Monitoring procedure for parameter change in causal time series
Jean-Marc Bardet, William Charky Kengne

To cite this version:

HAL Id: hal-00734210
https://hal.science/hal-00734210v2
Submitted on 21 Feb 2013

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

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Monitoring procedure for parameter change in causal time series

Jean-Marc BARDET and William KENGNE *

February 21, 2013

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(\( \infty \)), ARCH(\( \infty \)), TARCH(\( \infty \)), 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 a 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 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 [16], Inclan and Tiao [19], Kokoszka and Leipus [22], Kim et al. [21], Horváth and Shao [18], Aue et al. [3] or Kengne [20]. 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 the sequential change-point problem. For instance, consider the following sequential problems.

Example 1.1 (Industrial quality control). Consider an industrial system producing electronic objects. To know in real time the quality of production, some devices have been installed to rely the informations about the functioning. The performance is evaluated for each production and the system is automatically stopped if a disorder is detected. After an investigation of the root cause of the problem and a possible maintenance, the system is resetting.

*Supported by Université Paris 1 and AUF (Agence Universitaire de la Francophonie).
Example 1.2 (Monitoring in economic system). Consider an economic system described by the regression model $Y_t = X'_t \beta_0 + \xi_t$, $t = 1, 2, \cdots n$ where $Y_t$ is the variable of interest and $X_t$ the vector of explanatory variables. New data arrive steadily at time $t = n + 1, \cdots$; Chu et al. (1996) asked: “is yesterday’s model capable of explaining today’s data?” This question leads to construct a procedure which will test sequentially if the model with parameter $\beta_0$ is able to explain new data. If at time $k > n$ a change from $\beta_0$ to $\beta_1$ is detected, the economic system will now use the model $Y_t = X'_t \beta_1 + \xi_t$, $t = k + 1, \cdots$ and monitoring will continue to ensure that this new model can explain the new data that will arrive.

The sequential change-point problem presented in Example 1.1 is often designed by the terminology "fault detection control" or "engineering process control". The aim of the control scheme is to trigged an alarm and stop the system when disorder is detected. Since the diagnostic, the resetting and the malfunctioning can have a cost, the scheme will be able to reduce as soon as possible the false alarm and the delay of detecting disorder. For more review of this approach, see Pollak and Siegmund (1991), Basseville and Nikiforov (1993), Lai (1998), Lai and Shan (1999), Mei (2006), Moustakides (2008) · · ·

The problem presented in Example 1.2 is designed by the terminology "monitoring scheme" (see for instance Chu et al. [13]) ; the term "on-line segmentation" or "sequential segmentation" is often used in application in signal processing (see for instance Basseville and Nikiforov [6] p. 2). These terminologies refer to the sequential procedure where false alarms are less crucial (than in the previous example, because here, the system will deal with in the next stage); in such case, it is important to estimate the model described by the system after change.

This paper focus on such sequential change-point problem and we will follow the paradigm of Chu et al. [13], Berkes et al. (2004), Horváth et al. (2004), Zeileis et al. (2005), Aue et al. (2009), Na et al. (2011) · · ·, who have seen this problem as a classical hypothesis testing with a fixed probability of type I error.

An important turning on this topic was made with the paper of Chu et al.. They considered the sequential change in regression model and pointed out the effects of repeating retrospective test when new data are observed; this can increase the probability of type I 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. [27] introduced the generalized fluctuation test based on the recursive moving estimator which contains the test of Chu et al. [13] as a special case. Horváth et al. [17] 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. [7] considered sequential changes in the parameters of GARCH process. According to the fact that the functional limit theorem assumed by Chu et al. [13] is not satisfied by the squares of residuals of GARCH process, they developed a procedure based on quasi-likelihood scores. Na et al. [30] developed a monitoring procedure for 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 $M, f : \Re^N \rightarrow \Re$ be measurable functions, $(\xi_t)_{t \in \Z}$ be a sequence of centered independent and identically distributed (iid) random variables satisfying $\text{var}(\xi_0) = \sigma^2$ and let $\Theta$ be a fixed compact subset of
$\mathbb{R}^d$. Let $T \subset \mathbb{Z}$, and for any $\theta \in \Theta$, define

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

$$X_{t+1} = M_\theta((X_{t-i})_{i \in \mathbb{N}}) + f_\theta((X_{t-i})_{i \in \mathbb{N}}) \quad \text{for all } t \in T.$$  \hspace{1cm} (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($\infty$), ARCH($\infty$), TARCH($\infty$), ARMA-GARCH or bilinear processes) are included in $M_{\mathbb{Z}}(M, f)$. The off-line change detection for such class of models has already been studied in Bardet et al. [5] and Kengne [20].

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

**$H_0$**: $\theta_0^*$ 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 $M_{\mathbb{N}}(\theta_0^*, f_{\theta_0^*})$;

**$H_1$**: there exist $k^* > n$, $\theta_k^* \in \Theta$ such that $X_1, \cdots, X_n, X_{n+1}, \cdots, X_{k^*}, X_{k^*+1}, \cdots$ belongs to $M_{\{1, \ldots, k^*\}}(\theta_0^*, f_{\theta_0^*}) \cap M_{\{k^*+1, \ldots\}}(\theta_k^*, f_{\theta_k^*})$.

When new data arrive, Chu et al. [13] 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 new parameter is considered; otherwise, the monitoring scheme continues. In their procedure, Leisch et al. [27] 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. [13], the recursive estimators computed by Na et al. [30] are based on all the observations. As we will see in the next sections, their procedure cannot be effective in terms of detection delay in different frames. An application to financial data (Nikkei 225,
S&P 500 and FTSE 100 stock index) allows to detect changes in these data in accordance with historical and economic events.

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 famous financial data. The proofs of the main results are provided in Section 6.

2 Assumptions and definition of the quasi-likelihood estimator

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

Let $\theta \in \mathbb{R}^d$ and $M_\theta$ and $f_\theta$ be numerical functions such that for all $(x_i)_{i \in \mathbb{N}} \in \mathbb{R}^{dN}$, $M_\theta((x_i)_{i \in \mathbb{N}}) \neq 0$ and $f_\theta((x_i)_{i \in \mathbb{N}}) \in \mathbb{R}$. 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 $K \subseteq \mathbb{R}^d$ and for any $g : K \rightarrow \mathbb{R}^d$, $\|g\|_K = \sup_{\theta \in K}(\|g(\theta)\|)$;
3. for any set $K \subseteq \mathbb{R}^d$, $\mathcal{K}$ denotes the interior of $K$.

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

Assumption A$_1(\Psi_\theta, K)$: Assume that $\|\partial^i \Psi_\theta(0) / \partial \theta^i\|_\Theta < \infty$ and there exists a sequence of non-negative real numbers $(\alpha^{(i)}_j(\Psi_\theta, K))_{j \geq 1}$ such that $\sum_{j=1}^{\infty} \alpha^{(i)}_j(\Psi_\theta, | |) < \infty$ and

$$\| \partial^1 \Psi_\theta(x) - \partial^1 \Psi_\theta(y) \|_K \leq \sum_{j=1}^{\infty} \alpha^{(i)}_j(\Psi_\theta, K)|x_j - y_j|$$

for all $x, y \in \mathbb{R}^{dN}$.

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 A$_1(h_\theta, K)$: Assume that $\|\partial^i h_\theta(0) / \partial \theta^i\|_\Theta < \infty$ and there exists a sequence of non-negative real numbers $(\alpha^{(i)}_j(h_\theta, K))_{j \geq 1}$ such that $\sum_{j=1}^{\infty} \alpha^{(i)}_j(h_\theta, K) < \infty$ and

$$\| \partial^1 h_\theta(x) - \partial^1 h_\theta(y) \|_K \leq \sum_{j=1}^{\infty} \alpha^{(i)}_j(h_\theta, K)|x_j^2 - y_j^2|$$

for all $x, y \in \mathbb{R}^{dN}$.

The Lipschitz-type hypothesis A$_1(\Psi_\theta, K)$ are classical when studying the existence of solutions of the general model (see for instance [14]). Using a result of [4], for each model $\mathcal{M}_Z(M_\theta, f_\theta)$ it is interesting to define the following set:

$$\Theta(r) := \left\{ \theta \in \Theta, A_0(f_\theta, \{\theta\}) \text{ and } A_0(M_\theta, \{\theta\}) \text{ hold with } \sum_{j=1}^{\infty} \alpha^{(0)}_j(f_\theta, \{\theta\}) + (E |\xi_0|^r)^{1/r} \sum_{j=1}^{\infty} \alpha^{(0)}_j(M_\theta, \{\theta\}) < 1 \right\}$$

$$\cup \left\{ \theta \in \Theta, f_\theta = 0 \text{ and } A_0(h_\theta, \{\theta\}) \text{ holds with } (E |\xi_0|^r)^{2/r} \sum_{j=1}^{\infty} \alpha^{(0)}_j(h_\theta, \{\theta\}) < 1 \right\}.$$
Then, if $\theta \in \Theta(r)$ the existence of a unique causal, stationary and ergodic solution $X = (X_t)_{t \in \mathbb{Z}} \in \mathcal{M}_2(f_\theta, M_\theta)$ is ensured (see more details in [4]). The subset $\Theta(r)$ is defined as a union to consider accurately general causal models and ARCH-type models simultaneously.

