Off-line detection of multiple change points with the Filtered Derivative with p-Value method

Pierre R. Bertrand, Mehdi Fhima, Arnaud Guillin

Off-line detection of multiple change points with the Filtered Derivative with p-Value method

Pierre, R. BERTRAND[1,2], Mehdi FHIMA[2] and Arnaud GUILLIN[2]
[1] *INRIA Saclay*
[2] *Laboratoire de Mathématiques, UMR CNRS 6620*
& Université de Clermont-Ferrand II, France

Abstract

This paper deals with the off-line detection of multiple change points problem for time series of independent observations, with an unknown number of change points. We propose a sequential analysis method for detecting the change points with linear time and memory complexity. It is based on Filtered Derivative method which detects the right change points but also false ones. We have improve this method by computing p-values associated to all potential change points in order to eliminate false alarms which have p-value smaller than a fixed critical level. This algorithm is applied to detection of change points in the average daily volume of financial time series, and to segmentation of heartbeat time series. And it is compared with the Penalized Least Square Criterion procedure.

Keywords: Off-line detection of multiple change points, Filtered Derivative method, p-values, linear time and memory complexity.

Introduction

In a wide variety of applications including health and medicine, finance, civil engineering, one models time dependent systems by a sequence of random variables described by a finite number of structural parameters. These structural parameters can change abruptly and it is important to detect the unknown change points. Both on-line and off-line detection have their own relevance, but in this work we are concerned with off-line detection.

Statisticians have studied change point detections since the 1950's and there is a huge literature on this subject, see for e.g. the textbooks Basseville & Nikiforov [3], Brodsky & Darkhovsky [9], Csörgo & Horváth [11], Montgomery (1997) [22], and let us refer to Hušková & Meintanis [16], Kirch [17], Gombay & Serban [15] for an update overview and Birgé and Massart [8] for a good summary of the model selection approach.

Among the popular methods, we find the Penalized Least Square Criterion (PLSC). This algorithm is based on the minimization of the contrast function when the number of change point is known, see Bai and Perron [2], Lavielle and Moulines [19]. When the number of changes is unknown, many authors use the penalized version of the contrast function, see for e.g. Lavielle and Teyssière [20] or Lebarbier [21]. From a numerical point of view, the least square methods are based on dynamic programming algorithm which needs to compute a matrix. Therefore, the time and memory complexity of these algorithms is of order $\mathcal{O}(n^2)$ where n is the size of data sets. So, complexity becomes an important limitation with technological progress.

Indeed, recent measurement methods allow us to record and to stock large data sets. For example, in Section 4, we present change point analysis of heartbeat time series: It is presently possible to record the duration of each single heartbeat during a marathon race or for healthy people during 24 hours. This leads to data sets of size $n \geq 40,000$ or $n \geq 100,000$, respectively. Actually, this phenomenon is general: time dependent data are often recorded at very high frequency (VHF), which combined with size reduction of memory capacities allows recording of millions of data.

This technological progress leads us to revisit change point detection methods in the particular case of large or huge data sets. This framework constitutes the main novelty of this work: we have to develop embeddable algorithms with low time and memory complexity. Moreover, we can adopt an asymptotic point of view.

In this paper, we investigate the properties of a new off-line detection methods for multiple change points, so-called Filtered Derivative with p-value method (FDp-V). Filtered Derivative has been introduced by Basseville & Benveniste [4, 3], next Antoch and Huskova [1] propose an asymptotic study and Bertrand [5] gives some non asymptotic results. On one hand, the advantage of Filtered Derivative method is its time and memory complexity, both of order $\mathcal{O}(n)$. On the other hand, the drawback of Filtered Derivative method is that if it detects the right change points it also gives many false alarms. To avoid this drawback, we introduce a second step in order to disentangle right change points and false alarms. In this second step we calculate the p-value associated to each potential change point detected in the first step. Stress that the second step has still time and memory complexity of order $\mathcal{O}(n)$.

Our belief is that FDp-V method is quite general for large datasets. However, in this work, we restrict ourselves to detection of change points on mean and variance for a sequence of independent random variables and change point on slope and intercept for linear model. The rest of this paper is organized as follows: In Section 1, we describe Filtered Derivative with p-value Algorithm. Then, Section 2 is conscerned with theoretical results and FDp-V method for detecting changes on mean and variance. In Section 3, we present theoretical results of FDp-V method for detecting changes on slope and intercept for linear regression model. Finally, in Section 4, we give numerical simulations with a comparison with penalized least square algorithm, and we present some results on real data. Finally, an Appendix contains the proofs of our theorems and corrolaries.

1 Description of the Filtered Derivative with p-Value algorithm

In this section, we describe the the Filtered Derivative with p-value method (FDp-V). First, we precise the statistical model. Next, we describe the two steps of FDp-V method: Step 1 is based on Filtered Derivative and select the potential change points, whereas Step 2 calculate the p-value associated to each potential change point, for disentangling right change points and false alarms.

Our model:

Let $(X_t)_{t=1,\dots,n}$ be a sequence of independent r.v. with distribution $\mathcal{M}_{\theta(t)}$, where $\theta \in \mathbb{R}^d$ is a finite dimensional parameter. We assume that the maps $t \mapsto \theta(t)$ is piecewise constant, *i.e.* there exists a configuration of change points $0 = \tau_0 < \tau_1 < \cdots < \tau_K < \tau_{K+1} = n$ such that $\theta(t) = \theta_k$ for $\tau_k \leq t < \tau_{k+1}$. The integer K corresponds to the number of change times and $(K+1)$ to the number of segments. In summary, if $j \in [\tau_k, \tau_{k+1}[$, the r.v. X_j are independent and identically distributed with distribution $\mathcal{M}_{\theta_k}$.
We stress that the number of abrupt changes K is unknown, leading to a problem of model selection. There is a huge literature on change point analysis and model selection see for e.g. the monographs [3, 9]. Once again, their main drawback are usually time and memory complexity.

Filtered Derivative:

Filtered Derivative is defined as the difference between the estimators of the parameter θ computed on two sliding windows respectively at the right and at the left of the index k, both of size A, that is as the following function:

$$D(k, A) = \widehat{\theta}(k, A) - \widehat{\theta}(k - A, A), \tag{1}$$

where $\hat{\theta}(k, A)$ is an estimator of θ on the sliding box $[k + 1,\, k + A]$. Eventually, this method consists in filtering data by computing the estimators of the parameter θ before applying a discrete derivation. So this construction explains the name of the algorithm, so-called Filtered Derivative method.
The norm of the filtered derivative function, namely $\|D\|$ presents hats at the vicinity of parameter change points. However, in order to simplify the presentation, we assume that θ is a one dimensional parameter, for example the mean, in the rest of this section. Thus, the norm $\|D\|$ turns to be the absolute value $|D|$ and, the change points can be estimated as arguments of local maxima of $|D|$, see Figure 1 below. Moreover, the size of changes in θ is equal to the height of positive or negative hats of D.

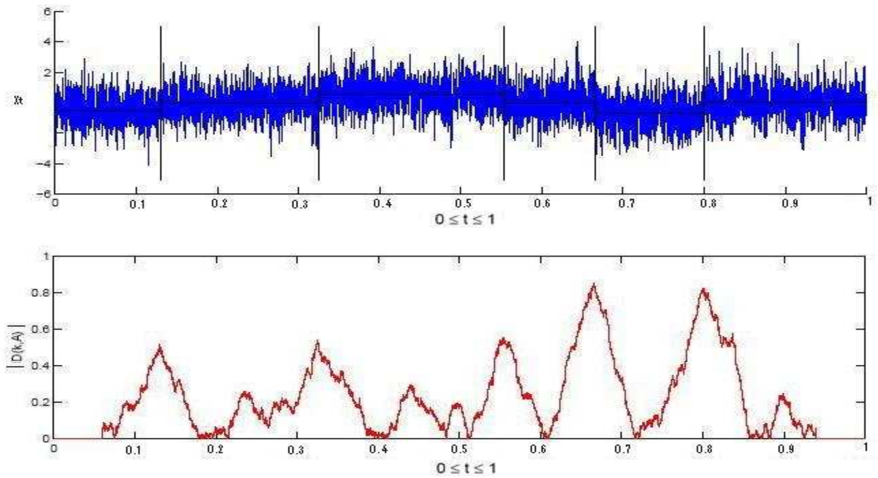

Figure 1: *Above: Gaussian random variables simulated with constant variance and presenting changes in the mean. Below: Filtered derivative function D.*

Nevertheless, we remark that the function $|D|$ gives not only the right hats but also false ones. Consequently we have introduced another idea in order to keep just the right change points. This objective is reached by splitting the detection procedure into two successive steps: In Step 1, we detect potential change points as local maxima of the filtered derivative function. In Step 2, we test wether a potential change point is a false alarm or not. Both steps use estimation of the p-value of existence of change point. The construction of two different statistical tests and the computation of p-values is detailed below.

Step 1: Detection of the potential change points

Detection of potential change points is based on the following test where we test the null hypothesis of no change in the parameter θ

$$(H_0): \ \theta_1 = \theta_2 = \cdots = \theta_{n-1} = \theta_n$$

against the alternative hypothesis indicating the existence of multiple changes

(H_1): There is an integer $K \in \mathbb{N}^*$ and $0 = \tau_0 < \tau_1 < \cdots < \tau_K < \tau_{K+1} = n$ such that

$$\theta_1 = \cdots = \theta_{\tau_1} \neq \theta_{\tau_1+1} = \cdots = \theta_{\tau_2} \cdots \neq \theta_{\tau_K+1} = \cdots = \theta_{\tau_{K+1}}.$$

where θ_i is the value of the parameter θ for $1 \leq i \leq n$.

In [4, 1, 5], potential change points are selected as corresponding to times τ_k where the absolute value of the filtered derivative $|D(\tau_k, A)|$ exceed a given threshold λ. However, the efficience of the approach is trongly linked to the choice of the threshold λ. Therefore, in this work, we have a slightly different approach: we fix a probability of type I error at level p_1^*, and we determine the corresponding critical value C_1 given by

$$\mathbb{P}\left(\max_{k\in[A:n-A]} |D(k,A)| > C_1 |\ H_0 \text{ is true}\right) = p_1^*.$$

Of course, such a probability is usually not available, so that we only have the asymptotic distribution of the maximum of $|D|$, which will be the main part of Section 2 and Section 3.

Then, roughly speaking, we select as potential change points, local maxima for which $|D(\tau_k, A)| > C_1$.

The second hyper parameter is the window size A. As pointed out in [1, 5], Filtered Derivative method works under the implicit assumption that the minimal distance between two successive change points is greater than $2A$. But, for the time being we have found no automatic choice of the window size. Thus, we need some *a priori* knowledge on the minimal length between two successive changes.
More formally, we have the following algorithm:

1. **Choice of the hyper parameters**

   - Choice of the window size A from information of the practitioners.
   - Choice of p_1^*.
     First we fix the significance level of type I error at p_1^*. Then, the expression of type I error, given in Section 2 and Section 3, fixes the value of the threshold C_1.

