Adaptive Detection of Multiple Change–Points in Asset Price Volatility

Marc Lavielle[1] and Gilles Teyssière[2]

[1] Université René Descartes and Université Paris–Sud, Laboratoire de Mathématiques. Marc.Lavielle@math.u-psud.fr
[2] Laboratoire de Statistique Appliquée et MOdélisation Stochastique (SAMOS), Université Paris 1. stats@gillesteyssiere.net

Summary. This chapter considers the multiple change–point problem for time series, including strongly dependent processes, with an unknown number of change–points. We propose an adaptive method for finding the segmentation, i.e., the sequence of change–points $\boldsymbol{\tau}$ with the optimal level of resolution. This optimal segmentation $\hat{\boldsymbol{\tau}}$ is obtained by minimizing a penalized contrast function $J(\boldsymbol{\tau}, \boldsymbol{y}) + \beta \text{pen}(\boldsymbol{\tau})$. For a given contrast function $J(\boldsymbol{\tau}, \boldsymbol{y})$ and a given penalty function $\text{pen}(\boldsymbol{\tau})$, the adaptive procedure for automatically choosing the penalization parameter β is such that the segmentation $\hat{\boldsymbol{\tau}}$ does not strongly depend on β. This algorithm is applied to the problem of detection of change–points in the volatility of financial time series, and compared with Vostrikova's (1981) binary segmentation procedure.

1 Introduction

The change–point analysis of volatility processes is a recent and important research topic in financial econometrics. Volatility processes, i.e., absolute and squared returns on asset prices, are characterized by a hyperbolic decay of their autocorrelation function (ACF), and then have been first considered as the realization of a strongly dependent, or long–range dependent or long–memory process. Most of the applied research works in this field resorted to the class of long–range dependent volatility processes introduced by Robinson (1991), developed by Granger and Ding (1995) and other authors, defined as

$$Y_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim \text{iid}, \quad E\varepsilon_0 = 0, \quad \text{Var}\,\varepsilon_0 = 1, \quad \sigma_t^2 = \omega + \sum_{j=1}^{\infty} \alpha_j Y_{t-j}^2, \quad (1)$$

with hyperbolically decaying positive weights $\alpha_j \asymp j^{-(1+\vartheta/2)}$, $\vartheta \in (0,1)$, $\sum_j \alpha_j \leqslant 1$. Volatility processes were then implicitly viewed as the realization of a homogeneous process defined by the single scaling parameter ϑ. The estimated intensity of long–range dependence $\hat{\vartheta}$ of asset price volatility is usually strong, near the stationarity limit.

Alternatively, the returns series are modeled by the Integrated ARCH(∞) volatility process, defined by equation (1) with exponentially decaying coefficients α_j and $\sum_j \alpha_j = 1$, i.e., a process with infinite variance since $\omega > 0$. This IARCH representation is incompatible with the conclusion on the presence of long–memory in volatility grounded on the hyperbolic decay of the sample ACF of power transformations of this returns process, since this sample ACF is not properly defined. This contradiction should have questioned the relevance of the hypothesis of a homogeneous volatility process as it was already well known in the statistical literature that change–point processes, nonstationary processes and long–range dependent processes might be confused; see e.g., Bhattacharya *et al.* (1983).

Pioneering works by Mikosch and Stărică (1999, 2003, 2004) advocated the change–point alternative for the analysis of volatility processes, and claimed that the empirical long–range dependence of volatility process was the consequence of nonstationarities. The standard short–range dependent volatility models still provide an accurate representation of the volatility process, provided that we estimate them on intervals of homogeneity. This idea of approximating nonstationary processes with locally stationary processes has been considered by Dalhaus (1997). The statistical theory for volatility processes with a change–point was developed very recently; see, besides the references mentioned above, Chu (1995), Kokoszka and Leipus (1999, 2000), Horváth, Kokoszka and Teyssière (2001), Kokoszka and Teyssière (2002), Berkes, Horváth and Kokoszka (2004), Berkes, Gombay, Horváth and Kokoszka (2004). Interested readers are referred to the chapter on GARCH volatility models by Giraitis, Leipus and Surgailis (2005) in this volume.

Furthermore, if we stick to the parametric framework of long–range dependent volatility processes, the estimates of the long memory parameter on the whole sample and on different subsamples significantly differ. Statistical tests for change in the memory parameter by Horváth (2001), Horváth and Shao (1999) and Beran and Terrin (1999) would reject the null hypothesis of constant long–memory parameter. Thus, even in this framework, the hypothesis of a homogeneous process might be too strong.

We consider here a semiparametric approach, i.e., without reference to a parametric volatility model. The time series cannot be modeled as a stationary process but rather as a piecewise stationary process. Some abrupt changes affect the variance of the time-series at random times, but the distribution of the data does not vary between two successive sudden changes. In what follows, we propose a method which allows to systematically detect these sudden changes and to locate their positions. This method also allows the estimation of the distribution of the data between the abrupt changes. This semiparametric approach is also of interest in a parametric framework, as it might suggest a partition of the series in intervals of homogeneity were stationary volatility models can be estimated; see Aggarwal *et al.* (1999).

The issue of multiple change–points detection has been first viewed as an extension of the single change–point problem by using Vostrikova's (1981) binary segmentation (BS) procedure, which consists in iteratively applying the single change–point detection procedure, i.e., apply first the test for change–point on the whole sample of observations, and if such a point is found, use the same testing procedure on the two resulting sub–segments and on subsequent partitions, until no further change–point is found. This method has been extended by Whitcher *et al.* (2002) to the case of long–range dependent processes by applying it to the discrete wavelet transform of the long–memory process with changes in variance. The BS method was also used by Berkes, Horváth, Kokoszka and Shao (2003) for adjudicating between multiple change–points and long–range dependence in levels.

The BS method is very simple, but has a serious drawback: the number of change–points might be overestimated and their location might be wrong, as one transforms the global problem of change–point detection in a sequence of local change–point detections. The resulting segmentation is not optimal.

We shall adopt here a global approach, where all the change–points are simultaneously detected by minimizing a penalized contrast function of the form

$$J(\boldsymbol{\tau}, \boldsymbol{y}) + \beta \mathrm{pen}(\boldsymbol{\tau}),$$

see Braun *et al.* (2000), Lavielle (1999), Lavielle and Ludeña (2000) and Yao (1988). Here, $J(\boldsymbol{\tau}, \boldsymbol{y})$ measures the fit of $\boldsymbol{\tau}$ with $\boldsymbol{y}$, with $\boldsymbol{y} = Y_1, \ldots, Y_n$. Its role is to locate the change–points as accurately as possible. The penalty term $\mathrm{pen}(\boldsymbol{\tau})$ only depends on the dimension $K(\boldsymbol{\tau})$ of the model $\boldsymbol{\tau}$ and increases with $K(\boldsymbol{\tau})$. Thus, it is used for determinating the number of change–points. The penalization parameter β adjusts the trade-off between the minimization of $J(\boldsymbol{\tau}, \boldsymbol{y})$ (obtained with a high dimension of $\boldsymbol{\tau}$), and the minimization of $\mathrm{pen}(\boldsymbol{\tau})$ (obtained with a small dimension of $\boldsymbol{\tau}$). Lavielle (1999) applied this method, with an arbitrary choice for β, to the series of French CAC 40 index and uncovered changes in the distribution of returns.

Asymptotic results concerning penalized least–squares estimates have been obtained in theoretical general contexts in Lavielle (1999) and Lavielle and Ludeña (2000), extending the previous results by Yao (1988). We shall show that this kind of contrast can also be useful in practice. The main problem is the optimal choice for a penalty function and a coefficient β. In the Gaussian case, Yao (1988) suggested the Schwarz criterion. A complete discussion of the most popular criteria (AIC, Mallow's C_p, BIC), and many other references can be found in Birgé and Massart (2001). In a more general context, we can use a contrast other than the least-squares criterion, since the variables are not necessarily Gaussian and independent. We propose an adaptive procedure for automatically choosing the penalty parameter β in section **2**. We present in section **3** the binary segmentation procedure. An application to financial time series, daily returns on the FTSE 100 index and on 30–minutes spaced returns on FX rates, and simulated returns from artificial financial markets based on microeconomic models with interactive agents, is considered in section **4**.

2 A Penalized Contrast Estimate for the Change–Point Problem

2.1 The Contrast Function

We assume that the process $\{Y_t\}$ 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 $J(\boldsymbol{\tau}, \boldsymbol{y})$.

