PIECEWISE FARIMA MODELS FOR LONG-MEMORY TIME SERIES

Li Song and Pascal Bondon

CNRS UMR 8506, Université Paris-Sud, France.

July 27, 2011

Abstract

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

Keywords: Break point, local stationarity, long-memory, FARIMA model.

AMS 2000 subject classification: Primary 62M10, secondary 60G35.

1 Introduction

During the last years, many researchers in statistics, applied probability and computer science have focused on models for internet traffic. Previous works by Leland et al. (1994) and Paxson and Floyd (1994) among others, have shown that the classical models for telephone traffic cannot be applied to internet traffic. After the seminal study of Leland et al. (1994), increasing evidence has been put up for the failure of traditional (Poisson-based) models to account for the long-range dependence (LRD) present at the large time scales in network traffic. At the same time, due to the many different network mechanisms and various source characteristics, short-range dependence also exists and plays a central role, see Chang et al. (1995), Eran et al. (1995) and Garrett and Willinger (1994). Therefore, a model like the famous fractional autoregressive integrated moving average (FARIMA) model introduced by Granger and Joyeux (1980) and Hosking (1981), is required to describe both short and long memories simultaneously.

Accurate estimation of a FARIMA model often requires a large sample of data taken over a long period of time, which increases the chance of structural changes of the time series over time. Therefore, structural changes and LRD may appear jointly frequently. In particular, as shown by Stoev et al. (2006), the assumption that traffic data can be modeled by a stationary process may be unrealistic; the long-memory coefficient, as well as other coefficients characterizing the process, may change with time. Previous models, like the fractional Brownian motion, see e.g. Norros (1994), and the FARIMA model fail to analyze structural changes that are ubiquitous in network traffic. This motivates introducing piecewise stationary processes with LRD.

Modeling time series by piecewise processes is closely related to structural changes detection. If there are structural changes in the data generating mechanism, estimates derived from a constant model are not meaningful, inferences can be severely biased, and forecasts lose accuracy. Much effort has been devoted to estimation, testing and computation for models involving structural changes. The earliest references can go back to Maguire et al. (1952) and Page (1955). Recent works include Bai and Perron (1998), Bai and Perron (2003) and Perron and Qu (2006) where the authors have addressed the multiple structural changes problem in a linear regression model and have established the consistency and the rates of convergence of the least-squares estimates of the break points (BPs) and the regression coefficients, see also Perron (2006) for a review. A natural method for fitting a piecewise parametric model to data consists in minimizing some criteria based on the likelihood of the model or on the residuals and whose arguments are the number and the locations of structural changes and the parameters of each stationary regime. For instance, Davis et al. (2006) and Davis et al. (2008) have studied piecewise autoregressive processes and piecewise generalized autoregressive conditionally heteroscedastic processes using the minimum description length criterion and have established the consistency of the BP estimates in the autoregressive case. The criterion based approach may encounter some difficulties when the length of the time series is large since the practical minimization of the criterion is a difficult task when the search space is huge. This is the case for instance with internet traffic series which have many thousands of data.

The literature addressing the issue of structural changes in LRD models is relatively sparse, partly because these two phenomena are easy to confuse, see Bhattacharya et al. (1983) and Künsch (1986). Due to the slowly decaying correlation structure of a LRD process, statistic tests commonly used for assessing the stability of the model over time may encounter some difficulties. Nunes et al. (1996) and Kuan and Hsu (1998) have shown that many well-known structural change tests may suggest that a change has occurred in a LRD process, even though this is not the case. Due to this difficulty, studies addressing structural changes in LRD processes consider a partial structural change model where only some coefficients are allowed to vary. For example, Lavielle and Moulines (2000) have derived the consistency and the rate of convergence of the least-squares BPs estimate for a mean change LRD process whose number of changes is known; Coulon and Swami (2001) and Ray and Tsay (2002) have considered a piecewise FARIMA process with known and constant autoregressive and moving average (ARMA) orders; Gil-Alana (2008) has estimated the BP locations for a piecewise fractionally integrated process with a known number of breaks. These partial structural changes models may be unrealistic in practice. Therefore, a model with more flexibility in modeling non-stationarity and long-memory as well as a corresponding estimation procedure for long time series seem to be needed.

This work proposes a parametric model which inherits the advantages of FARIMA processes, to model local stationary long-memory time series. We choose a piecewise FARIMA process assuming that the time series can be divided into regimes that are stationary. This 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. From a practical point of view, a change in the mean make easier the detection of the BP, see for instance Brown et al. (1975). In our model, we do not include change in the mean. As far as we know, no consistency results exist for piecewise ARMA models where the parameters and the orders can change. Here we propose a procedure to estimate the BPs as well as the orders and the parameters of each regime which can be applied to long time series with many thousands of data and BPs satisfying assumptions (A1) and (A2) in Section 3, and we concentrate on simulations results.

The rest of this paper is organized as follows. In Section 2, the model is presented and in Section 3, the model fitting methodology is described. Numerical simulation results are presented and discussed in Section 4. A real traffic data modeling is considered in Section 5 and concluding remarks can be found in Section 6.

2 Model description

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 $\tau_{m+1} = n + 1$. For $j = 1, \ldots, m + 1$, the jth block of $\{Y_t\}$ 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

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

B is the backward operator $BX_t = X_{t-1}$, $\{\epsilon_{t,j}\}$, $t \in \mathbb{Z}$, $j = 1, \ldots, m + 1$, is a sequence of iid zero-mean random variables with finite variance, $d_j \in (0, 1/2)$, 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 $d_j < 1/2$, $\sum_{k=0}^{\infty} \varphi_k(d_j)^2 < \infty$ and the series in (3) converges in the mean square sense. Since the sequence $\{\epsilon_{t,j}\}$, $t \in \mathbb{Z}$, is zero-mean and iid, the series in (3) converges also almost surely.

Let $p \geq \max(p_j)$, $q \geq \max(q_j)$, $\alpha_j = (d_j, \phi_{j,1}, \ldots, \phi_{j,p}, \theta_{j,1}, \ldots, \theta_{j,q})$ where $\phi_{j,k} = 0$ for $k > p_j$ and $\theta_{j,k} = 0$ for $k > q_j$. Vector α_j contains the parameters of the jth model defined in $[\tau_{j-1}, \tau_j)$. The piecewise FARIMA process $\{Y_t\}$ is characterized by the BPs τ_j and the parameters α_j for $j = 1, \ldots, m + 1$.

