Zuzana Prášková
Change point detection in vector autoregression

Kybernetika, Vol. 54 (2018), No. 6, 1122–1137

Persistent URL: `http://dml.cz/dmlcz/147600`

KYBERNETIKA — VOLUME *54* (2018), NUMBER 6, PAGES 1122–1137

CHANGE POINT DETECTION IN VECTOR AUTOREGRESSION

Zuzana Prášková

In the paper a sequential monitoring scheme is proposed to detect instability of parameters in a multivariate autoregressive process. The proposed monitoring procedure is based on the quasi-likelihood scores and the quasi-maximum likelihood estimators of the respective parameters computed from a training sample, and it is designed so that the sequential test has a small probability of a false alarm and asymptotic power one as the size of the training sample is sufficiently large. The asymptotic distribution of the detector statistic is established under both the null hypothesis of no change as well as under the alternative that a change occurs.

Keywords: vector autoregression, change point, quasi-maximum likelihood

Classification: 62M10, 62E20

1. INTRODUCTION

Consider a general vector autoregression model of order p, VAR(p),

$$\boldsymbol{y}_t = \boldsymbol{c} + \boldsymbol{\Phi}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p \boldsymbol{y}_{t-p} + \boldsymbol{\epsilon}_t, \quad t \in \mathbb{Z}$$

where $\boldsymbol{c}$ is a $d \times 1$ vector of constants, $\boldsymbol{\Phi}_1, \dots, \boldsymbol{\Phi}_p$ are $d \times d$ matrices of autoregression parameters, and $\boldsymbol{\epsilon}_t$ is a $d \times 1$ vector of errors with zero mean and a variance matrix that is denoted by $\boldsymbol{\Omega}$.

It is well known that any $d-$dimensional VAR(p) process can be represented by a $pd-$dimensional VAR(1) process (cf., e. g., Lütkepohl [16], p. 15) so in what follows, we will focus on a $d-$dimensional VAR(1) process only, i. e., on the process $\boldsymbol{y}_t = \boldsymbol{c} + \boldsymbol{\Phi} \boldsymbol{y}_{t-1} + \boldsymbol{\epsilon}_t$. We introduce a new parametrization

$$\boldsymbol{y}_t - \boldsymbol{\mu} = \boldsymbol{\Phi}(\boldsymbol{y}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\Omega}^{\frac{1}{2}} \, \mathbf{e}_t, \quad t \in \mathbb{Z} \tag{1}$$

if we put $\boldsymbol{e}_t = \boldsymbol{\Omega}^{-\frac{1}{2}} \boldsymbol{\epsilon}_t$ and $\boldsymbol{c} = (\boldsymbol{I} - \boldsymbol{\Phi})\boldsymbol{\mu}$ where $\boldsymbol{I}$ denotes the identity matrix.

The vector of parameters of the model is $\boldsymbol{\theta} = (\boldsymbol{\mu}', \boldsymbol{\phi}', \boldsymbol{\sigma}')'$ where $\boldsymbol{\mu} = (\boldsymbol{I} - \boldsymbol{\Phi})^{-1}\boldsymbol{c}$, $\boldsymbol{\phi} = \mathsf{vec}\, \boldsymbol{\Phi}$ and $\boldsymbol{\sigma} = \mathsf{vech}\, \boldsymbol{\Omega}$. In this notation, all the vectors are columns and for a vector $\boldsymbol{x}$, its transpose is denoted by $\boldsymbol{x}'$. Further, for a matrix $\boldsymbol{A}$, $\mathsf{vec}\boldsymbol{A}$ is a column vector that stacks columns of the matrix $\boldsymbol{A}$ one on top of another one in the order from left to right,

DOI: 10.14736/kyb-2018-6-1122

and for a squared symmetric matrix $\boldsymbol{A}$, $\mathsf{vech}\boldsymbol{A}$ is a column vector that stacks columns of the lower triangular submatrix of $\boldsymbol{A}$ one on top of another one in the order from left to right. In general, $\boldsymbol{\theta} = \boldsymbol{\theta}_t$. Obviously, for the $d-$dimensional process given in (1), $\boldsymbol{\theta} \in \mathbb{R}^r, r = \frac{3}{2}d(d+1)$.

In our next considerations we will deal with the sequential testing of the stability of parameters $\boldsymbol{\theta}_t$. For this we assume that a training sample of stable observations $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$ is available that serves to the calibration of the model, such that

$$\boldsymbol{\theta}_1 = \cdots = \boldsymbol{\theta}_m = \boldsymbol{\theta}_0.$$

New observations are arriving one after another; after each new observation arrives, we make a decision whether the condition of stability is violated (i. e., if a change occurs) or not.

The problem of an instability of parameters $\boldsymbol{\theta}_t$ is formulated as a sequential testing problem, that is, we test the null hypothesis of no change

$$H_0 : \boldsymbol{\theta}_t = \boldsymbol{\theta}_0, \quad t = 1, 2, \ldots, \tag{2}$$

against the alternative that a change occurs at time k^* after the monitoring started, i. e., at a break point $m + k^*$,

$$\begin{aligned} H_A : & \text{ there exists } k^* \geq 1, \\ & \boldsymbol{\theta}_t = \boldsymbol{\theta}_0 \,, \; 1 \leq t \leq m + k^*, \\ & \boldsymbol{\theta}_t = \boldsymbol{\theta}_1 \,, \; m + k^* < t < \infty, \quad \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_0, \end{aligned} \tag{3}$$

where $\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, k^* = k^*_m$ are unknown. The decision is based on a detector statistic constructed from all observations up to $m + k, k = 1, 2, \ldots$, and when it exceeds a critical level for the first time, we stop the monitoring. We utilize here the ideas that appeared in the papers by Chu et al. [5] and Horváth et al. [10]. For a detector statistic $Q(m, k)$ that will be considered below we define the stopping time

$$\begin{aligned} \tau(m) = \inf\{k \geq 1, |Q(m,k)| \geq c_\alpha\, q(m/k)\} \\ (= \infty, \text{ if } |Q(m,k)| < c_\alpha\, q(m/k) \text{ for all } \; k \geq 1) \end{aligned} \tag{4}$$

if we consider an open-end procedure, or, in case of a closed-end procedure,

$$\tau_m = \inf\{1 \leq k \leq \lfloor mT \rfloor : Q(k,m)/q(k/m) \geq c_\alpha\} \tag{5}$$

where $T > 0$ is a positive number that can depend on m, $T = T_m$, and we assume that $mT_m \to \infty$ as $m \to \infty$. In our case we will assume T to be fixed. By $\lfloor mT \rfloor$ we mean the upper bound for the maximum number of possible observations, usually specified a priori. The function $q(t)$ is a boundary function and $c_\alpha > 0$ is a constant such that

$$\lim_{m \to \infty} \mathsf{P}(\tau(m) < \infty | H_0) = \alpha$$

which means that α is the prescribed probability of false alarm, and

$$\lim_{m \to \infty} \mathsf{P}(\tau(m) < \infty | H_A) = 1$$

which means that the test is consistent.