Let K be some integer and let $\boldsymbol{\tau} = \{\tau_1, \tau_2, \ldots, \tau_{K-1}\}$ be an ordered sequence of integers satisfying $0 < \tau_1 < \tau_2 < \ldots < \tau_{K-1} < n$. For any $1 \leqslant k \leqslant K$, let $U(Y_{\tau_{k-1}+1}, \ldots, Y_{\tau_k}; \theta)$ 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 $\hat{\theta}(Y_{\tau_{k-1}+1}, \ldots, Y_{\tau_k})$, computed on segment k of $\boldsymbol{\tau}$, is defined as a solution to the following minimization problem:

$$U\left(Y_{\tau_{k-1}+1}, \ldots, Y_{\tau_k}; \hat{\theta}(Y_{\tau_{k-1}+1}, \ldots, Y_{\tau_k})\right) \leqslant U(Y_{\tau_{k-1}+1}, \ldots, Y_{\tau_k}; \theta) \ , \ \forall \theta \in \Theta. \tag{2}$$

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

$$G(Y_{\tau_{k-1}+1}, \ldots, Y_{\tau_k}) = U\left(Y_{\tau_{k-1}+1}, \ldots, Y_{\tau_k}; \hat{\theta}(Y_{\tau_{k-1}+1}, \ldots, Y_{\tau_k})\right). \tag{3}$$

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

$$J(\boldsymbol{\tau}, \boldsymbol{y}) = \frac{1}{n} \sum_{k=1}^{K} G(Y_{\tau_{k-1}+1}, \ldots, Y_{\tau_k}), \tag{4}$$

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

For the detection of changes in the variance of a sequence of random variables, the following contrast function, based on a Gaussian log–likelihood function, can be used:

$$J_n(\boldsymbol{\tau}, \boldsymbol{y}) = \frac{1}{n} \sum_{k=1}^{K} n_k \log(\hat{\sigma}_k^2), \tag{5}$$

where $n_k = \tau_k - \tau_{k-1}$ is the length of segment k, $\hat{\sigma}_k^2$ is the empirical variance computed on that segment k, $\hat{\sigma}_k^2 = n_k^{-1} \sum_{i=\tau_{k-1}+1}^{\tau_k} (Y_i - \bar{Y})^2$, and $\bar{Y}$ is the empirical mean of $Y_1, \ldots, Y_n$.

When the true number $K^\star$ of segments is known, the sequence $\hat{\boldsymbol{\tau}}_n$ of change–point instants that minimizes this kind of contrast function has the property, that under extremely general conditions, for any $1 \leqslant k \leqslant K^\star - 1$,

$$\mathrm{P}(|\hat{\tau}_{n,k} - \tau_k^\star| > \delta) \to 0, \quad \text{when } \delta \to \infty \text{ and } n \to \infty, \tag{6}$$

see Lavielle (1999), Lavielle and Ludeña (2000). In particular, this result holds for weakly and strongly dependent processes.

2.2 Penalty Function for the Change–Point Problem

When the number of change–points is unknown, we estimate it by minimizing a penalized version of the function $J(\boldsymbol{\tau}, \boldsymbol{y})$. 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

$$U(\boldsymbol{\tau}) = J(\boldsymbol{\tau}, \boldsymbol{y}) + \beta \text{pen}(\boldsymbol{\tau}). \tag{7}$$

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

If β is a function of n that goes to 0 at an appropriate rate as n goes to infinity, the estimated number of segments $K(\hat{\boldsymbol{\tau}}_n)$ converges in probability to $K^\star$ and condition (6) still holds; see Lavielle (1999), Lavielle and Ludeña (2000) for more details.

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 β must be fixed to some arbitrary value. When the parameter β is chosen to be very large, only the more significant abrupt changes are detected. However, a small value of β produces a high number of the estimated changes. Therefore, a trade-off must be made, i.e., we have to select a value of β 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.* (2000) 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, we suggest to use in practice the simplest penalty function $\text{pen}(\boldsymbol{\tau}) = K(\boldsymbol{\tau})$.

Remark 1. We can argue this specific choice for the penalty function with theoretical considerations. Indeed, precise results have been recently obtained by Birgé and Massart (2001) in the following model:

$$Y_i = s^\star(i) + \sigma\varepsilon_i, \quad 1 \leqslant i \leqslant n, \tag{8}$$

where $s^\star(i) = \sum_{k=1}^{K^\star} m_k \mathbf{1}_{\{\tau^\star_{k-1}+1 \leqslant i \leqslant \tau^\star_k\}}$ is a piecewise constant function. The sequence $\{\varepsilon_i\}$ is a sequence of Gaussian white noise, with variance 1. A penalized least-squares estimate is obtained by minimizing

$$J(\boldsymbol{\tau}, \boldsymbol{y}) = \frac{1}{n} \sum_{k=1}^{K(\boldsymbol{\tau})} \sum_{i=\tau_{k-1}+1}^{\tau_k} (Y_i - \bar{Y}_k)^2 + \beta \text{pen}(\boldsymbol{\tau}). \tag{9}$$

In a non asymptotic context, Birgé and Massart (2001) have shown that a penalty function of the form

$$\text{pen}(\boldsymbol{\tau}) = K(\boldsymbol{\tau})\left(1 + c\log\frac{n}{K(\boldsymbol{\tau})}\right), \quad \beta = \frac{2\sigma^2}{n}, \tag{10}$$

is optimal for minimizing $\mathbb{E}\left(\|\hat{s}_{\boldsymbol{\tau}} - s^\star\|^2\right)$, where the estimated sequence of means $\{\hat{s}_{\boldsymbol{\tau}}(i)\}$ is defined as $\hat{s}_{\boldsymbol{\tau}}(i) = \sum_{k=1}^{K(\boldsymbol{\tau})} \bar{Y}_k \mathbf{1}_{\{\tau_{k-1}+1 \leqslant i \leqslant \tau_k\}}$. Based on some numerical experiments, the authors suggest to use $c = 2.5$. Note that when the number $K^\star$ of segments is small in comparison with the length n of the series, this optimal penalty function is an almost linear function of K. Furthermore, Yao (1988) has proved the consistency of the Schwarz criterion for this model, with $\text{pen}(\boldsymbol{\tau}) = K(\boldsymbol{\tau})$ and $\beta = 2\sigma^2(\log n)/n$.

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 β.

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

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

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

$$\hat{\boldsymbol{\tau}}_K = \arg\min_{\boldsymbol{\tau}\in\mathcal{T}_K} J(\boldsymbol{\tau}, \boldsymbol{y}). \tag{11}$$

Note that the sequence $\{\hat{\boldsymbol{\tau}}_K, 1 \leqslant K \leqslant K_{MAX}\}$ 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 $j \geqslant i$ is $\mathcal{G}_{i,j} = G(Y_i, Y_{i+1}, \ldots Y_j)$, where $G(Y_i, \ldots Y_j)$ is the contrast function computed with $(Y_i, Y_{i+1}, \ldots Y_j)$. Thus, for any $1 \leqslant K \leqslant K_{MAX}$, we have to find a path $\tau_0 = 0 < \tau_1 < \tau_2 < \ldots, < \tau_{K-1} < \tau_K = n$ that minimizes the total cost

$$J(\boldsymbol{\tau}, \boldsymbol{y}) = \frac{1}{n}\sum_{k=1}^{K} \mathcal{G}_{\tau_{k-1},\tau_k}. \tag{12}$$

A dynamic programming algorithm can recursively compute the optimal paths $(\hat{\boldsymbol{\tau}}_K, 1 \leqslant K_{MAX})$, see Kay (1998). This algorithm requires $\mathcal{O}(n^2)$ operations. Then, let

$$J_K = J(\hat{\boldsymbol{\tau}}_K, \boldsymbol{y}), \tag{13}$$

$$p_K = \text{pen}(\boldsymbol{\tau}), \quad \forall \boldsymbol{\tau} \in \mathcal{T}_K. \tag{14}$$

As mentioned above, we suggest to use $p_K = K$.

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

$$\hat{\boldsymbol{\tau}}(\beta) = \arg\min(J(\boldsymbol{\tau}, \boldsymbol{y}) + \beta\text{pen}(\boldsymbol{\tau})) \tag{15}$$

$$= \hat{\boldsymbol{\tau}}_{\hat{K}(\beta)} \tag{16}$$

where

$$\hat{K}(\beta) = \arg\min_{K \geqslant 1} \{J_K + \beta p_K\}. \tag{17}$$

The way how the solution $\hat{K}(\beta)$ varies with the penalization parameter β is given by the following proposition:

Proposition 1. *There exists a sequence* $\{K_1 = 1 < K_2 < \ldots\}$*, and a sequence* $\{\beta_0 = \infty > \beta_1 > \ldots\}$*, with*

$$\beta_i = \frac{J_{K_i} - J_{K_{i+1}}}{p_{K_{i+1}} - p_{K_i}} \ , \quad i \geqslant 1, \tag{18}$$

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

Proof. For any $K \geqslant 1$, let $\hat{K}(\beta) = K$. Then

$$J_K + \beta p_K < \min_{L>K} (J_L + \beta p_L), \tag{19}$$

$$J_K + \beta p_K < \min_{L<K} (J_L + \beta p_L). \tag{20}$$

Thus, β must satisfy

$$\max_{L>K} \frac{J_K - J_L}{p_L - p_K} < \beta < \min_{L<K} \frac{J_L - J_K}{p_K - p_L}. \tag{21}$$

■

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

In summary, we propose the following procedure:

1. For $K = 1, 2, \ldots, K_{MAX}$, compute $\hat{\boldsymbol{\tau}}_K$, $J_K = J(\hat{\boldsymbol{\tau}}_K, \boldsymbol{y})$ and $p_K = \text{pen}(\hat{\boldsymbol{\tau}}_K)$,
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. 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 3. 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 significatively.

