



**HAL**  
open science

# A criterion for estimating the largest linear homoscedastic region in Gaussian data

Jean-Marc Bardet

► **To cite this version:**

Jean-Marc Bardet. A criterion for estimating the largest linear homoscedastic region in Gaussian data. 2023. hal-04062059

**HAL Id: hal-04062059**

**<https://hal.science/hal-04062059>**

Preprint submitted on 7 Apr 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Public Domain

# A criterion for estimating the largest linear homoscedastic region in Gaussian data

Jean-Marc Bardet<sup>1</sup>

bardet@univ-paris1.fr

<sup>1</sup> SAMM, Université Paris 1, 90 rue de Tolbiac, 75013 Paris, FRANCE

April 7, 2023

## Abstract

We construct a criterion to identify the largest homoscedastic region in a data set. This can be reduced to a one-sided non-parametric break detection, knowing that up to a certain index the output is governed by a linear homoscedastic model, while after this index it is different (*e.g.* a different model, different variables, different volatility, ...). We show the convergence of the estimator of this index, with asymptotic concentration inequalities that can be exponential. Monte Carlo experiments will also confirm its very good numerical performance.

*Keywords:* Change detection; Gaussian linear model; Model selection; Hydrography

## 1 Introduction

The idea of this work has its roots in a numerical work done in physical geography, but which could have been done in other fields of application. The aim was to compute the fractal dimension of networks of ravines and gullies in the French Alps (see Figure 1). For this purpose, we chose to use the correlation dimension introduced by Grassberger and Proccacia (see [8]), which consists, for a "small" radius  $r$ , in calculating the log proportion  $\log(C(r))$  of pairs of points  $(X_i, X_j)$  such that the distance between  $X_i$  and  $X_j$  is less than  $r$ . And it is known that if  $\lim_{r \rightarrow 0} \log(C(r))/\log(1/r) = D$  then  $D$  is the correlation dimension, which is a fractal dimension often equal to the Hausdorff dimension. To estimate such a fractal dimension, a log-log regression is usually performed by considering a cloud of points  $(\log(r_i), \log(C(r_i)))_{1 \leq i \leq n}$  (see Figure 1):

Consider this point cloud. It is clear that if  $r$  is too small (typically smaller than the smallest of the interdistances between the  $X_i$ ) the fraction  $C(r)$  tends to 0 and if  $r$  is too large the fraction  $C(r)$  tends to 1 by jumping in steps. The idea to estimate  $D$  will be to consider a "good" linearity zone, for example the one that at first sight would give for  $\log(r) \in [2, 4]$  (we will find  $[2.3, 3.4]$  after applying our criterion). The question is: how to find this optimal linearity zone with statistical criteria?

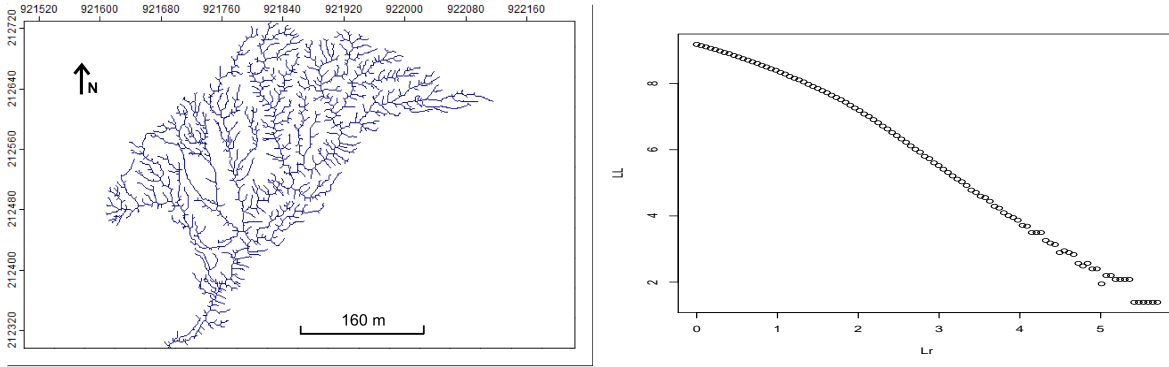


Figure 1: Network of ravines and gullies in the French Alps (left) and plot of  $\log(C(r_i))$  versus  $\log(r_i)$  where  $1 \leq r_i \leq 250$  in meters (right).

This question of the largest linearity zone is, of course, close to the problem of break detection. This problem can be simplified as follows: the outputs  $Y_i$  (real-valued) depend linearly on  $p$  exogenous variables  $Z_i^{(j)}$ ,  $j = 1, \dots, p$  with homoscedastic errors, for an index  $i \leq n$ , where  $n$  is unknown, while for  $i \geq n + 1$  this linear model is no longer valid. Much has been written about detecting a break in a linear model. The first thing that comes to mind is a Fisher test or what economists call a Chow test (see [6] or more generally in [3]). This consists of setting up a first linear model up to a certain index (e.g. individual or time) and then a second linear model after this index. The detection of the break index is done by minimizing the double sum of residual squares, which is also the Gaussian maximum likelihood estimator. This can also be done by minimizing the sum of the absolute residuals, as in [2].

But this framework is much more restrictive than the one we were considering. We would rather have a non-parametric breakthrough detection. We can then refer to work such as that of Carlstein [5] or Ben Hariz *et al.* [4]. Another method of non-parametric detection is to construct a criterion based on kernel density estimation, as in Arlot *et al.* [1]. However, these works assume an identical distribution before and after the break, which is not at all the case we are interested in (even in the case of  $i \leq t$ , which is the case of the linear model).

We have therefore constructed an ad hoc criterion for detecting this larger linear region, which applies both when there is one break (typically something other than the linear model after an  $n$  index) and when there are two breaks (other than the linear model before an  $n_1$  index and after an  $n_2$  index). The assumptions made outside the linear homoscedastic zone are completely nonparametric and are equally suited to an abrupt change in the model as a function of the exogenous variables or an abrupt change in the homoscedasticity of the model.

The criterion, whose precise definition is given in (3.7) in the case of one break and in (3.13) in the case of two breaks, is constructed from a minimisation of the residual sum of squares, carefully renormalised so as to choose the largest possible homoscedastic linear region, while increasing as soon as the linear region is exceeded. Under rather general assumptions, to which is added that of Gaussianity, asymptotic concentration inequalities at the breakpoint are obtained and the asymptotic normality of the least squares estimators of the linear model is also proved.

Finally, Monte Carlo experiments are carried out to demonstrate the accuracy of the new criterion. We first consider the case of a single break and, after optimizing the meta-parameters of the criterion, we place ourselves in different frameworks of changes in linearity or changes in volatility. The results are very convincing and the convergence as  $n$  increases towards the breakpoint is clear. By moving to the framework of a zone of homoscedastic linearity between 2 breakpoints  $n_1$  and  $n_2$ , the results of the simulations show the high quality of the performance of the criterion. Finally, the application to the data set of Figure 1 allows to identify a very convincing zone of homoscedastic linearity and to obtain a fractal dimension of 1.74.

The following section 2 is devoted to the assumptions. The definitions of the criteria and the main results are given in section 3. Numerical experiments are proposed in section 4 and proofs are established in section 5.

## 2 Assumptions

Let  $(\varepsilon_i)_{i \in \mathbb{N}}$  be a sequence of random variables such as  $\mathbb{E}(\varepsilon_i) = 0$  and  $\mathbb{E}(\varepsilon_i^2) = \sigma^2 < \infty$  for all  $i \in \mathbb{N}$ . Let  $p \in \mathbb{N}^*$ , known, and  $(Z_i^{(j)})_{1 \leq j \leq p, i \in \mathbb{N}}$  a family of known reals numbers. Let  $N \in \mathbb{N}^*$ ,  $N \geq p + 1$  such as  $\text{rank}((Z_i^{(j)})_{1 \leq j \leq p, 1 \leq i \leq N}) = p$ . Let  $(Y_i)_{i \in \mathbb{N}}$  be a sequence of random variables and assume that  $(Y_1, \dots, Y_N)$  is observed. In the sequel we will assume:

**Assumption A:** *there exists  $1 \leq n \leq N - 1$ , an unknown integer number such as*

$$Y_i = \sum_{j=1}^p \theta_j Z_i^{(j)} + \varepsilon_i \quad \text{for all } i \in \{1, \dots, n\}, \quad (2.1)$$

where  $\theta = (\theta_1, \dots, \theta_p)' \in \mathbb{R}^p$  is unknown.

The aim of your study is the estimation of  $n$  and  $\theta$ . For  $p \leq t \leq N$ , we define the ordinary least squares estimator (OLSE)  $\hat{\theta}_t$  of  $\theta$

$$\hat{\theta}_t := (Z(t)' Z(t))^{-1} Z(t)' Y(t) \quad (2.2)$$

where, for  $p \leq t \leq N$ ,

- $Z(t) := (Z_i^{(j)})_{1 \leq i \leq t, 1 \leq j \leq p}$  and we assume that  $Z(t)$  is a regular matrix (*i.e.* its rank is  $p$ );
- $Y(t) := (Y_i)_{1 \leq i \leq t}$  and  $\varepsilon(t) := (\varepsilon_i)_{1 \leq i \leq t}$ .

