

# DETECTION OF MULTIPLE CHANGE-POINTS IN MULTIVARIATE TIME SERIES

MARC LAVIELLE

Université René Descartes and Université Paris-Sud,  
Laboratoire de Mathématiques  
Marc.Lavielle@math.u-psud.fr

GILLES TEYSSIÈRE

Statistique Appliquée et MODélisation Stochastique  
CES, Université Paris 1 Panthéon-Sorbonne.  
stats@gillesteysièrè.net

July 2006

To appear in the *Lithuanian Mathematical Journal*, vol46, 2006

## Abstract

We consider the multiple change-point problem for multivariate time series, including strongly dependent processes, with an unknown number of change-points. We assume that the covariance structure of the series changes abruptly at some unknown common change-point times. The proposed adaptive method is able to detect changes in multivariate i.i.d., weakly and strongly dependent series. This adaptive method outperforms the Schwarz criteria, mainly for the case of weakly dependent data. We consider applications to multivariate series of daily stock indices returns and series generated by an artificial financial market.

## 1 Introduction

Detecting changes in multivariate time series is of interest if we believe that these series are correlated, and/or that the components of the multivariate vector processes are generated by the same process. This assumption is relevant for financial markets where correlated assets are traded. Empirical evidence, reported e.g., in Teyssièrè [36, 37], shows that several time series, i.e., Foreign Exchange (FX) rates returns, display the same degree of persistence in their volatilities and co-volatilities, a property that might be caused by a common non-stationarity of these series. The presence of strong dependence in asset price volatilities is still a matter of debate, although numerous works, see e.g., Mikosch and Stărică [33], Kokoszka and Teyssièrè [26], Lavielle and Teyssièrè [30], have shown that the strong persistence in volatility is likely to be a statistical artefact, i.e., mainly an effect of the concatenation of processes with different unconditional variances; see also Giraitis *et al.* [14] for a survey on volatility models.

From the point of view of the practitioner, change-points detection procedures are of interest, as we do not know which process does actually generate the data under investigation. Furthermore, economic

data are usually not stationary, and then it may be of interest to approximate an unknown and possibly nonstationary process with locally stationary processes; see e.g., Dalhaus [12].

The literature on change-point detection is rather huge: reference monographs include Basseville and Nikiforov [1], Brodsky and Darkhovsky [7], Csörgö and Horváth [11], Chen and Gupta [8]. The journal articles by Giraitis and Leipus [16, 15], Hawkins [19, 20], Chen and Gupta [9], Mia and Zhao [32], Sen and Srivastava [35] among others, are also of interest.

The statistical theory for weakly dependent volatility processes with a change-point was developed more recently; see, e.g., Chu [10], Kokoszka and Leipus [24, 25], Horváth, Kokoszka and Teyssière [21], Kokoszka and Teyssière [26], Berkes *et al.* [2]. The processes considered in these works are no longer i.i.d., but weakly dependent. For the case of strongly dependent time series, the reader is referred to the paper by Giraitis, Leipus and Surgailis [17], Lavielle [27], and the chapter by Kokoszka and Leipus [23] in the book on long-range dependence edited by Doukhan *et al.* (2003), which reviews the recent works on the issue of change-point detection for univariate dependent time series.

The occurrence of a single change-point in real data is rather rare, as data in economics, finance, hydrology, biology, electrical engineering, etc., display multiple changes, see e.g., Schechtman and Wolfe [34], Braun *et al.* [6], Lavielle and Moulines [29], Lavielle and Teyssière [30]. Thus, a statistical procedure able to reliably detect multiple changes is of practical interest. It has been often claimed that the testing procedure for single change-point can be extended to the multiple change-point case by using Vostrikova's [39] binary segmentation procedure, which consists in applying the single change-point detection procedure on the whole sample, split the sample at the detected change-point, and then apply iteratively the change-point detection procedure on the resulting two segments until no further change-point is found.

In Lavielle and Teyssière [30], we addressed the issue of global procedure vs local procedure, and found that the extension of single change-points procedures to the case of multiple change-point using Vostrikova's [39] binary segmentation procedure is misleading and yields an overestimation of the number of change-points.

A global approach means that all the change-points are simultaneously detected. These change-points are estimated by minimizing a penalized contrast  $J(\boldsymbol{\tau}, \mathbf{y}) + \beta \text{pen}(\boldsymbol{\tau})$  (see [3, 27, 40]). Here,  $J(\boldsymbol{\tau}, \mathbf{y})$  measures how the model obtained with the change-points sequence  $\boldsymbol{\tau}$  fits the observed series  $\mathbf{y}$ . Its role is to locate the change-points as accurately as possible. For detecting changes in the mean and/or the covariance matrix of a multivariate series, we propose to define the contrast  $J(\boldsymbol{\tau}, \mathbf{y})$  from the logarithm of a Gaussian likelihood, even if the observed series is not Gaussian. The penalty term  $\text{pen}(\boldsymbol{\tau})$  only depends on the dimension  $K(\boldsymbol{\tau})$  of the model  $\boldsymbol{\tau}$  and increases with  $K(\boldsymbol{\tau})$ . The penalization parameter  $\beta$  adjusts the trade-off between the minimization of  $J(\boldsymbol{\tau}, \mathbf{y})$  (obtained with a high dimension of  $\boldsymbol{\tau}$ ), and the minimization of  $\text{pen}(\boldsymbol{\tau})$  (obtained with a small dimension of  $\boldsymbol{\tau}$ ).

Asymptotic results have been obtained in theoretical general contexts in [27], extending the previous results of Yao [40]. We shall see that this approach is also very useful for practical applications, for detecting changes in the mean and/or variance of multivariate time series, with the restriction that the series have a common segmentation  $\boldsymbol{\tau}$ . An adaptive method is proposed for estimating the number of change-points. Numerical experiments show that the proposed method outperforms the Schwarz criterion and yields very good results.

For a multivariate time series, the algorithm of the detection procedure will be of order  $\mathcal{O}(mn^2)$ , where  $m$  is the dimension of the vector process, instead of the  $\mathcal{O}(n^2)$  order as in the univariate case.

## 2 A penalized contrast estimate for the multivariate change–point problem

### 2.1 The contrast function

We assume that the  $m$ –dimensional process  $\{\mathbf{Y}_t = (Y_{1,t}, \dots, Y_{m,t})'\}$  is abruptly changing and is characterized by a parameter  $\theta \in \Theta$  that remains constant between two changes. We will strongly use this assumption to define our contrast function  $J(\boldsymbol{\tau}, \mathbf{Y})$ .

Let  $K$  be some integer and let  $\boldsymbol{\tau} = \{\tau_1, \tau_2, \dots, \tau_{K-1}\}$  be an ordered sequence of integers satisfying  $0 < \tau_1 < \tau_2 < \dots < \tau_{K-1} < n$ . For any  $1 \leq k \leq K$ , let  $U(\mathbf{Y}_{\tau_{k-1}+1}, \dots, \mathbf{Y}_{\tau_k}; \theta)$  be a contrast function useful for estimating the unknown true value of the parameter in the segment  $k$ . In other words, the minimum contrast estimate  $\hat{\theta}(\mathbf{Y}_{\tau_{k-1}+1}, \dots, \mathbf{Y}_{\tau_k})$ , computed on the  $k^{\text{th}}$  segment of  $\boldsymbol{\tau}$ , is defined as a solution to the following minimization problem:

$$U\left(\mathbf{Y}_{\tau_{k-1}+1}, \dots, \mathbf{Y}_{\tau_k}; \hat{\theta}(\mathbf{Y}_{\tau_{k-1}+1}, \dots, \mathbf{Y}_{\tau_k})\right) \leq U(\mathbf{Y}_{\tau_{k-1}+1}, \dots, \mathbf{Y}_{\tau_k}; \theta), \quad \forall \theta \in \Theta. \quad (1)$$

For any  $1 \leq k \leq K$ , let  $G$  be defined as

$$G(\mathbf{Y}_{\tau_{k-1}+1}, \dots, \mathbf{Y}_{\tau_k}) = U\left(\mathbf{Y}_{\tau_{k-1}+1}, \dots, \mathbf{Y}_{\tau_k}; \hat{\theta}(\mathbf{Y}_{\tau_{k-1}+1}, \dots, \mathbf{Y}_{\tau_k})\right). \quad (2)$$

Then, define the contrast function  $J(\boldsymbol{\tau}, \mathbf{Y})$  as

$$J(\boldsymbol{\tau}, \mathbf{Y}) = \frac{1}{n} \sum_{k=1}^K G(\mathbf{Y}_{\tau_{k-1}+1}, \dots, \mathbf{Y}_{\tau_k}), \quad (3)$$

where  $\tau_0 = 0$  and  $\tau_K = n$ .

We consider in this paper changes in the covariance matrix of the sequence  $\{\mathbf{Y}_t\}$ . More precisely, we assume that there exists an integer  $K^*$ , a sequence  $\boldsymbol{\tau}^* = \{\tau_1^*, \tau_2^*, \dots, \tau_{K^*}^*\}$  with  $\tau_0^* = 0 < \tau_1^* < \dots < \tau_{K^*-1}^* < \tau_{K^*}^* = n$  and  $K^*$  ( $m \times m$ ) covariance matrices  $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \dots, \boldsymbol{\Sigma}_{K^*}$  such that  $\text{Cov}(\mathbf{Y}_t) = \mathbb{E}(\mathbf{Y}_t - \mathbb{E}(\mathbf{Y}_t))(\mathbf{Y}_t - \mathbb{E}(\mathbf{Y}_t))' = \boldsymbol{\Sigma}_k$  for  $\tau_{k-1}^* + 1 \leq t \leq \tau_k^*$ .

**Model M1:** There exist a  $m$ –vector  $\mu$  such that  $\mathbb{E}(\mathbf{Y}_t) = \mu$  for  $t = 1, 2, \dots, n$ . Furthermore,  $\boldsymbol{\Sigma}_k \neq \boldsymbol{\Sigma}_{k+1}$  for  $1 \leq k \leq K^* - 1$ .

For this simple case of changes in the covariance matrix without changes in the mean, which is of interest for multivariate volatility processes, the following contrast function, based on a Gaussian log–likelihood function, can be used:

$$J(\boldsymbol{\tau}, \mathbf{Y}) = \frac{1}{n} \sum_{k=1}^K n_k \log |\hat{\boldsymbol{\Sigma}}_{\tau_k}|, \quad (4)$$

where  $n_k = \tau_k - \tau_{k-1}$  is the length of the segment  $k$ ,  $\hat{\boldsymbol{\Sigma}}_{\tau_k}$  is the  $(m \times m)$  empirical covariance matrix computed on that segment  $k$ :

$$\hat{\boldsymbol{\Sigma}}_{\tau_k} = \frac{1}{n_k} \sum_{t=\tau_{k-1}+1}^{\tau_k} (\mathbf{Y}_t - \bar{\mathbf{Y}})(\mathbf{Y}_t - \bar{\mathbf{Y}})'. \quad (5)$$

Here  $\bar{\mathbf{Y}} = n^{-1} \sum_{t=1}^n \mathbf{Y}_t$  is the empirical mean of the  $m$ –dimensional series  $\mathbf{Y}_t$  computed on the complete series.

