STRUCTURAL CHANGES ESTIMATION FOR STRONGLY-DEPENDENT PROCESSES

Li Song and Pascal Bondon

CNRS UMR 8506, Université Paris XI, France.

December 24, 2011

Abstract

In this paper, we consider the problem of estimating multiple structural breaks in a long-memory FARIMA time series. The number of break points as well as their locations, the orders and the parameters of each regime are assumed to be unknown. A selection criterion based on the minimum description length (MDL) principle is proposed and a genetic algorithm is implemented for its optimization. Monte Carlo simulation results show the effectiveness of this criterion and an application to the Nile River data is considered.

Keywords : Structural change; minimum description length principle; long-memory; non-stationary time series; genetic algorithm.

AMS 2010 subject classification : Primary 62M10, secondary 60G35.

1 Introduction

There is a lot of interest in the estimation, test and simulation of models involving structural breaks in the literature. The problem of multiple structural changes in linear regression models with independent and identically distributed (iid) errors has been extensively studied; see for instance Andrews (1993), Bai and Perron (1998), and the reviews by Bhattacharya (1994) and Perron (2006). The same problem with short or long-memory correlated errors is considered by Hidalgo and Robinson (1996), Lavielle and Moulines (2000), Lazarová (2005), Bai and Perron (2006) and Perron and Qu (2006), among others.

In this article, we consider the problem of modeling a non-stationary time series by segmenting the series into blocks of stationary processes. More precisely, the time series is represented by a parametric model in which the parameters change values at fixed unknown times called break points (BPs). The problem is in estimating the BP number and the locations as well as in selecting an appropriate model and estimating its parameters for each stationary regime. This problem is addressed by Kitagawa and Akaike (1978), Fearnhead (2005) 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 (GARCH) processes. Here, we are interested in time series which exhibit long range dependence (LRD). 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 fractional autoregressive integrated moving-average (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 long-memory model accurately requires more data than estimating a short-memory model, which in turn, increases the chance of structural changes over time. Then, it may be unrealistic to assume that the data can be modeled by a stationary process with constant parameters. Much of the real data exhibit both structural changes and LRD, for instance hydrological data (Beran and Terrin, 1994, 1996), and internet traffic data (Stoev et al., 2006; Wang and Wang, 2006). In particular, Beran and Terrin (1996) and Ray and Tsay (2002) find that the yearly minima of the Nile river series, measured at the Rhoda Gauge near Cairo in the years 622–1284 AD, has a change in the LRD behavior at around the year 722 AD. Previous studies discussing structural changes in LRD processes include Gil-Alana (2008) who considers a simple linear regression with a fractional noise disturbance where the sub-series have different intercepts, slope coefficients and fractional orders. The number of BPs is supposed to be known and the estimation of the BP locations and the regression coefficients is based on the least squares principle proposed by Bai and Perron (1998) using a grid of values for the 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. 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 for estimating not only the BP number and locations, but also the ARMA orders and coefficients. A genetic algorithm (GA) is implemented for the optimization of this criterion. The results are illustrated by Monte Carlo experiments and our methodology is applied to investigate a structural change in the River Nile data.

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, we develop an automatic procedure based on a GA to minimize the criterion and we discuss its implementation in detail. In Section 5, Monte Carlo simulation results are presented and in Section 6, the yearly minima of the Nile river data is considered. Finally, concluding remarks are given in Section 7.

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 $j = 1, \ldots, m$, 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 $j = 1, \ldots, m+1$, the jth piece of the observed time series $\{Y_t\}$ is modeled by

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

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

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

$\{\epsilon_{t,j}\}$, $t \in \mathbb{Z}$, $j = 1, \ldots, m+1$, 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 $\{z \in \mathbb{C} : |z| \leq 1\}$. 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 $k \geq 1$. Since $d_j < 1/2$, $\sum_{k=0}^{\infty} \varphi_k(d_j)^2 < \infty$. Consequently, as the sequence $\{\epsilon_{t,j}\}$, $t \in \mathbb{Z}$, is zero-mean and iid, the series in (3) converges in the mean square sense and almost surely.

The parameters of the jth regime are $\alpha_j = (d_j, \phi_{j,1}, \ldots, \phi_{j,p_j}, \theta_{j,1}, \ldots, \theta_{j,q_j}, \sigma_j)$ 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 $\tau_1, \ldots, \tau_m$ and the parameters $\alpha_1, \ldots, \alpha_{m+1}$.

3 Model selection using MDL

Fitting model (1)–(2) to the data $y = (y_1, \ldots, y_n)$ consists in finding the "best" vector $\gamma = (m, \tau_1, \ldots, \tau_m, \alpha_1, \ldots, \alpha_{m+1})$. 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; see e.g. Rissanen (1978), Rissanen (1989), Barron et al. (1998), Hansen and Yu (2001) and Grünwald (2007).

We adopt the two-part description length method used by Rissanen; see e.g. Lee (2001) and Hansen and Yu (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 $\alpha_1, \ldots, \alpha_{m+1}$ are replaced by the maximum likelihood estimates (MLEs) $\widehat{\alpha}_1, \ldots, \widehat{\alpha}_{m+1}$ 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

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

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 $\widehat{\alpha}_1, \ldots, \widehat{\alpha}_{m+1}$. 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

$$\mathcal{L}_j(y_j; \widehat{\alpha}_j) = \frac{n_j}{2}\log_2(2\pi) + \frac{1}{2}\log_2(\det \widehat{V}_j) + \frac{\log_2 e}{2}\, \underline{y}_j' \widehat{V}_j^{-1} \underline{y}_j, \tag{7}$$

$\widehat{V}_j$ 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 $\underline{y}_j = (y_{\tau_{j-1}}, \ldots, y_{\tau_j-1})'$ 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

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

where functions L and $\mathcal{L}_j$ are defined by (5) and (7), respectively.

Applying the criterion proposed by Davis et al. (2006) 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

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

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 (BIC) 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

$$\mathrm{BIC} = \sum_{j=1}^{m+1} \left\{ \frac{p_j+q_j+2}{2}\log_2 n_j + \mathcal{L}_j(y_j; \widehat{\alpha}_j) \right\}. \tag{10}$$

The arguments which minimize (8), (9) and (10), respectively, are denoted generically by $(\widehat{m}, \widehat{\tau}_1, \ldots, \widehat{\tau}_{\widehat{m}}, \widehat{p}_1, \ldots, \widehat{p}_{\widehat{m}+1}, \widehat{q}_1, \ldots, \widehat{q}_{\widehat{m}+1})$. 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 (8), while $\log_2^+ 0 = \log_2^+ 1 = 0$ in (9). 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 ARMA orders of a FARIMA model. In Section 5, we compare the performances of criteria C, D and BIC under different scenarios.

4 Implementation

Since the search space is huge, the practical optimization of C is a complicated task and we use an automatic methodology based on a GA. A GA is a random search technique inspired by Darwin's theory of evolution, first proposed by Holland (1975). Recent general references on GA include Reeves and Rowe (2002) and Sivanandam and Deepa (2008), and GAs are applied successfully in statistical applications by Gaetan (2000), Baragona et al. (2004), Davis et al. (2006), Chiogna et al. (2008) and Davis et al. (2008), among others. To minimize C, we implement a GA in the following way. We encode $(m, \tau_1, \ldots, \tau_m, p_1, \ldots, p_{m+1}, q_1, \ldots, q_{m+1})$ in a vector g called *chromosome* which contains all the information about the model. The components of g are called *genes*. An initial population of N chromosomes are generated randomly. Then parent chromosomes are selected randomly from this population with probabilities inversely proportional to their C values and an offspring is reproduced by applying crossover or mutation operations. The process is repeated to create new generations. The GA continues until a convergence criterion has been met. A parallel implementation is performed using the *Island Model*, see e.g. Darrell et al. (1997).

We suppose that upper bounds P_0 and Q_0 on the orders p_j and q_j of the FARIMA regimes are known. To estimate the model properly, each regime should contain a sufficient number of observations which are called *minimum span*. For the jth piece, this number should depend on the orders p_j and q_j and be an increasing function of $p_j + q_j$, which is denoted by $\mathrm{sp}(p_j + q_j)$. Our implementation of the GA is an adaptation of the one used by Davis et al. (2006).

Chromosome representation We need to encode the information that characterizes a fitted model in a vector g. How to choose a proper chromosome representation is difficult and the performance results of the GA certainly depend on the encoding. Here $g = (g_1, \ldots, g_n)$ is of length n, the number of observations, with gene values g_t defined by

$$g_t = \begin{cases} p_j & \text{if } t = \tau_{j-1} \text{ and the AR order for the } j\text{th piece is } p_j, \\ q_j & \text{if } t = \tau_{j-1} + 1 \text{ and the MA order for the } j\text{th piece is } q_j, \\ -1 & \text{otherwise.} \end{cases}$$

Initialization Each chromosome in the initial population is generated as follows. First, select a value for p_1 and q_1 from $\{0, \ldots, P_0\}$ and $\{0, \ldots, Q_0\}$, respectively, with equal probabilities and set $g_1 = p_1$ and $g_2 = q_1$. Then the next $\mathrm{sp}(p_1 + q_1) - 2$ genes $(g_3, \ldots, g_{\mathrm{sp}(p_1+q_1)})$ are set to -1 to satisfy the minimum span constraint. Now the next gene, $g_{\mathrm{sp}(p_1+q_1)+1}$ is set as a BP with probability π_B or is set to -1 with probability $1 - \pi_B$. If it is set as a BP, then we randomly select a value for p_2 and q_2 as above, we set $g_{\mathrm{sp}(p_1+q_1)+1} = p_2$, $g_{\mathrm{sp}(p_1+q_1)+2} = q_2$, and the process is repeated. If the number of remaining genes in the chromosome is less than $\mathrm{sp}(p_j + q_j) - 2$, we

re-select (p_j, q_j) until $\text{sp}(p_j + q_j) - 2$ is less than or equal to this number. The procedure ends when the number of remaining genes is less than $\text{sp}(0) - 2$, in which case we consider that there is no break with the previous block, i.e. all of the last genes are set to -1.