2. **Computation of the filtered derivative function**
   The memory complexity results from the recording of the filtered derivative sequence $(D(k, A))_{A\leq k\leq n-A}$. Clearly, it induces memory complexity of order $\mathcal{O}(n)$. On the other hand, filtered derivative function can be calculated by recurrence, see (3) or (18) and (21). These recurrence formulas induces time complexity of order $\mathcal{O}(n)$.

3. **Determination of the potential change points**

   - Initialization:
     Set counter of potential change point $k = 0$ and $\widetilde{\tau}_k = \arg\max_{k\in[A,n-A]} |D(k, A)|$.
   - While ($|D(\widetilde{\tau}_k, A)| > C_1$) do
     - $k = k + 1$
     - $D(k, A) = 0$ for all $k \in (\widetilde{\tau}_k - A, \widetilde{\tau}_k + A)$

     (We increment the change point counter and we set the values of the function D to zero because the width of the hat is equal to $2A$).
   - Finally, we sort the vector $(\widetilde{\tau}_1, \dots, \widetilde{\tau}_{K_{\max}})$ in increasing order.

end of the Step 1 of the algorithm

Figure 2 below provides an example: the family of potential change points contains the right change points (surrounded in blue in Figure 2) but also false alarms (surrounded in green in Figure 2).

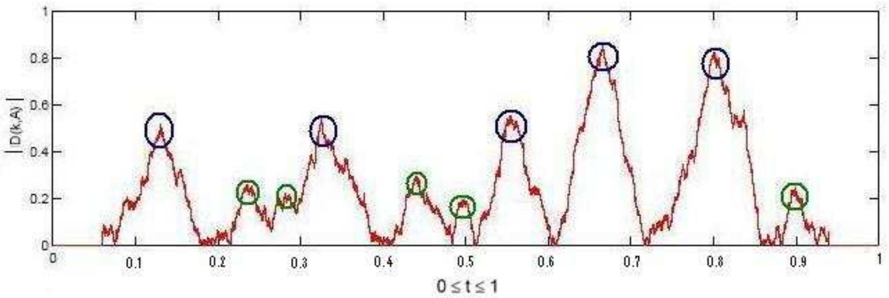

Figure 2: *Detection of potential change points*

Step 2: False alarms elimination

The potential change points $(\widetilde{\tau}_1, \dots, \widetilde{\tau}_{K_{\max}})$ present some false alarms. So, one of the main novelty of our work consists in eliminating these false detection and to keep only the right change points $(\widehat{\tau}_1, \dots, \widehat{\tau}_K)$. The idea rests on a second statistical hypothesis testing: For all potential change point $\widetilde{\tau}_k$, we test wether the parameter is the same for $k \in (\widetilde{\tau}_{k-1}+1, \widetilde{\tau}_k)$ and $k \in (\widetilde{\tau}_k + 1, \widetilde{\tau}_{k+1})$, or not. More formally, for all $1 \leq k \leq K_{\max}$, we apply the following hypothesis testing

$$(H_{0,k}) : \theta_k = \theta_{k+1} \qquad \text{versus} \qquad (H_{1,k}) : \theta_k \neq \theta_{k+1}$$

where θ_k is the estimator of θ on the segment $(\widetilde{\tau}_{k-1}+1, \widetilde{\tau}_k)$. By using this second test, we calculate new p-values $(\widetilde{p}_1, \dots, \widetilde{p}_{K_{\max}})$ associated respectively to each potential change points $(\widetilde{\tau}_1, \dots, \widetilde{\tau}_{K_{\max}})$. Then, we only keep the change points corresponding to a p-value lesser than a critical level denoted p_2^*. Consequently, Step 2 is much more selective and it permits us to keep the right change points, and so to deduce an estimator of the piecewise constant map $t \mapsto \theta_t$, see Figure 3 below.
The time and memory complexity of this second step of the algorithm is still $\mathcal{O}(n)$ because we only need to compute and to stock $(K_{\max}+1)$ estimators of parameter θ which are successively compared two by two in order to eliminate false alarms.

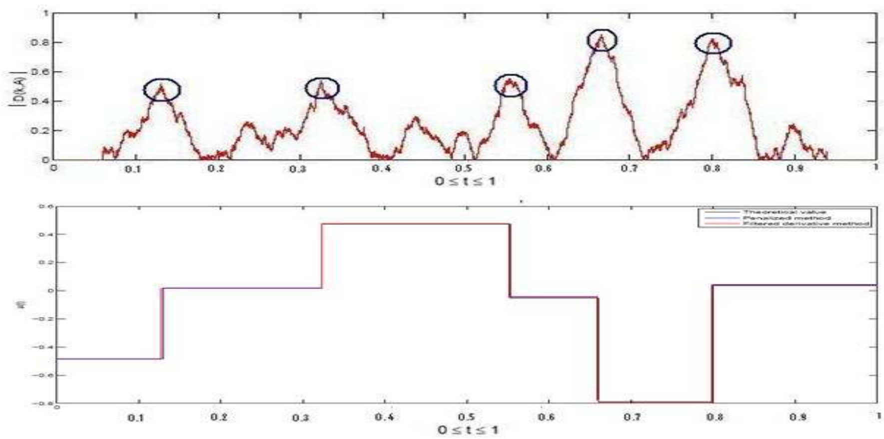

Figure 3: *Above: Detection of right change points. Below: Theoretical value of the piecewise-constant map $t \mapsto \mu_t$ (black), and its estimators given by PLSC method (blue) and FDp-V method (red).*

As a summary, FDp-V method lies in selecting in Step 1, a list of change points which can be too large, then, in Step 2, by making a statistical hypothesis testing on the selected change points to keep only the right ones. The first step has a low memory and calculation time complexity, i.e. of $\mathcal{O}(n)$. In the second step, the number of selected candidates is much less and exhibits also the advantage of a complexity of $\mathcal{O}(n)$ both in memory and computation time.

Let us finish this presentation of the algorithm by the following important remark: we will in the paper restrict ourselves to detection of change point of a dimension one parameter (mean, variance, slope...). Our algorithm however works in any (finite) dimension parameter space. The choice of the filtered derivative sequence is then crucial and it is often easier to use a filtered derivative for each parameter and divide type I error p_1^* for each of this parameter, allowing a separate treatment for change points in each parameters. The type I error p_1^* may then be seen as an upper bound of the real error.

2 Theoretical results

2.1 Change in the mean

Let $(X_i)_{i=1,\dots,n}$ be a sequence of independent r.v. with mean μ_i and a known variance σ^2. We assume that the map $i \mapsto \mu_i$ is piecewise constant, *i.e.* there exists a configuration $0 = \tau_0 < \tau_1 < \cdots < \tau_K < \tau_{K+1} = n$ such that $\mathbb{E}(X_i) = \mu_k$ for $\tau_k \le i < \tau_{k+1}$. The integer K corresponds to the number of changes. However, in any real life situation, the number of abrupt changes K is unknown, leading to a problem of model selection.
Filtered Derivative method applied to the mean is based on the difference between the empirical mean computed on two sliding windows respectively at the right and at the left of the index k, both of size A, see [1, 3]. This difference corresponds to a sequence $(D_1(k,A))_{A\le k\le n-A}$ defined by

$$D_1(k,A) \;=\; \hat{\mu}(k,A) - \hat{\mu}(k-A,A) \tag{2}$$

where $\hat{\mu}(k,A) = \frac{1}{A}\sum_{j=k+1}^{k+A} X_j$ is the empirical mean of X on the (sliding) box $[k+1,\, k+A]$. These quantities can easily be calculated by recurrence with complexity $\mathcal{O}(n)$. It suffices to remark that

$$AD_1(k+1,A) = AD_1(k,A) + X_{k+A+1} - 2X_k + X_{k-A+1}. \tag{3}$$

First we give in Theorem 1 the asymptotic behaviour of the maximum of $|D_1|$ under null hypothesis of no change in the mean and with size of the sliding windows tending to infinity at a certain rate. In the sequel, we denote by A_n the size of the sliding windows such that

$$\lim_{n\to+\infty} \frac{A_n}{n} = 0 \text{ and } \lim_{n\to+\infty} \frac{(\log n)^2}{A_n} = 0. \tag{4}$$

Theorem 1 (Change point in the mean with known variance)
Let $(X_i)_{i=1,\dots,n}$ be a sequence of independent identically distributed random variables with mean μ, variance σ^2 and assume that one of the following assumptions is satisfied

$(\mathcal{A}_1)$ $X_1 \sim \mathcal{N}(\mu,\sigma^2)$.

$(\mathcal{A}_2)$ $\exists\, t > 0$ *such as* $\mathbb{E}[\exp(tX_1)] < +\infty$, *and* $\lim_{n\to+\infty} \frac{(\log n)^3}{A_n} = 0.$

$(\mathcal{A}_3)$ $\exists\, p > 3$ *such as* $\mathbb{E}[|X_1|^p] < +\infty$, *and* $\lim_{n\to+\infty} \frac{n^{2/p}\log n}{A_n} = 0.$

Let D_1 be defined by (2). Then under the null hypothesis

$$\lim_{n\to+\infty} \mathbb{P}\left(\max_{k\in[A_n:n-A_n]} |D_1(k,A_n)| \le \frac{\sigma}{\sqrt{A_n}} c_n(x) \right) = \exp(-2e^{-x}), \tag{5}$$

$$c_n(x) = c\left(\frac{n}{A_n} - 1, x\right) \tag{6}$$

$$c(y,x) = \frac{1}{\sqrt{2\log y}}\left(x + 2\log y + \frac{1}{2}\log\log y - \frac{1}{2}\log\pi\right). \tag{7}$$

□

In applications, the variance σ^2 is unknown. For this reason we may replace it by its empirical estimator, $\widehat{\sigma}_n^2$. But, in order to keep the same result as in Theorem 1, the estimator $\widehat{\sigma}_n^2$ has to verify a certain condition given by the following theorem.

Theorem 2 (Change point in the mean with unknown variance)
We apply to the same notations and the same assumptions as in Theorem 1. Moreover, we assume that $\widehat{\sigma}_n$ is an estimator of σ satisfying

$$\lim_{n\to+\infty} |\sigma - \widehat{\sigma}_n| \log n \overset{P}{=} 0, \tag{8}$$

where the sign $\overset{P}{=}$ means convergence in probability. Then, under the null hypothesis,

$$\lim_{n\to+\infty} \mathbb{P}\left(\max_{k\in[A_n:n-A_n]} |D_1(k,A_n)| \le \frac{\widehat{\sigma}_n}{\sqrt{A_n}} c_n(x)\right) = \exp(-2e^{-x}). \tag{9}$$

□

Remark that condition (8) is not really restrictive, indeed as soon as $\widehat{\sigma}_n^2$ satisfies a CLT, the condition is verified. For example, with the usual empirical variance estimatore, a fourth order moment (of X) is sufficient.

