Time Series Segmentation Procedures to Detect, Locate and Estimate Change-Points

Ana Laura Badagián, Regina Kaiser, and Daniel Peña

Abstract This article deals with the problem of detecting, locating, and estimating the change-points in a time series process. We are interested in finding changes in the mean and the autoregressive coefficients in piecewise autoregressive processes, as well as changes in the variance of the innovations. With this objective, we propose an approach based on the Bayesian information criterion (BIC) and binary segmentation. The proposed procedure is compared with several others available in the literature which are based on cusum methods (Inclán and Tiao, J Am Stat Assoc 89(427):913–923, 1994), minimum description length principle (Davis et al., J Am Stat Assoc 101(473):229–239, 2006), and the time varying spectrum (Ombao et al., Ann Inst Stat Math 54(1):171–200, 2002). We computed the empirical size and power properties of the available procedures in several Monte Carlo scenarios and also compared their performance in a speech recognition dataset.

1 Introduction

In this article we consider the problem of modelling a nonstationary time series by segmenting it into blocks which are fitted by stationary processes. The segmentation aims to: (1) find the periods of stability and homogeneity in the behavior of the process; (2) identify the moments of change, called change-points; (3) represent the regularities and features of each piece; and (4) use this information in order to determine the pattern in the nonstationary time series.

Time series segmentation and change-point detection and location has many applications in several disciplines, such as neurology, cardiology, speech recognition, finance, and others. Consider questions like: What are the main features of the brain activity when an epileptic patient suffers a seizure? Is the heart rate variability reduced after ischemic stroke? What are the most useful phonetic features to recognizing speech data? Is the conditional volatility of the financial assets constant? These questions can often be answered by performing segmentation analysis. The

A.L. Badagián (✉) • R. Kaiser • D. Peña
Statistics Department, Universidad Carlos III de Madrid, Madrid, Spain
e-mail: abadagia@est-econ.uc3m.es; kaiser@est-econ.uc3m.es; dpena@est-econ.uc3m.es

J. Beran et al. (eds.), *Empirical Economic and Financial Research*,
Advanced Studies in Theoretical and Applied Econometrics 48,
DOI 10.1007/978-3-319-03122-4_3

reason is that many series in these fields do not behave as stationary, but can be represented by approximately stationary intervals or pieces.

Segmentation analysis aims to answer the following questions: Did a change occur? When did the changes occur? If more than one change occurs, how can we locate them? Whereas the first two questions refer to the problem of defining a statistical criteria for detecting, estimating, and locating a change-point, the last one is related with the difficult task of creating a strategy, implemented in an algorithm, in order to search for multiple change-points.

When multiple change-points are expected, as its number and location are usually unknown, the multiple searching issue is very intricate. It is a challenge to jointly estimate the number of structural breaks and their location, and also provide a estimation of the model representing each interval. This problem has received considerably less attention than the detection and estimation of a single change-point, due to the difficulty in handling the computations. Many algorithms exist to calculate the optimal number and location of the change-points, some of them were presented by Scott and Knott (1974), Inclán and Tiao (1994), Auger and Lawrence (1989), Jackson et al. (2005), and Davis et al. (2006)

The main contributions of this paper are: (a) proposing a procedure based on the BIC joint with the binary segmentation algorithm to look for changes in the mean, the autoregressive coefficients, and the variance of perturbation in piecewise autoregressive processes, by using a procedure; (b) comparing this procedure with several others available in the literature, which are based on cusum methods (Inclán and Tiao 1994; Lee et al. 2003), minimum description length (MDL) principle (Davis et al. 2006), and the time varying spectrum (Ombao et al. 2002). For that, we compute the empirical size and the power properties in several scenarios and we apply them to a speech recognition dataset.

The article is organized as follows. In Sect. 2 we present the change-point problem. Following, in Sect. 3, we briefly present cusum methods, Auto-PARM and Auto-SLEX procedures. The final part of this section is dedicated to the informational approach procedures and the proposed procedure based on BIC is presented. Section 4 presents different algorithms that are useful to search for multiple change-points. In Sect. 5 we compute and compare the size and the power of the presented approaches. In Sect. 6 they are applied to real data of speech recognition, and finally, the final section presents the conclusions.

2 The Change-Point Problem

The problem we deal is the following. Suppose that $x_1, x_2, \ldots, x_T$ is a time series process with m change-points at the moments $k_1^*, \ldots, k_m^*$, with $1 \le k_1^* \le \ldots \le k_m^* \le T$. The density function $f\ (x_t/\theta)$, with θ the vector of parameters, is assumed to be

$$f\left(x_t/\theta\right) = \begin{cases} f\left(x_t/\theta_1\right), & t = 1, \ldots, k_1^*, \\ f\left(x_t/\theta_2\right), & t = k_1^* + 1, \ldots, k_2^*, \\ \cdot & \\ \cdot & \\ \cdot & \\ f\left(x_t/\theta_{m+1}\right), & t = k_m^* + 1, \ldots, T. \end{cases} \quad \text{for } \theta_1 \neq \theta_2 \neq \ldots \neq \theta_{m+1}.$$

The values of $\theta_i,\ i = 1, 2, \ldots, m+1$ can be a priori known or unknown and the goal is to detect and locate $k_1^*, k_2^*, \ldots, k_m^*$, and also estimate θ_i's when they are unknown.