In the univariate case, monitoring in autoregressive sequences was considered by Hušková and Koubková [12], who used CUSUM statistics based on weighted residuals for monitoring changes in autoregression parameters. Gombay and Serban [8] considered a procedure based on maximum likelihood scores and normally distributed errors, but their test statistic depends on prescribed values of monitoring parameters. Lee et al. [15] and Hlávka et al. [11] considered monitoring procedures based on empirical characteristic functions. Carsoule and Frances [4] considered a procedure based on maximum likelihood scores under normality assumptions and sketched some theoretical results. Quasi-maximum likelihood based monitoring in random coefficient autoregression was considered in Li et al. [14], Na et al. [17] and Prášková [18].

These on-line procedures differ from an alternative approach to the change point analysis, called off-line or retrospective analysis. The off-line methods detect possible changes when the observations are available before the analysis and the sample size is known. The off-line methods in multivariate autoregression were studied under general assumptions by Dvořák [7] where recent results in the field are also reviewed.

In this paper we consider monitoring in the multivariate autoregressive process (1) based on the quasi-maximum likelihood score function and under general assumptions on the errors. The conditional (given $\boldsymbol{y}_0$) log-likelihood function based on the observations $\boldsymbol{y}_1, \dots, \boldsymbol{y}_k$ and normally distributed errors $\boldsymbol{\epsilon}_t$ is

$$l_k(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=1}^{k}\big[(\boldsymbol{y}_t - \boldsymbol{\mu} - \boldsymbol{\Phi}(\boldsymbol{y}_{t-1} - \boldsymbol{\mu}))'\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu} - \boldsymbol{\Phi}(\boldsymbol{y}_{t-1} - \boldsymbol{\mu})) + \log\det(\boldsymbol{\Omega})\big] = \sum_{t=1}^{k} g_t(\boldsymbol{\theta}) \tag{6}$$

but in the sequel we will drop the assumption of normality. Then we have the gradient vector

$$\boldsymbol{G}_t(\boldsymbol{\theta}) = \Big(\frac{\partial g_t(\boldsymbol{\theta})}{\partial \boldsymbol{\mu}}', \frac{\partial g_t(\boldsymbol{\theta})}{\partial \boldsymbol{\phi}}', \frac{\partial g_t(\boldsymbol{\theta})}{\partial \boldsymbol{\sigma}}'\Big)' = \begin{pmatrix} (\boldsymbol{I} - \boldsymbol{\Phi})'\boldsymbol{\Omega}^{-1}\boldsymbol{\epsilon}_t \\ (\boldsymbol{I} \otimes \boldsymbol{\Omega}^{-1})\mathsf{vec}[\boldsymbol{\epsilon}_t(\boldsymbol{y}_{t-1} - \boldsymbol{\mu})'] \\ \frac{1}{2}\boldsymbol{D}_d'(\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1})\mathsf{vec}[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t' - \boldsymbol{\Omega}] \end{pmatrix} \tag{7}$$

where $\boldsymbol{\epsilon}_t = \boldsymbol{y}_t - \boldsymbol{\mu} - \boldsymbol{\Phi}(\boldsymbol{y}_{t-1} - \boldsymbol{\mu})$ and $\boldsymbol{D}_d$ is the $d^2 \times d(d+1)/2$ duplication matrix, i. e., such matrix that $\boldsymbol{D}_d(\mathsf{vech}\boldsymbol{\Omega}) = \mathsf{vec}(\boldsymbol{\Omega})$. The symbol $\otimes$ denotes the Kronecker product.

Let $\boldsymbol{J}(\boldsymbol{\theta})$ be the matrix

$$\boldsymbol{J}(\boldsymbol{\theta}) = \mathsf{E}\,\boldsymbol{G}_t(\boldsymbol{\theta})\boldsymbol{G}_t(\boldsymbol{\theta})'. \tag{8}$$

We define the detector statistic

$$Q(m,k) = \frac{1}{\sqrt{m}}\widehat{\boldsymbol{J}}_m^{-\frac{1}{2}}\sum_{t=m+1}^{m+k}\boldsymbol{G}_t(\widehat{\boldsymbol{\theta}}_m) \tag{9}$$

where $\widehat{\boldsymbol{\theta}}_m$ is the quasi-maximum likelihood estimator (QMLE) of $\boldsymbol{\theta}_0$ based on the historical (training) observations, i. e., a solution of the equation

$$\sum_{t=1}^{m} \boldsymbol{G}_t(\widehat{\boldsymbol{\theta}}_m) = \boldsymbol{0} \tag{10}$$

and $\widehat{\boldsymbol{J}}_m$ is an estimator of $\boldsymbol{J}(\boldsymbol{\theta}_0)$ based on the historical observations.

Thus, we can expect that large values of $Q(m,k)$ will detect a violation of the null hypothesis and we reject H_0 when $|Q(m,k)| > c_\alpha q(k/m)$ for the first time, where q is a boundary function, c_α is the $\alpha-$critical value of the test and $|\cdot|$ denotes the maximum norm of a vector. In the sequel, we will consider the function

$$q(t) = (1+t)\left(\frac{t}{1+t}\right)^{\gamma}, \gamma \in [0, 1/2) \tag{11}$$

where γ is a constant that influences the ability of the test to detect a change, see, e. g., Horváth et al. [10] for details.