Crossover and mutation New chromosomes are generated from the initial population by choosing randomly either a crossover operation with probability π_C or a mutation operation with probability $1 - \pi_C$.

For the crossover operation, two parent chromosomes are selected randomly from the current population of chromosomes with probabilities inversely proportional to their C values. The genes of the offspring chromosome are inherited from his parents in the following way. For the two first genes (g_1, g_2), take on the corresponding values randomly from either the first or second parent with equal probabilities. Then the next $\text{sp}(g_1 + g_2) - 2$ genes are set to -1 to fulfil the minimum span constraint. For the next gene, say g_i, select one of the two parents with probability $1/2$ and we denote by g_i^p the corresponding gene of the selected parent. If $g_i^p \neq -1$ and $g_{i+1}^p \neq -1$, set $(g_i, g_{i+1}) = (g_i^p, g_{i+1}^p)$ and set the next $\text{sp}(g_i + g_{i+1}) - 2$ genes to -1. If $g_i^p = -1$ or $g_{i+1}^p = -1$, set $g_i = -1$. Then repeat the same gene-inheriting process to the next available gene.

For the mutation operation, the genes of the offspring chromosome are inherited from only one parent. For the two first genes (g_1, g_2), take on the genes of the parent with probability π_P or generate new orders with probability $1 - \pi_P$, and then apply the minimum span constraint. The next gene, say g_i, is fixed according to one of the three following possibilities. (a) with probability π_P, take on the parent's gene g_i^p. If $g_i^p \neq -1$ and $g_{i+1}^p \neq -1$, set $(g_i, g_{i+1}) = (g_i^p, g_{i+1}^p)$ and set the next $\text{sp}(g_i + g_{i+1}) - 2$ genes to -1. If $g_i^p = -1$ or $g_{i+1}^p = -1$, set $g_i = -1$. (b) with probability π_N, set $g_i = -1$. (c) with probability $1 - \pi_P - \pi_N$, generate new orders and then apply the minimum span constraint. The same process is applied to the next available gene.

Parallel implementation To speed up the convergence rate and to reduce the chance of converging to a local minimum, a parallel implementation of the GA is performed using the Island Model. In this model, the population of N chromosomes is divided into NI sub-populations called islands with N/NI chromosomes and a GA is run in each island. Moreover, the islands exchange individuals periodically according to the following migration policy. After every M_i generations, the worst M_N chromosomes from the jth island are eliminated and the best M_N chromosomes from the $(j-1)$th island are duplicated in the jth island, $j = 2, \ldots, NI$. For $j = 1$, the best M_N chromosomes are duplicated from the NIth island.

Stopping rule At the end of each migration in the Island Model, the overall best chromosome is noted. If this chromosome does not change for M_C consecutive migrations, or the total number of migrations exceeds M_M, this "best" chromosome is taken as the solution to the optimization problem and the whole GA ends.

5 Monte Carlo simulations

Here we compare the practical performances of criteria C, D and BIC through a set of Monte Carlo simulations. All the experiments are based on 1000 replications and we take $\sigma_j = 1$ for $j = 1, \ldots, m+1$ in (2). In the following, we use the standardized break fraction $\lambda_j = \tau_j/n$. The choice of the probabilities in a GA and the parameters in an Island Model depends on the data, and good guidance can be found in Sivanandam and Deepa (2008) and Darrell et al. (1997). In our simulations, the number of chromosomes is $N = 900$ except in Section 5.4 where $N = 2500$, the maximum orders are $P_0 = Q_0 = 3$, the minimum span function is $\mathrm{sp}(x) = 300 + 50x$, the probabilities in the GA are $\pi_B = 600/n$, $\pi_C = (n - 600)/n$, $\pi_P = \pi_N = 0.3$, the parameters in the Island Model are $NI = 30$ (in Section 5.4, $NI = 50$), $M_i = 10$ and $M_N = 2$, and the parameters in the stopping rule are $M_C = 5$ and $M_M = 30$.

In Section 5.1, we study the behavior of criterion C versus the BP number m, their locations and the orders of the FARIMA models. Section 5.2 concentrates on the case of a stationary model. In Section 5.3, we consider the case where the series has only one BP, while Section 5.4 investigates the case of multiple BPs. In model (1)–(2), it is assumed that the different segments are Gaussian and independent. Since in practice these two hypotheses may not be satisfied, we consider the non-Gaussian and dependent regime cases in Sections 5.5 and 5.6, respectively. In Section 5.7, we consider the case where the data generating process is not a piecewise stationary model, but the series has a slowly varying spectrum.

5.1 Behavior of criterion C

Here, we consider a piecewise FARIMA$(1, d, 1)$ model of length $n = 4000$ with two BPs at $\tau_1 = 1000$ ($\lambda_1 = 0.25$) and $\tau_2 = 2000$ ($\lambda_2 = 0.50$) whose parameters for each regime are given in table 1. Emphasis is put on the variations of criterion C when the parameters fluctuate in a neighborhood of the true values. For instance, we study the behavior of C as a function of the BP locations when the BP number and the orders are the true ones. When the variations of C with a given parameter close to the optimum are large, it is easier to estimate this parameter by minimizing C.

Parameters \ Regime	1	2	3
d_j	0.3	0.2	0.4
ϕ_j	0.7	-0.4	-0.7
θ_j	-0.6	-0.8	0.5

Table 1: Model parameters in section 5.1.

BP locations We fix the BP number to its true value $m = 2$ and the orders to their true values $(p_j, q_j) = (1, 1)$ for all j in (8), and we study the behavior of C versus (λ_1, λ_2) in a neighborhood of the true break fractions $(0.25, 0.50)$. The level curves of the sample means of C from all experiments are plotted in figure 1. There is only one minimum which is equal to 5778.223 and is reached at $(0.25, 0.50)$. Furthermore, small variations of (λ_1, λ_2) around $(0.25, 0.50)$ imply large variations in C.

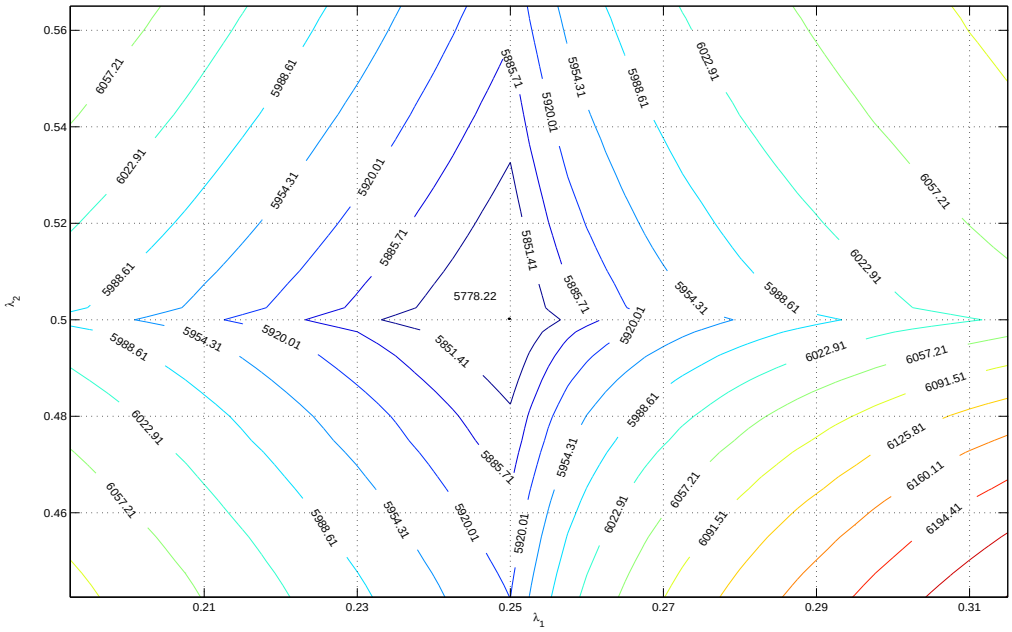

Figure 1: Level curves of C.

BP number We fix the orders to their true values in (8), and we study the behavior of C versus the BP number. First, we consider the case where m is overestimated taking $m = 3$ in (8). The sample means and standard errors (SEs) in parentheses of the estimated break fractions $(\widehat{\lambda}_1, \widehat{\lambda}_2, \widehat{\lambda}_3)$ and of C are given in table 2, where % denotes the corresponding percentage of experiments. In each experiment, two of the three estimated fractions correspond exactly to the true fractions. The other estimated fraction may be located in any stationary regime $X_{t,j}$. Therefore, when m is larger than the true value, the minimization of C gives the true BP locations and additional spurious BPs. Moreover, the sample means of C are much larger than the minimum value, 5778.223.

Consider now the case where m is underestimated. Figure 2 displays the sample means of C versus the break fraction λ_1 when $m = 1$ in (8). The global minimum 6030.586 is reached at one of the true fractions and there is a local minimum at the other true fraction.

