Monitoring procedure for parameter change in causal time series

Jean-Marc Bardet, William Chakry Kengne

Monitoring procedure for parameter change in causal time series

Jean-Marc BARDET and William KENGNE *

September 21, 2012

SAMM, Université Paris 1 Panthéon-Sorbonne, 90 rue de Tolbiac 75634-Paris Cedex 13, France.
E-mail: Jean-Marc.Bardet@univ-paris1.fr ; William-Charky.Kengne@malix.univ-paris1.fr

Abstract : We propose a new sequential procedure to detect change in the parameters of a process $X = (X_t)_{t\in\mathbb{Z}}$ belonging to a large class of causal models (such as AR(∞), ARCH(∞), TARCH(∞), ARMA-GARCH processes). The procedure is based on a difference between the historical parameter estimator and the updated parameter estimator, where both these estimators are based on the quasi-likelihood of the model. Unlike classical recursive fluctuation test, the updated estimator is computed without the historical observations. The asymptotic behavior of the test is studied and the consistency in power as well as an upper bound of the detection delay are obtained. Some simulation results are reported with comparisons to some other existing procedures exhibiting the accuracy of our new procedure. The procedure is also applied to the daily closing values of the Nikkei 225, S&P 500 and FTSE 100 stock index. We show in this real-data applications how the procedure can be used to solve off-line multiple breaks detection.

Keywords: Sequential change detection; Change-point; Causal processes; Quasi-maximum likelihood estimator; Weak convergence.

1 Introduction

In statistical inference, many authors have pointed out the danger of omitting the existence of changes in the data. Many papers have been devoted to the problem of test for parameter changes in time series models when all data are available, see for instance Horváth [15], Inclan and Tiao [17], Kokoszka and Leipus [20], Kim *et al.* [19], Aue *et al.* [4], Bardet *et al.* [5], Kengne [18]. These papers consider "retrospective" (off-line) changes *i.e.* changes in parameters when all data are available. Another point of view is the change detection when new data arrive; this is sequential change-point problem. An important turning on this topic was made in 1996 with the paper of Chu *et al.*. They considered the sequential change in regression model and pointed out the effects of repeating retrospective test every time when new data are observed; this can increase the probability of type 1 error of the test. They successfully applied a fluctuation test to solve the sequential change-point problem. Two procedures are developed based on cumulative sum (CUSUM) of residuals and recursive parameter fluctuations. Their idea has been generalized and several procedures are now based on this approach. Leisch *et al.* [23] introduced the generalized fluctuation test based on the recursive moving estimator which contains the test of Chu *et al.* [12] as a special case. Horváth *et al.* [16] introduced residual CUSUM monitoring procedure where the recursive parameter is based on the historical data. This procedure has been generalized by Aue *et al.* [2] to the class of linear model with dependent errors. Berkes *et al.* [6] considered sequential changes in the parameters of GARCH process. According to the fact that the functional limit theorem assumed by Chu *et al.* [12] is not satisfied by the squares of residuals of GARCH process, they developed a procedure based on quasi-likelihood scores. Na *et al.* [25] developed a monitoring procedure for

*Supported by Université Paris 1 and AUF (Agence Universitaire de la Francophonie).

the detection of parameter changes in general time series models. They show that under the null hypothesis of no change, their detector statistic converges weakly to a known distribution. However, the asymptotic behavior of this detector is unknown under the alternative of parameter changes.

In this new contribution, we consider a large class of causal time series and investigate the asymptotic behavior under the null hypothesis of no change but also under the alternative hypothesis of change. More precisely, let $M, f : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ be measurable functions, $(\xi_t)_{t\in\mathbf{Z}}$ be a sequence of centered independent and identically distributed (iid) random variables satisfying $\mathrm{var}(\xi_0) = \sigma^2$ and let Θ be a fixed compact subset of $\mathbb{R}^d$. Let $T \subset \mathbf{Z}$, and for any $\theta \in \Theta$, define

Class $\mathcal{M}_T(M_\theta, f_\theta)$: *The process $X = (X_t)_{t\in\mathbf{Z}}$ belongs to $\mathcal{M}_T(M_\theta, f_\theta)$ if it satisfies the relation:*

$$X_{t+1} = M_\theta\big((X_{t-i})_{i\in\mathbb{N}}\big)\xi_t + f_\theta\big((X_{t-i})_{i\in\mathbb{N}}\big) \quad \text{for all } t \in T. \tag{1}$$

The existence and properties of this general class of causal and affine processes were studied in Bardet and Wintenberger [4]. Numerous classical time series (such as AR(∞), ARCH(∞), TARCH(∞), ARMA-GARCH or bilinear processes) are included in $\mathcal{M}_{\mathbf{Z}}(M, f)$. The off-line change detection for such a class of models has already been studied in Bardet *et al.* [5] and Kengne [18].

Suppose now that we have observed $X_1, \cdots, X_n$ which are available historical data. We assume that the historical data depends on one parameter *i.e.* there exists $\theta_0^* \in \Theta$ such as $(X_1, \cdots, X_n)$ belongs to $\mathcal{M}_{\{1,\cdots,n\}}(M_{\theta_0^*}, f_{\theta_0^*})$. Then, we observe new data $X_{n+1}, X_{n+2} \cdots, X_k, \cdots$: the monitoring starts. For each new observation, we would like to know if a change occurs in the parameter θ_0^*. More precisely, we consider the following test :

$\mathbf{H_0}$: θ_0^* is constant over the observation $X_1, \cdots, X_n, X_{n+1}, \cdots$ *i.e.* the observations $X_1, \cdots, X_n, X_{n+1}, \cdots$ belong to $\mathcal{M}_{\mathbb{N}}(M_{\theta_0^*}, f_{\theta_0^*})$;

$\mathbf{H_1}$: there exist $k^* > n$, $\theta_1^* \in \Theta$ such that $X_1, \cdots, X_n, X_{n+1}, \cdots, X_{k^*}, X_{k^*+1}, \cdots$ belongs to $\mathcal{M}_{\{1,\cdots,k^*\}}(M_{\theta_0^*}, f_{\theta_0^*}) \bigcap \mathcal{M}_{\{k^*+1,\cdots\}}(M_{\theta_1^*}, f_{\theta_1^*})$.

When new data arrive, Chu *et al.* [12] proposed to compute an estimator of the parameter based on all the observations and to compare it to an estimator based on historical data. A large distance between both these estimators means that new data come from a model with different parameters. Then the null hypothesis H_0 is rejected and the monitoring stops; otherwise, the monitoring continues. In their procedure, Leisch *et al.* [23] suggested to compute the recursive estimators on a moving window with a fixed width. They fixed a monitoring horizon so that, the procedure will stop after a fixed number of steps even if no change is detected. As Chu *et al.* [12], the recursive estimators computed by Na *et al.* [25] are based on all the observations. As we will see in the next sections, their procedure cannot be effective in terms of detection delay or if a small change in the parameter occurs.

For any $k \geq 1$, $\ell, \ell' \in \{1, \cdots, k\}$ (with $\ell \leq \ell'$) let $\widehat{\theta}(X_\ell, \cdots, X_{\ell'})$ be the quasi-maximum likelihood estimator (QMLE in the sequel) of the parameter computed on $\{\ell, \cdots, \ell'\}$ (see its definition in (6)). When new data arrive at time $k \geq n$, we explore the segment $\{\ell, \ell+1, \cdots, k\}$ with $\ell \in \{n - v_n, n - v_n + 1 \cdots k - v_n\}$ (where $(v_n)_{n\in\mathbb{N}}$ is a fixed sequence of integer numbers) that the distance between $\widehat{\theta}(X_\ell, \cdots, X_k)$ and $\widehat{\theta}(X_1, \cdots, X_n)$ is the largest. If the norm $\|\widehat{\theta}(X_\ell, \cdots, X_k) - \widehat{\theta}(X_1, \cdots, X_n)\|$ is greater than a suitable critical value, then H_0 is rejected and the monitoring stops; otherwise, the monitoring continues. More precisely, we construct a detector that takes into account the distance between $\widehat{\theta}(X_\ell, \cdots, X_k)$ and $\widehat{\theta}(X_1, \cdots, X_n)$. It is shown that this detector is almost surely finite under the null hypothesis and almost surely diverges to infinity under the alternative. Hence, the consistency of our procedure follows. Simulations results compared to the procedure of Horváth *et al.* [16] (see also Aue *et al.* [2]) and Na *et al.* [25] show that our procedure outperforms in terms of test power and detection delay.

In the forthcoming Section 2 the assumptions and the definition of the quasi-likelihood estimator are provided. In Section 3 we present the monitoring procedure and the asymptotic results. Section 4 is devoted to a simulation study for AR(1) and GARCH$(1,1)$ processes. In Section 5 we apply our procedure to Nikkei 225, S&P 500 and FTSE 100 stock index. After conclusions in Section 6, the proofs of the main results are provided in Section 7.

2 Assumptions and definition of the quasi-likelihood estimator

2.1 Assumptions on the class of models $\mathcal{M}_{\mathbf{Z}}(f_\theta, M_\theta)$

Let $\theta \in I\!R^d$ and M_θ and f_θ be numerical functions such that for all $(x_i)_{i\in I\!N} \in I\!R^{I\!N}$, $M_\theta\big((x_i)_{i\in I\!N}\big) \neq 0$ and $f_\theta\big((x_i)_{i\in I\!N}\big) \in I\!R$. Denote $h_\theta := M_\theta^2$. We will use the following classical notations:

1. $\|\cdot\|$ applied to a vector denotes the Euclidean norm of the vector;
2. for any compact set $\mathcal{K} \subseteq I\!R^d$ and for any $g : \mathcal{K} \longrightarrow I\!R^{d'}$, $\|g\|_{\mathcal{K}} = \sup_{\theta\in\mathcal{K}}(\|g(\theta)\|)$;
3. for any set $\mathcal{K} \subseteq I\!R^d$, $\overset{\circ}{\mathcal{K}}$ denotes the interior of $\mathcal{K}$.

Throughout the sequel, we will assume that the functions $\theta \mapsto M_\theta$ and $\theta \mapsto f_\theta$ are twice continuously differentiable on Θ. Let $\Psi_\theta = f_\theta$, M_θ and $i = 0, 1, 2$, then for any compact set $\mathcal{K} \subset \Theta$ define

Assumption $\mathbf{A}_i(\Psi_\theta, \mathcal{K})$: *Assume that* $\|\partial^i\Psi_\theta(0)/\partial\theta^i\|_\Theta < \infty$ *and there exists a sequence of non-negative real numbers* $(\alpha_j^{(i)}(\Psi_\theta,\mathcal{K}))_{j\geq 1}$ *such that* $\sum\limits_{j=1}^{\infty} \alpha_j^{(i)}(\Psi_\theta, |) < \infty$ *and*

$$\Big\|\frac{\partial^i\Psi_\theta(x)}{\partial\theta^i} - \frac{\partial^i\Psi_\theta(y)}{\partial\theta^i}\Big\|_{\mathcal{K}} \leq \sum_{j=1}^{\infty} \alpha_j^{(i)}(\Psi_\theta,\mathcal{K})|x_j - y_j| \quad \textit{for all } x, y \in I\!R^{I\!N}.$$

In the sequel we refer to the particular case called "ARCH-type process" if $f_\theta = 0$ and if the following assumption holds with $h_\theta = M_\theta^2$:

Assumption $\mathbf{A}_i(h_\theta, \mathcal{K})$: *Assume that* $\|\partial^i h_\theta(0)/\partial\theta^i\|_\Theta < \infty$ *and there exists a sequence of non-negative real numbers* $(\alpha_j^{(i)}(h_\theta,\mathcal{K}))_{j\geq 1}$ *such as* $\sum\limits_{j=1}^{\infty} \alpha_j^{(i)}(h_\theta,\mathcal{K}) < \infty$ *and*

$$\Big\|\frac{\partial^i h_\theta(x)}{\partial\theta^i} - \frac{\partial^i h_\theta(y)}{\partial\theta^i}\Big\|_{\mathcal{K}} \leq \sum_{j=1}^{\infty} \alpha_j^{(i)}(h_\theta,\mathcal{K})|x_j^2 - y_j^2| \quad \textit{for all } x, y \in I\!R^{I\!N}.$$

The Lipschitz-type hypothesis $A_i(\Psi_\theta, \mathcal{K})$ are classical when studying the existence of solutions of the general model (see for instance [13]). Using a result of [4], for each model $\mathcal{M}_{\mathbf{Z}}(M_\theta, f_\theta)$ it is interesting to define the following set:

$$\Theta(r) := \Big\{\theta \in \Theta,\ A_0(f_\theta,\{\theta\}) \text{ and } A_0(M_\theta,\{\theta\}) \text{ hold with } \sum_{j\geq 1}\alpha_j^{(0)}(f_\theta,\{\theta\}) + (E\,|\xi_0|^r)^{1/r}\sum_{j\geq 1}\alpha_j^{(0)}(M_\theta,\{\theta\}) < 1\Big\}$$
$$\bigcup\Big\{\theta\in\Theta,\ f_\theta = 0 \text{ and } A_0(h_\theta,\{\theta\}) \text{ holds with } (E\,|\xi_0|^r)^{2/r}\sum_{j\geq 1}\alpha_j^{(0)}(h_\theta,\{\theta\}) < 1\Big\}.$$

Then, if $\theta \in \Theta(r)$ the existence of a unique causal, stationary and ergodic solution $X = (X_t)_{t\in\mathbf{Z}} \in \mathcal{M}_{\mathbf{Z}}(f_\theta, M_\theta)$ is ensured (see more details in [4]). The subset $\Theta(r)$ is defined as a reunion to consider accurately general causal models and ARCH-type models simultaneously.

Here there are assumptions required for studying QMLE asymptotic properties:

Assumption D(Θ): $\exists \underline{h} > 0$ such that $\inf\limits_{\theta\in\Theta}(|h_\theta(x)|) \geq \underline{h}$ for all $x \in I\!R^{I\!N}$.

Assumption Id(Θ): For all $(\theta,\theta')\in\Theta^2$,

$$\Big(f_\theta(X_0,X_{-1},\cdots)=f_{\theta'}(X_0,X_{-1},\cdots) \text{ and } h_\theta(X_0,X_{-1},\cdots)=h_{\theta'}(X_0,X_{-1},\cdots) \text{ a.s.}\Big) \Rightarrow \ \theta=\theta'.$$

Assumption Var(Θ): For all $\theta\in\Theta$, one of the families $\big(\dfrac{\partial f_\theta}{\partial\theta^i}(X_0,X_{-1},\cdots)\big)_{1\le i\le d}$ or $\big(\dfrac{\partial h_\theta}{\partial\theta^i}(X_0,X_{-1},\cdots)\big)_{1\le i\le d}$ is a.s. linearly independent.

Assumption $\mathbf{K}(f_\theta,M_\theta,\Theta)$: for i= 0, 1, 2, $\mathbf{A}_i(f_\theta,\Theta)$ and $\mathbf{A}_i(M_\theta,\Theta)$ (or $\mathbf{A}_i(h_\theta,\Theta)$) hold and there exists $\ell>2$ such that $\alpha_j^{(i)}(f_\theta,\Theta)+\alpha_j^{(i)}(M_\theta,\Theta)+\alpha_j^{(i)}(h_\theta,\Theta)=\mathcal{O}(j^{-\ell})$ for $j\in\mathbb{N}$.

Note that in this last assumption, as in [4], we use the convention that if $\mathbf{A}_i(M_\theta,\Theta)$ holds then $\alpha_\ell^{(i)}(h_\theta,\Theta)=0$ and if $\mathbf{A}_i(h_\theta,\Theta)$ holds then $\alpha_\ell^{(i)}(M_\theta,\Theta)=0$.

2.2 Two first examples

1. ARMA(p,q) processes.
   Consider the ARMA(p,q) process defined by:
   $$X_t+\sum_{i=1}^p a_i^* X_{t-i}=\sum_{j=0}^q b_j^*\xi_{t-j}\ ,\ t\in\mathbf{Z} \tag{2}$$
   with $b_0^*\neq 0$, $\theta_0^*=(a_1^*,\cdots,a_p^*,b_0^*,\cdots,b_q^*)\in\Theta\subset\mathbb{R}^{p+q+1}$ and (ξ_t) a white noise such as $E\left(\xi_0^2\right)=1$. When $\sum_{j=0}^q b_j^*X^j\neq 0$ and $1+\sum_{i=0}^p a_i^*X^i\neq 0$ for all $|X|\le 1$, this process can be also written as:
   $$X_t=b_0^*\xi_t+\sum_{j=1}^\infty \phi_j(\theta_0^*)\,X_{t-i}\ ,\ t\in\mathbf{Z}$$
   where $\theta\in\Theta\mapsto\phi_j(\theta)$ are functions only depending on θ and decreasing exponentially fast to 0 ($j\to\infty$). The process (2) belongs to the class $\mathcal{M}_{\mathbf{Z}}(M_{\theta_0^*},f_{\theta_0^*})$ where $f_\theta(x_1,\cdots)=\sum_{j\ge1}\phi_j(\theta)x_j$ and $M_\theta\equiv b_0^*$ for all $\theta\in\Theta$. Then Assumptions $\mathbf{D}(\Theta)$, $\mathbf{A}_0(f_\theta,\Theta)$, $\mathbf{A}_0(M_\theta,\Theta)$ hold with $\underline{h}=|b_0^*|>0$ and $\alpha_j^{(0)}(f_\theta,\Theta)=\|\phi_j(\theta)\|_\Theta$ while $\alpha_j^{(0)}(M_\theta,\Theta)=0$ for $j\in\mathbb{N}^*$. Assumption $\mathbf{K}(f_\theta,M_\theta,\Theta)$ holds since there exists $c>0$ and $C>0$ such as $|\phi_j|\le C\,e^{-cj}$ for $j\in\mathbb{N}$. Moreover, if (ξ_t) is a sequence of non-degenerate random variables (*i.e.* ξ_t are not equal to a constant), Assumptions $\mathbf{Id}(\Theta)$ and $\mathbf{Var}(\Theta)$ hold. Finally, for any $r\ge1$ such that $E\,|\xi_0|^r<\infty$, then
   $$\Theta(r)=\{\theta\in\mathbb{R}^{p+q+1},\ \sum_{j\ge1}|\phi_j(\theta)|<1\}.$$
   Note that if $\theta\in\Theta(r)$ with $r\ge1$ then the previous conditions of stationarity $\sum_{j=0}^q b_jX^j\neq0$ and $1+\sum_{i=0}^p a_iX^i\neq0$ for all $|X|\le1$ are satisfied.