**Model M2:** There exist  $K^*$   $m$ -vectors  $\mu_1, \dots, \mu_{K^*}$  such that  $\mathbb{E}(\mathbf{Y}_t) = \mu_k$  for  $\tau_{k-1}^* + 1 \leq t \leq \tau_k^*$ . Furthermore,  $(\mu_k, \boldsymbol{\Sigma}_k) \neq (\mu_{k+1}, \boldsymbol{\Sigma}_{k+1})$  for  $1 \leq k \leq K^* - 1$ .

For the detection of changes in the mean vector and/or the covariance matrix of a multivariate sequence of random variables, this contrast also reduces to

$$J(\boldsymbol{\tau}, \mathbf{Y}) = \frac{1}{n} \sum_{k=1}^K n_k \log |\widehat{\boldsymbol{\Sigma}}_{\tau_k}| \quad (6)$$

but the  $(m \times m)$  empirical covariance matrix  $\widehat{\boldsymbol{\Sigma}}_{\tau_k}$  is computed on segment  $k$  as

$$\widehat{\boldsymbol{\Sigma}}_{\tau_k} = \frac{1}{n_k} \sum_{t=\tau_{k-1}+1}^{\tau_k} (\mathbf{Y}_t - \bar{\mathbf{Y}}_{\tau_k})(\mathbf{Y}_t - \bar{\mathbf{Y}}_{\tau_k})' \quad (7)$$

where  $\bar{\mathbf{Y}}_{\tau_k} = n_k^{-1} \sum_{t=\tau_{k-1}+1}^{\tau_k} \mathbf{Y}_t$  is the empirical mean of the  $m$ -dimensional series  $\mathbf{Y}_t$  computed on that segment.

Asymptotic results for the minimum contrast estimate of  $\boldsymbol{\tau}^*$  can be obtained within the following asymptotic framework:

**A1** For any  $1 \leq i \leq m$  and any  $1 \leq t \leq n$ , define  $\eta_{t,i} = Y_{t,i} - \mathbb{E}(Y_{t,i})$ . There exists  $C > 0$  and  $1 \leq h < 2$  such that for any  $u \geq 0$  and any  $s \geq 1$ ,

$$\mathbb{E} \left( \sum_{t=u+1}^{u+s} \eta_{t,i} \right)^2 \leq C(\theta) s^h. \quad (8)$$

(**A1** holds with  $h = 1$  for weakly dependent sequences and  $1 < h < 2$  for strongly dependent sequences)

**A2** There exists a sequence  $0 < a_1 < a_2 < \dots < a_{K^*-1} < a_{K^*} = 1$  such that for any  $n \geq 1$  and for any  $1 \leq k \leq K^* - 1$ ,  $\tau_k^* = [na_k]$ .

When the true number  $K^*$  of segments is known, we have the following result concerning the rate of convergence of the minimum contrast estimator of  $\boldsymbol{\tau}^*$ :

**Theorem 2.1** *Assume that conditions **A1-A2** are satisfied. Under model **M1** (resp. model **M2**), let  $\hat{\boldsymbol{\tau}}_n$  be the time instants that minimize the empirical contrast  $J(\boldsymbol{\tau}, \mathbf{Y})$  defined in (4) (resp. (6)). Then, the sequence  $\{n \|\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}^*\|_\infty\}$  is uniformly tight in probability:*

$$\lim_{n \rightarrow \infty} \lim_{\delta \rightarrow \infty} \mathbb{P} \left( \max_{1 \leq k \leq K^*-1} |\hat{\tau}_{n,k} - \tau_k^*| > \delta \right) = 0. \quad (9)$$

(Here,  $J(\boldsymbol{\tau}, \mathbf{Y})$  is minimized over all possible sequences  $\boldsymbol{\tau}$  of length  $K^*$ )

**Proof:** The proof is a direct application of Theorem 2.4 by Lavielle [27]. We can easily check that hypotheses **H1-H2** of [27] are satisfied under models **M1** and **M2** and under hypotheses **A1-A2**. ■

This result means that the rate of convergence of  $\hat{\boldsymbol{\tau}}_n$  does not depend on the covariance structure of the sequence  $\{\mathbf{Y}_t\}$ . For strongly mixing sequences, as well as for strongly dependent sequences, the optimal rate is obtained since  $\|\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}^*\|_\infty = \mathcal{O}_P(1)$ .

## 2.2 Penalty functions for the change–point problem

When the number of change–points is unknown, we estimate it by minimizing a penalized version of the function  $J(\boldsymbol{\tau}, \mathbf{Y})$ . For any sequence of change–point instants  $\boldsymbol{\tau}$ , let  $\text{pen}(\boldsymbol{\tau})$  be a function of  $\boldsymbol{\tau}$  that increases with the number  $K(\boldsymbol{\tau})$  of segments of  $\boldsymbol{\tau}$ . Then, let  $\{\hat{\boldsymbol{\tau}}_n\}$  be the sequence of change–point instants that minimizes

$$U(\boldsymbol{\tau}) = J(\boldsymbol{\tau}, \mathbf{Y}) + \beta \text{pen}(\boldsymbol{\tau}). \quad (10)$$

The procedure is intuitively simple: the adjustment criteria must be compensated so that the over-segmentation would be penalized. However, this compensation must not be very important as a too large penalty function yields an underestimation of the number of segments.

If  $\beta$  is a function of  $n$  that goes to 0 at an appropriate rate as  $n$  goes to infinity, the following theorem states that the estimated number of segments converges in probability to  $K^*$  and that (9) still holds.

**Theorem 2.2** *Let  $\{\beta_n\}$  be a positive sequence of real numbers such that*

$$\beta_n \xrightarrow[n \rightarrow \infty]{} 0 \quad \text{and} \quad n^{2-h} \beta_n \xrightarrow[n \rightarrow \infty]{} \infty, \quad 1 \leq h < 2. \quad (11)$$

*Then, under **A1-A2**, the estimated number of segments  $K(\hat{\boldsymbol{\tau}}_n)$ , where  $\hat{\boldsymbol{\tau}}_n$  is the minimum penalized contrast estimate of  $\boldsymbol{\tau}^*$  obtained by minimizing  $J(\boldsymbol{\tau}, \mathbf{Y}) + \beta_n \text{pen}(\boldsymbol{\tau})$ , converges in probability to  $K^*$ .*

*(Here,  $J(\boldsymbol{\tau}, \mathbf{Y})$  is minimized over all possible sequences  $\boldsymbol{\tau}$  and over all possible  $1 \leq K \leq K_{max}$ , where  $K_{max}$  is some known upperbound of  $K^*$ )*

**Proof:** the proof is a direct application of Theorem 3.1 by Lavielle [27]. ■

In practice, asymptotic results are not very useful for selecting the penalty term  $\beta \text{pen}(\boldsymbol{\tau})$ . Indeed, given a real observed signal with a fixed and finite length  $n$ , the parameter  $\beta$  must be fixed to some arbitrary value. When the parameter  $\beta$  is chosen to be very large, only the more significant abrupt changes are detected. However, a small value of  $\beta$  produces a high number of estimated changes. Therefore, a trade-off must be made, i.e., we have to select a value of  $\beta$  which yields a reasonable level of resolution in the segmentation.

Various authors suggest different penalty functions according to the model they consider. For example, the Schwarz criterion is used by Braun *et al.* [6] for detecting changes in a DNA sequence.

Consider first the penalty function  $\text{pen}(\boldsymbol{\tau})$ . By definition,  $\text{pen}(\boldsymbol{\tau})$  should increase with the number of segments  $K(\boldsymbol{\tau})$ . Following the most popular information criteria such the AIC and the Schwarz criteria, the simplest penalty function  $\text{pen}(\boldsymbol{\tau}) = K(\boldsymbol{\tau})$  can be used. Furthermore, Yao [40] has proved the consistency of the Schwarz criterion for some models.

**Remark 2.3** *For the multivariate i.i.d. case, the penalization parameter for the Schwarz criterion is*

$$\beta = \frac{m(m+1)}{2} \frac{\log n}{n}. \quad (12)$$

*In order to reduce the computational cost of the algorithm and according to the required precision in the estimation, the change-points can be detected on a sub-grid  $d, 2d, 3d, \dots$  of  $1, 2, \dots, n$  (we used  $d = 10$  in*

the numerical experiments). For the multivariate i.i.d. case, where the function  $J(\boldsymbol{\tau}, \mathbf{Y})$  is estimated on a grid of length  $d$ , the penalization parameter for the Schwarz criterion is

$$\beta = \frac{m(m+1)}{2} \frac{\log(n/d)}{n}. \quad (13)$$

In the GARCH framework, the series are weakly dependent. For the univariate strongly dependent case, Lavielle and Moulines [29] suggest to use

$$\beta = \frac{4 \log n}{n^{1-\vartheta}}, \quad (14)$$

where  $\vartheta \in (0, 1)$  is the scaling parameter of the process. For weakly dependent data,  $\vartheta = 0$ . However, a practitioner dealing with real data does not know  $\vartheta$ , and has to estimate it from the observed data. If the process is not stationary, spectral based methods yield incorrect results that overestimate  $\vartheta$  and then artificially increase  $\beta$ . One possibility might be to use wavelet based methods, but a large sample is required; see Teyssière and Abry [38].

**Remark 2.4** *Precise results have been recently obtained by Birgé and Massart [3] in the following model:*

$$Y_i = s^*(i) + \sigma \varepsilon_i, \quad 1 \leq i \leq n, \quad (15)$$

where  $s^*(i) = \sum_{k=1}^{K^*} m_k \mathbf{1}_{\{\tau_{k-1}^*+1 \leq i \leq \tau_k^*\}}$  is a piecewise constant function. The sequence  $\{\varepsilon_i\}$  is a sequence of Gaussian white noise, with variance 1. Then,  $\mathbf{Y}$  is a Gaussian sequence with constant variance  $\sigma^2$  and mean  $m_k$  in segment  $k$ . A penalized least-squares estimate is obtained by minimizing

$$J(\boldsymbol{\tau}, \mathbf{Y}) = \frac{1}{n} \sum_{k=1}^{K(\boldsymbol{\tau})} \sum_{i=\tau_{k-1}+1}^{\tau_k} (Y_i - \bar{Y}_k)^2 + \beta \text{pen}(\boldsymbol{\tau}). \quad (16)$$

In a non asymptotic context, Birgé and Massart [3] have shown that a penalty function of the form

$$\text{pen}(\boldsymbol{\tau}) = K(\boldsymbol{\tau}) \left( 1 + c \log \frac{n}{K(\boldsymbol{\tau})} \right), \quad \beta = \frac{2\sigma^2}{n}, \quad (17)$$

is optimal for minimizing  $\mathbb{E}(\|\hat{s}_{\boldsymbol{\tau}} - s^*\|^2)$ , where the estimated sequence of means  $\{\hat{s}_{\boldsymbol{\tau}}(i)\}$  is defined as  $\hat{s}_{\boldsymbol{\tau}}(i) = \sum_{k=1}^{K(\boldsymbol{\tau})} \bar{Y}_k \mathbf{1}_{\{\tau_{k-1}+1 \leq i \leq \tau_k\}}$ . Based on some numerical experiments, the authors suggest to use  $c = 2.5$  for this model. Note that when the number  $K^*$  of segments is small in comparison with the length  $n$  of the series, this optimal penalty function is an almost linear function of  $K$ .