%	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$	$\widehat{\lambda}_3$	C
55.6	0.126 (0.015)	0.25 (0)	0.50 (0)	5809.211 (44.775)
19.3	0.25 (0)	0.375 (0.015)	0.50 (0)	5804.557 (46.225)
25.1	0.25 (0)	0.50 (0)	0.683 (0.097)	5811.703 (46.942)

Table 2: Break fractions estimation and C value when $m = 3$.

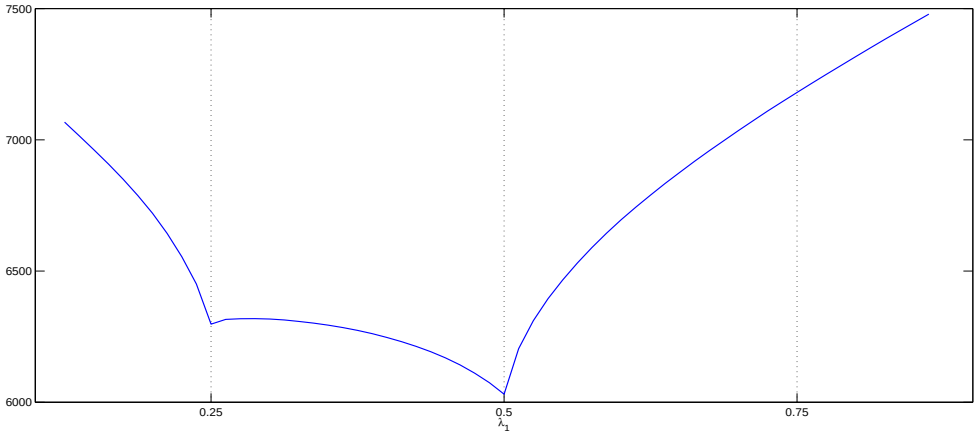

Figure 2: C versus λ_1 when $m = 1$.

Model orders We fix the BP number and the locations to their true values as well as $(p_1, q_1) = (p_2, q_2) = (1, 1)$ in (8) and we are interested in the selected orders $(\widehat{p}_3, \widehat{q}_3)$. For each experiment, C is calculated for $0 \leq p_3, q_3 \leq 3$ and table 3 presents the percentage of selection of each pair $(\widehat{p}_3, \widehat{q}_3)$ and the corresponding sample means of C. We observe that the true pair is selected in 90% of the cases and that C overestimates the orders in the remaining cases.

$(\widehat{p}_3, \widehat{q}_3)$	(0,0)	(0,1)	(0,2)	(0,3)	(1,0)	**(1,1)**	(1,2)	(1,3)
%	0	0	0	0	0	**89.8**	2.4	0.2
C	6634.352	6167.448	5897.957	5844.158	5820.933	**5778.223**	5781.335	5785.667
$(\widehat{p}_3, \widehat{q}_3)$	(2,0)	(2,1)	(2,2)	(2,3)	(3,0)	(3,1)	(3,2)	(3,3)
%	1.1	1.6	0.9	0.2	2.3	0.2	0.3	1.0
C	5786.853	5781.346	5784.685	5789.188	5782.369	5785.650	5789.157	5792.631

Table 3: Selected orders $(\widehat{p}_3, \widehat{q}_3)$ and sample means of C.

The results above show that the variations of C with the BP number and their locations are more important than the variations with the orders. Moreover, in most of cases where C selects wrong orders, the chosen orders are larger than the true ones.

5.2 Stationary model

We consider a stationary FARIMA$(0, d, 0)$ model where $d = 0.3$ and length n varies from 600 to 2000. We compare the performances of criteria C, D and BIC with regard to the detection of spurious BPs and the selection of the true orders as the time series length increases. We take $n \geq 600$ to get sensible MLEs of the parameters of the long-memory model (1)–(2), and $n = 2000$ is the maximum length of any stationary regime considered in this paper.

The three criteria choose either a model with zero or one BP. In table 4, we report the percentage of selection of $\widehat{m} = 0$ for each criterion. We see that C detects a spurious BP in 0.1% to 2.5% of the experiments, D in 13.3% to 19.6%, and BIC in 33.3% to 45.9%. For C and D, this percentage only slightly increases as n increases, which is not the case for BIC. This is not surprising that BIC has the worst performance since, according to (8), (9) and (10), BIC has the smallest penalty term on the BP number m.

Criterion \ n	600	800	1000	1200	1400	1600	1800	2000
C	98.4	97.5	98.7	98.2	98.8	99.5	99.9	99.1
D	80.7	80.4	82.5	83.4	84.2	84.4	85.2	86.7
BIC	64.6	62.3	66.7	62.5	60.1	57.8	54.6	54.1

Table 4: Percentage of selection of $\widehat{m} = 0$ (stationary model).

Table 5 gives the percentage of selection of a FARIMA$(0, d, 0)$ model (the percentage is calculated from all experiments and not only the ones where $\widehat{m} = 0$ is selected). We see that C clearly outperforms D and BIC. This may be explained as follows. According to (8), (9) and (10), C is the criterion with the highest penalty terms on the orders (p_j, q_j), followed respectively by D and BIC. Moreover, since $\log_2^+ 0 = \log_2^+ 1 = 0$, these penalties are the same in D for $p_j = 0$ and $p_j = 1$, and for $q_j = 0$ and $q_j = 1$. Hence, D and BIC overestimate the orders more frequently and more importantly than C does.

Criterion \ n	600	800	1000	1200	1400	1600	1800	2000
C	82.1	84.4	86.3	87.4	89.0	90.5	90.3	89.4
D	56.0	56.8	60.6	66.1	70.9	73.4	75.6	76.2
BIC	42.1	41.8	48.7	47.5	50.8	48.3	46.6	45.4

Table 5: Percentage of selection of a FARIMA$(0, d, 0)$ model (stationary model).

5.3 Single BP

We consider a piecewise FARIMA$(1, d, 1)$ model with one BP and different parameter sets in table 6 for the two regimes. Two time series lengths are considered, $n = 1000$ and $n = 2000$. When $n = 1000$, we take $\tau_1 = 500$ ($\lambda_1 = 0.5$), and when $n = 2000$, we take $\tau_1 = 500$ ($\lambda_1 = 0.25$), $\tau_1 = 1000$ ($\lambda_1 = 0.5$), and $\tau_1 = 1500$ ($\lambda_1 = 0.75$), respectively. We want to see the effects of the parameter changes between the two regimes and the sizes of the two regimes on the BP number selection, the preciseness of estimation of the break fraction, and the orders selection for criteria C, D and BIC.

First, we study the effects of the parameter changes between the two regimes. We take $n = 1000$ and $\lambda_1 = 0.5$. The parameters of the second regime are the same for each set, while the parameters of the first regime are different. Therefore, the magnitude of a parameter change is calculated with respect to the value of the parameter in the second regime. In sets 1 and 2, only parameters θ_j change between the two regimes, and the magnitude of change is 0.4 (50% of 0.8) for set 1, and 0.2 (25% of 0.8) for set 2. Sets 3 and 4 are designed similarly for parameters ϕ_j, the magnitude of change is 0.4 (100% of 0.4) for set 3, and 0.2 (50% of 0.4) for set 4. In sets 6, 7 and 8, only parameters d_j change, the magnitude of change being 0.3 (75% of 0.4) for set 6, 0.2 (50% of 0.4) for set 7, and 0.1 (25% of 0.4) for set 8. In set 5, all parameters change between the two regimes, the magnitude of change being 0.1 for each parameter (12.5% of 0.8 for θ_j, 25% of 0.4 for ϕ_j, 25% of 0.4 for d_j). Figure 3 displays typical realizations for each parameter set. Not surprisingly, the differences between the two regimes are more visible for set 1 than for set 2, for set 3 than for set 4, for set 6 than for set 7, and for set 7 than for set 8.

Set	$(d_1, \phi_1, \theta_1; d_2, \phi_2, \theta_2)$
1	(0.4, 0.4, -0.4; 0.4, 0.4, -0.8)
2	(0.4, 0.4, -0.6; 0.4, 0.4, -0.8)
3	(0.4, 0.8, -0.8; 0.4, 0.4, -0.8)
4	(0.4, 0.6, -0.8; 0.4, 0.4, -0.8)
5	(0.3, 0.5, -0.7; 0.4, 0.4, -0.8)
6	(0.1, 0.4, -0.8; 0.4, 0.4, -0.8)
7	(0.2, 0.4, -0.8; 0.4, 0.4, -0.8)
8	(0.3, 0.4, -0.8; 0.4, 0.4, -0.8)

Table 6: Model parameters in section 5.3.