3 Estimation procedure

The problem of fitting model (1)–(2) to data consists in finding $(m, \tau_1, \ldots, \tau_m, \alpha_1, \ldots, \alpha_{m+1})$. Instead of optimizing a selection criterion depending on all these parameters which is computationally difficult for long time series, we propose a four-step procedure. The main problem is to estimate the BPs accurately, which can be realized by detecting the changes in the parameter estimates. Stoev and Taqqu (2005) have revealed 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. After the differences between the parameter estimates in elementary intervals can be used to search the BPs which are dispersed into a few intervals.

In the following, K is the integer part of n/E, i.e. $K = \lfloor n/E \rfloor$ and we introduce the elementary intervals $I_k = ((k-1)E, kE]$ for $k = 1, \ldots, K-1$ and $I_K = ((K-1)E, n]$. We make the following assumptions:

(A1) There is no BP neither in I_1 nor in I_K.

(A2) At least $(2+\delta)E$ data separate two consecutive BPs for some $\delta > 0$.

Our estimation procedure consists in the following steps.

Step 1 : *Local estimation.* For each interval I_k, $k = 1, \ldots, K$, the model's parameters $\tilde{\alpha}_k$ are chosen by quasi Gaussian maximum-likelihood estimation (QMLE), see e.g. Beran (1994), and a pair $(\tilde{p}_k, \tilde{q}_k)$ is selected with the Bayes information criterion (BIC) as suggested by Torre et al. (2007). To catch the parameters changes with a comparatively small E, we choose QMLE since these estimates perform better than the two others popular estimates, namely the wavelet estimates, see e.g. Stoev et al. (2006), and the Whittle estimates, see e.g. Taqqu and Teverovsky (1997), when the data length is not long.

Step 2 : *Choose* $0 < \eta < \min\{0.5, \delta\}$ *and for* $m = 1, \ldots, \left\lfloor \frac{(K-2)E-1}{(2+\eta)E} \right\rfloor + 1$, *do Steps 2a,b,c.*

Step 2a : *Selection of the intervals with a BP.* If model (1)–(2) is suitable for the data, one expects that $\tilde{\alpha}_k$ 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}, $\tilde{\alpha}_k$ should be significantly different from both $\tilde{\alpha}_{k-1}$ and $\tilde{\alpha}_{k+1}$. Then, let $k_0 = 0$, $k_{m+1} = K+1$, and

$$(\hat{k}_1, \ldots, \hat{k}_m) = \underset{\{k_1, \ldots, k_m\}}{\operatorname{argmin}} \sum_{j=1}^{m+1} \sum_{k=k_{j-1}+1}^{k_j - 1} \Big(\psi_1(|\tilde{\alpha}_k - \bar{\alpha}_j|) + \psi_2(|\tilde{p}_k - \bar{p}_j| + |\tilde{q}_k - \bar{q}_j|) \Big), \tag{4}$$

where the minimum is taken over all possible m-tuples $(k_1, \ldots, k_m)$ satisfying $1 < k_1 < \cdots < k_m < K$ and assumption (A2) where δ is replaced by η, for any vector u with components u_i's, $|u| = \sum_i |u_i|$, $\bar{\alpha}_j = \dfrac{1}{k_j - k_{j-1} - 1} \displaystyle\sum_{k=k_{j-1}+1}^{k_j - 1} \tilde{\alpha}_k$, $\bar{p}_j$ (resp. $\bar{q}_j$) is the order which is the most frequently selected among the orders $\tilde{p}_k$ (resp. $\tilde{q}_k$) for $k = k_{j-1}+1, \ldots, k_j - 1$. In the case where

$\bar{p}_j$ (resp. $\bar{q}_j$) is not unique, the lowest order is chosen. For $j = 1, 2$, function $\psi_j : \mathbb{R}_+ \mapsto \mathbb{R}_+$ is strictly increasing and is defined by $\psi_j(x) = x^{a_j}[\ln(1+x)]^{b_j}$ where $a_j \geq 0$, $b_j \geq 0$ and $a_j b_j \neq 0$ (here, $x^0 = 1$ for any $x \geq 0$).

When a BP is located close to the upper bound of an elementary interval, minimizing (4) might lead to select this interval or the next one. A similar problem appears when a BP is close to the lower bound of an elementary interval. For this reason, we define the intervals containing a BP as being $(J_{\hat{k}_1}, \ldots, J_{\hat{k}_m})$ where $J_2 = (E, (2+\eta)E]$, $J_k = ((k-1-\eta)E, (k+\eta)E]$ for $k = 3, \ldots, K-2$ and $J_{K-1} = ((K-2-\eta)E, (K-1)E]$. The role of $\eta > 0$ is to avoid the possible erroneous selection when a BP is close to the limits of an elementary interval.

The choice of parameters a_j, b_j, η is discussed in Section 4.

Step 2b : *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. $((\hat{k}_{j-1}+\eta)E, (\hat{k}_j - 1 - \eta)E]$ where we set $\hat{k}_0 + \eta = 0$, and we define $\hat{\alpha}_p$ as the QMLE of α_j based on the data in this block where the orders (p_p, q_p) are selected by BIC. In the same way, let $\hat{\alpha}_n$ be the QMLE of α_{j+1} based on the data in the “next” block between $J_{\hat{k}_j}$ and $J_{\hat{k}_{j+1}}$, viz. $((\hat{k}_j + \eta)E, (\hat{k}_{j+1} - 1 - \eta)E]$ where we set $(\hat{k}_{m+1} - 1 - \eta)E = n$, and (p_n, q_n) be the orders selected by BIC. We treat $\hat{\alpha}_p$ and $\hat{\alpha}_n$ as two benchmarks. These estimates are more precise than any local estimate calculated in Step 2 since they involve more data. Suppose that $l \in J_{\hat{k}_j}$ is the BP τ_j. Then we can calculate the QMLE $\hat{\alpha}_{l_p}$ of α_j using the orders (p_p, q_p) and the QMLE $\hat{\alpha}_{l_n}$ of α_{j+1} using the order (p_n, q_n) based respectively on $((\hat{k}_{j-1} + \eta)E, l]$ and $(l, (\hat{k}_{j+1} - 1 - \eta)E]$. These estimates should be close to benchmarks $\hat{\alpha}_p$ and $\hat{\alpha}_n$, respectively. Hence, our choice of the BP estimate $\hat{\tau}_j$ is based on the following criterion

$$\hat{\tau}_j = \underset{l \in J_{\hat{k}_j}}{\operatorname{argmin}} \Big(\psi_1(|\hat{\alpha}_{l_p} - \hat{\alpha}_p|) + \psi_1(|\hat{\alpha}_{l_n} - \hat{\alpha}_n|) \Big). \tag{5}$$

Step 2c : *Estimation of the parameters of each stationary block.* Once $(\hat{\tau}_1, \ldots, \hat{\tau}_m)$ are obtained, the parameters α_j of the stationary sequence $X_{t,j}$ for $j = 1, \ldots, m+1$, can be estimated by QMLE and BIC, on the basis on the data in $(\hat{\tau}_{j-1}, \hat{\tau}_j]$, where $\hat{\tau}_0 = 1$ and $\hat{\tau}_{m+1} = n$. We denote by $\hat{\alpha}_j$ and $(\hat{p}_j, \hat{q}_j)$ the corresponding parameters and orders.