2. GARCH(p,q) processes.
   Consider the GARCH(p,q) process defined by:
   $$X_t=\sigma_t\,\xi_t\ ,\ \sigma_t^2=\alpha_0^*+\sum_{j=1}^p\alpha_j^*X_{t-j}^2+\sum_{j=1}^q\beta_j^*\sigma_{t-j}^2\ ,\quad t\in\mathbf{Z} \tag{3}$$
   with $E\left(\xi_0^2\right)=1$ and $\theta_0^*:=(\alpha_0^*,\cdots,\alpha_p^*,\beta_1^*,\cdots,\beta_q^*)\in\Theta$ where Θ is a compact subset of $]0,\infty[\times[0,\infty[^{p+q}$ such that $\sum_{j=1}^p\alpha_j+\sum_{j=1}^q\beta_j<1$ for all $\theta\in\Theta$. Then there exists (see Bollerslev [10] or Nelson and Cao [26]) a nonnegative sequence $(\psi_j(\theta_0^*))_{j\ge0}$ such that $\sigma_t^2=\psi_0(\theta_0^*)+\sum_{j\ge1}\psi_j(\theta_0^*)X_{t-j}^2$ with $\psi_0(\theta_0^*)=\alpha_0^*/(1-\sum_{j=1}^q\beta_j^*)$. This process belongs to the class $\mathcal{M}_{\mathbf{Z}}(M_{\theta_0^*},f_{\theta_0^*})$ where $f_\theta\equiv0$ and $M_\theta(x_1,\cdots)=\sqrt{\psi_0(\theta)+\sum_{j\ge1}\psi_j(\theta)x_j^2}$ for all $\theta\in\Theta$. Assumption $\mathbf{D}(\Theta)$ holds with $\underline{h}=\inf\limits_{\theta\in\Theta}(\psi_0(\theta))>0$. If there exists $0<\rho_0<1$ such that for any

$\theta \in \Theta$, $\sum_{j=1}^q \alpha_j + \sum_{j=1}^p \beta_j \leq \rho_0$ then the sequences $(\|\psi_j(\theta)\|_\Theta)_{j\geq 1}$, $(\|\psi_j'(\theta)\|_\Theta)_{j\geq 1}$ and $(\|\psi_j''(\theta)\|_\Theta)_{j\geq 1}$ decay exponentially fast (see Berkes *et al.* [7]) and Assumption $\mathbf{K}(f_\theta, M_\theta, \Theta)$ holds. Moreover, (ξ_t^2) is a sequence of non-degenerate random variables (*i.e.* ξ_t^2 are not equal to a constant), Assumptions $\mathbf{Id}(\Theta)$ and $\mathbf{Var}(\Theta)$ hold. Finally for $r \geq 2$ we obtain

$$\Theta(r) = \big\{\theta \in \Theta \ ; \ (E\,|\xi_0|^r)^{2/r} \sum_{j=1}^\infty \phi_j(\theta) < 1\big\}.$$

2.3 The quasi-maximum likelihood estimator

Let $k \geq n \geq 2$, if $(X_1, \cdots, X_k) \in \mathcal{M}_{\{1,\cdots,k\}}(M_\theta, f_\theta)$, then for $T \subset \{1, \cdots, k\}$, the conditional quasi-(log)likelihood computed on T is given by:

$$L(T,\theta) := -\frac{1}{2}\sum_{t\in T} q_t(\theta) \quad \text{with} \quad q_t(\theta) = \frac{(X_t - f_\theta^t)^2}{h_\theta^t} + \log(h_\theta^t) \tag{4}$$

where $f_\theta^t = f_\theta\big(X_{t-1}, X_{t-2}\ldots\big)$, $M_\theta^t = M_\theta\big(X_{t-1}, X_{t-2}\ldots\big)$ and $h_\theta^t = M_\theta^{t\,2}$. The classical approximation of this conditional log-likelihood is given by:

$$\widehat{L}(T,\theta) := -\frac{1}{2}\sum_{t\in T} \widehat{q}_t(\theta) \quad \text{where} \quad \widehat{q}_t(\theta) := \frac{\big(X_t - \widehat{f}_\theta^t\big)^2}{\widehat{h}_\theta^t} + \log\big(\widehat{h}_\theta^t\big) \tag{5}$$

with $\widehat{f}_\theta^t = f_\theta\big(X_{t-1}, \ldots, X_1, 0, 0, \cdots\big)$, $\widehat{M}_\theta^t = M_\theta\big(X_{t-1}, \ldots, X_1, 0, 0, \cdots\big)$ and $\widehat{h}_\theta^t = (\widehat{M}_\theta^t)^2$.
For $T \subset \{1, \cdots, k\}$, define the quasi maximum-likelihood estimator (QMLE) of θ computed on T by

$$\widehat{\theta}(T) := \underset{\theta\in\Theta}{\text{argmax}}(\widehat{L}(T,\theta)). \tag{6}$$

In Bardet and Wintenberger [4] it was established that if $(X_1, \cdots, X_n)$ is an observed trajectory of $X \subset \mathcal{M}_{\mathbf{Z}}(f_{\theta_0^*}, M_{\theta_0^*})$ with $\theta_0^* \in \overset{\circ}{\Theta}(4)$ and if Θ is a compact set such as Assumptions $\mathbf{A}_i(f_\theta, M_\theta, \Theta)$ (or $\mathbf{A}_i(h_\theta, \Theta)$) hold for $i = 0, 1, 2$ and under Assumptions $\mathbf{D}(\Theta)$, $\mathbf{Id}(\Theta)$, $\mathbf{Var}(\Theta)$, $\mathbf{K}(f_\theta, M_\theta, \Theta)$, then

$$\sqrt{n}\big(\widehat{\theta}(T_{1,n}) - \theta_0^*\big) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}\big(0\,,\, F\,G^{-1}\,F\big), \tag{7}$$

with

$$G := E\Big[\frac{\partial q_0(\theta_0^*)}{\partial\theta}\frac{\partial q_0(\theta_0^*)}{\partial\theta}'\Big] \quad \text{and} \quad F := E\Big[\frac{\partial^2 q_0(\theta_0^*)}{\partial\theta\partial\theta'}\Big], \tag{8}$$

where $'$ denotes the transpose and with q_0 defined in (4). Note that under assumptions $\mathbf{D}(\Theta)$ and $\mathbf{Var}(\Theta)$, G is symmetric positive definite (see [18]) and F is non-singular (see [4]). Also define the matrix

$$\widehat{G}(T) := \frac{1}{\text{Card}(T)}\sum_{t\in T}\Big(\frac{\partial\widehat{q}_t(\widehat{\theta}(T))}{\partial\theta}\Big)\Big(\frac{\partial\widehat{q}_t(\widehat{\theta}(T))}{\partial\theta}\Big)' \quad \text{and} \quad \widehat{F}(T) := -\frac{2}{\text{Card}(T)}\Big(\frac{\partial^2\widehat{L}_m(T,\widehat{\theta}(T))}{\partial\theta\partial\theta'}\Big). \tag{9}$$

Under the previous assumptions, $\widehat{G}(T_{1,n})$ and $\widehat{F}(T_{1,n})$ converge almost surely to G and F respectively. Hence,

$$\sqrt{n}\,\widehat{G}(T_{1,n})^{-1/2}\widehat{F}(T_{1,n})\big(\widehat{\theta}(T_{1,n}) - \theta_0^*\big) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}\big(0\,,\, I_d\big) \tag{10}$$

with I_d the identity matrix. This result will be the starting point of the following monitoring procedure.

3 The monitoring procedure and asymptotic results

3.1 The monitoring procedure

In the sequel, $(X_1, \cdots, X_n)$ is supposed to be the historical available observations belonging to the class $\mathcal{M}_{\{1,\cdots,n\}}(f_{\theta_0^*}, M_{\theta_0^*})$.
For $1 \leq \ell \leq \ell'$, denote

$$T_{\ell,\ell'} := \{\ell, \ell+1, \cdots, \ell'\}.$$

At a monitoring instant k, our idea is to evaluate the difference between $\widehat{\theta}(T_{\ell,k})$ and $\widehat{\theta}(T_{1,n})$ for any $\ell = n, \cdots, k$. More precisely, from (10), for any $k > n$ define the statistic (called the detector)

$$\widehat{C}_{k,\ell} := \sqrt{n}\,\frac{k-\ell}{k}\,\big\|\widehat{G}(T_{1,n})^{-1/2}\widehat{F}(T_{1,n})\big(\widehat{\theta}(T_{\ell,k}) - \widehat{\theta}(T_{1,n})\big)\big\|$$

for $\ell = n, \cdots, k$. Since the matrix $\widehat{G}(T_{1,n})$ is asymptotically symmetric and positive definite (see [18]), $\widehat{G}(T_{1,n})^{-1/2}$ exists for n large enough and $\widehat{C}_{k,\ell}$ is well defined. At the beginning of the monitoring and when ℓ is close to k, the length of $T_{\ell,k}$ is too small, therefore the numerical algorithm used to compute $\widehat{\theta}(T_{\ell,k})$ cannot converge. This can introduce a large distortion in the procedure. To avoid this, we introduce a sequence of integer numbers $(v_n)_{n\in\mathbb{N}}$ with $v_n << n$ and compute $\widehat{C}_{k,\ell}$ for $\ell \in \{n - v_n, n - v_n + 1, \cdots, k - v_n\}$. Thus, for any $k > n$ denote

$$\Pi_{n,k} := \{n - v_n, n - v_n + 1, \cdots, k - v_n\}.$$

For technical reasons, assume that,

$$v_n \to \infty \quad \text{and} \quad v_n/\sqrt{n} \to 0 \;\; (n \to \infty).$$

According to Remark 1 of [18], we can choose $v_n = [(\log n)^{\delta}]$ with $\delta > 1$.
Note that, if change does not occur at time $k > n$, for any $\ell \in \Pi_{n,k}$, the two estimators $\widehat{\theta}(T_{\ell,k})$ and $\widehat{\theta}(T_{1,n})$ are close and the detector $\widehat{C}_{k,\ell}$ is not large enough. Hence, the monitoring stops and reject H_0 at the first time $k > n$ where there exists $\ell \in \Pi_{n,k}$ satisfying $\widehat{C}_{k,\ell} > c$ for a fixed constant $c > 0$. To be more general, we will use a $b : (0,\infty) \mapsto (0,\infty)$, called a boundary function satisfying:

Assumption B: $b : (0,\infty) \mapsto (0,\infty)$ is a non-increasing and continuous function such as $\underset{0<t<\infty}{\mathrm{Inf}}\, b(t) > 0$.

Then the monitoring procedure stops at the first time $k > n$ such as there exists $\ell \in \Pi_{n,k}$ satisfying $\widehat{C}_{k,\ell} > b((k-\ell)/n)$. Hence define the stopping time:

$$\tau(n) := \mathrm{Inf}\{k > n \;/\; \exists \ell \in \Pi_{n,k},\; \widehat{C}_{k,\ell} > b((k-\ell)/n)\} = \mathrm{Inf}\Big\{k > n \;/\; \max_{\ell\in\Pi_{n,k}} \frac{\widehat{C}_{k,\ell}}{b((k-\ell)/n)} > 1\Big\}.$$

Therefore, we have

$$P\{\tau(n) < \infty\} = P\Big\{\max_{\ell\in\Pi_{n,k}} \frac{\widehat{C}_{k,\ell}}{b((k-\ell)/n)} > 1 \text{ for some } k > n\Big\} = P\Big\{\sup_{k>n}\max_{\ell\in\Pi_{n,k}} \frac{\widehat{C}_{k,\ell}}{b((k-\ell)/n)} > 1\Big\}. \tag{11}$$

The challenge is to choose a suitable boundary function $b(\cdot)$ such as for some given $\alpha \in (0,1)$

$$\lim_{n\to\infty} P_{H_0}\{\tau(n) < \infty\} = \alpha$$

and

$$\lim_{n\to\infty} P_{H_1}\{\tau(n) < \infty\} = 1$$

where the hypothesis H_0 and H_1 are specified in Section 1.
In the case where $b(\cdot)$ is a constant positive value, $b \equiv c$ with $c > 0$, these conditions lead to compute a threshold $c = c_\alpha$ depending on α. If change is detected under H_1 i.e. $\tau(n) < \infty$ and $\tau(n) > k^*$, then the detection delay is defined by

$$\widehat{d}_n = \tau(n) - k^*.$$

Using the previous notations, Na *et al.* [25] used the following detector

$$\widehat{D}_k := \sqrt{n}\big\|\widehat{G}(T_{1,n})^{-1/2}\widehat{F}(T_{1,n})\big(\widehat{\theta}(T_{1,k}) - \widehat{\theta}(T_{1,n})\big)\big\|.$$

At the step k of the monitoring, their recursive estimator is based on $X_1, \cdots, X_n, \cdots, X_k$. One can see that this estimator is highly influenced by the historical data. Assume that a change occurs at time $k^* \le k$, in the sequel of the procedure, the recursive estimator contents the observations $X_1, \cdots, X_n, \cdots, X_{k^*-1}$ which depends on θ_0^*. Then, one must wait longer before the difference between $\widehat{\theta}(X_1, \cdots, X_n)$ and $\widehat{\theta}(X_1, \cdots, X_n, \cdots, X_k)$ becomes significant at a step $k > k^*$. Therefore, their procedure cannot be effective in terms of detection delay. Moreover, if n tends to infinity, it is not almost sure that this change will be detected. These are confirmed by the results of simulations (see Section 4).

Berkes *et al.* (2004) considered an estimator based on historical data to compute the quasi-likelihood scores. They used the fact that the partial derivatives applied to a vector $\mathbf{u}$ is equal to 0 if and only if $\mathbf{u}$ is the true parameter of the model. So, when change occurs, their detector growths asymptotically to infinity. Therefore, their procedure is consistent. They proved this result for GARCH(p,q) models.

3.2 Asymptotic behaviour under the null hypothesis

Under H_0, the parameter θ_0^* does not change over the new observations. Thus we have the result

Theorem 3.1 *Assume* $\boldsymbol{D}(\Theta)$, $\boldsymbol{Id}(\Theta)$, $\boldsymbol{Var}(\Theta)$, $\boldsymbol{K}(f_\theta, M_\theta, \Theta)$, $\mathbf{B}$ *and* $\theta_0^* \in \overset{\circ}{\Theta}(4)$. *Under the null hypothesis* H_0, *then*

$$\lim_{n\to\infty} P\{\tau(n) < \infty\} = P\Big\{\sup_{t>1}\sup_{1<s<t} \frac{\|W_d(s) - sW_d(1))\|}{t\ b(s)} > 1\Big\}$$

where W_d *is a* d*-dimensional standard Brownian motion.*

In the simulations, we will use the most "natural" boundary function $b(\cdot) = c$ with c a positive constant since it satisfies the above assumptions imposed to $b(\cdot)$. In such case, the forthcoming corollary indicates that the asymptotic distribution of Theorem 3.1 can be easily computed:

Corollary 3.1 *Assume* $b(t) = c > 0$ *for* $t \geq 0$. *Under the assumptions of Theorem 3.1,*

$$\lim_{n\to\infty} P\{\tau(n) < \infty\} = P\Big\{\sup_{t>1}\sup_{1<s<t} \frac{1}{t}\|W_d(s) - sW_d(1))\| > c\Big\} = P\{U_d > c\}$$

where $U_d = \sup_{0<u<1}\ f(u)\,\|W_d(u)\|$ *with* $f(u) = \dfrac{\sqrt{9-u}+\sqrt{1-u}}{\sqrt{9-u}+3\sqrt{1-u}}\Big(\dfrac{2}{3-u+\sqrt{(9-u)(1-u)}}\Big)^{1/2}$.

Remark 3.1 *By the law of the iterated logarithm, it comes that*

$$\sup_{1<s<t} \frac{1}{t}\|W_d(s) - sW_d(1))\| \xrightarrow[t\to\infty]{a.s.} \|W_d(1))\|.$$

So, the two distributions $\sup_{1<s<t}\frac{1}{t}\|W_d(s) - sW_d(1))\|$ *as* $t \to \infty$ *(resp.* $t \to 1$*) and* $f(u)\,\|W_d(u)\|$ *as* $u \to 1$ *(resp.* $u \to 0$*) are equal. It is easy to show (see proof of Corollary 3.1) that*

$$\sup_{t>1}\sup_{1<s<t} \frac{1}{t}\|W_d(s) - sW_d(1))\| \overset{\mathcal{D}}{=} \sup_{0<u<1} f(u)\,\|W_d(u)\|.$$

Therefore, at a nominal level $\alpha \in (0,1)$, take $c = c(\alpha)$ be the $(1-\alpha)$-quantile of the distribution of $U_d = \sup_{0<u<1}\ f(u)\,\|W_d(u)\|$ which can be computed through Monte-Carlo simulations. Table 1 shows the $(1-\alpha)$-quantile of this distribution for $\alpha = 0.01, 0.05, 0.10$ and $d = 1, \cdots, 5$.