The paper is further organized as follows. In Section 2 we formulate the basic assumptions and main results. The proofs of the main theorems together with some auxiliary assertions are given in Section 3. In Section 4 results of a short simulation study are presented. In the proofs we will also use the Euclidean and Frobenius norm of a vector or a matrix, respectively, which will be denoted by $||\cdot||$. All the used matrix operations are taken from Lütkepohl [16].

2. ASSUMPTIONS AND MAIN RESULTS

First, let us introduce the following assumptions.

A1 $\det(\boldsymbol{I} - \boldsymbol{\Phi} z) \neq 0, |z| \leq 1;\quad \boldsymbol{\Omega}$ is positive definite;

A2 $\{\boldsymbol{e}_t\}$ is a (strictly) stationary end ergodic sequence of martingale differences with respect to $\mathcal{F}_t = \sigma\{\boldsymbol{e}_s, s \leq t\}$;

A3 $\mathsf{E}\, \boldsymbol{e}_t \boldsymbol{e}_t' | \mathcal{F}_{t-1} = \boldsymbol{I}$;

A4 $\mathsf{E}\, e_{it} e_{jt} e_{kt} = 0\ \forall (i,j,k), \forall t$ or $\mathsf{E}\, e_{it} e_{jt} e_{kt} | \mathcal{F}_{t-1} = K < \infty\ \forall (i,j,k)$;

A5 $\mathsf{E}\, ||\boldsymbol{e}_t||^4 < \infty$.

Alternatively, we can consider another set of conditions:

B1 $\det(\boldsymbol{I} - \boldsymbol{\Phi} z) \neq 0, |z| \leq 1;\quad \boldsymbol{\Omega}$ is positive definite;

B2 $\mathsf{E}\, \boldsymbol{e}_t = \boldsymbol{0},\ \forall t,\ \mathsf{E}\, \boldsymbol{e}_t \boldsymbol{e}_s' = \boldsymbol{I}\delta(t-s)\ \forall t, s$;

B3 $\mathsf{E}\, e_{it} e_{jt} e_{kt} = 0\ \forall t,\ \forall (i,j,k)$;

B4 $\sup_t \mathsf{E}\, ||\boldsymbol{e}_t||^{4+\kappa} < \infty$ for some $\kappa > 0$;

B5 $\{\boldsymbol{y}_t\}$ is a stationary and ergodic strong mixing sequence with the mixing coefficient $\alpha_n = O(n^{-(1+\epsilon)(1+\frac{4}{\kappa})})$ for some $\epsilon > 0$.

Remark 2.1. It is obvious that assumptions A1 and B1 are identical but we keep them in both groups of conditions for completeness.

Notice also that under A1–A5, $\{\boldsymbol{G}_t(\boldsymbol{\theta})\}$ is a strictly stationary and ergodic sequence of martingale differences while under B1–B5, $\{\boldsymbol{G}_t(\boldsymbol{\theta})\}$ is a strictly stationary and ergodic strong mixing sequence with the mixing coefficient α_n.

It is well known that under A1–A3 or B1–B2 and B5, $\{\boldsymbol{y}_t\}$ is stationary with the infinite moving average representation

$$\boldsymbol{y}_t = \boldsymbol{\mu} + \sum_{j=0}^{\infty} \boldsymbol{\Phi}^j \boldsymbol{\epsilon}_{t-j}, \quad t \in \mathbb{Z}, \tag{12}$$

and the series on the right-hand side of (12) is component-wise absolutely summable (see, e. g., Hamilton [9], Chapter 10.1). Then, using A1–A5 or B1–B5, we get from (7) and (8) that $\boldsymbol{J}(\boldsymbol{\theta})$ is regular and

$$\boldsymbol{J}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{J}_{11}(\boldsymbol{\theta}) & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{J}_{22}(\boldsymbol{\theta}) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{J}_{33}(\boldsymbol{\theta}) \end{pmatrix}$$

where

$$\begin{aligned}
\boldsymbol{J}_{11}(\boldsymbol{\theta}) &= (\boldsymbol{I} - \boldsymbol{\Phi})'\boldsymbol{\Omega}^{-1}(\boldsymbol{I} - \boldsymbol{\Phi}), \\
\boldsymbol{J}_{22}(\boldsymbol{\theta}) &= \boldsymbol{\Gamma}_y(0) \otimes \boldsymbol{\Omega}^{-1}, \quad \boldsymbol{\Gamma}_y(0) = \mathsf{E}\,(\boldsymbol{y}_1 - \boldsymbol{\mu})(\boldsymbol{y}_1 - \boldsymbol{\mu})', \\
\boldsymbol{J}_{33}(\boldsymbol{\theta}) &= \frac{1}{4}\boldsymbol{D}_d'(\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1})\boldsymbol{V}(\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1})\boldsymbol{D}_d, \\
\boldsymbol{V} &= \mathsf{Var}\ (\mathsf{vec}[\boldsymbol{\epsilon}_1\boldsymbol{\epsilon}_1' - \boldsymbol{\Omega}]).
\end{aligned}$$

Further, we will need some assumptions on the parameter space.

C1 $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ where $\boldsymbol{\Theta}$ is a compact subset of $\mathbb{R}^r$;

C2 $\boldsymbol{\theta}_0$ is an interior point of $\boldsymbol{\Theta}$.

In addition, we will also assume that

C3 $\inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\det \boldsymbol{\Omega}(\boldsymbol{\theta})| \geq \delta > 0$.

Asymptotic properties of the QMLE are summarized in the following lemma.

Lemma 2.2. Let $\{\boldsymbol{y}_t\}$ be the sequence defined in (1) with the parameter $\boldsymbol{\theta}_0$ which is unknown. Let $\widehat{\boldsymbol{\theta}}_m$ be the QMLE of $\boldsymbol{\theta}_0$ based on $\boldsymbol{y}_0, \ldots, \boldsymbol{y}_m$. Then, under A1–A5 or B1–B5, together with C1–C3, as $m \to \infty$, $\widehat{\boldsymbol{\theta}}_m \to \boldsymbol{\theta}_0$ a.s. and

$$\sqrt{m}(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{J}(\boldsymbol{\theta}_0)\boldsymbol{I}(\boldsymbol{\theta}_0)^{-1}) \tag{13}$$

where $\boldsymbol{J}(\boldsymbol{\theta}_0)$ is defined by (8) and $\boldsymbol{I}(\boldsymbol{\theta}_0)$ is the information matrix.