Step 3 : $m = 0$. We fit a FARIMA model to the series $\{Y_t\}$, $t = 1, \ldots, n$, estimating the model's parameters by QMLE and selecting the ARMA orders by BIC.

Step 4 : *Selection of the BP number.* For $m = 0, \ldots, \left\lfloor \frac{(K-2)E-1}{(2+\eta)E} \right\rfloor + 1$, we compute the sum of squared residuals of the fitted model (1)–(2) with m BPs, $S_n(\hat{\tau}_1, \ldots, \hat{\tau}_m)$, and the Gaussian $\log_2$-likelihood $\log_2 L_j$ of the jth segment for $j = 1, \ldots, m+1$. For $m = 0$, we take the model fitted in Step 3 and for $m > 0$, we employ the models fitted in Step 2c. These quantities are useful for calculating the following BP number selection criteria. The first three criteria are based on the Schwarz criterion (Schwarz, 1978) and differ in the severity of their penalty for over-specification. The last criterion is based on the minimum description length (MDL) principle introduced by Rissanen (1978). Following Yao (1988), the selected number of BPs $\hat{m}$ minimizes

$$\mathrm{C}_1(m) = \ln\left[S_n(\hat{\tau}_1, \ldots, \hat{\tau}_m)/n\right] + p^\star \frac{\ln n}{n}, \tag{6}$$

where $p^\star = \sum_{i=1}^{m+1}(p_i + q_i) + 2m + 1$ is the total number of parameters. A criterion proposed by Yao and Au (1989) takes the form

$$\mathrm{C}_2(m) = \ln\left[S_n(\hat{\tau}_1, \ldots, \hat{\tau}_m)/n\right] + m\frac{C_n}{n}, \tag{7}$$

where C_n satisfies some constraints, and Liu et al. (1997) introduced criterion

$$\mathrm{C}_3(m) = \ln\left[S_n(\hat{\tau}_1, \ldots, \hat{\tau}_m)/(n - p^\star)\right] + p^\star \frac{c_0(\ln n)^{2+\gamma_0}}{n}, \tag{8}$$

where $c_0 > 0$ and $\gamma_0 > 0$. Finally, following Davis et al. (2008) amounts to minimize

$$\mathrm{C}_4(m) = \log_2^+ m + (m+1)\log_2 n + \sum_{j=1}^{m+1}\left\{\log_2^+ \hat{p}_j + \log_2^+ \hat{q}_j + \frac{\hat{p}_j + \hat{q}_j + 2}{2}\log_2 \hat{n}_j - \log_2 L_j\right\},$$

where $\hat{n}_j = \hat{\tau}_j - \hat{\tau}_{j-1}$, $\log_2^+ x = 0$ if $x \in [0,1]$ and $\log_2^+ x = \log_2 x$ if $x > 1$. We compare these four criteria in Section 4.

Remark 1. Predefining a suitable length E for the elementary sub-series is not always an easy task: on the one hand, due to LRD, a reasonable number of observations are needed to obtain precise parameter estimates, and then E can't be too short; on the other, the probability of meeting a BP increases as E grows. Hence some restriction should be put on E, and E is chosen by empirical experience.

Remark 2. To reduce the complexity, $\hat{\alpha}_{l_p}$ and $\hat{\alpha}_{l_n}$ in Step 2b 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.

Remark 3. When the series is stationary, i.e., the true BP number is zero, the estimated parameters of each stationary block in Step 2c are almost the same and coincide with the parameters obtained in Step 3. Therefore, $S_n(\hat{\tau}_1, \ldots, \hat{\tau}_m)/n$ and $\sum_{j=1}^{m+1}\log_2 L_j$ do not vary too much with m. Hence, the four selection criteria in Step 4 are minimum for the true number of BPs $m = 0$. Simulations results (not reported here) illustrate this.

4 Simulations

Here we illustrate the estimation procedure by Monte Carlo simulations and show its effectiveness. All the experiments are based on 1000 replications of a piecewise FARIMA process (1)–(2) where the sequence $\{\epsilon_{t,j}\}$, $t \in \mathbb{Z}$, $j = 1, \ldots, m+1$, is Gaussian with unit variance. In all the simulations the maximum value of the ARMA orders considered in BIC is 7, we take $C_n = c_1\, n^{0.9}$ in (7), $\gamma_0 = 2$ in (8) and we choose c_0 and c_1 to get the same penalty in (6), (7) and (8) for $n = 2000$. This gives $c_0 = (\ln 2000)^{-3}$, but since m, p_i, q_i are unknown we take arbitrarily $2 \ln 2000 = c_1\, 2000^{0.9}$. The influence of a_j, b_j and η is discussed in Section 4.1. Since (4) and (5) measure the differences between the parameters of the different regimes $X_{t,j}$ in (2), we study in Section 4.2 the performance of our method when these parameters are very different and when they are close in a piecewise model. In Sections 4.2, we take $(p_j, q_j) = (1, 1)$ for all regimes, while Section 4.3 investigates the case where the orders (p_j, q_j) change. Lastly in Section 4.4, we investigate the case of close BPs.

In the following, we use the standardized break fraction $\lambda_j = \tau_j/n$. In Sections 4.2 and 4.3, $n = 40000$ and we have four BPs at $\lambda_1 = 0.2025$, $\lambda_2 = 0.4000$, $\lambda_3 = 0.6775$ and $\lambda_4 = 0.8675$. In Section 4.4, $n = 20000$ and we also have four BPs with the same break fractions as above. We take $E = 2000$, therefore $K = 20$ in Sections 4.2 and 4.3 and $K = 10$ in Sections 4.4.