### 2.3 An adaptive choice for the penalization parameter

For a given contrast function  $J$  and a given penalty function  $\text{pen}(\boldsymbol{\tau})$ , the problem now reduces to the choice for the parameter  $\beta$ .

Let  $K_{MAX}$  be an upper bound on the dimension of  $\boldsymbol{\tau}$ . For any  $1 \leq K \leq K_{MAX}$ , let  $\mathcal{T}_K$  be the set of all the models of dimension  $K$ :

$$\mathcal{T}_K = \{\boldsymbol{\tau} = (\tau_0, \dots, \tau_K) \in \mathbb{N}^{K+1}, \tau_0 = 0 < \tau_1 < \tau_2 < \dots < \tau_{K-1} < \tau_K = n\}.$$

By definition the best model  $\hat{\boldsymbol{\tau}}_K$  of dimension  $K$  minimizes the contrast function  $J$ :

$$\hat{\boldsymbol{\tau}}_K = \arg \min_{\boldsymbol{\tau} \in \mathcal{T}_K} J(\boldsymbol{\tau}, \mathbf{Y}). \quad (18)$$

Note that the sequence  $\{\hat{\tau}_K, 1 \leq K \leq K_{MAX}\}$  can easily be computed. Indeed, let  $\mathcal{G}$  be the upper triangular matrix of dimension  $(n \times n)$  such that the element  $(i, j)$ , for  $j \geq i$  is  $\mathcal{G}_{i,j} = G(\mathbf{Y}_i, \mathbf{Y}_{i+1}, \dots, \mathbf{Y}_j)$ , where  $G(\mathbf{Y}_i, \dots, \mathbf{Y}_j)$  is the contrast function computed with  $(\mathbf{Y}_i, \mathbf{Y}_{i+1}, \dots, \mathbf{Y}_j)$ . Thus, for any  $1 \leq K \leq K_{MAX}$ , we have to find a path  $\tau_0 = 0 < \tau_1 < \tau_2 < \dots, < \tau_{K-1} < \tau_K = n$  that minimizes the total cost

$$J(\boldsymbol{\tau}, \mathbf{Y}) = \frac{1}{n} \sum_{k=1}^K \mathcal{G}_{\tau_{k-1}, \tau_k}. \quad (19)$$

A dynamic programming algorithm can recursively compute the optimal paths  $(\hat{\tau}_K, 1 \leq K \leq K_{MAX})$ , see Kay [22]. Since we consider that the elements of the  $m$ -dimensional vector process  $\{\mathbf{Y}_t\}$  share the same segmentation  $\boldsymbol{\tau}$ , this algorithm requires  $\mathcal{O}(n^2)$  operations as in the unidimensional case.

Assuming that the penalization  $\text{pen}(\boldsymbol{\tau})$  only depends on the dimension of the model, that is on the number  $K$  of segments, let

$$J_K = J(\hat{\tau}_K, \mathbf{Y}), \quad (20)$$

$$p_K = \text{pen}(\boldsymbol{\tau}), \quad \forall \boldsymbol{\tau} \in \mathcal{T}_K \quad (21)$$

$$\hat{p}_K = \text{pen}(\hat{\tau}_K). \quad (22)$$

Thus, for any penalization parameter  $\beta > 0$ , the solution  $\hat{\tau}(\beta)$  minimizes the penalized contrast:

$$\hat{\tau}(\beta) = \arg \min_{\boldsymbol{\tau}} (J(\boldsymbol{\tau}, \mathbf{Y}) + \beta \text{pen}(\boldsymbol{\tau})) \quad (23)$$

$$= \hat{\tau}_{\hat{K}(\beta)} \quad (24)$$

where

$$\hat{K}(\beta) = \arg \min_{K \geq 1} \{J_K + \beta p_K\}. \quad (25)$$

The solution  $\hat{K}(\beta)$  is a piecewise constant function of  $\beta$ . More precisely, if  $\hat{K}(\beta) = K$ ,

$$J_K + \beta p_K < \min_{L \neq K} (J_L + \beta p_L). \quad (26)$$

Thus,  $\beta$  satisfies

$$\max_{L > K} \frac{J_K - J_L}{p_L - p_K} < \beta < \min_{L < K} \frac{J_L - J_K}{p_K - p_L}. \quad (27)$$

Then, there exists a sequence  $\{K_1 = 1 < K_2 < \dots\}$ , and a sequence  $\{\beta_0 = \infty > \beta_1 > \dots\}$ , with

$$\beta_i = \frac{J_{K_i} - J_{K_{i+1}}}{p_{K_{i+1}} - p_{K_i}}, \quad i \geq 1, \quad (28)$$

such that  $\hat{K}(\beta) = K_i, \forall \beta \in [\beta_i, \beta_{i-1})$ . Furthermore, the subset  $\{(p_{K_i}, J_{K_i}), i \geq 1\}$  is the convex hull of the set  $\{(p_K, J_K), K \geq 1\}$ .

The estimated sequence  $\hat{\tau}(\beta)$  should not strongly depend on the choice for the penalization coefficient  $\beta$ . In other words, a small change of  $\beta$  should not lead to a radically different solution  $\hat{\tau}$ . This stability of the solution with respect to the choice for  $\beta$  will be ensured if we only retain the largest intervals  $[\beta_i, \beta_{i-1}), i \geq 1$ .

In summary, we propose the following procedure:

1. for  $K = 1, 2, \dots, K_{MAX}$ , compute  $\hat{\tau}_K, J_K = J(\hat{\tau}_K, \mathbf{Y})$  and  $\hat{p}_K = \text{pen}(\hat{\tau}_K)$ ,

2. compute the sequences  $\{K_i\}$  and  $\{\beta_i\}$ , and the lengths  $\{l_{K_i}\}$  of the intervals  $[\beta_i, \beta_{i-1})$ ,
3. retain the greatest value(s) of  $K_i$  such that  $l_{K_i} \gg l_{K_j}$ , for  $j > i$ .

**Remark 2.5** *Choosing the largest interval usually underestimates the number of changes. Indeed, this interval usually corresponds to a very small number of change-points and we only detect the most drastic changes with such a penalty function. This explains why we should better look for the highest dimension  $K_i$  such that  $l_{K_i} \gg l_{K_j}$ , for any  $j > i$ , to recover the smallest details.*

Instead of computing only one configuration of change-points, this method allows us to put forward different solutions with different dimensions. Indeed, it would be an illusion to believe that a completely blind method can give the “best” solution in any situation. If two dimensions  $K_i$  and  $K_j$  satisfy the criteria suggested in step 3, it is more suitable to propose these two solutions to the user, instead of removing one of them with an arbitrary criterion.

**Remark 2.6** *A classical and natural graphical method for selecting the dimension  $K$  can be summarized as follows:*

- i) examine how the contrast  $J_K$  decreases when  $K$  (that is,  $p_K$ ) increases,*
- ii) select the dimension  $K$  for which  $J_K$  ceases to decrease significantly.*

*In other words, this heuristic approach looks for the maximum curvature in the plot  $(p_K, J_K)$ . The second derivative of this curve is directly related to the length of the intervals  $([\beta_i, \beta_{i-1}), i \geq 1)$ . Indeed, if we represent the points  $(p_K, J_K)$ , for  $1 \leq K \leq K_{MAX}$ ,  $\beta_i$  is the slope between the points  $(p_{K_i}, J_{K_i})$  and  $(p_{K_{i+1}}, J_{K_{i+1}})$ . Thus, looking for where  $J_K$  ceases to decrease means looking for a break in the slope of this curve. Now, the variation of the slope at the point  $(p_K, J_K)$  is precisely the length  $l_{K_i}$  of the interval  $[\beta_i, \beta_{i-1})$ .*

## 2.4 An automatic procedure for estimating $K$

The proposed method requires a careful inspection of the sequence of lengths  $\{l_i\}$  but is difficult to automate. We propose another approach for model selection which yields very good results and which is much more easier to automate for practical applications.

The idea of this method is to model how the sequence  $\{J_K\}$  decreases when there is no change-points in the series  $\{\mathbf{Y}_t\}$  and to look for which values of  $K$  this model fits the observed contrast sequence.

Without any changes in the variance, the joint distribution of the sequence  $\{J_K\}$  is very difficult to compute in a closed form, but some Monte-Carlo experiments show that this sequence decreases as  $c_1 K + c_2 K \log(K)$ .

A numerical example is displayed in Figure 1. We have simulated ten sequences of i.i.d. Gaussian variables and computed the series  $\{J_K\}$  for each of them. The fit with a function  $c_1 K + c_2 K \log(K)$  is always almost perfect ( $r^2 > 0.999$ ). Nevertheless, the coefficients  $c_1$  and  $c_2$  are different for each of these series.



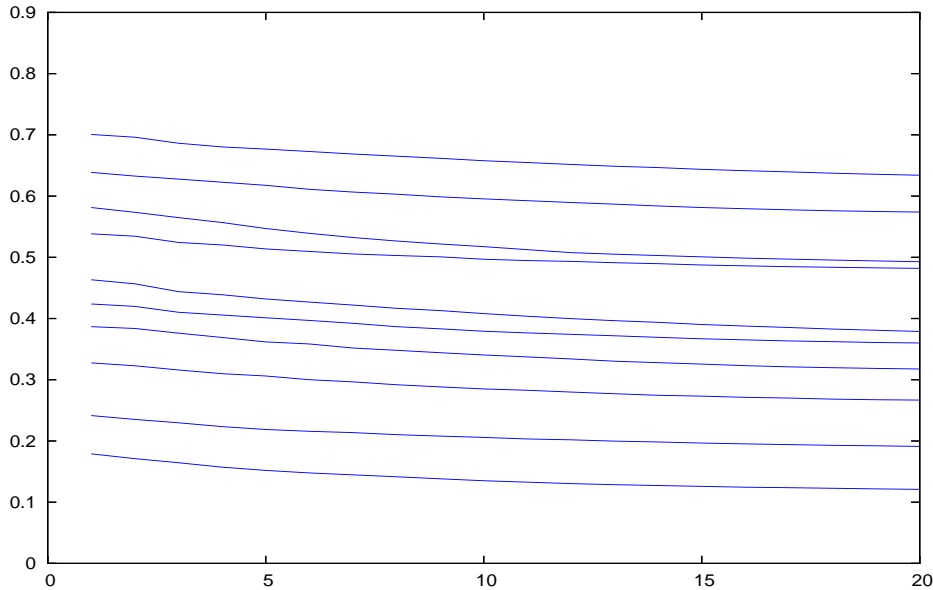


Figure 1: Ten sequences of contrast functions  $\{J_K\}$  computed from ten sequences of i.i.d. bivariate Gaussian variables, with correlation coefficient  $\rho = 0.5$

Thus, we propose the following algorithm:

**Algorithm 1** For  $i = 1, 2, \dots$ ,

1. fit the model

$$J_K = c_1 K + c_2 K \log(K) + e_K,$$

to the sequence  $\{J_K, K \geq K_i\}$ , assuming that  $\{e_K\}$  is a sequence of i.i.d. centered Gaussian random variables,

2. evaluate the probability that  $J_{K_i-1}$  follows also this model, i.e., estimate the probability

$$\mathcal{P}_{K_i} = P(e_{K_i-1} \geq J_{K_i-1} - \hat{c}_1(K_i - 1) + \hat{c}_2(K_i - 1) \log(K_i - 1)), \quad (29)$$

under this estimated model.