P r o o f. It follows from Theorem 1 in Boubacar Mainassara and Francq [3]. □

The QMLE of $\boldsymbol{\theta}_0$ based on the observations $\boldsymbol{y}_0, \dots, \boldsymbol{y}_m$ can be computed from the following relations:

$$\widehat{\boldsymbol{\mu}}_m = (\boldsymbol{I} - \widehat{\boldsymbol{\Phi}}_m)^{-1}(\overline{\boldsymbol{y}}_m - \widehat{\boldsymbol{\Phi}}_m \overline{\boldsymbol{y}}_{m(1)}),$$

$$\widehat{\boldsymbol{\Phi}}_m = \sum_{t=1}^{m}(\boldsymbol{y}_t - \widehat{\boldsymbol{\mu}}_m)(\boldsymbol{y}_{t-1} - \widehat{\boldsymbol{\mu}}_m)' \left[\sum_{t=1}^{m}(\boldsymbol{y}_{t-1} - \widehat{\boldsymbol{\mu}}_m)(\boldsymbol{y}_{t-1} - \widehat{\boldsymbol{\mu}}_m)'\right]^{-1},$$

$$\widehat{\boldsymbol{\Omega}}_m = \frac{1}{m}\sum_{t=1}^{m}(\boldsymbol{y}_t - \widehat{\boldsymbol{\mu}}_m - \widehat{\boldsymbol{\Phi}}_m(\boldsymbol{y}_{t-1} - \widehat{\boldsymbol{\mu}}_m))(\boldsymbol{y}_t - \widehat{\boldsymbol{\mu}}_m - \widehat{\boldsymbol{\Phi}}_m(\boldsymbol{y}_{t-1} - \widehat{\boldsymbol{\mu}}_m))'$$

$$= \frac{1}{m}\sum_{t=1}^{m}\widehat{\boldsymbol{\epsilon}}_t\widehat{\boldsymbol{\epsilon}}_t',$$

$$\widehat{\boldsymbol{V}}_m = \frac{1}{m}\sum_{t=1}^{m}\mathsf{vec}(\widehat{\boldsymbol{\epsilon}}_t\widehat{\boldsymbol{\epsilon}}_t')(\mathsf{vec}(\widehat{\boldsymbol{\epsilon}}_t\widehat{\boldsymbol{\epsilon}}_t'))' - \mathsf{vec}\widehat{\boldsymbol{\Omega}}_m(\mathsf{vec}\widehat{\boldsymbol{\Omega}}_m)',$$

$$\widehat{\boldsymbol{\Gamma}}_y(0) = \frac{1}{m}\sum_{t=1}^{m}(\boldsymbol{y}_t - \widehat{\boldsymbol{\mu}}_m)(\boldsymbol{y}_t - \widehat{\boldsymbol{\mu}}_m)'$$

where we have denoted $\overline{\boldsymbol{y}}_m = \sum_{t=1}^{m}\boldsymbol{y}_t/m$ and $\overline{\boldsymbol{y}}_{m(1)} = \sum_{t=1}^{m}\boldsymbol{y}_{t-1}/m$, $\widehat{\boldsymbol{\epsilon}}_t = \boldsymbol{y}_t - \widehat{\boldsymbol{\mu}}_m - \widehat{\boldsymbol{\Phi}}_m(\boldsymbol{y}_{t-1} - \widehat{\boldsymbol{\mu}}_m)$.

Now we can formulate the basic result concerning the asymptotic distribution of the test statistic under H_0.

Theorem 2.3. Let assumptions A1–A5 or B1–B5, together with C1–C3 be satisfied. Let Q_{mk} be the statistic defined by (9) and let $\widehat{\boldsymbol{J}}_m$ be an estimator of $\boldsymbol{J}(\boldsymbol{\theta}_0)$ such that $\widehat{\boldsymbol{J}}_m - \boldsymbol{J}(\boldsymbol{\theta}_0) = o_p(1)$ as $m \to \infty$. Then, under H_0, as $m \to \infty$,

$$\mathsf{P}\left(\sup_{1\le k<\infty}\left|\frac{Q(m,k)}{q(m/k)}\right| \le x\right) \longrightarrow \mathsf{P}\left(\sup_{0\le t\le 1}\left|\frac{\boldsymbol{W}(t)}{t^\gamma}\right| \le x\right) \tag{14}$$

$$\mathsf{P}\left(\sup_{1\le k\le \lfloor mT\rfloor}\left|\frac{Q(m,k)}{q(m/k)}\right| \le x\right) \longrightarrow \mathsf{P}\left(\sup_{0\le t\le T/(T+1)}\left|\frac{\boldsymbol{W}(t)}{t^\gamma}\right| \le x\right) \tag{15}$$

where $\{\boldsymbol{W}(t), t \in [0,1]\}$ is an r-dimensional standard Wiener process on $[0,1]$, $r = \frac{3}{2}d(d+1)$.

P r o o f. The proof is postponed to the next section. □

Next, we will study the asymptotic distribution of the test statistic under the alternative hypothesis H_A.

Let us consider the process

$$\boldsymbol{y}_t = \boldsymbol{\mu}_0 + \boldsymbol{\Phi}_0(\boldsymbol{y}_{t-1} - \boldsymbol{\mu}_0) + \boldsymbol{\Omega}_0^{\frac{1}{2}}\, \boldsymbol{e}_t, \quad t = 1, \dots, m + k^* \tag{16}$$

$$= \boldsymbol{\mu}_1 + \boldsymbol{\Phi}_1(\boldsymbol{y}_{t-1} - \boldsymbol{\mu}_1) + \boldsymbol{\Omega}_1^{\frac{1}{2}}\, \boldsymbol{e}_t, \quad t = m + k^* + 1, \dots \tag{17}$$