4.1 Influence of a_j, b_j and η on Step 2

First, we set $\eta = 0.1$ and we investigate the influence of a_j and b_j. For this, we consider a piecewise FARIMA model of length $n = 10000$ with one BP at $\lambda_1 = 0.5$ whose first regime is a fractional noise with $d_1 = 0.4$ and second regime is a FARIMA$(1, d, 1)$ process with $(d_2, \phi_2, \theta_2) = (0.2, -0.7, 0.4)$. Figure 1 displays a typical realization of this process.

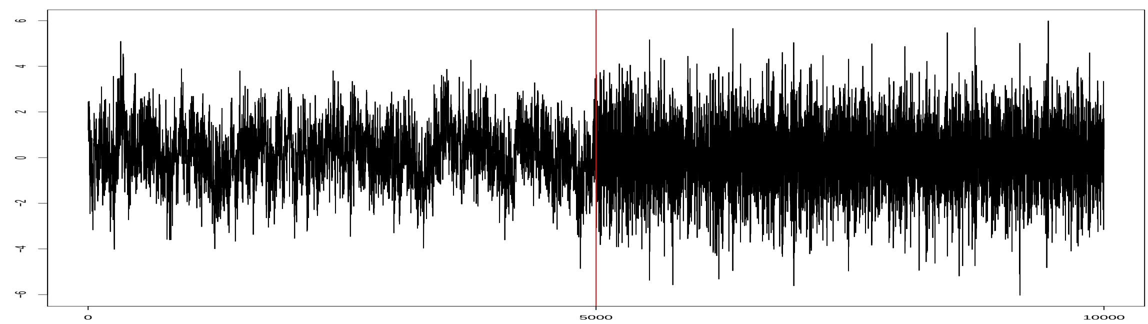

Figure 1: A realization of a single BP model. The vertical line indicates the true BP location.

Table 1 presents the mean-squared error (MSE) of the estimated break fraction $\hat{\lambda}_1$ using different values of (a_j, b_j). For all these values, the MSE is small and is almost the same. For larger values of (a_j, b_j), the MSE increases. For instance, when $(a_1, b_1) = (0, 1)$ and $(a_2, b_2) = (3, 3)$, the MSE is 3.0576e-3. We recommend to choose (a_j, b_j) in $[0, 1]$ and since the minimum MSE in Table 1 is obtained when $(a_1, b_1) = (0, 1)$ and $(a_2, b_2) = (0.5, 0)$, we shall use these values in all the simulations.

(a_2, b_2) \ (a_1, b_1)	(0,0.5)	(0.5,0)	(0.5,0.5)	(0,1)	(1,0)
(0,0.5)	7.4345e-4	8.1673e-4	5.1781e-4	4.7666e-4	9.1121e-4
(0.5,0)	6.4828e-4	7.7263e-4	5.2996e-4	4.1374e-4	5.0555e-4
(0.5,0.5)	9.1133e-4	1.0130e-3	7.5807e-4	5.0381e-4	8.7364e-4
(0,1)	7.1380e-4	7.3450e-4	5.1239e-4	4.8932e-4	6.4700e-4
(1,0)	8.6068e-4	8.8714e-4	5.7340e-4	4.3425e-4	8.0118e-4

Table 1: MSE of $\hat{\lambda}_1$ in terms of (a_j, b_j).

For the choice of η, we use a single BP piecewise FARIMA$(1, d, 1)$ model of length $n = 10000$ with different parameters for the two regimes given in Table 2 and different BP positions, namely $2.9E$, $2.95E$, $3.0E$, $3.05E$ and $3.1E$. The corresponding elementary intervals containing the BPs are I_3, I_3, I_3, I_4 and I_4, respectively.

Set \ Parameters	d_1	ϕ_1	θ_1	d_2	ϕ_2	θ_2
(a)	0.2	0.6	-0.2	0.3	-0.5	0.4
(b)	0.2	0.5	-0.2	0.2	-0.5	0.3
(c)	0.4	0.3	-0.5	0.1	-0.7	0.2

Table 2: Model parameters (single BP model).

Table 3 gives the number of selection of the right elementary interval in Step 2a for the different BP positions in the 1000 simulations. Although this number varies with the parameters of the FARIMA model, more than 90% of the results are the right ones when the BP locates (at least) $0.1E$ far from the upper bound of I_3. Thereby, we shall use $\eta = 0.1$ in Step 2a in all the simulations.

Set	$2.90E$	$2.95E$	$3.00E$	$3.05E$	$3.10E$
(a)	903	839	573	881	951
(b)	918	864	786	877	947
(c)	936	950	779	916	959

Table 3: Number of selection of the right elementary interval in Step 2a.

4.2 Regimes with the same ARMA orders

In this simulation, the parameters of the first three regimes $X_{t,j}$ in (2) are very different while the parameters of the last three regimes are close. These parameters are given in Table 4 and Figure 2 displays a typical realization of this model.

Parameters \ Regime	1	2	3	4	5
d_j	0.15	0.40	0.20	0.35	0.20
ϕ_j	0.80	0.80	-0.70	-0.30	-0.40
θ_j	-0.50	0.60	0.40	0.50	0.80

Table 4: Model parameters (regimes with the same ARMA orders).

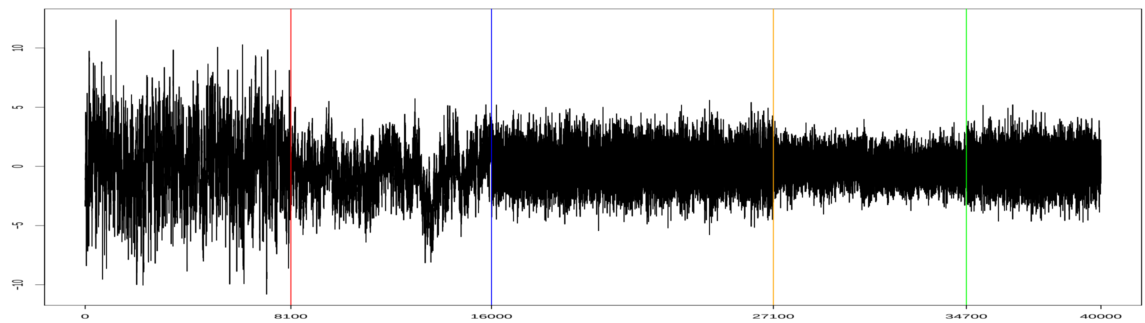

Figure 2: A realization of a piecewise FARIMA process whose regimes have the same ARMA orders. The vertical lines indicate the true BP locations.