	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$
$\alpha = 0.01$	2.583	3.035	3.335	3.631	3.914
$\alpha = 0.05$	1.954	2.432	2.760	3.073	3.334
$\alpha = 0.10$	1.652	2.156	2.486	2.784	3.028

Table 1: Empirical $(1-\alpha)$-quantile of the distribution of U_d, for $d = 1, \cdots, 5$.

3.3 Asymptotic behaviour under the alternative hypothesis

Under the alternative H_1, the parameter changes from θ_0^* to θ_1^* at $k^* > n$, where $\theta_1^* \in \Theta$ and $\theta_0^* \neq \theta_1^*$. Then

Theorem 3.2 *Assume* $\boldsymbol{D}(\Theta)$*,* $\boldsymbol{Id}(\Theta)$*,* $\boldsymbol{Var}(\Theta)$*,* $\boldsymbol{K}(f_\theta, M_\theta, \Theta)$ *and* $\mathbf{B}$*. Under the alternative* H_1*, if* $\theta_1^* \neq \theta_0^*$ *and* $\theta_0^*, \theta_1^* \in \overset{\circ}{\Theta}(4)$ *then for* $k^* = k^*(n)$ *such as* $\limsup_{n\to\infty} k^*(n)/n < \infty$ *and* $k_n = k^*(n) + n^\delta$ *with* $\delta \in (1/2, 1)$*,*

$$\max_{\ell \in \Pi_{n,k_n}} \frac{\widehat{C}_{k_n,\ell}}{b((k_n-\ell)/n)} \xrightarrow[n\to\infty]{a.s.} \infty.$$

The forthcoming Corollary 3.2 can be immediately deduced from the relation (11).

Corollary 3.2 *Under assumptions of Theorem 3.2,*

$$\lim_{n\to\infty} P\{\tau(n) < \infty\} = 1.$$

Remark 3.2 *We know that the monitoring is stopped and rejects* H_0 *at the first time* k *where*

$$\max_{\ell \in \Pi_{n,k}} \frac{\widehat{C}_{k,\ell}}{b((k-\ell)/n)} > 1.$$

Therefore, it follows from Theorem 3.2 that under the hypothesis H_1*, the detection delay* $\widehat{d}_n$ *of the procedure can be bounded by* $\mathcal{O}_P(n^{1/2+\varepsilon})$ *for any* $\varepsilon > 0$ *(or even by* $\mathcal{O}_P\big(\sqrt{n}(\log n)^a\big)$ *with* $a > 0$ *using the same kind of proof).*

3.4 Examples

3.4.1 AR(∞) processes

Consider the generalization of ARMA(p,q) processes defined in (2) *i.e.* a AR(∞) processes defined by:

$$X_t = \phi_0(\theta_0^*) + \sum_{j\geq 1} \phi_j(\theta_0^*) X_{t-j} + \xi_t \ , \ t \in \mathbf{Z} \tag{12}$$

with $\theta_0^* \in \overset{\circ}{\Theta}$, where we can chose Θ as a compact subset of $\Theta(4) \subset I\!\!R^d$ where

$$\Theta(4) = \big\{\theta \in I\!\!R^d; \ \sum_{j\geq 1} |\phi_j(\theta)| < 1\big\}.$$

This process belongs to the class $\mathcal{M}_{\mathbf{Z}}(M_{\theta_0^*}, f_{\theta_0^*})$ where $f_\theta(x_1, \cdots) = \sum_{j\geq 1} \phi_j(\theta) x_j$ and $M_\theta \equiv \phi_0(\theta)$ for all $\theta \in \Theta$ and therefore $\alpha_j^{(0)}(f_\theta, \Theta) = \|\phi_j(\theta)\|_\Theta$ and $\alpha_j^{(0)}(M_\theta, \Theta) = 0$ for $j \in I\!\!N^*$. Then

- Assumption $\mathbf{D}(\Theta)$ holds if $\underline{h} = \inf_{\theta\in\Theta}(|\phi_0(\theta)|) > 0$;
- Assumption $\mathbf{K}(f_\theta, M_\theta, \Theta)$ holds if there exists $\ell > 2$ and and if $\theta \mapsto \phi_j(\theta)$ are twice differentiable functions satisfying $\max\big(\|\psi_j(\theta)\|_\Theta, \|\phi_j'(\theta)\|_\Theta, \|\phi_j''(\theta)\|_\Theta\big) = O(j^{-\ell})$ for $j \in I\!\!N$.
- if (ξ_t) is a sequence of non-degenerate random variables (*i.e.* ξ_t are not equal to a constant), Assumptions $\mathbf{Id}(\Theta)$ and $\mathbf{Var}(\Theta)$ hold.

Case of AR(p) process
Assume that

$$X_t = \phi_0^* + \sum_{j=1}^{p} \phi_j^* X_{t-j} + \xi_t \quad \text{with } p \in I\!\!N^*.$$

The true parameter of the model is denoted by $\theta_0^* = (\phi_0^*, \phi_1^*, \cdots, \phi_p^*) \in \Theta$ where $\Theta = \{\theta = (\phi_0, \phi_1, \cdots, \phi_p) \in I\!\!R^{p+1} \ / \sum_{j=1}^{p} |\phi_j| < 1\}$. Then, $\Theta(r) = \Theta$ for any $r \geq 1$. Assume that a trajectory $(X_1, \cdots, X_k)$ has been observed, for any $t = 1, \cdots, k$ and $\theta \in \Theta$ we have, $\widehat{q}_t(\theta) = \big(X_t - \phi_0 - \sum_{j=1}^{p} \phi_j X_{t-j}\big)^2$, $\dfrac{\partial \widehat{q}_t(\theta)}{\partial \theta} = -2\big(X_t - \phi_0 - \sum_{j=1}^{p} \phi_j X_{t-j}\big) \cdot (1, X_{t-1}, X_{t-2}, \cdots, X_{t-p})$. Moreover, $\dfrac{\partial^2 \widehat{q}_t(\theta)}{\partial \phi_0 \partial \phi_0} = 2$, for $j = 1, \cdots, p$, $\dfrac{\partial^2 \widehat{q}_t(\theta)}{\partial \phi_0 \partial \phi_j} = 2X_{t-j}$ and for $1 \leq i, j \leq p$, $\dfrac{\partial^2 \widehat{q}_t(\theta)}{\partial \phi_i \partial \phi_j} = 2X_{t-i}X_{t-j}$.

3.4.2 ARCH(∞) processes

Consider the generalization GARCH(p, q) processes defined in (3) *i.e.* a ARCH(∞) processes defined by:

$$X_t = \sigma_t \, \xi_t \quad \text{and} \quad \sigma_t^2 = \psi_0(\theta_0^*) + \sum_{j=1}^{\infty} \psi_j(\theta_0^*) X_{t-j}^2 \ , \ t \in \mathbf{Z} \tag{13}$$

with $\theta_0^* \in \overset{\circ}{\Theta}$, where we can chose Θ as a compact subset of $\Theta(4) \subset I\!R^d$ where

$$\Theta(4) = \big\{\theta \in I\!R^d; \ (E\,|\xi_0|^4)^{1/2} \sum_{j=1}^{\infty} |\phi_j(\theta)| < 1\big\}.$$

This process, introduced by Robinson [28], belongs to the class $\mathcal{M}_{\mathbf{Z}}(f_{\theta_0^*}, M_{\theta_0^*},)$ where $f_\theta(x_1, \cdots) \equiv 0$ and $M_\theta^2(x_1, \cdots) = \psi_0(\theta) + \sum_{j \geq 1} \psi_j(\theta) x_j^2$ for all $\theta \in \Theta$ and therefore $\alpha_j^{(0)}(f_\theta, \Theta) = 0$ and $\alpha_j^{(0)}(h_\theta, \Theta) = \|\phi_j(\theta)\|_\Theta$ for $j \in I\!N^*$ (X is of course a ARCH-type process). Then

- Assumption $\mathbf{D}(\Theta)$ holds if $\underline{h} = \inf_{\theta \in \Theta}(\psi_0(\theta)) > 0$;
- Assumption $\mathbf{K}(f_\theta, M_\theta, \Theta)$ holds if there exists $\ell > 2$ and and if $\theta \mapsto \phi_j(\theta)$ are twice differentiable functions satisfying $\max\big(\|\psi_j(\theta)\|_\Theta, \|\psi_j'(\theta)\|_\Theta, \|\psi_j''(\theta)\|_\Theta\big) = O(j^{-\ell})$ for $j \in I\!N$.
- if (ξ_t^2) is a sequence of non-degenerate random variables (*i.e.* ξ_t^2 are not equal to a constant), Assumptions $\mathbf{Id}(\Theta)$ and $\mathbf{Var}(\Theta)$ hold.

Case of GARCH$(1,1)$ process

Assume that

$$X_t = \sigma_t \xi_t \quad \text{with} \ \ \sigma_t^2 = \alpha_0^* + \alpha_1^* X_{t-1}^2 + \beta_1^* \sigma_{t-1}^2$$

with $\theta_0^* = (\alpha_0^*, \alpha_1^*, \beta_1^*) \in \Theta \subset]0, \infty[\times[0, \infty[^2$ and satisfying $\alpha_1^* + \beta_1^* < 1$. The ARCH($\infty$) representation is $\sigma_t^2 = \alpha_0^*/(1 - \beta_1^*) + \alpha_1^* \sum_{j \geq 1} (\beta_1^*)^{j-1} X_{t-j}^2$. If a trajectory $(X_1, \cdots, X_k)$ has been observed, for any $t = 1, \cdots, k$ and $\theta \in \Theta$ we have,

$$\widehat{h}_\theta^t = \alpha_0/(1 - \beta_1) + \alpha_1 X_{t-1}^2 + \alpha_1 \sum_{j=2}^{t} \beta_1^{j-1} X_{t-j}^2 \ \text{ and } \ \widehat{q}_t(\theta) = X_t^2 / \ \widehat{h}_\theta^t + \log(\widehat{h}_\theta^t).$$

Thus, it follows that $\dfrac{\partial \widehat{q}_t(\theta)}{\partial \theta} = \dfrac{1}{\widehat{h}_\theta^t}\Big(1 - \dfrac{X_t^2}{\widehat{h}_\theta^t}\Big)\Big(\dfrac{\partial \widehat{h}_\theta^t}{\partial \alpha_0}, \dfrac{\partial \widehat{h}_\theta^t}{\partial \alpha_1}, \dfrac{\partial \widehat{h}_\theta^t}{\partial \beta_1}\Big)$ with $\dfrac{\partial \widehat{h}_\theta^t}{\partial \alpha_1} = X_{t-1}^2 + \sum_{j=2}^{t} \beta_1^{j-1} X_{t-j}^2$ $\dfrac{\partial \widehat{h}_\theta^t}{\partial \alpha_0} = 1/(1 - \beta_1)$, and $\dfrac{\partial \widehat{h}_\theta^t}{\partial \beta_1} = \alpha_0/(1 - \beta_1)^2 + \alpha_1 X_{t-2}^2 + \alpha_1 \sum_{j=3}^{t} (j-1)\beta_1^{j-2} X_{t-k}^2$.

Let $\theta = (\alpha_0, \alpha_1, \beta_1) = (\theta_1, \theta_2, \theta_3) \in \Theta$, for $1 \leq i, j \leq 3$, we have

$$\frac{\partial^2 \widehat{q}_t(\theta)}{\partial \theta_i \partial \theta_j} = \frac{1}{(\widehat{h}_\theta^t)^2}\Big(\frac{2X_t^2}{\widehat{h}_\theta^t} - 1\Big)\frac{\partial \widehat{h}_\theta^t}{\partial \theta_i}\frac{\partial \widehat{h}_\theta^t}{\partial \theta_j} + \frac{1}{\widehat{h}_\theta^t}\Big(1 - \frac{X_t^2}{\widehat{h}_\theta^t}\Big)\frac{\partial^2 \widehat{h}_\theta^t}{\partial \theta_i \partial \theta_j}$$

with $\dfrac{\partial^2 \widehat{h}_\theta^t}{\partial \alpha_0^2} = 0$, $\dfrac{\partial^2 \widehat{h}_\theta^t}{\partial \alpha_0 \partial \alpha_1} = 0$, $\dfrac{\partial^2 \widehat{h}_\theta^t}{\partial \alpha_1^2} = 0$, $\dfrac{\partial^2 \widehat{h}_\theta^t}{\partial \alpha_1 \partial \beta_1} = X_{t-2}^2 + \sum_{j=3}^{t} (j-1)\beta_1^{j-2} X_{t-j}^2$, $\dfrac{\partial^2 \widehat{h}_\theta^t}{\partial \alpha_0 \partial \beta_1} = 1/(1 - \beta_1)^2$ and $\dfrac{\partial \widehat{h}_\theta^t}{\partial \beta_1^2} = 2\alpha_0/(1 - \beta_1)^3 + 2\alpha_1 X_{t-3}^2 + \alpha_1 \sum_{j=4}^{t} (j-1)(j-2)\beta_1^{j-3} X_{t-j}^2$.

3.4.3 TARCH(∞) processes

The process X is called Threshold ARCH(∞) (TARCH(∞) in the sequel) if it satisfies

$$X_t = \sigma_t \xi_t \quad \text{and} \quad \sigma_t = b_0(\theta_0^*) + \sum_{j=1}^{\infty} \Big[b_j^+(\theta_0^*) \max(X_{t-j}, 0) - b_j^-(\theta_0^*) \min(X_{t-j}, 0) \Big] \, , \; t \in \mathbf{Z} \tag{14}$$

where the parameters $b_0(\theta)$, $b_j^+(\theta)$ and $b_j^-(\theta)$ are assumed to be non negative real numbers and $\theta \in \overset{\circ}{\Theta}$ where Θ is a compact subset of $\Theta(4)$ where

$$\Theta(4) = \Big\{ \theta \in I\!\!R^d \; \Big/ \; \left(E\, |\xi_0|^4 \right)^{1/4} \sum_{j=1}^{\infty} \max\left(b_j^-(\theta), b_j^+(\theta) \right) < 1 \Big\}$$

since $\alpha_j^{(0)}(M, \{\theta\}) = \max\left(b_j^-(\theta), b_j^+(\theta) \right)$. This class of processes is a generalization of the class of TGARCH(p,q) processes (introduced by Rabemananjara and Zakoïan [27]). Then,

- Assumption $\mathbf{D}(\Theta)$ holds if $\underline{h} = \inf_{\theta \in \Theta} b_0(\theta) > 0$;
- Assumption $\mathbf{K}(f_\theta, M_\theta, \Theta)$ holds if there exists $\ell > 2$ and and if $\theta \mapsto b_j^-(\theta)$ and $\theta \mapsto b_j^+(\theta)$ are twice differentiable functions satisfying

$$\max\left(\|b_j^-(\theta)\|_\Theta, \|b_j^+(\theta)\|_\Theta, \|\frac{\partial}{\partial\theta} b_j^-(\theta)\|_\Theta, \|\frac{\partial}{\partial\theta} b_j^+(\theta)\|_\Theta, \|\frac{\partial^2}{\partial\theta^2} b_j^-(\theta)\|_\Theta, \|\frac{\partial^2}{\partial\theta^2} b_j^+(\theta)\|_\Theta \right) = O(j^{-\ell}) \quad \text{for} \quad j \in I\!\!N.$$

Unfortunately, for TARCH(∞) it is not possible to provide simple conditions for obtaining Assumptions $\mathbf{Id}(\Theta)$ and $\mathbf{Var}(\Theta)$ as for AR(∞) or ARCH(∞) processes.

4 Some simulation and numerical experiments

First remark that, at a time $k > n$, we need to compute $\widehat{C}_{k,\ell}$ for all $\ell \in \Pi_{n,k}$ to test whether change occurs or not. On can see that, the computational time is very long and increases with k. To reduce it, we introduce an integer sequence (u_n) (satisfying $u_n/\sqrt{n} \to 0$ as $n \to \infty$; typically $u_n = [\ln(n)]$) and compute $\widehat{C}_{k,\ell}$ only for

$$\ell \in \Pi^0_{n,k} := \{ n - v_n, n - v_n + u_n, n - v_n + 2u_n, \cdots, k - v_n \}.$$

We have $\Pi^0_{n,k} \subset \Pi_{n,k}$ and for any $t = \frac{\ell}{n}$ with $\ell \in \Pi_{n,k}$, we can find an integer j_ℓ such that $n - v_n + j_\ell u_n \in \Pi^0_{n,k}$ and $n - v_n + j_\ell u_n \leq \ell \leq n - v_n + (j_\ell + 1) u_n$. This implies that $\frac{n - v_n + j_\ell u_n}{n} \leq t \leq \frac{n - v_n + j_\ell u_n}{n} + \frac{u_n}{n}$. Thus, we have asymptotically (as $n \to \infty$), $t \sim \frac{n - v_n + j_\ell u_n}{n}$. It shows that the previous asymptotic results still hold by computing $\widehat{C}_{k,\ell}$ for $\ell \in \Pi^0_{n,k}$. The condition $u_n/\sqrt{n} \to 0$ ensures that the Theorem 3.2 still holds by choosing $k_n = k^*(n) + n^\delta$ with $\delta \in (1/2, 1)$. In practice, the use of $\Pi^0_{n,k}$ can introduce a distortion in the detection delay. But, the new detection delay must be between $\widehat{d}_n$ and $\widehat{d}_n + u_n$ (where $\widehat{d}_n$ is the detection delay obtained by using $\Pi_{n,k}$). In the sequel, we use $u_n = [\ln(n)]$.
Moreover, if $b \equiv c > 0$ is a constant function, according to (11), we have

$$P\{\tau(n) < \infty\} = P\Big\{ \sup_{k > n} \max_{\ell \in \Pi^0_{n,k}} \widehat{C}_{k,\ell} > c \Big\}. \tag{15}$$

Thus, denote

$$\widehat{C}_k = \max_{\ell \in \Pi^0_{n,k}} \widehat{C}_{k,\ell} \quad \text{for any} \quad k > n.$$

The procedure is monitored from $k = n+1$ to $k = n+500$. The set $\{n+1, \cdots, n+500\}$ is called monitoring period. According to the Remark 1 of [18], $v_n = [(\log n)^\delta]$ (with $1 \leq \delta \leq 3$) is chosen. We evaluated the performance of the procedure with $v_n = [\log n]$, $[(\log n)^{3/2}]$, $[(\log n)^2]$, $[(\log n)^3]$ and we recommend to use $v_n = [(\log n)^{3/2}]$ for linear model and $v_n = [(\log n)^2]$ for ARCH-type model. The nominal level used in the sequel is $\alpha = 0.05$.