Even if the following definitions of the criteria are valid for any  $p \leq n \leq N$ , an additional and usual assumption will be added for obtaining asymptotic results:

**Assumption B:** *there exists  $0 < \tau < 1$ , an unknown real numbers such as for any  $N \in \mathbb{N}^*$ ,*

$$n = \lceil \tau N \rceil.$$

In the sequel, we will consider a general scenario of change for the time series  $(Y_i)_{1 \leq i \leq N}$ , which satisfies Assumption A. Indeed, we assume

**Assumption C:** *There exists a sequence of random variables  $(\delta_t)_{t \in \mathbb{N}}$ , independent to  $(\varepsilon_t)_{t \in \mathbb{N}}$  and satisfying*

$$Y_i = \sum_{j=1}^p \theta_j Z_i^{(j)} + \varepsilon_i + \delta_i \quad \text{for all } i \in \{n+1, \dots, N\} \quad (2.3)$$

and there exists  $C > 0$  and a positive random variable  $k_0$  such as for any  $k \geq k_0$

$$\Delta_k := \sum_{i=n+1}^{n+k} \delta_i^2 \geq Ck \quad \text{and} \quad \max_{n+1 \leq i \leq n+k} \delta_i^2 \leq \frac{C}{2p}k \quad \text{almost surely.} \quad (2.4)$$

The framework of (2.3) and more particularly of (2.4) is very general and may correspond to:

- A non-parametric change of the trend of  $(Y_i)$ . A typical non-parametric example is the following:  $Y_i = \theta_0 + \theta_1 x_i + \varepsilon_i$  for  $i = 1, \dots, n$  and  $Y_i = g(i) + \varepsilon_i$  for  $i = n+1, \dots, N$ , with  $|g(n+k) - (\theta_0 + \theta_1 x_{n+k})| \geq C > 0$  for  $k \in \mathbb{N}^*$ .
- A non-parametric change of variance of  $(Y_i)_{1 \leq i \leq N}$ , which corresponds for instance to an increasing of the volatility. A typical non-parametric example is the following:  $Y_i = \theta_0 + \theta_1 x_i + \varepsilon_i$  for  $i = 1, \dots, n$  and  $Y_i = \theta_0 + \theta_1 x_i + \xi_i$  for  $i = n+1, \dots, N$ , with  $\mathbb{E}(\xi_i) = 0$  and  $\sum_{i=n+1}^{n+k} \text{Var}(\xi_i) > k \sigma^2$  for  $k$  large enough.

### 3 Detecting change with a criterion of selection of individuals

#### 3.1 Definition of a criterion of individual selection

First, define for  $k \in \{p, \dots, N\}$ ,

- for any  $u \in \mathbb{R}^k$ ,  $\|u\|^2 = u' u$  where  $u'$  denotes the transpose of  $u$ .
- $[Z]_k := \{Z(k)\theta, \theta \in \mathbb{R}^p\}$  which is a linear subspace of  $\mathbb{R}^k$ ;
- $H(k) := Z(k)(Z'(k)Z(k))^{-1}Z'(k)$  which is the orthogonal projection matrix on  $[Z]_k$ ;

For estimating  $n$  under the previous assumptions, we consider a new criterion,  $\widehat{Cr}_a$ :

**Definition of  $\widehat{Cr}_a$ :** Let  $a : \mathbb{N} \mapsto ]0, \infty[$  be a decreasing function and for  $p+1 \leq t \leq N$  define

$$\widehat{Cr}_a(t) = \frac{a(t)}{t-p} \|Y(t) - H(t)Y(t)\|^2. \quad (3.1)$$

It is well known that under Assumption A and for  $t \leq n$ ,

$$\widehat{Cr}_a(t) = a(t) \widehat{\sigma}^2(t) \quad \text{where} \quad \widehat{\sigma}^2(t) := \frac{1}{t-p} \|Y(t) - \widehat{Y}(t)\|^2 \quad \text{and} \quad \widehat{Y}(t) = H(t)Y(t),$$

the usual estimated variance and predicted vector obtained from ordinary least square estimation.

From this definition, we define the following estimators:

$$\widehat{n} := \arg \min_{p+1 \leq t \leq N} \widehat{Cr}(t) \quad \text{and} \quad \widehat{\tau} := \frac{\widehat{n}}{N}, \quad (3.2)$$

since we chose the criterion to be minimized.

As a consequence the function  $a(\cdot)$  should be a decreasing function because we would like to select the "largest" acceptable linear zona  $\{1, \dots, t\}$ . Indeed, if  $t \leq n$ , we would like that  $\widehat{Cr}_a(t+k) < \widehat{Cr}_a(t)$  for any  $k \in \{1, \dots, n-t\}$  with high probability. But it is well known that under Assumption A,  $\widehat{Cr}_a(t) \sim a(t) \sigma^2$  almost surely when  $n \rightarrow \infty$  and  $t \rightarrow \infty$  with  $t \leq n$  and this implies that  $a(\cdot)$  has to be a decreasing function.

However we can not chose any decreasing function  $a(\cdot)$  since Assumption C holds. Indeed, for  $t \geq n+1$ , we expect that the probability that  $\widehat{Cr}_a(t) \leq \widehat{Cr}_a(n)$  becomes small when  $t-n$  increases. A first answer to this question can be given by considering the following computation. Indeed, we have

$$\bullet \text{ for } t \leq n, \quad \widehat{Cr}_a(t) = \frac{a(t)}{t-p} \|(I(t) - H(t)) \varepsilon(t)\|^2 \quad (3.3)$$

$$\bullet \text{ for } t \geq n+1, \quad \widehat{Cr}_a(t) = \frac{a(t)}{t-p} \|(I(t) - H(t)) \varepsilon(t) + (I(t) - H(t)) D(t)\|^2 \quad (3.4)$$

where  $I(t)$  is the identity matrix of size  $t$ , and

$$D(t) = (0, \dots, 0, \delta_{n+1}, \dots, \delta_t)'. \quad (3.5)$$

**Lemma 1.** *Using Assumptions A, B and C, there exists a positive random variable  $n_0$  such as for any  $t \geq n+1$  and  $n \geq n_0$ ,*

$$\|(I(t) - H(t)) D(t)\|^2 \geq \frac{C}{2}(t-n).$$

A first condition on the function  $a(\cdot)$  can be established by considering the expectation of the criterion. Indeed, for detecting the change, we expect:  $\mathbb{E}(\widehat{Cr}(n+k+1)) > \mathbb{E}(\widehat{Cr}_a(n+k))$  for any  $k \geq 0$ .

Thus, with  $F_k = \mathbb{E}(\|(I(n+k) - H(n+k)) D(n+k)\|^2)$ , this condition implies:

$$\begin{aligned} a(n+k+1) \left( \sigma^2 + \frac{F_{k+1}}{n+k+1-p} \right) &> a(n+k) \left( \sigma^2 + \frac{F_k}{n+k-p} \right) \\ \implies a(n+k+1) &> a(n+k) \left( 1 + \frac{F_k}{\sigma^2(n+k-p)} \right) \left( 1 + \frac{F_{k+1}}{\sigma^2(n+k-p)} \right)^{-1} \end{aligned}$$

Using approximations, we deduce that there exists  $c_k$  a bounded sequence such as

$$a(n+k+1) \geq a(n+k) \left( 1 - \frac{c_k}{(n+k-p)} \right). \quad (3.6)$$

Hence we chose a decreasing sequence  $(a(t))_{t \in \mathbb{N}^*}$  satisfying (3.6) and our choice is  $(a_{\alpha, \beta}(t))_{t \geq p+1}$  where  $\beta > 1$ ,  $\beta - \frac{1}{2} < \alpha < \beta$  defined by

$$\begin{aligned} a_{\alpha, \beta}(t) &:= 1 + \frac{N^\alpha}{(t-p)^\beta} \text{ for any } t \in \{p, \dots, N\} \\ \implies \widehat{Cr}_{\alpha, \beta}(t) &:= \frac{1}{t-p} \left( 1 + \frac{N^\alpha}{(t-p)^\beta} \right) \|Y(t) - H(t) Y(t)\|^2. \quad (3.7) \end{aligned}$$

The term  $N^\alpha$  and the relations  $\beta > 1$  and  $\frac{1}{2} < \beta - \frac{1}{2} < \alpha < \beta$  will be explained with details by the proof of the forthcoming Theorem 3.1. Roughly speaking, we could say that  $t \mapsto (t-p)^{-\beta}$

is the decreasing function and  $N^\alpha$  with  $\alpha > 0$  insures that the probability to select a "small"  $\hat{n}$  decreases to 0 when  $N \rightarrow \infty$  while if  $\alpha = 0$  this is possible. Moreover, after applying usual Taylor expansions, we have for  $t \rightarrow \infty$  such as  $t^\beta N^{-\alpha} \rightarrow \infty$  (which is implied by  $t \geq n$  and  $n = \lceil \tau N \rceil$  from Assumption B),

$$\frac{a_{\alpha,\beta}(t+1)}{a_{\alpha,\beta}(t)} = 1 - \beta \frac{N^\alpha}{(t-p)^{\beta+1}} (1 + o(1)).$$

and thus (3.6) is satisfied.