In other words, this heuristic approach looks for the maximum curvature in the plot $(p_K\ ,\ J_K)$. Proposition 1 states that the second derivative of this curve is directly related to the length of the intervals $([\beta_i, \beta_{i-1}), i \geqslant 1)$. Indeed, if we represent the points $(p_K\ ,\ J_K)$, for $1 \leqslant K \leqslant K_{MAX}$, β_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*

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 $c_1K + c_2K\log(K)$.

A numerical example is displayed 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 $c_1K + c_2K\log(K)$ is always almost perfect ($r^2 > 0.999$). Nevertheless, the coefficients c_1 and c_2 are different for each of these series. Thus, we propose the following algorithm:

Algorithm 1
For $i = 1, 2, \ldots,$

1. *Fit the model*
$$J_K = c_1K + c_2K\log(K) + e_K,$$
to the sequence $\{J_K, K \geqslant K_i\}$*, assuming that* $\{e_K\}$ *is a sequence of iid centered Gaussian random variables,*
2. *Evaluate the probability that* J_{K_i-1} *follows also this model, i.e., estimate the probability*
$$\mathcal{P}_{K_i} = P\left(e_{K_i-1} \geqslant J_{K_i-1} - \hat{c}_1(K_i-1) + \hat{c}_2(K_i-1)\log(K_i-1)\right), \quad (22)$$
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* α*. We set* $\alpha = 10^{-5}$ *in the numerical examples.*

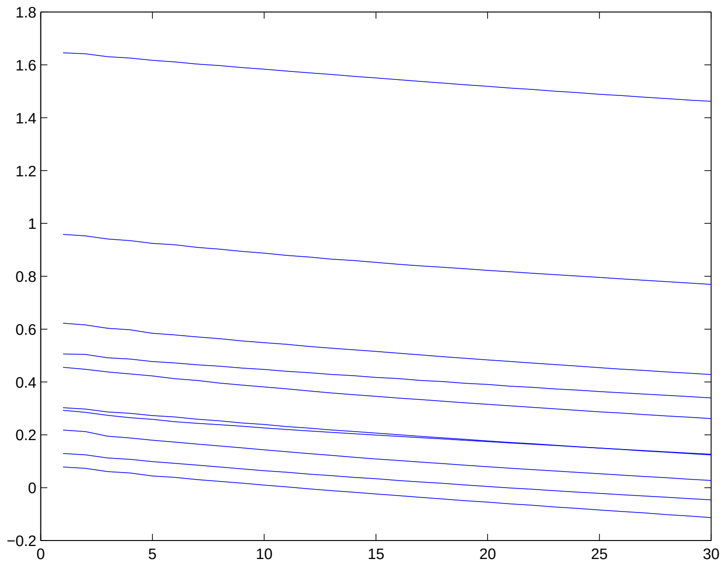

Fig. 1. Ten sequences of contrast functions (J_K) computed from ten sequences of i.i.d. Gaussian variables

3 An Alternative Method: The Binary Segmentation Procedure

We present here the local approach for finding multiple change–points, i.e., finding the configuration $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_{K-1})$ with break dates $\{0 < \tau_1 < \ldots < \tau_{K-1} < n\}$, which rely on single change–points tests. Since financial time series are very large, i.e., over several thousands of observations, single change–point tests are of limited practical interest.

The binary segmentation procedure, studied by Vostrikova (1981), is the standard method for detecting multiple change–points by using a test for single change–point: we split the series at the point detected by the single change–point test, i.e., the point where the test statistic reaches its maximum over the critical value, and repeat the detection procedure on the new segments until no further change–point is found. However, the problem of optimal resolution in the segmentation $\boldsymbol{\tau}$ is not solved as no penalized objective function is considered.

Remark 4. Most applied econometrics research papers supposedly using the multiple change–point tests by Lavielle (1999) and Lavielle and Moulines (2000) do in fact resort to the binary segmentation algorithm, which is the less we can say missleading.

3.1 Weakly Dependent Processes

We present here the tests for single change–point in the variance of time series by Inclán and Tiao (1994) and Kokoszka and Leipus (1999). The test by Inclán and Tiao (1994) for change in the variance of a weakly dependent process $\{Y_t\}$ is based on the process $\{D_n(h), h \in [0,1]\}$ defined as

$$D_n(h) := \frac{\sum_{j=1}^{[nh]} Y_j^2}{\sum_{j=1}^{n} Y_j^2} - \frac{[nh]}{n}, \quad h \in [0,1]. \tag{23}$$

Under the null hypothesis of constant unconditional variance, the process $\{D_n(h), h \in [0,1]\}$ converges to a Brownian bridge on $[0,1]$. A test for constancy of the unconditional variance is based on the following functional of the process $\{D_n(h)\}$, which under this null hypothesis of constant unconditional variance converges in distribution to the supremum of a Brownian bridge on $[0,1]$

$$\sqrt{n/2} \sup_{0 \leqslant h \leqslant 1} |D_n(h)| \xrightarrow{d} \sup_{0 \leqslant h \leqslant 1} \left|W^0(h)\right|. \tag{24}$$

where $W^0(h)$ is the Brownian bridge on the unit interval $[0,1]$ defined as $W^0(h) = W(h) - hW(1)$, $W(h)$ is the Wiener process.

Kokoszka and Leipus (1999) made the assumption that the process $\{Y_t\}$ follows an ARCH(∞) process defined as

$$\begin{aligned} Y_t &= \sigma_t \varepsilon_t, \quad \varepsilon_t \sim \text{iid}, \quad E\varepsilon_0 = 0, \quad \operatorname{Var} \varepsilon_0 = 1, \\ \sigma_t^2 &= \omega + \sum_{j=1}^{\infty} \alpha_j Y_{t-j}^2, \quad t = 1, \ldots, t_0, \\ \sigma_t^2 &= \omega^\star + \sum_{j=1}^{\infty} \alpha_j^\star Y_{t-j}^2, \quad t = t_0 + 1, \ldots, n, \end{aligned} \tag{25}$$

with the assumption that the unconditional variance of the process changes at an unknown time t_0, i.e.,

$$\Delta(n) = \frac{\omega}{1 - \sum_{j=1}^{\infty} \alpha_j} - \frac{\omega^\star}{1 - \sum_{j=1}^{\infty} \alpha_j^\star} \neq 0. \tag{26}$$

The null hypothesis is $H_0 : \omega = \omega^\star,\ \alpha_j = \alpha_j^\star$ for all j, while under the alternative hypothesis $H_A : \omega \neq \omega^\star$ or $\alpha_j \neq \alpha_j^\star$ for some j. The change–point test is based on the process $\{U_n(h), h \in [0,1]\}$ defined as

$$U_n(h) := \sqrt{n} \frac{[nh](n - [nh])}{n^2} \left(\frac{1}{[nh]} \sum_{j=1}^{[nh]} Y_j^2 - \frac{1}{n - [nh]} \sum_{j=[nh]+1}^{n} Y_j^2 \right), \tag{27}$$

which under H_0 converges to the process $\left\{\sigma W^0(h), h \in [0,1]\right\}$, i.e.,

$$U_n(h) \overset{\mathcal{D}[0,1]}{\longrightarrow} \sigma W^0(h), \tag{28}$$

where $\overset{\mathcal{D}[0,1]}{\longrightarrow}$ means weak convergence in the space $\mathcal{D}[0,1]$ endowed with the Skorokhod topology. We consider here as test statistic the functional based on the process $\{U_n(h), h \in [0,1]\}$

$$\sup_{0 \leqslant h \leqslant 1} |U_n(h)| / \sigma \overset{d}{\longrightarrow} \sup_{0 \leqslant h \leqslant 1} \left|W^0(h)\right|, \tag{29}$$

where the long–run variance σ^2 is usually estimated by nonparametric kernel methods. We use here the heteroskedastic and autocorrelation consistent (HAC) estimator by Newey and West (1987) with the truncations order $q = 0, 2, 5, 10, 15$.

The location of the change–point $\hat{\tau}$ is detected by the CUSUM–type estimator based on the same process $\{U_n(h), h \in [0,1]\}$, and defined by

$$\hat{\tau} = [n\hat{h}], \quad \hat{h} = \min\left\{h : |U_n(h)| = \max_{0<h \leqslant 1} |U_n(h)|\right\}. \tag{30}$$

This estimator is consistent if $\Delta(n) \to 0$ as $n \to \infty$ but at a slower rate than $n^{1/2}$ as

$$|\Delta(n)| n^{1/2} \to \infty, \quad n \to \infty,$$

see Kokoszka and Leipus (2000) for further details.

3.2 Strongly Dependent Processes

In the previous section, the process $\{Y_t\}$ was assumed weakly dependent. Whitcher *et al.* (2002) proposed to deal with long–range dependent processes with an unknown number of change–points in the unconditional variance, by applying the BS procedure to the discrete wavelet transform of the long–memory process $\{Y_t\}$.