Then, the estimated number of segments will be the largest value of  $K_i$  such that the  $P$ -value  $\mathcal{P}_{K_i}$  is smaller than a given threshold  $\alpha$ . We set  $\alpha = 10^{-7}$  and  $K_{MAX} = 20$  in the numerical examples.

**Remark 2.7** A new method was introduced by Lavielle and Ludeña in [28] for estimating the number of significant coefficients in non ordered model selection problems. The contrast used for this thresholding problem is based on a convenient random centering of the partial sums of the ordered observations. This contrast also behaves as  $c_1 K + c_2 K \log(K)$  but the constants  $c_1$  and  $c_2$  can be computed in a closed form in this case. Indeed, it is shown in this paper that this contrast decreases as  $-K - K \sum_{k=K+1}^n 1/k \approx -K - K \log(n/K)$ .

### 3 Numerical experiments with simulated data

#### 3.1 The i.i.d. case

We evaluate the performance of the change-point detection procedure with different penalty functions by considering the case of a bivariate i.i.d. Normal process ( $m = 2$ )  $\{\mathbf{Y}_t = (Y_{1,t}, Y_{2,t})'\}$ . Under the null hypothesis of constant covariance matrix, the process is defined as

$$\begin{pmatrix} Y_{1,t} \\ Y_{2,t} \end{pmatrix} \sim N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{pmatrix} \right], \quad (30)$$

where  $\Sigma_{\tau_k} = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{pmatrix}$  is constant over all intervals  $[\tau_k, \tau_{k+1}]$ .

We study the size and power of the procedure for samples of size  $n = 500$  and  $n = 1000$ , under the null hypothesis of no change-point and two alternative hypotheses with changes in  $\Sigma_t$

1. **DGP 0:** no change in the parameters of  $\Sigma_t$

$$\Sigma_t = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{pmatrix}, \quad t = 1, \dots, n, \quad \sigma_1^2 = \sigma_2^2 = 1, \sigma_{1,2} = 0.5, \quad |\Sigma| = 0.75, \quad (31)$$

We consider as alternatives two Data Generating Processes with double change-points in the covariance matrix at times  $\tau_1$  and  $\tau_2$ ,

2. **DGP 1:** two large changes in the parameters of  $\Sigma_t$

$$\begin{aligned} \Sigma_t &= \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{pmatrix}, \quad t = 1, \dots, \tau_1, \quad \sigma_1^2 = \sigma_2^2 = 1, \sigma_{1,2} = 0.5, \quad |\Sigma_t| = 0.75, \\ \Sigma_t &= \begin{pmatrix} \bar{\sigma}_1^2 & \bar{\sigma}_{1,2} \\ \bar{\sigma}_{1,2} & \bar{\sigma}_2^2 \end{pmatrix}, \quad t = \tau_1 + 1, \dots, \tau_2, \quad \bar{\sigma}_1^2 = 1, \bar{\sigma}_2^2 = 2, \bar{\sigma}_{1,2} = \frac{1}{\sqrt{2}}, \quad |\Sigma_t| = 1.5, \\ \Sigma_t &= \begin{pmatrix} \tilde{\sigma}_1^2 & \tilde{\sigma}_{1,2} \\ \tilde{\sigma}_{1,2} & \tilde{\sigma}_2^2 \end{pmatrix}, \quad t = \tau_2 + 1, \dots, n, \quad \tilde{\sigma}_1^2 = 2, \tilde{\sigma}_2^2 = \frac{1}{\sqrt{2}}, \tilde{\sigma}_{1,2} = 1, \quad |\Sigma_t| = \sqrt{2} - 1, \end{aligned}$$

3. **DGP 2:** one large change and one small change in the parameters of  $\Sigma_t$

$$\begin{aligned} \Sigma_t &= \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{pmatrix}, \quad t = 1, \dots, \tau_1, \quad \sigma_1^2 = \sigma_2^2 = 1, \sigma_{1,2} = 0.5, \quad |\Sigma_t| = 0.75, \\ \Sigma_t &= \begin{pmatrix} \bar{\sigma}_1^2 & \bar{\sigma}_{1,2} \\ \bar{\sigma}_{1,2} & \bar{\sigma}_2^2 \end{pmatrix}, \quad t = \tau_1 + 1, \dots, \tau_2, \quad \bar{\sigma}_1^2 = 1, \bar{\sigma}_2^2 = 2, \bar{\sigma}_{1,2} = \sqrt{1.3}, \quad |\Sigma_t| = 0.7, \\ \Sigma_t &= \begin{pmatrix} \tilde{\sigma}_1^2 & \tilde{\sigma}_{1,2} \\ \tilde{\sigma}_{1,2} & \tilde{\sigma}_2^2 \end{pmatrix}, \quad t = \tau_2 + 1, \dots, n, \quad \tilde{\sigma}_1^2 = 1.5, \tilde{\sigma}_2^2 = 2.2, \tilde{\sigma}_{1,2} = \sqrt{1.5}, \quad |\Sigma_t| = 1.5. \end{aligned}$$

**Remark 3.1** *What really matters is the magnitude of the change in the components of  $\Sigma_{\tau_k}$  and not in the determinant  $|\Sigma_{\tau_k}|$ :  $|\Sigma_{\tau_k}|$  might remain constant but changes are detected provided that the components of  $\Sigma_{\tau_k}$  substantially change.*

### 3.2 The weakly dependent case: non-homogeneous multivariate GARCH processes

We consider here the constant conditional correlation bivariate GARCH, introduced by Bollerslev [4], and defined as:

$$\begin{pmatrix} Y_{1,t} \\ Y_{2,t} \end{pmatrix} = \mathbf{\Sigma}_t^{\frac{1}{2}} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}, \quad \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \sim N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right], \quad (32)$$

where the diagonal components of  $\mathbf{\Sigma}_t$  are time varying and are univariate GARCH(1,1) processes, i.e.,

$$\mathbf{\Sigma}_t = \begin{pmatrix} \sigma_{1,t}^2 & \rho\sigma_{1,t}\sigma_{2,t} \\ \rho\sigma_{1,t}\sigma_{2,t} & \sigma_{2,t}^2 \end{pmatrix}, \quad \rho \in (-1, 1), \quad \begin{aligned} \sigma_{1,t}^2 &= \omega_1 + \beta_1\sigma_{1,t-1}^2 + \alpha_1Y_{1,t-1}^2 \\ \sigma_{2,t}^2 &= \omega_2 + \beta_2\sigma_{2,t-1}^2 + \alpha_2Y_{2,t-1}^2 \end{aligned}. \quad (33)$$

The coefficient of correlation  $\rho$  is constant. This simple specification imply that the conditional covariance matrix is always positive definite. Readers interested in the formal statistical properties of multivariate ARCH-type processes are referred to Boussama [5] and Doukhan, Teyssi re and Winant [13].

We study the performance of the change-point procedure for two cases

1. **DGP 3:** the null hypothesis of constancy of the parameters of the GARCH(1,1) components of  $\mathbf{\Sigma}_t$ . We set  $\omega_1 = 0.1$ ,  $\beta_1 = 0.3$ ,  $\alpha_1 = 0.2$ ,  $\omega_2 = 0.15$ ,  $\beta_2 = 0.2$ ,  $\alpha_2 = 0.2$  and  $\rho = 0.5$ .
2. **DGP 4:** the alternative hypothesis that the parameters of the functional form of the diagonal elements of  $\mathbf{\Sigma}_t$  are constant only on some intervals. This locally stationary bivariate GARCH process is defined as

$$\begin{aligned} \sigma_{1,t}^2 &= \omega_1 + \beta_1\sigma_{1,t-1}^2 + \alpha_1Y_{1,t-1}^2 \\ \sigma_{2,t}^2 &= \omega_2 + \beta_2\sigma_{2,t-1}^2 + \alpha_2Y_{2,t-1}^2 \end{aligned}, \quad \rho = 0.5, \quad t = 1, \dots, \tau_1, \quad (34)$$

$$\begin{aligned} \sigma_{1,t}^2 &= \bar{\omega}_1 + \bar{\beta}_1\sigma_{1,t-1}^2 + \bar{\alpha}_1Y_{1,t-1}^2 \\ \sigma_{2,t}^2 &= \bar{\omega}_2 + \bar{\beta}_2\sigma_{2,t-1}^2 + \bar{\alpha}_2Y_{2,t-1}^2 \end{aligned}, \quad \rho = 0.3, \quad t = \tau_1 + 1, \dots, \tau_2, \quad (35)$$

$$\begin{aligned} \sigma_{1,t}^2 &= \tilde{\omega}_1 + \tilde{\beta}_1\sigma_{1,t-1}^2 + \tilde{\alpha}_1Y_{1,t-1}^2 \\ \sigma_{2,t}^2 &= \tilde{\omega}_2 + \tilde{\beta}_2\sigma_{2,t-1}^2 + \tilde{\alpha}_2Y_{2,t-1}^2 \end{aligned}, \quad \rho = 0.7, \quad t = \tau_2 + 1, \dots, n. \quad (36)$$

We set  $\omega_1 = 0.1$ ,  $\beta_1 = 0.3$ ,  $\alpha_1 = 0.2$ ,  $\omega_2 = 0.15$ ,  $\beta_2 = 0.2$ ,  $\alpha_2 = 0.2$ ,  $\tilde{\omega}_1 = \bar{\omega}_1 = 0.2$ ,  $\tilde{\beta}_1 = \bar{\beta}_1 = 0.1$ ,  $\tilde{\alpha}_1 = \bar{\alpha}_1 = 0.1$ ,  $\tilde{\omega}_2 = \bar{\omega}_2 = 0.05$ ,  $\tilde{\beta}_2 = \bar{\beta}_2 = 0.3$ ,  $\tilde{\alpha}_2 = \bar{\alpha}_2 = 0.2$ , i.e., at time  $\tau_1$  all parameters of the process change, while at time  $\tau_2$ , only the coefficient of correlation  $\rho$  changes.

For  $n = 500$ , we set  $\tau_1 = 200$  and  $\tau_2 = 350$ , while for  $n = 1000$ , we set  $\tau_1 = 400$  and  $\tau_2 = 700$ .

Tables 1, 2 and 3 below report the performance of the change-point procedure using the Schwarz criteria, equation (13), for **DGP0** to **DGP 4**.

**Remark 3.2** For the bivariate GARCH case with a single change-point, we made a simulation based comparison between this procedure and the extension to the multivariate case of univariate change-point procedures, e.g., the Generalized Likelihood Ratio (GLR) test; see Kokoszka and Teyssi re [26]. The Gaussian semiparametric procedure considered here appears largely faster as estimating the parameters of a bivariate GARCH(1,1) model on a moving window for the GLR test is computing time consuming.