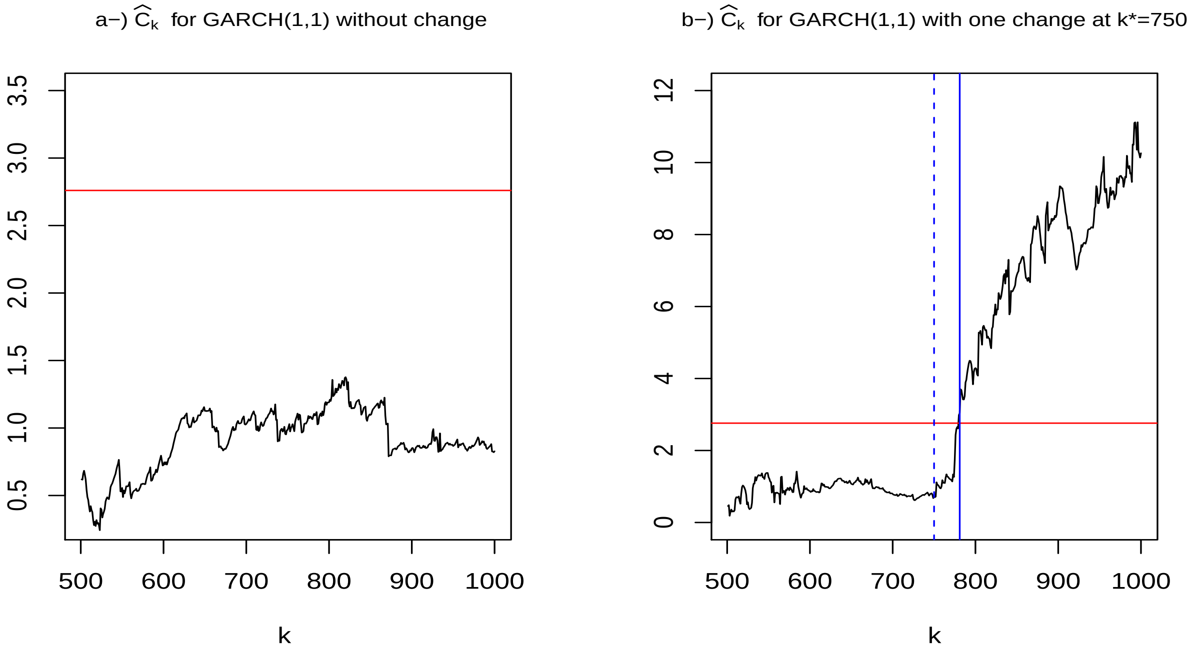

Figure 1: Typical realization of the statistics $\widehat{C}_k$ for $n = 500$ and $k = 501, \cdots, 1000$. a-) The parameter $\theta_0^* = (0.01, 0.3, 0.2)$ is constant ; b-) the parameter $\theta_0^* = (0.01, 0.3, 0.2)$ changes to $\theta_1^* = (0.05, 0.5, 0.2)$ at $k^* = 750$. The horizontal solid line represents the limit of the critical region, the vertical dotted line indicates where the change occurs and the vertical solid line indicates the time where the monitoring procedure detecting the change.

4.1 An illustration

We consider GARCH(1,1) process : $X_t = \sigma_t \xi_t$ with $\sigma_t^2 = \alpha_0^* + \alpha_1^* X_{t-1}^2 + \beta_1^* \sigma_{t-1}^2$. Thus, the parameter of model is $\theta_0^* = (\alpha_0^*, \alpha_1^*, \beta_1^*)$. The historical available data are $X_1, \cdots, X_{500}$ (therefore $n = 500$) and the monitoring period is $\{501, \cdots, 1000\}$. At the nominal level $\alpha = 0.05$, the critical values of the procedure is $C_\alpha = 2.760$. The Figure 1 is a typical realization of the statistic $(\widehat{C}_k)_{500<k\leq 1000}$. We consider a scenario without change (Figure 1 a-)) and a scenario with change at $k^* = n + 250 = 750$ (Figure 1 b-)). Figure 1 a-) shows that, the detector $\widehat{C}_k$ is bellow under the horizontal line which represents the limit of the critical region. On Figure 1 b-) we can see that, before change occurs, $\widehat{C}_k$ is bellow under the horizontal line and increases with a high speed after change. Such growth over a long period indicates that something happening in the model.

4.2 Monitoring mean shift in times series

Let $(X_1, \cdots, X_n)$ be an (historical) observation of a process $X = (X_t)_{t \in \mathbf{Z}}$. We assume that X satisfy

$$\begin{cases} X_t = \mu_0 + \epsilon_t & \text{for} \quad 1 \leq t \leq k^* \\ X_t = \mu_1 + \epsilon_t & \text{for} \quad t > k^* \end{cases}$$

with $k^* > n$, $\mu_0 \neq \mu_1$ and (ϵ_t) a zero mean stationary time series belongs to a class $\mathcal{M}_{\mathbf{Z}}(f_\theta, M_\theta)$. Under H_0, $k^* = \infty$. The monitoring procedure start at $k = n + 1$ and the aim is to test mean shift over the new observation $X_{n+1}, X_{n+2}, \cdots$. This problem can be see as monitoring changes in linear model (see Horváth *et al.* [16], Aue *et al.* [2]) with constant regressor. The empirical mean

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

is a consistent estimator of μ_0. The recursive residual is defined by

$$\widehat{\epsilon}_k = X_k - \overline{X}_n \ ; \ \text{for } k > n.$$

Horváth *et al.* [16] and Aue *et al.* [2] proposed the detector

$$\widehat{Q}_k = \frac{1}{\widehat{\sigma}_n} \frac{1}{c\sqrt{n}(\frac{k}{n})(1-\frac{n}{k})^{\gamma}} \Big| \sum_{i=n+1}^{k} \widehat{\epsilon}_i \Big| \quad k > n;\, c > 0;\, 0 \leq \gamma < 1/2$$

where $\widehat{\sigma}_n^2$ is an consistent estimator of the long-run variance

$$\sigma^2 = \lim_{n\to\infty} \frac{1}{n} Var(\sum_{i=1}^{n} \epsilon_i).$$

If the process (ϵ_t) are uncorrelated (for instance GARCH-type model), empirical variance of the historical data can be used as estimator of σ^2. If (ϵ_t) are correlated, the popular Bartlett estimator (see [8]) can be used. Under some regular conditions, it hold that (see [16] and [2])

$$\lim_{n\to\infty} P\{\tau(n) < \infty\} = P\Big\{ \sup_{0<s<1} \frac{|W_1(s)|}{s^{\gamma}} > c \Big\}.$$

Hence, at a nominal level $\alpha = 0.05$, the critical value of the test is the $(1-\alpha)$-quantile of the distribution of $\sup_{0<s<1} \|W_1(s)\|/s^{\gamma}$. When $\gamma = 0$, these quantiles are known (see Table 1 of [25] for values obtained through a Monte Carlo simulation).

We compare our procedure to the previous residuals CUSUM one (with $\gamma = 0$) in two situations

1. (ϵ_t) is an AR(1) process; $\epsilon_t = \phi_1^* \epsilon_{t-1} + \xi_t$ with $\phi_1^* = 0.2$;
2. (ϵ_t) is a GARCH$(1,1)$ process; $\epsilon_t = \sigma_t \xi_t$ with $\sigma_t^2 = \alpha_0^* + \alpha_1^* \epsilon_{t-1}^2 + \beta_1^* \sigma_{t-1}^2$ and $(\alpha_0^*, \alpha_1^*, \beta_1^*) = (0.01, 0.3, 0.2)$.

The historical sample size are $n = 500$ and $n = 1000$. These procedure are evaluated at times $k = n+100, n+200, n+300, n+400, n+500$, while the change occurs at $k^* = n+50$ or $k^* = n+250$. Tables 2 and 3 indicate the empirical levels and the empirical powers based of 200 replications. The elementary statistics of the empirical detection delay are reported in Tables 4.

The results of Table 2 and Table 3 show that both procedure based on detectors $\widehat{C}_k$ and $\widehat{Q}_k$ are more conservative. One can also see that, increasing length n of historical data reduces the size distortion of these procedures. This is due to the fact that the length of monitoring period is fixed and not increasing with n.
Under H_1, the change has been detected before the monitoring time $k = n+500$. But, as we mentioned above, the challenge of this problem is to minimize the detection delay. For this criteria, it is seen in Table 4 that for the mean shift in AR process, our procedure works well as Horváth *et al.'s* procedure when the change occurs at the beginning of the monitoring ($k^* = n+50$) and it is little better when the change occurs long time after the beginning of the monitoring ($k^* = n+250$). For the mean shift in GARCH process, our test procedure outperforms the Horváth *et al.'s* test in terms of mean, Q_1, median and Q_3 of detection delay.

4.3 Monitoring parameter change in AR(1) and GARCH(1,1) models

In this subsection, we present some simulations results for monitoring parameter change in AR(1) and GARCH(1,1) models and compare our procedure to that proposed by Na *et al.* [25]. First recall that if the boundary function $b \equiv c > 0$ is constant, this procedure is based on the relation

$$P\{\tau(n) < \infty\} = P\Big\{ \sup_{k>n} \widehat{D}_k > c \Big\}$$

		k	$n+100$	$n+200$	$n+300$	$n+400$	$n+500$
Empirical levels	$n=500$	$\widehat{C}_k$	0.000	0.000	0.010	0.015	0.015
		$\widehat{Q}_k$	0.000	0.005	0.005	0.010	0.015
	$n=1000$	$\widehat{C}_k$	0.000	0.000	0.000	0.005	0.010
		$\widehat{Q}_k$	0.000	0.000	0.005	0.000	0.010
Empirical powers	$n=500$; $k^*=n+50$	$\widehat{C}_k$	0.310	1	1	1	1
		$\widehat{Q}_k$	0.335	1	1	1	1
	$n=500$; $k^*=n+250$	$\widehat{C}_k$	0.000	0.000	0.190	1	1
		$\widehat{Q}_k$	0.000	0.000	0.130	0.965	1
	$n=1000$; $k^*=n+50$	$\widehat{C}_k$	0.075	1	1	1	1
		$\widehat{Q}_k$	0.095	1	1	1	1
	$n=1000$; $k^*=n+250$	$\widehat{C}_k$	0.000	0.000	0.135	1	1
		$\widehat{Q}_k$	0.000	0.000	0.075	0.980	1

Table 2: Empirical levels and powers for monitoring means shift in AR(1) with $\phi_1^*=0.2$. The empirical levels are computed when $\mu_0=0$ and the empirical powers are computed when the mean $\mu_0=0$ changes to $\mu_1=1.2$.

		k	$n+100$	$n+200$	$n+300$	$n+400$	$n+500$
Empirical levels	$n=500$	$\widehat{C}_k$	0.005	0.015	0.030	0.055	0.060
		$\widehat{Q}_k$	0.000	0.000	0.005	0.010	0.010
	$n=1000$	$\widehat{C}_k$	0.000	0.005	0.005	0.010	0.015
		$\widehat{Q}_k$	0.000	0.000	0.000	0.010	0.010
Empirical powers	$n=500$; $k^*=n+50$	$\widehat{C}_k$	1	1	1	1	1
		$\widehat{Q}_k$	1	1	1	1	1
	$n=500$; $k^*=n+250$	$\widehat{C}_k$	0.010	0.015	1	1	1
		$\widehat{Q}_k$	0.000	0.000	0.920	1	1
	$n=1000$; $k^*=n+50$	$\widehat{C}_k$	0.995	1	1	1	1
		$\widehat{Q}_k$	0.985	1	1	1	1
	$n=1000$; $k^*=n+250$	$\widehat{C}_k$	0.000	0.000	0.980	1	1
		$\widehat{Q}_k$	0.000	0.000	0.765	1	1

Table 3: Empirical levels and powers for monitoring means shift in GARCH(1,1) with $(\alpha_0^*,\alpha_1^*,\beta_1^*)=(0.01,0.3,0.2)$. The empirical levels are computed when $\mu_0=0$ and the empirical powers are computed when the mean $\mu_0=0$ changes to $\mu_1=0.3$.

$\widehat{d}_n$			Mean	SD	Min	Q_1	Med	Q_3	Max
AR(1)	$n = 500$; $k^* = n + 50$	$\widehat{C}_k$	54.74	14.95	18	44	54	64	103
		$\widehat{Q}_k$	53.78	14.72	16	43	54	63	102
	$n = 500$; $k^* = n + 250$	$\widehat{C}_k$	63.14	23.18	12	45	61	77	135
		$\widehat{Q}_k$	72.70	21.47	7	56	71.5	90	139
	$n = 1000$; $k^* = n + 50$	$\widehat{C}_k$	75.84	14.19	37	66	75	83	114
		$\widehat{Q}_k$	72.60	13.23	41	63	73	82	111
	$n = 1000$; $k^* = n + 250$	$\widehat{C}_k$	76.24	19.15	23	60	76	89	140
		$\widehat{Q}_k$	86.82	22.57	27	70	85	100	151
GARCH(1,1)	$n = 500$; $k^* = n + 50$	$\widehat{C}_k$	20.21	6.15	1	16	20	24	35
		$\widehat{Q}_k$	27.06	4.52	16	24	27	30	44
	$n = 500$; $k^* = n + 250$	$\widehat{C}_k$	25.53	8.04	3	20	25	31	50
		$\widehat{Q}_k$	35.40	10.01	13	28	35	41	62
	$n = 1000$; $k^* = n + 50$	$\widehat{C}_k$	28.43	7.41	6	24	28	33	51
		$\widehat{Q}_k$	36.98	5.09	21	33	37	40	48
	$n = 1000$; $k^* = n + 250$	$\widehat{C}_k$	31.16	8.52	4	26	33	39	53
		$\widehat{Q}_k$	44.35	10.04	14	37	45	50	71

Table 4: Elementary statistics of the empirical detection delay for monitoring mean shift in AR(1) and GARCH(1,1).

where

$$\widehat{D}_k := \sqrt{n}\left\|\widehat{G}(T_{1,n})^{-1/2}\widehat{F}(T_{1,n})\big(\widehat{\theta}(T_{1,k})-\widehat{\theta}(T_{1,n})\big)\right\|.$$

Na *et al.* show that under H_0,

$$\lim_{n\to\infty} P\{\tau(n)<\infty\} = \lim_{n\to\infty} P\Big\{\sup_{k>n}\widehat{D}_k > c\Big\} = P\Big\{\sup_{0<s<1}\|W_d(s)\| > c\Big\}.$$

Hence, at a nominal level α, the critical value of their procedure is the $(1-\alpha)$-quantile of the distribution of $\sup_{0<s<1}\|W_d(s)\|$ which can be found in Table 1 of [25].

The comparisons are made in the followings situations.

1. For **AR(1) model** : $X_t = \phi_1^* X_{t-1} + \xi_t$. Under H_0, $\theta_0 = \phi_1^* = 0.2$. Under H_1, θ_0 changes to $\theta_1 = -0.5$ at k^*.
2. For **GARCH(1,1) model** : $X_t = \sigma_t \xi_t$ with $\sigma_t^2 = \alpha_0^* + \alpha_1^* X_{t-1}^2 + \beta_1^* \sigma_{t-1}^2$. Under H_0, $\theta_0 = (\alpha_0^*, \alpha_1^*, \beta_1^*) = (0.01, 0.3, 0.2)$, while under H_1, $\theta_0 = (0.01, 0.3, 0.2)$ changes to $\theta_1 = (0.05, 0.5, 0.2)$ at k^*.

The historical sample size are $n = 500$ and $n = 1000$. These procedure are evaluated at times $k = n+100, n+200, n+300, n+400, n+500,$, while the change occurs at $k^* = n+50$ or $k^* = n+250$. Tables 5 and 6 indicate the empirical levels and the empirical powers based of 200 replications. The elementary statistics of the empirical detection delay are reported in Tables 7.

		k	$n+100$	$n+200$	$n+300$	$n+400$	$n+500$
Empirical levels	$n=500$	$\widehat{C}_k$	0.000	0.000	0.010	0.010	0.035
		$\widehat{D}_k$	0.000	0.000	0.000	0.000	0.025
	$n=1000$	$\widehat{C}_k$	0.000	0.000	0.000	0.010	0.025
		$\widehat{D}_k$	0.000	0.000	0.000	0.000	0.020
Empirical powers	$n=500$; $k^*=n+50$	$\widehat{C}_k$	0.335	1	1	1	1
		$\widehat{D}_k$	0.175	0.985	1	1	1
	$n=500$; $k^*=n+250$	$\widehat{C}_k$	0.000	0.000	0.180	990	1
		$\widehat{D}_k$	0.000	0.000	0.095	0.865	1
	$n=1000$; $k^*=n+50$	$\widehat{C}_k$	0.065	0.995	1	1	1
		$\widehat{D}_k$	0.090	0.975	1	1	1
	$n=1000$; $k^*=n+250$	$\widehat{C}_k$	0.000	0.000	0.140	0.990	1
		$\widehat{D}_k$	0.000	0.000	0.075	0.855	0.995

Table 5: Empirical levels and powers for monitoring parameter change in AR(1) process. The empirical levels are computed when $\theta_0 = \phi_1^* = 0.2$ is constant and the empirical powers are computed when $\theta_0 = 0.2$ changes to $\theta_1 = -0.5$.