where (16) represents the process before a change, while (17) describes the behaviour of the process after the change that occurred at time $m+k^*$. Parameters of the process after the change are $\boldsymbol{\theta}_1 = (\boldsymbol{\mu}_1', \boldsymbol{\phi}_1', \boldsymbol{\sigma}_1')'$ with the same meaning as under H_0. The error process $\{\boldsymbol{e}_t\}$ is unchanged and we will assume that assumptions A1, B1, C1–C3 are satisfied for $\boldsymbol{\mu}_1, \boldsymbol{\Phi}_1$ and $\boldsymbol{\Omega}_1$, too. We will assume that $\boldsymbol{\theta}_0 \neq \boldsymbol{\theta}_1$. Let $\{\widetilde{\boldsymbol{y}}_t\}$ be a process that solves equation (17) for all $t \in \mathbb{Z}$ and $\{\widetilde{\boldsymbol{G}}_t(\boldsymbol{\theta})\}$ be the corresponding gradient sequence on $\boldsymbol{\Theta}$. Then, under our assumptions, $\{\widetilde{\boldsymbol{y}}_t\}$ and $\{\widetilde{\boldsymbol{G}}_t(\boldsymbol{\theta})\}$ are stationary and ergodic, and in general,

$$\mathsf{E}\, \widetilde{\boldsymbol{G}}_t(\boldsymbol{\theta}_0) \neq \mathbf{0}. \tag{18}$$

Theorem 2.4. Suppose that the alternative hypothesis (16) and (17) holds and assumptions A1–A5 or B1–B5, together with C1–C3 are satisfied for both the parameters $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_0 \neq \boldsymbol{\theta}_1$, and condition (18) holds. Suppose that $k^* = k^*_m$ and $\limsup_{m\to\infty} \frac{k^*_m}{m} < T$. Then, as $m \to \infty$,

$$\sup_{1 \le k < \infty} \left| \frac{Q(m,k)}{q(m/k)} \right| \xrightarrow{\mathsf{P}} \infty \tag{19}$$

and

$$\sup_{1 \le k \le \lfloor mT \rfloor} \left| \frac{Q(m,k)}{q(m/k)} \right| \xrightarrow{\mathsf{P}} \infty. \tag{20}$$

Proof. The proof is postponed to the next section. □

3. PROOFS

We start this section with the following lemma.

Lemma 3.1. Under A1–A5 or B1–B5, together with C1–C3,

$$\mathsf{E} \sup_{\theta \in \boldsymbol{\Theta}} ||\boldsymbol{G}_t(\boldsymbol{\theta})\boldsymbol{G}_t(\boldsymbol{\theta})'|| < \infty \;\forall t \tag{21}$$

and

$$\mathsf{E} \sup_{\theta \in \boldsymbol{\Theta}} ||H_t(\theta)|| < \infty \;\forall t \tag{22}$$

where

$$H_t(\boldsymbol{\theta}) = \frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}.$$

P r o o f. We will prove (22), only, since the proof of (21) is analogous. Notice that $\{\boldsymbol{H}_t(\boldsymbol{\theta})\}$ is a strictly stationary and ergodic sequence and that $H_t(\boldsymbol{\theta})$ is a symmetric, block-wise divided matrix with the blocks $\frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial\boldsymbol{\mu}\partial\boldsymbol{\mu}'}, \ldots, \frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial\boldsymbol{\sigma}\partial\boldsymbol{\sigma}'}$, thus we can consider each block separately. Throughout the computations we will repeatedly use the relations $||\boldsymbol{A}|| = \sqrt{\mathsf{Tr}\boldsymbol{A}'\boldsymbol{A}}$ and $\mathsf{Tr}(\boldsymbol{A}\otimes\boldsymbol{B}) = \mathsf{Tr}\boldsymbol{A}\mathsf{Tr}\boldsymbol{B}$, respectively, where Tr denotes the trace of a matrix and $\boldsymbol{A}, \boldsymbol{B}$ are matrices of the respective sizes.

First, let us consider $\frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial\boldsymbol{\mu}\partial\boldsymbol{\mu}'} = -(\boldsymbol{I}-\boldsymbol{\Phi})'\boldsymbol{\Omega}^{-1}(\boldsymbol{I}-\boldsymbol{\Phi})$. Due to assumptions A1/B1 and C1–C3, $||(\boldsymbol{I}-\boldsymbol{\Phi})'\boldsymbol{\Omega}^{-1}(\boldsymbol{I}-\boldsymbol{\Phi})||$ is a continuous function on the compact set Θ, thus

$$\mathsf{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial\boldsymbol{\mu}\partial\boldsymbol{\mu}'}\right\| < \infty. \tag{23}$$

Next, let us consider

$$\frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}'} = -(\boldsymbol{y}_{t-1}-\boldsymbol{\mu})(\boldsymbol{y}_{t-1}-\boldsymbol{\mu})'\otimes\boldsymbol{\Omega}^{-1}.$$

Then,

$$||(\boldsymbol{y}_{t-1}-\boldsymbol{\mu})(\boldsymbol{y}_{t-1}-\boldsymbol{\mu})'\otimes\boldsymbol{\Omega}^{-1}|| = ||\boldsymbol{y}_t-\boldsymbol{\mu}||^2||\boldsymbol{\Omega}^{-1}||.$$

Again, due to assumptions C1–C3, $||\boldsymbol{\Omega}^{-1}||$ is uniformly bounded on $\boldsymbol{\Theta}$. From the moving average representation (12),

$$||\boldsymbol{y}_{t-1}-\boldsymbol{\mu}|| \le \sum_{j=0}^{\infty}||\boldsymbol{\Phi}^j||\cdot||\boldsymbol{\epsilon}_{t-1-j}||.$$

From the Jordan decomposition of $\boldsymbol{\Phi}$ it follows that there exists a $0<\rho<1$ such that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}||\boldsymbol{\Phi}^j|| \le Cj^d\rho^j \tag{24}$$

with a positive constant C (for details see, e. g., Dvořák [7], Theorem 2.1). Thus, combining this result and the stationarity of $\{\boldsymbol{\epsilon}_t\}$, we can conclude that

$$\begin{aligned}\mathsf{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}||\boldsymbol{y}_{t-1}-\boldsymbol{\mu}||^2||\boldsymbol{\Omega}^{-1}|| &\le C\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}j^dk^d\rho^{j+k}\mathsf{E}\,||\boldsymbol{\epsilon}_{t-1-j}||.||\epsilon_{t-1-k}||\\ &\le C\left(\sum_{j=0}^{\infty}j^d\rho^j\right)^2\mathsf{E}\,||\boldsymbol{\epsilon}_1||^2<\infty.\end{aligned} \tag{25}$$