Then, in general, the change-point problem consists of testing

$$\begin{aligned} &H_0 : x_t \sim f\left(x_t/\theta\right), t = 1, \ldots, T \\ &H_1 : x_t \sim f\left(x_t/\theta_1\right), t = 1, \ldots, k_1^*,\ x_t \sim f\left(x_t/\theta_2\right), t = k_1^* + 1, \ldots, k_2^*, \ldots \\ &\qquad \ldots, x_t \sim f\left(x_t/\theta_{m+1}\right), t = k_m^* + 1, \ldots, T, \text{for } \theta_1 \neq \theta_2 \neq \ldots \neq \theta_{m+1}. \end{aligned} \tag{1}$$

If the distributions $f\left(x_t/\theta_1\right),\ f\left(x_t/\theta_2\right), \ldots, f\left(x_t/\theta_m + 1\right)$ belong to a common parametric family, then the change-point problem in (1) is equivalent to test the null hypothesis:

$$\begin{aligned} H_0 : &\quad \theta_1 = \theta_2 = \ldots = \theta_{m+1} = \theta \\ H_1 : &\quad \theta_1 = \ldots = \theta_{k_1^*} \neq \theta_{k_1^*+1} = \ldots = \theta_{k_2^*} \neq \ldots \\ &\quad \ldots \neq \theta_{k_{m-1}+1} = \ldots = \theta_{k_m} \neq \theta_{k_m+1} = \ldots = \theta_T. \end{aligned} \tag{2}$$

Most of the parametric methods proposed in the literature for change-point problems consider a normal model. If the density function is constant over time, the change-point problem consists on testing whether the mean or the variance registered a change over the period analyzed.

3 Segmentation Procedures to Detect, Locate, and Estimate Change-Points

There are many approaches for solving the problem of detecting, estimating, and locating a change-point for independent or linear autocorrelated random variables that can be based on parametric (Chen and Gupta 2001, 2011) and non-parametric methods (Brodsky and Darkhovsky 1993; Heiler 1999, 2001). The main idea consists of minimizing a loss function which involves some criteria or statistic selected to measure the goodness of the segmentation performed. The computation of those statistics is useful to detect a potential change-point, by comparing the corresponding statistic computed under the hypothesis of no changes with the one assuming a change-point at the most likely period (Kitagawa and Gersch 1996; Chen and Gupta 1997; Al Ibrahim et al. 2003; Davis et al. 2006).

3.1 Cusum Methods

One of the statistics most often used to segment a time series is the cumulative sum or cusum (Page 1954). In fact, many procedures for change-point detection are based on cusum statistics (Inclán and Tiao 1994; Lee et al. 2003; Kokoszka and Leipus 1999; Lee et al. 2004 among others). The procedure in Inclán and Tiao (1994) is useful to test the null hypothesis of constant unconditional variance of a Gaussian uncorrelated process x_t, against the alternative of multiple change-points. The test statistic is defined as:

$$IT = \sqrt{T/2}\,\max_k D_k \tag{3}$$

where

$$D_k = \frac{\sum_{t=1}^{k} x_t^2}{\sum_{t=1}^{T} x_t^2} - \frac{k}{T}, \tag{4}$$

with $0 < k < T$. The asymptotic distribution of the statistic IT is the maximum of a Brownian bridge ($B(k)$):

$$IT \rightarrow_{D[0,1]} \max\{B(k) : k \in [0, 1]\}$$

This establishes a Kolmogorov–Smirnov type asymptotic distribution. The null hypothesis is rejected when the maximum value of the function IT is greater than the critical value and the change-point is located at period $k = \hat{k}$ where the maximum is achieved:

$$\hat{k} = \{k : IT > \text{c.v.}\}.$$

where c.v. is the corresponding critical value.

3.2 Automatic Procedure Based on Parametric Autoregressive Model (Auto-PARM)

In Davis et al. (2006) an automatic procedure called Auto-PARM is proposed for modelling a nonstationary time series by segmenting the series into blocks of different autoregressive processes.