Table 1: Average number of detected change-points and their location using the Schwarz criteria,  $n = 500$ ,  $\tau_1 = 200$ ,  $\tau_2 = 350$  (based on 5000 replications). Standard errors are between parentheses

DGP	Number of change-points	$\hat{\tau}_1$	$\hat{\tau}_2$
DGP 0	0.2590 (0.59)	—	—
DGP 1	2.3148 (0.67)	191.1580 (52.11)	324.4140 (59.33)
DGP 2	2.3310 (0.66)	187.6320 (41.33)	325.5490 (62.23)
DGP 3	2.1626 (1.47)	—	—
DGP 4	3.8324 (1.55)	145.6920 (73.07)	243.7830 (100.99)

Table 2: Average number of detected change-points and their location using the Schwarz criteria,  $n = 1000$ ,  $\tau_1 = 400$ ,  $\tau_2 = 700$  (based on 5000 replications). Standard errors are between parentheses

DGP	Number of change-points	$\hat{\tau}_1$	$\hat{\tau}_2$
DGP 0	0.1354 (0.43)	—	—
DGP 1	2.2102 (0.51)	385.6060 (70.00)	666.3900 (97.71)
DGP 2	2.2010 (0.50)	385.5100 (63.74)	670.5840 (94.02)
DGP 3	2.4684 (1.68)	—	—
DGP 4	4.2904 (1.83)	287.1220 (143.60)	466.998 (208.85)

Table 3: Average number of detected change-points using the Schwarz criteria for  $n = 100$  (based on 5000 replications). Standard errors are between parentheses

DGP	Number of change-points
DGP 0	1.2678 (1.39)
DGP 3	2.1618 (1.68)

**Remark 3.3** *In the weakly dependent case, i.e., **DGP 3** and **DGP 4**, the use of the Schwarz criteria defined by equation (13) yields an overestimation of the number of change-points.*

**Remark 3.4** *For the weakly dependent case **DGP 4**, if the parameters  $\omega_j$ ,  $\beta_j$  and  $\alpha_j$  change so that the unconditional variances  $\sigma_j^2$  of the corresponding GARCH(1,1) process, equal to  $\omega_j/(1-\beta_j-\alpha_j)$ ,  $j = 1, 2$ , remains constant as well as  $\rho$  before and after the change, then the number of detected change-points tends to the number of change-points detected under the null hypothesis of **DGP 3**. This is similar to what is observed for change-points tests in the univariate case; see Kokoszka and Teyssière [26]. However, in the multivariate case, we expect that such configuration of constancy of the unconditional variance does not occur for all components of  $\Sigma_t$ , so that the procedure remains useful.*

Tables 4, 5 and 6 below report the performance of the adaptive method. Although this automatic procedure is not suitable for a simulation exercise, as it requires some interaction with the user for selecting the segmentation, it has a better performance than the BIC criteria, mainly for the case of weakly dependent data, which is of interest for analyzing financial data. A comparison between the results reported in tables 3 and 6 show that for very small samples ( $n = 100$ ) generated by an homogeneous process, the use of the Schwartz criteria yields an overestimation of the number of change-points, while the adaptive method provides reliable results. For larger sample sizes, the adaptive method yields accurate

estimates of the change–point times: the bias and standard errors of the estimated change–point locations are far smaller when using the adaptive method than when using the Schwartz criteria.

Table 4: Average number of detected change–points and their location using the adaptive method,  $n = 500$ ,  $\tau_1 = 200$ ,  $\tau_2 = 350$  (based on 5000 replications). Standard errors are between parentheses

DGP	Number of change–points	$\hat{\tau}_1$	$\hat{\tau}_2$
DGP 0	0.1248 (0.62)	—	—
DGP 1	1.7974 (0.52)	236.1160 (69.32)	345.6080 (27.48)
DGP 2	1.8290 (0.61)	196.7920 (25.61)	342.9910 (43.26)
DGP 3	0.2962 (0.90)	—	—
DGP 4	1.5650 (0.83)	217.1770 (64.31)	330.1390 (61.25)

Table 5: Average number of detected change–points and their location using the adaptive method,  $n = 1000$ ,  $\tau_1 = 400$ ,  $\tau_2 = 700$  (based on 5000 replications). Standard errors are between parentheses

DGP	Number of change–points	$\hat{\tau}_1$	$\hat{\tau}_2$
DGP 0	0.1312 (0.62)	—	—
DGP 1	1.9968 (0.19)	405.4200 (47.37)	697.6540 (29.17)
DGP 2	2.0508 (0.32)	396.2610 (33.38)	693.3310 (57.41)
DGP 3	0.3130 (0.84)	—	—
DGP 4	2.0554 (0.74)	402.9410 (83.95)	671.697 (103.005)

Table 6: Average number of detected change–points using the adaptive method for  $n = 100$  (based on 5000 replications). Standard errors are between parentheses

DGP	Number of change–points
DGP 0	0.1442 (0.66)
DGP 3	0.1314 (0.59)

## 4 Application to multivariate financial time series

Since the method works well for detecting changes in multivariate GARCH processes that are widely used for modeling multivariate series of asset returns, we apply this method to multivariate series of returns from real and artificial financial markets.

### 4.1 The bivariate series of returns on FTSE 100 and S&P500 indices

We consider the bivariate process of daily returns  $(r_{1,t}, r_{2,t})$ , where  $r_{i,t} = 100 \times \log(Y_{i,t}/Y_{i,t-1})$ , where  $Y_{i,t}$  are observed on a common time scale between January 1986 and November 2002, i.e.,  $n = 4225$ .

The results obtained with the adaptive method and  $d = 1$  are summarized in Table 7 and Figure 2. The possible number of segments obtained by penalization are 1, 4, 6, 11, 19 and 24. The estimated

number of segments is clearly  $\hat{K} = K_3 = 6$ . Indeed, this solution is obtained for  $\beta \in [85.57, 203.22)$ : the length of this interval is  $l_3 = 117.65$ . Furthermore, the  $P$ -value for  $K_3 = 6$  is  $P_3 = 8.81e-10$  which is much smaller than the  $P$ -values obtained with largest values of  $K$ . We see that the fit obtained for the series  $(J_K, 6 \leq K \leq 30)$  is excellent, but this model yields an underestimated prediction of  $J_5$ .

$K_i$	$\beta_i$	$\beta_{i-1}$	$l_i$	$P_i$
1	692.97	Inf	Inf	-
4	203.22	598.20	394.98	1.76 e-011
6	85.57	203.22	117.65	8.81 e-010
11	41.46	65.88	24.42	0.0134
19	29.68	37.03	7.35	0.2808
24	22.18	27.60	5.42	0.0068

Table 7: FTSE 100 and S&P 500 indices: intervals of the penalization parameter and the  $P$ -values

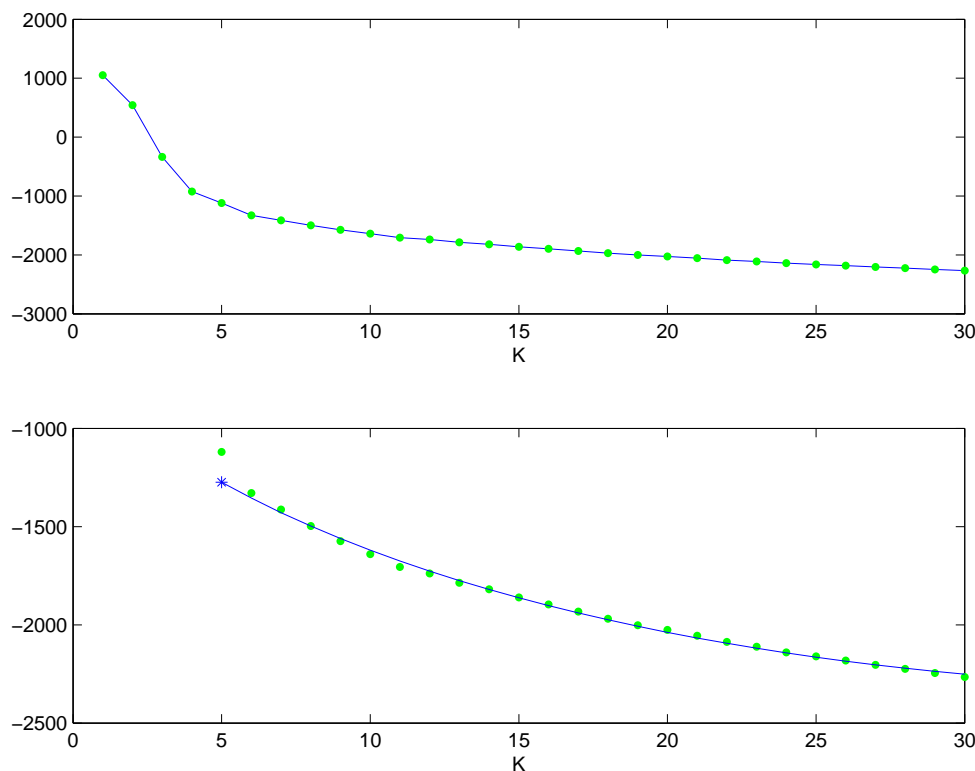


Figure 2: FTSE 100 and S&P500 indices: adaptive detection with  $\alpha = 10^{-7}$ . Above: the contrast sequence  $\{J_K, 1 \leq K \leq 30\}$ ; Below:  $\bullet$  : the contrast sequence  $\{J_K, 6 \leq K \leq 30\}$ ;  $-$  : the fitted model for  $\{J_K, 6 \leq K \leq 30\}$ ;  $*$  : the predicted value of  $J_5$  with this fitted model

Using the adaptive method, we obtain 6 segments and  $\hat{\tau} = \{448, 508, 1715, 2826, 4119\}$  while with the Schwarz criteria,  $\hat{\tau} = \{447, 457, 508, 672, 951, 962, 1152, 1224, 1580, 1715, 2466, 2750, 2967, 2978, 3161, 3215, 3995, 4116\}$ . Very similar results are obtained with a sub-grid ( $d = 5$  or  $d = 10$ ).

The segmentation selected by the adaptive method matches the following events:

- $\hat{\tau}_1 = 448$ ,  $\hat{\tau}_2 = 508$ : the increase and decrease of volatility after October 1987 stock market crash,
- $\hat{\tau}_4 = 2826$ : the Asian crisis of Summer 1997,
- $\hat{\tau}_5 = 4119$ : May 2002, an increase in volatility due to the collapse of the Internet bubble, and the Enron and WorldCom fraudulent bankruptcies, which cast doubts on the actual accounting position of several firms, and increased the level of uncertainty.