**Remark 1.** Another possible choice of a decreasing sequence  $(a(t))_{t \in \mathbb{N}^*}$  satisfying (3.6) could be  $a(t) = \exp(c N^\alpha t^{-\beta})$  with  $c > 0$ . From the proof of Theorem 3.1, such a choice makes quite impossible a selection of a "small"  $\hat{n}$ . However such a choice could imply some numerical difficulties since the computation of  $\exp(N^\alpha)$  with  $\alpha > 1/2$  is explosive when  $N$  increases.

### 3.2 Consistency of the estimators

Now, the consistency of the criterion can be established:

**Theorem 3.1.** Let  $\hat{n}$ ,  $\hat{\tau}$  obtained in (3.2) from the criterion  $\widehat{Cr}_{\alpha,\beta}$  defined in (3.7) with  $1 < \alpha < \beta < \alpha + \frac{\alpha}{2\beta} < \alpha + \frac{1}{2}$ . Then, under Assumptions A, B and C and if  $(\varepsilon_t)_t$  is a Gaussian sequence, for any  $0 < \gamma < 1 - 2(\beta - \alpha)$ ,

$$\mathbb{P}(|\hat{\tau} - \tau| \geq N^{-\gamma}) \xrightarrow{N \rightarrow \infty} 0. \quad (3.8)$$

From this theorem, we deduce that the smaller  $\beta - \alpha$  the larger  $\gamma$  and the more accurate the estimation  $\hat{n}$ . Moreover, intermediate relationships (5.4), (5.7) and (5.8) imply that the larger  $\alpha$  and  $\beta$  the more accurate the estimation  $\hat{n}$ . However, simulations in Section 4 will temperate these purposes.

**Corollary 1.** Under the assumptions of Theorem 3.1, if we consider  $\tilde{n} = \arg \min_{N^\varepsilon \leq t \leq N} \widehat{Cr}_{\alpha,\beta}(t)$  with  $0 < \varepsilon < 1$ , there exists  $C > 0$  and  $\rho > 0$  such as

$$\mathbb{P}(|\tilde{n} - n| \geq N^{1-\gamma}) \leq e^{-CN^\rho}. \quad (3.9)$$

**Corollary 2.** Define  $\tilde{\theta}_N = \hat{\theta}_{\hat{n}-N^\delta}$  with  $1 - 2(\beta - \alpha) < \delta < 1$ , where  $\hat{\theta}_t$  is defined in (2.2). Under the assumptions of Theorem 3.1, then:

$$\tilde{\theta}_N \xrightarrow[N \rightarrow \infty]{\mathcal{P}} \theta \quad \text{and} \quad (Z(\hat{n} - N^\delta)' Z(\hat{n} - N^\delta))^{1/2} (\tilde{\theta}_N - \theta) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2). \quad (3.10)$$

The choice of  $\delta$  when  $1 - 2(\beta - \alpha) < \delta < 1$  in the previous corollary is asymptotically not crucial since it does not modify the convergence rate of the estimator of  $\theta$  which is almost given by  $(Z(n)' Z(n))^{1/2}$ . A possible choice is  $\delta = 1 - (\beta - \alpha)$ .

### 3.3 Application to the detection of the largest homoscedastic linear zone in a sequence of data indexed by time

We now consider the following problem:

**Assumption A'**: there exist two **unknown** change instants  $n_1$  and  $n_2$  in the law of  $(Y_1, \dots, Y_n)$  and a **known** integer number  $n_0$  such as  $n_0 \in \{n_1, \dots, n_2 - 1\}$  and

$$Y_i = \sum_{j=1}^p \theta_j Z_i^{(j)} + \varepsilon_i \quad \text{for all } i \in \{n_1, n_1 + 1, \dots, n_2 - 1\},$$

where  $\theta = (\theta_1, \dots, \theta_p)' \in \mathbb{R}^p$  is unknown.

**Remark:** Knowing such  $n_0$  is crucial to avoid selecting a potential linear region contained in  $\{1, \dots, n_1 - 1\}$  or in  $\{n_2, \dots, N\}$ .

The previous Assumptions B and C can be transform according to this new framework. Thus we suppose:

**Assumption B'**: there exists  $0 < \tau_1 < \tau_2 < 1$ , two unknown real numbers, and  $\tau_0$  a known real number such as  $\tau_1 < \tau_0 < \tau_2$ , satisfying for any  $N \in \mathbb{N}^*$ ,

$$n_1 = [\tau_1 N], \quad n_0 = [\tau_1 N] \quad \text{and} \quad n_2 = [\tau_2 N].$$

**Assumption C'**: there exists two sequences of random variables  $(\delta_t^{(1)})_{t \in \mathbb{N}}$  and  $(\delta_t^{(2)})_{t \in \mathbb{N}}$ , independent to  $(\varepsilon_t)_{t \in \mathbb{N}}$  and satisfying

$$\begin{cases} Y_i = \sum_{j=1}^p \theta_j Z_i^{(j)} + \varepsilon_i + \delta_i^{(1)} & \text{for all } i \in \{1, \dots, n_1 - 1\} \\ Y_i = \sum_{j=1}^p \theta_j Z_i^{(j)} + \varepsilon_i + \delta_i^{(2)} & \text{for all } i \in \{n_2, \dots, N\} \end{cases} \quad (3.11)$$

and there exists  $C > 0$  and a positive random variable  $k_0$  such as for any  $k \geq k_0$

$$\sum_{i=1}^k (\delta_{n_1-i}^{(1)})^2 \geq Ck \quad \text{and} \quad \sum_{i=1}^k (\delta_{n_2+i}^{(2)})^2 \geq Ck \quad \text{almost surely.} \quad (3.12)$$

Then, under those assumptions, it is possible to prove that an appropriated version of the new criterion  $\widehat{Cr}'_{\alpha, \beta}$  is able to determinate the "largest linear zone" (in time) in  $\{1, \dots, N\}$ .

**Theorem 3.2.** Under Assumptions A', B' and C', define for  $1 \leq t_1 \leq n_0 \leq t_2 - p \leq N$ :

$$\widehat{Cr}'(t_1, t_2) := \frac{a_{\alpha, \beta}(t_2 - t_1)}{t_2 - t_1 - p} \left\| Y(t_1, t_2) - H(t_1, t_2) Y(t_1, t_2) \right\|^2, \quad (3.13)$$

where  $Y(t_1, t_2) := (Y_i)_{t_1 \leq i \leq t_2}$ ,  $H(t_1, t_2) := Z(t_1, t_2)(Z'(t_1, t_2)Z(t_1, t_2))^{-1}Z'(t_1, t_2)$  and  $Z(t_1, t_2) := (Z_i^{(j)})_{t_1 \leq i \leq t_2, 1 \leq j \leq p}$  (assume that  $Z(t_1, t_2)$  is a regular matrix for any  $t_1 < t_2 + p$ ). Then, define

$$(\widehat{n}_1, \widehat{n}_2) := \arg \min_{1 \leq t_1 \leq n_0 \leq t_2 - p \leq N} \widehat{Cr}'(t_1, t_2). \quad (3.14)$$

Then for any  $0 < \gamma < 1 - 2(\beta - \alpha)$ ,

$$\mathbb{P}\left(|\widehat{n}_1 - n_1| + |\widehat{n}_2 - n_2| \geq N^{1-\gamma}\right) \xrightarrow{N \rightarrow \infty} 0. \quad (3.15)$$

Now, as in the previous subsection, we could add two corollaries to this theorem:



**Corollary 3.** Under the assumptions of Theorem 3.2, if we consider  $(\tilde{n}_1, \tilde{n}_2) := \arg \min_{N^\varepsilon \leq t_1 \leq n_0 \leq t_2 - p \leq N - N^\varepsilon} \widehat{Cr}'(t_1)$  with  $0 < \varepsilon < 1$ , there exists  $C > 0$  and  $\rho > 0$  such as

$$\mathbb{P}(|\tilde{n}_1 - n_1| + |\tilde{n}_2 - n_2| \geq N^{1-\gamma}) \leq e^{-CN^\rho}. \quad (3.16)$$

**Corollary 4.** For any  $\delta$  such as  $1 - 2(\beta - \alpha) < \delta < 1$ , define

$$\tilde{\theta}_N = (Z(\hat{n}_1 + N^\delta, \hat{n}_2 - N^\delta)' Z(\hat{n}_1 + N^\delta, \hat{n}_2 - N^\delta))^{-1} Z(\hat{n}_1 + N^\delta, \hat{n}_2 - N^\delta)' Y(\hat{n}_1 + N^\delta, \hat{n}_2 - N^\delta).$$