Let k_j the breakpoint between the j-th and the $(j+1)$-st AR processes, with $j = 1, \ldots, m$, $k_0 = 1$ and $k_m < T$. Thus, the j-th piece of the series is modelled as:

$$X_t = x_{t,j}, \qquad k_{j-1} \leq t < k_j, \tag{5}$$

where $\{x_{t,j}\}$ is an AR(p_j) process.

$$x_{t,j} = \gamma_j + \phi_{j1} x_{t-1,j} + \ldots + \phi_{j,p_j} x_{t-p_j,j} + \sigma_j \epsilon_t,$$

where $\boldsymbol{\theta}_j := \left(\gamma_j, \phi_{j1}, \ldots, \phi_{j,p_j}, \sigma_j^2\right)$ is the parameter vector corresponding to this AR(p_j) process and the sequence $\{\epsilon_t\}$ is iid with mean 0 and variance 1. This model assumes that the behavior of the time series is changing at various times. Such a change might be a shift in the mean, a change in the variance, and/or a change in the dependence structure of the process.

The best segmentation is the one that makes the maximum compression of the data possible measured by the MDL principle of Rissanen (1989). *MDL* is defined as[1]:

$$MDL\,(m, k_1, \ldots, k_m, p_1, \ldots, p_{m+1}) = \tag{6}$$
$$\log m + (m+1)\log T + \sum_{j=1}^{m+1} \log p_j + \sum_{j=1}^{m+1} \frac{p_j + 2}{2} \log T_j + \sum_{j=1}^{m+1} \frac{T_j}{2} \log\left(2\pi\hat{\sigma}_j^2\right).$$

where m is the number of change-points located at $k_1, k_2, \ldots, k_m$, T_j is the number of observation in each segment j, p_j is the order of the autoregressive model fitted to the segment j, and $\hat{\sigma}_j^2$ is the Yule Walker estimator of σ_j^2 (Brockwell and Davis 1991).

3.3 Automatic Procedure Based on Smooth Localized Complex EXponentials (Auto-SLEX) Functions

In Adak (1998), Donoho et al. (1998), Ombao et al. (2002), and Maharaj and Alonso (2007) the segmentation is performed by using a cost function based on the spectrum, called evolutionary spectrum, because the calculation is made by the spectrum of each stationary interval. Ombao et al. (2002) created SLEX vectors which are calculated by applying a projection operator on the Fourier vectors, to get a basis which is simultaneously orthogonal and localized in time and frequency and is useful to compute the spectrum of nonstationary time series.

[1]For more details see Davis et al. (2006).

The cost function of the block $S_j = [k_j, k_{j+1}]$ is given by

$$\text{Cost}\,(S) = \sum_{S_j} \log \hat{\alpha}_{S_j} + \beta \sqrt{m_j}, \tag{7}$$

where $\hat{\alpha}_{S_j}$ is the SLEX periodogram, β is a penalty parameter generally equal to 1 (Donoho et al. 1998), and m_j is the number of breaks in the block. The cost for a particular segmentation of the time series is the sum of the costs at all the blocks defining that segmentation. The best segmentation is the one having the smallest cost.

3.4 Informational Approach

Information criteria, which commonly are useful as a measure of goodness of fit of a model, can be used to detect and estimate change-points. The first and most popular of the information criteria is the Akaike information criterion (AIC), which was introduced in 1973 for model selection in statistics. This criterion has found many applications in time series, outliers detection, robustness and regression analysis. AIC is defined as:

$$\text{AIC} = T \log \hat{\sigma}^2_{MV} + 2p.$$

where $\hat{\sigma}^2_{MV}$ is the maximum likelihood estimator of σ^2, and p is the number of free parameters. A model that minimizes the AIC is considered the appropriate model. The limitation of the minimum estimated AIC is that it is not an asymptotically consistent estimator of the model order (Schwarz 1978).

Another information criterion was introduced by Schwarz in 1978, and commonly is referred to as BIC or SIC. The fundamental difference with the AIC is the penalization function, which penalizes more the number of model parameters and leads to an asymptotically consistent estimate of the order of the true model. BIC is defined as

$$\text{BIC} = T \log \hat{\sigma}^2_{MV} + p \log T,$$

where $\hat{\sigma}^2_{MV}$ is the maximum likelihood estimator of σ^2, p is the number of free parameters, and T is the length of the time series. In this setting, we have two models corresponding to the null and the alternative hypotheses.

Let $\text{BIC}_0(T)$ the BIC under H_0 in (2) where no changes occur in the process along whole the sample and $\text{BIC}_1(k)$ the criterion assuming that there is a change-point at $t = k$, where k could be, in principle, $1, 2, \ldots, T$.