In Table 5, we present the BP number selection in Step 4 in the 1000 simulations. We see that C_2 works better than the other criteria and select the true BP number in at least 90% of the cases. Criteria C_3 and C_4 give the right answer in 80% and 75% of the cases, respectively, while C_1 performs relatively poorly. No criteria underestimate the BP number.

Table 6 presents the sample means $\hat{\mu}(\hat{\lambda}_j)$, the standard errors $\hat{\sigma}(\hat{\lambda}_j)$ and the MSEs of the

$\hat{m}$ / Criterion	4	5	6	7	8	9
C_1	476	216	144	71	63	30
C_2	933	53	10	2	2	0
C_3	804	113	47	18	15	3
C_4	744	164	51	29	6	6

Table 5: Selected number of BPs in Step 4 (regimes with the same ARMA orders).

BP estimations in Step 2b when $\hat{m} = 4$. All the estimated BPs are close to the true BPs, and the MSEs are quite small. Note that the estimations of the third and the fourth BP are not as good as the estimations of the first two BPs since the MSEs are larger. The reason is that the selection of the intervals with a BP in Step 2a encounters more difficulties when the parameters of the regimes are close.

jth BP	1	2	3	4
λ_j	0.2025	0.4000	0.6775	0.8675
$\hat{\mu}(\hat{\lambda}_j)$	0.2023	0.3999	0.6836	0.8714
$\hat{\sigma}(\hat{\lambda}_j)$	0.0022	0.0044	0.0126	0.0104
MSE	5.1459e-6	1.9338e-5	2.0739e-4	1.4047e-4

Table 6: Estimated BPs in Step 2b (regimes with the same ARMA orders).

When $\hat{m} = 4$, our procedure detected the right model order (1,1) in 95% of the cases for all the regimes. Table 7 gives the model coefficient estimation results. We see that since the BP locations and the model orders are well recognized, the model coefficient estimation is quite precise.

4.3 Regimes with different ARMA orders

This simulation is designed to check the procedure's validity when the model order (p_j, q_j) changes with the regime. The orders and the parameters of the different regimes $X_{t,j}$ in (2) are given in Table 8 and Figure 3 displays a typical realization of this model.

We see in Table 9 that all criteria in Step 4 overestimate the number of BPs. The best criterion is C_2 with 88% of right decisions, and the worst is C_1 with 57% of right decisions.

Table 10 presents the BP estimation results in Step 2b. The first two BP estimations are

Model coefficient	Regime				
	1	2	3	4	5
d_j	0.15	0.40	0.20	0.35	0.20
$\hat{\mu}(\hat{d}_j)$	0.1466	0.4031	0.2031	0.3492	0.1737
$\hat{\sigma}(\hat{d}_j)$	0.0370	0.0376	0.0238	0.0303	0.0518
MSE	1.3804e-3	1.4209e-3	5.7697e-4	9.1525e-4	3.3736e-3
ϕ_j	0.80	0.80	-0.70	-0.30	-0.40
$\hat{\mu}(\hat{\phi}_j)$	0.8008	0.7999	-0.6947	-0.3043	-0.3936
$\hat{\sigma}(\hat{\phi}_j)$	0.0252	0.0208	0.0689	0.0255	0.0208
MSE	6.3652e-4	4.3185e-4	4.7669e-3	6.5256e-4	4.3151e-4
θ_j	-0.50	0.60	0.40	0.50	0.80
$\hat{\mu}(\hat{\phi}_j)$	-0.4977	0.6015	0.4001	0.5009	0.7792
$\hat{\sigma}(\hat{\phi}_j)$	0.0164	0.0338	0.0254	0.0371	0.0369
MSE	2.7363e-4	1.1436e-3	6.4195e-4	1.3768e-3	1.7877e-3

Table 7: Estimated parameters in Step 2c when $\hat{m} = 4$ and the right model orders are selected (regimes with the same ARMA orders).

Parameters \ Regime	1	2	3	4	5
p_j	1	2	1	1	0
q_j	2	1	1	0	1
d_j	0.20	0.45	0.15	0.35	0.10
ϕ_j	0.70	(-0.3, -0.5)	0.60	0.80	-
θ_j	(-0.70, -0.40)	0.6	-0.70	-	0.80

Table 8: Model parameters (regimes with different ARMA orders).

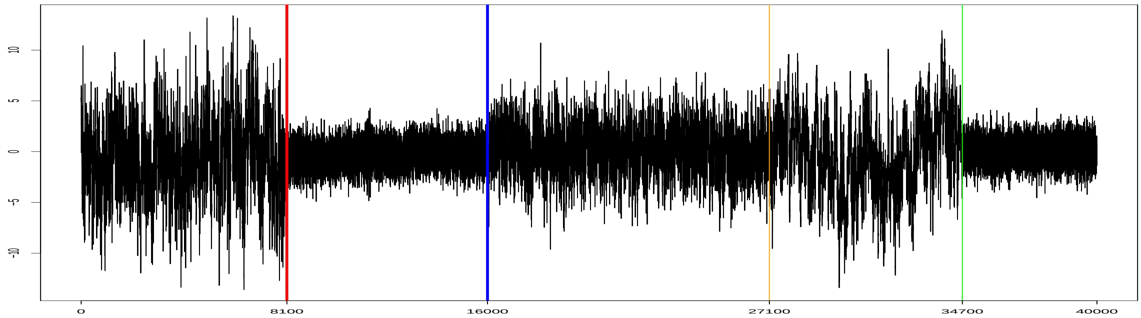

Figure 3: A realization of a piecewise FARIMA process whose regimes have different ARMA orders. The vertical lines indicate the true BPs locations.

Criterion \ $\hat{m}$	4	5	6	7	8	9
C_1	577	205	118	60	32	8
C_2	882	94	21	3	0	0
C_3	726	171	69	21	10	3
C_4	740	155	53	35	13	4

Table 9: Selected number of BPs in Step 4 (regimes with different ARMA orders).

less precise than the ones in Table 6 for the model whose regimes have the same ARMA orders. Nevertheless, the MSEs are quite small.

jth BP	1	2	3	4
λ_j	0.2025	0.4000	0.6775	0.8675
$\hat{\mu}(\hat{\lambda}_j)$	0.2016	0.3960	0.6716	0.8678
$\hat{\sigma}(\hat{\lambda}_j)$	0.0145	0.0244	0.0407	0.0041
MSE	2.1297e-4	6.0852e-4	1.6903e-3	1.7009e-5

Table 10: Estimated BPs in Step 2b (regimes with different ARMA orders).

