applications to biology, control, and artificial intelligence.

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

G. Kitagawa and H. Akaike. A procedure for the modeling of non-stationary time series. *Ann. Inst. Statist. Math.*, 30(1): 351–363, 1978.

T. C. M. Lee. An introduction to coding theory and the two-part minimum description length principle. *International Statistical Review*, 69(2):169–183, 2001.

W. Palma, P. Bondon, and J. Tapia. Assessing influence in gaussian long-memory models. *Comput. Statist. Data Anal.*, 52(9):4487–4501, 2008.

B. K. Ray and R. S. Tsay. Bayesian methods for change-point detection in long-range dependent processes. *J. Time Ser. Anal.*, 23(6):687–705, 2002.

J. Rissanen. Modeling by shortest data description. *Automatica*, 14(5):465–471, 1978.

J. Rissanen. A universal prior for integers and estimation by minimum description length. *Ann. Statist.*, 11(2):416–431, 1983.

J. Rissanen. *Stochastic complexity in statistical inquiry*, volume 15 of *World Scientific Series in Computer Science.* World Scientific Publishing Co. Inc., Teaneck, NJ, 1989.

M. S. Taqqu and V. Teverovsky. Robustness of Whittle-type estimators for time series with long-range dependence. *Comm. Statist. Stochastic Models*, 13(4):723–757, 1997.

B. Whitcher, S. D. Byers, P. Guttorp, and D. B. Percival. Testing for Homogeneity of Variance in Time Series: Long Memory, Wavelets and the Nile River. *Water Resources Research*, 38(5):1000–1029, 2002.