2.2 Change in the variance

Now, we consider the case where we have a set of observations $(X_i)_{i=1,\dots,n}$ and we wish to know whether their variance has changed at an unknown time. If μ is known, then the problem is very simple. Testing (H_0) against (H_1) means that we are looking for a change in the mean of the sequence $\left((X_i-\mu)^2\right)_{i=1,\dots,n}$.
Filtered Derivative method applied to the variance is based on the difference between the empirical variance computed on two sliding windows respectively at the right and at the left of the index k, both of size A_n which satisfy condition (4). This difference is in fact a sequence of random variables denoted by $(D_2(k,A_n))_{A_n\le k\le n-A_n}$ and defined as follows

$$D_2(k,A_n) \quad = \quad \hat{\sigma}^2(k,A_n) - \hat{\sigma}^2(k-A_n,A_n) \tag{10}$$

where $\hat{\sigma}^2(k,A_n) = \dfrac{1}{A_n}\displaystyle\sum_{j=k+1}^{k+A_n}(X_j-\mu)^2$ is the empirical variance of X on the box $[k+1,k+A_n]$.
By using Theorem 1, we can deduce directly, under null hypothesis (H_0), the asymptotic distribution of the maximum of $|D_2|$. This gives straightforwardly the following corollary.

Corollary 1 (Change point in the variance with known mean)
Let $(X_i)_{i=1,\dots,n}$ be a sequence of independent identically distributed random variables with mean μ and assume that one of the following assumptions is satisfied

$(\mathcal{A}_4)$ $\exists\, t>0$ *such as* $\mathbb{E}[\exp(t(X_1-\mu)^2)] < +\infty$, *and* $\displaystyle\lim_{n\to+\infty}\frac{(\log n)^3}{A_n} = 0.$

$(\mathcal{A}_5)$ $\exists\, p>3$ *such as* $\mathbb{E}\left[|X_1-\mu|^{2p}\right] < +\infty$, *and* $\displaystyle\lim_{n\to+\infty}\frac{n^{2/p}\log n}{A_n} = 0.$

Let D_2 be defined by (10) *and $\nu^2 = Var\left[(X_1-\mu)^2\right]$. Then, under the null hypothesis,*

$$\lim_{n\to+\infty} \mathbb{P}\left(\max_{k\in[A_n:n-A_n]} |D_2(k,A_n)| \le \frac{\nu}{\sqrt{A_n}} c_n(x) \right) = \exp(-2e^{-x}), \tag{11}$$

where $c_n(.)$ is defined by (6). □

In practical situations we only rarely know the value of the constant mean. However, $\hat{\mu}(k,A_n)$ is a consistent estimator on the box $[k+1,\ k+A_n]$ for μ under (H_0) (and (H_1)). In the definition of the sequence D_2, μ is replaced by its estimator $\hat{\mu}(k,A_n)$ on the box $[k+1,\ k+A_n]$, and we let

$$\widehat{D}_2(k,A_n) \quad = \quad \widetilde{\sigma}(k,A_n) - \widetilde{\sigma}(k-A_n,A_n) \tag{12}$$

where

$$\widetilde{\sigma}(k,A_n) = \frac{1}{A_n}\sum_{j=k+1}^{k+A_n} (X_j - \hat{\mu}(k,A_n))^2$$

is the empirical variance of X on the (sliding) box $[k+1,\ k+A_n]$ with unknown mean. So, in order to obtain the same asymptotic distribution as the one obtained in Corollary 1, we must add extra conditions on the estimators $\hat{\mu}(k,A_n)$. These new conditions are given in the next corollary.

Corollary 2 (Change point in the variance with unknown mean)
With the notations and assumptions of Corollary 1, we suppose moreover that

$$\lim_{n\to+\infty} \max_{0\le k\le n-A_n} |\mu - \hat{\mu}(k,A_n)| \,(A_n \log n)^{\frac{1}{4}} \overset{a.s}{=} 0 \tag{13}$$

where the sign $\overset{a.s}{=}$ means almost surely convergence. Then under the null hypothesis

$$\lim_{n\to+\infty} \mathbb{P}\left(\max_{k\in[A_n:n-A_n]} |\widehat{D}_2(k,A_n)| \le \frac{\nu}{\sqrt{A_n}} c_n(x) \right) = \exp(-2e^{-x}). \tag{14}$$

□

Let us remark that (13) is not a very stringent condition.

2.3 Step 2: calculus of p-values

In this subsection, we recall p-value formula associated to the second test in order to eliminate false alarms. Let us stress that the only novelty of this subsection is the idea to divide the detection of abrupt change into two steps, see section 2. Since the r.v. X_i are independent, the calculus of p-value relies on well known results that can be found in any statistical textbook.

First, consider the Gaussian case and let us introduce some notations: For $1 \le k \le K_{\max}$, let $(X_{\widetilde{\tau}_{k-1}+1},\dots,X_{\widetilde{\tau}_k})$ and $(X_{\widetilde{\tau}_k+1},\dots,X_{\widetilde{\tau}_{k+1}})$ two successive samples of i.i.d. Gaussian random variables such that

$$X_{\widetilde{\tau}_{k-1}+1} \sim \mathcal{N}(\mu_k,\sigma_k^2) \qquad \text{and} \qquad X_{\widetilde{\tau}_k+1} \sim \mathcal{N}(\mu_{k+1},\sigma_{k+1}^2).$$

We can use Fisher's F statistic to determine the p-value of the existence of a change on the variance at time $\widetilde{\tau}_k$ under the null assumption (H0): $\sigma_k^2 = \sigma_{k+1}^2$. Then, we can use Student statistic to determine the p-value of a change on the mean at time $\widetilde{\tau}_k$, that is under the null assumption (H0): $\mu_k = \mu_{k+1}$.

Secondly, consider the general case. Since $A_n \to \infty$ and by construction $\widetilde{\tau}_{k+1} - \widetilde{\tau}_k \ge A_n$, we can apply CLT as soon as the r.v. X_i satisfy Lindeberg condition. Thus, the empirical mean and the empirical variance converge to the corresponding ones in the Gaussian case. Eventually, if one of the assumptions $(\mathcal{A}_i)$ for $i = 1$, 2 or 3, then we can still apply Fisher and Student statistics to compute the p-value.

3 Linear regression

In this Section, we consider linear regression model:

$$Y_i \quad = \quad aX_i + b + e_i, \text{ for } 1 \le i \le n, \tag{15}$$

where the terms $(e_i)_{i=1,\dots,n}$ are independent and identically distributed Gaussian random errors with zero-mean and variance σ^2 and, to simplify, $(X_i)_{i=1,\dots,n}$ are equidistant time points given by

$$X_i = i\Delta \qquad \text{with} \qquad \Delta > 0. \tag{16}$$

Our aim is to detect change points on the parameters (a, b) of the linear model. As the Filtered Derivative is a local method, and to simplify notations, we will restrict ourselves here to one change point.

The two following subsections give the result for detection of potential change points on the slope and on the intercept. Subsection 3.3 provide formulas for calculating the p-value during Step 2.

3.1 Change in the slope

Filtered Derivative method applied to the slope is based on the differences between estimated values of the slope a computed on two sliding windows at the right and at the left of the index k, both of size A_n. These differences, for $k \in [A_n, n - A_n]$, form a sequence of random variables, given by

$$D_3(k, A_n) \quad = \quad \hat{a}(k, A_n) - \hat{a}(k - A_n, A_n) \tag{17}$$

where

$$\hat{a}(k, A_n) = \left[A \sum_{j=k+1}^{k+A} X_j Y_j - \sum_{j=k+1}^{k+A} X_j \sum_{j=k+1}^{k+A} Y_j \right] \left[A \sum_{j=k+1}^{k+A} X_j^2 - \left(\sum_{j=k+1}^{k+A} X_j \right)^2 \right]^{-1}, \tag{18}$$

is the estimator of the slope a on the (sliding) box $[k+1,\, k+A_n]$. Let us stress that these quantities can be calculated by recurrence with complexity $\mathcal{O}(n)$.
Our first result gives the asymptotic distribution of the maximum of $|D_3|$ under the null hypothesis of no change on the linear regression.

Theorem 3 (Change point in the slope)
Let $(X_i)_{i=1,\dots,n}$ and $(Y_i)_{i=1,\dots,n}$ be given by (16) and (15) where e_i is a family of i.i.d. mean zero Gaussian r.v. with variance σ^2. Let D_3 be defined by (17) and assume that A_n satisfies condition (4). Then under the null hypothesis

$$\lim_{n \to +\infty} \mathbb{P} \left(\max_{k \in [A_n : n - A_n]} |D_3(k, A_n)| \le \frac{2\sqrt{6}\sigma}{\Delta \sqrt{A_n(A_n^2 - 1)}} d_n(x) \right) = \exp(-2e^{-x}), \tag{19}$$

with $d_n(x) = c\,(A_n, x)$ and $c(.,.)$ is defined by (7). □

3.2 Change in the intercept

In this subsection, we are concerned with detection of change points in the intercept with known slope a. To do this, we calculate the differences between estimators of the intercept b computed on two sliding windows respectively at the right and at the left of the index k, both of size A_n. For $A_n \leq k \leq n - A_n$, these differences form a sequence of random variables given by

$$D_4(k, A_n) = \hat{b}(k, A_n) - \hat{b}(k - A_n, A_n) \tag{20}$$

where

$$\hat{b}(k, A_n) = \frac{1}{A} \sum_{j=k+1}^{k+A} Y_j - a \times \frac{1}{A} \sum_{j=k+1}^{k+A} X_j, \tag{21}$$

is the estimator of the intercept b on the (sliding) box $[k+1,\ k+A_n]$. By applying Theorem 1, we get the asymptotic distribution of the maximum of $|D_4|$ under the null hypothesis of no change on the linear regression.

Corollary 3 (Change point in the intercept)
With the notations and assumptions of Theorem 3, we have under the null hypothesis

$$\lim_{n \to +\infty} \mathbb{P}\left(\max_{k \in [A_n : n - A_n]} |D_4(k, A_n)| \leq \frac{\sigma}{\sqrt{A_n}} c_n(x) \right) = \exp(-2e^{-x}), \tag{22}$$

where $c_n(.)$ is defined by (6). □

In real applications the slope is often unknown. In this case, we replace it in (20) by its empirical estimator $\widehat{a}_n$. This leads to the definition

$$\widehat{D}_4(k, A_n) = \widetilde{b}(k, A_n) - \widetilde{b}(k - A_n, A_n) \tag{23}$$

where

$$\widetilde{b}(k, A_n) = \frac{1}{A} \sum_{j=k+1}^{k+A} Y_j - \widehat{a}_n \times \frac{1}{A} \sum_{j=k+1}^{k+A} X_j,$$

is the estimator of the intercept b on the (sliding) box $[k+1,\ k+A_n]$ with unknown slope. Naturally, we must assume that the estimator of the slope satisfy a certain convergence condition which is given in the following corollary.