Here there are assumptions required for studying QMLE asymptotic properties:

**Assumption D(\theta):** $\exists \Delta_0 > 0$ such that $\inf_{\theta \in \Theta} \{\phi_\theta(x)\} > \Delta_0$ for all $x \in \mathbb{R}^N$.

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

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

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

**Assumption K(f, M_\theta, \Theta):** for $i = 0$, $1$, $2$, $A_i(f_\theta, \Theta)$ and $A_i(M_\theta, \Theta)$ (or $A_i(h_\theta, \Theta)$) 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 $A_i(M_\theta, \Theta)$ holds then $\alpha_{j}^{(i)}(h_\theta, \Theta) = 0$ and if $A_i(h_\theta, \Theta)$ holds then $\alpha_{j}^{(i)}(M_\theta, \Theta) = 0$.

### 2.2 Two first examples

1. **ARMA(p, q) processes.**
   
   Consider the ARMA(p, q) process defined by:

   $$
   X_t + \sum_{i=1}^{p} a_i^* X_{t-i} = \sum_{j=0}^{q} b_j^* \xi_{t-j} , \quad t \in \mathbb{Z} \tag{2}
   $$

   with $b_0^* \neq 0$, $b_0^* = (a_1^*, \ldots, a_p^*, b_0^*, \ldots, b_q^*) \in \Theta \subset \mathbb{R}^{p+q+1}$ and $(\xi_t)$ a white noise such as $E(\xi_t^2) = 1$. When $\sum_{j=0}^{q} b_j^* \xi^j \neq 0$ and $1 + \sum_{i=0}^{p} a_i^* X^i \neq 0$ for all $|X| \leq 1$, this process can be also written as:

   $$
   X_t = b_0^* \xi_t + \sum_{j=1}^{\infty} \phi_j(b_0^*) X_{t-j} , \quad t \in \mathbb{Z}
   $$

   where $\theta \in \Theta \mapsto \phi_j(\theta)$ (functions only depending on $\theta$ and decreasing exponentially fast to 0 ($j \to \infty$). The process (2) belongs to the class $\mathcal{M}_2(M_{\theta}^{b_0^*}, f_{\theta}^{b_0^*})$ where $f_{\theta}(x_1, \cdots) = \sum_{j \geq 1} \phi_j(\theta)x_j$ and $M_\theta \equiv b_0^*$ for all $\theta \in \Theta$. Then Assumptions D(\theta), A(f_\theta, \Theta), A(M_\theta, \Theta) hold with $\Delta = |b_0^*| > 0$ and $\alpha_j^{(i)}(f_\theta, \Theta) = \|\phi_j(\theta)\|_{\Theta}$ while $\alpha_j^{(i)}(M_\theta, \Theta) = 0$ for $j \in \mathbb{N}^*$. Assumption K(f_\theta, M_\theta, \Theta) holds since there exists $c > 0$ and $C > 0$ such that $|\phi_j| \leq C e^{-cj}$ for $j \in \mathbb{N}$. Moreover, if $(\xi_t)$ is a sequence of non-degenerate random variables (i.e. $\xi_t$ is not equal to a constant), Assumptions Id(\theta) and Var(\theta) hold. Finally, for any $r \geq 1$ such that $E|\xi_0|^r < \infty$, then

   $$
   \Theta(r) = \left\{ \theta \in \mathbb{R}^{p+q+1}, \sum_{j \geq 1} |\phi_j(\theta)| < 1 \right\}
   $$

   5
Note that if $\theta \in \Theta(r)$ with $r \geq 1$ then the previous conditions of stationarity $\sum_{j=0}^{q} a_j X_j \neq 0$ and $1 + \sum_{j=0}^{p} a_j X_j \neq 0$ for all $|X| \leq 1$ are satisfied.

2. GARCH($p, q$) processes.

Consider the GARCH($p, q$) process defined by:

$$X_t = \sigma_t \xi_t, \quad \sigma_t^2 = \alpha_0^* + \sum_{j=1}^{p} \alpha_j^* X_{t-j}^2 + \sum_{j=1}^{q} \beta_j^* \sigma_{t-j}^2, \quad t \in \mathbb{Z}$$

(3)

with $E(\xi_t^2) = 1$ and $\theta_0^* := (\alpha_0^*, \ldots, \alpha_p^*, \beta_1^*, \ldots, \beta_q^*) \in \Theta$ where $\Theta$ 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 [11] or Nelson and Cao [31]) a nonnegative sequence $(\psi_j(\theta_0^*))_{j \geq 0}$ such that $\sigma_t^2 = \psi_0(\theta_0^*) + \sum_{j \geq 1} \psi_j(\theta_0^*) X_{t-j}$ with $\psi_0(\theta_0^*) = \alpha_0^*/(1 - \sum_{j=1}^{q} \beta_j^*)$.

This process belongs to the class $M_{\mathbb{Z}}(M_{\theta_0^*}, f_{\theta_0})$ where $f_{\theta} \equiv 0$ and $M_{\theta}(x_1, \ldots) = \sqrt{\psi_0(\theta) + \sum_{j \geq 1} \psi_j(\theta) x_j^2}$ for all $\theta \in \Theta$. Assumption $D(\Theta)$ holds with $L = \inf_{\theta \in \Theta}(\psi_0(\theta)) > 0$. If there exists $0 < \rho_0 < 1$ such that for any $\theta \in \Theta$, $\sum_{j=1}^{q} \alpha_j + \sum_{j=1}^{q} \beta_j \leq \rho_0$ 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. [8]) and Assumption $K(f_{\theta}, M_{\theta}, \Theta)$ holds. Moreover, $(\xi_t^2)$ is a sequence of non-degenerate random variables (i.e. $\xi_t^2$ is not equal to a constant), Assumptions $\text{Id}(\Theta)$ and $\text{Var}(\Theta)$ hold. Finally for $r \geq 2$ we obtain

$$\Theta(r) = \left\{ \theta \in \Theta : (E(\xi_0^r))^2/r \sum_{j=1}^{\infty} \phi_j(\theta) < 1 \right\}.$$  

2.3 The quasi-maximum likelihood estimator

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

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

(4)

where $f_{\theta}^2 = f_{\theta}(X_{t-1}, X_{t-2}, \ldots)$, $M_{\theta}^2 = M_{\theta}(X_{t-1}, X_{t-2}, \ldots)$ and $h_{\theta}^2 = M_{\theta}^2$. The classical approximation of this conditional log-likelihood (see more details in Bardet and Wintenberger [4]) is given by:

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

(5)

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

$$\hat{\theta}(T) := \arg\max_{\theta \in \Theta}(\hat{L}(T, \theta)).$$

(6)

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

$$\sqrt{n}(\hat{\theta}(T_{1,n}) - \theta_0^*) \xrightarrow{D} N(0, F G^{-1} F),$$

(7)
with
\[ G := E \left[ \frac{\partial^2 q_0(\theta_0^\prime)}{\partial \theta \partial \theta^\prime} \right] \quad \text{and} \quad F := E \left[ \frac{\partial^2 q_0(\theta_0^\prime)}{\partial \theta \partial \theta^\prime} \right], \tag{8} \]
where $'$ denotes the transpose and with $q_0$ defined in (4). Note that under assumptions $D(\Theta)$ and $\text{Var}(\Theta)$, $G$ is symmetric positive definite (see [20]) and $F$ is non-singular (see [4]). Also define the matrix
\[ \hat{G}(T) := \frac{1}{\text{Card}(T)} \sum_{i \in T} \left( \frac{\partial \hat{q}(\hat{\theta}(T))}{\partial \theta} \right) \left( \frac{\partial \hat{q}(\hat{\theta}(T))}{\partial \theta} \right)^\prime \quad \text{and} \quad \hat{F}(T) := -\frac{2}{\text{Card}(T)} \left( \frac{\partial^2 \hat{L}_n(T, \hat{\theta}(T))}{\partial \theta \partial \theta^\prime} \right). \tag{9} \]
Under the previous assumptions, $\hat{G}(T_{1,n})$ and $\hat{F}(T_{1,n})$ converge almost surely to $G$ and $F$ respectively. Hence,
\[ \sqrt{n} \hat{G}(T_{1,n})^{-1/2} \hat{F}(T_{1,n}) \left( \hat{\theta}(T_{1,n}) - \theta_0^* \right) \xrightarrow{\mathcal{D}} N(0, I_d) \tag{10} \]
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 procedure evaluates the difference between $\hat{\theta}(T_{\ell,k})$ and $\hat{\theta}(T_{1,n})$ for any $\ell = n, \cdots, k$. More precisely, for any $k > n$ define the statistic (called the detector)
\[ \hat{C}_{k,\ell} := \sqrt{n} \frac{k - \ell}{k} \| \hat{G}(T_{1,n})^{-1/2} \hat{F}(T_{1,n}) \left( \hat{\theta}(T_{\ell,k}) - \hat{\theta}(T_{1,n}) \right) \| \]
for $\ell = n, \cdots, k$. Since the matrix $\hat{G}(T_{1,n})$ is asymptotically symmetric and positive definite (see [20]), $\hat{G}(T_{1,n})^{-1/2}$ exists for $n$ large enough and $\hat{C}_{k,\ell}$ is well defined. At the beginning of the monitoring scheme and when $\ell$ is close to $k$, the length of $T_{\ell,k}$ is too small, therefore the numerical algorithm used to compute $\hat{\theta}(T_{\ell,k})$ 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 $\hat{C}_{k,\ell}$ 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
\[ v_n \to \infty \quad \text{and} \quad v_n/\sqrt{n} \to 0 \quad (n \to \infty). \]
According to Remark 1 of [20], 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 $\hat{\theta}(T_{\ell,k})$ and $\hat{\theta}(T_{1,n})$ are close and the detector $\hat{C}_{k,\ell}$ is not too large. Hence, the monitoring scheme rejects $H_0$ at the first time $k > n$ where there exists $\ell \in \Pi_{n,k}$ satisfying $\hat{C}_{k,\ell} > c$ 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 \( \inf_{0 < t < \infty} b(t) > 0. \)