We apply the three criteria to fit a piecewise FARIMA model (1)–(2) to each realization. For all sets and all experiments, C and D select zero or one BP, BIC selects zero, one or two BPs, two BPs being selected only for some sets in a few experiments : 2.9% for set 2, 3.6% for set 4, 10.3%

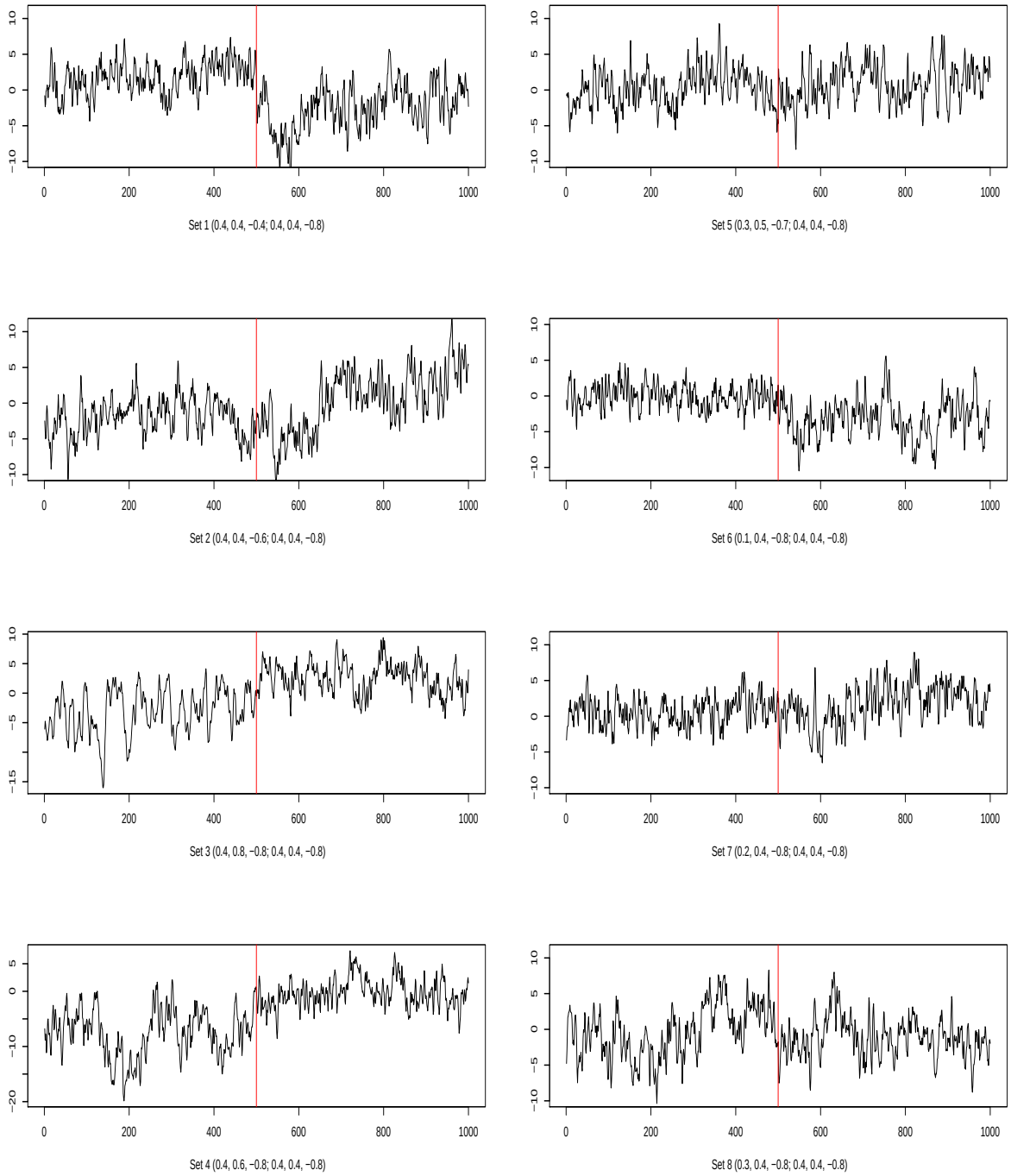

Figure 3: Realizations of a single BP model for different parameters ($n = 1000$, $\lambda_1 = 0.5$). The vertical line indicates the BP location.

for set 5, 4.1% for set 7 and 6.5% for set 8. The BP location estimation results are reported in table 7, where % denotes the percentage of experiments for which one BP is selected, Mean, SE and MSE denote respectively the sample mean, the SE and the mean-squared error of the estimated break fraction $\widehat{\lambda}_1$. For all criteria, the percentage of selection of the true BP number and the preciseness of estimation of the break fraction decrease as the (relative) magnitude of the parameter change decreases (compare sets 1 and 2, sets 3 and 4, and sets 6, 7 and 8). For each set, the best percentage is obtained with C followed by D and BIC in decreasing order. Nevertheless, when the magnitude of the parameter change decreases, the differences between the percentages decrease. For all sets, except sets 5 and 8, the true BP number is found in more than 57% of the experiments by the three criteria. We see on figure 3, that there are less differences between the two regimes for the sets 5 and 8 than for the other sets. This may explain why the three criteria encounters more difficulties in these two cases. Concerning, the preciseness of estimation of the break fraction, C outperforms D which itself outperforms BIC both in terms of bias and variance, for all parameter sets. Comparing sets 1, 4 and 7 in which only parameters θ_j, ϕ_j and d_j, respectively, change between the two regimes with the same relative magnitude of change, we see that a change in the LRD parameter of a FARIMA process with a fixed ARMA part seems to be more difficult to estimate than a change in the ARMA part only, using any of the three criteria. Furthermore, a small change in the ARMA part does not seem to be very helpful for finding a BP when the LRD parameter changes slightly also. Indeed, the results for set 5 in which all parameters change between the two regimes are only slightly better than the results for set 8 in which only d_j parameters change with the same relative magnitude of change.

Set	C				D				BIC			
	%	Mean	SE	MSE	%	Mean	SE	MSE	%	Mean	SE	MSE
1	81.5	0.499	0.017	2.76e-4	75.5	0.498	0.019	3.81e-4	69.9	0.503	0.025	6.54e-4
2	65.1	0.498	0.025	6.21e-4	62.2	0.495	0.027	7.75e-4	57.7	0.496	0.031	9.59e-4
3	89.3	0.500	0.011	1.14e-4	84.4	0.501	0.017	2.82e-4	80.4	0.502	0.022	5.14e-4
4	69.8	0.500	0.023	5.13e-4	65.3	0.503	0.023	5.35e-4	61.0	0.505	0.028	8.67e-4
5	49.8	0.499	0.030	9.34e-4	48.6	0.497	0.036	1.01e-3	46.3	0.495	0.044	1.98e-3
6	65.5	0.502	0.020	4.18e-4	62.3	0.496	0.024	5.85e-4	59.8	0.505	0.034	1.15e-3
7	58.6	0.501	0.028	7.66e-4	58.2	0.504	0.033	1.02e-3	57.9	0.504	0.039	1.95e-3
8	44.8	0.499	0.031	9.78e-4	44.6	0.506	0.037	1.66e-3	44.3	0.506	0.048	2.43e-3

Table 7: BP estimation (independent Gaussian segments, $n = 1000$, $\lambda_1 = 0.5$).

In table 8, we present the percentages of true orders selection for criteria C, D and BIC when each criterion selects $\widehat{m} = 1$. We see that C has the best percentages for each set (the smallest

percentage being 67%), followed respectively by D and BIC, the differences between C and BIC being around 15%. On the other hand, when the true BP number is selected, the percentages of true orders selection do not seem to be related to the magnitude of the parameter change. Indeed, in table 6, the parameters for the second regime are the same for all sets and only the parameters for the first regime differ. Nevertheless, in table 8, for each criterion, the percentages of true orders selection for the second regime are almost the same for all sets. When the true orders are not selected, the three criteria tend to overestimate the orders but C behaves better than D and BIC since it selects closer orders to the true orders (the detailed results are not reproduced here). The tendency of C to overestimate the orders in some cases is coherent with the results in table 3.

Set	C		D		BIC	
	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2
1	67.32	83.61	63.82	77.80	55.72	67.11
2	79.31	82.76	69.90	73.75	59.28	65.41
3	75.92	86.45	71.27	78.51	56.48	67.14
4	78.80	83.12	73.52	81.10	63.56	65.39
5	78.88	85.22	70.80	78.70	64.75	62.02
6	69.90	85.40	57.75	83.66	51.51	66.60
7	82.20	84.90	71.13	81.20	64.67	66.81
8	91.96	87.72	81.65	83.48	68.65	66.72

Table 8: Percentage of true orders selection (independent Gaussian segments, $n = 1000$, $\lambda_1 = 0.5$).

Now, we investigate the effect of the length of the time series and the BP location on the BP number selection, the preciseness of estimation of the break fraction, and the orders selection for criteria C, D and BIC. We fix $n = 2000$, and we take respectively $\lambda_1 = 0.25, 0.5, 0.75$. The results for sets 3 and 8 are listed in the following since they correspond respectively to the best and the worst case when $n = 1000$ and $\lambda_1 = 0.5$. Indeed, the results in table 7 for set 3 show that the percentages of selection of the true BP number are the largest and the MSEs are the smallest for the three criteria, while for set 8, these percentages are the smallest and the MSEs are the largest.

The BP number selection results are given in table 9 for set 3 and table 10 for set 8. For both sets and any λ_1, the best percentage of selection of $\widehat{m} = 1$ is obtained with C followed by D and BIC in decreasing order, as in table 7. For each criterion, the percentages of selection of a given $\widehat{m}$ are nearly the same for every λ_1. Comparing the results for $\lambda_1 = 0.5$ in tables 9 and 10 with the ones in table 7, we see that the percentage of selection of the true BP number increases

as the length of the blocks increases. This is especially the case for set 8. On the other hand, as noticed in Section 5.2, spurious BPs are more likely to be detected as the length of the time series increases. This is actually the case for BIC which selects $\widehat{m} = 2$ in 0% and 6.5% of the experiments for sets 3 and 8, respectively, when $n = 1000$, while these percentages are 10.6% and 20.4%, respectively, when $n = 2000$. As already noticed, criteria C and D are more robust with regard to the detection of spurious BPs since $\widehat{m} = 2$ is never selected when $n = 1000$, and is selected in less than 5% of the experiments when $n = 2000$.