Corollary 4 (Change point in the intercept with unknown slope)
Under the same notations and the same assumptions than in Corollary 3. Moreover, we assume that the estimator $\hat{a}_n$ of a satisfy the following condition

$$\lim_{n \to +\infty} |a - \hat{a}_n| A_n^{\frac{3}{2}} \Delta_n \sqrt{\log n} \overset{a.s}{=} 0 \tag{24}$$

where the sign $\overset{a.s}{=}$ means almost surely convergence. Then under the null hypothesis

$$\lim_{n \to +\infty} \mathbb{P}\left(\max_{k \in [A_n : n - A_n]} |\widehat{D}_4(k, A_n)| \leq \frac{\sigma}{\sqrt{A_n}} c_n(x) \right) = \exp(-2e^{-x}). \tag{25}$$

□

3.3 Step 2: calculus of p-values

Let us give here p-value formulae associated to Step 2 in linear regression model (16) and (15). More precisely, we are concerned with detection of right change points on the slope or intercept. Recall that Step 2 of FDp-V method has been introduced in order to eliminate false alarms. Indeed, at Step 1, a time $X_{\widetilde{\tau}_k}$ has been selected as a potential change point. At Step 2, we test whether the slopes and intercepts of two data sets at left and right of the potential change point $X_{\widetilde{\tau}_k}$ are significantly different or not, and we measure this by the corresponding p-value.

Before going further, let us introduce some notations. For $1 \leq k \leq K_{\max}$, let $\mathcal{R}_k = \big((X_{\widetilde{\tau}_{k-1}+1}, Y_{\widetilde{\tau}_{k-1}+1}), \ldots, (X_{\widetilde{\tau}_k}, Y_{\widetilde{\tau}_k})\big)$ and $\mathcal{R}_{k+1} = \big((X_{\widetilde{\tau}_k+1}, Y_{\widetilde{\tau}_k+1}), \ldots, (X_{\widetilde{\tau}_{k+1}}, Y_{\widetilde{\tau}_{k+1}})\big)$ two successive samples of observations such that the relationship between variables X and Y is given by (16) and (15), or more explicitly by

$$Y_j \quad = \quad a_k X_j + b_k + e_j \qquad\qquad \text{for } \widetilde{\tau}_{k-1} + 1 \leq j \leq \widetilde{\tau}_k.$$

By using definition of the error terms,we can easily see to that slope estimator $\widehat{a}_k$ and intercept estimator $\widehat{b}_k$ have Gaussian distribution given by

$$\widehat{a}_k \sim \mathcal{N}\left(a_k, \sigma^2_{a_k}\right) \text{ with } \frac{\sigma^2_{a_k}}{\sigma^2} = \left(\sum_{j=\widetilde{\tau}_{k-1}+1}^{\widetilde{\tau}_k} (X_j - \overline{X}_k)^2\right)^{-1},$$

$$\widehat{b}_k \sim \mathcal{N}\left(b_k, \sigma^2_{b_k}\right) \text{ with } \frac{\sigma^2_{b_k}}{\sigma^2} = \frac{1}{n_k} + \overline{X}_k \left(\sum_{j=\widetilde{\tau}_{k-1}+1}^{\widetilde{\tau}_k} (X_j - \overline{X}_k)^2\right)^{-1},$$

respectively, where $\overline{X}_k$ is the empirical mean of the sequence $\left(X_{\widetilde{\tau}_{k-1}+1}, \ldots, X_{\widetilde{\tau}_k}\right)$. In the sequel, we denote by $\widehat{\sigma}^2_{a_k}$ and $\widehat{\sigma}^2_{b_k}$ the empirical variance of respectively the random variables $\widehat{a}_k$ and $\widehat{b}_k$.

Comparing slope

We want to test if the samples $\mathcal{R}_k$ and $\mathcal{R}_{k+1}$ present a change in slope or not.

$$\left(H^a_{0,k}\right) : \; a_k = a_{k+1} \qquad \text{against} \qquad \left(H^a_{1,k}\right) : \; a_k \neq a_{k+1}.$$

Then, the p-value $\widetilde{p}_{k,a}$ associated to the potential change point $\widetilde{\tau}_k$ in order to eliminate false alarms for changes in slope is given by

$$\widetilde{p}_{k,a} = 1 - \phi_a \left(\frac{|\widehat{a}_k - \widehat{a}_{k+1}|}{\sqrt{\frac{\widehat{\sigma}^2_{a_k}}{n_k} + \frac{\widehat{\sigma}^2_{a_{k+1}}}{n_{k+1}}}} \right)$$

where ϕ_a is a Student T-distribution with υ_a degrees of freedom

$$\upsilon_a = N\left(\widehat{\sigma}^2_{a_k}, \widehat{\sigma}^2_{a_{k+1}}, n_k, n_{k+1}\right) = \left\lfloor \frac{\left(\frac{\widehat{\sigma}^2_{a_k}}{n_k} + \frac{\widehat{\sigma}^2_{a_{k+1}}}{n_{k+1}}\right)^2}{\left(\frac{\widehat{\sigma}^2_{a_k}}{n_k\sqrt{n_k - 1}}\right)^2 + \left(\frac{\widehat{\sigma}^2_{a_{k+1}}}{n_{k+1}\sqrt{n_{k+1}-1}}\right)^2} \right\rfloor$$

and $\lfloor x \rfloor$ is the integer part of x.

Comparing intercept

Now, we consider the case where we test the null hypothesis that the intercepts are all identical

$$(H^b_{0,k}) \; b_k = b_{k+1} \qquad \text{against} \qquad (H^b_{1,k}) \; b_k \neq b_{k+1}$$

Then, the p-value $\widetilde{p}_{k,b}$ associated to the potential change point $\widetilde{\tau}_k$ in order to eliminate false alarms for changes in intercept is given by

$$\widetilde{p}_{k,b} = 1 - \phi_b \left(\frac{|\widehat{b}_k - \widehat{b}_{k+1}|}{\sqrt{\frac{\widehat{\sigma}^2_{b_k}}{n_k} + \frac{\widehat{\sigma}^2_{b_{k+1}}}{n_{k+1}}}} \right)$$

where ϕ_b is a Student T-distribution with $\upsilon_b = N\left(\widehat{\sigma}^2_{b_k}, \widehat{\sigma}^2_{b_{k+1}}, n_k, n_{k+1}\right)$ degrees of freedom.

4 Numerical results

4.1 A toy model: off-line detection of abrupt changes in the mean of independent Gaussian random variables with known variance

Numerical simulation

At first, we give an example on one sample. In the next paragraph, this example is plainly confirmed by Monte-Carlo simulations. To begin with, for $n = 5000$ we have simulated one replication of a sequence of Gaussian random variable $(X_1, \ldots, X_n)$ with variance $\sigma^2 = 1$ and mean $\mu_i = g(i/n)$ where g is a piecewise-constant function with five change points such as $\delta_k \in [0.5, 1.25]$ where $\delta_k := |\mu_k - \mu_{k+1}|$ represents the size of change in the mean. Then, on this sample, we have computed the function $k \to |D_1(A, k)|$ with $A = 300$, see Figure 1.
Both estimators penalized least square criterion (PLSC) and Filtered Derivative with p-values $p_1^* = 0.05$ and $p_2^* = 10^{-4}$ provide right results, see Figure 3 and the Monte-Carlo simulation below.

Monte-Carlo simulation

In this paragraph, we have made $M = 1000$ simulations of independent copies of sequences of Gaussian r.v. $X_0^{(k)}, \ldots, X_n^{(k)}$ with variance $\sigma^2 = 1$ and mean $\mu(i) = g(i/n)$, for $k = 1, \ldots, M$. On each sample, we apply the FDp-V method and the PLSC method. We find the right number of changes in 98.1% of all cases for the first method and in 97.9% for the second one, see Figure 4.

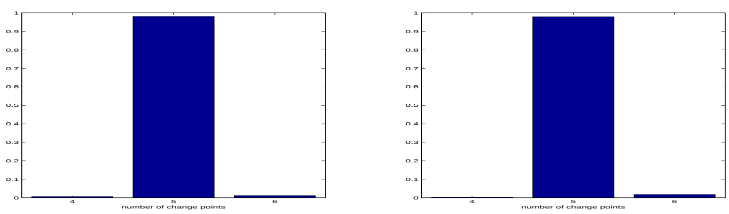

Figure 4: *Distribution of the estimated number of change points $\widehat{K}$ for $M = 1000$ realizations. Left: Using PLSC method. Right: Using Filtered Derivative method.*

Then, we compute the mean square errors. There are two kinds of mean square errors:

- Mean Integrate Square Error: $\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} |\widehat{g}(i/n) - g(i/n)|^2\right) = \mathbb{E}\|\widehat{g} - g\|^2_{L^2([0,1])}$.The estimate function is obtained in two steps : first we estimate the configuration of change points $(\hat{\tau}_k)_{k=1,\ldots,\hat{K}}$, then we estimate the value of $\widehat{g}$ between two successive change points as the empirical mean.

- Square Error on Change Points: $\mathbb{E}\left(\sum_{k=1}^{K}|\widehat{\tau}_k - \tau_k|^2\right)$, in the case where we have found the right number of change points.

Table 1 gives the result of Monte Carlo simulation mean errors, and alsothe comparison between the mean time complexity and the mean memory complexity. We have written the two programs in Matlab and have runned it with computer system which has the following characteristics: 1.8GHz processor and 512MB memory.

	Square Error on Change Points	Mean Integrated Squared Error
FDp-V method	1.1840×10^{-4}	0.0107
PLSC method	1.2947×10^{-4}	0.0114
	Memory allocation (in Megabytes)	CPU time (in second)
FDp-V method	0.04 MB	0.005 s
PLSC method	200 MB	240 s

Table 1: Errors and complexities given by FDp-V method and PLSC method.

Numerical conclusion

On one hand, both methods have the same accuracy in terms of percentage of selection of the right model, Square Error on the configuration of change points or Mean Integrate Square Error. On the other hand, the Filtered Derivative with p-Value is less expensive in terms of time complexity and memory complexity, see Table 1. of computer memory, while Filtered derivative method only needs 0.008%. This plainly confirms the difference of time and memory complexity, *i.e.* $\mathcal{O}(n^2)$ *versus* $\mathcal{O}(n)$.

4.2 Off-line detection of changes in the slope of simple linear regression

In this subsection, we consider the problem of detecting the change points in the slope of a simple linear model corrupted by an additive Gaussian noise. At first, for $n = 1400$ we have simulated two replications of sequences $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$ defined by (16) and (15) with $\Delta = 1$, $\sigma = 30$ and slope $a_i = h(i)$ where h is a piecewise-constant function with four change points. For the first replication, the size of changes in the slope is small, i.e. $\nu_k \in [0.75, 1]$ where $\nu_k := |a_k - a_{k+1}|$. The second replication presents biggest changes such as $\nu_k \in [3, 5]$. Next, we have plot X versus Y, see scatter plots 5 and 8.

Then, in order to detect changes in the slope of these simulated data, we have computed the function $k \mapsto |D_3(k, A)|$ with $A = 100$, see Figures 6 and 9.

Finally, by applying FDp-V procedure with p-values $p_1^* = 0.05$ and $p_2^* = 10^{-10}$, we obtain a right localization of the change points and so a right estimation of the piecewise-constant function h, see Figures 7 and 10.