Then, under the assumptions of Theorem 3.1, then:

$$\tilde{\theta}_N \xrightarrow[N \rightarrow \infty]{\mathcal{P}} \theta \quad \text{and} \quad (Z(\hat{n}_1 + N^\delta, \hat{n}_2 - N^\delta)' Z(\hat{n}_1 + N^\delta, \hat{n}_2 - N^\delta))^{1/2} (\tilde{\theta}_N - \theta) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$

## 4 Numerical applications

### 4.1 Case of a linear homoscedastic zone with break on one side

In this first subsection, we consider the following framework:  $(Y_1, \dots, Y_N)$  is observed, and there exists  $n$ , unknown integer number, such as

$$Y_i = \begin{cases} \sum_{j=1}^p \theta_j Z_i^{(j)} + \varepsilon_i & \text{for all } i \in \{1, \dots, n\}; \\ \sum_{j=1}^p \theta_j Z_i^{(j)} + \varepsilon_i + \delta_i & \text{for all } i \in \{n+1, \dots, N\}. \end{cases}$$

where  $\theta = (\theta_1, \dots, \theta_p)' \in \mathbb{R}^p$  is unknown. More precisely, we will consider the following particular cases:

- The size  $N$  of the observed sample is such as:  $N = 100, 200, 400, 800$  and  $1600$ .
- The error  $(\varepsilon_i)$  of the homoscedastic linear model is a sequence of Gaussian centered i.i.d.r.v. with  $\sigma^2 = \text{Var}(\varepsilon_0)$  is such as  $\sigma = 1$  and  $\sigma = 2$ .
- The unknown change date  $n$  is such as  $n = [0.4 \times N]$ .
- Three distinct frameworks are considered:

1. **(M1)** Trend break in a time series:

$$Y_i = \begin{cases} 3 + 0.1 \times i + \varepsilon_i & \text{for all } i \in \{1, \dots, n\}; \\ 3 + 0.1 \times i + (2\sqrt{3 + (i-n)} - 4) + \varepsilon_i & \text{for all } i \in \{n+1, \dots, N\}. \end{cases}$$

2. **(M2)** Time series volatility break:

$$Y_i = \begin{cases} 3 + 0.1 \times i + \varepsilon_i & \text{for all } i \in \{1, \dots, n\}; \\ 3 + 0.1 \times i + 2 * \varepsilon_i & \text{for all } i \in \{n+1, \dots, N\}. \end{cases}$$

3. **(M3)** Model change for an homoscedastic linear model:

$$Y_i = \begin{cases} -4 + 2 Z_i^{(1)} - 5 Z_i^{(2)} + 3 Z_i^{(3)} + \varepsilon_i & \text{for all } i \in \{1, \dots, n\}; \\ -8 - Z_i^{(1)} + 3 \log(|Z_i^{(3)}|) + \varepsilon_i & \text{for all } i \in \{n+1, \dots, N\}. \end{cases}$$

where  $(Z_i^{(1)})$  and  $(Z_i^{(3)})$  are observed i.i.d.r.v. following respectively a uniform distribution  $\mathcal{U}([-1, 1])$  and a Student distribution  $t(4)$ , and  $Z_i^{(2)} = (Z_i^{(1)})^2$ .

	$N$	100	200	400	800	1600
(M1)	$\sigma = 1$	0.1026	0.0695	0.0472	0.0311	0.0202
	$\sigma = 2$	0.1851	0.0728	0.0480	0.0313	0.0213
(M2)	$\sigma = 1$	0.1037	0.0672	0.0515	0.0341	0.0213
	$\sigma = 2$	0.1018	0.0702	0.0486	0.0325	0.0233
(M3)	$\sigma = 1$	0.0972	0.0665	0.0456	0.0330	0.0211
	$\sigma = 2$	0.0975	0.0710	0.0423	0.0308	0.0184

Table 1: Square roots of the MSE for the estimators  $\hat{\tau}$  minimizing the criterion  $\widehat{Cr}$  for the three different models (M1), (M2) and (M3), and the different values of parameters  $\sigma$  and  $N$ .

We applied the  $\widehat{Cr}_{\alpha,\beta}$  criterion defined in (3.7). After an important empirical study not reproduced here, we have decided on the parameters  $\alpha = 2$  and  $\beta = 2.1$ , and we have added a constant to compensate for the fact that  $n$  is not very large (which does not change the asymptotic results of the previous section). Hence the chosen criterion is:

$$\widehat{Cr}(t) = \frac{1}{t-p} \left( 1 + \frac{N^2}{10(t-p)^{2.1}} \right) \|Y(t) - H(t)Y(t)\|^2.$$

After performing 1000 independent replications of the scatterplot for the different models and parameters, and applying the largest linear area detection criterion  $\widehat{Cr}$ , we obtained the results recorded in Table 1:

**Conclusions of Monte Carlo experiments:** We first observe the consistency of the estimator  $\hat{\tau}$  as  $N$  becomes large. And the rate of convergence is close to  $\sqrt{N}$ . We also note two important points:

- The variance  $\sigma^2$  of the noise has very little effect on the rate of convergence of the estimator, except in one case;
- For a fixed  $N$ , the square root of the MSE is almost the same whatever model is chosen. And for all the other different models we tried, we also find almost the same values!

## 4.2 Case of a linear homoscedastic zone with breaks on both sides

Now we consider the following framework:  $(Y_1, \dots, Y_N)$  is observed, and there exists  $n_1$  and  $n_2$ , unknown integer numbers, such as

$$Y_i = \begin{cases} \sum_{j=1}^p \theta_j Z_i^{(j)} + \varepsilon_i & \text{for all } i \in \{n_1, \dots, n_2 - 1\}; \\ \sum_{j=1}^p \theta_j Z_i^{(j)} + \varepsilon_i + \delta_i^{(1)} & \text{for all } i \in \{1, \dots, n_1 - 1\} \\ \sum_{j=1}^p \theta_j Z_i^{(j)} + \varepsilon_i + \delta_i^{(2)} & \text{for all } i \in \{n_2, \dots, N\}. \end{cases}$$

where  $\theta = (\theta_1, \dots, \theta_p)' \in \mathbb{R}^p$  is unknown. More precisely, we will consider the following particular cases:

	$N$	100		200		400		800	
		$\widehat{\tau}_1$	$\widehat{\tau}_2$	$\widehat{\tau}_1$	$\widehat{\tau}_2$	$\widehat{\tau}_1$	$\widehat{\tau}_2$	$\widehat{\tau}_1$	$\widehat{\tau}_2$
(MM1)	$\sigma = 1$	0.0696	0.1107	0.0563	0.0930	0.0315	0.0555	0.0220	0.0405
	$\sigma = 2$	0.0735	0.1463	0.0516	0.1279	0.0382	0.1020	0.0258	0.0775
(MM2)	$\sigma = 1$	0.0492	0.0671	0.0309	0.0383	0.0146	0.0140	0.0112	0.0057
	$\sigma = 2$	0.0574	0.1550	0.0362	0.0811	0.0185	0.0183	0.0121	0.0080

Table 2: Square roots of the MSE for the estimators  $\widehat{\tau}_1$  and  $\widehat{\tau}_2$  minimizing the criterion  $\widehat{C}r'$  for the two different models (MM1) and (MM2), and the different values of parameters  $\sigma$  and  $N$ .

- The size  $N$  of the observed sample is such as:  $N = 100, 200, 400$  and  $800$ .
- The error ( $\varepsilon_i$ ) of the homoscedastic linear model is a sequence of Gaussian centered i.i.d.r.v. with  $\sigma^2 = \text{Var}(\varepsilon_0)$  is such as  $\sigma = 1$  and  $\sigma = 2$ .
- The unknown change dates  $n_1$  and  $n_2$  are such as  $n_1 = [0.4 \times N]$  and  $n_2 = [0.7 \times N]$ , and  $n_0 = [0.5 \times N]$  is known.
- Two distinct frameworks are considered:

1. **(MM1)** Trend and volatility breaks in a time series:

$$Y_i = \begin{cases} 3 + 0.1 \times i + 2 * \varepsilon_i & \text{for all } i \in \{1, \dots, n_1 - 1\}; \\ 3 + 0.1 \times i + \varepsilon_i & \text{for all } i \in \{n_1, \dots, n_2 - 1\}; \\ 3 + 0.1 \times i + (2 \sqrt{3 + (i - n)} - 4) + \varepsilon_i & \text{for all } i \in \{n_2, \dots, N\}. \end{cases}$$

2. **(MM2)** Model changes for homoscedastic linear model:

$$Y_i = \begin{cases} -4 + 2 Z_i^{(1)} - 5 Z_i^{(2)} + 3 Z_i^{(3)} + \varepsilon_i & \text{for all } i \in \{1, \dots, n_1 - 1\}; \\ -8 - Z_i^{(1)} + 3 \log(|Z_i^{(3)}|) + \varepsilon_i & \text{for all } i \in \{n_1, \dots, n_2 - 1\}; \\ -4 + 2 Z_i^{(1)} - 5 Z_i^{(2)} + 3 Z_i^{(3)} + 2 * \varepsilon_i & \text{for all } i \in \{n_2, \dots, N\}. \end{cases}$$

where  $(Z_i^{(1)})$  and  $(Z_i^{(3)})$  are observed i.i.d.r.v. following respectively a uniform distribution  $\mathcal{U}([-1, 1])$  and a Student distribution  $t(4)$ , and  $Z_i^{(2)} = (Z_i^{(1)})^2$ .