Then the monitoring scheme rejects \( H_0 \) at the first time \( k > n \) such as there exists \( \ell \in \Gamma_{n,k} \) satisfying \( \hat{C}_{k,\ell} > b((k - \ell)/n) \). Hence define the stopping time:

\[
\tau(n) := \inf\{ k > n / \exists \ell \in \Gamma_{n,k}, \hat{C}_{k,\ell} > b((k - \ell)/n) \} = \inf\{ k > n / \max_{\ell \in \Gamma_{n,k}} \frac{\hat{C}_{k,\ell}}{b((k - \ell)/n)} > 1 \}.
\]

Therefore, we have

\[
P\{\tau(n) < \infty\} = P\{ \max_{\ell \in \Gamma_{n,k}} \frac{\hat{C}_{k,\ell}}{b((k - \ell)/n)} > 1 \text{ for some } k > n \} = P\{ \sup_{k>n} \max_{\ell \in \Gamma_{n,k}} \frac{\hat{C}_{k,\ell}}{b((k - \ell)/n)} > 1 \}.
\]

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 \( c = c_0 \) depending on \( \alpha \). If change is detected under \( H_1 \) i.e. \( \tau(n) < \infty \) and \( \tau(n) > k^* \), then the detection delay is defined by

\[
\hat{d}_n = \tau(n) - k^*.
\]

Using the previous notations, Na et al. [30] used the following detector

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

At the step \( k \) of the monitoring scheme, their recursive estimator is based on \( X_1, \ldots, X_n, \ldots, X_k \). One can see that this estimator is highly influenced by the historical data. Assume that a change occurs at time \( k^* \leq k \), in the sequel of the procedure, the recursive estimator contains the observations \( X_1, \ldots, X_n, \ldots, X_{k-1} \) which depends on \( \theta_0^* \). Then, one must wait longer before the difference between \( \hat{\theta}(X_1, \ldots, X_n) \) and \( \hat{\theta}(X_1, \ldots, X_n, \ldots, X_k) \) 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 \( u \) is equal to 0 if and only if \( u \) 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 \( \theta_0^* \) does not change over the new observations. Thus we have the result

**Theorem 3.1.** Assume \( D(\Theta), \ Id(\Theta), \ Var(\Theta), \ K(f_0, M_0, \Theta), \ \text{and} \ \theta_0^* \in \Theta(4). \) Under the null hypothesis \( H_0 \), then

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

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,

\[
\lim_{n \to \infty} P\{\tau(n) < \infty\} = P\left( \sup_{t \geq 1} \sup_{1 < s < t} \frac{1}{t} \|W_d(s) - sW_d(1)\| > c \right) = P\{U_d > c\}
\]

where \( U_d = \sup_{0 < u < 1} f(u) \|W_d(u)\| \) with \( f(u) = \frac{\sqrt{9 - u + \sqrt{1 - u}}}{\sqrt{9 - u + 3\sqrt{1 - u}} \left( \frac{2}{3 - u + (9 - u)(1 - u)} \right)^{1/2}} \).

**Remark 3.1.**

1. By the law of the iterated logarithm, it comes that

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

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

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

2. Under the null hypothesis, it holds that \( \hat{\theta}(T_{1,n}) \xrightarrow{a.s.} \theta_0^n \) (see [4]). Thus denote

\[
\hat{C}_{k,\ell}^{(0)} := \sqrt{n} \frac{k - \ell}{k} \|\hat{G}(T_{1,n})^{-1/2}\hat{F}(T_{1,n})(\hat{\theta}(T_{k,\ell}) - \theta_0^n)\|.
\]

Under the assumptions of Theorem 3.1, one can easily show that

\[
\sup_{k > n} \max_{\ell \in \Pi_{n,k}} \frac{1}{b((k - \ell)/n)} \left| \hat{C}_{k,\ell} - \hat{C}_{k,\ell}^{(0)} \right| = o_P(1) \quad \text{as} \quad n \to \infty.
\]

Thus, the Theorem 3.1 still holds if the stopping time \( \tau(n) \) is computed by using the detector \( \hat{C}_{k,\ell}^{(0)} \). Hence, if the parameter \( \theta_0^n \) of the historical observations is known, then use the detector \( \hat{C}_{k,\ell}^{(0)} \) instead of \( \hat{C}_{k,\ell} \). But let us note that this situation is infrequent in practice.

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 \( d = 1, \cdots, 5 \).

<table>
<thead>
<tr>
<th>( d )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha = 0.01 )</td>
<td>2.583</td>
<td>3.035</td>
<td>3.335</td>
<td>3.631</td>
<td>3.914</td>
</tr>
<tr>
<td>( \alpha = 0.05 )</td>
<td>1.954</td>
<td>2.432</td>
<td>2.760</td>
<td>3.073</td>
<td>3.334</td>
</tr>
<tr>
<td>( \alpha = 0.10 )</td>
<td>1.652</td>
<td>2.156</td>
<td>2.486</td>
<td>2.784</td>
<td>3.028</td>
</tr>
</tbody>
</table>

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 $\theta_0^* \to \theta_1^*$ at $k^* > n$, where $\theta_1^* \in \Theta$ and $\theta_0^* \neq \theta_1^*$. Then

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

$$\max_{\ell \in \mathbb{N}_{n,k}} \frac{\hat{C}_{kn, \ell}}{b((k_n - \ell)/n)} \xrightarrow{a.s.} \infty.$$

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 scheme rejects $H_0$ at the first time $k$ where

$$\max_{\ell \in \mathbb{N}_{n,k}} \frac{\hat{C}_{kn, \ell}}{b((k_n - \ell)/n)} > 1.$$

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

3.4 Examples

3.4.1 AR(\infty) processes

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

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

(12)

with $\theta_0^* \in \hat{\Theta}$, where we can chose $\Theta$ as a compact subset of $\Theta(4) \subset \mathbb{R}^d$ where

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

This process belongs to the class $\mathcal{M}_\mathbb{Z}(M_{\theta^*}, \beta_{\theta^*})$ 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 $j \in \mathbb{N}^*$. Then

- Assumption $D(\Theta)$ holds if $\mathcal{H} = \inf_{\theta \in \Theta}(|\phi_0(\theta)|) > 0$;
- Assumption $K(f_\theta, M_\theta, \Theta)$ holds if there exists $\ell > 2$ and and if $\theta \mapsto \phi_j(\theta)$ are twice differentiable functions satisfying max $(\|\psi_j(\theta)\|_\Theta, \|\phi_j'(\theta)\|_\Theta, \|\phi_j''(\theta)\|_\Theta) = O(j^{-\ell})$ for $j \in \mathbb{N}$.
- if $(\xi_t)$ is a sequence of non-degenerate random variables, Assumptions $Id(\Theta)$ and $\text{Var}(\Theta)$ hold.

Case of AR($p$) process

Assume that

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

### 3.4.2 ARCH(\( \infty \)) processes

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

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

(13) with \( \theta_0^* \in \Theta \), where we can choose \( \Theta \) as a compact subset of \( \Theta(4) \subset \mathbb{R}^d \) where

\[
\Theta(4) = \{ \theta \in \mathbb{R}^d; (E|\xi_0|^4)^{1/2} \sum_{j=1}^{\infty} |\phi_j(\theta)| < 1 \}. \]

This process, introduced by Robinson [34], belongs to the class \( \mathcal{M}_\mathbb{Z}(f_{\theta_0^*}, M_{\theta^*_0}) \) where \( f_0(x_1, \cdots) \equiv 0 \) and \( M_j(x_1, \cdots) = \psi_0(\theta) + \sum_{j \geq 1} \psi_j(\theta)x_j^2 \) for all \( \theta \in \Theta \) and therefore \( \alpha_j^{(0)}(f_0, \Theta) = 0 \) and \( \alpha_j^{(0)}(h_0, \Theta) = \|\phi_j(\theta)\|_\Theta \) for \( j \in \mathbb{N}^* \) (\( X \) is of course a ARCH-type process). Then

- Assumption \( D(\Theta) \) holds if \( \hat{\lambda} = \inf_{\theta \in \Theta} (\psi_0(\theta)) > 0 \);
- Assumption \( K(f_0, M_0, \Theta) \) holds if there exists \( \ell > 2 \) and if for \( \theta \mapsto \phi_j(\theta) \) are twice differentiable functions satisfying \( \max \left( \|\psi_j(\theta)\|_\Theta, \|\psi_j'(\theta)\|_\Theta, \|\psi_j''(\theta)\|_\Theta \right) = O(j^{-\ell}) \) for \( j \in \mathbb{N} \).
- if \((\xi_t^2)\) is a sequence of non-degenerate random variables, Assumptions \( \text{Id}(\Theta) \) and \( \text{Var}(\Theta) \) hold.