Table 11 gives the most frequently selected orders in Step 2c for each stationary regime identified in Step 2b. We see that the true orders are well identified in at least 85% of the cases.

Table 12 displays the model coefficient estimation results for each regime when $\hat{m} = 4$ and the right model orders are selected for all regimes. As for the model regimes with the same ARMA orders, the estimation results are quite precise since the BP locations and the model

Order \ Regime	1	2	3	4	5
$(\hat{p}_j,\, \hat{q}_j)$	(1,2)	(2,1)	(1,1)	(1,0)	(0,1)
Frequency	972	866	919	889	929

Table 11: Most frequently selected orders in Step 2c (regimes with different ARMA orders).

orders are well identified.

Model coefficient	Regime				
	1	2	3	4	5
d_j	0.20	0.45	0.15	0.35	0.10
$\hat{\mu}(\hat{d}_j)$	0.1981	0.4404	0.1502	0.3539	0.0870
$\hat{\sigma}(\hat{d}_j)$	0.0362	0.0307	0.0357	0.0311	0.0356
MSE	1.3168e-3	1.0298e-3	1.2711e-3	9.6522e-4	1.4321e-3
ϕ_j	0.70	(-0.3, -0.5)	0.60	0.80	-
$\hat{\mu}(\hat{\phi}_j)$	0.6998	(-3004, -0.5007)	0.6010	0.7968	-
$\hat{\sigma}(\hat{\phi}_j)$	0.0302	(0.0122, 0.0113)	0.0304	0.0246	-
MSE	9.1087e-4	(1.4768e-4, 1.2739e-4)	9.2666e-4	6.1208e-4	-
θ_j	(-0.70, -0.40)	0.6	-0.70	-	0.80
$\hat{\mu}(\hat{\phi}_j)$	(-0.6982, -0.3985)	0.5889	-0.6972	-	0.7895
$\hat{\sigma}(\hat{\phi}_j)$	(0.0149, 0.0130)	0.0323	0.0200	-	0.0283
MSE	(2.2323e-4, 1.7146e-4)	1.1650e-3	4.0883e-4	-	9.1108e-4

Table 12: Estimated parameters in Step 2c when $\hat{m} = 4$ and the right model orders are selected (regimes with different ARMA orders).

4.4 Close BPs

Here we test the procedure when the interval between two BPs is only about $2E$. For the purpose of comparison, we consider a piecewise FARIMA$(1, d, 1)$ process with four BPs, the same break fractions and the same regime parameters as in Section 4.2.

Since here $K = 10$, it follows from assumption (A2) that the maximum possible number of BPs is four. In Step 4, all criteria find four BPs in more than 90% of the cases, but this may be due to the fact that in this case, no overestimation is possible.

Table 13 presents the BP estimation results in Step 2b. These results are just slightly less good than those in Table 6. This is probably because to minimize criterion (5), two benchmarks $\hat{\alpha}_p$ and $\hat{\alpha}_n$ are required and these benchmarks are estimated with fewer data in this simulation and are therefore more imprecise.

jth BP	1	2	3	4
λ_j	0.2025	0.4000	0.6775	0.8675
$\hat{\mu}(\hat{\lambda}_j)$	0.2028	0.4002	0.6855	0.8748
$\hat{\sigma}(\hat{\lambda}_j)$	0.0040	0.0065	0.0145	0.0135
MSE	1.6112e-5	4.4328e-5	2.72961e-4	2.3580e-4

Table 13: Estimated BPs in Step 2b (close BP model).

When $\hat{m} = 4$, the right model order (1,1) is detected in 90% of the cases for all the regimes. Table 14 presents the parameter estimation results. The bias, standard errors and MSEs are larger than those in Table 7. This is not surprising since QMLE estimates are based on fewer data in this case.

Model	Regime				
coefficient	1	2	3	4	5
d_j	0.15	0.40	0.20	0.35	0.20
$\hat{\mu}(\hat{d}_j)$	0.1446	0.3750	0.1953	0.3359	0.1408
$\hat{\sigma}(\hat{d}_j)$	0.0520	0.0569	0.0286	0.0391	0.0721
MSE	2.7251e-3	3.8492e-3	8.3896e-4	1.7232e-3	8.6888e-3
ϕ_j	0.80	0.80	-0.70	-0.30	-0.40
$\hat{\mu}(\hat{\phi}_j)$	0.8008	0.8050	-0.6983	-0.3091	-0.3878
$\hat{\sigma}(\hat{\phi}_j)$	0.0366	0.0312	0.0126	0.0333	0.0285
MSE	1.3373e-3	9.9668e-4	1.6105e-3	1.1858e-3	8.1263e-4
θ_j	-0.50	0.60	0.40	0.50	0.80
$\hat{\mu}(\hat{\phi}_j)$	-0.4937	0.5825	0.3916	0.4871	0.7555
$\hat{\sigma}(\hat{\phi}_j)$	0.0249	0.0504	0.0342	0.0488	0.05300
MSE	6.1991e-4	2.8492e-3	1.2372e-3	2.5427e-3	2.8090e-3

Table 14: Estimated parameters in Step 2c when $\hat{m} = 4$ and the right model orders are selected (closed BP model).

5 Application to traffic data

We consider the first OC48c Packet-over-SONET data set published by the NLANR MNA team. These 32000 data are the number of IP bytes collected at the Indianapolis router node on August 14, 2002, per 30 millisecond time intervals during 16 minutes. Figure 4 plots the series. We use $E = 2000$, which corresponds to an observation time of one minute for each elementary sub-series, and then $K = 16$. For each block we estimate the autocorrelations and the periodogram, and we find a similar behaviour for all blocks. Figures 5 and 6 plot the autocorrelations and the smoothed periodogram with Daniell filters of the first block. The autocorrelations decrease slowly and are significant for large lags (more than 500), and the periodogram blows up at the origin which are two important features of a LRD process. However, these features can also be present for a short memory process affected by a regime change or a smoothly varying trend, leading to so-called spurious long-memory. Therefore we use a test proposed by Qu (2009) to check whether this LRD behaviour is true or spurious. This test is in the frequency domain, and here we use trimming parameter $\varepsilon = 0.02$ and $m = [E^{0.67}]$ as recommended by Qu (2009) and Robinson (1995). For all elementary sub-series except one of them, this test does not reject the null hypothesis that the process is a long-memory process at the 1% confidence level. All these evidences suggest that the traffic data is LRD.