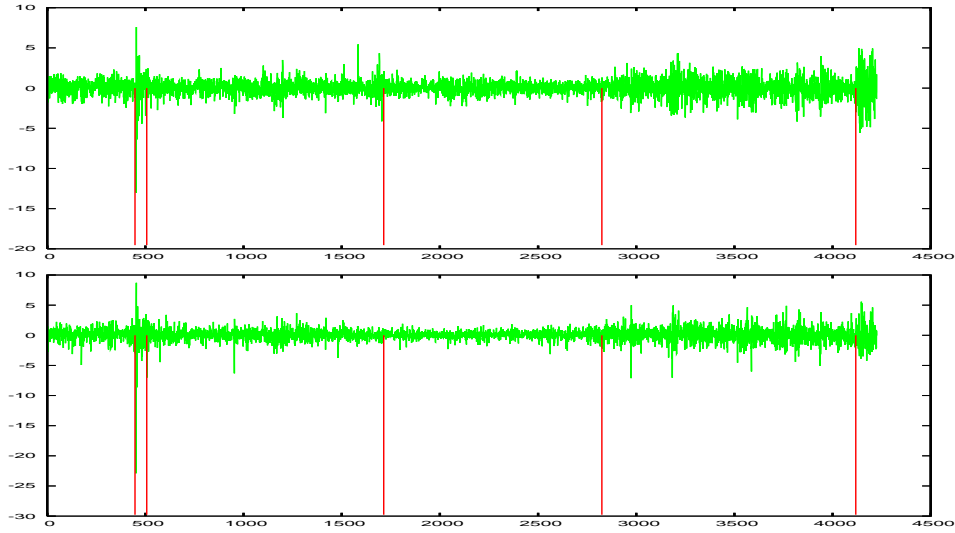


Figure 3: Adaptive detection. Above: The series of returns on FTSE 100 with the estimated change-points represented by vertical lines; Below: The series of returns on S&P 500 with the estimated change-points represented by vertical lines

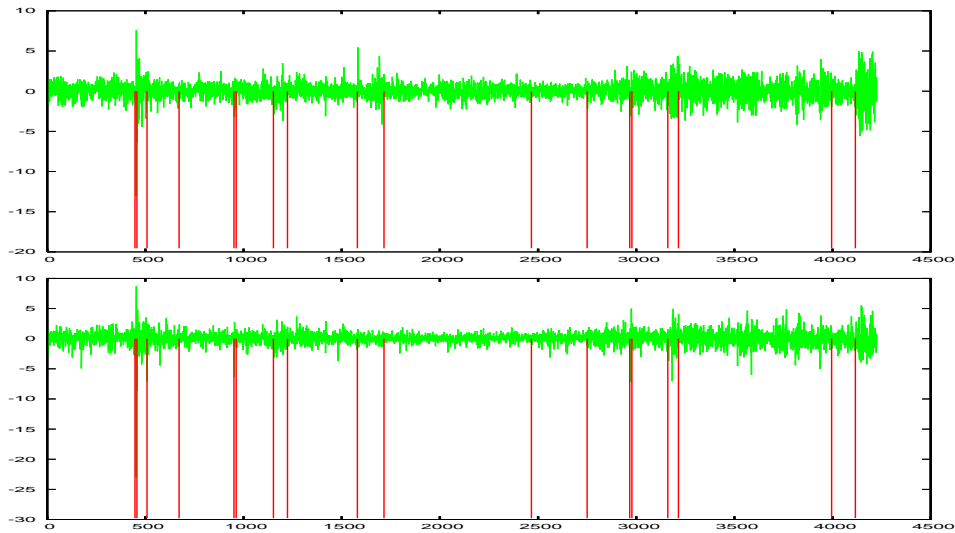


Figure 4: Schwarz criteria. Above: The series of returns on FTSE 100 with the estimated change-points represented by vertical lines; Below: The series of returns on S&P 500 with the estimated change-points represented by vertical lines

It appears that the adaptive procedure detects only major changes, i.e., crashes in stock market volatility.

As a simple illustration of the segmentation, figure 5 below displays the sample autocorrelation function (ACF) of the series of absolute returns on S&P 500 ( $|r_S|$ ), FTSE 100 ( $|r_F|$ ), and the co-volatility of the two series  $\sqrt{|r_S r_F|}$ , for the whole sample and for the segment [508 : 1715]. The apparent strong dependence of these three series, observed on the whole sample, appears to be a statistical artefact: there is indeed a form of strong dependence, as the sample ACF does not vanish to zero for higher lags, but the ACF behaviour is different from the one of the whole sample.

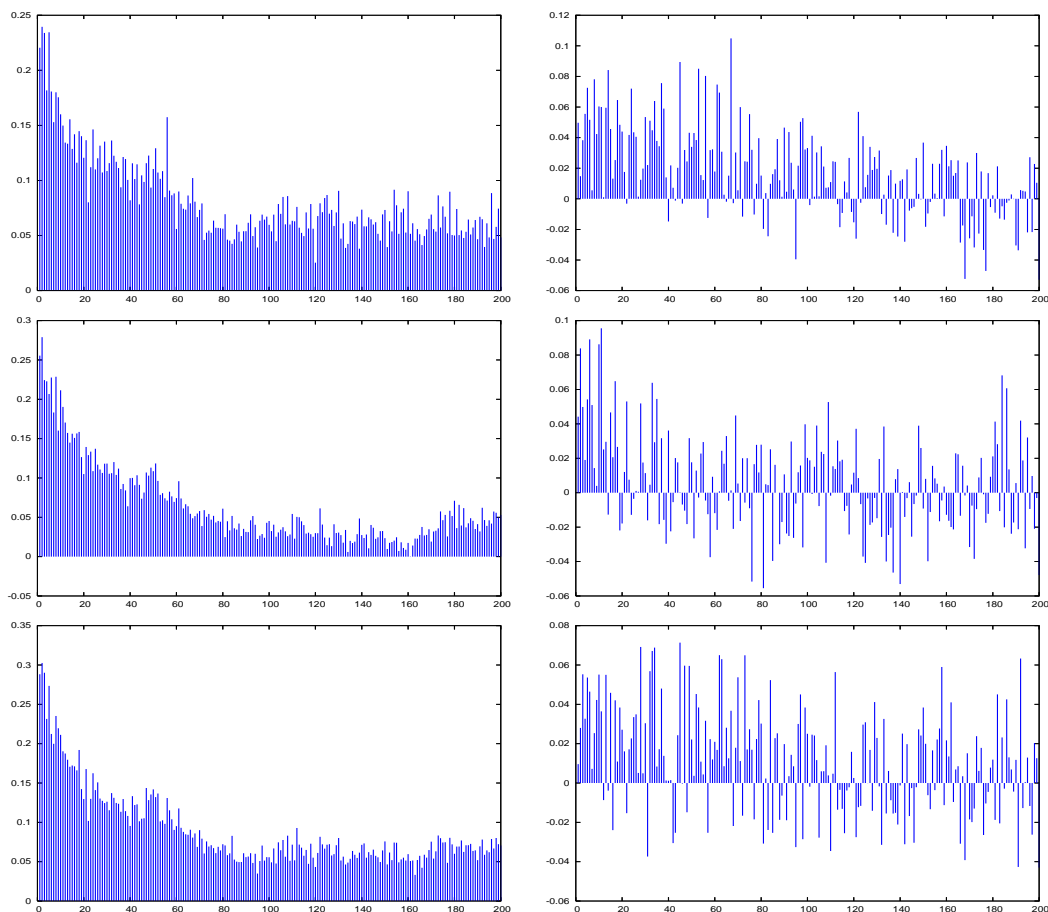


Figure 5: Left column: From top to bottom the sample autocorrelations on absolute returns on S&P 500 ( $|r_S|$ ), absolute returns on FTSE 100 ( $|r_F|$ ), and the sequence of their co-volatility  $\sqrt{|r_S r_F|}$  for the whole sample.

Right Column: The sample ACF of these series for the time interval [508 : 1715]



## 4.2 A trivariate series of stock returns in the financial sector

We consider here three series of log of returns on shares of three banks: Citybank (City), Bank of America (BoA) and Banque Nationale de Paris (BNP) observed at daily frequency from April 8 1999 until April 2002; i.e., 752 observations. These series have been studied in the univariate framework by Kokoszka and Teyssière [26]. With the adaptive method we obtain  $\hat{\tau} = \{512, 612, 634\}$ , the last two change-points matching the huge variations caused by the events of September 11, 2001.

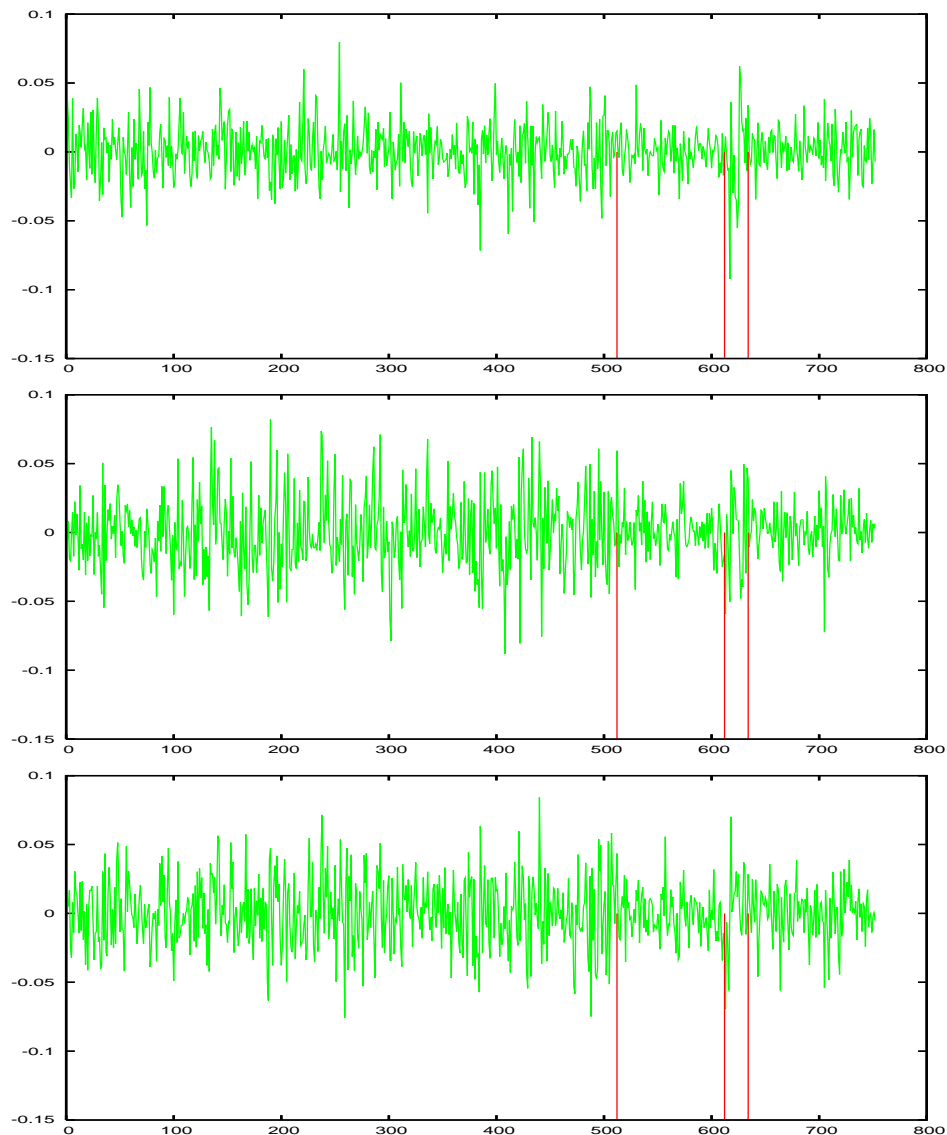


Figure 6: Adaptive method:

Above: The series of returns on BNP with the estimated change-points represented by vertical lines;  
Middle: The series of returns on BoA with the estimated change-points represented by vertical lines;  
Below: The series of returns on City with the estimated change-points represented by vertical lines.