We applied the  $\widehat{C}r'(t_1, t_2)$  criterion defined in (3.13) with the same data-driven choice of parameters obtained in the previous one-sided case:

$$\widehat{C}r'(t_1, t_2) = \frac{1}{t_2 - t_1 - p} \left( 1 + \frac{N^2}{10(t_2 - t_1 - p)^{2.1}} \right) \|Y(t_2 - t_1) - H(t_2 - t_1)Y(t_2 - t_1)\|^2. \quad (4.1)$$

Thus, the criterion  $\widehat{C}r'$  is applied and Table 2 is obtained after performing 1000 independent replications.

**Conclusions of Monte Carlo experiments:** As in the previous case, the convergence rates of the estimators seem to be close to  $\sqrt{N}$ . However, in contrast to the one-sided break detection, the variance  $\sigma^2$  of the noise has a non-negligible influence on the convergence rate of the estimator as well as on the chosen model.

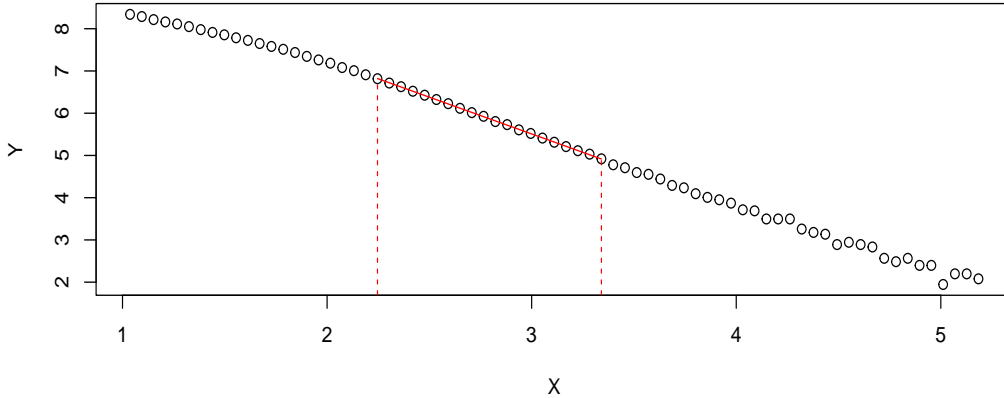


Figure 2: Correlation dimension of a network of ravines and gullies in the French Alps: selection of the largest homoscedastic linear zone of the log-log plot  $(X_i, Y_i)$  where  $Y_i = \log(C(r_i))$ ,  $X_i = \log(r_i)$  and  $1 \leq r_i \leq 250$  in meters (right).

### 4.3 Application to the estimation of the correlation dimension of a network of talwegs and gullies

Let us return to the real data that motivated this work and that were presented in the introduction. It concerns the estimation of a fractal dimension, the correlation dimension, in a network of ravines and gullies in the French Alps. We then apply the criterion  $\widehat{C}r'$  defined in (4.1) to this data set composed of  $(X_i, Y_i)_{1 \leq i \leq N}$  where  $Y_i = \log(C(r_i))$ ,  $X_i = \log(r_i)$  and  $1 \leq r_i \leq 250$ , and  $N = 73$ , the aim being to find the largest zone of homoscedastic linearity to best estimate the slope, which is the estimator of the correlation dimension of the network. Figure 2 shows the point cloud and the optimal linear portion obtained by the  $\widehat{C}r'$  criterion. The results of the log-log regression are very convincing: the multiple R-squared coefficient is now 0.9998, the p-value of the Fisher test is  $2 \cdot 10^{-16}$  and finally the estimator of the slope, which is also the estimator  $\widehat{D}$  of the correlation dimension, is almost 1.741.

## 5 Proofs

*Proof of Theorem 3.1.* For ease of writing, denote now:

$$n_p := n - p \quad \text{and} \quad t_p := t - p.$$

We are going to study successively the cases  $t < n$  and  $t > n$ . Indeed, we have the property:

$$\begin{aligned} \mathbb{P}(|\widehat{\tau} - \tau| \geq N^{-\gamma}) &\leq \mathbb{P}\left(\max \{\widehat{C}r(t), t \text{ such as } |t - n| \geq N^{1-\gamma}\} < \widehat{C}r(n)\right) \\ &\leq \sum_{t=p+1}^{n-N^{1-\gamma}} \mathbb{P}\left(\widehat{C}r(t) < \widehat{C}r(n)\right) + \sum_{t=n+N^{1-\gamma}}^N \mathbb{P}\left(\widehat{C}r(t) < \widehat{C}r(n)\right). \end{aligned} \quad (5.1)$$

1. Let  $t < n$  such as  $|t - n| \geq N^{1-\gamma}$ , i.e.  $p + 1 \leq t \leq [\tau N] - N^{1-\gamma}$ .

Denote  $J_{t,n}$  the subspace of  $\mathbb{R}^n$  such as  $J_{t,n} = \{(x_1, \dots, x_t, 0, \dots, 0), (x_1, \dots, x_t) \in \mathbb{R}^t\}$  and  $P_{J_{t,n}}$  the orthogonal projection on  $J_{t,n}$  in  $\mathbb{R}^n$ . Since  $I(n) - H(n) = P_{[Z]_n^\perp}$ , matrix of the orthogonal projection on  $[Z]_n^\perp$  in  $\mathbb{R}^n$ , then  $V_t = (I(t) - H(t))\varepsilon(t) = P_{[Z]_n^\perp} P_{J_{t,n}} \varepsilon(n)$ . But we have  $P_{[Z]_n^\perp} \varepsilon(n) = P_{[Z]_n^\perp} P_{J_{t,n}} \varepsilon(n) + P_{[Z]_n^\perp} P_{J_{t,n}^\perp} \varepsilon(n)$ , with  $W_{n-t} = P_{[Z]_n^\perp} P_{J_{t,n}} \varepsilon(n) = P_{[Z]_n^\perp \cap J_{t,n}} \varepsilon(n)$ . Therefore, using Cochran' Theorem,  $\|V_t\|^2$  follows a chi-square distribution with  $t_p = (t - p)$  degrees of freedom and  $\|W_{n-t}\|^2$  follows a chi-square distribution with  $(n - t)$  degrees of freedom, where  $V_t$  and  $W_{n-t}$  are independent Gaussian variables.

But, since  $t \leq n$ :

$$\begin{aligned} \mathbb{P}\left(\widehat{C}r(t) < \widehat{C}r(n)\right) &= \mathbb{P}\left(\frac{a_{\alpha,\beta}(t)}{t_p} \|(I(t) - H(t))\varepsilon(t)\|^2 < \frac{a_{\alpha,\beta}(n)}{n_p} \|(I(n) - H(n))\varepsilon(n)\|^2\right) \\ &= \mathbb{P}\left(\frac{a_{\alpha,\beta}(t)}{t_p} \|V_t\|^2 < \frac{a_{\alpha,\beta}(n)}{n_p} (\|V_t\|^2 + \|W_{n-t}\|^2)\right) \\ &= \mathbb{P}\left(\|W_{n-t}\|^2 - A \|V_t\|^2 > 0\right) \quad \text{with } A := \frac{n_p a_{\alpha,\beta}(t)}{t_p a_{\alpha,\beta}(n)} - 1. \end{aligned}$$

For  $U$  a random variable such as there exists  $\ell > 0$  satisfying  $\mathbb{E}(e^{\ell U}) < \infty$ , we have:

$$\mathbb{P}(U > 0) \leq \exp \left[ \inf_{x \geq 0} \{ \log (\mathbb{E}(e^{xU})) \} \right]. \quad (5.2)$$

For  $\xi$  a chi-square variable with  $m$  freedom degrees,  $\mathbb{E}(e^{x\xi}) = (1 - 2x)^{-m/2}$ . Therefore, with  $U = \|W_{n-t}\|^2 - A \|V_t\|^2$  and the independence between  $V_t$  and  $W_{n-t}$ , we obtain from (5.2),

$$\begin{aligned} \mathbb{P}\left(\widehat{C}r(t) > \widehat{C}r(n)\right) &\leq \exp \left( \inf_{x \geq 0} \left\{ \log \left[ \frac{1}{(1 + 2xA)^{t_p/2}} \times \frac{1}{(1 - 2x)^{(n_p - t_p)/2}} \right] \right\} \right) \\ &\leq \exp \left( -\frac{1}{2} \sup_{0 \leq x \leq 1/2} \{f(x)\} \right) \\ &\quad \text{with } f(x) := t_p \log(1 + 2xA) + (n_p - t_p) \log(1 - 2x). \end{aligned}$$