The rejection of H_0 is based on the principle of minimum information criterion. That is, we do not reject H_0 if $\text{BIC}_0(T) < \min_k \text{BIC}_1(k)$, because the BIC computed assuming no changes is smaller than the BIC calculated supposing the existence of a change-point at the most likely k, that is, in the value of k where the minimum BIC is achieved. On the other hand, H_0 is rejected if $\mathrm{BIC}_0(T) > \mathrm{BIC}_1(k)$ for some k and estimate the position of the change-point k^* by $\hat{k}$ such that

$$\mathrm{BIC}(\hat{k}) = \min_{2<k<T} \mathrm{BIC}_1(k).$$

In Chen and Gupta (1997) a procedure which combine BIC and the binary segmentation is proposed[2] to test for multiple change-points in the marginal variance, assuming independent observations. In this article BIC is used for locating the number of breaks in the variance of stock returns. Liu et al. (1997) modified the BIC by adding a larger penalty function and Bai and Perron (1998) considered criteria based on squared residuals. In the following section we present the approach of Chen and Gupta (1997) for testing a single change-point in the variance of independent normal data. In Al Ibrahim et al. (2003) the BIC is used to detect change-points in the mean and autoregressive coefficients of an AR(1).

3.5 A Proposed Procedure to Detect Changes in Mean, Variance, and Autoregressive Coefficients in AR Models

In this section, we propose an informational approach procedure for detecting changes in mean, variance, and autoregressive coefficients for AR(p) processes. Let $x_1, x_2, \ldots, x_T$ be the T consecutive observations from a Gaussian autoregressive process of order p given by:

$$x_t = \begin{cases} c_1 + \phi_{11}x_{t-1} + \ldots + \phi_{1p}x_{t-p} + \sigma_1\epsilon_t, & -\infty < t \le k_1 \\ c_2 + \phi_{21}x_{t-1} + +\phi_{2p}x_{t-p} + \sigma_2\epsilon_t, & k_1 < t \le k_2 \\ \cdot & \\ \cdot & \\ \cdot & \\ c_m + \phi_{m1}x_{t-1} + \ldots + \phi_{mp}x_{t-p} + \sigma_m\epsilon_t, & k_{m-1} < t \le k_m \\ c_{m+1} + \phi_{m+1,1}x_{t-1} + \ldots + \phi_{m+1,p}x_{t-p} + \sigma_{m+1}\epsilon_t, & k_m < t \le \infty \end{cases} \tag{8}$$

The null hypothesis is that

$$H_0 : c_1 = \ldots = c_{m+1}, \;\; \phi_{11} = \ldots = \phi_{m+1,1}, \;\; \phi_{1p} = \ldots = \phi_{m+1,p} \;\text{ and}$$
$$\sigma_1^2 = \ldots = \sigma_{m+1}^2.$$

Under the null hypothesis, the formula for the BIC, denoted as $\mathrm{BIC}_0(T)$, is given by:

[2]Binary segmentation is a searching procedure in order to detect multiple change-points in one time series. We will explain it in Sect. 4.

$$BIC_0(T) = (T-p)\,\hat{\sigma}_0^2 + (p+2)\log(T-p)\,, \tag{9}$$

where $\hat{\sigma}_0^2 = \frac{1}{T-p}\sum_{t=p+1}^{T}\left(x_t - \hat{c}_1 - \hat{\phi}_1 x_{t-1} - \ldots - \hat{\phi}_p x_{t-p}\right)^2$, $\hat{c}, \hat{\phi}_1, \ldots, \hat{\phi}_p$ are the conditional maximum likelihood estimators of σ^2, c_1, and the autoregressive parameters, respectively.

The $\mathrm{BIC}_1(k)$ for the piecewise AR(p) model under the alternative hypothesis is given by:

$$\mathrm{BIC}_1(k) = (k_1 - 1)\log\hat{\sigma}_1^2 + \ldots + (T - k_m)\log\hat{\sigma}_{m+1}^2 + (m+1)(p+2)\log T. \tag{10}$$

where $\hat{\sigma}_1^2 = \frac{1}{k_1-1}\sum_{t=2}^{k_1}(x_t - \tilde{c}_1 - \tilde{\phi}_{11}x_{t-1} - \ldots - \tilde{\phi}_{1p}x_{t-p})^2, \ldots, \hat{\sigma}_{m+1}^2 = \frac{1}{T-k_m}\sum_{t=k_m+1}^{T}(x_t - \tilde{c}_{m+1} - \tilde{\phi}_{m+1,1}x_{t-1} - \ldots - \tilde{\phi}_{m+1,p}x_{t-p})^2$, $\tilde{c}_1,\ldots,\tilde{c}_{m+1}$, $\tilde{\phi}_{11},\ldots,\tilde{\phi}_{m+1,p}$ are the conditional maximum likelihood estimators of the variances, $\sigma_1^2,\ldots,\sigma_{m+1}^2$, the constants, $c_1,\ldots,\ c_{m+1}$ and the autoregressive parameters, $\phi_{11},\ldots,\phi_{m+1,p}$, respectively.

H_0 is not rejected if $\mathrm{BIC}_0(T) < \min_k \mathrm{BIC}_1(k) + c_\alpha$, where c_α, and α have the relationship $1-\alpha = P[\mathrm{BIC}_0(T) < \min_k \mathrm{BIC}_1(k) + c_\alpha / H_0]$.