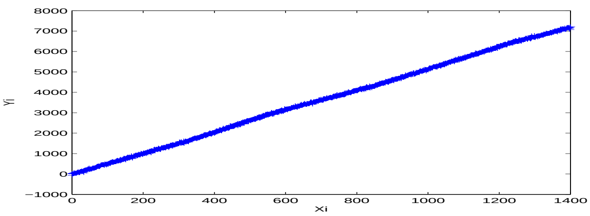

Figure 5: *Case 1: Scatter plot of the simulated data* (X_i, Y_i) *for* $1 \leq i \leq n$.

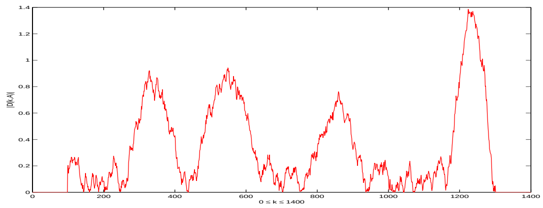

Figure 6: *Case 1: The hat function* $|D_3|$.

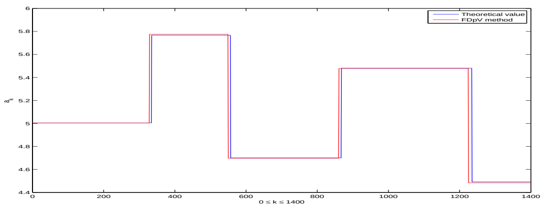

Figure 7: *Case 1: Theoretical value of the piecewise-constant function h (blue), and its estimator given by FDp-V method (red).*

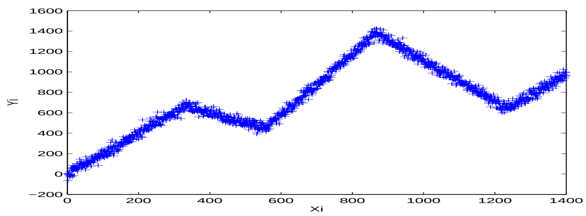

Figure 8: *Case 2: Scatter plot of the simulated data* (X_i, Y_i) *for* $1 \leq i \leq n$.

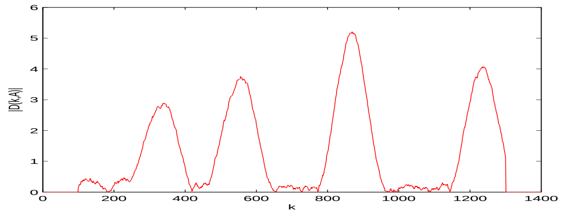

Figure 9: *Case 2: The hat function* $|D_3|$.

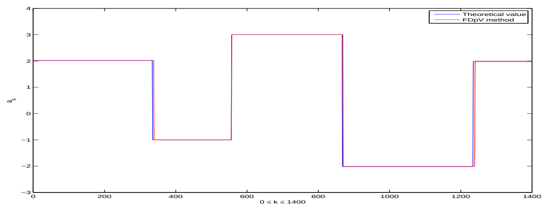

Figure 10: *Case 2: Theoretical value of the piecewise-constant function h (blue), and its estimator given by FDp-V method (red).*

4.3 Application to real data

In this subsection, we apply our algorithm to detect changes points in the mean of two real samples drawn from on one hand health and wellbeing, and on the other hand finance. Then, we show and analyze the obtained results.

An Application to Change Detection of Heartbeat Time Series

In this paragraph, we give an example of application of FDp-V method to heartbeat time series analysis. ECG signal analysis has a long story after the implementation of monitoring by Holter at the beginning of the fifties. However, we consider here the RR interval, which provides an accurate measure of the length of each single heartbeat and corresponds to the instantaneous speed of the heart engine, see [14]. From the beginning of 21st century, the size reduction of the measurement devices allow us to record heartbeat time series for healthy people in ecological situation over a long period of time: marathon runners, individuals daily (24 hours) records, etc. We then obtain large data sets of more than 40.000 observations that allow us to characterize the variations of the heartbeat, see Figure 11 below.

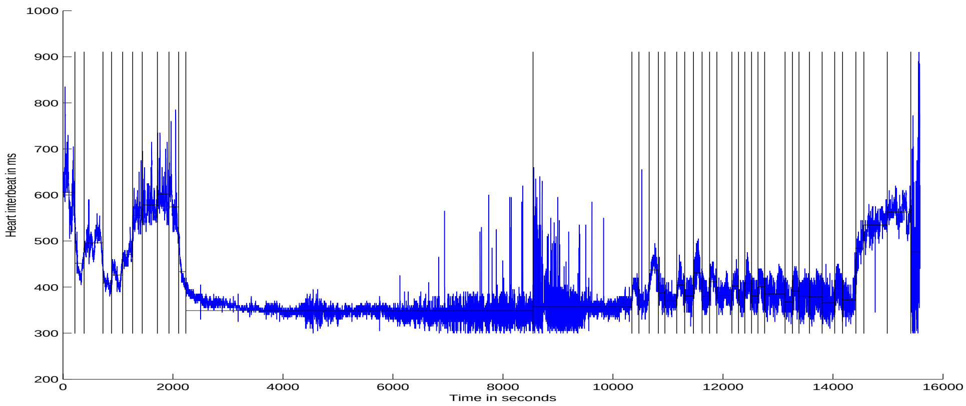

Figure 11: *Segmentation of heartbeat time series of a marathon runner*

The previous hearbeat time series of marathon runner has been recorded by Véronique Billat and team UBIAE (U. 902 INSERM and Évry Génopole) during Paris Marathon 2006. Data are then been preprocessed by using the "CLEANING TACHOGRAM" algorithm developed by Nadia Khalfa at INRIA Saclay in 2009. "CLEANING TACHOGRAM" cancelled aberrant data, based on physiological considerations rather than statistical procedure. This work is part of the project "PHYSIOSTAT" supported by a DIM Digitéo.

In this case, the FD-pV method offer the advantage of being a fast and automatic procedure of segmentation of a large dataset, on the mean in this example, but possibly on the slope or the variance. Combined with the "CLEANING TACHOGRAM" algorithm, we then have an entirely automatic procedure to obtain apparently homogeneous segment. The next step of our study will be to detect change on hidden structural parameters. This will be the subject of forthcoming studies.

Changes in the average daily volume

Trading volume represents number of shares or contracts traded in a financial market during a specific period. Average traded volume is an important indicator in technical analysis as it is used to measure the worth of a market move. If the markets move significantly up or down, the perceived strength of that move depends on the volume of trading in that period. The higher the volume during that price move, the more significant the move. Therefore, the detection of abrupt changes in the average daily volume provide relevant information for financial engineer, trader, etc. Then, we consider here a daily volume of *Carbone Lorraine* compagny observed during 02 January 2009. These data has been kindly furnished by Charles-Albert Lehalle from Crédit Agricole Cheuvreux, Groupe CALYON (Paris). The results obtained with our algorithm for $A = 300$, $p_1^* = 0.05$ and $p_2^* = 10^{-5}$ are illustrated in Figure 12. It appears that FDp-V procedure detects majors changes observed after each huge variations. In future works, we will investigate sequential detection of change points in the average daily volume in connection with the worth of a market move.

Conclusion

It clearly appears that both methods, namely: FDp-V and PLSC, give right results with practically the same precision. But, when we compare the complexity, we remark that the FDp-V method is less expensive in terms of time and memory complexity. Consequently, our method is faster (time) and cheaper (memory), and so it is more adapted to segment random

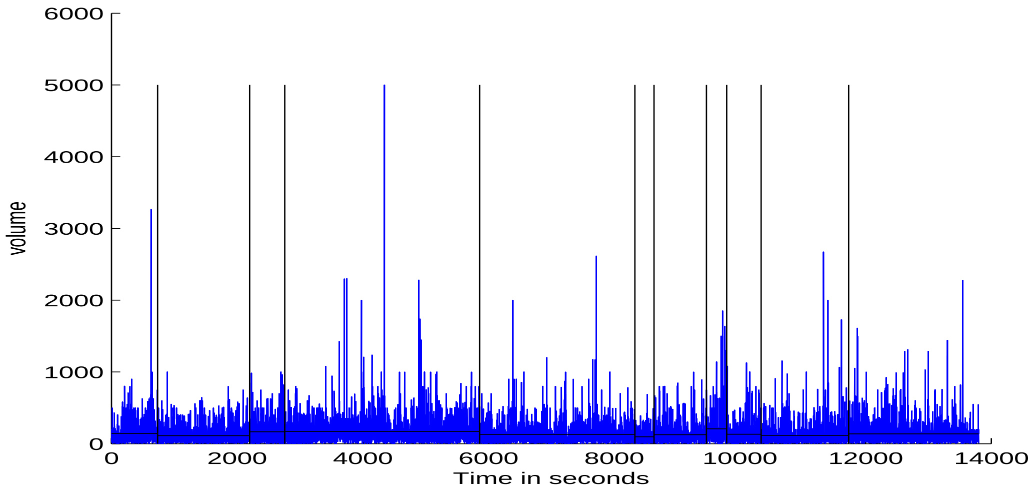

Figure 12: *Segmentation of the daily volume of Carbone Lorraine compagny observed during 02 January 2009.*

signals with large data sets.
In future works, we will develop the Filtered Derivative with p-value method in order to detect abrupt changes in parameters of weakly and strongly dependent series. In particularly, we will consider the detection problem on the Hurst parameter of the multifractional Brownian motion which is a long memory process and apply it to physiological data as in Billat *et al.* [7].
Let us also mention that whereas Filtered FDp-V method is based on sliding window and could be adapted to sequential detection, see for instance Bertrand [5], Bertrand & Fleury [6], and Bertrand & Fhima citeBertrand:Fhima:2009.

References

[1] Antoch, J., and Hušková, M. Procedures for the detection of multiple changes in series of independent observations. In *Asymptotic statistics (Prague, 1993)*, Contrib. Statist. Physica, Heidelberg, 1994, pp. 3–20.

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

[3] Basseville, M., and Nikiforov, I. V. *Detection of abrupt changes: theory and application.* Prentice Hall Information and System Sciences Series. Prentice Hall Inc., Englewood Cliffs, NJ, 1993.

[4] Benveniste, A., and Basseville, M. Detection of abrupt changes in signals and dynamical systems: some statistical aspects. In *Analysis and optimization of systems, Part 1 (Nice, 1984)*, vol. 62 of *Lecture Notes in Control and Inform. Sci.* Springer, Berlin, 1984, pp. 145–155.

[5] Bertrand, P. R. A local method for estimating change points: the “hat-function”. *Statistics 34*, 3 (2000), 215–235.

[6] Bertrand, P. R., and Fleury, G. Detecting small shift on the mean by finite moving average. *International Journal of Statistics and Management System 3* (2008), 56–73.

[7] Billat, V. L., Hamard, L., Meyer, Y., and Wesfreid, E. Detection of changes in the fractal scaling of heart rate and speed in a marathon race. *Physica A 388* (2009), 3798Ð3808.

[8] Birgé, L., and Massart, P. Minimal penalties for Gaussian model selection. *Probab. Theory Related Fields 138*, 1-2 (2007), 33–73.

[9] Brodsky, B. E., and Darkhovsky, B. S. *Nonparametric methods in change-point problems*, vol. 243 of *Mathematics and its Applications.* Kluwer Academic Publishers Group, Dordrecht, 1993.