4 Detecting Change–Points in the Volatility of Financial Time Series

We consider two series, the FTSE 100 index and the US dollar–Japanese yen intra–day foreign exchange (FX) rate.

4.1 Application to The FTSE 100 Index

The FTSE 100 index, or Footsie, consists of 100 blue chip stocks that trade on the London Stock Exchange. This series of 4381 observations has been observed between January 1984 and November 2002. Figure 2 displays the series of indices in levels:

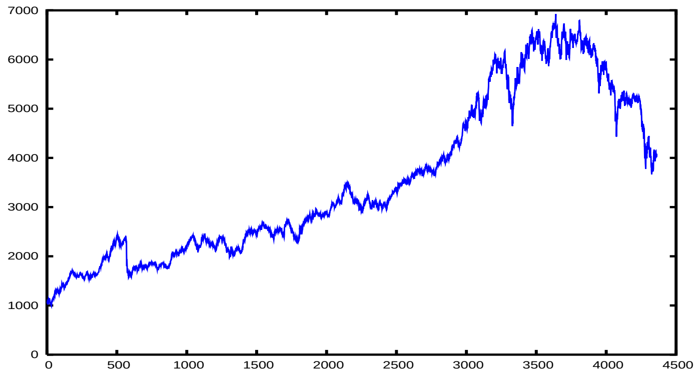

Fig. 2. The series of FTSE 100 indices

Table 1 below displays the sequence of change–points K_i, lengths l_{K_i} and P–values given by Algorithm **1**, see also equation (22).

Table 1. Sequences of number of change–points K_i, lengths l_{K_i} and corresponding P–values $\mathcal{P}_{K_i}$ given by Algorithm **1**

K_i	l_{K_i}	$\mathcal{P}_{K_i}$
1	∞	5.0000e-05
3	152.9601	9.7200e-07
4	68.9379	6.8018e-04
6	50.9085	9.1889e-07
7	32.4764	6.5439e-06
8	7.4296	2.8738e-01
11	5.6321	2.6108e-01
13	5.6107	1.3535e-03
15	5.4152	3.7485e-02

Figure 3 below shows that Algorithm **1** is able to pick the main changes in the unconditional variance of the series of returns on FTSE 100 .

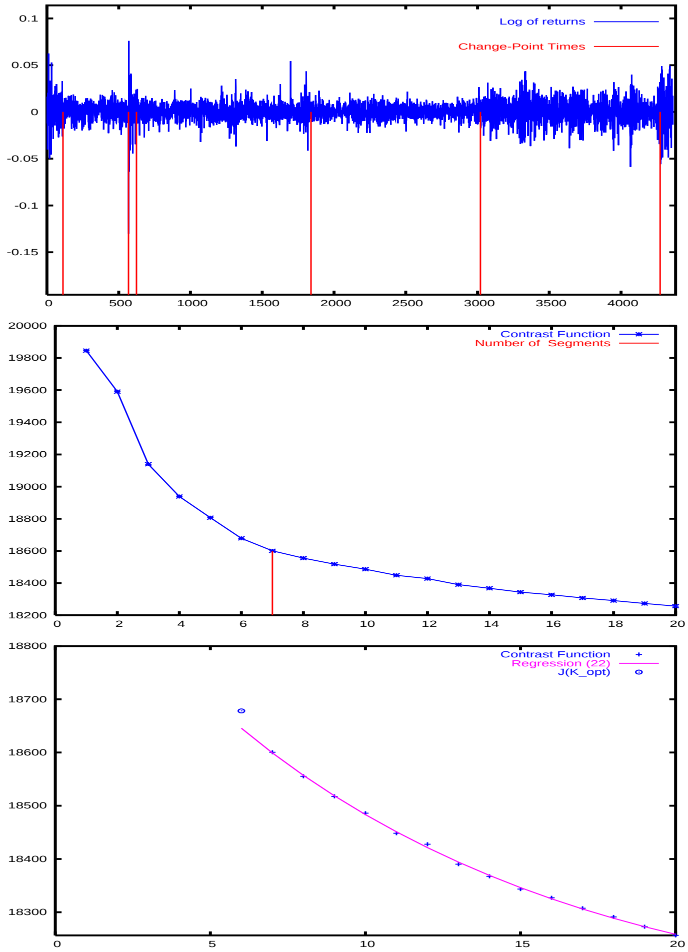

Fig. 3. Above: the series with the estimated change–points represented by vertical lines; Middle: The sequence of contrasts $(J_K, 1 \leqslant K \leqslant K_{MAX})$, the vertical line indicates the estimated number of segments $(\hat{K}, J_{\hat{K}})$; Below: the sequence of contrasts $\{J_K, \hat{K} \leqslant K \leqslant K_{MAX}\}$ are indicated with $+$, the fitted function $\hat{c}_1(K) + \hat{c}_2 K \log(K)$ is in solid line and $J_{\hat{K}}$ is represented with a circle

We obtain the segmentation $\hat{\boldsymbol{\tau}} = \{112, 568, 624, 1840, 3020, 4272\}$. The point $\boldsymbol{\tau}_1 = 112$ matches a change in the sampling frequency, as we have weekly data before January 1986, and daily observations after that date. Thus, the procedure detects this heterogeneity in the process. The point $\boldsymbol{\tau}_2 = 568$ is simply the 14^{th} October 1987, i.e., the stock market crash, while the increase of volatility after $\boldsymbol{\tau}_5 = 3020$ (June 26, 1997) indicates the conjunction of two opposite phenomena: the Footsie has broken the psychological 5,000 barrier in August 1997, as a consequence of a series of positive earnings for the companies composing the index. On the other side, the Asian crisis of Summer 1997 increased the uncertainty, and then the volatility, as the extent of the consequences of this crisis on economic activity were unpredictable.

Figure 4 below displays the sample autocorrelation function (ACF) for the whole series of absolute returns on the FTSE index:

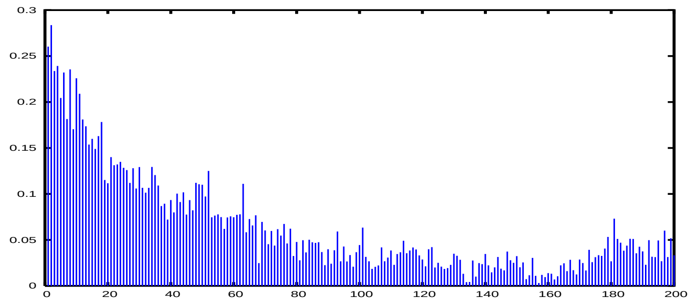

Fig. 4. Sample ACF of the absolute value of returns $|r_t|$ on FTSE

This sample ACF has a hyperbolic decay which is similar to the one of a strongly dependent process: the ACF are always positive, with a plateau for the larger orders of autocorrelation. However, when displaying the sample ACF for the sub–intervals defined by Algorithm **1**, we get the following pictures, the shape of which are very different from Figure 4:

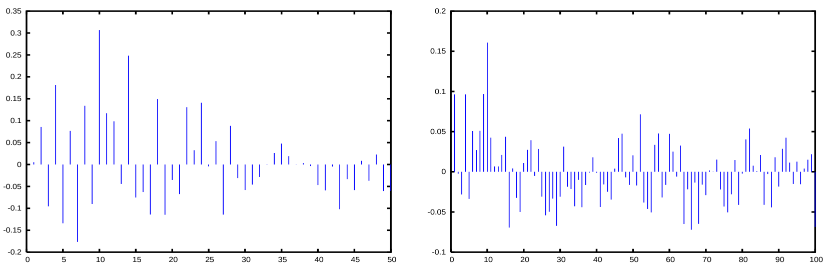

Fig. 5. Sample ACFs of the absolute value of returns $|r_t|$ on FTSE for the time interval $[1, 112]$ (left), and for the time interval $[113, 568]$ (right)

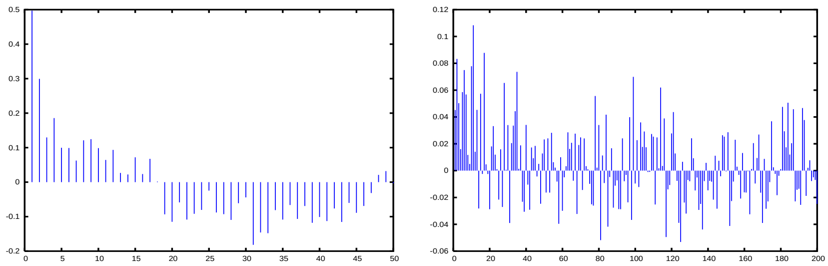

Fig. 6. Sample ACF of the absolute value of returns $|r_t|$ on FTSE for the time interval $[569, 624]$ (left), and for the time interval $[625, 1840]$ (right)

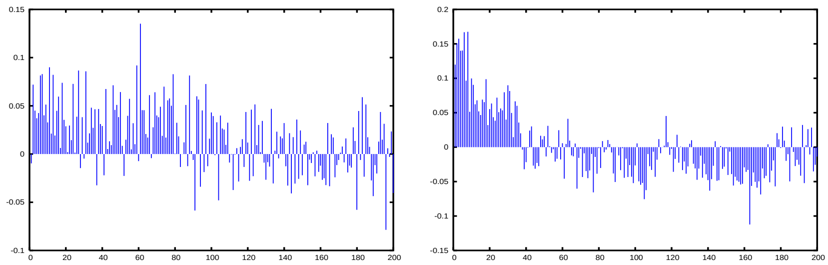

Fig. 7. Sample ACF of the absolute value of returns $|r_t|$ on FTSE for the time interval $[1841, 3020]$ (left), and for the time interval $[3021, 4272]$ (right)

The sample ACF of the absolute value of returns $|r_t|$ on FTSE for the time interval $[4273, 4380]$ displays a pattern similar to Figure 6 (right) and is not displayed here.