4 Multiple Change-Point Problem

When multiple change-points are expected, as its number and location are usually unknown, it is a challenge to jointly estimate the number of structural breaks, their location, and also provide a estimation of the model representing each interval. Many algorithms exist to calculate the optimal number and location of the change-points, some of them were presented by Scott and Knott (1974), Inclán and Tiao (1994), Davis et al. (2006), and Stoffer et al. (2002).

Binary segmentation (Scott and Knott 1974; Sen and Srivastava 1975; Vostrikova 1981) addresses the issue of multiple change-points detection as an extension of the single change-point problem. The segmentation procedure sequentially or iteratively applies the single change-point detection procedure, i.e. it applies the test to the total sample of observations, and if a break is detected, the sample is then segmented into two sub-samples and the test is reapplied. This procedure continues until no further change-points are found. This simple method can consistently estimate the number of breaks (e.g., Bai 1997; Inclán and Tiao 1994) and is computationally efficient, resulting in an $O(T\log T)$ calculation (Killick et al. 2012). In practice, binary segmentation becomes less accurate with either small changes or changes that are very close on time. Inclán and Tiao (1994) applied a such of modified binary segmentation in its Iterative Cusums of Square (ICSS) algorithm, by sequentially applying the statistic IT presented in Sect. 3.1.

In Davis et al. (2006) a genetic algorithm is used for detecting the optimal number and location of multiple change-points by minimizing the MDL. These algorithms make a population of individuals or chromosomes “to evolve” subject to random actions similar to those that characterize the biologic evolution (i.e., crossover and genetic mutation), as well as a selection process following a certain criteria which determines the most adapted (or best) individuals who survive the process, and the less adapted (or the “worst” ones), who are ruled out. In general, usual methods for applying genetic algorithm encode each parameter using binary coding or gray coding. Parameters are concatenated together in a vector to create a chromosome which evolve to a solution of the optimization problem.

Finally, other algorithms set a priori the segmentation structure. For instance, some procedures perform a dyadic segmentation to detect multiple change-points. Under this structure, time series can be divided into a number of blocks which are a power of 2. The algorithm begins setting the smallest possible size of the segmented blocks or the maximum number of blocks. Ideally, the block size should be small enough so that one can ensure the stationary behavior, but not too small to guarantee good properties of the estimates. Stoffer et al. (2002) recommended a block size greater or equal than 2^8. Then, the following step is to segment the time series in $2^8, 2^7, \ldots, 2^1, 2^0$ blocks, which is equivalent to consider different resolution levels $j = 8, 7, \ldots, 1, 0$, respectively. At each level j, we compare a well-defined cost function computed in that level j (father block) with respect to that computed in the level $j - 1$ (two children blocks). The best segmentation is that which minimizes the cost function.

Some papers focusing on multiple change-point problem for autocorrelated data are Andreou and Ghysels (2002) and Al Ibrahim et al. (2003). In Andreou and Ghysels (2002) an algorithm similar to ICSS (Inclán and Tiao 1994) is applied to detect multiple change-points in financial time series using cusum methods. In the first step the statistic is applied to the total sample and if a change-point is detected, the sampled is segmented and the test is applied again to each subsample up to five segments. Other algorithms are applied in this paper, using a grid search approach or methods based on dynamic programming. In Al Ibrahim et al. (2003) the binary segmentation algorithm combined with the BIC is used for piecewise autoregressive models.

Given the merits of binary segmentation saving a lot of computational time and the better performance with respect to ICSS algorithm, in order to design the simulation experiments, and, for empirical applications below, we propose to combine the BIC statistic assuming the model in Eq. (8) with binary segmentation (referred as BICBS).

5 Monte Carlo Simulation Experiments

In this section we evaluate the performance of the methods presented above, by computing the empirical size and power under different hypotheses. We have used four methods: ICSS (Inclán and Tiao 1994), BICBS (BIC for model in (8) with binary segmentation), Auto-PARM (Davis et al. 2006), and Auto-SLEX (Ombao et al. 2002). In the tables below, where these procedures are compared, the results for BICBS, which is the proposed procedure, are highlighted with bold font.

5.1 Empirical Size

First, we compute the empirical size, that is, how many times the corresponding methodology incorrectly segments a stationary process. The length of the simulated series is set equal to 4,096. Table 1 presents the results for 1,000 replications for a Gaussian white noise with unitary variance, and for AR(1) and MA(1) stationary processes.

All the procedures analyzed seems to appear undersized in finite samples. Applying them to stationary processes we obtain only one block or segment in most of the cases, and only a very small percentage of processes are segmented in two blocks. For example, for ICSS, BICBS, and Auto-PARM the rate of wrong segmented stationary processes is almost zero. The hypothesis that the type of autocorrelation (i.e., autoregressive and moving average) could influence the segmentation is rejected, given that the results for MA(1) and AR(1) processes are similar leading to the conclusion that the type of serial correlation seems to be not important for the size of these procedures.