Classical computations show that the supremum of  $f$  is obtained for

$$x_0 := \frac{t_p A - (n_p - t_p)}{2 n_p A} = \frac{1}{2} \frac{t_p (a_{\alpha,\beta}(t) - a_{\alpha,\beta}(n))}{(n_p a_{\alpha,\beta}(t) - t_p a_{\alpha,\beta}(n))}$$

and straightforward computations show that  $0 < x_0 < 1/2$ . As a consequence, we obtain:

$$\begin{aligned} f(x_0) &= t_p \log \left( \frac{a_{\alpha,\beta}(t)}{a_{\alpha,\beta}(n)} \right) + (n_p - t_p) \log \left( \frac{(n_p - t_p) a_{\alpha,\beta}(t)}{n_p a_{\alpha,\beta}(t) - t_p a_{\alpha,\beta}(n)} \right) \\ &= n_p \log(1 + b_{\alpha,\beta}(n, t)) - (n_p - t_p) \log \left( 1 + \frac{n_p}{n_p - t_p} b_{\alpha,\beta}(n, t) \right) \\ &\quad \text{with } b_{\alpha,\beta}(n, t) := \frac{a_{\alpha,\beta}(t)}{a_{\alpha,\beta}(n)} - 1 \geq 0. \end{aligned}$$

The function  $t \in \{p + 1, \dots, n\} \mapsto b_{\alpha,\beta}(n, t) = a_{\alpha,\beta}^{-1}(n) \left( 1 + \frac{N^\alpha}{(t-p)^\beta} \right) - 1$  is a decreasing function. Moreover, since  $\alpha < \beta$ , if  $t$  (or  $t_p = t - p$ ) is negligible with respect to  $N^{\alpha/\beta}$  then with  $b_{\alpha,\beta}(n, t) \underset{N \rightarrow \infty}{\sim} N^{\alpha} t_p^{-\beta} \underset{N \rightarrow \infty}{\rightarrow} \infty$ . And if  $N^{\alpha/\beta}$  is negligible with respect to  $t_p$ , then  $b_{\alpha,\beta}(n, t) \underset{N \rightarrow \infty}{\rightarrow} 0$ . Thus, we will now consider 3 cases inducing 3 different behaviors of  $b_{\alpha,\beta}(n, t)$  and therefore of  $x_0$  and

$f(x_0)$ :  $t_p \in \{[n_p/2], [n_p/2] + 1, \dots, n_p - [N^{1-\gamma}]\}$ ,  $t_p \in \{[(3N^\alpha)^{1/\beta}], [(3N^\alpha)^{1/\beta}] + 1, \dots, [n_p/2] - 1\}$  and  $t_p \in \{1, 2, \dots, [(3N^\alpha)^{1/\beta}] - 1\}$ .

Since  $\alpha < \beta$ , we will also use a Taylor-Lagrange expansion of  $b_{\alpha,\beta}(n, t)$ , which implies that for any  $1 \leq t_p \leq n_p$ , there exists  $\theta_{t_p, n_p} \in [t_p, n_p]$  such as

$$b_{\alpha,\beta}(n, t) \underset{N \rightarrow \infty}{\sim} N^\alpha (t_p^{-\beta} - n_p^{-\beta}) \underset{N \rightarrow \infty}{\sim} \beta \frac{N^\alpha \theta_{t_p, n_p}^{\beta-1}}{n_p^\beta t_p^\beta} (n_p - t_p). \quad (5.3)$$

i. Let  $t_p \in \{[n_p/2], [n_p/2] + 1, \dots, n_p - [N^{1-\gamma}]\}$ . Then  $b_{\alpha,\beta}(n, t) \leq (3^\beta - 1) N^\alpha (\tau N)^{-\beta} \xrightarrow{N \rightarrow \infty} 0$  and from (5.3),  $n_p(n_p - t_p)^{-1} b_{\alpha,\beta}(n, t) \xrightarrow{N \rightarrow \infty} 0$ . Therefore with Taylor expansions, for  $N$  large enough,

$$\begin{aligned} f(x_0) &\underset{N \rightarrow \infty}{\sim} \frac{1}{2} \frac{t_p n_p}{n_p - t_p} b_{\alpha,\beta}^2(n, t) \\ &\underset{N \rightarrow \infty}{\sim} \frac{\beta^2}{2} \frac{N^{2\alpha} \theta_{t_p, n_p}^{2\beta-2}}{n_p^{2\beta-1} t_p^{2\beta-1}} (n_p - t_p) \\ &\geq \frac{\beta^2}{2^\beta} N^{2\alpha} n_p^{2\beta} (n_p - t_p) \\ &\geq \frac{\beta^2}{2^\beta \tau^{2\beta}} N^{2\alpha-2\beta} N^{1-\gamma}. \end{aligned}$$

Then for  $t_p \in \{[n_p/2], [n_p/2] + 1, \dots, n_p - [N^{1-\gamma}]\}$  and  $N$  large enough, using the assumption  $1 - \gamma - 2(\beta - \alpha) > 0$ ,

$$\mathbb{P}(\widehat{C}r(t) > \widehat{C}r(n)) \leq \exp\left(-\frac{\beta^2}{2^{\beta+1} \tau^{2\beta}} N^{1-\gamma-2(\beta-\alpha)}\right). \quad (5.4)$$

ii. Let  $t_p \in \{[(3N^\alpha)^{1/\beta}], [(3N^\alpha)^{1/\beta}] + 1, \dots, [n_p/2] - 1\}$ . Simple computations imply that  $0 \leq b_{\alpha,\beta}(n, t) < \frac{1}{2}$  when  $N$  is large enough. From the definition of  $f(x_0)$ , we obtain:

$$\begin{aligned} f(x_0) &= t_p \log(1 + b_{\alpha,\beta}(n, t)) - (n_p - t_p) \log\left(1 + \frac{t_p}{n_p - t_p} \frac{b_{\alpha,\beta}(n, t)}{1 + b_{\alpha,\beta}(n, t)}\right) \\ &\geq t_p \left(\log(1 + b_{\alpha,\beta}(n, t)) - \frac{b_{\alpha,\beta}(n, t)}{1 + b_{\alpha,\beta}(n, t)}\right), \end{aligned} \quad (5.5)$$

using  $\log(1 + u) \leq u$  for  $u > -1$ . Using also  $\log(1 + u) \geq u - \frac{1}{2} u^2$  for  $u \geq 0$ , we obtain from (5.5),

$$\begin{aligned} f(x_0) &\geq t_p \left(b_{\alpha,\beta}(n, t) - \frac{1}{2} b_{\alpha,\beta}^2(n, t) - \frac{b_{\alpha,\beta}(n, t)}{1 + b_{\alpha,\beta}(n, t)}\right) \\ &\geq \frac{1}{2} t_p \frac{(1 - b_{\alpha,\beta}(n, t)) b_{\alpha,\beta}^2(n, t)}{(1 + b_{\alpha,\beta}(n, t))} \\ &\geq \frac{1}{6} t_p b_{\alpha,\beta}^2(n, t). \end{aligned} \quad (5.6)$$

Using (5.3),  $\beta > 1$  and  $t_p \in \{[(3N^\alpha)^{1/\beta}], [(3N^\alpha)^{1/\beta}] + 1, \dots, [n_p/2] - 1\}$ , we deduce for  $N$  large enough,

$$b_{\alpha,\beta}(n, t) \geq \frac{1}{2} \beta \frac{N^\alpha}{n_p^\beta t_p} (n_p - t_p) \geq \frac{\beta}{2 \tau^\beta} N^{\alpha-\beta}.$$

Therefore, from (5.6) and always with  $N$  large enough:

$$\begin{aligned} f(x_0) &\geq \frac{1}{6} (3N^\alpha)^{1/\beta} \left( \frac{\beta}{2\tau^\beta} N^{\alpha-\beta} \right)^2 \\ &\geq \frac{3^{1/\beta} \beta^2}{24\tau^{2\beta}} N^{\alpha/\beta-2(\beta-\alpha)}. \end{aligned}$$

Then if  $t_p \in \{[(3N^\alpha)^{1/\beta}], [(3N^\alpha)^{1/\beta}] + 1, \dots, [n_p/2] - 1\}$  and for  $N$  large enough,

$$\mathbb{P}(\widehat{Cr}(t) > \widehat{Cr}(n)) \leq \exp\left(-\frac{3^{1/\beta} \beta^2}{48\tau^{2\beta}} N^{\alpha/\beta-2(\beta-\alpha)}\right), \quad (5.7)$$

which explains the condition  $\frac{\alpha}{2\beta} > \beta - \alpha$ .

iii. Let  $t_p \in \{1, 2, \dots, [(3N^\alpha)^{1/\beta}] - 1\}$ . Simple computations imply that  $b_{\alpha,\beta}(n, t) > \frac{1}{4}$  when  $N$  is large enough. But it can be easily proved that  $\log(1+u) - u(1+u)^{-1} \geq \frac{1}{10} \log(1+u)$  for all  $u > 1/4$ . Therefore, from (5.5),

$$f(x_0) \geq \frac{1}{10} t_p \log(1 + b_{\alpha,\beta}(n, t)).$$

But for  $t_p \in \{1, 2, \dots, [(3N^\alpha)^{1/\beta}] - 1\}$  and when  $N$  is large enough,  $b_{\alpha,\beta}(n, t) \geq \frac{1}{2} \frac{N^\alpha}{t_p^\beta}$  and thus

$$f(x_0) \geq \frac{1}{10} t_p \log\left(1 + \frac{1}{2} \frac{N^\alpha}{t_p^\beta}\right).$$