[10] Chen, X. Inference in a simple change-point model. *Scienta Sinica A31* (1988), 654–667.

[11] Csörgo, M., and Horváth, L. *Limit Theorem in Change-Point Analysis.* J. Wiley, New York., 1997.

[12] Csörgö, M., and Révész, P. *Strong Approximations in Probability and Statistics.* Akadémiai Kiadö, Budapest, 1981.

[13] Csáki, E., and Gonchigdanzan, K. Almost sure limit theorems for the maximum of stationary gaussian sequences. *Stat. Prob. Letters 58* (2002), 195–203.

[14] force of the European Soc. Cardiology and the North American Society of Pacing, T., and Electrophysiology. Heart rate variability. standards of measurement, physiological interpretation, and clinical use. *Circulation 93* (1996), 1043–1065.

[15] Gombay, E., and Serban, D. Monitoring parameter change in ar(p) time series models. *Journal of Multivariate Analysis 100* (2009), 715–725.

[16] Hušková, M., and Meintanis, S. Change point analysis based on the empirical characteristic functions of ranks. *Sequential Analysis 25* (2006b), 421–436.

[17] Kirch, C. Bootstrapping sequential change-point tests. *Sequential Analysis 27* (2008), 330–349.

[18] Komlós, J., Major, P., and Tusnády, G. An approximation of partial sums of independent rv's and the sample df. *I. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 32* (1975), 111–131.

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

[20] Lavielle, M., and Teyssière, G. Detection of multiple change points in multivariate time series. *Lithuanian Math. J. 46* (2006), 287–306.

[21] Lebarbier, E. Detecting multiple change-points in the mean of gaussian process by model selection. *Signal Processing 85* (2005), 717–736.

[22] Montgomery, D. *Introduction to Statistical Quality Control, 3rd Edition.* John Wiley & Sons, New York, 1997.

[23] Pickands III, J. Asymptotic properties of the maximum in a stationary gaussian process. *Trans. Amer. Math. Soc 145* (1969), 75–86.

[24] Qualls, C., and Watanabe, H. Asymptotic properties of gaussian processes. *Ann. Math. Statist 43* (1972), 580–596.

A Proof

Proof of Theorem 1

Under the null hypothesis (H_0) the filtered derivative can be rewritten as follows

$$\frac{\sqrt{A_n}}{\sigma} D_1(k, A_n) = \frac{S_{k-A_n} - 2S_k + S_{k+A_n}}{\sqrt{A_n}},$$

where $S_k = \sum_{j=1}^{k} \xi_j$, with $(\xi_j)_{j\in[1,n]}$ a sequence of i.i.d r.v such as $\mathbb{E}[\xi_1] = 0$, $\mathbb{E}[\xi_1^2] = 1$, $S_0 = 0$.

To achieve our goal, we state three lemmas. Firstly, we show in Lemma 1 that if a positive sequence (a_n) has the following asymptotic distribution

$$\lim_{n\to+\infty} \mathbb{P}(a_n \le c_n(x)) = \exp(-2e^{-x}),$$

where $c_n(.)$ is defined by (6) and if there is a second positive sequence (b_n) which converges almost surely (a.s) to (a_n) with rate of convergence of order $\mathcal{O}\left(\sqrt{\log n}\right)$, then (b_n) has the same asymptotic distribution as (a_n). Secondly, we prove in Lemma 2 that under one of the assumptions $(\mathcal{A}_i)$ with $i \in \{1,2,3\}$, the maximum of the increment $\frac{\sqrt{A_n}}{\sigma}|D_1(k, A_n)|$ converges a.s to the maximum of discrete Wiener process' increment with rate $\mathcal{O}\left(\sqrt{\log n}\right)$. Also, we display in Lemma 3 that the maximum of **discrete** Wiener process' increment converges a.s to the maximum of **continuous** Wiener process' increment with rate $\mathcal{O}\left(\sqrt{\log n}\right)$. Then, by applying Qualls and Watanabe [24, Theorem 5.2, p. 594] we deduce the asymptotic distribution of the maximum of continuous Wiener process' increment. Finally, by combining these results, we get directly (5).

To begin with, let us state the first lemma.

Lemma 1

Let (a_n) and (b_n) two sequences of positive r.v.'s and we denote $\eta_n = |a_n - b_n|$.
We assume that

1. $\lim_{n\to+\infty} \mathbb{P}(a_n \le c_n(x)) = \exp(-2e^{-x})$.
2. $\lim_{n\to+\infty} \eta_n\sqrt{\log n} \overset{a.s}{=} 0$.

where $c_n(.)$ is defined by (6). Then

$$\lim_{n\to+\infty} \mathbb{P}(b_n \le c_n(x)) = \exp(-2e^{-x}). \tag{26}$$

◊

Proof of Lemma 1

Without any restriction, we can consider the case where $\eta_n > 0$. We denote by $|\Delta x|$ an infinitesimally small change in x. Then, for n large enough and $|\Delta x|$ small enough, $c_n(x + |\Delta x|)$ and $c_n(x - |\Delta x|)$ satisfy the following inequalities

$$c_n(x + |\Delta x_n|) \ \ge \ c_n(x) + \frac{|\Delta x_n|}{\sqrt{2\log n}}, \tag{27}$$

$$c_n(x - |\Delta x_n|) \ \le \ c_n(x) - \frac{|\Delta x_n|}{\sqrt{2\log n}}. \tag{28}$$

Following Chen [10], we supply a lower and an upper bounds of $\mathbb{P}(b_n \le c_n(x))$. On the one hand, by using the inequality (27), the upper bound results from the following calculations

$$\begin{aligned}
\mathbb{P}\left(b_n \leq c_n(x)\right) &\leq \mathbb{P}\left(b_n \leq c_n\left(x+|\Delta x|\right) - \frac{|\Delta x|}{\sqrt{2\log n}}\right) \\
&\leq \mathbb{P}\left(a_n - \eta_n \leq c_n\left(x+|\Delta x|\right) - \frac{|\Delta x|}{\sqrt{2\log n}}\right) \\
&\leq \mathbb{P}\left(a_n \leq c_n\left(x+|\Delta x|\right)\right) + \mathbb{P}\left(-\eta_n \leq -\frac{|\Delta x|}{\sqrt{2\log n}}\right) \\
&= \mathbb{P}\left(a_n \leq c_n\left(x+|\Delta x|\right)\right) + \mathbb{P}\left(\eta_n \geq \frac{|\Delta x|}{\sqrt{2\log n}}\right).
\end{aligned}$$

On the other hand, by using the inequality (28), the lower bound results from analogous calculations

$$\mathbb{P}\left(a_n \leq c_n\left(x-|\Delta x_n|\right)\right) \leq \mathbb{P}\left(b_n \leq c_n(x)\right) + \mathbb{P}\left(\eta_n \geq \frac{|\Delta x_n|}{\sqrt{2\log n}}\right).$$

Then, by putting together the two previous bounds, we obtain

$$\begin{aligned}
\mathbb{P}\left(a_n \leq c_n\left(x-|\Delta x_n|\right)\right) - \mathbb{P}\left(\eta_n \geq \frac{|\Delta x_n|}{\sqrt{2\log n}}\right) &\leq \mathbb{P}\left(b_n \leq c_n(x)\right) \\
&\leq \mathbb{P}\left(a_n \leq c_n\left(x+|\Delta x_n|\right)\right) + \mathbb{P}\left(\eta_n \geq \frac{|\Delta x_n|}{\sqrt{2\log n}}\right).
\end{aligned}$$

Finally, by taking the limit in n and as $|\Delta x|$ is arbitrary small, we deduce (26). This finishes the proof of Lemma 1. ♦

Now, we show that the maximum of $\frac{\sqrt{A_n}}{\sigma}|D_1(k,A_n)|$ converges a.s to the maximum of discrete Wiener process' increment with rate of convergence of order $\mathcal{O}\left(\sqrt{\log n}\right)$. This is stated in Lemma 2 below (which gives also corrections to previous results of [10]):

Lemma 2
Let $(W_t,\ t\geq 0)$ be a standard Wiener process and $\left(Z_{A_n}(q), 0\leq q\leq \frac{n}{A_n}-1\right)$ be the discrete sequence defined by

$$Z_{A_n}(q) = \begin{cases} W_{q-1} - 2W_q + W_{q+1} & \text{if } 1\leq q\leq \frac{n}{A_n}-1, \\ 0 & \text{else}. \end{cases} \tag{29}$$

Let $(X_i)_{i=1,\dots,n}$ be a sequence of independent identically distributed random variables with mean μ, variance σ^2, D_1 be defined by (2), and denote

$$\eta_{1,n} = \left| \max_{0\leq k\leq n-A_n} \frac{\sqrt{A_n}}{\sigma}|D_1(k,A_n)| - \max_{0\leq q\leq \frac{n}{A_n}-1} |Z_{A_n}(q)| \right|.$$

Moreover, we suppose that one of the assumptions $(\mathcal{A}_i)$, with $i\in\{1,2,3\}$ is in force. Then there exits a Wiener process $(W_t)_t$ such that

$$\lim_{n\to+\infty} \eta_{1,n}\sqrt{\log n} \overset{a.s}{=} 0. \tag{30}$$

◊

Proof of Lemma 2

We consider a new discrete sequence, $(B(k, A_n), 0 \leq k \leq n - A_n)$, obtained by scaling from the sequence $\left(Z_{A_n}(q), 0 \leq q \leq \frac{n}{A_n} - 1\right)$. It is defined as follows

$$B(k, A_n) = \begin{cases} \dfrac{W_{k-A_n} - 2W_k + W_{k+A_n}}{\sqrt{A_n}} & \text{if } A_n \leq k \leq n - A_n \\ 0 & \text{else} \end{cases}$$

Then

$$\eta_{1,n} \stackrel{\mathcal{D}}{=} \left| \max_{0 \leq k \leq n-A_n} \frac{\sqrt{A_n}}{\sigma} |D(k, A_n)| - \max_{0 \leq k \leq n-A_n} |B(k, A_n)| \right|,$$

where the sign $\stackrel{\mathcal{D}}{=}$ means equality in law. Depending on which assumption $(\mathcal{A}_i)$ is in force, we have three different proofs:

1. Assuming $(\mathcal{A}_1)$.
   This is the simplest case. We can choose a standard Wiener process, $(W_t,\ t \geq 0)$, such that $S_k = W_k$ at all the integers $k \in [0, n - A_n]$. Hence, $\frac{\sqrt{A_n}}{\sigma} D(k, A_n) = B(k, A_n)$. Then, we can deduce (30).