λ_1	C			D			BIC		
	$\widehat{m}$=0	$\widehat{m}$=1	$\widehat{m}$=2	$\widehat{m}$=0	$\widehat{m}$=1	$\widehat{m}$=2	$\widehat{m}$=0	$\widehat{m}$=1	$\widehat{m}$=2
0.25	3.4	94.6	2.0	9.1	87.3	3.6	7.6	83.3	9.1
0.5	4.0	92.9	3.1	9.2	86.7	4.1	8.3	81.1	10.6
0.75	3.3	93.5	3.2	7.5	88.1	4.4	7.9	81.2	10.9

Table 9: BP number selection (independent Gaussian segments, $n = 2000$) : Set 3.

λ_1	C			D			BIC		
	$\widehat{m}$=0	$\widehat{m}$=1	$\widehat{m}$=2	$\widehat{m}$=0	$\widehat{m}$=1	$\widehat{m}$=2	$\widehat{m}$=0	$\widehat{m}$=1	$\widehat{m}$=2
0.25	39.4	57.4	3.2	40.4	54.3	5.3	25.0	54.2	20.8
0.5	39.7	56.5	3.8	41.5	53.8	4.7	26.7	52.9	20.4
0.75	39.6	56.5	3.9	40.5	54.0	5.5	26.1	53.7	20.2

Table 10: BP number selection (independent Gaussian segments, $n = 2000$) : Set 8.

The BP location estimation results are given in table 11 for set 3 and table 12 for set 8, for the experiments where one BP is selected. For each criterion and each set, the MSE only varies slightly with λ_1, the best and the worse preciseness being obtained in all cases for $\lambda_1 = 0.5$ and $\lambda_1 = 0.25$, respectively. The MSEs are smaller than in table 7, which confirms the intuition that the longer the regimes, the better the estimation of the break fraction (see in particular the results for $\lambda_1 = 0.5$).

λ_1	C			D			BIC		
	Mean	SE	MSE	Mean	SE	MSE	Mean	SE	MSE
0.25	0.251	0.010	1.11e-4	0.252	0.015	2.46e-4	0.254	0.021	4.46e-4
0.5	0.500	0.008	8.91e-5	0.499	0.014	2.13e-4	0.501	0.019	3.83e-4
0.75	0.749	0.010	1.10e-4	0.749	0.014	2.25e-4	0.747	0.020	4.33e-4

Table 11: BP location estimation (independent Gaussian segments, $n = 2000$) : Set 3.

λ_1	C			D			BIC		
	Mean	SE	MSE	Mean	SE	MSE	Mean	SE	MSE
0.25	0.250	0.023	5.53e-4	0.252	0.028	8.15e-4	0.254	0.037	1.43e-3
0.5	0.500	0.022	5.07e-4	0.503	0.027	7.68e-4	0.504	0.035	1.21e-3
0.75	0.749	0.023	5.30e-4	0.748	0.028	8.18e-4	0.745	0.036	1.35e-3

Table 12: BP location estimation (independent Gaussian segments, $n = 2000$) : Set 8.

In tables 13 and 14, we present respectively for sets 3 and 8, the percentages of true orders selection for criteria C, D and BIC when each criterion selects $\widehat{m} = 1$. The results show that C has the best percentages for each set, followed respectively by D and BIC. Compared with the results in table 8, we see that the ability of each criterion, in particular D, to select the true orders increases in almost all cases for both sets. When $\lambda_1 = 0.5$, the difference is notable for each criterion for the first regime in set 3. As in the case $n = 1000$ and $\lambda_1 = 0.5$, when the true orders are not selected, the three criteria tend to overestimate the orders, C selecting closer orders to the true orders than D and BIC (the detailed results are not reproduced here).

λ_1	C		D		BIC	
	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2
0.25	76.95	90.69	72.01	84.23	57.78	78.53
0.5	85.41	89.21	79.12	82.36	70.26	76.81
0.75	86.23	87.10	82.66	81.90	77.52	69.54

Table 13: Percentage of true orders selection (independent Gaussian segments, $n = 2000$) : Set 3.

λ_1	C		D		BIC	
	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2
0.25	89.98	89.24	78.36	86.31	70.11	78.29
0.5	92.08	88.43	83.37	83.96	78.18	75.20
0.75	92.16	87.93	84.25	83.00	81.75	67.48

Table 14: Percentage of true orders selection (independent Gaussian segments, $n = 2000$) : Set 8.

As a result, C outperforms D and BIC in terms of percentage of selection of the true BP number, preciseness of estimation of the break fraction, and percentage of selection of the true orders. For all criteria, the percentage of selection of the true BP number and the preciseness of

estimation of the break fraction decrease as the (relative) magnitude of the parameter change decreases and increase as the length of the regimes increases.

5.4 Multiple BPs

In this simulation, we consider a piecewise FARIMA model of length $n = 4000$ with three BPs at $\tau_1 = 600$ ($\lambda_1 = 0.15$), $\tau_2 = 1600$ ($\lambda_2 = 0.40$) and $\tau_3 = 2400$ ($\lambda_3 = 0.60$). The orders of the four FARIMA regimes are different and the parameters are given in table 15. Figure 4 displays a typical realization of this model.

Parameters \ Regime	1	2	3	4
d_j	0.30	0.10	0.40	0.20
ϕ_j	0	0.60	0	0.80
θ_j	0	-0.70	-0.50	0

Table 15: Model parameters in section 5.4.

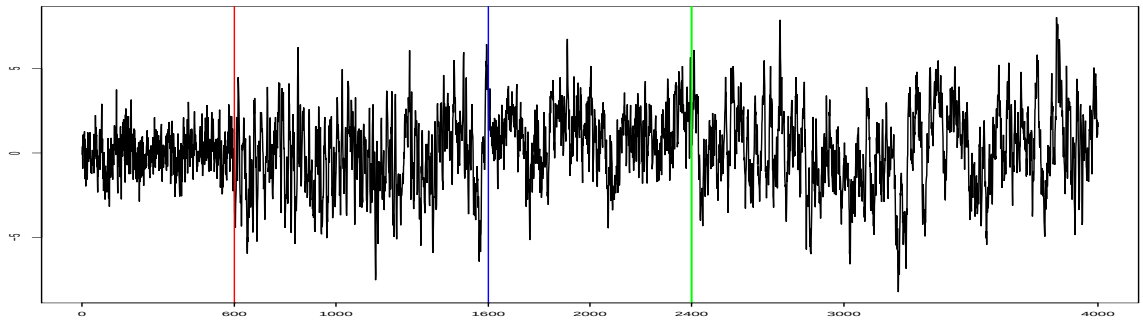

Figure 4: A realization of the model in table 15. The vertical lines indicate the BP locations.

We compare criteria C, D and BIC for the selection of a piecewise FARIMA model (1)–(2). The sample means and MSEs in parentheses of the estimated break fractions are given in table 16, where % denotes the percentage of experiments for which $\widehat{m}$ BPs are selected. We see that C and D never overestimate the BP number and underestimate it in 20% and 28% of the cases, respectively. Conversely, BIC never underestimates the BP number and overestimates it in 33% of the experiments. When the BP number is underestimated either by C or D, the estimated break fractions correspond to true BPs and the two criteria estimate the break fractions with good preciseness (MSE lower than 2e-4). In the same way, when the BP number is overestimated by BIC, the true break fractions are some of the estimated break fractions and again, the preciseness of estimation is good. This behavior was observed for C in Section 5.1.

Now, when the true BP number is selected, the three criteria estimate the break fractions with good preciseness, C performing the best in terms of bias and variance.

Criterion	$\widehat{m}$	%	$\widehat{\lambda}_j$, $j = 1, \ldots, \widehat{m}$
C	2	16.7	0.1502 (3.089e-6), 0.4004 (2.343e-5)
	2	3.2	0.1503 (2.734e-6), 0.6005 (2.949e-5)
	3	80.1	0.1502 (2.202e-6), 0.4001 (1.232e-5), 0.6002 (3.062e-5)
D	2	13.3	0.1483 (2.328e-5), 0.4024 (9.453e-5)
	2	14.6	0.1507 (1.201e-5), 0.5974 (1.001e-4)
	3	72.1	0.1509 (1.027e-5), 0.4026 (7.857e-5), 0.6021 (8.081e-5)
BIC	3	66.9	0.1533 (3.865e-5), 0.4040 (1.178e-4), 0.5929 (4.568e-4)
	4	33.1	0.1448 (5.261e-5), 0.3928 (1.034e-4), 0.6107 (7.765e-4) 0.8019 (–)

Table 16: Break fractions estimation (multiple BPs model).

In table 17, we present the percentages of true orders selection for the four regimes when C, D and BIC detect three BPs, respectively. We see that C has the best percentages, followed respectively by D and BIC. When the true orders are not selected, the three criteria tend to overestimate the orders but C selects closer orders to the true orders than D and BIC (the detailed results are not reproduced here).

Criterion \ Regime	1	2	3	4
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

Table 17: Percentages of true orders selection (multiple BPs model).

The results observed in Section 5.3 for a single BP and constant orders piecewise FARIMA model are confirmed for a multiple BPs model whose FARIMA regimes have different orders. Thus, it is preferable to use C than D and BIC.