For all the sub–samples, the sample ACF displayed in figures 5–7 do not indicate the same degree of persistence as the one observed for the whole sample in Figure 4: some autocorrelations are negative, and these ACF do not display the "plateau effect" for the higher orders. Thus, as mentioned in the introduction of this chapter, the hypothesis of homogeneity and stationarity of the returns process is inappropriate, and the global procedure for finding the optimal resolution for the process is able to pick the nonstationarities of the process out.

Choosing the level of resolution just below, i.e., with 6 segments, would have given $\hat{\boldsymbol{\tau}} = \{112, 568, 624, 3048, 4272\}$. However, considering the period between $t = 624$ and $t = 3048$, i.e., between the 5^{th} of January 1988 and the 4^{th} of August 1997 as homogeneous is rather unlikely. Figure 8 below displays the sample ACF for the absolute returns for this time interval.

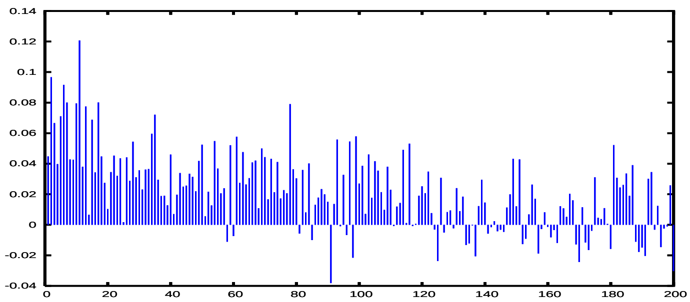

Fig. 8. Sample ACF of the absolute value of returns $|r_t|$ on FTSE for the time interval $[624, 3047]$

The sample ACF resembles the one of a long–range dependent process, with a mild degree of persistence. However, we have seen before that the sample ACF for the two sub–periods $[624, 1847]$ and $[1848, 3019]$ have a different shape, which indicates that the mild persistence for the interval $[624, 3047]$ is a statistical artefact. Thus, the resolution with 7 segments looks preferable.

We compare the selected segmentation with the one obtained with the binary segmentation procedure. Table 2 below reports the segmentation yield by the BS method, using both statistics given by equations (24) and (29). We observe that the two statistics give a quite similar segmentation, which however has a higher resolution than the one yield by Algorithm **1**. The segmentation given by the BS method includes the points found by Algorithm **1**, or points close to those of $\hat{\boldsymbol{\tau}}$.

Table 2. Segmentation found by the BS method, Inclán and Tiao (1994), henceforth IT, and Kokoszka and Leipus (1999), henceforth KL

KL statistic	Change–point date	IT statistic	Change–point date
9.0585	110	5.5376	110
12.2929	570	6.1560	570
13.2107	648	10.0611	648
1.8179	1062	1.6352	1062
1.3997	1113	1.9091	1273
2.6116	1273	2.0503	1324
2.5349	1324	2.3089	1703
2.6729	1703	2.6606	1838
2.8486	1838	1.5992	1943
2.3037	2117	2.3103	2117
2.1048	2458	2.2611	2458
4.2924	3000	4.3451	3000
1.6788	3173	1.3736	3173
11.2412	3284	7.4494	3284
4.8933	3418	2.4585	3418
3.2717	3654	1.9844	3654
2.6650	3761	1.7056	3761
5.0372	3955	2.5301	3955
3.2697	4128	1.7171	4128
8.6904	4270	4.3607	4270

If we try to refine the segmentation by choosing a number of change–points similar to the BS method, we get the following picture

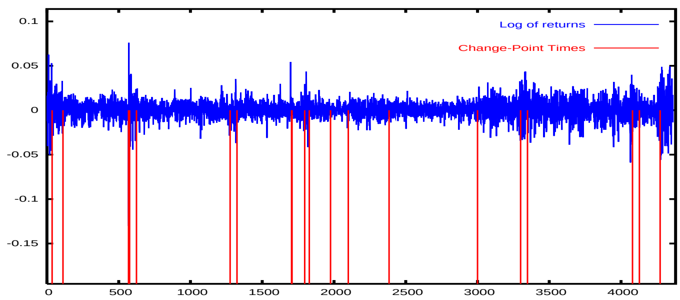

Fig. 9. The series with the 20 estimated change–points represented by vertical lines

We capture more and more details of the variations of the process, but the gain is rather marginal, as the main variations are captured with 7 segments.

4.2 Application to the US Dollar–Japanese Yen FX Rate

We consider here a sample of 30 minute–spaced observations observed in the year 1996. These data, provided by Olsen & Associates are in ϑ time, i.e., all intra–day seasonal components have been removed.

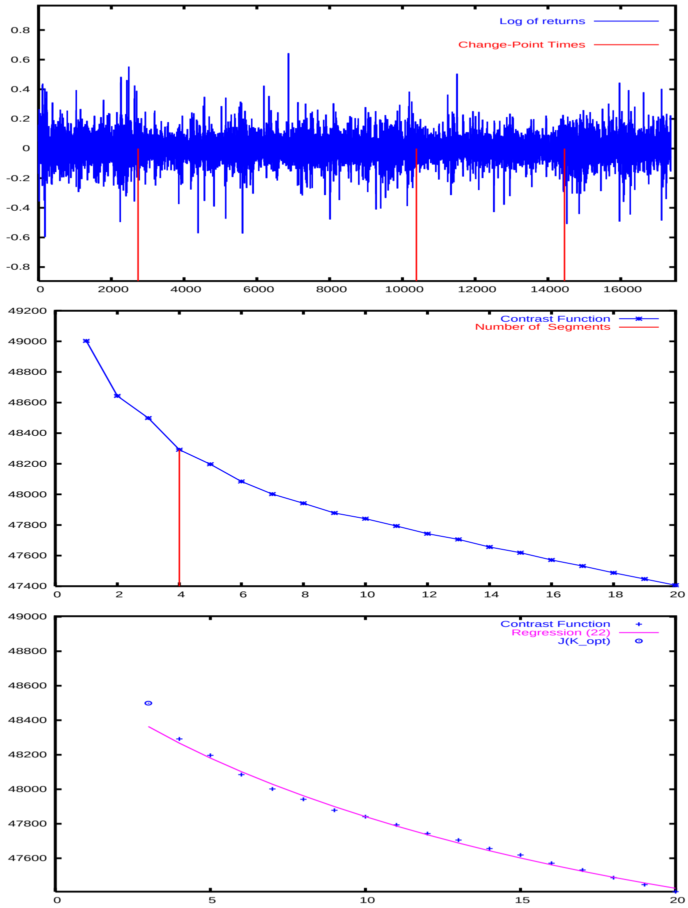

Fig. 10. Above: the series with the estimated change–points represented by vertical lines; Middle: The sequence of contrasts $(J_K, 1 \leqslant K \leqslant K_{MAX})$, the vertical line indicates the estimated number of segments $(\hat{K}, J_{\hat{K}})$; Below: the sequence of contrasts $\{J_K, \hat{K} \leqslant K \leqslant K_{MAX}\}$ are indicated with $+$, the fitted function $\hat{c}_1(K) + \hat{c}_2 K \log(K)$ is in solid line and $J_{\hat{K}}$ is represented with a circle

Figure 10 displays the series with the estimated change–points, the contrast function J_K, and the fitted function $\hat{c}_1(K) + \hat{c}_2 K \log(K)$.

The detected number of change points is rather low for the sample size considered, but we have to keep in mind that this series represents only a year of observations, so that structural changes are rather rare, even for data sampled with a high frequency.

Table 3 below displays the sequence of the number of change–points K_i found by Algorithm **1**.

Table 3. Sequences of number of change–points K_i, lengths l_{K_i} and corresponding P–values $\mathcal{P}_{K_i}$ given by Algorithm **1**

K_i	l_{K_i}	$\mathcal{P}_{K_i}$
1	∞	5.0000e-05
2	183.5677	6.8982e-05
4	72.2762	3.1535e-05
7	21.4666	2.4397e-04
9	16.3938	9.4880e-03
12	1.7163	1.7154e-01

The chapter by Teyssière and Abry (2005) in this volume considers the wavelet analysis of this series: they compare the wavelet estimator of the degree of persistence for the absolute returns of this series with the local Whittle and log–periodogram estimators. The discrepancy between the estimation results obtained with the wavelet based estimator and the ones obtained from the spectral based estimators is interpreted as the consequence of nonstationarities in the returns process.