5.2 Power for Piecewise Stationary Processes

We compute the power of the methods by counting how many times the corresponding methodology correctly segments piecewise stationary processes in 1,000 replications. Two stationary segments or blocks are assumed. We observe if the procedure finds the correct number of segments or blocks and if the changes occur in a narrow interval centered on the correct breakpoint ($k^* \pm 100$). For a time series of length $T = 4096$, we evaluate the performance of the procedures when the data present serial correlation and the perturbation's variance changes. The simulated process is an AR(1) with autoregressive parameter $\phi \in (-1, 1)$ changing the perturbation variance from 1 to 2 in $k^* = 2048$.

Table 1 Size of ICSS, BICBS, Auto-PARM, and Auto-SLEX

Processes	ICSS	**BICBS**	Auto-PARM	Auto-SLEX
White noise	0.000	**0.04**	0.000	0.000
AR(1) $\phi \in (-1, 1)$	0.000	**0.000**	0.005	0.025
MA(1) $\theta \in (-1, 1)$	0.000	**0.000**	0.001	0.011

Table 2 Power of the procedures segmenting piecewise autoregressive processes with $\phi \in (-1, 1)$, where the perturbation's variance changes from 1 to 2 in $t = 2048$

Processes	ICSS	**BICBS**	Auto-PARM	Auto-SLEX
Precise detection	0.951	**0.960**	0.961	0.923
Oversegmentation	0.001	**0.040**	0.039	0.077
No segmentation	0.048	**0.000**	0.000	0.000

In Table 2 we present the results, where the autoregressive coefficient is generated as $\phi \in (-1, 1)$, and the perturbation term is a white noise with unitary variance in the first piece ($t = 1, \ldots, 2048$), shifting to 2 in the second piece ($t = 2049, \ldots, 4096$).

All the procedures obtained excellent results when the perturbation's term variance changes, where the best results were for Auto-PARM and BICBS.

Finally, we analyze the performance of the tests detecting multiple change-points in three processes. The first one is given by:

$$x_t = \begin{cases} \epsilon_t, & 1 < t \leq 1365 \\ 2\epsilon_t, & 1366 < t \leq 2730 \\ 0.5\epsilon_t, & 2731 < t \leq 4096, \end{cases} \tag{11}$$

where we are interested in changes in the scale of the perturbation term, when the process does not have autocorrelation. The second is:

$$x_t = \begin{cases} 0.5x_{t-1} + \epsilon_t, & 1 < t \leq 1365 \\ 0.8x_{t-1} + \epsilon_t, & 1366 < t \leq 2730 \\ -0.5x_{t-1} + \epsilon_t, & 2731 < t \leq 4096, \end{cases} \tag{12}$$

where it is introduced first order autocorrelation in the process and the change-points are due to the autoregressive coefficient. The third process is given by:

$$x_t = \begin{cases} 0.5x_{t-1} + \epsilon_t, & 1 < t \leq 1365 \\ 0.8x_{t-1} + \epsilon_t, & 1366 < t \leq 2730 \\ 0.8x_{t-1} + 2\epsilon_t, & 2731 < t \leq 4096, \end{cases} \tag{13}$$

where also is introduced autocorrelation in the data and there is both a change-point in the autoregressive coefficient and another one in the variance of the perturbation. It is assumed that $\epsilon_t \sim$ N(0,1) and $x_0 = 0$. The results are presented in Table 3.

When multiple change-points are present in the time series, some procedures performed well only if the data have no serial correlation [process (11)]. That is the case of ICSS, BICBS, and Auto-PARM. Auto-SLEX detected the change-point, but with a big rate of oversegmentation. For autocorrelated data, the procedures with the best performance were BICBS and Auto-PARM, with powers greater than 0.91. ICSS has smaller power and often it does not segment or only finds one of the two

Table 3 Proportion of detected change-points in piecewise stationary processes with two changes presented in Eqs. (11)–(13)

	ICSS	**BICBS**	Auto-PARM	Auto-SLEX
Process with no autocorrelation as in (11)				
Precise detection	0.999	**0.910**	1.000	0.626
One change-point	0.000	**0.000**	0.000	0.000
Oversegmentation	0.000	**0.005**	0.000	0.372
No segmentation	0.001	**0.085**	0.000	0.000
Process AR(1) as in (12)				
Precise detection	0.673	**0.992**	0.995	0.029
One change-point	0.000	**0.000**	0.000	0.000
Oversegmentation	0.001	**0.001**	0.000	0.914
No segmentation	0.326	**0.007**	0.005	0.057
Process AR(1) in (13)				
Precise detection	0.753	**0.910**	0.954	0.023
One change-point	0.206	**0.028**	0.045	0.000
Oversegmentation	0.013	**0.062**	0.001	0.945
No segmentation	0.000	**0.000**	0.000	0.032

change-points that the process exhibits. Finally, Auto-SLEX performed badly, again detecting more than the right number of change-points.