5.5 Non-Gaussian time series

In (7), it is assumed that the segments $\{X_{t,j}\}$ are Gaussian. However, in finance and communication networks where LRD models are popular, the data are known to feature heavy tails, see for instance Embrechts et al. (1997) and Karasaridis and Hatzinakos (2001). Commonly used heavy-tailed distributions, like the Lévy and the Cauchy distributions belong to the class

of stable distributions. Here, we consider as in Section 5.3 a piecewise FARIMA$(1, d, 1)$ model of length $n = 1000$ with one BP at $\tau_1 = 500$ ($\lambda_1 = 0.5$) and the parameter sets 3 and 8 in table 6, but now the independent innovations $\{\epsilon_{t,j}\}$, $t \in \mathbb{Z}$, $j = 1, \ldots, m+1$ follow a non-Gaussian stable distribution. We compare the performances of criteria C, D and BIC. The characteristic function of a random variable Z with a stable distribution is for all $t \in \mathbb{R}$,

$$\mathrm{E}[\exp(itZ)] = \begin{cases} \exp\{it\mu - c|t|^\alpha[1 - i\beta\tan(\pi\alpha/2)\,\mathrm{sgn}(t)]\} & \text{if } \alpha \neq 1, \\ \exp\{it\mu - c|t|[1 + i\beta 2\pi^{-1}\,\mathrm{sgn}(t)\ln|t|]\} & \text{if } \alpha = 1, \end{cases}$$

where $\mathrm{sgn}(t) = t/|t|$ if $t \neq 0$ and $\mathrm{sgn}(0) = 0$, and the parameters $\mu \in \mathbb{R}$, $c > 0$, $\alpha \in (0, 2]$ and $\beta \in [-1, 1]$ are known as the location, scale, exponent and skewness parameters, respectively, see for instance Ibragimov and Linnik (1971, Chapter 2). In our simulations, we take $(\mu, c) = (0, 1)$ and we consider the three following cases, $(\alpha, \beta) = (1, 0), (0.5, 1), (1.5, 1)$ which correspond respectively to the Cauchy and Lévy distributions, and to a typical distribution for Internet data (see Ge et al., 2003).

We use the three criteria to fit a piecewise FARIMA model (1)–(2) to each realisation. As in the Gaussian case, C and D select zero or one BP for both sets in all experiments, BIC selects zero, one or two BPs, two BPs being selected only for set 8 in a few experiments : 2.5% when $(\alpha, \beta) = (1, 0)$, 3.8% when $(\alpha, \beta) = (0.5, 1)$, and 2.1% when $(\alpha, \beta) = (1.5, 1)$. The BP estimation results are given in table 18 for set 3 and table 19 for set 8, where % denotes the percentage of experiments for which one BP is selected. As in the Gaussian case, for both sets, the best percentage is obtained with C followed by D and BIC in decreasing order. Now, for the three criteria and both sets, these percentages are larger than in the Gaussian case, especially for set 8 where only parameters d_j change slightly between the two regimes. This may indicate that the long-memory parameter has more influence on the structure of an heavy-tailed time series than a Gaussian time series. As in the Gaussian case, for both sets, C outperforms D and BIC in terms of bias and variance of the estimate of the break fraction. Nevertheless, for the three criteria and both sets, the preciseness is not as good as in the Gaussian case, especially for set 3.

(α, β)	C				D				BIC			
	%	Mean	SE	MSE	%	Mean	SE	MSE	%	Mean	SE	MSE
(1,0)	96.7	0.512	0.057	3.39e-3	92.3	0.483	0.062	4.12e-3	81.6	0.479	0.075	6.00e-3
(0.5,1)	95.3	0.521	0.049	2.91e-3	89.8	0.478	0.059	4.03e-3	82.7	0.476	0.081	6.88e-3
(1.5,1)	97.2	0.501	0.012	4.41e-4	91.9	0.496	0.059	3.61e-3	81.2	0.490	0.059	3.78e-3

Table 18: BP estimation (independent non-Gaussian segments, $n = 1000$, $\lambda_1 = 0.5$) : Set 3.

In tables 20 and 21, we present respectively for sets 3 and 8, the percentages of true orders selection for criteria C, D and BIC when each criterion selects $\widehat{m} = 1$. The results show that C

(α,β)	C				D				BIC			
	%	Mean	SE	MSE	%	Mean	SE	MSE	%	Mean	SE	MSE
(1,0)	86.4	0.500	0.077	5.98e-3	80.2	0.504	0.078	6.25e-3	79.5	0.488	0.080	6.36e-3
(0.5,1)	86.5	0.510	0.079	6.30e-3	79.8	0.511	0.087	8.13e-3	78.7	0.492	0.086	7.80e-3
(1.5,1)	86.2	0.507	0.064	4.20e-3	81.9	0.512	0.063	4.48e-3	77.3	0.489	0.096	9.73e-3

Table 19: BP estimation (independent non-Gaussian segments, $n = 1000$, $\lambda_1 = 0.5$) : Set 8.

has the best percentages for each set, followed respectively by D and BIC. Compared with the results in table 8, we see that the ability of each criterion to select the true orders decreases in most cases for both sets. In particular, when $(\alpha, \beta) = (1.5, 1)$, the difference is notable for each criterion for the first regime in set 3. As in the Gaussian case, when the true orders are not selected, the three criteria tend to overestimate the orders, C selecting closer orders to the true orders than D and BIC (the detailed results are not reproduced here).

(α,β)	C		D		BIC	
	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2
(1,0)	78.28	54.29	65.79	48.95	56.81	41.67
(0.5,1)	79.75	65.90	72.64	54.72	55.53	41.38
(1.5,1)	32.72	76.44	25.23	62.18	17.25	38.91

Table 20: Percentage of true orders selection (independent non-Gaussian segments, $n = 1000$, $\lambda_1 = 0.5$) : Set 3.

(α,β)	C		D		BIC	
	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2
(1,0)	71.41	64.05	66.56	62.02	62.85	60.37
(0.5,1)	73.64	64.86	70.32	61.87	59.53	52.19
(1.5,1)	88.86	82.51	86.14	81.47	62.34	62.74

Table 21: Percentage of true orders selection (independent non-Gaussian segments, $n = 1000$, $\lambda_1 = 0.5$) : Set 8.

As a result, for each criterion, the percentage of selection of the true BP number increases, the preciseness of estimation of the break fraction decreases, and the percentage of selection of the true orders decreases, when the series is non-Gaussian. Moreover, C still outperforms D and BIC in all these aspects.

5.6 Non-independent segments

In (6), the different blocks $\{X_{t,j}\}$ are assumed to be independent. To study the influence of this assumption on the performances of criteria C, D and BIC, we consider a case where the pieces are correlated. More precisely, we assume that the observed time series $\{Y_t\}$ is modeled by (1) where $\{X_{t,j}\}$, $t \in \mathbb{Z}$, is now the FARIMA(p_j, d_j, q_j) process defined by the difference equation

$$\Phi_j(B)X_{t,j} = \Theta_j(B)(1-B)^{-d_j}\sigma_j\epsilon_t, \tag{11}$$

$\{\epsilon_t\}$, $t \in \mathbb{Z}$, being a sequence of iid zero-mean random variables. The distribution of ϵ_t can be either Gaussian or heavy-tailed, and we consider as in Sections 5.3 and 5.5 a piecewise FARIMA$(1, d, 1)$ model of length $n = 1000$ with one BP at $\tau_1 = 500$ ($\lambda_1 = 0.5$) and the parameter sets 3 and 8 in table 6. The results obtained with the different heavy-tailed stable distributions considered in Section 5.5 are similar, and we present here only the case $(\mu, c, \alpha, \beta) = (0, 1, 1.5, 1)$ which is of practical interest for Internet data.

A piecewise FARIMA model (1)–(2) is fitted to each realisation using the three criteria. As in the independent Gaussian and heavy-tailed cases, C and D select zero or one BP for both sets in all experiments, but now this is also the case for BIC. The BP estimation results are given in table 22 for set 3 and table 23 for set 8, where % denotes the percentage of experiments for which one BP is selected. When the innovations are Gaussian, for the three criteria and both sets, the percentages of selection of the true BP number and the preciseness of estimation of the break fraction decrease with respect to the case of independent segments presented in table 7. For set 3, C still outperforms D and BIC. For set 8, the percentages of selection of the true BP number decrease seriously and are smaller than 25% for the three criteria. This indicates that detecting a BP when parameters d_j only change slightly between two non-independent Gaussian regimes is almost impossible with these criteria. When the innovations are heavy-tailed, the results are similar to the ones in tables 18 and 19 for independent innovations.

$\{\epsilon_t\}$	C				D				BIC			
	%	Mean	SE	MSE	%	Mean	SE	MSE	%	Mean	SE	MSE
Gaussian	82.6	0.500	0.020	4.11e-4	78.9	0.504	0.038	1.57e-3	72.6	0.507	0.044	2.18e-3
$(\alpha, \beta) = (1.5, 1)$	97.8	0.509	0.058	3.43e-3	90.9	0.512	0.072	5.72e-3	82.4	0.523	0.058	4.28e-3

Table 22: BP estimation (non-independent segments, $n = 1000$, $\lambda_1 = 0.5$) : Set 3.

In tables 24 and 25, we present respectively for sets 3 and 8, the percentages of true orders selection for criteria C, D and BIC when each criterion selects $\widehat{m} = 1$. The results for both sets are similar to those obtained for independent regimes in table 8 for Gaussian innovations, and in tables 20 and 21 for heavy-tailed innovations. In the same way, when the true orders are not selected, the three criteria behave as in the case of independent regimes (the detailed results are

$\{\epsilon_t\}$	C				D				BIC			
	%	Mean	SE	MSE	%	Mean	SE	MSE	%	Mean	SE	MSE
Gaussian	22.5	0.497	0.040	2.26e-3	23.1	0.503	0.047	2.46e-3	23.7	0.513	0.066	4.36e-3
$(\alpha, \beta) = (1.5, 1)$	85.8	0.507	0.073	6.45e-3	82.5	0.513	0.084	7.19e-3	72.9	0.526	0.095	1.04e-2

Table 23: BP estimation (non-independent segments, $n = 1000$, $\lambda_1 = 0.5$) : Set 8.

not reproduced here).