Concerning $\frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial\boldsymbol{\sigma}\partial\boldsymbol{\sigma}'}$, we have

$$\frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial\boldsymbol{\sigma}\partial\boldsymbol{\sigma}'} = \boldsymbol{D}_d'\frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial\boldsymbol{\omega}\partial\boldsymbol{\omega}'}\boldsymbol{D}_d,$$

$$\frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial\boldsymbol{\omega}\partial\boldsymbol{\omega}'} = \frac{1}{2}(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1}) - \frac{1}{2}(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'\boldsymbol{\Omega}^{-1}) - \frac{1}{2}(\boldsymbol{\Omega}^{-1}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1}) \tag{26}$$

where $\boldsymbol{\omega} = \mathsf{vec}(\boldsymbol{\Omega})$, see Lütkepohl [16], p. 91. From here we get

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} ||\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1}|| \le C,$$

$$||\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'\boldsymbol{\Omega}^{-1}|| = ||\boldsymbol{\Omega}^{-1}|| \cdot ||\boldsymbol{\Omega}^{-1}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'\boldsymbol{\Omega}^{-1}|| \le ||\boldsymbol{\Omega}^{-1}||^3||\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'|| = ||\boldsymbol{\Omega}^{-1}||^3||\boldsymbol{\epsilon}_t||^2.$$

The last term in (26) is treated in the same way, hence

$$\mathsf{E} \sup_{\boldsymbol{\theta}\in\Theta} \left\| \frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial\boldsymbol{\sigma}\partial\boldsymbol{\sigma}'} \right\| < \infty. \tag{27}$$

The remaining blocks are (cf. Lütkepohl [16], p. 91)

$$\frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial\boldsymbol{\mu}\partial\boldsymbol{\phi}'} = -(\boldsymbol{I}-\boldsymbol{\Phi})'\boldsymbol{\Omega}^{-1}((\boldsymbol{y}_{t-1}-\boldsymbol{\mu})\otimes\boldsymbol{I}) - (\boldsymbol{\epsilon}_t'\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{I})\frac{\partial\mathsf{vec}(\boldsymbol{\Phi}')}{\partial\boldsymbol{\phi}'},$$

$$\frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial\boldsymbol{\sigma}\partial\boldsymbol{\mu}'} = -\frac{1}{2}\boldsymbol{D}_d'(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1})[(\boldsymbol{I}\otimes\boldsymbol{\epsilon}_t)(\boldsymbol{I}-\boldsymbol{\Phi}) + (\boldsymbol{\epsilon}_t\otimes\boldsymbol{I})(\boldsymbol{I}-\boldsymbol{\Phi})],$$

$$\frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial\boldsymbol{\sigma}\partial\boldsymbol{\phi}'} = -\frac{1}{2}\boldsymbol{D}_d'(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1})[(\boldsymbol{I}\otimes\boldsymbol{\epsilon}_t(\boldsymbol{y}_{t-1}-\mu)')\frac{\partial\mathsf{vec}(\boldsymbol{\Phi}')}{\partial\boldsymbol{\phi}'} + (\boldsymbol{\epsilon}_t(\boldsymbol{y}_{t-1}-\boldsymbol{\mu})'\otimes\boldsymbol{I})]$$

and can be treated analogously. □

Condition (22) implies the uniform strong law of large numbers for strictly stationary and ergodic sequences of random elements $\{\boldsymbol{H}_t\}$ with values in the space of continuous functions on $\boldsymbol{\Theta}$, i.e.,

$$\sup_{\theta\in\boldsymbol{\Theta}} \left\| \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{H}_t(\boldsymbol{\theta}) - \mathsf{E}\,\boldsymbol{H}_1(\boldsymbol{\theta}) \right\| \xrightarrow{a.s.} 0 \tag{28}$$

as $n \to \infty$, see, e.g., Straumann [19], Chapter 2.2. Result (28) holds in the maximum norm, too.

Lemma 3.2. Under assumptions A1–A5 or B1–B5, together with C1–C3 and under H_0,

$$\sup_{1\le k<\infty} \frac{1}{m^{1/2}q(k/m)} \left| \sum_{i=m+1}^{m+k} \boldsymbol{G}_i(\widehat{\boldsymbol{\theta}}_m) - \left(\sum_{i=m+1}^{m+k} \boldsymbol{G}_i(\boldsymbol{\theta}_0) - \frac{k}{m}\sum_{i=1}^{m} \boldsymbol{G}_i(\boldsymbol{\theta}_0) \right) \right| = o_p(1). \tag{29}$$

P r o o f. Due to relation (10) we have

$$\sum_{t=m+1}^{m+k} \boldsymbol{G}_t(\widehat{\boldsymbol{\theta}}_m) - \left(\sum_{t=m+1}^{m+k} \boldsymbol{G}_t(\boldsymbol{\theta}_0) - \frac{k}{m}\sum_{t=1}^{m} \boldsymbol{G}_t(\boldsymbol{\theta}_0) \right)$$
$$= \sum_{t=m+1}^{m+k} (\boldsymbol{G}_t(\widehat{\boldsymbol{\theta}}_m) - \boldsymbol{G}_t(\boldsymbol{\theta}_0)) - \frac{k}{m}\sum_{t=1}^{m} (\boldsymbol{G}_t(\widehat{\boldsymbol{\theta}}_m) - (\boldsymbol{G}_t(\boldsymbol{\theta}_0)). \tag{30}$$