In summary, Monte Carlo simulation experiments showed that Auto-PARM and the proposed BICBS have the better performance, with high power in the different simulation experiments. Thus, the proposed method provides an intuitive and excellent tool to detect and locate the change-points and has the advantage with respect to Auto-PARM of the simplicity, without the need of a complex searching method as the genetic algorithm.

6 Application to a Speech Recognition Dataset

The performance of the procedures is illustrated by applying them to a speech dataset consisting in the recordings of the word GREASY with 5,762 observations. GREASY has been analyzed by Ombao et al. (2002) and Davis et al. (2006). The resulting segmentations of the four procedures are presented in Fig. 1. Breakpoints are showed with vertical dashed lines.

GREASY appears in the figure as nonstationary, but it could be segmented into approximately stationary blocks. Note that in the behavior of the time series we can identify blocks corresponding to the sounds G, R, EA, S, and Y (Ombao et al. 2002). Auto-SLEX was the procedure which found more breakpoints also for this time series. The performance of ICSS, BICBS, and Auto-PARM seems to be better, finding 6–13 change-points, most of them limiting intervals corresponding to the sounds compounding the word GREASY.

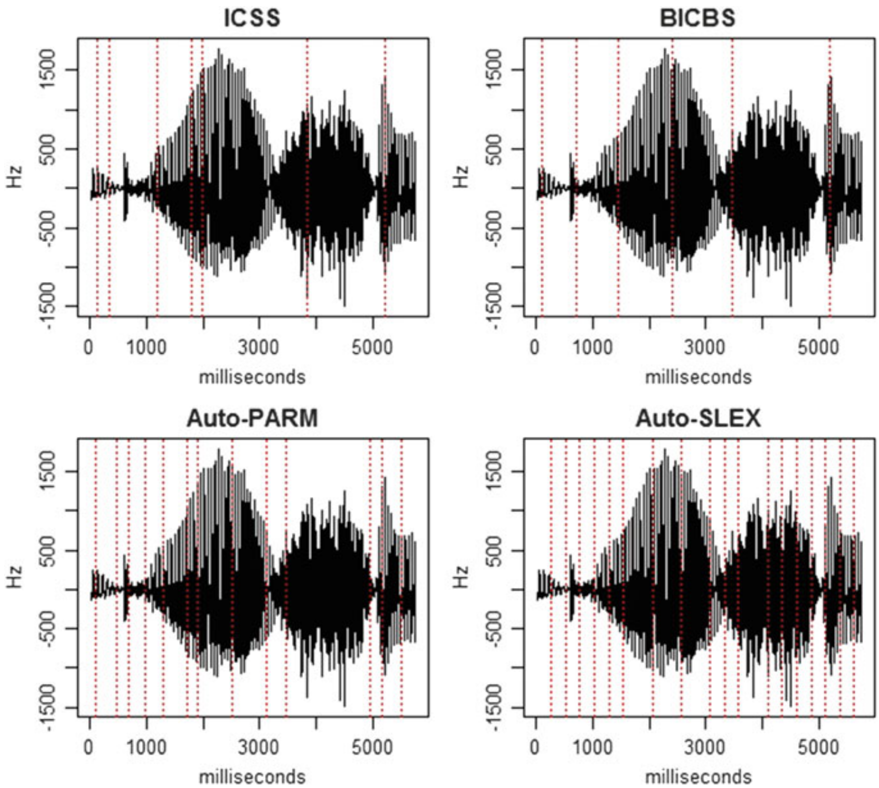

Fig. 1 Changepoints of GREASY estimated by ICSS, BICBS, Auto-PARM, and Auto-SLEX

Table 4 Standard deviation, AIC, BIC, and number of change-point in the segmentation by each methodology

	ICSS	BICBS	Auto-PARM	Auto-SLEX
Std. dev.	*51.97*	52.44	118.32	137.84
AIC	*4.0409*	*4.0409*	4.0486	4.0898
BIC	4.0763	*4.0759*	4.1178	4.1712
# change-points	7	6	13	18

In order to compare the goodness of the segmentation, we compute the standard deviation, Akaike and Bayesian Information criteria for the resulting segmentation by each method. We present the results in Table 4, where the best values of the statistics proposed are highlighted with italic font.

Although the segmentation with less standard deviation is reached by ICSS, the information criteria selected as the best the segmentation performed by BICBS.

Conclusions

In this paper we handled the problem of detecting, locating, and estimating a single or multiple change-points in the marginal mean and/or the marginal variance for both uncorrelated and serial correlated data. By combining the BIC with binary segmentation we propose a very simple procedure, which does not need a complex searching algorithms, with excellent performance in several simulation experiments.

References

Adak, S. (1998). Time-dependent spectral analysis of nonstationary time series. *Journal of the American Statistical Association, 93*(444), 1488–1489.