### 4.3 Artificial financial markets

In the two previous examples, the change–point process is unknown, and the explanation of these changes looks exogenous, i.e., caused by external events. We consider here a multivariate process with known and endogenous change–point process. We study simulated multivariate series from an artificial financial market, i.e., a dynamic system which models financial markets with interacting agents. This bivariate process  $(P_{1,t}, P_{2,t})$  of Foreign Exchange (FX) rates has been introduced by Teyssière [36], a framework which could be used for all univariate artificial markets.

This multivariate processes generates common breaks in the volatility process, and displays the same type of dependence as the one observed in multivariate financial time series; see Teyssière [36]. Note that [18] suggested that common breaks might explain the common persistence of asset prices volatility.

The main hypothesis of these models is that agents are heterogeneous, they differ by their forecast functions of asset prices, but have a herding behavior, as they tend to follow the “market opinion”, i.e., the predominant forecast about asset prices.

Since Bachelier, the classical hypothesis is that financial markets are “efficient”, i.e., conditionally on the information set  $I_t$  at time  $t$ , the best forecast of a asset price is

$$E(P_{t+1}|I_t) = P_t. \quad (37)$$

Furthermore, prices are assumed to be independent. If the first claim, equation (37), is acceptable, the second assumption on independence is wrong: Mandelbrot [31] gives numerous examples of “traders” using this dependence for building their strategies. Note that forecasting prices does not make much sense, although forecasting their volatility does; see e.g., Mandelbrot [31].

We consider a market with two types of traders, chartists and fundamentalists, who differ by their forecast function of the future price of the Foreign Exchange (FX). This forecast function for an agent  $i$  of the future price of the Foreign Exchange  $j$ ,  $j = 1, \dots, m$ , is a function of the information set  $I_t$ , and is denoted by  $E^i(P_{j,t+1}|I_t)$ . Chartists extrapolate the next price  $P_{j,t+1}$  by using a linear function of the previous prices:

$$E^c(P_{j,t+1}|I_t) = \sum_{l=0}^{M^c} h_{j,l} P_{j,t-l}, \quad j = 1, \dots, m, \quad (38)$$

where  $h_{j,l}$ ,  $l = 0, \dots, M^c$  are constants,  $M^c$  is the ‘memory’ of the chartists, while fundamentalists forecast that this next price does not depart too much from a series of “fundamentals” prices  $\bar{P}_{j,t}$ :

$$E^f(P_{j,t+1}|I_t) = \bar{P}_{j,t} + \sum_{l=1}^{M^f} \nu_{j,l} (P_{j,t-l+1} - \bar{P}_{j,t-l}), \quad j = 1, \dots, m, \quad (39)$$

where  $\nu_{j,l}$ ,  $l = 1, \dots, M^f$  are positive constants, representing the degree of reversion to the fundamentals,  $M^f$  is the ‘memory’ of the fundamentalists. We assume that the series of ‘fundamentals’  $\bar{P}_{j,t}$ , which can be thought as the price if it were only to be explained by a set of relevant variables, follows a random walk:

$$\bar{P}_{j,t} = \bar{P}_{j,t-1} + \varepsilon_{j,t}, \quad \varepsilon_{j,t} \sim N(0, \sigma_{\varepsilon_j}^2), \quad j = 1, \dots, m. \quad (40)$$

Agents have the possibility of investing at home in a risk free asset or investing abroad in risky assets. We denote by  $\rho_{j,t}$  the foreign interest rate, by  $d_{j,t}^i$  the demand by the  $i^{th}$  agent for foreign currency  $j$ , and by  $r$  the domestic interest rate, with  $\rho_{j,t} > r$ . The exchange rate  $P_{j,t}$  and the foreign interest rate

$\rho_{j,t}$  are considered by agents as independent random variables, with  $E\rho_{j,t} = \rho_j$  and  $\text{Var}(\rho_{j,t}) = \sigma_{\rho_j}^2$ . We denote by  $W_{j,t+1}^i$  the cumulated wealth of individual  $i$  at time  $t + 1$ , with the asset  $j$ :

$$W_{j,t+1}^i = (1 + \rho_{j,t+1})P_{j,t+1}d_{j,t}^i + (W_{j,t}^i - P_{j,t}d_{j,t}^i)(1 + r), \quad j = 1, \dots, m. \quad (41)$$

Agents  $i$  have a standard mean–variance utility function:

$$U(W_{j,t+1}^i) = E(W_{j,t+1}^i) - \lambda_j \text{Var}(W_{j,t+1}^i), \quad (42)$$

where  $\lambda_j$  denotes the risk aversion coefficient, then

$$E(W_{j,t+1}^i | I_t) = (1 + \rho_j)E^i(P_{j,t+1} | I_t)d_{j,t}^i + (W_{j,t}^i - P_{j,t}d_{j,t}^i)(1 + r), \quad (43)$$

$$\text{Var}(W_{j,t+1}^i | I_t) = (d_{j,t}^i)^2 \zeta_{j,t}, \quad \zeta_{j,t} = \text{Var}(P_{j,t+1}(1 + \rho_{j,t+1})), \quad (44)$$

where  $E^i(\cdot | I_t)$  denotes the forecast of an agent of type  $i$ .

Let  $k_t$  be the proportion of fundamentalists at time  $t$ . We assume that the opinion process  $\{k_t\}$  is the same for both markets, as structural changes in opinions do not affect a particular market: the same swing in opinions from chartists to fundamentalists, and vice versa, affects linked markets, e.g., FX markets. Agents are trading several FX rates, and a fundamentalist (resp. chartist) on one market will act as fundamentalist (resp. chartist) on the others markets. Then,  $k_t E^f(P_{j,t+1} | I_t) + (1 - k_t) E^c(P_{j,t+1} | I_t)$  is the forecast for market  $j$ .

Demand  $d_{j,t}^i$  of agent  $i$  for market  $j$  is found by maximizing the expected utility function. First order condition for each market  $j$  gives:

$$d_{j,t} = \frac{(1 + \rho_j) (k_t E^f(P_{j,t+1} | I_t) + (1 - k_t) E^c(P_{j,t+1} | I_t)) - (1 + r) P_{j,t}}{2\zeta_{j,t} \lambda_j}. \quad (45)$$

Denote by  $X_{j,t}$  the exogenous supply of foreign exchange, then the equilibrium condition for each market  $j$  requires that this supply matches the demand, i.e.,  $X_{j,t} = d_{j,t}$ , so that

$$P_{j,t} = \frac{1 + \rho_j}{1 + r} (k_t E^f(P_{j,t+1} | I_t) + (1 - k_t) E^c(P_{j,t+1} | I_t)) - \frac{2\zeta_{j,t} \lambda_j X_{j,t}}{1 + r}. \quad (46)$$

From equation (46), the dynamics of the price process  $\{P_t\}$  depends on the evolution of the process  $\{k_t\}$ , i.e., the proportion of fundamentalists, which governs the transition between the two forecast functions  $E^f(P_{j,t+1} | I_t)$  and  $E^c(P_{j,t+1} | I_t)$ . Several mechanisms for the evolution of the opinion process  $\{k_t\}$  have been proposed in the literature:  $\{k_t\}$  might be either the outcome of an epidemiologic model for the diffusion of opinions, or a learning process based on the comparison of accuracy of the forecast functions using Theil's  $U$ -statistic (see [36]), or a decision based on the accumulated profit gained with each forecast function, etc.

If we assume that  $m = 2$ , i.e., agents are trading two foreign currencies, the bivariate foreign exchange rate process depends on a pair of foreign interest rates  $(\rho_1, \rho_2)$ . We assume that  $2\zeta_{j,t} \lambda_j X_{j,t} / (1 + \rho_j) = \gamma_j \bar{P}_{j,t}$  for  $j = 1, 2$ , and  $M^f = M^c = 1$ , the equilibrium price for the bivariate model is given by

$$\begin{pmatrix} P_{1,t} \\ P_{2,t} \end{pmatrix} = \begin{pmatrix} \frac{k_t - \gamma_1}{A_1} \bar{P}_{1,t} - \frac{k_t \nu_{1,1}}{A_1} \bar{P}_{1,t-1} + \frac{(1 - k_t) h_{1,1}}{A_1} P_{1,t-1} \\ \frac{k_t - \gamma_2}{A_2} \bar{P}_{2,t} - \frac{k_t \nu_{2,1}}{A_2} \bar{P}_{2,t-1} + \frac{(1 - k_t) h_{2,1}}{A_2} P_{2,t-1} \end{pmatrix}, \quad (47)$$

with

$$A_j = \frac{1 + r}{1 + \rho_j} - (1 - k_t) h_{j,0} - k_t \nu_{j,1}. \quad (48)$$

We assume that the bivariate process of fundamentals  $(\bar{P}_{1,t}, \bar{P}_{2,t})$  is positively correlated as follows:

$$\begin{pmatrix} \bar{P}_{1,t} \\ \bar{P}_{2,t} \end{pmatrix} = \begin{pmatrix} \bar{P}_{1,t-1} \\ \bar{P}_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}, \quad \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \sim N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{pmatrix} \right], \quad (49)$$

with  $\sigma_{1,2} > 0$ . In the example considered here, we set  $\sigma_{1,2}$  so that the coefficient of correlation between the two processes  $\varepsilon_{1,t}$  and  $\varepsilon_{2,t}$  is equal to 0.75, a choice motivated by empirical results; see [37, 36].

The parameters of the process are set as follows: the three annual interest rates are equal to  $\rho_1 = 0.07$ ,  $\rho_2 = 0.08$ ,  $r = 0.04$ ,  $\gamma_j = 0.90 - (1+r)/(1+\rho_j)$ ,  $j = 1, 2$ ,  $h_{1,0} = 0.63$ ,  $h_{2,0} = 0.65$ ,  $h_{j,1} = 1 - h_{j,0}$ ,  $j = 1, 2$ ,  $\nu_{1,1} = 0.60$ ,  $\nu_{2,1} = 0.55$ . The choice for  $\gamma_j$  is such that the process is more volatile when chartists' opinion is predominant, while fundamentalists are stabilizing the market. With another choice of the parameters, e.g.,  $\gamma_j = 1 - (1+r)/(1+\rho_j)$ , we obtain the reverse result, as chartists are now stabilizing the market. We consider the former choice as it looks more sensible.

We simulate a bivariate series  $(P_{1,t}, P_{2,t})'$  and detect the changes in the covariance matrix  $\Sigma$  of the bivariate series of returns  $(r_{1,t}, r_{2,t})'$ , with  $r_{j,t} = 100 \times \log(P_{j,t}/P_{j,t-1})$ ,  $j = 1, 2$ . With the adaptive method we find the segmentation  $\hat{\tau} = \{409, 1016, 2115, 2657, 3431\}$ , illustrated by Figure 7 below, while with the Schwarz criteria for i.i.d. observations, equation (12), we found the following segmentation  $\hat{\tau} = \{112, 169, 463, 1016, 1947, 2115, 2656, 2727, 2830, 3036, 3431\}$ . Note that we obtain this adaptive segmentation without resorting to the grid. Using a grid with  $d = 10$ , we obtain  $\hat{\tau} = \{410, 1020, 2120, 2660, 3440\}$ , which is not a too bad approximation.

Figure 7 shows that the adaptive change-point detection procedure is able to pick up the changes in the volatility generated by the variation of the opinion process  $\{k_t\}$  governing the evolution of the volatility of the bivariate price process.