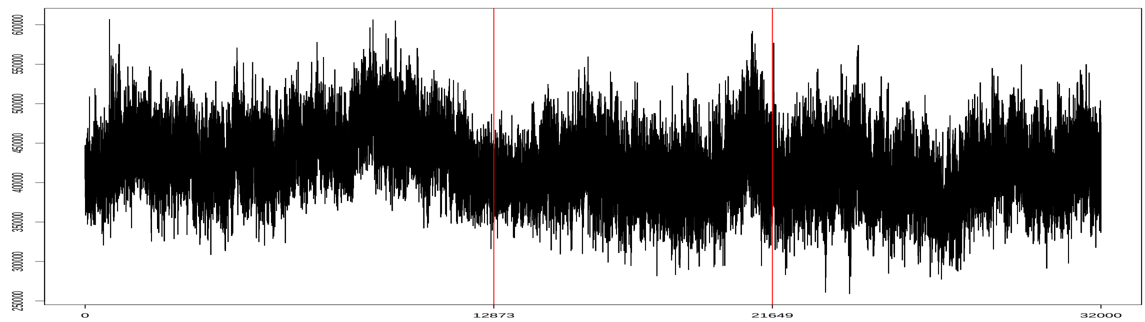

Figure 4: Internet traffic data. The vertical lines indicate the two estimated BP locations in Step 2b.

We fit a piecewise FARIMA model to the data. In our four-step procedure, we take the same parameters a_j, b_j, η, C_n, c_0 and γ_0 as in Section 4, and the maximum value of the ARMA orders considered in BIC is 7. In Table 15, we present the values of the different criteria in Step 4 for $m = 0, \ldots, 7$. All criteria are minimum when $m = 2$.

The two estimated BPs are indicated by vertical lines in Figure 4. For the first regime, a FARIMA$(2, d, 3)$ model with $d = 0.4829$ is selected, for the second regime, a FARIMA$(4, d, 2)$ model where $d = 0.3463$ is chosen, and for the last regime a FARIMA$(2, d, 5)$ model with

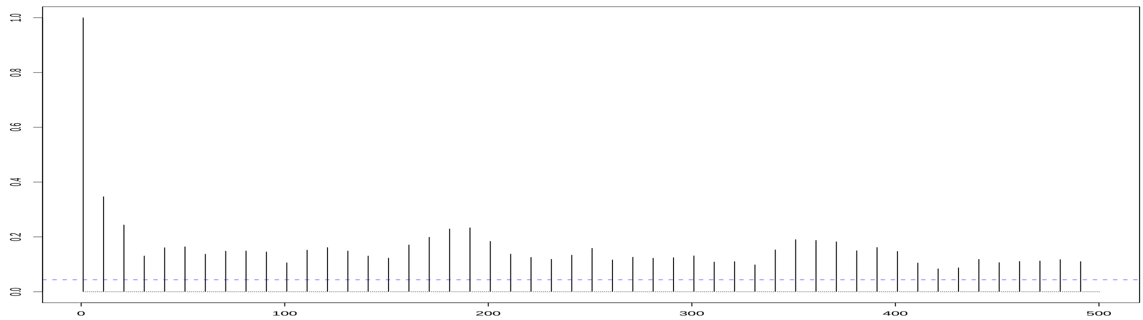

Figure 5: Auto-correlations of the internet traffic data.

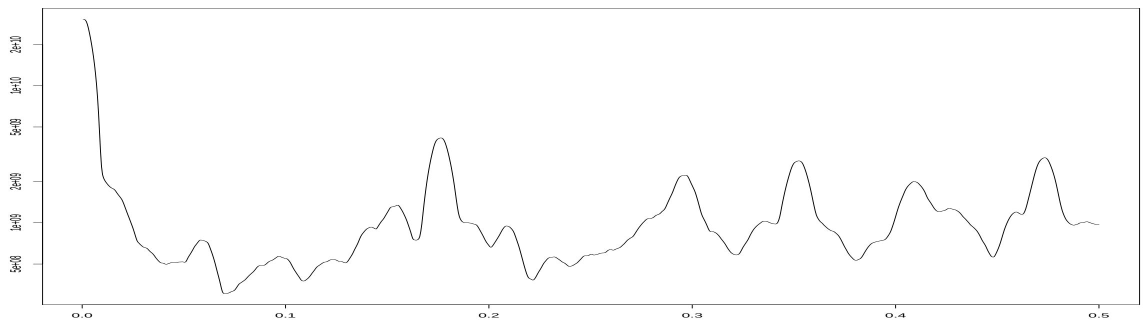

Figure 6: Smoothed periodogram of the internet traffic data.

Criterion \ m	0	1	2	3
C_1	20.9439	20.8159	**20.7972**	20.7993
C_2	20.9428	20.8154	**20.8008**	20.8058
C_3	20.9454	20.8246	**20.8090**	20.8151
C_4	3.785509e5	3.781614e5	**3.781608e5**	3.78258e5

Criterion \ m	4	5	6	7
C_1	20.8090	20.8103	20.8352	20.8131
C_2	20.8190	20.8236	20.8512	20.8311
C_3	20.8284	20.8333	20.8628	20.8464
C_4	3.782925e5	3.78323e5	3.78445e5	3.78414e5

Table 15: BP number selection (internet traffic data).

$d = 0.4622$ is preferred. The corresponding parameter estimates, as well as the local estimates obtained in Step 1 are shown in Figure 7. The dotted lines fluctuate around the solid lines.

6 Conclusion

In this article, we have proposed a piecewise FARIMA model and a methodology for fitting it to a local stationary long-memory time series. This model is able to capture the structural break properties of the series, it is flexible and allows to model simultaneously long and short range dependence. The model fitting consists in a four-step procedure designed to estimate both the BPs and the parameters. Simulations have shown good performances of the method. When applying our methodology to internet traffic data recorded during 16 minutes and sampled every 30 milliseconds, a piecewise FARIMA model with two BPs was selected.

Acknowledgements

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

References

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. Critical values for multiple structural change tests. *Econom. J.*, 6(1): 72–78, 2003.

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. L. Brown, J. Durbin, and J. M. Evans. Techniques for testing the constancy of regression relationships over time. *J. Roy. Statist. Soc. Ser. B*, 37:149–192, 1975.

H. Chang, M. Devetsikiotis, I. Lambadaris, and A. R. Kaye. Modeling and simulation of self-similar variable bit rate compressed video: a unified approach. *SIGCOMM Comput. Commun. Rev.*, 25(4):114–125, 1995.