As a consequence,  $t_p \in \{1, 2, \dots, [(3N^\alpha)^{1/\beta}] - 1\}$  and when  $N$  is large enough,

$$\mathbb{P}(\widehat{Cr}(t) > \widehat{Cr}(n)) \leq \left(1 + \frac{1}{2} \frac{N^\alpha}{t_p^\beta}\right)^{-t_p/20}. \quad (5.8)$$

**2.** Let  $t > n$  such as  $|t - n| > N^{1-\gamma}$ , i.e.  $[\tau N] + N^{1-\gamma} \leq t \leq N$ . Then:

$$\begin{aligned} \mathbb{P}(\widehat{Cr}(t) < \widehat{Cr}(n)) &= \mathbb{P}\left(\frac{a_{\alpha,\beta}(t)}{t_p} \|(I(t) - H(t))(\varepsilon(t) + D(t))\|^2 < \frac{a_{\alpha,\beta}(n)}{n_p} \|(I(n) - H(n))\varepsilon(n)\|^2\right) \\ &= \mathbb{P}\left(\frac{a_{\alpha,\beta}(t)}{t_p} \|(V_n + W_{t-n}) + P_{[Z]_t^\perp} D(t)\|^2 < \frac{a_{\alpha,\beta}(n)}{n_p} \|V_n\|^2\right) \\ &= \mathbb{P}\left(\frac{a_{\alpha,\beta}(t)}{t_p} (\|V_n + P_{[Z]_t^\perp} D(t)\|^2 + \|W_{t-n}\|^2) < \frac{a_{\alpha,\beta}(n)}{n_p} \|V_n\|^2\right) \\ &= \mathbb{P}\left((\|V_n - B P_{[Z]_t^\perp} D(t)\|^2 - B(B+1) \|P_{[Z]_t^\perp} D(t)\|^2 - B \|W_{t-n}\|^2) > 0\right), \end{aligned}$$

where  $B = n_p a_{\alpha,\beta}(t) (t_p a_{\alpha,\beta}(n) - n_p a_{\alpha,\beta}(t))^{-1}$ , and  $V_n$  and  $W_{t-n}$  defined as in the first part of the proof and  $D(t)$  defined in (3.5). The conditional probability given  $D(t)$  can be bounded using again (5.2). Indeed, given  $D(t)$ ,  $\sigma^{-2} \|V_n - B P_{[Z]_t^\perp} D(t)\|^2$  follows a noncentral chi-squared distribution with  $n_p$  degrees of freedom, independent to  $\sigma^{-2} \|W_{t-n}\|^2$  which follows a chi-squared distribution with  $(t_p - n_p)$  degrees of freedom. Moreover, given  $D(t)$ , the moment generating function of  $\sigma^{-2} \|V_n - B P_{[Z]_t^\perp} D(t)\|^2$  is given by  $(1-2x)^{n_p/2} \exp(\lambda x (1-2x)^{-1})$  where  $\lambda = \sigma^{-2} B^2 \|P_{[Z]_t^\perp} D(t)\|^2$ .

Thus,

$$\begin{aligned} \mathbb{P}\left(\widehat{Cr}(t) < \widehat{Cr}(n) \mid D(t)\right) &\leq \exp\left(\inf_{x \geq 0} \left\{ \log \left[ \frac{\exp\left(\frac{B^2 \|P_{[Z]_t^\perp} D(t)\|^2 x}{\sigma^2(1-2x)}\right)}{(1-2x)^{n_p/2}} \times \frac{\exp\left(-x \frac{B(B+1) \|P_{[Z]_t^\perp} D(t)\|^2}{\sigma^2}\right)}{(1+2Bx)^{(t_p-n_p)/2}} \right] \right\}\right) \\ &\leq \exp\left(-\frac{1}{2} \sup_{0 \leq x \leq 1/2} \{g(x)\}\right) \end{aligned}$$

$$\text{with } g(x) := n_p \log(1-2x) + (t_p - n_p) \log(1+2Bx) + \frac{2B \|P_{[Z]_t^\perp} D(t)\|^2 x}{\sigma^2} \frac{(1-2(B+1)x)}{1-2x}.$$

Using  $n_p \sim \tau N$  and  $t_p > n_p$ , we obtain the following expansion:

$$B = \frac{n_p}{t_p - n_p} \left(1 - \beta \frac{N^\alpha}{n_p^\beta} (1 + o(1))\right).$$

As a consequence we have  $B \sim \frac{n_p}{t_p - n_p}$  and also  $B \geq \frac{\tau}{2(1-\tau)}$  for  $N$  large enough. Moreover, for any  $u \in [-1/2, 1]$ ,  $\log(1+u) \geq u - u^2$ . Therefore, for  $x$  such as  $x \in [0, \max(1/4; 1/2B)]$  and  $N$  large enough,

$$\begin{aligned} g(x) &\geq n_p(-2x - 4x^2) + (t_p - n_p)(2Bx - 4B^2x^2) + \frac{2B \|P_{[Z]_t^\perp} D(t)\|^2 x}{\sigma^2} (1 - 2(B+1)x) \\ &\geq -4\beta \frac{N^\alpha}{n_p^\beta} (t_p - n_p)x - 4 \frac{t_p n_p}{t_p - n_p} x^2 + \frac{C}{\sigma^2} (t_p - n_p) Bx (1 - 2(B+1)x), \end{aligned}$$

using Assumption C and Lemma 1, implying  $\|P_{[Z]_t^\perp} D(t)\|^2 \geq \frac{1}{2} C (t_p - n_p)$  for  $N$  large enough. For a given  $\varepsilon > 0$  such as  $\varepsilon \leq \max\left(\frac{\tau}{8(1-\tau)}; \frac{1}{2}\right)$ , for  $N$  large enough,

$$\begin{aligned} g(\varepsilon/B) &\geq -4\beta \frac{N^\alpha}{n_p^\beta} \frac{(t_p - n_p)^2}{n_p} \varepsilon - 4 \frac{t_p}{n_p} (t_p - n_p) \varepsilon^2 + \frac{C}{\sigma^2} \left(1 - 2\varepsilon - 4 \frac{1-\tau}{\tau} \varepsilon\right) (t_p - n_p) \varepsilon \\ &\geq -4\beta \frac{1-\tau}{\tau} \frac{N^\alpha}{n_p^\beta} (t_p - n_p) \varepsilon - \frac{4}{\tau} (t_p - n_p) \varepsilon^2 + \frac{C}{\sigma^2} \left(1 - \frac{4}{\tau} \varepsilon\right) (t_p - n_p) \varepsilon \\ &\geq (t_p - n_p) \varepsilon \left(-\frac{5}{\tau} \varepsilon + \frac{C}{\sigma^2} \left(1 - \frac{4}{\tau} \varepsilon\right)\right). \end{aligned}$$

Therefore, for  $\varepsilon < C\tau(4C+5\sigma^2)^{-1}$ , for instance  $\varepsilon = \frac{C}{2}\tau(4C+5\sigma^2)^{-1}$ , we obtain:

$$\sup_{0 \leq x \leq 1/2} \{g(x)\} \geq g(\varepsilon/B) \geq \frac{C^2 \tau}{4\sigma^2(4C+5\sigma^2)} (t_p - n_p).$$

Then, for  $[\tau N] + N^{1-\gamma} \leq t \leq N$  and  $N$  large enough,

$$\begin{aligned} \mathbb{P}\left(\widehat{Cr}(t) < \widehat{Cr}(n) \mid D(t)\right) &\leq \exp\left(-\frac{C^2 \tau}{8\sigma^2(4C+5\sigma^2)} (t_p - n_p)\right) \\ \implies \mathbb{P}\left(\widehat{Cr}(t) < \widehat{Cr}(n)\right) &\leq \exp\left(-\frac{C^2 \tau}{8\sigma^2(4C+5\sigma^2)} N^{1-\gamma}\right). \end{aligned} \quad (5.9)$$