By using the mean value theorem for the jth component G_{tj} of the gradient vector $\boldsymbol{G}_t(\boldsymbol{\theta})$ we get

$$G_{tj}(\widehat{\boldsymbol{\theta}}_m) - G_{tj}(\boldsymbol{\theta}_0) = \boldsymbol{H}_{tj}(\widetilde{\boldsymbol{\theta}}_{tj})(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0) \tag{31}$$

where $\boldsymbol{H}_{tj}(\boldsymbol{\theta})$ means the jth row of the matrix $\boldsymbol{H}_t(\boldsymbol{\theta})$ and $\widetilde{\boldsymbol{\theta}}_{tj}$ lies between $\boldsymbol{\theta}_0$ and $\widehat{\boldsymbol{\theta}}_m$, $|\widetilde{\boldsymbol{\theta}}_{tj} - \boldsymbol{\theta}_0| \leq |\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0|$. Denoting the jth row of $\boldsymbol{Z}(\boldsymbol{\theta}) = \mathsf{E}\,\boldsymbol{H}_1(\boldsymbol{\theta})$ by $\boldsymbol{Z}_j(\boldsymbol{\theta})$, we can further write

$$\begin{aligned}
&\sum_{t=m+1}^{m+k} (\boldsymbol{G}_{tj}(\widehat{\boldsymbol{\theta}}_m) - \boldsymbol{G}_{tj}(\boldsymbol{\theta}_0)) - \frac{k}{m}\sum_{t=1}^{m}(\boldsymbol{G}_{tj}(\widehat{\boldsymbol{\theta}}_m) - (\boldsymbol{G}_{tj}(\boldsymbol{\theta}_0)) \\
&= \left[\sum_{t=m+1}^{m+k} (\boldsymbol{H}_{tj}(\widetilde{\boldsymbol{\theta}}_{tj}) - \boldsymbol{Z}_j(\widetilde{\boldsymbol{\theta}}_{tj})) - \frac{k}{m}\sum_{t=1}^{m}(\boldsymbol{H}_{tj}(\widetilde{\boldsymbol{\theta}}_{tj}) - \boldsymbol{Z}_j(\widetilde{\boldsymbol{\theta}}_{tj}))\right](\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0).
\end{aligned} \tag{32}$$

Combining this last relation with the uniform strong law of large numbers (28) and the fact that $|\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0| = O_p(m^{-1/2})$, which follows from Lemma 1, we obtain

$$\begin{aligned}
&\sup_{1\leq k<\infty} \frac{1}{m^{1/2}q(k/m)}\left|\sum_{i=m+1}^{m+k} \boldsymbol{G}_i(\widehat{\boldsymbol{\theta}}_m) - \left(\sum_{i=m+1}^{m+k} \boldsymbol{G}_i(\boldsymbol{\theta}_0) - \frac{k}{m}\sum_{i=1}^{m} \boldsymbol{G}_i(\boldsymbol{\theta}_0)\right)\right| \\
&\leq \max\left(\max_{1\leq k\leq m} \frac{1}{m^{1/2}q(k/m)} km^{-\frac{1}{2}}, \sup_{m+1\leq k<\infty} \frac{1}{m^{1/2}q(k/m)} km^{-\frac{1}{2}}\right) o_p(1) = o_p(1).
\end{aligned} \tag{33}$$

□

Lemma 3.3. Under assumptions A1–A5 or B1–B5, together with C1–C3 and under H_0, for every $K > 0$, as $m \to \infty$,

$$\left\{\frac{1}{\sqrt{m}}\boldsymbol{J}(\boldsymbol{\theta}_0)^{-1/2}\sum_{t=1}^{\lfloor m\tau\rfloor} \boldsymbol{G}_t(\boldsymbol{\theta}_0), \tau \in [0, K]\right\} \Longrightarrow \{\boldsymbol{W}_{\boldsymbol{I}}(\tau), \tau \in [0, K]\} \tag{34}$$

where $\{\boldsymbol{W}_{\boldsymbol{I}}(\tau)\}$ is a multivariate Gaussian process such that $\mathsf{E}\,\boldsymbol{W}_{\boldsymbol{I}}(s)\boldsymbol{W}'_{\boldsymbol{I}}(t) = \min(s,t)\boldsymbol{I}$ and $\Longrightarrow$ means the convergence in the Skorokhod space $\mathcal{D}^d[0, K]$.

P r o o f . Under assumptions A1–A5, the result follows from the functional central limit theorem for multivariate martingale differences (see, e. g., Davidson [6], Theorem 27.17), and under B1–B5 from the functional central limit theorem for strong mixing sequences of stationary random vectors, see Kuelbs and Philipp [13], Theorem 4. □

It follows from the previous lemma that

$$\frac{1}{\sqrt{m}}\sum_{t=1}^{m} \boldsymbol{G}_t(\boldsymbol{\theta}_0) = O_p(1). \tag{35}$$

P r o o f o f T h e o r e m 2.3.
From Lemma 3.2 it follows that

$$\max_{1\le k<\infty}\left|\frac{\boldsymbol{J}^{-1/2}(\boldsymbol{\theta}_0)}{\sqrt{m}q(m/k)}\sum_{t=m+1}^{m+k}\boldsymbol{G}_t(\widehat{\boldsymbol{\theta}}_m)\right|$$

$$=\max_{1\le k<\infty}\left|\frac{\boldsymbol{J}^{-1/2}(\boldsymbol{\theta}_0)}{\sqrt{m}q(m/k)}\left(\sum_{t=m+1}^{m+k}\boldsymbol{G}_t(\boldsymbol{\theta}_0)-\frac{k}{m}\sum_{t=1}^{m}\boldsymbol{G}_t(\boldsymbol{\theta}_0)\right)\right|+o_p(1). \tag{36}$$

From (34) we have for any $K>0$ and $\tau\in[0,K]$, as $m\to\infty$,

$$\frac{1}{\sqrt{m}}\boldsymbol{J}(\boldsymbol{\theta}_0)^{-1/2}\left(\sum_{t=m+1}^{m+\lfloor m\tau\rfloor}\boldsymbol{G}_t(\boldsymbol{\theta}_0)-\tau\sum_{t=1}^{m}\boldsymbol{G}_t(\boldsymbol{\theta}_0)\right)\Longrightarrow\boldsymbol{W}_{\boldsymbol{I}}(1+\tau)-(1+\tau)\boldsymbol{W}_{\boldsymbol{I}}(1). \tag{37}$$