$\{\epsilon_t\}$	C		D		BIC	
	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2
Gaussian	83.17	85.96	72.31	75.45	58.47	61.02
$(\alpha, \beta) = (1.5, 1)$	46.83	69.84	36.77	61.56	21.43	33.91

Table 24: Percentage of true orders selection (non-independent segments, $n = 1000$, $\lambda_1 = 0.5$) : Set 3.

$\{\epsilon_t\}$	C		D		BIC	
	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2
Gaussian	91.11	80.44	83.47	82.49	62.92	63.58
$(\alpha, \beta) = (1.5, 1)$	87.76	73.19	85.16	75.23	57.48	61.27

Table 25: Percentage of true orders selection (non-independent segments, $n = 1000$, $\lambda_1 = 0.5$) : Set 8.

As a result, for each criterion, the non-independence of the segments has few impact on the performances, except in the Gaussian case where parameters d_j only change slightly between two regimes, but detecting a BP in this case is also a difficult task for independent regimes. Hence, the hierarchy between the three criteria is the same for independent and non-independent segments.

5.7 Slowly varying FARIMA model

Here we analyse the results obtained with criteria C, D and BIC when we fit a piecewise FARIMA model (1)–(2) to a time series which is not piecewise stationary. We consider a locally stationary FARIMA model of length $n = 2000$ defined by the difference equation

$$Y_t = (1 - B)^{-d(t/n)} \epsilon_t, \tag{12}$$

where $\{\epsilon_t\}$, $t \in \mathbb{Z}$, is a sequence of iid zero-mean Gaussian random variables with unit variance, and $d(u) = 0.25 + 0.2\cos(2\pi u)$ for $u \in [0, 1]$. Following Dahlhaus (1997), the time-varying

spectral density of $\{Y_t\}$ is defined by

$$f(u,\nu) = \frac{1}{2\pi}|1 - e^{-i\nu}|^{-2d(u)}, \qquad -\pi \leq \nu \leq \pi.$$

A typical realization of model (12) is shown in figure 5, and $\log f(u,\nu)$ is plotted in figure 6(a).

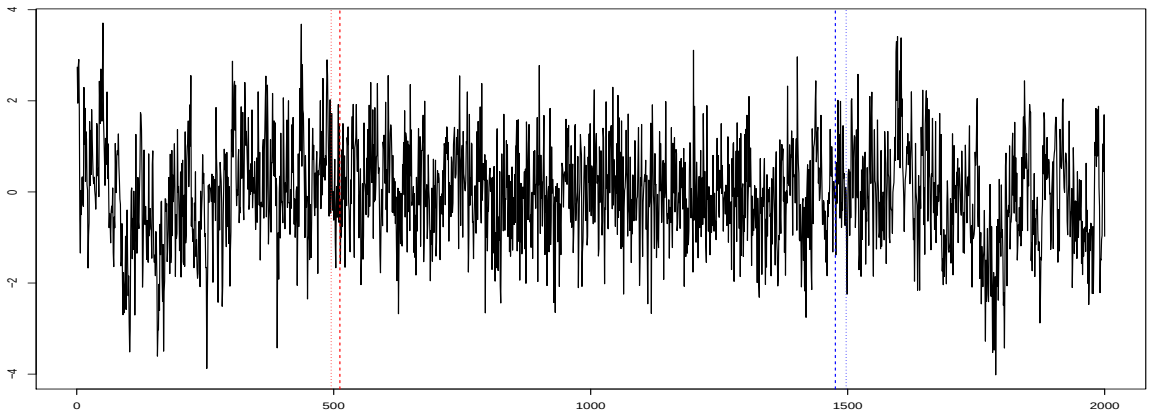

Figure 5: A realization of model (12). The vertical lines indicate the estimated BP locations by criterion C (dotted lines) and D (dashed lines).

For the realization in figure 5, C and D find two BPs located at $(\widehat{\tau}_1, \widehat{\tau}_2) = (495, 1497)$ and $(\widehat{\tau}_1, \widehat{\tau}_2) = (512, 1476)$, respectively, and BIC fit a stationary model. For C and D, the different regimes are identified as FARIMA$(0, d, 0)$ processes, while BIC fits a FARIMA$(1, d, 1)$ process. Since there is no structural break in model (12), we use as Ombao et al. (2001) and Davis et al. (2006) the averaged squared error (ASE) with respect to the time-varying log-spectrum to quantify the performances of the criteria. The ASE is defined by

$$\text{ASE} = \frac{1}{n(M_J/2+1)} \sum_{t=1}^{n} \sum_{k=0}^{M_J/2} \left| \log f(t/n, \nu_k) - \log \widehat{f}(t/n, \nu_k) \right|^2,$$

where $\widehat{f}(u,\nu_k)$ is an estimate of $f(u,\nu_k)$ obtained by fitting model (1)–(2) to a realization, J is a specified scale $J < [\log_2 n]$, $M_J = [n/2^J]$ and $\nu_k = 2\pi k/M_J$. Here, we use $J = 4$.

The experiment is repeated for 1000 independent realizations of model (12). The sample means and SEs in parentheses of the estimated break fractions and ASE for criteria C, D and BIC are given in table 26, where % denotes the percentage of experiments for which $\widehat{m}$ BPs are selected. In the majority of cases, C and D detect two BPs located around 0.25 and 0.75, respectively, while BIC chooses a stationary model. Observe that 0.25 and 0.75 correspond to the maxima of the absolute value of the first derivative of function $d(\cdot)$. Thereby, C and D find a spurious BP when $d(\cdot)$ has the largest variation. Even when C, D and BIC select one BP, this BP is close either to 0.25 or 0.75. Notice that contrarily to D and BIC, C never selects a stationary model. For each criterion, the smallest ASE is obtained when a piecewise model with

two BPs is fitted. Moreover, C outperforms D and BIC in the sense that the sample mean and SE of the ASE from all experiments is the smallest when C is used.

Criterion	$\widehat{m}$	%	$\widehat{\lambda}_j$, $j = 1, \ldots, \widehat{m}$	ASE
C	1	21.9	0.223 (0.029)	0.035 (0.003)
	1	12.6	0.778 (0.052)	0.039 (0.006)
	2	65.5	0.259 (0.045), 0.733 (0.045)	0.011 (0.004)
	all	100.0	–	0.019 (0.007)
D	0	10.6	–	0.121 (0.013)
	1	17.1	0.221 (0.026)	0.038 (0.005)
	1	18.7	0.726 (0.044)	0.040 (0.006)
	2	53.6	0.261 (0.046), 0.736 (0.047)	0.015 (0.005)
	all	100.0	–	0.035 (0.009)
BIC	0	46.4	–	0.128 (0.015)
	1	13.3	0.205 (0.048)	0.051 (0.006)
	1	12.1	0.795 (0.052)	0.049 (0.010)
	2	28.2	0.221 (0.051), 0.713 (0.074)	0.022 (0.006)
	all	100.0	–	0.079 (0.012)

Table 26: Break fractions estimation and ASE (slowly varying model).

The average of log-spectrum estimate obtained from all realizations by C, D and BIC are displayed in figure 6(b), (c) and (d), respectively. Figure 6(b) looks quite similar to figure 6(a) where the true time-varying log-spectrum of model (12) is plotted. This is not the case for figures 6(c) and (d).

In table 27, we present the percentages of orders $(0, 0)$ selection for the three criteria. We see that when a piecewise FARIMA model (1)–(2) with at least one BP is fitted ($\widehat{m} \neq 0$), the most frequently selected orders for each regime are $(0, 0)$, which is coherent with model (12). Now, when a stationary model is selected either by D or BIC, the orders selection turns out to be incoherent in most cases.

$\widehat{m}$	0	1		2		
Regime	0	1	2	1	2	3
C	–	78.84	84.64	78.17	89.62	77.25
D	44.84	65.29	73.91	72.69	77.48	74.16
BIC	26.57	54.88	49.55	62.54	69.32	67.41

Table 27: Percentages of orders $(0, 0)$ selection (slowly varying model).