#### Case of GARCH\((1,1)\) process

Assume that

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

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

\[
\sigma_t^2 = \alpha_0^*/(1 - \beta_1^*) + \alpha_1^* \sum_{j=1}^{\infty} \beta_1^{*j-1} X_{t-j}^2
\]

If a trajectory \((X_1, \cdots, X_k)\) has been observed, for any \( t = 1, \cdots, k \) and \( \theta \in \Theta \) we have,

\[
\hat{\theta}_t(\theta) = \alpha_0/(1 - \beta_1) + \alpha_1 X_{t-1}^2 + \alpha_1 \sum_{j=2}^{t} \beta_1^{j-1} X_{t-j}^2 \quad \text{and} \quad \hat{\theta}_t(\theta) = X_t^2 / \hat{\theta}_t^t + \log(\hat{\theta}_t^t).
\]

Thus, it follows that

\[
\frac{\partial \hat{\theta}_t(\theta)}{\partial \theta} = \frac{1}{\hat{\theta}_t^t} \left( 1 - \frac{X_t^2}{\hat{\theta}_t} \right) \left( \frac{\partial \hat{\theta}_t^t}{\partial \alpha_0} \frac{\partial \hat{\theta}_t}{\partial \alpha_1} \frac{\partial \hat{\theta}_t}{\partial \beta_1} \right) \quad \text{with} \quad \frac{\partial \hat{\theta}_t^t}{\partial \alpha_1} = X_{t-1}^2 + \sum_{j=2}^{t} \beta_1^{j-1} X_{t-j}^2 \quad \text{and} \quad \frac{\partial \hat{\theta}_t}{\partial \beta_1} = \frac{1}{1 - \beta_1}.
\]

And

\[
\frac{\partial \hat{\theta}_t}{\partial \beta_1} = \alpha_0/(1 - \beta_1)^2 + \alpha_1 X_{t-2}^2 + \alpha_1 \sum_{j=3}^{t} (j-1) \beta_1^{j-2} X_{t-j}^2
\]

\(11\)}
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

\[
\frac{\partial^2 \hat{h}_t(\theta)}{\partial \theta_i \partial \theta_j} = \frac{1}{(h_0')^2} \left( 2X_t^2 - 1 \right) \frac{\hat{h}_t(\theta)}{\partial \theta_i} \frac{\hat{h}_t(\theta)}{\partial \theta_j} + \frac{1}{h_0'} \left( 1 - \frac{X_t^2}{h_0'} \right) \frac{\partial^2 \hat{h}_0(\theta)}{\partial \theta_i \partial \theta_j}
\]

with \( \frac{\partial^2 \hat{h}_0(\theta)}{\partial \alpha_0^2} = 0 \), \( \frac{\partial^2 \hat{h}_0(\theta)}{\partial \alpha_1^2} = 0 \), \( \frac{\partial^2 \hat{h}_0(\theta)}{\partial \alpha_1 \partial \beta_1} = 0 \), \( \frac{\partial^2 \hat{h}_0(\theta)}{\partial \alpha_1 \partial \beta_1} = X_t^2 + \sum_{j=3}^{t} (j-1)\beta_1^{j-2}X_{t-j}^2 \), \( \frac{\partial^2 \hat{h}_0(\theta)}{\partial \alpha_0 \partial \beta_1} = 1/(1 - \beta_1)^2 \) and

\[
\frac{\partial^2 \hat{h}_0(\theta)}{\partial \beta_1^2} = 2\alpha_0/(1 - \beta_1)^3 + 2\alpha_1X_t^2 + \alpha_1 \sum_{j=4}^{t} (j-1)(j-2)\beta_1^{j-3}X_{t-j}^2.
\]

### 3.4.3 TARCH(\( \infty \)) processes

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

\[
X_t = \sigma_t \xi_t \quad \text{and} \quad \sigma_t = b_0(\Theta_t) + \sum_{j=1}^{\infty} \left[ b_j^+(\Theta_t) \max(X_{t-j}, 0) - b_j^- (\Theta_t) \min(X_{t-j}, 0) \right], \quad t \in \mathbb{Z}
\]

(14)

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

\[
\Theta(4) = \left\{ \theta \in \mathbb{R}^l \cap (E |\xi|)^{1/4} \sum_{j=1}^{\infty} \max(b_j^-(\theta), b_j^+(\theta)) < 1 \right\}
\]

since \( \alpha_j^0(M, \{\theta\}) = \max(b_j^-(\theta), b_j^+(\theta)) \). This class of processes is a generalization of the class of TGARCH(\( p,q \)) processes (introduced by Rabemananjara and Zakoian [33]). Then,

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

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

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

### 4 Some simulation and numerical experiments

First remark that, at a time \( k > n \), we need to compute \( \hat{C}_{k,t} \) for all \( \ell \in \Pi_{n,k}^0 \) 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 \( u_n = \lfloor \ln(n) \rfloor \)) and compute \( \hat{C}_{k,t} \) only for

\[
\ell \in \Pi_{n,k}^0 := \{ n-v_n, n-v_n+u_n, n-v_n+2u_n, \ldots, k-v_n \}.
\]

We have \( \Pi_{n,k}^0 \subset \Pi_{n,k} \) and for any \( t = \frac{\ell}{n} \) with \( \ell \in \Pi_{n,k} \), we can find an integer \( j_t \) such that \( n-v_n+j_t u_n \in \Pi_{n,k}^0 \) and \( n-v_n+j_t u_n \leq \ell \leq n-v_n+(j_t+1)u_n \). This implies that \( \frac{n-v_n+j_t u_n}{n} \leq t \leq \frac{n-v_n+(j_t+1)u_n}{n} \). Thus, we
have asymptotically (as \( n \to \infty \)), \( t \sim \frac{n-u_n+250}{\sqrt{n}} \). It shows that the previous asymptotic results still hold by computing \( \hat{C}_{k,\ell} \) for \( \ell \in \Pi_{n,k}^0 \). The condition \( u_n/\sqrt{n} \to 0 \) ensures that the Theorem 3.2 still holds by choosing \( k_n = k^*(n) \) with \( \delta \in (1/2, 1) \). In practice, the use of \( \Pi_{n,k}^0 \) can introduce a distortion in the detection delay. But, the new detection delay must be between \( \hat{d_n} \) and \( d_n + u_n \) (where \( \hat{d_n} \) 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

\[
P\{\tau(n) < \infty\} = P\left\{ \sup_{k>n} \max_{\ell \in \Pi_{n,k}^0} \hat{C}_{k,\ell} > c \right\}.
\]

Thus, denote

\[
\hat{C}_k = \max_{\ell \in \Pi_{n,k}^0} \hat{C}_{k,\ell} \quad \text{for any} \quad k > n.
\]

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 [20], \( v_n = [(\log n)^{\delta}] \) (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 \( \alpha = 0.05 \).

### 4.1 An illustration

We consider a 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 the 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 \( \hat{C}_{k,500-k \leq 1000} \). 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 \( \hat{C}_k \) is under the horizontal line which represents the limit of the critical region. On Figure 1 b-) we can see that, before change occurs, \( \hat{C}_k \) is 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 \( X = (X_t)_{t \in \mathbb{Z}} \). We assume that \( X \) satisfy

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

with \( k^* > n \), \( \mu_0 \neq \mu_1 \) and \( \epsilon_t \) a zero mean stationary time series belongs to a class \( \mathcal{MZ}(f_0, M_0) \). Under \( H_0 \), \( k^* = \infty \). The monitoring procedure start at \( k = n + 1 \) and the aim is to test mean shift over the new observations \( X_{n+1}, X_{n+2}, \cdots \).