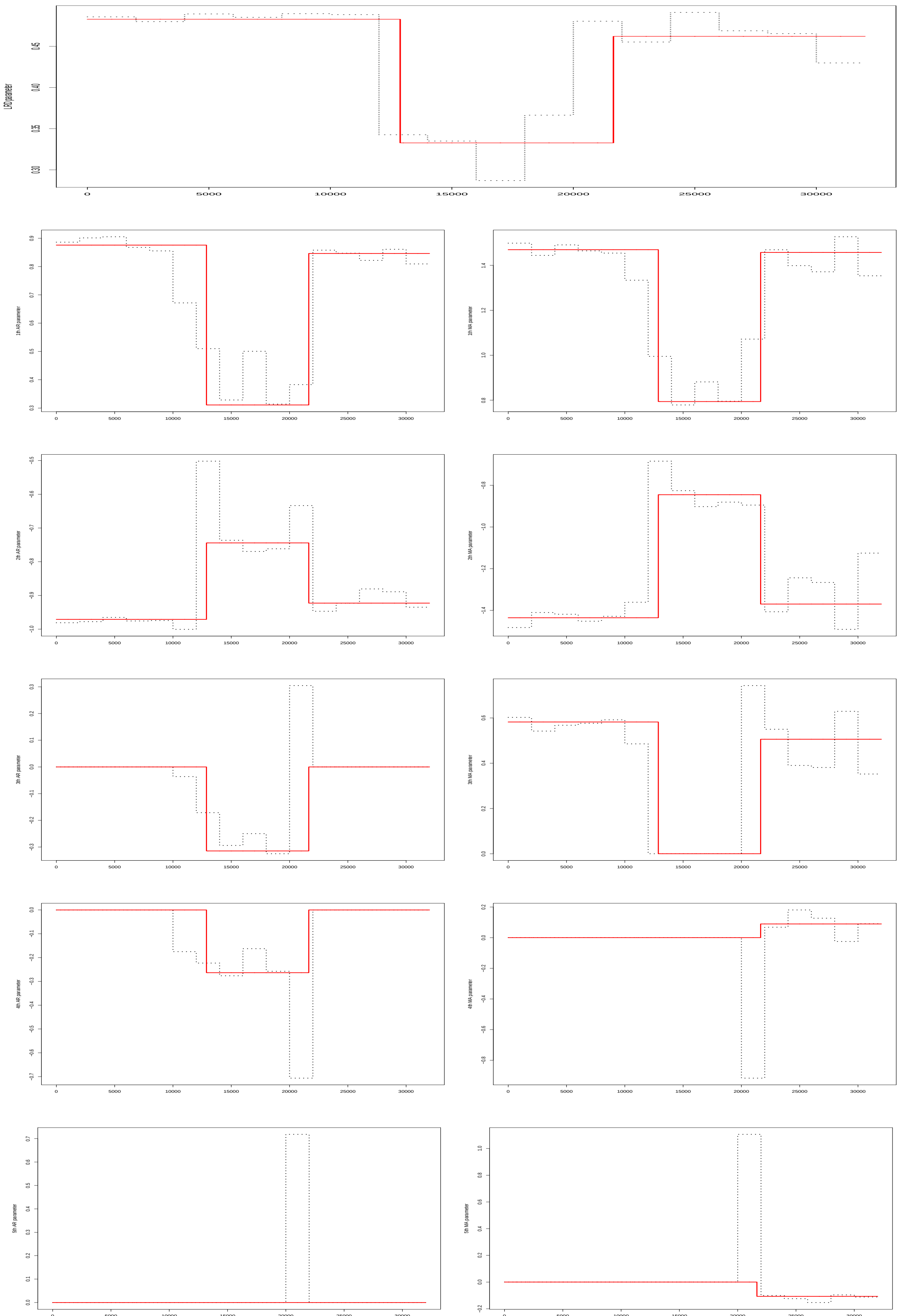

Figure 7: Estimated parameters for the Internet traffic data. Parameter estimates in Step 2c (solid line) and local parameter estimates in Step 1 (dotted lines).

M. Coulon and A. Swami. Least squares detection of multiple changes in fractional ARIMA processes. In *Proceedings of the International Conference on Acoustic, Speech and Signal*, volume 5, pages 3177–3180, 2001.

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.

J. Eran, R. Sherman, M. S. Taqqu, and W. Willinger. Long-range dependence in variable-bit-rate video traffic. *IEEE Trans. Commun*, 43:1566–1579, 1995.

M. W. Garrett and W. Willinger. Analysis, modeling and generation of self-similar VBR video traffic. *ACM SIGCOMM Computer Communication*, 24(4):269–280, 1994.

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

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

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

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.

W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson. On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Trans. Netw.*, 2(1):1–15, 1994.

J. Liu, S. Wu, and J. V. Zidek. On segmented multivariate regressions. *Statistica Sinica*, 7(2): 497–525, 1997.

B. A. Maguire, E. S. Pearson, and A. H. A. Wynn. The time intervals between industrial accidents. *Biometrika*, 39(1-2):168–180, 1952.

I. Norros. A storage model with self-similar input. *Queueing Systems Theory Appl.*, 16(3-4): 387–396, 1994.

L. C. Nunes, C. M. Kuan, and P. Newbold. Spurious break. *Economics Letters*, 50(2):736–749, 1996.

E. S. Page. A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42: 523–527, 1955.

V. Paxson and S. Floyd. Wide-area traffic: the failure of Poisson modeling. In *Proceedings of the Conference on Communications Architectures, Protocols and Applications*, pages 257–268, New York, NY, USA, 1994. ACM Press.

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

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

Z. Qu. A test against spurious long memory. Unpublished paper. Department of Economics, Boston University, USA, April 2009.

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.

P. M. Robinson. Gaussian semiparametric estimation of long range dependence. *Ann. Statist.*, 23(5):1630–1661, 1995.

G. Schwarz. Estimating the dimension of a model. *Ann. Statist.*, 6(2):461–464, 1978.

S. Stoev and M. S. Taqqu. Asymptotic self-similarity and wavelet estimation for long-range dependent fractional autoregressive integrated moving average time series with stable innovations. *J. Time Ser. Anal.*, 26(2):211–249, 2005.

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.

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.

Y.-C Yao. Estimating the number of change-points via Schwarz' criterion. *Statistics & Probability Letters*, 6(3):181–189, 1988.

Y-C. Yao and S. T. Au. Least-squares estimation of a step function. *Sankhyā Ser. A*, 51(3): 370–381, 1989.