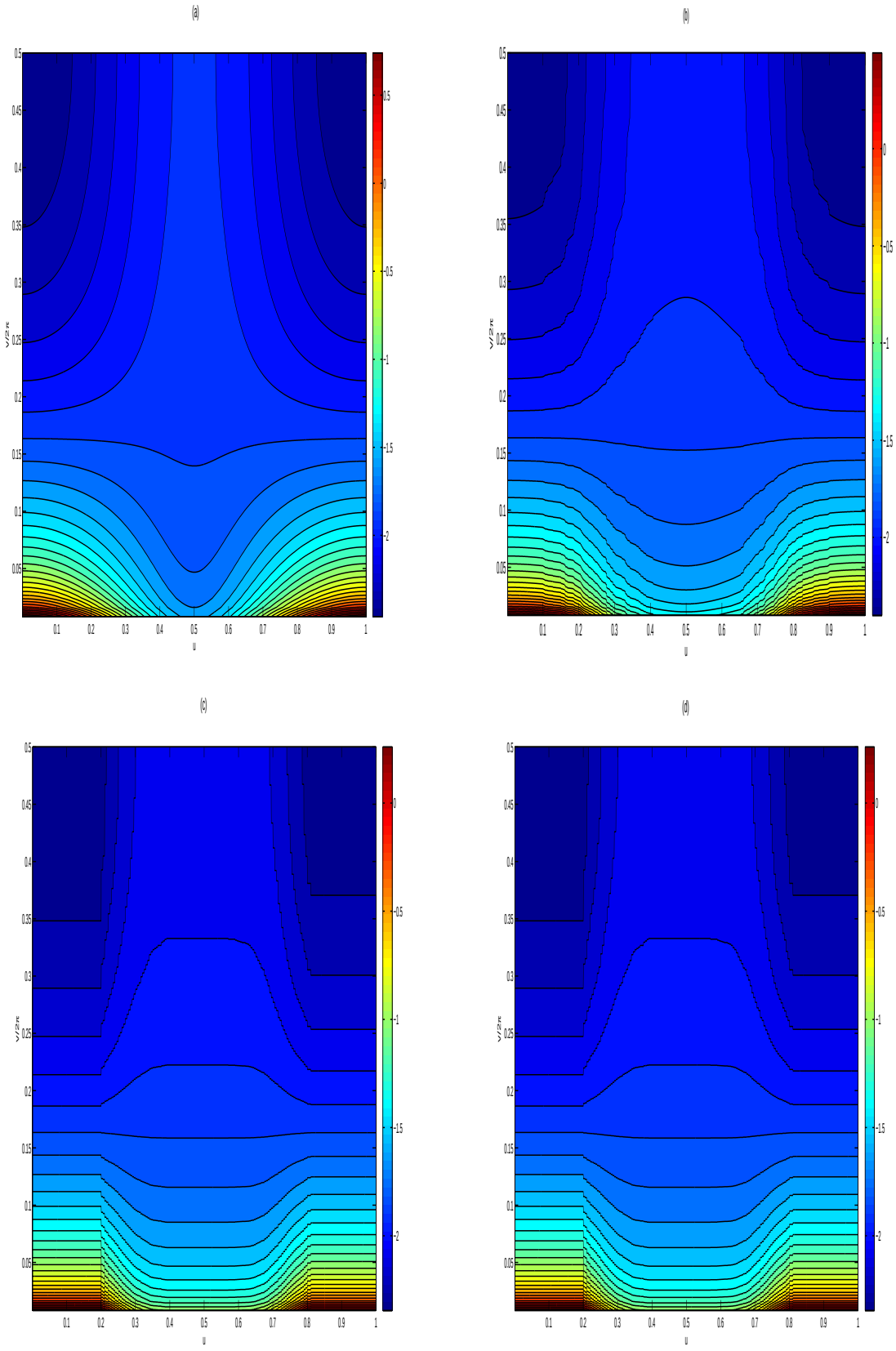

Figure 6: (a) True time-varying log-spectrum of model (12). Average of log-spectrum estimate obtained from 1000 realizations by (b) C; (c) D; (d) BIC.

As a result, the small values of the ASE and the similarities between figures 6(a) and 6(b) show that criterion C chooses a more adequate piecewise stationary FARIMA model (1)–(2) than D and BIC to fit slowly varying long-memory data.

6 Application to the River Nile data

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 7. 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.

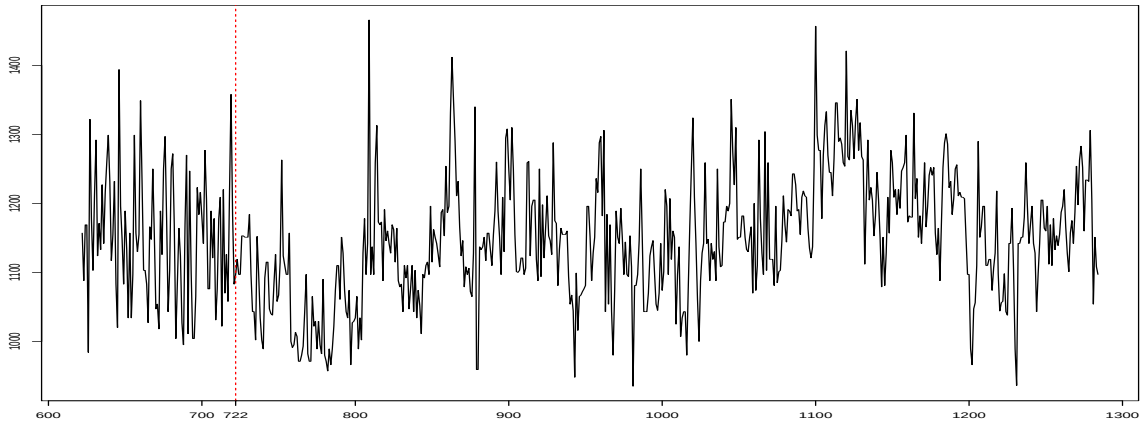

Figure 7: Nile river data (622–1284 AD). The vertical dashed line indicates the estimated BP location by criteria C, D and BIC.

We apply criteria C, D and BIC to fit a piecewise FARIMA model (1)–(2) to the Nile river data, and the parameters in the GA are the same as in Section 5 ($N = 900$, $NI = 30$). The three criteria choose a two-regimes model with one BP $\widehat{\tau}_1$ at 722 AD. For the first regime, a FARIMA$(2, \widehat{d}, 2)$ process with $\widehat{d} = 0.03$ is selected, while a FARIMA$(0, \widehat{d}, 0)$ process with $\widehat{d} = 0.45$ 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.

7 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 criterion based on MDL in the set of feasible solutions via a GA. Numerical experiments involving Gaussian and non-Gaussian piecewise FARIMA processes with independent and non-independent segments have shown that this criterion outperforms BIC and the criterion proposed by Davis et al. (2006) in terms of percentage of selection of the true BP number, preciseness of estimation of the break fractions, and percentage of selection of the true orders. In addition, when fitting a piecewise FARIMA process to a slowly varying long-memory time series, the smallest ASE and the best likeness between the true and the estimated time-varying log-spectrum are obtained with our criterion. When applying our methodology to the Nile River data a two regime piecewise FARIMA model with a BP at 722 AD is selected.

Acknowledgements

The authors are grateful to one anonymous referee whose constructive comments led to substantial improvements of the paper.

References

D. W. K. Andrews. Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):821–856, 1993.

J. Bai and P. Perron. Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–78, 1998.

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.

R. Baragona, F. Battaglia, and D. Cucina. Fitting piecewise linear threshold autoregressive models by means of genetic algorithms. *Comput. Statist. Data Anal.*, 47(2):277–295, 2004.

A. Barron, J. Rissanen, and B. Yu. The minimum description length principle in coding and modeling. *IEEE Trans. Inform. Theory*, 44(6):2743–2760, 1998. Information theory: 1948–1998.

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.

P. K. Bhattacharya. Some aspects of change-point analysis. In *Change-point problems (South Hadley, MA, 1992)*, volume 23 of *IMS Lecture Notes Monogr. Ser.*, pages 28–56. Inst. Math. Statist., Hayward, CA, 1994.

M. Chiogna, C. Gaetan, and G. Masarotto. Automatic identification of seasonal transfer function models by means of iterative stepwise and genetic algorithms. *J. Time Ser. Anal.*, 29(1):37–50, 2008.

R. Dahlhaus. Fitting time series models to nonstationary processes. *Ann. Statist.*, 25(1):1–37, 1997.

W. Darrell, R. Soraya, and B. Robert. Island model genetic algorithms and linearly separable problems. In *Evolutionary Computing*, pages 109–125. Springer Berlin / Heidelberg, London, UK, 1997.

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.

P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events*, volume 33 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1997.

P. Fearnhead. Exact Bayesian curve fitting and signal segmentation. *IEEE Trans. Signal Process.*, 53(6):2160–2166, 2005.

C. Gaetan. Subset ARMA model identification using genetic algorithms. *J. Time Ser. Anal.*, 21(5):559–570, 2000.

X. Ge, G. Zhu, and Y. Zhu. On the testing for alpha-stable distributions of ethernet network traffic. *Journal of Electronics (China)*, 20(4):309–312, 2003.

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.

P. Grünwald. *The Minimum Description Length Principle*. Monographs on Statistics and Applied Probability. MIT Press, Cambridge, MA, 2007.

M. H. Hansen and B. Yu. Model selection and the principle of minimum description length. *J. Amer. Statist. Assoc.*, 96(454):746–774, 2001.

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

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.

I. A. Ibragimov and Y. V. Linnik. *Independent and stationary sequences of random variables*. Wolters-Noordhoff Publishing, Groningen, 1971.

A. Karasaridis and D. Hatzinakos. Network Heavy Traffic Modeling using α-Stable Self-Similar Processes. *IEEE Trans. Commun*, 49(7):1203–1214, 2001.

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

M. Lavielle and E. Moulines. Least-squares estimation of an unknown number of shifts in a time series. *J. Time Ser. Anal.*, 21(1):33–59, 2000.

Š. Lazarová. Testing for structural change in regression with long memory processes. *J. Econometrics*, 129(1-2):329–372, 2005.

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.

H. C. Ombao, J. A. Raz, R. von Sachs, and B. A. Malow. Automatic statistical analysis of bivariate nonstationary time series. *J. Amer. Statist. Assoc.*, 96(454):543–560, 2001.

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

P. Perron. Dealing with structural breaks. In T. C. Mills and K. Patterson, editors, *Palgrave handbook of econometrics: volume 1: econometric theory*, pages 278–352. Palgrave Macmillan, UK, 2006.

P. Perron and Z. Qu. Estimating restricted structural change models. *J. Econometrics*, 134(2): 373–399, 2006.

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.

C. R. Reeves and J. E. Rowe. *Genetic Algorithms - Principles and Perspectives: A Guide to GA Theory*. Springer, New York, first edition, 2002. Operations Research/Computer Science Interfaces Series.

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.

S. N. Sivanandam and S. N. Deepa. *Introduction to genetic algorithms*. Springer, Berlin, 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.

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.

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

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.