Al Ibrahim, A., Ahmed, M., & BuHamra, S. (2003). Testing for multiple change-points in an autoregressive model using SIC criterion. In *Focus on applied statistics* (pp. 37–51). New York: Nova Publishers.

Andreou, E., & Ghysels, E. (2002). Detecting multiple breaks in financial market volatility dynamics. *Journal of Applied Econometrics, 17*(5), 579–600.

Auger, I., & Lawrence, C. (1989). Algorithms for the optimal identification of segment neighborhoods. *Bulletin of Mathematical Biology, 51*(1), 39–54.

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

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

Brockwell, P., & Davis, R. (1991). *Time series: Theory and methods*. New York: Springer.

Brodsky, B., & Darkhovsky, B. (1993). *Nonparametric methods in change-point problems* (Vol. 243). Dordrecht: Springer.

Chen, J., & Gupta, A. (1997). Testing and locating variance changepoints with application to stock prices. *Journal of the American Statistical Association, 92*(438), 739–747.

Chen, J., & Gupta, A. (2011). *Parametric statistical change point analysis: With applications to genetics, medicine, and finance*. Boston: Birkhauser.

Chen, J., & Gupta, A. K. (2001). On change point detection and estimation. *Communications in Statistics-Simulation and Computation, 30*(3), 665–697.

Davis, R., Lee, T., & Rodriguez-Yam, G. (2006). Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association, 101*(473), 229–239.

Donoho, D., Mallat, S., & von Sachs, R. (1998). *Estimating covariances of locally stationary processes: Rates of convergence of best basis methods* (Vol. 517, pp. 1–64). Technical Report. Stanford, CA: Department of Statistics, Stanford University.

Heiler, S. (1999). *A survey on nonparametric time series analysis*. Fakultät für Wirtschaftswissenschaften und Statistik, Universität Konstanz.

Heiler, S. (2001). Nonparametric time series analysis: Nonparametric regression, locally weighted regression, autoregression, and quantile regression. In *A course in time series analysis*. Wiley and sons, (pp. 308–347).

Inclán, C., & Tiao, G. (1994). Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association, 89*(427), 913–923.

Jackson, B., Scargle, J., Barnes, D., Arabhi, S., Alt, A., Gioumousis, P., et al. (2005). An algorithm for optimal partitioning of data on an interval. *Signal Processing Letters, IEEE, 12*(2), 105–108.

Killick, R., Fearnhead, P., & Eckley, I. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association, 107*(500), 1590–1598.

Kitagawa, G., & Gersch, W. (1996). *Smoothness priors analysis of time series*. New York: Springer.
Kokoszka, P., & Leipus, R. (1999). Testing for parameter changes in ARCH models. *Lithuanian Mathematical Journal, 39*(2), 182–195.
Lee, S., Ha, J., Na, O., & Na, S. (2003). The cusum test for parameter change in time series models. *Scandinavian Journal of Statistics, 30*(4), 781–796.
Lee, S., Tokutsu, Y., & Maekawa, K. (2004). The CUSUM test for parameter change in regression models with ARCH errors. *Journal of the Japanese Statistical Society, 34*, 173–188.
Liu, J., Wu, S., & Zidek, J. (1997). On segmented multivariate regression. *Statistica Sinica, 7*, 497–526.
Maharaj, E., & Alonso, A. (2007). Discrimination of locally stationary time series using wavelets. *Computational Statistics & Data Analysis, 52*(2), 879–895.
Ombao, H., Raz, J., von Sachs, R., & Guo, W. (2002). The SLEX model of a non-stationary random process. *Annals of the Institute of Statistical Mathematics, 54*(1), 171–200.
Page, E. (1954). Continuous inspection schemes. *Biometrika, 41*(1/2), 100–115.
Rissanen, J. (1989). *Stochastic complexity in statistical inquiry theory*. Singapore: World Scientific Publishing.
Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461–464.
Scott, A., & Knott, M. (1974). A cluster analysis method for grouping means in the analysis of variance. *Biometrics, 30*(3), 507–512.
Sen, A., & Srivastava, M. (1975). On tests for detecting change in mean. *The Annals of Statistics, 3*(1), 98–108.
Stoffer, D., Ombao, H., & Tyler, D. (2002). Local spectral envelope: An approach using dyadic treebased adaptive segmentation. *Annals of the Institute of Statistical Mathematics, 54*(1), 201–223.
Vostrikova, L. (1981). "Disorder" in multidimensional random processes. *Soviet Mathematics Doklady 24*, 55–59.

Springer
http://www.springer.com/978-3-319-03121-7

Empirical Economic and Financial Research
Theory, Methods and Practice
Beran, J.; Feng, Y.; Hebbel, H. (Eds.)
2015, XVIII, 503 p. 72 illus., 27 illus. in color., Hardcover
ISBN: 978-3-319-03121-7