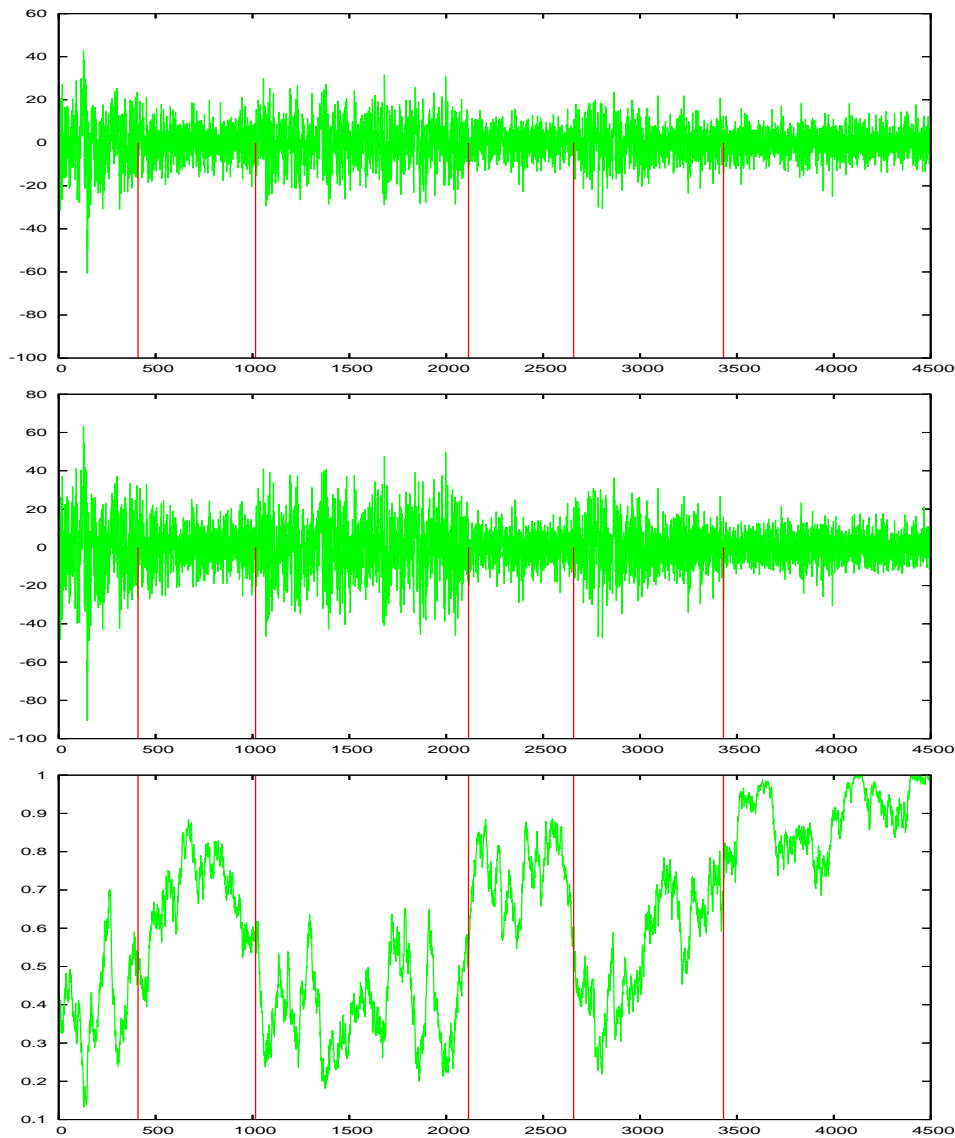


Figure 7: Above: The series  $r_{1,t}$  with the estimated change-points represented by vertical lines; Middle: The series  $r_{2,t}$  with the estimated change-points represented by vertical lines; Below: the series  $k_t$  with the estimated change-points.

**Acknowledgments:** We thank both the referee and the editor, Alfredas Račkauskas, for comments which helped to improve the presentation. We are grateful to Alfredas Račkauskas for proposing us to publish the French translation.

## References

- [1] Basseville, M. and Nikiforov, N. (1993). *The Detection of Abrupt Changes - Theory and Applications*. Prentice-Hall: Information and System Sciences Series.

- [2] Berkes, I., Gombay, E., Horváth, L. and Kokoszka, P.S. (2004). Sequential change–point detection in GARCH( $p, q$ ) models. *Econometric Theory*, **20**, 1140–1167.
- [3] Birgé, L. and Massart, P. (2001) Gaussian model selection. *J. Eur. Math. Soc.* **3** 203–268.
- [4] Bollerslev, T. (1990). Modeling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *Review of Economics and Statistics*, **72**, 498–505.
- [5] Boussama, F. (1998). Ergodicité, mélange, et estimation dans les modèles GARCH. PhD Thesis, University Paris 7.
- [6] Braun, J.V., Braun, R.K. and Muller, H.G. (2000). Multiple changepoint fitting via quasilikehood, with application to DNA sequence segmentation. *Biometrika*, **87**, 301–314.
- [7] Brodsky, B. and Darkhovsky, B. (1993). *Nonparametric Methods in Change–Point Problems*. Kluwer Academic Publishers, The Netherlands.
- [8] Chen, J. and Gupta, A.K. (2000). *Parametric Statistical Change Point Analysis*. Birkhäuser Verlag, Basel.
- [9] Chen, J. and Gupta, A.K. (2004). Statistical inference of covariance change points in Gaussian models. *Statistics*, **38**, 17–28.
- [10] Chu, C.-S.J. (1995). Detecting parameter shift in GARCH models. *Econometric Reviews*, **14**, 241–266.
- [11] Csörgö, M. and Horváth, L. (1997). *Limit Theorems in Change–Point Analysis*. Wiley.
- [12] Dalhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics*, **25**, 1–37.
- [13] Doukhan, P., Teyssière, G. and Winant, P. (2006). A LARCH( $\infty$ ) vector valued process, in: Bertail, P., Doukhan, P. and Soulier, Ph. (Eds.), *Dependence in Probability and Statistics*. Lecture Notes in Statistics, Vol. 187. Springer Verlag, pp. 245–258.
- [14] Giraitis, L., Leipus, R. and Surgailis, D. (2005). Recent advances in ARCH modelling, in: Teyssière, G. and Kirman, A., (Eds.), *Long–Memory in Economics*, 3–38. Springer Verlag, Berlin.
- [15] Giraitis, L. and Leipus, R. (1992). Testing and estimating in the change–point problem of the spectral function. *Lithuanian Mathematical Journal*, **32**, 20–38.
- [16] Giraitis, L. and Leipus, R. (1990). Functional CLT for nonparametric estimates of the spectrum and change–point problem for a spectral function. *Lithuanian Mathematical Journal*, **30**, 674–679.
- [17] Giraitis, L., Leipus, R. and Surgailis, D. (1996). The change–point problem for dependent observations. *Journal of Statistical Planning and Inference*, **53**, 297–310.
- [18] Granger, C.W.J. and Hyung, N. (1999). Occasional structural breaks and long–memory. *Preprint*.
- [19] Hawkins, D.M. (1977). Testing a sequence of observations for a shift in location. *Journal of the American Statistical Association*, **72**, 180–186.
- [20] Hawkins, D.M. (2001). Fitting multiple change–point models to data. *Computational Statistics and Data Analysis*, **37**, 323–341.
- [21] Horváth, L., Kokoszka, P.S. and Teyssière, G. (2001). Empirical process of the squared residuals of an ARCH sequence. *The Annals of Statistics*, **29**, 445–469.

- [22] Kay, S.M. (1998). *Fundamentals of Statistical Signal Processing: Detection Theory*. Prentice Hall: Signal Processing Series.
- [23] Kokoszka, P.S. and Leipus, R. (2003). Detection and estimation of changes in regime, in: P. Doukhan, G. Oppenheim and M.S. Taqqu (Eds.), *Long-Range Dependence: Theory and Applications*, 325–337. Birkhauser.
- [24] Kokoszka, P.S. and Leipus, R. (2000). Change–point estimation in ARCH models. *Bernoulli*, **6**, 513–539.
- [25] Kokoszka, P.S. and Leipus, R. (1999). Testing for parameter changes in ARCH models. *Lithuanian Mathematical Journal*, **39**, 231–247.
- [26] Kokoszka, P.S. and Teyssière, G. (2002). Change–point detection in GARCH models: Asymptotic and bootstrap tests. *Preprint*.
- [27] Lavielle, M. (1999). Detection of multiple changes in a sequence of dependent variables. *Stochastic Processes and their Applications*, **83**, 79–102.
- [28] Lavielle, M. and Ludeña, C. (2005). Random thresholds for linear model selection. *Publication of the INRIA*, **5572**, <http://www.inria.fr/rrrt/rr-5572.html>.
- [29] Lavielle, M. and Moulines, E. (2000). Least–squares estimation of a unknown number of shifts in a time series. *Journal of Time Series Analysis*, **21**, 33–59.
- [30] Lavielle, M. and Teyssière, G. (2005). Adaptive detection of multiple change–points in asset price volatility, in: G. Teyssière and A. Kirman (Eds.), *Long–Memory in Economics*, 129–156. Springer Verlag, Berlin.
- [31] Mandelbrot, B.B. and Hudson R.L. (2004). *The Misbehavior of Markets: A Fractal View of Risk, Ruin, and Reward*. Basic Books, New-York.
- [32] Mia, B.Q. and Zhao, L.C. (1988). Detection of change points using rank methods. *Communications in Statistics - Theory and Methods*, **17**, 3207–3217.
- [33] Mikosch, T. and Stărică, C. (2003) Long–range dependence effects and ARCH modeling, in: Doukhan, P., Oppenheim, G. and Taqqu, M.S. (Eds.), *Long–Range Dependence: Theory and Applications*. Birkhauser, Boston, pp. 439–459.
- [34] Schechtman, E. and Wolfe, D.A. (1985). Multiple change points problem - nonparametric procedures for estimation of the points of change. *Communications in Statistics - Simulation and Computation*, **14**, 615–631.
- [35] Sen, A. and Srivastava, M.S. (1975). On tests for detecting change in the mean. *The Annals of Statistics*, **3**, 96–103.
- [36] Teyssière, G. (2003). Interaction models for common long–range dependence in asset price volatility, in: Rangarajan, G. and Ding, M. (Eds.), *Processes with Long Range Correlations: Theory and Applications*, Lecture Notes in Physics, Vol. 621. Springer Verlag, Berlin, pp. 251–269.
- [37] Teyssière, G. (1997). Modelling exchange rates volatility with multivariate long–memory ARCH processes. *Preprint*.
- [38] Teyssière, G. and Abry, P. (2005). Wavelet analysis of nonlinear long–range dependent processes. Applications to financial time series, in: G. Teyssière and A. Kirman (Eds.), *Long Memory in Economics*, 173–238. Springer Verlag.

- [39] Vostrikova, L.Ju. (1981). Detection of “disorder” in multidimensional random processes. *Soviet Mathematics Doklady*, **24**, 55–59.
- [40] Yao, Y.C. (1988). Estimating the number of change–points via Schwarz criterion. *Statistics & Probability Letters*, **6**, 181–189.