2. Assuming $(\mathcal{A}_2)$.
   We have

   $$0 \leq \eta_{1,n} \leq \max_{0 \leq k \leq n-A_n} |D(k, A_n) - B(k, A_n)| \leq \frac{4}{\sqrt{A_n}} \max_{0 \leq k \leq n-A_n} |S_k - W_k|$$

   and after

   $$0 \leq \eta_{1,n} \sqrt{\log n} \leq \frac{4(\log n)^{\frac{3}{2}}}{\sqrt{A_n}} \times \frac{\max_{0 \leq k \leq n-A_n} |S_k - W_k|}{\log n}$$

   However, $\lim_{n \to +\infty} \frac{(\log n)^3}{A_n} = 0$ and according to Komlós *et al* [18, Theorem 3, p.34], there is a Wiener process, $(W_t,\ t \geq 0)$, such as $\lim_{n \to +\infty} \frac{\max_{0 \leq k \leq n} |S_k - W_k|}{\log n} < +\infty$. Then, we can deduce (30).

3. Assuming $(\mathcal{A}_3)$.
   By using the same tricks than in the proof of the case $(\mathcal{A}_2)$, we can show that

   $$0 \leq \eta_{1,n} \sqrt{\log n} \leq \frac{4n^{\frac{1}{p}} \sqrt{\log n}}{\sqrt{A}} \times \frac{\max_{0 \leq k \leq n-A_n} |S_k - W_k|}{n^{\frac{1}{p}}}$$

   However, $\lim_{n \to +\infty} \dfrac{n^{\frac{2}{p}} \log n}{A_n} = 0$ and according to Komlós *et al* [18, Theorem 3, p.34], there is a Brownian motion, $(W_t,\ t \geq 0)$, such as $\lim_{n \to +\infty} \dfrac{\max_{0 \leq k \leq n} |S_k - W_k|}{n^{\frac{1}{p}}} < +\infty$. Then, we can deduce (30).

This finishes the proof of Lemma 2. ♦

In order to apply Qualls and Watanabe [24, Theorem 5.2, p. 594] theorem, we have to consider a continuous version of the process Z_{A_n}. For this reason, we define the continuous process $\left(Z_{A_n}(t), t \in \left[0, \dfrac{n}{A_n} - 1\right]\right)$ such as

$$Z_{A_n}(t) = W_{t-1} - 2W_t + W_{t+1}. \tag{31}$$

Then, in Lemma 3, we show that the maximum of $|Z_{A_n}(q)|$ converges a.s to the maximum of $|Z_{A_n}(t)|$ with rate of convergence of order $\mathcal{O}\left(\sqrt{\log n}\right)$.

Lemma 3
Let Z_{A_n} be defined by (29) and set $\eta_{2,n} = \left| \sup_{t\in\left[0,\frac{n}{A_n}-1\right]} |Z_{A_n}(t)| - \max_{0\le q\le \frac{n}{A_n}-1} |Z_{A_n}(q)| \right|.$

Then

$$\lim_{n\to+\infty} \eta_{2,n}\sqrt{\log n} \overset{a.s}{=} 0. \qquad (32)$$

◊

Proof of Lemma 3
We have

$$0 \le \eta_{2,n} \le 4 \sup_{t\in\left[0,\frac{n}{A_n}-1\right]} |W_{t+1}-W_t| \le 4 \sup_{t\in\left[0,\frac{n}{A_n}-\frac{1}{A_n}\right]} \sup_{s\in\left[0,\frac{1}{A_n}\right]} |W_{t+s}-W_t|.$$

This implies

$$0 \le \eta_{2,n}\sqrt{\log n} \le 4\sqrt{\log n} \sup_{t\in\left[0,\frac{n}{A_n}-\frac{1}{A_n}\right]} \sup_{s\in\left[0,\frac{1}{A_n}\right]} |W_{t+s}-W_t|$$

and after

$$\mathbb{P}\left(\eta_{2,n}\sqrt{\log n}\ge\delta\right) \le \mathbb{P}\left(\sup_{t\in\left[0,\frac{n}{A_n}-\frac{1}{A_n}\right]} \sup_{s\in\left[0,\frac{1}{A_n}\right]} |W_{t+s}-W_t| \ge \frac{\delta\sqrt{A_n}}{4\sqrt{\log n}}\frac{1}{\sqrt{A_n}} \right)$$

According to Csörgö and Révész [12, Lemma 1.2.1, p. 29], and by taking $T = n/A_n$, $h = 1/A_n$, $\varepsilon = 1$, $\nu = \dfrac{\delta\sqrt{A_n}}{4\sqrt{\log n}}$, and C a non negative real, we deduce that

$$\mathbb{P}\left(\eta_{2,n}\sqrt{\log n}\ge\delta\right) \le C\frac{n}{A_n}\exp\left(-\frac{\delta^2 A_n}{48\log n}\right).$$

Next, by using $\lim_{n\to+\infty}\frac{(\log n)^2}{A_n} = 0$, we can deduce

$$\sum_{n\ge1}\frac{n}{A_n}\exp\left(-\frac{\delta^2 A_n}{48\log n}\right) < +\infty,$$

and after

$$\sum_{n\ge1}\mathbb{P}\left(\eta_{2,n}\sqrt{\log n}\ge\delta\right) < +\infty.$$

Then, according to Borel-Cantelli lemma, we can deduce (32). This finishes the proof of Lemma 3. ♦

End of the proof of Theorem 1

Next, we apply Qualls and Watanabe [24, Theorem 5.2, p. 594] to the continuous process $|Z_{A_n}(t)|$. We obtain the asymptotic distribution of its maximum given by

$$\lim_{n\to+\infty}\mathbb{P}\left(\sup_{t\in\left[0,\frac{n}{A}-1\right]} |Z_{A_n}(t)| \le c_n(x) \right) = \exp(-2e^{-x}). \qquad (33)$$

This result can be proved by applying Theorem 5.2 of Qualls and Watanabe [24] to the centered stationary Gaussian process $(Z_{A_n}(t), t\ge0)$. The covariance function of $Z_{A_n}(t)$ is given by

$$\rho(\tau) = \frac{Cov\left(Z_{A_n}(t), Z_{A_n}(t+\tau)\right)}{\sqrt{Var\left(Z_{A_n}(t)\right)}\sqrt{Var\left(Z_{A_n}(t+\tau)\right)}} = \begin{cases} 1-|\tau| & \text{if } 0\le|\tau|\le1,\\ -1+\dfrac{|\tau|}{2} & \text{if } 1<|\tau|\le2,\\ 0 & \text{if } 2<|\tau|<+\infty. \end{cases}$$

So, the conditions of Theorem 5.2 of Qualls and Watanabe are satisfied in the following way: $\alpha = 1$, $H_\alpha = 1$ (according to Pickands [23, p. 77]), and $\widetilde{\sigma}^{-1}(x) = 2x^{-2}$. This finishes the proof of (33).

Eventually, by combining Lemma 1, Lemma 3 and result (33), we get the asymptotic distribution of the maximum of the sequence $\left(|Z_{A_n}(q)|, 0 \leq q \leq \frac{n}{A_n} - 1\right)$. Then, by using Lemma 1 and Lemma 2, we immediately get the distribution of the maximum of the filtered derivative sequence $(|D_1(A_n, k)|, 0 \leq k \leq n - A_n)$. ■

Proof of Theorem 2

Fix $\varepsilon > 0$, the key argument is to divide Ω into two complementary events

$$\Omega_{1,n} = \{|\sigma - \widehat{\sigma}_n| \log n \leq \sigma\varepsilon\} \qquad \text{and} \qquad \Omega_{2,n} = \{|\sigma - \widehat{\sigma}_n| \log n > \sigma\varepsilon\}.$$

Then, we have

$$\begin{aligned}\mathbb{P}\left(\max_{k\in[A_n:n-A_n]} |D_1(k, A_n)| \leq \frac{\widehat{\sigma}_n}{\sqrt{A_n}} c_n(x)\right) &= \mathbb{P}\left(\max_{k\in[A_n:n-A_n]} |D_1(k, A_n)| \leq \frac{\widehat{\sigma}_n}{\sqrt{A_n}} c_n(x) \text{ and } \Omega_{1,n}\right) \\ &+ \mathbb{P}\left(\max_{k\in[A_n:n-A_n]} |D_1(k, A_n)| \leq \frac{\widehat{\sigma}_n}{\sqrt{A_n}} c_n(x) \text{ and } \Omega_{2,n}\right)\end{aligned}$$

On the one hand, we remark that

$$\mathbb{P}\left(\max_{k\in[A_n:n-A_n]} |D_1(k, A_n)| \leq \frac{\widehat{\sigma}_n}{\sqrt{A_n}} c_n(x) \text{ and } \Omega_{2,n}\right) \leq \mathbb{P}(\Omega_{2,n}),$$

which combined with assumption (8) implies that

$$\lim_{n\to+\infty} \mathbb{P}(\Omega_{2,n}) = 0. \tag{34}$$

On the other hand, for all $\omega \in \Omega_{1,n}$, we have $\widehat{\sigma}_n = \sigma(1+\lambda_n(\omega))$ with $|\lambda_n(\omega)| \leq \frac{\varepsilon}{\log n}$. Therefore, $\lim_{n\to+\infty} \lambda_n \overset{a.s}{=} 0$. Next, by setting $a_n = \frac{\sqrt{A_n}}{\sigma} \max_{k\in[A_n:n-A_n]} |D_1(k, A_n)|$, we get

$$\begin{aligned}\mathbb{P}\left(\max_{k\in[A_n:n-A_n]} |D_1(k, A_n)| \leq \frac{\widehat{\sigma}_n}{\sqrt{A_n}} c_n(x) \text{ and } \Omega_{1,n}\right) &= \mathbb{P}(a_n \leq c_n(x)(1 + \lambda_n(\omega))) \\ &= \mathbb{P}(a_n - \eta_{3,n} \leq c_n(x))\end{aligned}$$

with $\eta_{3,n} = c_n(x)\lambda_n$. Therefore, after having checked that

$$\lim_{n\to+\infty} \eta_{3,n}\sqrt{\log n} \overset{a.s}{=} 0, \tag{35}$$

we can apply Lemma 1 which combined with Theorem 1 implies that

$$\mathbb{P}\left(\max_{k\in[A_n:n-A_n]} |D_1(k, A_n)| \leq \frac{\widehat{\sigma}_n}{\sqrt{A_n}} c_n(x) \text{ and } \Omega_{1,n}\right) = \exp(-2e^{-x})$$

Eventually, combined with (34) this implies (9). To finish the proof, it just remains to verify that (35) is satisfied. Indeed,

$$\eta_{3,n}\sqrt{\log n} \leq \varepsilon \frac{c_n(x)}{\sqrt{\log n}},$$

and after having replaced $c_n(x)$ by its expression (6), we can easily verify that

$$\lim_{n\to+\infty} \eta_{3,n}\sqrt{\log n} \overset{a.s}{=} 0.$$

This finishes the proof of Theorem 2. ♦

Proof of Corollary 2
Let D_2 and $\widehat{D}_2$ be defined respectively by (10) and (12), and set

$$\eta_{4,n} = \left| \max_{0\le k\le n-A_n} \frac{\sqrt{A_n}}{\nu} \left|D_2(k,A_n)\right| - \max_{0\le k\le n-A_n} \frac{\sqrt{A_n}}{\nu} \left|\widehat{D}_2(k,A_n)\right| \right|.$$