The considered AR and GARCH processes have zero mean. Contrary to the mean shift studied above, this mean is not estimated. For AR model, it appears in Table 5 that both procedures based on detector $\widehat{C}_k$ and $\widehat{D}_k$ are conservative. This is not the case for GARCH model (Table 6). The high size distortions when $n = 500$ is due to the difficulty to estimate the parameter of GARCH model. This size distortion decreases when n increases and Corollary 3.1 ensures that with infinite monitoring period, the empirical level tends to the nominal one as $n \to \infty$.

For both the cases of AR and GARCH processes, the procedure based on detector $\widehat{C}_k$ detect the change after the monitoring time $k = n+500$. Unlike Na *et al.* [25], we consider a scenario of GARCH model with moderate change in parameter, it is seen in Table 6 that the procedure based on detector $\widehat{D}_k$ provides unsatisfactory results. At the monitoring time $k = n+500$, it is not sure that the change must be detected even when $k^* = n+50$. This

		k	$n+100$	$n+200$	$n+300$	$n+400$	$n+500$
Empirical levels	$n=500$	$\widehat{C}_k$	0.010	0.025	0.040	0.095	0.105
		$\widehat{D}_k$	0.010	0.015	0.040	0.040	0.055
	$n=1000$	$\widehat{C}_k$	0.000	0.000	0.030	0.045	0.055
		$\widehat{D}_k$	0.000	0.000	0.010	0.015	0.035
Empirical powers	$n=500$; $k^*=n+50$	$\widehat{C}_k$	0.890	1	1	1	1
		$\widehat{D}_k$	0.390	0.855	0.930	0.965	0.985
	$n=500$; $k^*=n+250$	$\widehat{C}_k$	0.010	0.030	0.825	1	1
		$\widehat{D}_k$	0.010	0.020	0.270	0.805	0.915
	$n=1000$; $k^*=n+50$	$\widehat{C}_k$	0.835	1	1	1	1
		$\widehat{D}_k$	0.310	0.970	0.990	0.995	0.995
	$n=1000$; $k^*=n+250$	$\widehat{C}_k$	0.000	0.005	0.685	1	1
		$\widehat{D}_k$	0.000	0.000	0.250	0.955	0.990

Table 6: Empirical levels and powers for monitoring parameter change in GARCH(1,1) process. The empirical levels are computed when $\theta_0 = (\alpha_0^*, \alpha_1^*, \beta_1^*) = (0.01, 0.3, 0.2)$ is constant (hypothesis H_0) and the empirical powers are computed when $\theta_0 = (0.01, 0.3, 0.2)$ changes to $\theta_1 = (0.05, 0.5, 0.2)$ (hypothesis H_1).

$\widehat{d}_n$			Mean	SD	Min	Q_1	Med	Q_3	Max
AR(1)	$n=500$; $k^*=n+50$	$\widehat{C}_k$	55.36	18.75	9	42	56	67	121
		$\widehat{D}_k$	71.54	38.44	2	52.75	69	89	167
	$n=500$; $k^*=n+250$	$\widehat{C}_k$	66.81	25.27	5	49	65	83	149
		$\widehat{D}_k$	97.80	39.42	21	68	89	123	222
	$n=1000$; $k^*=n+50$	$\widehat{C}_k$	75.13	19.87	24	62	74	90	147
		$\widehat{D}_k$	87.70	28.72	14	66	85	109	195
	$n=1000$; $k^*=n+250$	$\widehat{C}_k$	76.89	26.16	15	56	77	96	172
		$\widehat{D}_k$	101.20	37.97	20	75	96	129	245
GARCH(1,1)	$n=500$; $k^*=n+50$	$\widehat{C}_k$	29.41	15.84	4	22	31	40	98
		$\widehat{D}_k$	86.05	90.50	2	36	61	99	416
	$n=500$; $k^*=n+250$	$\widehat{C}_k$	38.02	19.33	5	27	37	44	113
		$\widehat{D}_k$	87.72	50.96	1	49.25	79	112	236
	$n=1000$; $k^*=n+50$	$\widehat{C}_k$	41.96	13.93	3	32	41	48	94
		$\widehat{D}_k$	71.29	37.12	6	46	66	88	287
	$n=1000$; $k^*=n+250$	$\widehat{C}_k$	44.99	17.16	5	35	41	52	117
		$\widehat{D}_k$	75.78	35.10	7	52	71	95	198

Table 7: Elementary statistics of the empirical detection delay for monitoring parameter change in AR(1) and GARCH(1,1).

is not surprising according to the comment of subsection 3.1.

Table 7 indicates the distribution of the detection delay $\widehat{d}_n$. We can see in Table 7 (even in Table 4) that for our procedure, the relation $\widehat{d}_{1000} \leq \sqrt{1000/500}\ \widehat{d}_{500}$ is globally satisfied (from Theorem 3.2, we deduced that $\widehat{d}_n = \mathcal{O}_P\big(n^{1/2}\log n\big)$ when is large enough). It is also seen that, mean, Q_1, median and Q_3 of our test are shorter than Na *et al.*'s one. The results of Table 5, 6 and 7 show that, our test is uniformly better and the procedure based on detector $\widehat{C}_k$ is highly recommended.

5 Real-Data Applications

We consider the returns of the daily closing values of the Nikkei 225 stock index (from January 2, 1995 to October 19, 1998), S&P 500 and FTSE 100 (from January 2, 2004 to June 11, 2012). These data are available on Yahoo! Finance at `http://finance.yahoo.com/`. They are represented on Figure 2 and Figure 6. These series are known to represent ARCH effect and GARCH(1,2) (resp. GARCH(1,1)) can be used to capture it in returns of Nikkei 225 (resp. S&P 500 and FTSE 100), see the book of Francq and Zakoïan 2010. Let us take the observations going from January 2, 1995 to December 31, 1996 (resp. January 2, 2004 to December 30, 2005) as a historical data for the Nikkei 225 (resp. S&P 500 and FTSE 100) stock index. These periods are known to be stable in financial community. To verify it, we apply three procedures to test for parameter change in the historical observations. The null hypothesis is that the parameter is constant over the observations against the parameter changes alternative.

- The first test is proposed by Kengne [18]. Define the asymptotic covariance matrix (which take into account the change possibility) of the estimator $\widehat{\theta}_n(T_{1,n})$ by

$$\widehat{\Sigma}_{n,k} := \frac{k}{n}\widehat{F}_n(T_{1,k})\widehat{G}_n(T_{1,k})^{-1}\widehat{F}_n(T_{1,k})\mathbf{1}_{\det(\widehat{G}_n(T_{1,k}))\neq 0} + \frac{n-k}{n}\widehat{F}_n(T_{k,n})\widehat{G}_n(T_{k,n})^{-1}\widehat{F}_n(T_{k,n})\mathbf{1}_{\det(\widehat{G}_n(T_{k,n}))\neq 0}.$$

The test is based on the statistic

$$\widehat{Q}_n := \max\big(\widehat{Q}_n^{(1)}, \widehat{Q}_n^{(2)}\big) \quad \text{where}$$

$$\widehat{Q}_n^{(1)} := \max_{v_n \leq k \leq n-v_n} \widehat{Q}_{n,k}^{(1)} \quad \text{with} \quad \widehat{Q}_{n,k}^{(1)} := \frac{k^2}{n}\big(\widehat{\theta}_n(T_{1,k}) - \widehat{\theta}_n(T_n)\big)'\widehat{\Sigma}_{n,k}\big(\widehat{\theta}_n(T_k) - \widehat{\theta}_n(T_n)\big),$$

$$\widehat{Q}_n^{(2)} := \max_{v_n \leq k \leq n-v_n} \widehat{Q}_{n,k}^{(2)} \quad \text{with} \quad \widehat{Q}_{n,k}^{(2)} := \frac{(n-k)^2}{n}\big(\widehat{\theta}_n(T_{k,n}) - \widehat{\theta}_n(T_{1,n})\big)'\widehat{\Sigma}_{n,k}\big(\widehat{\theta}_n(T_{k,n}) - \widehat{\theta}_n(T_{k,n})\big).$$

This test is applied with $v_n = (\log n)^\delta$ where $2 \leq \delta \leq 5/2$.

- The second test (see Lee and Song [22]) is based on the statistic

$$\widehat{Q}_n^{(0)} := \max_{v_n \leq k \leq n-v_n}\Big(\frac{k^2}{n}\big(\widehat{\theta}_n(T_{1,k}) - \widehat{\theta}_n(T_{1,n})\big)'\widehat{\Sigma}_{n,n}\big(\widehat{\theta}_n(T_{1,k}) - \widehat{\theta}_n(T_{1,n})\big)\Big)$$

with $v_n = (\log n)^2$.

- The third procedure is the CUSUM test see Kulperger and Yu [21].

At a nominal level $\alpha \in (0,1)$, each of these procedure rejects null hypothesis if the test statistic is greater than a critical value C_α. Table 8 provides the results of these tests to the historical data that we have chosen.

	$\widehat{Q}_n$	$\widehat{Q}_n^0$	CUSUM
Nikkei 225	3.35 (3.98)	2.31 (3.45)	0.98 (1.36)
S&P 500	2.13 (3.47)	2.01 (3.06)	0.93 (1.36)
FTSE 100	1.95 (3.47)	2.31 (3.06)	1.13 (1.36)

Table 8: Results of test for parameter changes in the historical data of Nikkei 225 (from January 2, 1995 to December 31, 1996), S&P 500 and FTSE 100 (from January 2, 2004 to December 30, 2005). Figures in brackets the critical values of the procedure at the nominal level $\alpha = 0.05$.

Note that, these series are very closed to a nonstationary process, in the sense that $\sum_{j=1}^{q}\alpha_j + \sum_{j=1}^{p}\beta_j \simeq 1$ (see (3)). Therefore, it would be difficult to compute the estimator $\widehat{\theta}_n(T_{k,l})$ $(1 \leq k < l \leq n)$. For the statistics $\widehat{Q}_n$ and $\widehat{Q}_n^{(0)}$, we consider only the time k that the computation of $\widehat{\theta}_n(T_{1,k})$ and $\widehat{\theta}_n(T_{k,n})$ converges. This certainly introduces distortions on these tests. On the other hand, the CUSUM procedure needs to compute only the estimator $\widehat{\theta}_n(T_{1,n})$ which convergence is obtained. According to these results, we conclude that the parameter does not change over these historical observations.

For Nikkei 225 data, monitoring starting at January 2, 1997. Figure 2 shows the realization of the sequence $(\widehat{C}_k)$ for k going from January 2, 1997 to October 19, 1998. Monitoring procedure stops at October 27, 1997. Recall that, the monitoring scheme can be used as an alarm system. When it triggered, we need to apply retrospective test to estimate the breakpoint. According to Kengne [18], the test based on $\widehat{Q}_n$ and $\widehat{Q}_n^{(0)}$ are more powerful than the CUSUM test. Thus, in the retrospective procedure, we applied these two tests and considered the one which provides more significant result (in terms of p-value).
Retrospective procedure is applied to the observations going from January 2, 1995 to October 27, 1997 and break is detected at $\widehat{t}_N \simeq$ September 17, 1997; see Figure 2. This change corresponds to the Asian financial crisis (1997-1998) where the turmoil period started at July 1997.

We are going to see for S&P 500 and FTSE 100 data how multiple changes can be monitored. For these series, monitoring starts at January 2, 2006. Figure 3 represents the sequence $(\widehat{C}_k)$ for k going from January 2, 2006 to December 31, 2008. According to the Figure 3, the monitoring stops at November 16, 2007 and September 4, 2007 for S&P 500 and FTSE 100 data respectively. Retrospective procedure is applied to the series going from January 2, 2006 to November 16, 2007 (for S&P 500 data) and the series going from January 2, 2006 to September 4, 2007 (for FTSE 100 data). Breaks are detected at $\widehat{t}_{S,1} \simeq$ June 18, 2007 (for S&P 500 data) and $\widehat{t}_{F,1} \simeq$ July 6, 2007 (for FTSE 100 data); see Figure 6. These breaks correspond to the beginning of the Subprime Crisis in US.
After monitored the first change, we need to update the procedure. The new historical data are the series going from $\widehat{t}_{S,1}$ to November 16, 2007 (for S&P 500 data) and the series going from $\widehat{t}_{F,1}$ to September 4, 2007 (for FTSE 100 data). Therefore, monitoring continues at November 19, 2007 (for S&P 500 data) and at September 5, 2007 (for FTSE 100 data). Figure 4 shows the curve of the sequence $(\widehat{C}_k)$. The monitoring stops at October 17, 2008 and November 10, 2008 for S&P 500 and FTSE 100 data respectively. The retrospective test is applied and the break point estimation are $\widehat{t}_{S,2} \simeq$ August 14, 2008 and $\widehat{t}_{F,2} \simeq$ September 17, 2008 respectively for these two series; see Figure 6. These breaks correspond to the Lehman Brothers Bankruptcy which affects the worldwide financial system.

After that, the procedure is updated and monitoring continues at October 20, 2008 and November 11, 2008 for S&P 500 and FTSE 100 data. Figure 5 shows the sequence $(\widehat{C}_k)$. The monitoring stopped at March 17, 2009 (S&P 500) and March 9, 2009 (FTSE 100) and retrospective test detected change at $\widehat{t}_{S,3} \simeq$ January 5, 2009 (in S&P 500) and $\widehat{t}_{F,3} \simeq$ December 29, 2008 (in FTSE 100 data); see Figure 6. These breaks correspond to the worldwide governments intervention to solve the financial crisis.
The procedure continues until June 2012, other breaks are detected at $\widehat{t}_{S,4} \simeq$ 26 June 2009, $\widehat{t}_{S,5} \simeq$ 5 April 2010, $\widehat{t}_{S,6} \simeq$ 27 September 2010, $\widehat{t}_{S,7} \simeq$ 19 July 2011, $\widehat{t}_{S,8} \simeq$ 11 January 2012 (for S&P 500 data) and $\widehat{t}_{F,4} \simeq$ 30 June 2009, $\widehat{t}_{F,5} \simeq$ 27 July 2011, $\widehat{t}_{F,6} \simeq$ 21 December 2011 (for FTSE 100 data). They are represented on Figure 6. These breaks correspond to the turmoils periods in the $2010 - 2012^{+}$ Greece and European debt crisis.

Summary of the real-data applications

Both monitoring procedure (based on detector $\widehat{C}_k$) and retrospective test have been applied to detect breaks in the Nikkei 225, S&P 500 and FTSE 100 stock index. The following results are obtained :

1. For the Nikkei 225, from January 2, 1995 to October 19, 1998 ; break is detected at
   - $\widehat{t}_N \simeq$ 17 September 1997 which correspond to the turmoil period of the Asian financial crisis (1997-1998).

See also Figure 2.

2. For the S&P 500 and FTSE 100, from January 2, 2004 to June 11, 2012 ; break are detected at ($\widehat{t}_{S,i}$ and $\widehat{t}_{F,i}$ are referred to the breakpoint in the S&P 500 and FTSE 100 respectively)
   - $\widehat{t}_{S,1} \simeq$ 18 June 2007 and $\widehat{t}_{F,1} \simeq$ 6 July 2007 which correspond to the beginning of the Subprime Crisis in US;
   - $\widehat{t}_{S,2} \simeq$ 14 August 2008 and $\widehat{t}_{F,2} \simeq$ 17 September 2008 which correspond to the Lehman Brothers Bankruptcy;
   - $\widehat{t}_{S,3} \simeq$ 5 January 2009 and $\widehat{t}_{F,3} \simeq$ 29 December 2008 which correspond worldwide governments intervention to solve the financial crisis;
   - $\widehat{t}_{S,4} \simeq$ 26 June 2009, $\widehat{t}_{S,5} \simeq$ 5 April 2010, $\widehat{t}_{S,6} \simeq$ 27 September 2010, $\widehat{t}_{S,7} \simeq$ 19 July 2011, $\widehat{t}_{S,8} \simeq$ 11 January 2012 and $\widehat{t}_{F,4} \simeq$ 30 June 2009, $\widehat{t}_{F,5} \simeq$ 27 July 2011, $\widehat{t}_{F,6} \simeq$ 21 December 2011. These breaks indicates the turmoils periods in the $2010 - 2012^{+}$ Greece and European debt crisis.

See also Figure 6.

6 Conclusions

This paper is devoted to the sequential change-point detection in the parameters of a large class of causals models. We construct a monitoring scheme based on a difference between the historical parameter estimate and the updated parameter estimate computed without the historical observations. The simulation results for the monitoring mean shift in AR(1) and GARCH(1,1) show that our procedure works well (in terms of power and detection delay) as Horváth *et al.'s* procedure when the change occurs at the beginning of the monitoring and it is preferable when the change occurs long time after the beginning of the monitoring. The simulation results for the monitoring parameter change in AR(1) and GARCH(1,1) show that our procedure outperforms (in terms of power and detection delay) the Na *et al.*'s one. The real-data study shows the applicability of the procedure on the daily closing values of the Nikkei 225, S&P 500 and FTSE 100 stock index. It is shown how this monitoring scheme with retrospective test can be used to off-line multiple breaks detection.