This problem can be seen as monitoring changes in linear model (see Horváth et al. [17], Aue et al. [2]) with constant regressor. The empirical mean \( \overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_t \) is a consistent estimator of \( \mu_0 \) and the recursive residuals are defined by \( \hat{\epsilon}_k = X_k - \overline{X}_n \); for \( k > n \). Horváth et al. [17] and Aue et al. [2] proposed the
Figure 1: Typical realization of the statistics $\hat{C}_k$ for GARCH(1,1), $n = 500$ and $k = 501, \ldots , 1000$. 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.
CUSUM detector

\[ \hat{Q}_k = \frac{1}{a_n c \sqrt{n}} \frac{1}{(1 - \frac{n}{k})^\gamma} \left| \sum_{i=n+1}^k \hat{e}_i \right| \quad k > n, \ c > 0, \ 0 \leq \gamma < 1/2, \]  

where \( \hat{\sigma}_n^2 \) is a consistent estimator of the long-run variance \( \sigma^2 = \lim_{n \to \infty} \frac{1}{n} \text{Var}(\sum_{i=1}^n \epsilon_i) \). If the process \( \epsilon_t \) are uncorrelated (for instance GARCH-type model), empirical variance of the historical data can be used as estimator of \( \sigma^2 \). If \( \epsilon_t \) are correlated, the popular Bartlett estimator (see [9]) can be used. Under some regular conditions, it hold that (see [17] and [2])

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

Hence, at a nominal level \( \alpha = 0.05 \), the critical value of the test is the \( (1 - \alpha) \)-quantile of the distribution of \( \sup_{0 < s < 1} \frac{|W_1(s)|}{s^\gamma} \). When \( \gamma = 0 \), these quantiles are known (see Table 1 of [30] for values obtained through a Monte Carlo simulation).

We compare our procedure to this CUSUM one (see (16) with \( \gamma = 0 \)) in two situations

1. \( \epsilon_t \) is an AR(1) process; \( \epsilon_t = \phi_1 \epsilon_{t-1} + \xi_t \) with \( \phi_1 = 0.2 \);
2. \( \epsilon_t \) 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 procedures 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 the procedures based on detectors \( \hat{C}_k \) and \( \hat{Q}_k \) are conservative. One can also see that the larger \( n \) (length of historical data) the smaller the distortion size of these procedures. This is due to the fact that the length of monitoring period is fixed and does not increase 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 can be seen in Table 4 that in the case of 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 \( (k^* = n + 50) \); but our procedure is a little more accurate when the change occurs a long time after the beginning of the monitoring \( (k^* = n + 250) \). For the case of the mean shift in GARCH process, our test procedure outperforms the Horváth et al.’s test in terms of mean and quantiles of the detection delay.

4.3 Monitoring parameter changes in AR(1) and GARCH(1,1) processes

In this subsection, we present some simulations results for monitoring parameter changes in AR(1) and GARCH(1,1) models and compare our procedure to the one proposed by Na et al. [30]. If the boundary
<table>
<thead>
<tr>
<th>$k$</th>
<th>$n + 100$</th>
<th>$n + 200$</th>
<th>$n + 300$</th>
<th>$n + 400$</th>
<th>$n + 500$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical levels</td>
<td>$n = 500$</td>
<td>$\hat{C}_k$</td>
<td>0.000</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>$n = 1000$</td>
<td>$\hat{C}_k$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>$n = 500$</td>
<td>$\hat{Q}_k$</td>
<td>0.000</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>$n = 1000$</td>
<td>$\hat{Q}_k$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

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 $\mu_1 = 1.2$.

<table>
<thead>
<tr>
<th>$k$</th>
<th>$n + 100$</th>
<th>$n + 200$</th>
<th>$n + 300$</th>
<th>$n + 400$</th>
<th>$n + 500$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical levels</td>
<td>$n = 500$</td>
<td>$\hat{C}_k$</td>
<td>0.310</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>$n = 500$</td>
<td>$\hat{Q}_k$</td>
<td>0.335</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>$n = 500$</td>
<td>$\hat{C}_k$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.190</td>
</tr>
<tr>
<td></td>
<td>$n = 500$</td>
<td>$\hat{Q}_k$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.130</td>
</tr>
<tr>
<td></td>
<td>$n = 1000$</td>
<td>$\hat{C}_k$</td>
<td>0.075</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>$n = 1000$</td>
<td>$\hat{Q}_k$</td>
<td>0.095</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>$n = 1000$</td>
<td>$\hat{C}_k$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.135</td>
</tr>
<tr>
<td></td>
<td>$n = 1000$</td>
<td>$\hat{Q}_k$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.075</td>
</tr>
</tbody>
</table>

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 $\mu_1 = 0.3$. 

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

Table 4: Elementary statistics of the empirical detection delay for monitoring mean shift in AR(1) and GARCH(1,1).
function \( b(\cdot) \equiv c > 0 \) with a real number \( c > 0 \), Na et al. show that under \( H_0 \),
\[
\lim_{n \to \infty} P\{ \tau(n) < \infty \} = \lim_{n \to \infty} P\left\{ \sup_{k > n} \hat{D}_k > c \right\} = P\left\{ \sup_{0 < s < 1} \|W_d(s)\| > c \right\},
\]
where
\[
\hat{D}_k := \sqrt{n} \| \hat{G}(T_{1,n})^{-1/2} \hat{F}(T_{1,n}) (\hat{\theta}(T_{1,k}) - \hat{\theta}(T_{1,n})) \|.
\]

Hence, at a nominal level \( \alpha \), 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 [30].

The comparisons between their procedure based on \( \hat{D}_k \) and ours based on \( \hat{C}_{k,\ell} \) are made in the following 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 \( H_1 \), \( \theta_0 \) 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 sizes of historical samples are \( n = 500 \) and \( n = 1000 \). The procedures 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.

<table>
<thead>
<tr>
<th></th>
<th>( k = n + 100 )</th>
<th>( k = n + 200 )</th>
<th>( k = n + 300 )</th>
<th>( k = n + 400 )</th>
<th>( k = n + 500 )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Empirical levels</strong></td>
<td>( n = 500 )</td>
<td>( \hat{C}_k )</td>
<td>0.000</td>
<td>0.000</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>( \hat{D}_k )</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>( n = 1000 )</td>
<td>( \hat{C}_k )</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>( \hat{D}_k )</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>Empirical powers</strong></td>
<td>( n = 500 ; k^* = n + 50 )</td>
<td>( \hat{C}_k )</td>
<td>0.335</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>( \hat{D}_k )</td>
<td>0.175</td>
<td>0.985</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>( n = 500 ; k^* = n + 250 )</td>
<td>( \hat{C}_k )</td>
<td>0.000</td>
<td>0.000</td>
<td>0.180</td>
</tr>
<tr>
<td></td>
<td>( \hat{D}_k )</td>
<td>0.000</td>
<td>0.000</td>
<td>0.095</td>
<td>0.865</td>
</tr>
<tr>
<td></td>
<td>( n = 1000 ; k^* = n + 50 )</td>
<td>( \hat{C}_k )</td>
<td>0.065</td>
<td>0.995</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>( \hat{D}_k )</td>
<td>0.090</td>
<td>0.975</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>( n = 1000 ; k^* = n + 250 )</td>
<td>( \hat{C}_k )</td>
<td>0.000</td>
<td>0.000</td>
<td>0.140</td>
</tr>
<tr>
<td></td>
<td>( \hat{D}_k )</td>
<td>0.000</td>
<td>0.000</td>
<td>0.075</td>
<td>0.855</td>
</tr>
</tbody>
</table>

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,
Empirical levels

\[
\begin{array}{ccccccc}
 k & n + 100 & n + 200 & n + 300 & n + 400 & n + 500 \\
\hline
 C_k & 0.010 & 0.025 & 0.040 & 0.095 & 0.105 \\
 D_k & 0.010 & 0.015 & 0.040 & 0.040 & 0.055 \\
\end{array}
\]

Empirical powers

\[
\begin{array}{ccccccc}
 k & n = 500 ; k^* = n + 50 & C_k & 0.890 & 1 & 1 & 1 \\
 D_k & 0.390 & 0.855 & 0.930 & 0.965 & 0.985 \\
\end{array}
\]

\[
\begin{array}{ccccccc}
 k & n = 500 ; k^* = n + 250 & C_k & 0.010 & 0.030 & 0.825 & 1 & 1 \\
 D_k & 0.010 & 0.020 & 0.270 & 0.805 & 0.915 \\
\end{array}
\]

\[
\begin{array}{ccccccc}
 k & n = 1000 ; k^* = n + 50 & C_k & 0.835 & 1 & 1 & 1 & 1 \\
 D_k & 0.310 & 0.970 & 0.990 & 0.995 & 0.995 \\
\end{array}
\]

\[
\begin{array}{ccccccc}
 k & n = 1000 ; k^* = n + 250 & C_k & 0.000 & 0.005 & 0.685 & 1 & 1 \\
 D_k & 0.000 & 0.000 & 0.250 & 0.955 & 0.990 \\
\end{array}
\]

Table 6: Empirical levels and powers for monitoring parameter change in GARCH(1,1) process. The empirical levels are computed when \( \theta_0 = (0.01, 0.3, 0.2) \) 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 \( H_1 \)).

For both the cases of AR and GARCH processes, the procedure based on detector \( \hat{C}_k \) detects the change before the monitoring time \( k = n + 500 \). Unlike Na et al. [30], we consider a scenario of GARCH model with moderate change in parameter and it can be seen in Table 6 that the procedure based on detector \( \hat{D}_k \) 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 is not surprising according to the comment of subsection 3.1.