Finally, using (5.1), (5.4), (5.7), (5.8) and (5.9) we deduce

$$\begin{aligned} \mathbb{P}(|\widehat{\tau} - \tau| \geq N^{-\gamma}) &\leq \sum_{t_p=1}^{[(3N^\alpha)^{1/\beta}] - 1} \left(1 + \frac{1}{2} \frac{N^\alpha}{t_p^\beta}\right)^{-t_p/20} + \sum_{t_p=[(3N^\alpha)^{1/\beta}]}^{[n_p/2] - 1} \exp\left(-\frac{3^{1/\beta} \beta^2}{48 \tau^{2\beta}} N^{\alpha/\beta - 2(\beta-\alpha)}\right) \\ &\quad + \sum_{t_p=[n_p/2]}^{n_p - [N^{1-\gamma}]} \exp\left(-\frac{\beta^2}{8\tau^{2\beta}} N^{1-\gamma - 2(\beta-\alpha)}\right) + \sum_{t_p=n_p + [N^{1-\gamma}]}^{N-p} \exp\left(-\frac{C^2 \tau}{8\sigma^2(4C+5\sigma^2)} N^{1-\gamma}\right) \\ &\leq C N^{-\alpha/20} \xrightarrow{N \rightarrow \infty} 0, \end{aligned} \quad (5.10)$$



with  $C > 0$ .  $\square$

*Proof of Corollary 1.* If  $0 < \varepsilon < \alpha/\beta$ , then using the previous proof we have for  $N$  large enough,

$$\begin{aligned} \mathbb{P}(|\tilde{n} - n| \geq N^{1-\gamma}) &\leq \sum_{t_p=N^\varepsilon}^{[(3N^\alpha)^{1/\beta}]^{-1}} \left(1 + \frac{1}{2} \frac{N^\alpha}{t_p^\beta}\right)^{-t_p/20} + \sum_{t_p=[(3N^\alpha)^{1/\beta}]}^{[n_p/2]-1} \exp\left(-\frac{3^{1/\beta} \beta^2}{48 \tau^{2\beta}} N^{\alpha/\beta-2(\beta-\alpha)}\right) \\ &\quad + \sum_{t_p=[n_p/2]}^{n_p-[N^{1-\gamma}]} \exp\left(-\frac{\beta^2}{8\tau^{2\beta}} N^{1-\gamma-2(\beta-\alpha)}\right) + \sum_{t_p=n_p+[N^{1-\gamma}]}^{N-p} \exp\left(-\frac{C^2 \tau}{8\sigma^2 (4C + 5\sigma^2)} N^{1-\gamma}\right) \\ &\leq N \left[ e^{-C_1 N^\varepsilon} + e^{-C_2 N^{\alpha/\beta-2(\beta-\alpha)}} + e^{-C_3 N^{1-\gamma-2(\beta-\alpha)}} + e^{-C_4 N^{1-\gamma}} \right] \\ &\leq e^{-C N^\rho}, \end{aligned}$$

where  $C_1, C_2, C_3, C_4$  and  $C$  are positive real numbers and  $0 < \rho < \min(\varepsilon, 1 - \gamma - 2(\beta - \alpha))$ .

The case  $\varepsilon \geq \alpha/\beta$  can be studied identically except that the first sum is omitted.  $\square$

*Proof of Corollary 2.* Let  $x \in \mathbb{R}$  and denote  $W_N = (Z'(\hat{n} - N^\delta) Z(\hat{n} - N^\delta))^{1/2} (\tilde{\theta}_N - \theta)$ . Then:

$$\mathbb{P}(W_N \leq x) = \mathbb{P}(W_N \leq x \mid \hat{n} - N^\delta \leq n) \mathbb{P}(\hat{n} - N^\delta \leq n) + \mathbb{P}(W_N \leq x \cap \hat{n} - N^\delta \leq n).$$

But given  $\hat{n} - N^\delta \leq n$ , under Assumption D, it is well known (see for instance Huber, 1981) that  $\tilde{\theta}_N$  is asymptotically Gaussian and therefore  $W_N \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2)$ . Moreover, from the proof of Theorem 3.1,  $\mathbb{P}(\hat{n} \geq n + N^{1-\gamma}) \leq N \exp(-C_4 N^{1-\gamma})$  with  $C_4 > 0$  and therefore

$$\mathbb{P}(\hat{n} - N^\delta \leq n) = 1 - \mathbb{P}(\hat{n} \leq n + N^\delta) \xrightarrow[N \rightarrow \infty]{} 1.$$

Moreover  $\mathbb{P}(W_N \leq x \cap \hat{n} - N^\delta \leq n) \leq \mathbb{P}(\hat{n} - N^\delta \leq n) \xrightarrow[N \rightarrow \infty]{} 0$ . Thus, with  $F_W$  the cumulative distribution function of  $W \sim \mathcal{N}'(0, \sigma^2)$ , we have

$$\mathbb{P}(W_N \leq x) \xrightarrow[N \rightarrow \infty]{} F_W(x),$$

and the asymptotic normality of  $\tilde{\theta}_N$  can be deduced. The convergence  $\tilde{\theta}_N \xrightarrow[N \rightarrow \infty]{\mathcal{P}} \theta$  is just a consequence of the asymptotic normality with Assumption D.  $\square$

*Proof of Theorem 3.2.* From Assumption B', we deduce that we can consider  $N$  large enough, such as  $n_0 \geq n_1 + N^{1-\gamma}$  and  $n_0 + p \leq n_2 - N^{1-\gamma}$ . Now, using the notation  $c(t_1, t_2) = \mathbb{P}(\widehat{C}r'(t_1, t_2) < \widehat{C}r'(n_1, n_2))$  and as it was already established in the proof of Theorem (3.1), we have

$$\begin{aligned} &\mathbb{P}(|\hat{n}_1 - n_1| + |\hat{n}_2 - n_2| \geq N^{1-\gamma}) \\ &\leq \mathbb{P}\left(\max\{\widehat{C}r'(t_1, t_2), t_1 \leq n_0 \leq t_2 - p, |t_1 - n_1| \geq N^{1-\gamma} \text{ and } |t_2 - n_2| \geq N^{1-\gamma}\} < \widehat{C}r'(n_1, n_2)\right) \\ &\leq \sum_{t_1=p}^{n_1-p-N^{1-\gamma}} \left\{ \sum_{t_2=n_0+p}^{n_2-N^{1-\gamma}} c(t_1, t_2) + \sum_{t_2=n_2+N^{1-\gamma}}^N c(t_1, t_2) \right\} \\ &\quad + \sum_{t_1=n_1+N^{1-\gamma}}^{n_0} \left\{ \sum_{t_2=n_0+p}^{n_2-N^{1-\gamma}} c(t_1, t_2) + \sum_{t_2=n_2+N^{1-\gamma}}^N c(t_1, t_2) \right\} \\ &\leq I_{11} + I_{12} + I_{21} + I_{22}. \end{aligned}$$

From a time translation, the cases of  $t_1$  and  $t_2$  defined in  $I_{11}$ ,  $I_{12}$ ,  $I_{21}$  and  $I_{22}$  are similar to the one described in Theorem 3.1, *i.e.* at least  $t_1$  or  $t_2$  are in the "linear" zone. Therefore, these sums all satisfy  $I. \xrightarrow[N \rightarrow \infty]{} 0$ .

We deduce that:

$$\mathbb{P}(|\hat{n}_1 - n_1| + |\hat{n}_2 - n_2| \geq N^{1-\gamma}) \xrightarrow[N \rightarrow \infty]{} 0.$$

□

*Proofs of Corollaries 3 and 4.* See the proofs of Corollaries 1 and 2. □

## References

- [1] Arlot, A., Celisse, A. and Harchaoui, Z. (2019). A Kernel Multiple Change-point Algorithm via Model Selection. *Journal of Machine Learning Research*, 20, 1-56.
- [2] Bai, J. (1998). Estimation of multiple-regime regressions with least absolute deviation, *Journal of Statistical Planning and Inference*, **74**, 103–134.
- [3] Bai J. and Perron P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* 66, 47-78.
- [4] Ben Hariz, S., Wylie J. and Zhang Q. (2007). Optimal rate of convergence for nonparametric change-point estimators for nonstationary sequences. *The Annals of Statistics* 35, 1802-1826.
- [5] Carlstein, E. (1988). Nonparametric change-point estimation. *The Annals of Statistics* 16, 188-197.
- [6] Chow, G. (1960). Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica*, 28, p. 591-605.
- [7] Drygas, H. (1976). Weak and strong consistency of the least squares estimators in regression models. *Z. Wahrscheinlichkeitstheorie Verw. Geb.* 34, 119-127.
- [8] Grassberger, P. and Proccacia, I. (1983). Measuring the strangeness of strange attractors. *Physica D* 9, 189-208.
- [9] Huber, P.J. (1981). *Robust statistics*, Wiley Series in Probability and Mathematical Statistics, Wiley & Sons Inc., New York.
- [10] Lai T.L., Robbins, H. and Wei C.Z. (1978). Strong consistency of least squares estimates in multiple regression. *Proc. Natl. Acad. Sci. USA* 75, 3034-3036.
- [11] Lavielle, M. (1999). Detection of multiple changes in a sequence of dependent variables. *Stoc. Proc. Appl.* 83, 79-102.
- [12] Lavielle, M. and Moulines, E. (2000). Least-squares estimation of an unknown number of shifts in a time series. *J. of Time Series Anal.* 21, 33-59.

- [13] Thommeret, N., Bailly, J.-S., Bardet, J.-M., Kaiser, B. and Puech, C. (2014). Dimensions fractales de réseaux vectoriels: méthodes d'estimation et robustesse des résultats. *Cybergeo: European Journal of Geography*