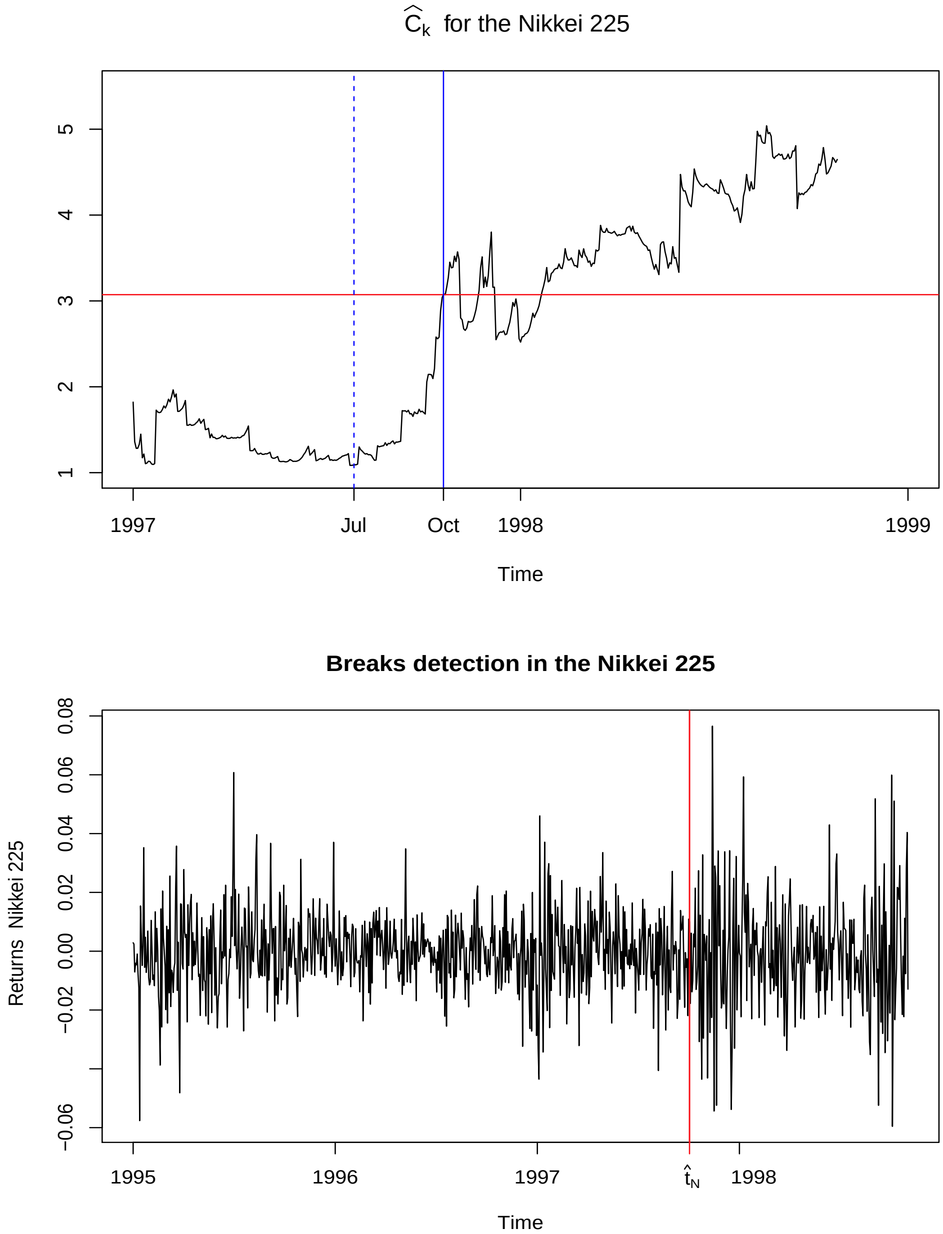

Figure 2: The top figure is a realization of the statistics $\widehat{C}_k$ with k going from January 2, 1995 to October 19, 1998 for Nikkei 225 data; the historical data considered are the series going from January 2, 1995 to December 31, 1996. The horizontal solid line represents the limit of the critical region, the vertical dotted line indicates the date of the beginning of the Asian financial crisis (1997-1998) and the vertical solid line indicates the time where the monitoring procedure will stop. The bottom figure is the returns of Nikkei 225 data from January 2, 1995 to October 19, 1998; the vertical solid line indicates the date where break have been detected using retrospective test after the monitoring stops.

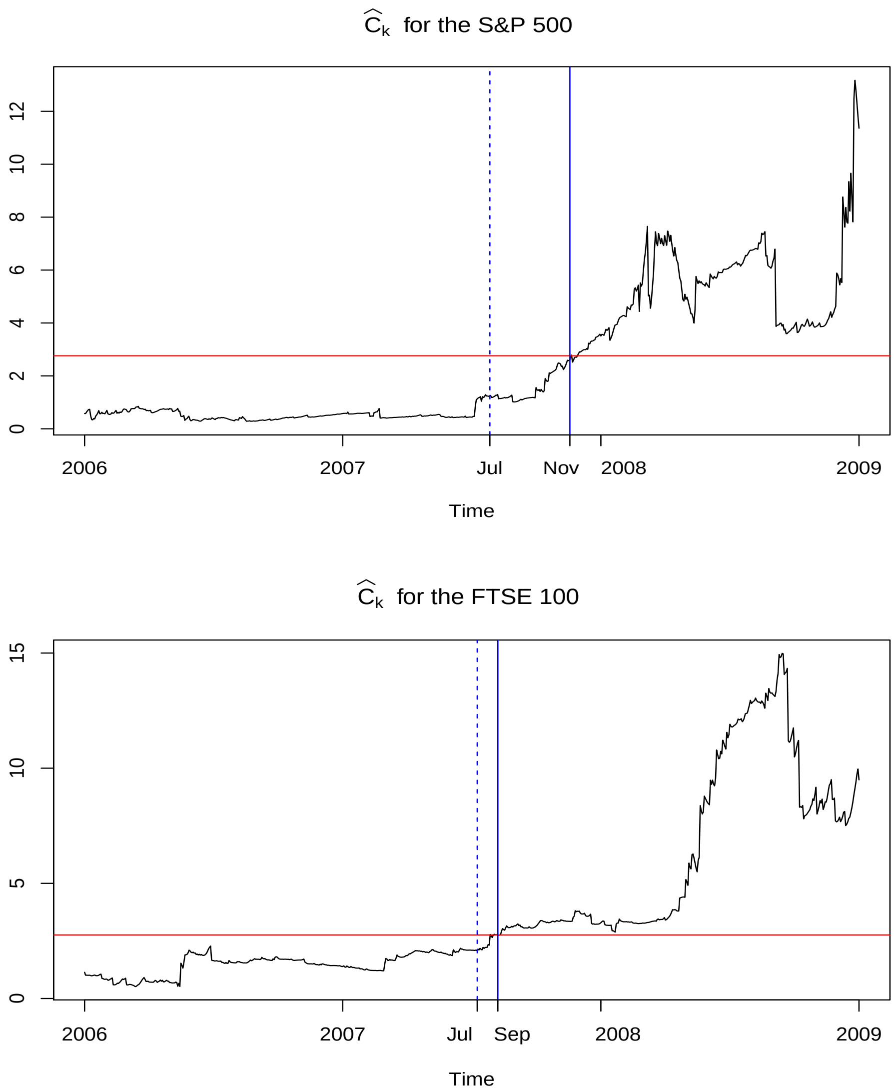

Figure 3: Realization of the statistics $\widehat{C}_k$ with k going from January 2, 2006 to December 31, 2008 for S&P 500 and FTSE 100 data; the historical data considered are the series going from January 2, 2004 to December 30, 2005. The horizontal solid line represents the limit of the critical region, the vertical dotted line indicates the date of the beginning of the Subprime Crisis in US and the vertical solid line indicates the time where the monitoring procedure stopped.

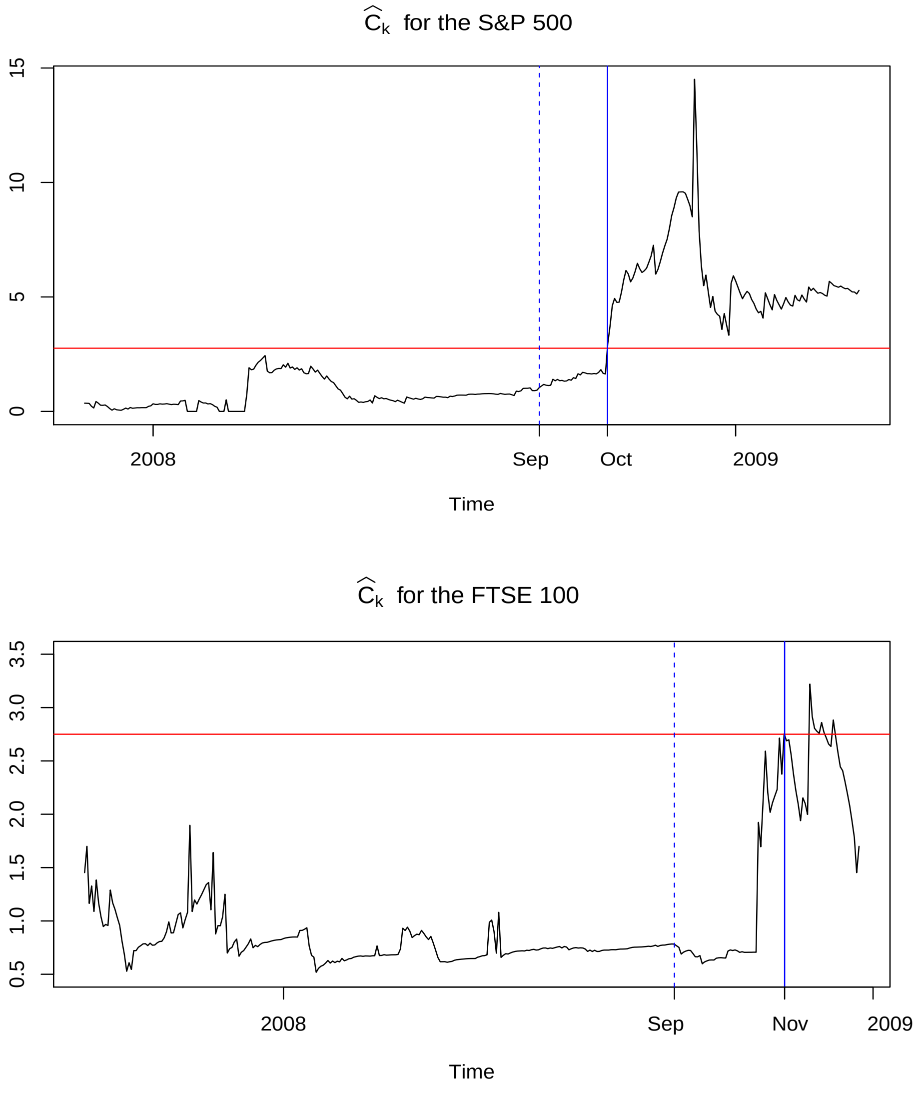

Figure 4: Realization of the statistics $\widehat{C}_k$ for S&P 500 and FTSE 100 data; the historical data are the series going from June 18, 2007 to November 16, 2007 (for S&P 500 data) and July 6, 2007 to September 4, 2007 (for FTSE 100 data). The horizontal solid line represents the limit of the critical region, the vertical dotted line indicates the date of the Lehman Brothers Crisis and the vertical solid line indicates the time where the monitoring procedure stopped.

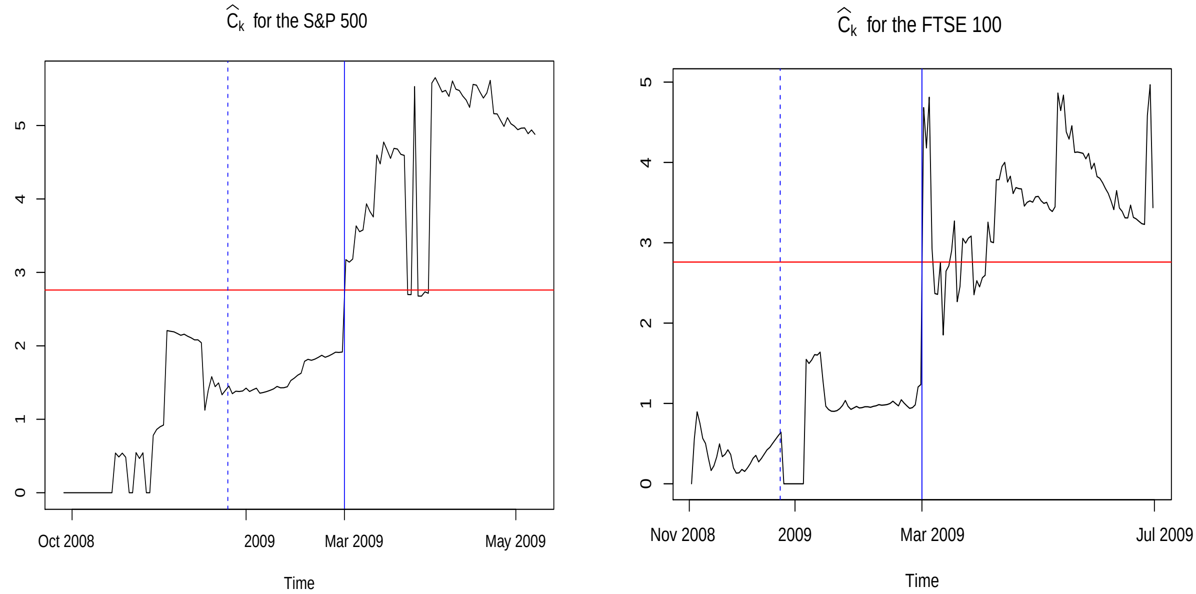

Figure 5: Realization of the statistics $\widehat{C}_k$ for S&P 500 and FTSE 100 data; the historical data are the series going from August 14, 2008 to October 17, 2008 (S&P 500) and September 17, 2008 to November 10, 2008 (FTSE 100). The horizontal solid line represents the limit of the critical region, the vertical dotted line indicates the date of the beginning of stabilization in financial system due to the worldwide governments intervention and the vertical solid line indicates the time where the monitoring procedure stopped.

7 Proofs of the main results

Let us prove first some useful lemmas. In the sequel, for any $x \in I\!\!R$, $[x]$ denotes the integer part of x. Let $(\psi_n)_n$ and $(r_n)_n$ be sequences of random variables. Throughout this section, we use the notation $\psi_n = o_P(r_n)$ to mean : for all $\varepsilon > 0$, $P(|\psi_n| \geq \varepsilon |r_n|) \to 0$ as $n \to \infty$. Write $\psi_n = O_P(r_n)$ to mean : for all $\varepsilon > 0$, there exists $C > 0$ such that $P(|\psi_n| \geq C|r_n|) \leq \varepsilon$ for n large enough.

Recall that $(X_1, \cdots, X_n)$ is an observed trajectory of a process $\mathcal{M}_{\mathbf{Z}}(M_{\theta_0^*}, f_{\theta_0^*})$.
Let $k \geq n \geq 2$ and $T_{1,n} = \{1, \cdots, n\}$, $T_{\ell,k} = \{\ell, \ell+1, \cdots, k\}$ with $\ell \in \Pi_{n,k} = \{v_n, v_n+1, \cdots, k-v_n\}$, and define

$$C_{k,\ell} := \sqrt{n}\,\frac{k-\ell}{k}\big\|G^{-1/2}F \cdot \big(\widehat{\theta}(T_{\ell,k}) - \widehat{\theta}(T_{1,n})\big)\big\|,$$

with $\widehat{\theta}$ defined in (6).