Table 7 indicates the distribution of the detection delay \( \hat{d}_n \). We can see in Table 7 (even in Table 4) that for our procedure, the relation \( \hat{d}_{1000} \leq \sqrt{1000/500} \hat{d}_{500} \) is globally satisfied (from Theorem 3.2, we deduced that \( \hat{d}_n = O_P(n^{1/2} \log n) \) when \( n \) is large enough). Moreover, elementary statistics (mean and quantiles) show that the detection delay using our procedure is shorter than using the 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 \( \hat{C}_k \) could be recommended in this frame.

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...
<table>
<thead>
<tr>
<th>$\tilde{d}_n$</th>
<th>AR(1)</th>
<th>GARCH(1,1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 500$ ; $k^* = n + 50$</td>
<td>$\hat{C}_k$</td>
<td>55.36</td>
</tr>
<tr>
<td>$n = 500$ ; $k^* = n + 250$</td>
<td>$\hat{D}_k$</td>
<td>71.54</td>
</tr>
<tr>
<td>$n = 1000$ ; $k^* = n + 50$</td>
<td>$\hat{C}_k$</td>
<td>66.81</td>
</tr>
<tr>
<td>$n = 1000$ ; $k^* = n + 250$</td>
<td>$\hat{D}_k$</td>
<td>38.44</td>
</tr>
</tbody>
</table>

Table 7: Elementary statistics of the empirical detection delay for monitoring parameter change in AR(1) and GARCH(1,1).
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 Franq and Zakoïan 2010. Consider the observations going from January 2, 1995 to December 31, 1996 (resp. January 2, 2004 to December 30, 2005) as the historical data for the Nikkei 225 (resp. S&P 500 and FTSE 100) stock index. These periods are known to be stable in the 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 [20]. Define the asymptotic covariance matrix (which take into account the change possibility) of the estimator $\hat{\theta}_n(T_{1,n})$ by

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

The test is based on the statistic

$$\hat{Q}_n := \max(\hat{Q}^{(1)}_n, \hat{Q}^{(2)}_n)$$

where

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

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

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

- The second test (see Lee and Song [26]) is based on the statistic

$$\hat{Q}^{(0)}_n := \max_{v_n \leq k \leq n-v_n} \left( \frac{k^2}{n} (\hat{\theta}_n(T_{1,k}) - \hat{\theta}_n(T_{1,n}))^T \tilde{\Sigma}_{n,n} (\hat{\theta}_n(T_{1,k}) - \hat{\theta}_n(T_{1,n})) \right)$$

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

- The third procedure is the CUSUM test see Kulperger and Yu [23].

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 $C_\alpha$. Table 8 provides the results of these tests to the historical data that we have chosen.

<table>
<thead>
<tr>
<th></th>
<th>$Q_n$</th>
<th>$Q_n^{(0)}$</th>
<th>CUSUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nikkei 225</td>
<td>3.35</td>
<td>(3.98)</td>
<td>2.31</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>2.13</td>
<td>(3.47)</td>
<td>2.01</td>
</tr>
<tr>
<td>FTSE 100</td>
<td>1.95</td>
<td>(3.47)</td>
<td>2.31</td>
</tr>
</tbody>
</table>

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 $\hat{\theta}_n(T_{k,l})$ ($1 \leq k < l \leq n$). For the statistics $\hat{Q}_n$ and $\hat{Q}_n^{(0)}$, we consider only the time $k$ that the computation of $\hat{\theta}_n(T_{1,k})$ and $\hat{\theta}_n(T_{k,n})$ converges. This
certainly introduces distortions on these tests. On the other hand, the CUSUM procedure needs to compute only the estimator $\hat{\theta}_n(T_{1,n})$ 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 $(\hat{C}_k)$ 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 [20], the test based on $\hat{Q}_n$ and $\hat{Q}_n^{(0)}$ 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 $\hat{t}_N \simeq$ 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 $(\hat{C}_k)$ 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 $\hat{t}_{S,1} \simeq$ June 18, 2007 (for S&P 500 data) and $\hat{t}_{F,1} \simeq$ 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 $\hat{t}_{S,1}$ to November 16, 2007 (for S&P 500 data) and the series going from $\hat{t}_{F,1}$ 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 $(\hat{C}_k)$. 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 $\hat{t}_{S,2} \simeq$ August 14, 2008 and $\hat{t}_{F,2} \simeq$ September 17, 2008 respectively for these two series; see Figure 6. These breaks correspond to the beginning of the Subprime Crisis in US.

The procedure continues until June 2012, other breaks are detected at $\hat{t}_{S,4} \simeq$ 26 June 2009, $\hat{t}_{S,5} \simeq$ 5 April 2010, $\hat{t}_{S,6} \simeq$ 27 September 2010, $\hat{t}_{S,7} \simeq$ 19 July 2011, $\hat{t}_{S,8} \simeq$ 11 January 2012 (for S&P 500 data) and $\hat{t}_{F,4} \simeq$ 30 June 2009, $\hat{t}_{F,5} \simeq$ 27 July 2011, $\hat{t}_{F,6} \simeq$ 21 December 2011 (for FTSE 100 data). They are represented on Figure 6. These breaks correspond to the turmoil periods in the 2010 – 2012+ Greece and European debt crisis.

**Summary of the real-data applications**

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

   See also Figure 6.
Figure 2: The top figure is a realization of the statistics $\hat{C}_k$ 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 $\hat{C}_k$ 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 $\hat{C}_k$ 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 $\hat{C}_k$ 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.
Figure 6: Break detection in the returns of S&P 500 and FTSE 100 data using monitoring procedure based on $\hat{C}_k$. The verticals lines indicate the dates where breaks have been detected.
6 Proofs of the main results

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

Recall that \((X_1, \cdots, X_n)\) is an observed trajectory of a process \(M_Z(M_{\theta^*_n}, f_{\theta^*_n})\).

Let \(k \geq n \geq 2\) and \(T_{\ell,n} = \{1, \cdots, n\}, T_{\ell,k} = \{\ell, \ell + 1, \cdots, k\}\) with \(\ell \in \Pi_{n,k} = \{v_n, v_n + 1, \cdots, k - v_n\}\), and define

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

with \(\hat{\theta}\) defined in (6).

Lemma 6.1. Under the assumptions of Theorem 3.1,

\[
\sup_{k > n} \max_{\ell \in \Pi_{n,k}} \frac{1}{b((k - \ell)/n)} |\hat{C}_{k,\ell} - C_{k,\ell}| = o_P(1) \text{ as } n \to \infty.
\]

Proof. For any \(n \geq 1\), we have

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

Now, proceed similarly as in the proof of Lemma 3 of [20].

Lemma 6.2. Under the assumptions of Theorem 3.1

\[
\sup_{k > n} \max_{\ell \in \Pi_{n,k}} \frac{1}{b((k - \ell)/n)} \sqrt{n} \frac{k - \ell}{k} \|G^{-1/2}F \cdot (\hat{\theta}(T_{\ell,k}) - \hat{\theta}(T_{1,n})) - 2 \left( \frac{\partial}{\partial \theta} L(T_{\ell,k}, \theta^*_n) - \frac{k - \ell}{n} \frac{\partial}{\partial \theta} L(T_{1,n}, \theta^*_n) \right) \| = o_P(1) \text{ as } n \to \infty.
\]

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

\[
\frac{2}{\text{Card}(T)} \frac{\partial}{\partial \theta} \hat{L}(T, \theta^*_n) = \hat{F}(T) \cdot (\hat{\theta}(T) - \theta^*_n) \text{ where } \hat{F}(T) = -2 \left( \frac{1}{\text{Card}(T)} \frac{\partial^2 \hat{L}(T, \hat{\theta}_i(T))}{\partial \theta \partial \theta_i} \right)_{1 \leq i \leq d}
\]

for some \(\hat{\theta}_i(T)\) between \(\hat{\theta}(T)\) and \(\theta^*_n\).