The BS procedure finds a far larger number of change–points, i.e., 103 for the KL statistic and 95 for the IT statistic. The graphical representation of the segmentation yield by algorithm **1** looks however more sensible.

In fact, the segmentations yield by competing statistical methods are very different. Mikosch and Stărică (1999) and Granger and Hyung (2004) studied the series of S&P 500, using respectively parametric and semiparametric tests. While Mikosch and Stărică (1999) found a rather parsimonious segmentation, the number of change–points found by Granger and Hyung (2004) is huge.

However, this parsimonious segmentation looks relevant when comparing the sample ACF of the absolute returns of the whole series and of the sub–intervals defined by Algorithm **1**, see figures 11–13: while the sample ACF, computed on the whole sample, Figure 11, is similar to the one of a strongly dependent process, the patterns of the sample ACFs for the sub–intervals defined by Algorithm **1** show that the persistence on these sub–intervals is far smaller than the one for the whole sample.

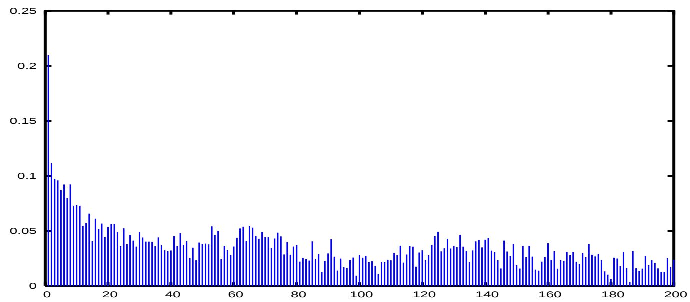

Fig. 11. Sample ACF of the absolute value of returns $|r_t|$ on USD–JPY FX rate

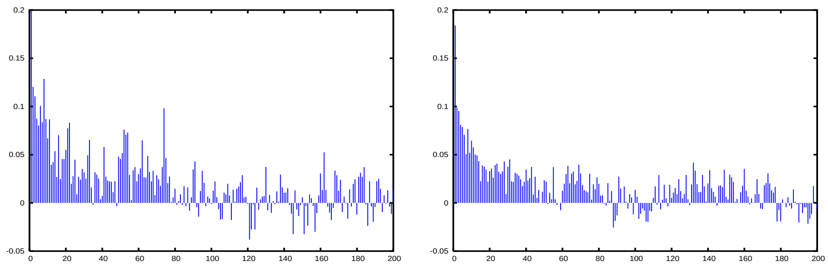

Fig. 12. Sample ACF of the absolute value of returns $|r_t|$ on USD–JPY FX rate for the time interval $[1, 2736]$ (left), and for the time interval $[2737, 10386]$ (right)

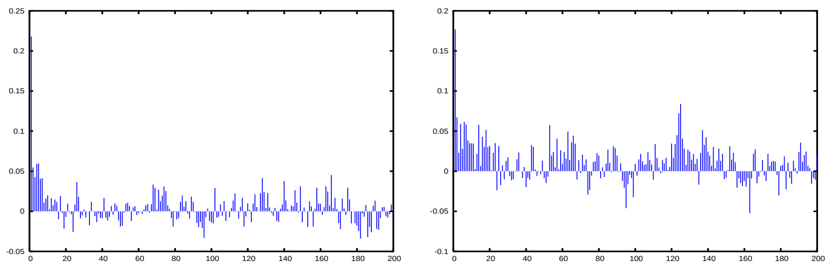

Fig. 13. Sample ACF of the absolute value of returns $|r_t|$ on USD–JPY FX rate for the time interval $[10387, 14454]$ (left), and for the time interval $[14455, 17508]$ (right)

4.3 Application to Micro–Simulated Data

We consider here simulated series from an artificial financial market, i.e., a dynamic system which models financial markets with interacting agents. Although these models do not resort to statistical distributions leading to the generation of long–range dependent processes, the volatility series generated by these models display the same dependence properties as the ones of the volatility of asset prices.

We consider that agents i on financial markets differ by their forecasting function $E^i(P_{t+1}|I_t)$ of the future price as a function of the information set I_t. Chartists extrapolate the exchange rate P_{t+1} by using is a linear function of the previous prices, i.e.,

$$E^c\left(P_{t+1}|I_t\right) = \sum_{j=0}^{M^c} h_j P_{t-j}, \tag{31}$$

where h_j, $j = 0, \ldots, M^c$ are constants, M^c is the 'memory' of the chartists, while fundamentalists forecast this next price as:

$$E^f\left(P_{t+1}|I_t\right) = \bar{P}_t + \sum_{j=1}^{M^f} \nu_j(P_{t-j+1} - \bar{P}_{t-j}), \tag{32}$$

where ν_j, $j = 1, \ldots, M^f$ 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' $\bar{P}_t$, which can be thought as the price if it were only to be explained by a set of relevant variables, follows a random walk:

$$\bar{P}_t = \bar{P}_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2). \tag{33}$$

Agents have the possibility of investing at home in a risk free asset or investing abroad in a risky asset. We denote by ρ_t the foreign interest rate, by d_t^i the demand by the i^{th} agent for foreign currency, and by r the domestic interest rate, with $\rho_t > r$. The exchange rate P_t and the foreign interest rate ρ_t are considered by agents as independent random variables, with $\rho_t \sim N(\rho, \sigma_\rho^2)$. The cumulated wealth of individual i at time $t+1$, W_{t+1}^i is given by:

$$W_{t+1}^i = (1 + \rho_{t+1})P_{t+1}d_t^i + (W_t^i - P_t d_t^i)(1 + r). \tag{34}$$

Agents i have a standard mean–variance utility function:

$$U(W_{t+1}^i) = E(W_{t+1}^i) - \lambda \mathrm{Var}(W_{t+1}^i), \tag{35}$$

where λ denotes the risk aversion coefficient, and

$$E(W_{t+1}^i|I_t) = (1 + \rho)E^i(P_{t+1}|I_t)d_t^i + (W_t^i - P_t d_t^i)(1 + r), \tag{36}$$

$$\mathrm{Var}(W_{t+1}^i|I_t) = (d_t^i)^2 \zeta_t, \quad \zeta_t = \mathrm{Var}\left(P_{t+1}(1 + \rho_{t+1})\right). \tag{37}$$

Demand d_t^i is found by maximizing utility. First order condition gives

$$(1+\rho)E^i(P_{t+1}|I_t) - (1+r)P_t - 2\zeta_t\lambda d_t^i = 0, \tag{38}$$

where $E^i(.|I_t)$ denotes the forecast of an agent of type i. Let k_t be the proportion of fundamentalists at time t, the market demand is:

$$d_t = \frac{(1+\rho)\left(k_t E^f(P_{t+1}|I_t) + (1-k_t)E^c(P_{t+1}|I_t)\right) - (1+r)P_t}{2\zeta_t\lambda}. \tag{39}$$

Now consider the exogenous supply of foreign exchange X_t, then the market is in equilibrium if aggregate supply is equal to aggregate demand, i.e., $X_t = d_t$, which gives

$$P_t = \frac{1+\rho}{1+r}\left(k_t E^f(P_{t+1}|I_t) + (1-k_t)E^c(P_{t+1}|I_t)\right) - \frac{2\zeta_t\lambda X_t}{1+r}. \tag{40}$$

From equation (40), 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 $E^f(P_{t+1}|I_t)$ and $E^c(P_{t+1}|I_t)$. Several mechanisms for the evolution of the opinion process $\{k_t\}$ have been proposed in the literature, which are either based on epidemiologic phenomenon, or on a preference given to the most performing forecasting function, or on the accumulated wealth gained with each forecast function, etc. Interested readers are referred to Teyssière (2003) and the chapter by Gaunersdorfer and Hommes (2005) in this volume.

We consider a multivariate extension of this model, i.e., a bivariate process $(P_{1,t}, P_{2,t})$ of foreign exchange rates. This is motivated by the fact that structural changes do not affect singles markets, i.e., the same swing in opinions from chartists to fundamentalists affects linked markets. It has been suggested in the 1999 version of the work by Granger and Hyung that these common breaks might explain the common persistence of asset prices volatility. Indeed, the bivariate common break process by Teyssière (2003) used here, generates the same type of dependence as the one observed in multivariate financial time series.