Lemma 7.1 *Under assumptions of Theorem 3.1,*

$$\sup_{k>n}\ \max_{\ell\in\Pi_{n,k}} \frac{1}{b((k-\ell)/n)}\big|\widehat{C}_{k,\ell} - C_{k,\ell}\big| = o_P(1) \quad \textit{as } n \to \infty.$$

Proof. For any $n \geq 1$, we have

$$\sup_{k>n}\ \max_{\ell\in\Pi_{n,k}} \frac{1}{b((k-\ell)/n)}\big|\widehat{C}_{k,\ell} - C_{k,\ell}\big| = \frac{1}{\inf_{s>0} b(s)}\ \sup_{k>n}\ \max_{\ell\in\Pi_{n,k}} \big|\widehat{C}_{k,\ell} - C_{k,\ell}\big|.$$

Now, proceed similarly as in the proof of Lemma 5.3 of [18]. ■

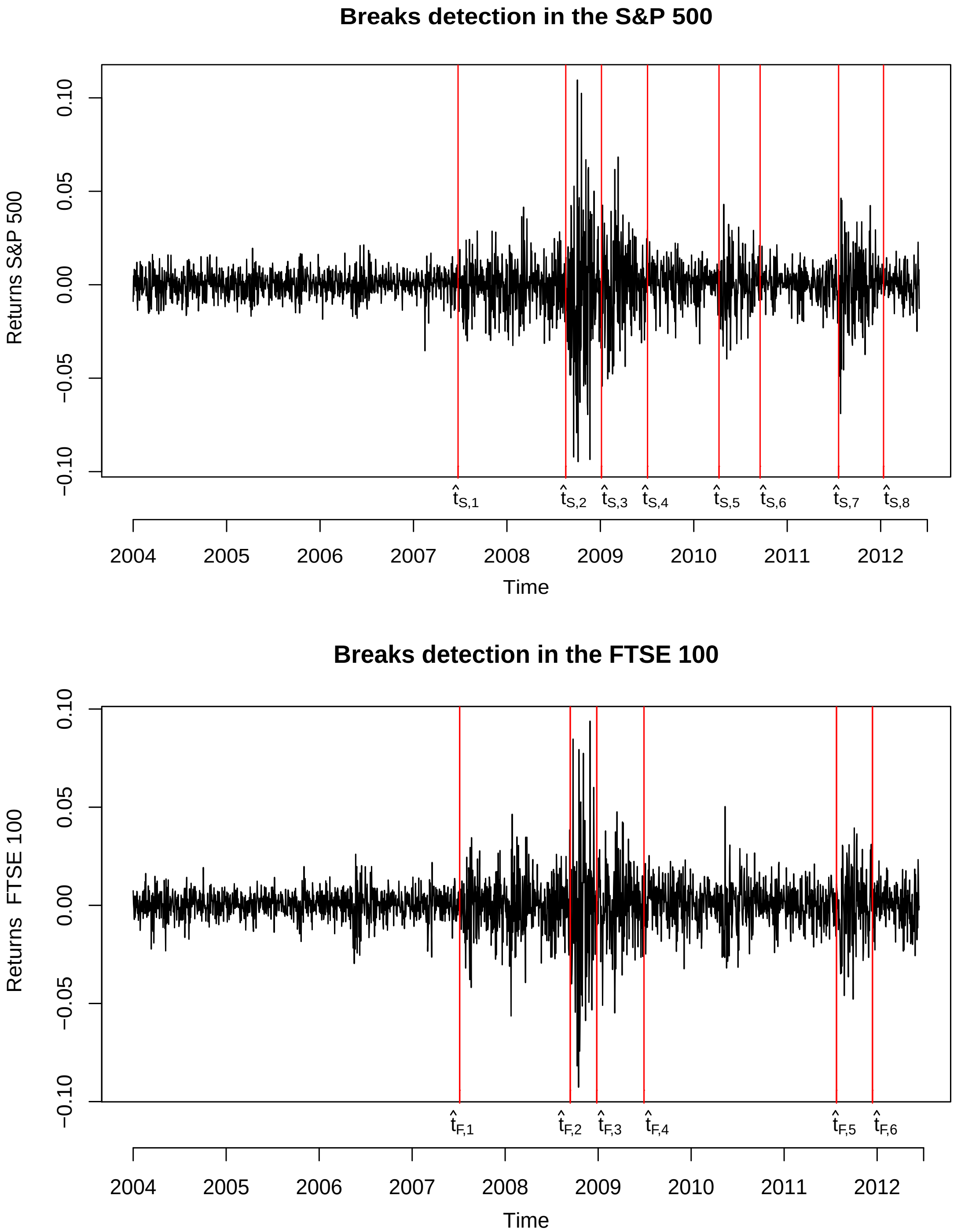

Figure 6: Break detection in the returns of S&P 500 and FTSE 100 data using monitoring procedure based on $\widehat{C}_k$. The verticals lines indicate the dates where breaks have been detected.

Lemma 7.2 *Under assumptions of Theorem 3.1*

$$\sup_{k>n}\max_{\ell\in\Pi_{n,k}}\frac{1}{b((k-\ell)/n)}\frac{\sqrt{n}}{k}\big\|(k-\ell)F\cdot\big(\widehat{\theta}(T_{\ell,k})-\widehat{\theta}(T_{1,n})\big)-2\big(\frac{\partial}{\partial\theta}L(T_{\ell,k},\theta_0^*)-\frac{k-\ell}{n}\frac{\partial}{\partial\theta}L(T_{1,n},\theta_0^*)\big)\big\|$$
$$=o_P(1)\quad as\ n\to\infty.$$

Proof. Let $k\geq n$ and $T\subset\{1,\cdots,k\}$. By applying the Taylor expansion to the coordinates of $\partial\widehat{L}(T,\cdot)/\partial\theta$, and using the fact that $\partial\widehat{L}(T,\widehat{\theta}(T))/\partial\theta=0$ we have

$$\frac{2}{\mathrm{Card}(T)}\frac{\partial}{\partial\theta}\widehat{L}(T,\theta_0^*)=\widetilde{F}(T)\cdot(\widehat{\theta}(T)-\theta_0^*)\ \text{ where }\ \widetilde{F}(T)=-2\big(\frac{1}{\mathrm{Card}(T)}\frac{\partial^2\widehat{L}(T,\widetilde{\theta}_i(T))}{\partial\theta\partial\theta_i}\big)_{1\leq i\leq d}$$

for some $\widetilde{\theta}_i(T)$ between $\widehat{\theta}(T)$ and θ_0^*.
Hence for any $\ell\in\Pi_{n,k}$

$$F\,(\widehat{\theta}(T_{\ell,k})-\theta_0^*)=\frac{2}{k-\ell}\frac{\partial}{\partial\theta}L(T_{\ell,k},\theta_0^*)+\big(F-\widetilde{F}(T_{\ell,k})\big)\big(\widehat{\theta}(T_{\ell,k})-\theta_0^*\big)$$
$$+\frac{2}{k-\ell}\big(\frac{\partial}{\partial\theta}\widehat{L}(T_{\ell,k},\theta_0^*)-\frac{\partial}{\partial\theta}L(T_{\ell,k},\theta_0^*)\big).$$

and

$$F\,(\widehat{\theta}(T_{1,n})-\theta_0^*)=\frac{2}{n}\frac{\partial}{\partial\theta}L(T_{1,n},\theta_0^*)+\big(F-\widetilde{F}(T_{1,n})\big)\big(\widehat{\theta}(T_{1,n})-\theta_0^*\big)+\frac{2}{n}\big(\frac{\partial}{\partial\theta}\widehat{L}(T_{1,n},\theta_0^*)-\frac{\partial}{\partial\theta}L(T_{1,n},\theta_0^*)\big).$$

Therefore, for any $\ell\in\Pi_{n,k}$

$$\begin{aligned}&\frac{\sqrt{n}}{k}\Big((k-\ell)F\,\big(\widehat{\theta}(T_{\ell,k})-\widehat{\theta}(T_{1,n})\big)-2\big(\frac{\partial}{\partial\theta}L(T_{\ell,k},\theta_0^*)-\frac{k-\ell}{n}\frac{\partial}{\partial\theta}L(T_{1,n},\theta_0^*)\big)\Big)\\&\quad=\sqrt{n}\frac{k-\ell}{k}\big(F-\widetilde{F}(T_{\ell,k})\big)\big(\widehat{\theta}(T_{\ell,k})-\theta_0^*\big)+2\frac{\sqrt{n}}{k}\big(\frac{\partial}{\partial\theta}\widehat{L}(T_{\ell,k},\theta_0^*)-\frac{\partial}{\partial\theta}L(T_{\ell,k},\theta_0^*)\big)\\&\qquad-\sqrt{n}\frac{k-\ell}{k}\big(F-\widetilde{F}(T_{1,n})\big)\big(\widehat{\theta}(T_{1,n})-\theta_0^*\big)-2\frac{k-\ell}{k}\frac{1}{\sqrt{n}}\big(\frac{\partial}{\partial\theta}\widehat{L}(T_{1,n},\theta_0^*)-\frac{\partial}{\partial\theta}L(T_{1,n},\theta_0^*)\big).\end{aligned}\tag{16}$$

For $k>n$ and with some $\ell_k\in\Pi_{n,k}$, we have

$$\max_{\ell\in\Pi_{n,k}}\frac{1}{b((k-\ell)/n)}\sqrt{n}\frac{k-\ell}{k}\|F-\widetilde{F}(T_{\ell,k})\|\,\|\widehat{\theta}(T_{\ell,k})-\theta_0^*\|\leq\frac{1}{\operatorname{Inf}_{s>0}b(s)}\sqrt{k-\ell_k}\|F-\widetilde{F}(T_{\ell_k,k})\|\,\|\widehat{\theta}(T_{\ell_k,k})-\theta_0^*\|.$$

According to [4] and [5], $\|F-\widetilde{F}(T_{\ell_k,k})\|=o_P(1)$ and $\|\widehat{\theta}(T_{\ell_k,k})-\theta_0^*\|=O_P(1/\sqrt{k-\ell_k})$ as $k-\ell_k\to\infty$. Hence

$$\sup_{k>n}\max_{\ell\in\Pi_{n,k}}\frac{1}{b((k-\ell)/n)}\sqrt{n}\frac{k-\ell}{k}\|F-\widetilde{F}(T_{\ell,k})\|\,\|\widehat{\theta}(T_{\ell,k})-\theta_0^*\|=o_P(1)\ \text{ as }\ n\to\infty.\tag{17}$$

Similar arguments imply that

$$\sup_{k>n}\max_{\ell\in\Pi_{n,k}}\frac{1}{b((k-\ell)/n)}\sqrt{n}\frac{k-\ell}{k}\|F-\widetilde{F}(T_{1,n})\|\,\|\widehat{\theta}(T_{1,n})-\theta_0^*\|=o_P(1)\ \text{ as }\ n\to\infty.\tag{18}$$

For $k>n$ and for some $\ell_k\in\Pi_{n,k}$, we have

$$\max_{\ell\in\Pi_{n,k}}\frac{1}{b((k-\ell)/n)}\frac{\sqrt{n}}{k}\|\frac{\partial}{\partial\theta}\widehat{L}(T_{\ell,k},\theta_0^*)-\frac{\partial}{\partial\theta}L(T_{\ell,k},\theta_0^*)\|$$
$$\leq\frac{1}{\operatorname{Inf}_{s>0}b(s)}\frac{1}{\sqrt{k-\ell_k}}\|\frac{\partial}{\partial\theta}\widehat{L}(T_{\ell_k,k},\theta_0^*)-\frac{\partial}{\partial\theta}L(T_{\ell_k,k},\theta_0^*)\|.$$

According to [4], $\frac{1}{\sqrt{k-\ell_k}}\|\frac{\partial}{\partial\theta}\widehat{L}(T_{\ell_k,k},\cdot)-\frac{\partial}{\partial\theta}L(T_{\ell_k,k},\cdot)\|_\Theta=o_P(1)$ as $k-\ell_k\to\infty$. Hence

$$\sup_{k>n}\max_{\ell\in\Pi_{n,k}}\frac{1}{b((k-\ell)/n)}\frac{\sqrt{n}}{k}\|\frac{\partial}{\partial\theta}\widehat{L}(T_{\ell,k},\theta_0^*)-\frac{\partial}{\partial\theta}L(T_{\ell,k},\theta_0^*)\|=o_P(1)\ \text{ as }\ n\to\infty.\tag{19}$$

Similar arguments show that

$$\sup_{k>n}\max_{\ell\in\Pi_{n,k}}\frac{1}{b((k-\ell)/n)}\frac{k-\ell}{k}\frac{1}{\sqrt{n}}\|\frac{\partial}{\partial\theta}\widehat{L}(T_{1,n},\theta_0^*)-\frac{\partial}{\partial\theta}L(T_{1,n},\theta_0^*)\|=o_P(1)\quad\text{as}\quad n\to\infty. \tag{20}$$

Thus, Lemma 7.2 follows from (16), (17), (18), (19) and (20). ■

Lemma 7.3 *Under assumptions of Theorem 3.1*

$$\sup_{k>n}\max_{\ell\in\Pi_{n,k}}\frac{1}{b((k-\ell)/n)}\sqrt{n}\frac{k-\ell}{k}\|F\cdot\big(\widehat{\theta}(T_{\ell,k})-\widehat{\theta}(T_{1,n})\big)\|\ \xrightarrow[n\to\infty]{\mathcal{D}}\ \sup_{t>1}\sup_{0<s<t}\frac{\|W_G(s)-sW_G(1)\|}{t\ b(s)}$$

where W_G is a d-dimensional Gaussian centered process with covariance matrix $E\left(W_G(s)W_G(\tau)'\right)=\min(s,\tau)G$.

Proof. We are going to apply Lemma 7.2 for specifying the asymptotic behaviour of $\widehat{\theta}(T_{\ell,k})-\widehat{\theta}(T_{1,n})$.

For $k>n$ and $\ell\in\Pi_{n,k}$, we have

$$2\frac{\sqrt{n}}{k}\big(\frac{\partial}{\partial\theta}L(T_{\ell,k},\theta_0^*)-\frac{k-\ell}{n}\frac{\partial}{\partial\theta}L(T_{1,n},\theta_0^*)\big)=-\frac{n}{k}\frac{1}{\sqrt{n}}\Big(\sum_{i=\ell+1}^{k}\frac{\partial q_i(\theta_0^*)}{\partial\theta}-\frac{k-\ell}{n}\sum_{i=1}^{n}\frac{\partial q_i(\theta_0^*)}{\partial\theta}\Big).$$

Now we are going to proceed in two steps.

Step 1. Let $T>1$. We have

$$\begin{aligned}&\max_{n<k<nT}\max_{\ell\in\Pi_{n,k}}\frac{1}{b((k-\ell)/n)}\frac{2\sqrt{n}}{k}\|\frac{\partial}{\partial\theta}L(T_{\ell,k},\theta_0^*)-\frac{k-\ell}{n}\frac{\partial}{\partial\theta}L(T_{1,n},\theta_0^*)\|\\&=\max_{n<k<nT}\max_{\ell\in\Pi_{n,k}}\frac{1}{b((k-\ell)/n)}\frac{\sqrt{n}}{k}\|\sum_{i=\ell}^{k}\frac{\partial q_i(\theta_0^*)}{\partial\theta}-\frac{k-\ell}{n}\sum_{i=1}^{n}\frac{\partial q_i(\theta_0^*)}{\partial\theta}\|\\&=\max_{t\in\{1,1+\frac{1}{n},\cdots,T\}}\max_{s\in\{1-\frac{v_n}{n},2-\frac{v_n}{n},\cdots,t-\frac{v_n}{n}\}}\frac{1}{b(([nt]-[ns])/n)}\frac{n}{[nt]}\|\frac{1}{\sqrt{n}}\Big(\sum_{i=[ns]+1}^{[nt]}\frac{\partial q_i(\theta_0^*)}{\partial\theta}-\frac{[nt]-[ns]}{n}\sum_{i=1}^{n}\frac{\partial q_i(\theta_0^*)}{\partial\theta}\Big)\|.\end{aligned}$$

Define the set $S:=\{(t,s)\in[1,T]\times[1,T]/\ s<t\}$. According to [4], $\big(\frac{\partial q_i(\theta_0^*)}{\partial\theta}\big)_{t\in\mathbf{Z}}$ is a stationary ergodic martingale difference sequence with covariance matrix G. By Cramér-Wold device (see [9] p. 206), it holds that

$$\frac{1}{\sqrt{n}}\sum_{i=[ns]+1}^{[nt]}\frac{\partial q_i(\theta_0^*)}{\partial\theta}\ \xrightarrow[n\to\infty]{\mathcal{D}(S)}\ W_G(t-s).$$

with $\xrightarrow[n\to\infty]{\mathcal{D}(S)}$ means the weak convergence on the Skorohod space $\mathcal{D}(S)$. Hence

$$\frac{1}{\sqrt{n}}\Big(\sum_{i=[ns]+1}^{[nt]}\frac{\partial q_i(\theta_0^*)}{\partial\theta}-\frac{[nt]-[ns]}{n}\sum_{i=1}^{n}\frac{\partial q_i(\theta_0^*)}{\partial\theta}\Big)\ \xrightarrow[n\to\infty]{\mathcal{D}(S)}\ W_G(t-s)-(t-s)W_G(1).$$

Therefore

$$\begin{aligned}&\max_{n<k<nT}\max_{\ell\in\Pi_{n,k}}\frac{1}{b((k-\ell)/n)}\frac{2\sqrt{n}}{k}\|\frac{\partial}{\partial\theta}L(T_{\ell,k},\theta_0^*)-\frac{k-\ell}{n}\frac{\partial}{\partial\theta}L(T_{1,n},\theta_0^*)\|\\&\qquad\xrightarrow[n\to\infty]{\mathcal{D}}\ \sup_{1<t<T}\ \sup_{1<s<t}\frac{\|W_G(t-s)-(t-s)W_G(1)\|}{t\ b(t-s)}\\&\qquad\xrightarrow[n\to\infty]{\mathcal{D}}\ \sup_{1<t<T}\ \sup_{1<s<t}\frac{\|W_G(s)-s\,W_G(1)\|}{t\ b(s)}.\end{aligned}\tag{21}$$

Step 2. We will show that the limit distribution (as $n,T\to\infty$) of

$$\sup_{k>nT}\max_{\ell\in\Pi_{n,k}}\frac{1}{b((k-\ell)/n)}\frac{2\sqrt{n}}{k}\|\frac{\partial}{\partial\theta}L(T_{\ell,k},\theta_0^*)-\frac{k-\ell}{n}\frac{\partial}{\partial\theta}L(T_{1,n},\theta_0^*)\|$$

exists and is equal to the limit distribution (as $T \to \infty$) of

$$\sup_{t>T} \sup_{1<s<t} \frac{\|W_G(s) - s\, W_G(1)\|}{t\; b(s)}.$$

Let $k > nT$. We have

$$\max_{\ell \in \Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \frac{\sqrt{n}}{k} \|\frac{\partial}{\partial \theta} L(T_{\ell,k}, \theta_0^*)\| \leq \frac{1}{\operatorname{Inf}_{s>0} b(s)} \frac{\sqrt{n}}{k} \| \sum_{i=\ell_k+1}^{k} \frac{\partial q_i(\theta_0^*)}{\partial \theta} \| \quad \text{for some} \quad \ell_k \in \Pi_{n,k}.$$