Hence for any \(\ell \in \Pi_{n,k}\)

\[
F(\hat{\theta}(T_{\ell,k}) - \theta^*_n) = \frac{2}{k - \ell} \frac{\partial}{\partial \theta} L(T_{\ell,k}, \theta^*_n) + (F - \hat{F}(T_{\ell,k})) (\hat{\theta}(T_{\ell,k}) - \theta^*_n)
\]

\[
+ \frac{2}{k - \ell} \left( \frac{\partial}{\partial \theta} \hat{L}(T_{\ell,k}, \theta^*_n) - \frac{\partial}{\partial \theta} L(T_{\ell,k}, \theta^*_n) \right).
\]

and

\[
F(\hat{\theta}(T_{1,n}) - \theta^*_n) = \frac{2}{n} \frac{\partial}{\partial \theta} L(T_{1,n}, \theta^*_n) + (F - \hat{F}(T_{1,n})) (\hat{\theta}(T_{1,n}) - \theta^*_n) + \frac{2}{n} \left( \frac{\partial}{\partial \theta} \hat{L}(T_{1,n}, \theta^*_n) - \frac{\partial}{\partial \theta} L(T_{1,n}, \theta^*_n) \right).
\]
Therefore, for any \( \ell \in \Pi_{n,k} \)

\[
\sqrt{\frac{1}{k}} \left( (k - \ell) F \left( \hat{\theta}(T_{\ell,k}) - \hat{\theta}(T_{1,n}) \right) - 2 \left( \frac{\partial}{\partial \theta} L(T_{\ell,k}, \theta_0^*) - \frac{k - \ell}{n} \frac{\partial}{\partial \theta} L(T_{1,n}, \theta_0^*) \right) \right) \\
= \sqrt{\frac{1}{k}} \left( (F - \hat{F}(T_{\ell,k})) (\hat{\theta}(T_{\ell,k}) - \theta_0^*) + 2 \sqrt{\frac{1}{k}} \left( \frac{\partial}{\partial \theta} \hat{L}(T_{\ell,k}, \theta_0^*) - \frac{\partial}{\partial \theta} L(T_{\ell,k}, \theta_0^*) \right) \\
- \sqrt{\frac{1}{k}} (F - \hat{F}(T_{1,n})) (\hat{\theta}(T_{\ell,k}) - \theta_0^*) - 2 \frac{k - \ell}{k} \frac{1}{\sqrt{n}} \left( \frac{\partial}{\partial \theta} \hat{L}(T_{1,n}, \theta_0^*) - \frac{\partial}{\partial \theta} L(T_{1,n}, \theta_0^*) \right) \right). \tag{17}
\]

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

\[
\max_{\ell \in \Pi_{n,k}} \frac{1}{b((k - \ell)/n)} \sqrt{n} \frac{k - \ell}{k} \left\| F - \hat{F}(T_{\ell,k}) \right\| \left\| \hat{\theta}(T_{\ell,k}) - \theta_0^* \right\| \leq \frac{1}{\text{Inf}(s) \sqrt{k - \ell_k}} \left\| F - \hat{F}(T_{\ell_k,k}) \right\| \left\| \hat{\theta}(T_{\ell_k,k}) - \theta_0^* \right\|.
\]

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

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

Similar arguments imply that

\[
\sup_{k > n} \max_{\ell \in \Pi_{n,k}} \frac{1}{b((k - \ell)/n)} \sqrt{n} \frac{k - \ell}{k} \left\| F - \hat{F}(T_{1,n}) \right\| \left\| \hat{\theta}(T_{1,n}) - \theta_0^* \right\| = o_P(1) \text{ as } n \to \infty. \tag{19}
\]

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

\[
\max_{\ell \in \Pi_{n,k}} \frac{1}{b((k - \ell)/n)} \sqrt{n} \frac{k - \ell}{k} \left\| \frac{\partial}{\partial \theta} \hat{L}(T_{\ell_k,k}, \theta_0^*) - \frac{\partial}{\partial \theta} L(T_{\ell_k,k}, \theta_0^*) \right\| \leq \frac{1}{\text{Inf}(s) \sqrt{k - \ell_k}} \left\| \frac{\partial}{\partial \theta} \hat{L}(T_{\ell_k,k}, \theta_0^*) - \frac{\partial}{\partial \theta} L(T_{1,n}, \theta_0^*) \right\|. 
\]

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

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

Similar arguments show that

\[
\sup_{k > n} \max_{\ell \in \Pi_{n,k}} \frac{1}{b((k - \ell)/n)} \frac{k - \ell}{k} \left\| \frac{\partial}{\partial \theta} \hat{L}(T_{1,n}, \theta_0^*) - \frac{\partial}{\partial \theta} L(T_{1,n}, \theta_0^*) \right\| = o_P(1) \text{ as } n \to \infty. \tag{21}
\]

Thus, Lemma 6.2 follows from (17), (18), (19), (20) and (21).

**Lemma 6.3.** Under the assumptions of Theorem 3.1

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

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

**Proof.** We are going to apply Lemma 6.2 for specifying the asymptotic behaviour of \( \hat{\theta}(T_{\ell,k}) - \hat{\theta}(T_{1,n}) \).
For $k > n$ and $\ell \in \Pi_{n,k}$, we have

\[
2\sqrt{n} \left( \frac{\partial}{\partial \theta} L(T_{\ell,k}, \theta_0^n) - \frac{k - \ell}{n} \frac{\partial}{\partial \theta} L(T_{1,n}, \theta_0^n) \right) = -\frac{1}{k} \sum_{i=\ell+1}^{k} \frac{\partial q_i(\theta_0^n)}{\partial \theta} - \frac{k - \ell}{n} \sum_{i=1}^{n} \frac{\partial q_i(\theta_0^n)}{\partial \theta}.
\]

Now we are going to proceed in two steps.

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

\[
\max_{n < k < nT} \max_{\ell \in \Pi_{n,k}} \frac{1}{b((k - \ell)/n)} \left( \frac{2\sqrt{n}}{k} \frac{\partial}{\partial \theta} L(T_{\ell,k}, \theta_0^n) - \frac{k - \ell}{n} \frac{\partial}{\partial \theta} L(T_{1,n}, \theta_0^n) \right)
\]

\[
= \max_{n < k < nT} \max_{\ell \in \Pi_{n,k}} \frac{1}{b((k - \ell)/n)} \sqrt{n} \left( \sum_{i=\ell}^{k} \frac{\partial q_i(\theta_0^n)}{\partial \theta} - \frac{k - \ell}{n} \sum_{i=1}^{n} \frac{\partial q_i(\theta_0^n)}{\partial \theta} \right)
\]

\[
= \max_{\ell \in \{1, 1 + \frac{1}{T}, \ldots, T\}} \max_{s \in \{1 - \frac{n}{T}, 2 - \frac{2}{T}, \ldots, T - \frac{n}{T}\}} \frac{1}{b(\ell / n)} \sqrt{n} \left( \sum_{i=\ell}^{n} \frac{\partial q_i(\theta_0^n)}{\partial \theta} - \frac{n}{n} \sum_{i=1}^{n} \frac{\partial q_i(\theta_0^n)}{\partial \theta} \right)
\]

Define the set $S := \{(t, s) \in [1, T] \times [1, T] / s < t \}$. According to [4], $(\frac{\partial q_i(\theta_0^n)}{\partial \theta})_{t \in \mathbb{Z}}$ is a stationary ergodic martingale difference sequence with covariance matrix $G$. By Cramér-Wold device (see [10] p. 206), it holds that

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

with $\mathcal{D}(S)$ means the weak convergence on the Skorohod space $\mathcal{D}(S)$. Hence

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

Therefore

\[
\max_{n < k < nT} \max_{\ell \in \Pi_{n,k}} \frac{1}{b((k - \ell)/n)} \left( \frac{2\sqrt{n}}{k} \frac{\partial}{\partial \theta} L(T_{\ell,k}, \theta_0^n) - \frac{k - \ell}{n} \frac{\partial}{\partial \theta} L(T_{1,n}, \theta_0^n) \right)
\]

\[
\overset{\mathcal{D}}{\to} \sup_{1 < t < T} \sup_{1 < s < t} \left\| W_G(t - s) - (t - s)W_G(1) \right\| \frac{\sqrt{n}}{t b(t - s)}
\]

\[
\overset{\mathcal{D}}{\to} \sup_{1 < t < T} \sup_{1 < s < t} \left\| W_G(s) - s W_G(1) \right\| \frac{\sqrt{n}}{t b(s)}
\]

(22)

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

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

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

\[
\sup_{1 < s < t} \sup_{t > T} \left\| W_G(s) - s W_G(1) \right\| \frac{\sqrt{n}}{t b(s)}.
\]

Let $k > nT$. We have

\[
\max_{\ell \in \Pi_{n,k}} \frac{1}{b((k - \ell)/n)} \sqrt{n} \left\| \frac{\partial}{\partial \theta} L(T_{\ell,k}, \theta_0^n) \right\| \leq \frac{1}{\inf_{s > 0} b(s)} \sqrt{n} \left\| \sum_{i=\ell+1}^{k} \frac{\partial q_i(\theta_0^n)}{\partial \theta} \right\| \text{ for some } \ell_k \in \Pi_{n,k}.
\]
It comes from the Hájek-Rényi-Chow inequality (see [12]) that, for any $\varepsilon > 0$

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

Hence

$$\sup_{k>nT} \max_{\ell \in \Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \frac{2\sqrt{n}}{k} \left\| \frac{k-\ell}{n} \frac{\partial}{\partial \theta} L(T_{t,k}, \theta_0^*) \right\| = a_p(1) \text{ as } T, n \to \infty. \quad (23)$$

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

$$\sup_{k>nT} \max_{\ell \in \Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \frac{2\sqrt{n}}{k} \left\| \frac{k-\ell}{n} \frac{\partial}{\partial \theta} L(T_{t,k}, \theta_0^*) \right\| = \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial q_i(\theta_0^*)}{\partial \theta} \right\| \times \sup_{k>nT} \max_{\ell \in \Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \frac{k-\ell}{k}$$

$$= \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial q_i(\theta_0^*)}{\partial \theta} \right\| \times \sup_{k>nT} \frac{1}{b((k-v_n)/n)} \frac{k-v_n}{k}$$

$$= \frac{1}{\inf_{s>0} b(s)} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial q_i(\theta_0^*)}{\partial \theta} \right\| \frac{\sup_{s>0} \|W_G(1)\|}{\inf_{s>0} b(s)} \Rightarrow D_{n \to \infty} \frac{\sup_{s>0} \|W_G(1)\|}{\inf_{s>0} b(s)}.$$ \quad (24)

using again the Cramér-Wold device. It comes from (23) and (24) that

$$\sup_{k>nT} \max_{\ell \in \Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \frac{2\sqrt{n}}{k} \left\| \frac{k-\ell}{n} \frac{\partial}{\partial \theta} L(T_{t,k}, \theta_0^*) \right\| \Rightarrow D_{T,n \to \infty} \frac{1}{\inf_{s>0} b(s)} \|W_G(1)\|. \quad (25)$$