We then consider that the opinion process $\{k_t\}$ is the same for both markets. This bivariate foreign exchange rate process depends on a pair of foreign interest rates (ρ_1, ρ_2). We assume that $2\zeta_{i,t}\lambda X_{i,t}/(1+\rho_i) = \gamma_i \bar{P}_{i,t}$ for $i = 1, 2$, and $M^f = M^c = 1$, the equilibrium price for the bivariate model is given by

$$\begin{pmatrix} P_{1,t} \\ P_{2,t} \end{pmatrix} = \begin{pmatrix} \frac{k_t-\gamma}{A_1}\bar{P}_{1,t} - \frac{k_t\nu_{1,1}}{A_1}\bar{P}_{1,t-1} + \frac{(1-k_t)h_{1,1}}{A_1}P_{1,t-1} \\ \frac{k_t-\gamma}{A_2}\bar{P}_{2,t} - \frac{k_t\nu_{2,1}}{A_2}\bar{P}_{2,t-1} + \frac{(1-k_t)h_{2,1}}{A_2}P_{2,t-1} \end{pmatrix}, \tag{41}$$

with

$$A_i = \frac{1+r}{1+\rho_i} - (1-k_t)h_{i,0} - k_t\nu_{i,1}. \tag{42}$$

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

$$\begin{pmatrix} \bar{P}_{1,t} \\ \bar{P}_{2,t} \end{pmatrix} = \begin{pmatrix} \bar{P}_{1,t-1} \\ \bar{P}_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}, \quad \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{1,1}^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_{2,2}^2 \end{pmatrix}\right], \quad (43)$$

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 Teyssière (1997, 2003).

We generate here a bivariate series of returns. Figure 14 below displays the two generated series of returns $r_{1,t}, r_{2,t}$, with $r_{1,t} = \ln(P_{1,t}/P_{1,t-1})$ and $r_{2,t} = \ln(P_{2,t}/P_{2,t-1})$, and the detected changes in their unconditional variance. We can see that the detected changes for the series $r_{2,t}$ are very close to some of the detected changes for the series $r_{1,t}$, i.e., $\hat{\boldsymbol{\tau}} = \{676, 868, 1360, 1584, 2408, 4032, 4148\}$ for the series $r_{1,t}$, while $\hat{\boldsymbol{\tau}} = \{1360, 1580, 4144\}$ for the series $r_{2,t}$, which is not very surprising as the opinion process $\{k_t\}$ is common for both processes $r_{1,t}$ and $r_{2,t}$. The joint detection of change–points in multivariate time series is considered in a subsequent paper; see Lavielle and Teyssière (2005).

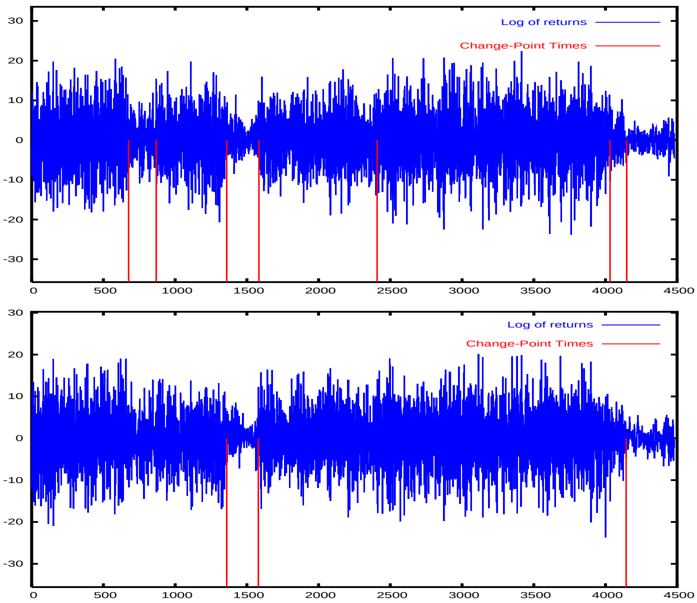

Fig. 14. The two jointly simulated series, $r_{1,t}$ above and $r_{2,t}$ below, with the estimated change–points represented by vertical lines

We focus on the second returns series $r_{2,t}$, with the lowest resolution level: its sample ACF, see Figure 15, is similar to the one of a strongly dependent process. However, the sample ACF for the sub–intervals defined by Algorithm **1**, see figures 16 and 17, display a different pattern with both positive and negative autocorrelations, a property similar to what has been observed with the two previous examples.

Thus, Algorithm **1** is able to detect the nonstationarities of the returns process generated by the artificial financial market.

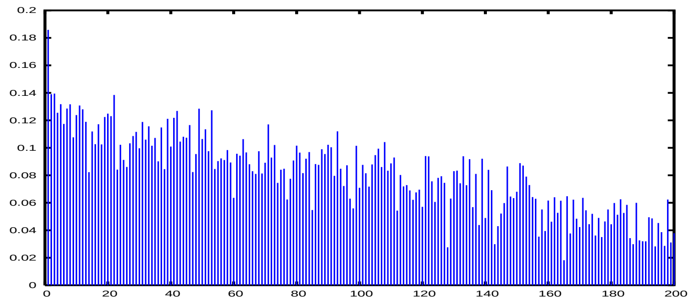

Fig. 15. Sample ACF of the absolute value of simulated returns $|r_{2,t}|$

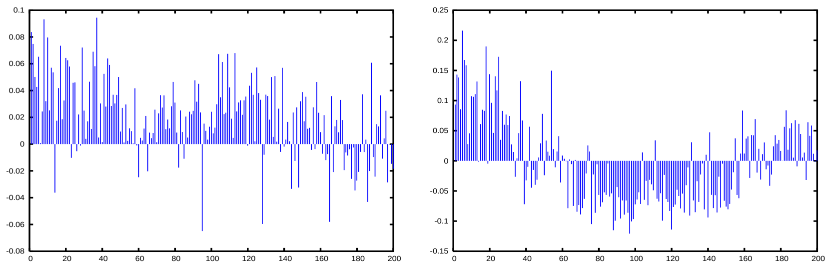

Fig. 16. Sample ACF of the absolute value of simulated returns $|r_{2,t}|$ for the time interval $[1, 1360]$ (left), and for the time interval $[1361, 1580]$ (right)

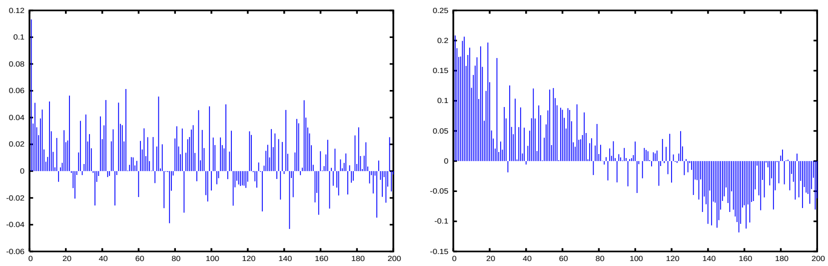

Fig. 17. Sample ACF of the absolute value of simulated returns $|r_{2,t}|$ for the time interval $[1581, 4144]$ (left), and for the time interval $[4145, 4500]$ (right)

5 Conclusion: Detecting Break in the Variance of Returns or in the Mean of Absolute Returns?

Since we checked the adequacy of the resolution by looking at the ACF of the sequence of absolute returns, one might think that changes in the volatility are uncovered by detecting changes in the mean of the absolute returns series, i.e., instead of the contrast function given by equation (5) one might consider the following contrast function based again on a Gaussian likelihood function

$$J_n(\boldsymbol{\tau}, \boldsymbol{y}) = \frac{1}{n} \sum_{k=1}^{K} \| Y_{\tau_k} - \bar{Y}_{\tau_k} \|. \tag{44}$$

Applying again Algorithm **1** to the series of absolute returns $|r_t|$ on the FTSE 100 index, we select different values for the number of change–points K_i from 9 to 12, and obtained the following segmentations:

- $\hat{\boldsymbol{\tau}} = \{112, 568, 576, 624, 3300, 3348, 4080, 4092, 4272\}$,
- $\hat{\boldsymbol{\tau}} = \{112, 568, 576, 624, 3300, 3348, 4080, 4092, 4284, 4304\}$,
- $\hat{\boldsymbol{\tau}} = \{112, 568, 576, 624, 3300, 3340, 3348, 4080, 4092, 4284, 4304\}$,
- $\hat{\boldsymbol{\tau}} = \{112, 568, 576, 624, 1856, 3004, 3312, 3348, 4080, 4092, 4284, 4304\}$,

i.e., we have to consider a large number of segments, 13, for splitting the interval $[624, 3300]$. Thus, considering the series of absolute returns is not suitable for finding the optimal resolution of the volatility series.

Alternatively, one might detect both changes in the mean and the variance for the series of absolute returns by considering the following contrast function:

$$J_n(\boldsymbol{\tau}, \boldsymbol{y}) = \frac{1}{n} \sum_{k=1}^{K} \frac{\| Y_{\tau_k} - \bar{Y}_{\tau_k} \|}{\hat{\sigma}_k^2} + n_k \log(\hat{\sigma}_k^2). \tag{45}$$

In that case, algorithm **1** selected the following partition with 12 change–points:

$$\hat{\boldsymbol{\tau}} = \{112, 568, 576, 624, 1796, 1828, 3020, 3304, 3348, 4080, 4128, 4272\},$$

which has a level of resolution far higher than the one obtained when detecting changes in the unconditional variance of the returns process. We obtain similar results for the series of returns on US dollar–Japanese Yen FX rate, and the series of results generated by the artificial microeconomic model. Thus, the straightforward and natural way for detecting changes in the volatility is to consider the contrast function defined by equation (5).