It comes from the Hájek-Rényi-Chow inequality (see [11]) that, for any $\varepsilon > 0$

$$\lim_{T\to\infty} \lim_{n\to\infty} P\big(\sup_{k>nT} \frac{\sqrt{n}}{k} \| \sum_{i=\ell_k+1}^{k} \frac{\partial q_i(\theta_0^*)}{\partial \theta} \| > \varepsilon\big) = 0.$$

Hence

$$\sup_{k>nT} \max_{\ell \in \Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \frac{2\sqrt{n}}{k} \|\frac{\partial}{\partial \theta} L(T_{\ell,k}, \theta_0^*)\| = o_P(1) \quad \text{as} \quad T, n \to \infty. \tag{22}$$

Moreover, since the function $b(\cdot)$ is non-increasing, for any $n, T > 1$, we have:

$$\begin{aligned}
\sup_{k>nT} \max_{\ell \in \Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \frac{2\sqrt{n}}{k} \|\frac{k-\ell}{n} \frac{\partial}{\partial \theta} L(T_{1,n}, \theta_0^*)\| &= \|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial q_i(\theta_0^*)}{\partial \theta}\| \times \sup_{k>nT} \max_{\ell \in \Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \frac{k-\ell}{k} \\
&= \|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial q_i(\theta_0^*)}{\partial \theta}\| \times \sup_{k>nT} \frac{1}{b((k-v_n)/n)} \frac{k-v_n}{k} \\
&= \frac{1}{\operatorname{Inf}_{s>0} b(s)} \|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial q_i(\theta_0^*)}{\partial \theta}\| \\
&\xrightarrow[n\to\infty]{\mathcal{D}} \frac{1}{\operatorname{Inf}_{s>0} b(s)} \|W_G(1)\|,
\end{aligned} \tag{23}$$

using again the Cramèr-Wold device. It comes from (22) and (23) that

$$\sup_{k>nT} \max_{\ell \in \Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \frac{2\sqrt{n}}{k} \|\frac{\partial}{\partial \theta} L(T_{\ell,k}, \theta_0^*) - \frac{k-\ell}{n} \frac{\partial}{\partial \theta} L(T_{1,n}, \theta_0^*)\| \xrightarrow[T,n\to\infty]{\mathcal{D}} \frac{1}{\operatorname{Inf}_{s>0} b(s)} \|W_G(1)\|. \tag{24}$$

Furthermore, since the coordinates of W_G are Brownian motions, by the law of the iterated logarithm there exists $t_0 > \exp(1)$ such as

$$s > t_0 \Rightarrow \|W_G(s)\| \leq \sqrt{s} \log(s) \quad \text{almost surely.}$$

Thus, for any $t > t_0$, we obtain almost surely

$$\sup_{1<s<t} \|W_G(s)\| \leq \sup_{1<s<t_0} \|W_G(s)\| + \sqrt{t} \log(t).$$

Therefore, for T large enough, we have

$$\sup_{t>T} \sup_{1<s<t} \frac{\|W_G(s)\|}{t\; b(s)} \leq \frac{1}{\inf_{s>0} b(s)} \big(\frac{1}{T} \sup_{1<s<t_0} \|W_G(s)\| + \sup_{t>T} \frac{\log(t)}{\sqrt{t}}\big) \xrightarrow[T\to\infty]{\text{a.s.}} 0. \tag{25}$$

Finally, since $b(\cdot)$ is non-increasing, for any $T > 1$, we have

$$\sup_{t>T} \sup_{1<s<t} \frac{\|s W_G(1)\|}{t\; b(s)} = \|W_G(1)\| \sup_{t>T} \frac{1}{t} \sup_{1<s<t} \frac{s}{b(t)} = \|W_G(1)\| \sup_{t>T} \frac{1}{b(t)} = \frac{1}{\operatorname{Inf}_{s>0} b(s)} \|W_G(1)\|. \tag{26}$$

It comes from (25) and (26) that the limit of (21) satisfies when $T \to \infty$,

$$\sup_{t>T} \sup_{1<s<t} \frac{\|W_G(s) - s W_G(1)\|}{t\; b(s)} \xrightarrow[T\to\infty]{\mathcal{D}} \frac{1}{\operatorname{Inf}_{s>0} b(s)} \|W_G(1)\|. \tag{27}$$

From **Step 1** and **Step 2** (the relations (21), (24) and (27)), it comes that

$$\sup_{k>nT} \max_{\ell\in\Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \frac{2\sqrt{n}}{k} \| \frac{\partial}{\partial\theta} L(T_{\ell,k},\theta_0^*) - \frac{k-\ell}{n} \frac{\partial}{\partial\theta} L(T_{1,n},\theta_0^*) \| \xrightarrow[n\to\infty]{\mathcal{D}} \sup_{t>T} \sup_{1<s<t} \frac{\|W_G(s)-sW_G(1)\|}{t\, b(s)}.$$

Hence, Lemma 7.3 follows from Lemma 7.2. ■

Proof of Theorem 3.1

We know that

$$\begin{aligned} P\{\tau(n)<\infty\} &= P\Big\{ \sup_{k>n} \max_{\ell\in\Pi_{n,k}} \frac{\widehat{C}_{k,\ell}}{b((k-\ell)/n)} > 1 \Big\} \\ &= P\Big\{ \sup_{k>n} \max_{\ell\in\Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \sqrt{n}\frac{k-\ell}{k} \|\widehat{G}(T_{1,n})^{-1/2}\widehat{F}(T_{1,n})\cdot\big(\widehat{\theta}(T_{\ell,k})-\widehat{\theta}(T_{1,n})\big)\| > 1 \Big\}. \end{aligned}$$

Since $\widehat{G}(T_{1,n}) \xrightarrow[n\to\infty]{a.s.} G$ and $\widehat{F}(T_{1,n}) \xrightarrow[n\to\infty]{a.s.} F$, it comes from Lemma 7.1 and 7.3 that

$$\sup_{k>n} \max_{\ell\in\Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \sqrt{n}\frac{k-\ell}{k} \|\widehat{G}(T_{1,n})^{-1/2}\widehat{F}(T_{1,n})\cdot\big(\widehat{\theta}(T_{\ell,k})-\widehat{\theta}(T_{1,n})\big)\|$$
$$\xrightarrow[n\to\infty]{\mathcal{D}} \sup_{t>1} \sup_{1<s<t} \frac{\|G^{-1/2}(W_G(s)-sW_G(1))\|}{t\, b(s)}.$$

Since the covariance matrix of $\{W_G(s)\ ;\ s\geq 0\}$, is $\min(s,\tau)G$, the covariance matrix of $\{G^{-1/2}W_G(s)\ ;\ s\geq 0\}$ is $\min(s,\tau)I_d$ (where I_d is the d-dimensional identity matrix). Hence Theorem 3.1 follows. ■

Proof of Corollary 3.1

Since $b\equiv c$ a positive constant, it follows immediately from Theorem 3.1 that

$$\lim_{n\to\infty} P\{\tau(n)<\infty\} = P\Big\{ \sup_{t>1} \sup_{1<s<t} \frac{1}{t} \|W_d(s)-sW_d(1))\| > c \Big\}.$$

Now, it suffices to show that $\sup_{t>1} \sup_{1<s<t} \frac{1}{t}\|W_d(s)-sW_d(1))\| \stackrel{\mathcal{D}}{=} U_d$.

For any $t>1$, we have

$$\sup_{1<s<t} \frac{1}{t}\|W_d(s)-sW_d(1))\| \stackrel{\mathcal{D}}{=} \sup_{1<s<t} \frac{s}{t}\|W_d(\frac{s-1}{s})\| = \sup_{0<u<1-1/t} \frac{1}{t(1-u)}\|W_d(u)\|.$$

Thus

$$\sup_{t>1} \sup_{1<s<t} \frac{1}{t}\|W_d(s)-sW_d(1))\| \stackrel{\mathcal{D}}{=} \sup_{t>1} \sup_{0<u<1-1/t} \frac{1}{t(1-u)}\|W_d(u)\| = \sup_{0<v<1} \sup_{0<u<v} \frac{1-v}{1-u}\|W_d(u)\|.$$

But, $\|W_d(u)\| \stackrel{\mathcal{D}}{=} v^{1/2}\left\|W_d\big(\frac{u}{v}\big)\right\|$. Therefore with $u=u'v$,

$$\sup_{0<v<1} \sup_{0<u<v} \frac{1-v}{1-u}\|W_d(u)\| = \sup_{0<v<1} \sup_{0<u'<1} \frac{(1-v)v^{1/2}}{1-u'v}\|W_d(u')\|.$$

It remains to compute $\sup_{0<v<1} \frac{(1-v)v^{1/2}}{1-u'v}$. Classical computations show that this supremum is obtained by $v = 2\big(3-u'+\sqrt{(9-u')(1-u')}\big)^{-1}$ and therefore

$$\sup_{t>1} \sup_{1<s<t} \frac{1}{t}\|W_d(s)-sW_d(1))\| \stackrel{\mathcal{D}}{=} \sup_{0<u'<1} f(u')\,\|W_d(u')\|$$
$$\text{with} \quad f(u') = \frac{\sqrt{9-u'}+\sqrt{1-u'}}{\sqrt{9-u'}+3\sqrt{1-u'}} \Big(\frac{2}{3-u'+\sqrt{(9-u')(1-u')}}\Big)^{1/2}. \qquad (28)$$

Hence,

$$\sup_{t>1} \sup_{1<s<t} \frac{1}{t} \|W_d(s) - sW_d(1))\| \stackrel{\mathcal{D}}{=} U_d$$

■

Proof of Theorem 3.2

Denote $k_n = k^* + n^\delta$ for $\delta \in (1/2, 1)$. For n large enough, we have $v_n < n^\delta$ and thus $k_n - v_n = k^* + n^\delta - v_n \geq k^*$. Moreover, since $k^* > n$ then $k^* \in \Pi_{n,k}$ for n large enough.

In addition, since $k^* = k^*(n) \geq n$ and $\limsup_{n\to\infty} k^*(n)/n < \infty$, there exists $c_0 > 1$ such that $k^* \leq c_0 n$ for n large enough. Hence, according to assumption **B**, there exists a constant $c > 0$ such that

$$\begin{aligned}
\max_{\ell \in \Pi_{n,k_n}} \frac{\widehat{C}_{k_n,\ell}}{b((k_n-\ell)/n)} &= \max_{\ell \in \Pi_{n,k_n}} \frac{1}{b((k_n-\ell)/n)} \sqrt{n} \frac{k_n-\ell}{k_n} \|\widehat{G}(T_{1,n})^{-1/2} \widehat{F}(T_{1,n}) \cdot \big(\widehat{\theta}_{k_n}(T_{\ell,k_n}) - \widehat{\theta}(T_{1,n})\big) \\
&\geq \frac{1}{b((k_n-k^*)/n)} \sqrt{n} \frac{k_n-k^*}{k_n} \|\widehat{G}(T_{1,n})^{-1/2} \widehat{F}(T_{1,n}) \cdot \big(\widehat{\theta}_{k_n}(T_{k^*,k_n}) - \widehat{\theta}(T_{1,n})\big) \\
&\geq c\, \sqrt{n}\, \frac{n^\delta}{k^*+n^\delta} \big\|\widehat{G}(T_{1,n})^{-1/2} \widehat{F}(T_{1,n}) \big(\widehat{\theta}_{k_n}(T_{k^*,k_n}) - \widehat{\theta}(T_{1,n})\big)\big\| \\
&\geq c\, \frac{n^{1/2+\delta}}{c_0 n + n^\delta} \big\|\widehat{G}(T_{1,n})^{-1/2} \widehat{F}(T_{1,n}) \big(\widehat{\theta}_{k_n}(T_{k^*,k_n}) - \widehat{\theta}(T_{1,n})\big)\big\| \\
&\geq c\, \frac{n^{\delta-1/2}}{(c_0+1)} \big\|\widehat{G}(T_{1,n})^{-1/2} \widehat{F}(T_{1,n}) \big(\widehat{\theta}_{k_n}(T_{k^*,k_n}) - \widehat{\theta}(T_{1,n})\big)\big\|. \qquad (29)
\end{aligned}$$

According to [4] and [18], $\widehat{G}(T_{1,n}) \xrightarrow[n\to\infty]{a.s.} G$, $\widehat{F}(T_{1,n}) \xrightarrow[n\to\infty]{a.s.} F$, $\widehat{\theta}(T_{1,n}) \xrightarrow[n\to\infty]{a.s.} \theta_0^*$ and $\widehat{\theta}_{k_n}(T_{k^*,k_n}) \xrightarrow[n\to\infty]{a.s.} \theta_1^*$. Since G is symmetric positive definite, F is invertible, $\theta_0^* \neq \theta_1^*$ and $\delta > 1/2$, then (29) implies that

$$\max_{\ell \in \Pi_{n,k_n}} \frac{\widehat{C}_{k_n,\ell}}{b((k_n-\ell)/n)} \xrightarrow[n\to\infty]{a.s.} \infty.$$

■

References

[1] Aue, A., Hörmann, S., Horváth, L. and Reimherr, M. Break detection in the covariance structure of multivariate time series models. *Ann. Statist. 37(6B)*, (2009), 4046-4087.

[2] Aue, A., Horváth, L., Husková, M., and Kokoszka, P. Change point monitoring in linear models. *Econometrics Journal 9*, (2006), 373-403.

[3] Aue, A., Horváth, L. and Reimherr, M. Delay times of sequential procedures for multiple time series regression models. *Journal of Econometrics 149*, (2009), 174-190.

[4] Bardet, J.-M. and Wintenberger, O. Asymptotic normality of the quasi-maximum likelihood estimator for multidimensional causal processes. *Ann. Statist. 37*, (2009), 2730-2759.

[5] Bardet, J.-M. , Kengne, W. and Wintenberger, O. Detecting multiple change-points in general causal time series using penalized quasi-likelihood. *Electronic Journal of Statistics 6*, (2012), 435-477.

[6] Berkes, I., Gombay E., Horváth, L., and Kokoszka, P. Sequential change-point detection in GARCH(p,q) models. *Econometric Theory 20*, (2004), 1140-1167.

[7] Berkes, I., Horváth, L., and Kokoszka, P. GARCH processes: structure and estimation. *Bernoulli 9* (2003), 201–227.

[8] Berkes, I., Horváth, L., Kokoszka, P. and Shao, Q.-M. Almost sure convergence of the Bartlett estimator. *Period Math Hung 51* (2005), 11–25.

[9] Billingsley, P. *Convergence of Probability Measures*. John Wiley & Sons Inc., New York, 1968.

[10] Bollerslev, T. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 31*, (1986), 307-327.

[11] Chow, Y.S. A martingale inequality and the law of large numbers. *Proceedings of the American Mathematics Society , 11*, (1960), 107-111.

[12] Chu, C-SJ., Stinchcombe, M., and White, W. Monitoring structural change. *Econometrica, 64*, (1996), 1045-1065.

[13] Doukhan, P. and Wintenberger, O. Weakly dependent chains with infinite memory. *Stochastic Process. Appl. 118*, (2008) 1997-2013.

[14] Francq, C., and Zakoïan, J.-M. GARCH Models: Structure, Statistical Inference and Financial Applications. *Wiley, Chichester, UK* (2010).

[15] Horváth, L. Change in autoregressive processes. *Stochastic Processes. Appl. 44*, (1993), 221-242.

[16] Horváth, L., Hušková, M., Kokoszka, P., Steinebach, J. Monitoring changes in linear models. *Journal of Statistical Planning and Inference 126*, (2004), 225-251.

[17] Inclan, C., Tiao, G. C. Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association 89*, (1994), 913-923.

[18] Kengne W. Testing for parameter constancy in general causal time-series models. *J. Time Ser. Anal. 33*, (2012), 503-518.

[19] Kim, S., Cho, S. and Lee, S. On the CUSUM test for parameter changes in GARCH(1, 1) models. *Comm. Statist. Theory Methods 29*, (2000), 445-462.

[20] Kokoszka, P. and R. Leipus Testing for parameter changes in ARCH models. *Lithuanian Mathematical Journal 39*, (1999), 182-195.

[21] Kulperger R. and Yu, H. High moment partial sum processes of residuals in GARCH models and their applications. *Ann. Statist. 33*, (2005), 2395-2422.

[22] Lee, S., Song, J. Test for parameter change in ARMA models with GARCH innovations. *Statistics & Probability Letters 78*, (2008), 1990–1998.

[23] Leisch, F., Hornik, K., and Kuan, C-H. Monitoring structural changes with the generalized fluctuation test. *Econometric Theory 16*, (2000), 835-854.

[24] Moustakides, G. V. Sequential change detection revisited. *Ann. Statist. 36(2)*, (2008), 787-807.

[25] Na, O. , Lee, Y. and Lee, S. Monitoring parameter change in time series models. *Stat Methods Appl 20*, (2011), 171-199.

[26] Nelson, D. B. and Cao, C. Q. Inequality Constraints in the Univariate GARCH Model. *Journal of Business & Economic Statistics 10*, (1992), 229-235.

[27] Rabemananjara, R. and Zakoïan, J.M. Threshold ARCH models and asymmetries in volatility. *Journal of Applied Econometrics 8*, (1993), 31-49.

[28] Robinson, P. M. Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *J. Econometrics 47*, (1991), 67-84.

[29] Zeileis, A., Leisch F., Kleiber, C., and Hornik, K. Monitoring structural change in dynamic econometric models. *J. Appl. Econometrics 20*, (2005), 99-121.