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

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

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

$$\sup_{t>T} \sup_{1<s<t} \frac{\|W_G(s)\|}{t b(s)} \leq \frac{1}{\inf_{s>0} b(s)} \left( \frac{1}{T} \sup_{1<s<t_0} \|W_G(s)\| + \sup_{t>T} \frac{\log(t)}{\sqrt{t}} \right) \Rightarrow 0. \quad (26)$$

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

$$\sup_{t>T} \sup_{1<s<t} \frac{\|s W_G(1)\|}{t b(s)} = \|W_G(1)\| \sup_{t>T} \frac{1}{t} \sup_{1<s<t} \frac{s}{b(t)} = \|W_G(1)\| \sup_{t>T} \frac{1}{t} \sup_{s>0} \frac{s}{b(s)} \Rightarrow D_{T \to \infty} \frac{1}{\inf_{s>0} b(s)} \|W_G(1)\|. \quad (27)$$

It comes from (26) and (27) that the limit of (22) satisfies when $T \to \infty$,

$$\sup_{t>T} \sup_{1<s<t} \frac{\|W_G(s) - s W_G(1)\|}{t b(s)} \Rightarrow D_{T \to \infty} \sup_{s>0} \frac{1}{\inf_{s>0} b(s)} \|W_G(1)\|. \quad (28)$$

From Step 1 and Step 2 (the relations (22), (25) and (28)), it comes that

$$\sup_{k>nT} \max_{\ell \in \Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \frac{2\sqrt{n}}{k} \left\| \frac{k-\ell}{n} \frac{\partial}{\partial \theta} L(T_{t,k}, \theta_0^*) \right\| \Rightarrow D_{n \to \infty} \sup_{t>T} \sup_{1<s<t} \frac{\|W_G(s) - s W_G(1)\|}{t b(s)}.$$
Hence, Lemma 6.3 follows from Lemma 6.2.

**Proof of Theorem 3.1**

We know that

\[
P\{\tau(n) < \infty\} = P\left\{ \sup_{k > n} \max_{\ell \in \Pi_{n, k}} \frac{\hat{C}_{k, \ell}}{b((k - \ell)/n)} > 1 \right\}
\]

\[
= P\left\{ \sup_{k > n} \max_{\ell \in \Pi_{n, k}} \frac{1}{b((k - \ell)/n)} \sqrt{n} \frac{k - \ell}{k} \left\| \hat{G}(T_{1, n})^{-1/2} \hat{F}(T_{1, n}) \cdot (\hat{\theta}(T_{k, k}) - \hat{\theta}(T_{1, n})) \right\| > 1 \right\}.
\]

Since \( \hat{G}(T_{1, n}) \xrightarrow{a.s.} G \) and \( \hat{F}(T_{1, n}) \xrightarrow{a.s.} F \), it comes from Lemma 6.1 and 6.3 that

\[
\sup_{k > n} \max_{\ell \in \Pi_{n, k}} \frac{1}{b((k - \ell)/n)} \sqrt{n} \frac{k - \ell}{k} \left\| \hat{G}(T_{1, n})^{-1/2} \hat{F}(T_{1, n}) \cdot (\hat{\theta}(T_{k, k}) - \hat{\theta}(T_{1, n})) \right\| \xrightarrow{p} \sup_{t > 1} \sup_{1 < s < t} \|G^{-1/2}(W_G(s) - sW_G(1))\| t b(s).
\]

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

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

Now, it suffices to show that \( \sup_{t > 1} \sup_{1 < s < t} \frac{1}{t} \|W_d(s) - sW_d(1)\| \xrightarrow{P} U_d. \)

For any \( t > 1 \), we have

\[
\sup_{1 < s < t} \frac{1}{t} \|W_d(s) - sW_d(1)\| \xrightarrow{P} \sup_{0 < u \in 1 - \frac{1}{t}} \frac{s}{t} \|W_d(s - 1/s)\| = \sup_{0 < u \in 1 - \frac{1}{t}} \frac{s}{t} (1 - u) \|W_d(u)\|.
\]

Thus

\[
\sup_{t > 1} \sup_{1 < s < t} \frac{1}{t} \|W_d(s) - sW_d(1)\| \xrightarrow{P} \sup_{0 < u \in 1 - \frac{1}{t}} \frac{1}{t} (1 - u) \|W_d(u)\| = \sup_{0 < u \in 1 - \frac{1}{t}} \frac{1 - u}{u} \|W_d(u)\|.
\]

But, \( \|W_d(u)\| \equiv \sqrt{v} \sqrt{\|W_d(u')\|}. \) Therefore with \( u = u'v \),

\[
\sup_{0 < u \in 1} \frac{1 - u}{u} \|W_d(u)\| = \sup_{0 < v \in 1 - u'} \frac{(1 - v)}{1 - u'} \sqrt{v} \sqrt{\|W_d(u')\|}.
\]

It remains to compute \( \sup_{0 < v \in 1} \frac{(1 - v)}{1 - u'} \sqrt{v} \sqrt{u'} \). Classical computations show that this supremum is obtained by

\[
v = 2 \left(3 - u' + \sqrt{(9 - u')(1 - u')}\right)^{-1}
\]

and therefore

\[
\sup_{t > 1} \sup_{1 < s < t} \frac{1}{t} \|W_d(s) - sW_d(1)\| \xrightarrow{P} \sup_{0 < u' \in 1} f(u') \|W_d(u')\|
\]

with \( f(u') = \frac{\sqrt{9 - u' + \sqrt{1 - u'}}}{\sqrt{9 - u' + 3\sqrt{1 - u'}}} (\frac{2}{3 - u' + \sqrt{(9 - u')(1 - u')}})^{1/2}. \) (29)
Hence,
\[
\sup_{t>1} \sup_{1<s<t} \frac{1}{t} \| W_d(s) - sW_d(1) \| \overset{D}{=} U_d
\]

\section*{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
\[
\max_{\ell \in \Pi_{n,k} \setminus b((k_n - \ell)/n)} \frac{\hat{C}_{k_n,\ell}}{b((k_n - \ell)/n)} = \max_{\ell \in \Pi_{n,k} \setminus b((k_n - \ell)/n)} \frac{1}{b((k_n - \ell)/n)} \sqrt{n} \frac{k_n - \ell}{k_n} \| \hat{G}(T_{1,n})^{-1/2} \hat{F}(T_{1,n}) \cdot (\hat{\theta}_{k_n, (T_k, k_n)} - \hat{\theta}(T_{1,n})) \|
\geq c \sqrt{n} \frac{n^\delta}{k_n + n^\delta} \| \hat{G}(T_{1,n})^{-1/2} \hat{F}(T_{1,n}) (\hat{\theta}_{k_n, (T_k, k_n)} - \hat{\theta}(T_{1,n})) \|
\geq c \frac{n^{1/2 + \delta}}{c_0 n + n^\delta} \| \hat{G}(T_{1,n})^{-1/2} \hat{F}(T_{1,n}) (\hat{\theta}_{k_n, (T_k, k_n)} - \hat{\theta}(T_{1,n})) \|
\geq c \frac{n^{3/2 - \delta}}{c_0^2 n + n^{2\delta}} \| \hat{G}(T_{1,n})^{-1/2} \hat{F}(T_{1,n}) (\hat{\theta}_{k_n, (T_k, k_n)} - \hat{\theta}(T_{1,n})) \|.
\] (30)

According to [4] and [20], \( \hat{G}(T_{1,n}) \overset{a.s.}{\to}_{n \to \infty} G \), \( \hat{F}(T_{1,n}) \overset{a.s.}{\to}_{n \to \infty} F \) and \( \hat{\theta}_{k_n, (T_k, k_n)} \overset{a.s.}{\to}_{n \to \infty} \theta_1^* \) and \( \hat{\theta}_{k_n, (T_k, k_n)} \overset{a.s.}{\to}_{n \to \infty} \theta_1^* \).

Since \( G \) is symmetric positive definite, \( F \) is invertible, \( \theta_0^* \neq \theta_1^* \) and \( \delta > 1/2 \), then (30) implies that
\[
\max_{\ell \in \Pi_{n,k} \setminus b((k_n - \ell)/n)} \frac{\hat{C}_{k_n,\ell}}{b((k_n - \ell)/n)} \overset{a.s.}{\to}_{n \to \infty} \infty.
\]

\section*{Acknowledgements}
The authors are grateful to the anonymous referee for its report which helped to improve this paper.

\section*{References}