References

1. Aggarwal, R., Inclán, C. and Leal, R. (1999). Volatility in emerging stock markets. *Journal of Financial and Quantitative Analysis*, **34,** 1–17.

2. Basseville, M. and Nikiforov, N. (1993). *The Detection of Abrupt Changes - Theory and Applications*. Prentice-Hall: Information and System Sciences Series.
3. Battacharya, P.K. (1987). Maximum Likelihood estimation of a change–point in the distribution of independent random variables: general multiparameter case. *Journal of Multivariate Analysis*, **32,** 183–208.
4. Beran, J. and Terrin, N. (1999). Correction: Testing for a change of the long–memory parameter. *Biometrika*, **86,** 233.
5. Beran, J. and Terrin, N. (1996). Testing for a change of the long–memory parameter. *Biometrika*, **83,** 627–638.
6. Berkes, I., Horváth, L. and Kokoszka, P.S. (2004). Testing for parameter constancy in GARCH(p, q) models. *Statistics and Probability Letters*, in press.
7. 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.
8. Berkes, I., Horváth, L., Kokoszka, P.S. and Shao, Q-M. (2003). On discriminating between long–range dependence and changes in mean. *Preprint*.
9. Bhattacharya, R.N., Gupta, V.K. and Waymire, E. (1983). The Hurst effect under trends. *Journal of Applied Probability*, **20,** 649–662.
10. Birgé, L. and Massart, P. (2001). A generalized C_p criterion for Gaussian model selection. *Preprint Université Paris-Sud*.
11. 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.
12. Brodsky. B. and Darkhovsky, B. (1993). *Nonparametric Methods in Change–Point Problems*. Kluwer Academic Publishers, The Netherlands.
13. Chu, C.-S.J. (1995). Detecting parameter shift in GARCH models. *Econometric Reviews*, **14,** 241–266.
14. Csörgö, M. and Horváth, L. (1997). *Limit Theorems in Change–Point Analysis*. Wiley.
15. Dalhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics*, **25,** 1–37.
16. Gaunersdorfer, A. and Hommes, C.H. (2005). A nonlinear structural model for volatility clustering, in: Teyssière, G. and Kirman, A., (Eds.), *Long–Memory in Economics*. Springer Verlag, Berlin. Appears in this volume.
17. Gijbels, I., Hall, P. and Kneip, A. (1999). On the estimation of jump points in smooth curves. *The Annals of the Institute of Statistical Mathematics*, **51,** 231–251.
18. Giraitis, L., Leipus, R. and Surgailis, D. (2005). Recent advances in ARCH modelling, in: Teyssière, G. and Kirman, A., (Eds.), *Long–Memory in Economics*. Springer Verlag, Berlin. Appears in this volume.
19. Giraitis, L. and Leipus, R. (1992). Testing and estimating in the change–point problem of the spectral function. *Lithuanian Mathematical Journal*, **32,** 20–38.
20. 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.
21. Giraitis, L., Leipus, R. and Surgailis, D. (1996). The change–point problem for dependent observations. *Journal of Statistical Planning and Inference*, **53,** 297–310.
22. Granger, C.W.J. (2002). Long memory, volatility, risk and distribution. *Preprint*.

23. Granger, C.W.J. (2000). Current perspectives on long memory processes. *Academia Economic Papers*, **28,** 1–16.
24. Granger, C.W.J. and Hyung, N. (2004). Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance*, **11,** 399–421.
25. Granger, C.W.J. and Hyung, N. (1999). Occasional structural breaks and long–memory. *Preprint.*
26. Granger, C.W.J. and Ding, Z. (1995). Some properties of absolute returns, an alternative measure of risk. *Annales d'Économie et de Statistique*, **40,** 67–91.
27. Hawkins, D.L. (1977). Testing a sequence of observations for a shift in location. *Journal of the American Statistical Association*, **72,** 180–186.
28. Horváth, L. (2001). Change–point detection in long–memory processes. *Journal of Multivariate Analysis*, **78,** 218–234.
29. 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.
30. Horváth, L. and Shao, Q-M. (2001). Limit theorems for quadratic forms with application to the Whittle's estimate. *Annals of Applied Probability,* **9,** 146–187.
31. Inclán, C. and Tiao, G.C. (1994). Use of cumulative sums of squares for retrospective detection of changes in variance. *Journal of the American Statistical Association*, **89,** 913–923.
32. Kay, S.M. (1998). *Fundamentals of Statistical Signal Processing: Detection Theory.* Prentice Hall: Signal Processing Series.
33. Kim, H.J. (1994). Likelihood ratio and cumulative sum tests for a change–point in linear regression. *Journal of Multivariate Analysis*, **51,** 54–70.
34. Kokoszka, P.S. and Leipus, R. (1999). Testing for parameter changes in ARCH models. *Lithuanian Mathematical Journal*, **39,** 231–247.
35. Kokoszka, P.S. and Leipus, R. (2000). Change–point estimation in ARCH models. *Bernoulli*, **6,** 513–539.
36. Kokoszka, P.S. and Leipus, R. (2002). Detection and estimation of changes in regime, in: M.S. Taqqu, G. Oppenheim and Doukhan P. (Eds.), *Long-Range Dependence: Theory and Applications*, 325–337. Birkhauser.
37. Kokoszka, P.S. and Teyssière, G. (2002). Change–point detection in GARCH models: Asymptotic and bootstrap tests. *Preprint.*
38. Lavielle, M. (1999). Detection of multiple changes in a sequence of dependent variables. *Stochastic Processes and their Applications*, **83,** 79–102.
39. Lavielle, M. (1998). Optimal segmentation of random processes. *IEEE Transactions on Signal Processing*, **46,** 1365–1373.
40. Lavielle, M. (1993). Detection of changes in the spectrum of a multidimensional process. *IEEE Transactions on Signal Processing*, **41,** 742–749.
41. Lavielle, M. and Teyssière, G. (2005). Detection of multiple change–points in multivariate time series. *Preprint.*
42. Lavielle, M. and Lebarbier, E. (2001). An application of MCMC methods to the multiple change–points problem. *Signal Processing*, **81,** 39–53.
43. Lavielle, M. and Ludeña, C. (2000). The multiple change–points problem for the spectral distribution. *Bernoulli*, **6,** 845–869.
44. Lebarbier, E. (2001). Adaptive estimation of a piecewise continuous signal in Gaussian noise. *Preprint Université Paris-Sud.*
45. Lombard, F. (1987). Rank tests for change point problems. *Biometrika*, **74,** 615–624.

46. Mia, B.Q. and Zhao, L.C. (1988). Detection of change points using rank methods. *Communications in Statistics - Theory and Methods*, **17,** 3207–3217.
47. Mikosch, T. and Stărică, C. (2004). Change of structure in financial time series and the GARCH model. *REVSTAT Statistical Journal*, **2,** 41–73.
48. 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.
49. Mikosch, T. and Stărică, C. (1999). Change of structure in financial time series, long range dependence and the GARCH model. *Preprint.*
50. Newey, W.K. and West, K.D. (1987). A simple positive definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, **55,** 703–708.
51. Robinson, P.M. (1991). Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics*, **47,** 67–84.
52. 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.
53. Sen, A. and Srivastava, M.S. (1975). On tests for detecting change in the mean. *The Annals of Statistics*, **3,** 96–103.
54. Stărică, C. and Granger, C.W.J. (2001). Nonstationarities in stock returns. *Review of Economics and Statistics*, forthcoming.
55. Teyssière, G. and Abry, P.(2005). Wavelet analysis of nonlinear long–range dependent processes. Applications to financial time series in: Teyssière, G. and Kirman, A., (Eds.), *Long–Memory in Economics.* Springer Verlag, Berlin. Appears in this volume.
56. Teyssière, G. (2003). Interaction models for common long–range dependence in asset price volatility, in: Rangarajan, G., Ding, M. (Eds.), *Processes with Long Range Correlations: Theory and Applications,* Lecture Notes in Physics, Vol. 621. Springer Verlag, Berlin, pp. 251–269.
57. Teyssière, G. (1997). Modelling exchange rates volatility with multivariate long–memory ARCH processes. *Preprint.*
58. Vostrikova, L.Ju. (1981). Detection of "disorder" in multidimensional random processes. *Soviet Mathematics Doklady*, **24,** 55–59.
59. Whitcher, B., Byers, S.D., Guttorp, P. and Percival D.B. (2002). Testing for homogeneity of variance in time series: long memory, wavelets and the nile river. *Water Resources Research*, **38,** 10.1029/2001WR000509.
60. Yao, Y.C. (1988). Estimating the number of change–points via Schwarz criterion. *Statistics & Probability Letters*, **6,** 181–189.