The key argument is to prove that

$$\lim_{n\to+\infty} \eta_{4,n}\sqrt{\log n} \overset{a.s}{=} 0. \tag{36}$$

by using assumption (13). Then, we apply Lemma 1 which combined with Corollary 1 implies (14). So, to finish the proof we must verify (36).
We have

$$0 \le \eta_{4,n} \le \frac{\sqrt{A_n}}{\nu} \max_{0\le k\le n-A_n} \left| D_2(k,A_n) - \widehat{D}_2(k,A_n) \right|$$

which implies

$$0 \le \eta_{4,n} \le \frac{\sqrt{A_n}}{\nu} \max_{0\le k\le n-A_n} \left| |\mu - \widehat{\mu}_k|^2 - |\mu - \widehat{\mu}_{k-A_n}|^2 \right|$$

and after

$$0 \le \eta_{4,n}\sqrt{\log n} \le \frac{2}{\nu} \max_{0\le k\le n-A_n} |\mu - \widehat{\mu}_k|^2 \sqrt{A_n \log n}.$$

Therefore, by using condition (13), we get

$$\lim_{n\to+\infty} \eta_{4,n}\sqrt{\log n} \overset{a.s}{=} 0.$$

This finishes the proof of Corollary 2. ■

B Proof for linear regression

Proof of Theorem 3
First we note that, under the null hypothesis (H_0), the sequence $(D_3(k,A_n))_{A_n\le k\le n-A_n}$ satisfy

$$AD_3(k,A_n) = S_k^{-2} \sum_{j=k+1}^{k+A_n} \left(X_j - \overline{X}_k\right)\varepsilon_j - S_{k-A_n}^{-2} \sum_{j=k-A_n+1}^{k} \left(X_j - \overline{X}_{k-A_n}\right)\varepsilon_j$$

where

$$\overline{X}_k = A_n^{-1} \sum_{j=k+1}^{k+A_n} X_j \quad \text{and} \quad S_k^2 = A_n^{-1} \sum_{j=k+1}^{k+A_n} \left(X_j - \overline{X}_k\right)^2,$$

are respectively the empirical mean and the empirical variance of X on the (sliding) box $[k+1,\, k+A_n]$. By using the definition (16), we see that

$$\overline{X}_k = \Delta\left(k + \frac{A_n+1}{2}\right) \qquad \text{and} \qquad S_k^2 = \Delta^2\left(\frac{A_n^2-1}{12}\right).$$

Therefore, we can deduce

$$D_3(k,A_n) = \frac{12}{\Delta \times A_n \times (A_n^2-1)} \sum_{j=k-A_n+1}^{k+A_n} \gamma(j-k,A_n)\varepsilon_j$$

where

$$\gamma(i,A_n)=\begin{cases} i-\frac{A_n+1}{2} & \text{if } i>0,\\ -i-\frac{A_n-1}{2} & \text{if } i\le 0.\end{cases} \tag{37}$$

Remark that the mean and the variance of the Gaussian sequence $(D_3(k,A_n))_{A_n\le k\le n-A_n}$ verify

$$\mathbb{E}\left[D_3(k,A_n)\right]=0 \qquad \text{and} \qquad Var\left[D_3(k,A_n)\right]=\frac{24\sigma^2}{\Delta^2\times A_n\times(A_n^2-1)}.$$

Moreover the variance does not depend on k. Then, $(D_3(k,A_n))_{A_n\le k\le n-A_n}$ is a centered stationary Gaussian sequence. Next, theorem 3 becomes an application of Csáki and Gonchigdanzan [13, Theorem 2.1, p. 3] which gives the asymptotic distribution of the maximum of standardized stationary Gaussian sequences $(Z_k)_{k\ge1}$ with covariance $r_n=\text{cov}\,(Z_1,Z_{n+1})$ under the condition $r_n\log(n)\to0$. Let us define the standardized version of D_3 as

$$D_3^{\text{Std}}(k,A)=\frac{D_3(k,A)}{\sqrt{\text{Var}\left[D_3(k,A)\right]}} \qquad \text{for } A_n\le k\le n-A_n, \tag{38}$$

and its covariance

$$\Gamma(k_1,k_2) \;=\; \text{cov}\left(D_3^{\text{Std}}(k_1,A),D_3^{\text{Std}}(k_2,A)\right).$$

The following lemma provides the value of the covariance:

Lemma 4
Let $\left(D_3^{Std}(k,A_n)\right)_{A_n\le k\le n-A_n}$ the standardized stationary Gaussian sequence defined by (38)*. Then, its covariance matrix denoted $(\Gamma(k_1,k_2))_{1\le k_1,k_2\le n}$ is given by*

$$\Gamma(k_1,k_2)=\frac{6}{A_n(A_n^2-1)}\times\begin{cases} f_1\left(|k_2-k_1|,A_n\right) & \text{if } 0\le|k_2-k_1|<A_n\\ f_2\left(|k_2-k_1|,A_n\right) & \text{if } A_n\le|k_2-k_1|\le 2A_n-1\\ 0 & \text{if } 2A_n-1<|k_2-k_1|<+\infty\end{cases}$$

where

$$f_1\left(p,A_n\right)=\frac{1}{6}(A_n-p)(A_n^2-2A_np-2p^2-1)+\frac{1}{12}p(3A_n^2+2p^2-6A_np+1)$$

and

$$f_2\left(p,A_n\right)=-\frac{1}{12}(2A_n-p)(A_n^2-10A_np-2p^2-1).$$

◊

Proof of Lemma 4
First we note that, by symmetry property of covariance matrix, we can restrict ourselves to the case $k_2-k_1\ge0$. Next, let us distinguish three different expressions of the covariance according to the value of k_2-k_1.

- If $k_2-k_1>2A_n-1$, then $\Gamma(k_1,k_2)=0$.
- If $0\le k_2-k_1\le 2A_n-1$, then

$$C^{-1}\Gamma(k_1,k_2)=\sum_{j=k_2-A_n+1}^{k_1+A_n}\gamma(j-k_1,A_n)\gamma(j-k_2,A_n) \text{ with } C=\frac{6}{A_n(A_n^2-1)}$$

By replacing $j-k_1$ with u, we get

$$C^{-1}\Gamma(k_1,k_2)=\sum_{u=-A_n+1}^{A_n-k_2+k_1}\gamma(u,A_n)\gamma(u+k_2-k_1,A_n).$$

But, in order to use formula (37) we must find the case where the sign of u and $u+k_2-k_1$ remain constant where $u\in[-A+1,A-k_2+k_1]$. That is why we must distinguish the following subcases

– If $0 \leq k_2 - k_1 \leq A_n - 1$, then

$$\begin{aligned} C^{-1}\Gamma(k_1,k_2) &= \sum_{u=-A_n+1}^{k_1-k_2} \gamma(\underbrace{u}_{\leq 0}, A_n)\gamma(\underbrace{u+k_2-k_1}_{\leq 0}, A_n) \\ &+ \sum_{u=k_1-k_2+1}^{0} \gamma(\underbrace{u}_{\leq 0}, A_n)\gamma(\underbrace{u+k_2-k_1}_{>0}, A_n) \\ &+ \sum_{u=1}^{A_n-k_2+k_1} \gamma(\underbrace{u}_{>0}, A_n)\gamma(\underbrace{u+k_2-k_1}_{>0}, A_n) \end{aligned}$$

and then $C^{-1}\Gamma_{k_1,k_2} = f_1\left(k_2 - k_1, A_n\right)$.

– If $A_n \leq k_2 - k_1 \leq 2A_n - 1$, then

$$C^{-1}\Gamma(k_1,k_2) = \sum_{u=-A_n+1}^{A_n-k_2+k_1} \gamma(\underbrace{u}_{\leq 0}, A_n)\gamma(\underbrace{u+k_2-k_1}_{>0}, A_n)$$

which implies $C^{-1}\Gamma(k_1,k_2) = f_2\left(k_2 - k_1, A_n\right)$.

This finishes the proof of Lemma 4. ♦

Hence, by applying Theorem 2.1 of Csáki and Gonchigdanzan to the sequence $\left(D_3^{\mathrm{Std}}(k,A_n)\right)_{A_n \leq k \leq n-A_n}$, we can deduce that

$$\lim_{n\to+\infty} \mathbb{P}\left(\max_{k\in[A_n:n-A_n]} |D_3^{\mathrm{Std}}(k,A_n)| \leq d_n(x) \right) = \exp(-2e^{-x}). \tag{39}$$

Then, by using (38), we obtain (19). This finishes the proof of Theorem 3. ■

Proof of Corollary 3

First we note that, under the null hypothesis (H_0), the Filtered Derivative applied to the intercept satisfy

$$D_4(k,A_n) = A_n^{-1} \sum_{j=k+1}^{k+A_n} e_j - A_n^{-1} \sum_{j=k-A_n+1}^{k} e_j.$$

Therefore, the sequence $(D_4(k,A_n))_{A_n\leq k\leq n-A_n}$ corresponds to a sequence of Filtered Derivative applied to the mean for a sample of independent and identically distributed Gaussian random variables with zero-mean and variance σ^2. Then, by applying Theorem 1 under assumption $\mathcal{A}_1$, we obtain (22). This finishes the proof of Corollary 3. ■

Proof of Corollary 4

Let D_4 and $\widehat{D}_4$ be defined respectively by (20) and (23), and set

$$\eta_{5,n} = \left| \max_{0\leq k\leq n-A_n} \frac{\sqrt{A_n}}{\sigma} |D_4(k,A_n)| - \max_{0\leq k\leq n-A_n} \frac{\sqrt{A_n}}{\sigma} \left|\widehat{D}_4(k,A_n)\right| \right|.$$

The key argument is to prove that

$$\lim_{n\to+\infty} \eta_{5,n}\sqrt{\log n} \overset{a.s}{=} 0. \tag{40}$$

By using assumption (24). Then, we apply Lemma 1 which combined with Corollary 3 implies (25). So, to finish the proof we must verify (40). Next, we have

$$0 \leq \eta_{5,n} \leq \frac{\sqrt{A_n}}{\sigma} \max_{0\leq k\leq n-A_n} \left| D_4(k,A_n) - \widehat{D}_4(k,A_n) \right|$$

which implies

$$0 \leq \eta_{5,n} \leq \frac{1}{\sqrt{A_n}\sigma} |a - \widehat{a}_n| \max_{0 \leq k \leq n - A_n} \left| \sum_{j=k+1}^{k+A_n} X_j - \sum_{j=k-A_n+1}^{k} X_j \right|.$$

Then, by using (16), we show that

$$0 \leq \eta_{5,n} \leq \frac{\Delta_n}{\sqrt{A_n}\sigma} |a - \widehat{a}_n| \max_{0 \leq k \leq n - A_n} \underbrace{\sum_{j=k-A_n+1}^{k+A_n} j}_{=A_n^2}$$

and after

$$\eta_{5,n} \sqrt{\log n} \leq |a - \hat{a}_n| A_n^{\frac{3}{2}} \Delta_n \sqrt{\log n}.$$

Therefore, by using condition (24), we get

$$\lim_{n \to +\infty} \eta_{5,n} \sqrt{\log n} \overset{a.s}{=} 0.$$

This finishes the proof of Corollary 4. ■