By using the Hájek–Rényi inequality for vector mixingales (cf. Bai and Perron [1]) and proceeding as in Lemma 6.6 in Berkes et al. [2], we get

$$\max_{1\le k<\infty}\left|\frac{\boldsymbol{J}^{-1/2}(\boldsymbol{\theta}_0)}{\sqrt{m}q(m/k)}\sum_{t=m+1}^{m+k}\boldsymbol{G}_t(\widehat{\boldsymbol{\theta}}_m)\right|\xrightarrow{\mathcal{D}}\sup_{0<t<\infty}\left|\frac{\boldsymbol{W}_{\boldsymbol{I}}(1+t)-(1+t)\boldsymbol{W}_{\boldsymbol{I}}(1)}{q(t)}\right|$$

with $q(t)$ defined by (11). It follows from the properties of the Wiener process that $\boldsymbol{W}_{\boldsymbol{I}}(1+t)-(1+t)\boldsymbol{W}_{\boldsymbol{I}}(1)\stackrel{\mathcal{D}}{=}(1+t)\boldsymbol{W}(t/1+t)$ with $\{\boldsymbol{W}(t),t>0\}$ denoting the standard multivariate Brownian motion. Replacing $\boldsymbol{J}(\boldsymbol{\theta}_0)$ by $\widehat{\boldsymbol{J}}_m(\widehat{\boldsymbol{\theta}}_m)$ we conclude the proof of the first assertion after careful computations. The proof of the second assertion is similar. □

Concerning the estimator of the matrix $\boldsymbol{J}(\boldsymbol{\theta}_0)$, we can use the ergodic properties of the sequence $\{\boldsymbol{G}_t(\boldsymbol{\theta})\boldsymbol{G}_t(\boldsymbol{\theta})'\}$ and define

$$\widehat{\boldsymbol{J}}_m(\widehat{\boldsymbol{\theta}}_m)=\frac{1}{m}\sum_{t=1}^{m}\boldsymbol{G}_t(\widehat{\boldsymbol{\theta}}_m)\boldsymbol{G}_t(\widehat{\boldsymbol{\theta}}_m)'.$$

The block-diagonal property of the matrix $\boldsymbol{J}(\boldsymbol{\theta})$ enables to test a change either in all parameters or in $\boldsymbol{\mu},\boldsymbol{\Phi},\boldsymbol{\Omega}$ separately.

Another estimator of $\boldsymbol{J}(\boldsymbol{\theta}_0)$ can be found in Boubacar Mainassara and Francq [3].

P r o o f o f T h e o r e m 2.4.
First, notice that for $k^*<k$,

$$\sum_{t=m+1}^{m+k}\boldsymbol{G}_t(\widehat{\boldsymbol{\theta}}_m)=\sum_{t=m+1}^{m+k^*}\boldsymbol{G}_t(\widehat{\boldsymbol{\theta}}_m)+\sum_{t=m+k^*+1}^{m+k}\boldsymbol{G}_t(\widehat{\boldsymbol{\theta}}_m) \tag{38}$$

and the first sum satisfies the null hypothesis, thus, it suffices to study the second sum. Repeating the recursion in (17), we get for $j=1,\dots,k-k^*$

$$\boldsymbol{y}_{m+k^*+j}-\boldsymbol{\mu}_1=\sum_{\nu=0}^{j-1}\boldsymbol{\Phi}_1^{\nu}\boldsymbol{\epsilon}_{m+k^*+j-\nu}+\boldsymbol{\Phi}_1^{j}(\boldsymbol{y}_{m+k^*}-\boldsymbol{\mu}_1)$$

where $(\boldsymbol{y}_{m+k^*} - \boldsymbol{\mu}_1) = O_p(1)$ since $\boldsymbol{y}_{m+k^*}$ satisfies (12) under H_0 and we suppose finite means. From here we can conclude that $\boldsymbol{y}_{m+k^*+j} - \boldsymbol{\mu}_1$ can be approximated by $\widetilde{\boldsymbol{y}}_{m+k^*+j} - \boldsymbol{\mu}_1$ which is a stationary solution of (17), with the remainder $O_p(\boldsymbol{\Phi}_1^j)$. The gradient $\boldsymbol{G}_t(\boldsymbol{\theta})$ for $t = m + k^* + j, j = 1, \ldots, k - k^*$, can be approximated by the respective gradient $\widetilde{\boldsymbol{G}}_t(\boldsymbol{\theta})$ in the same way. Combining this with (24) we get

$$\sum_{t=m+k^*+1}^{m+k} \boldsymbol{G}_t(\boldsymbol{\theta}) = \sum_{t=m+k^*+1}^{m+k} \widetilde{\boldsymbol{G}}_t(\boldsymbol{\theta}) + O_p(1).$$

Since $\widetilde{\boldsymbol{G}}_t(\boldsymbol{\theta})$ satisfies the conditions of Lemma 3.1, we can apply the mean-value theorem (component-wise) to $\widetilde{\boldsymbol{G}}_t(\widehat{\boldsymbol{\theta}}_m)$ in a neighbourhood of $\boldsymbol{\theta}_0$. For the $j-$th component of $\widetilde{\boldsymbol{G}}_t(\boldsymbol{\theta})$ we get

$$\widetilde{G}_{tj}(\widehat{\boldsymbol{\theta}}_m) - \widetilde{G}_{tj}(\boldsymbol{\theta}_0) = \widetilde{\boldsymbol{H}}_{tj}(\overline{\boldsymbol{\theta}}_{tj})(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0) \tag{39}$$

where $\widetilde{\boldsymbol{H}}_{tj}(\boldsymbol{\theta})$ means the jth row of the matrix $\widetilde{\boldsymbol{H}}_t(\boldsymbol{\theta})$ defined as in Lemma 3.1 and $\overline{\boldsymbol{\theta}}_{tj}$ lies between $\boldsymbol{\theta}_0$ and $\widehat{\boldsymbol{\theta}}_m$, $|\overline{\boldsymbol{\theta}}_{tj} - \boldsymbol{\theta}_0| \le |\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0|$. Next, utilizing the ergodicity and (18) we have, as $m \to \infty$,

$$\sum_{t=m+k^*+1}^{m+k} \boldsymbol{G}_t(\widehat{\boldsymbol{\theta}}_m) = (k - k^*)(\mathsf{E}\,\widetilde{\boldsymbol{G}}_t(\boldsymbol{\theta}_0) + O_p(m^{-1/2})) + Op(1). \tag{40}$$

Then, assuming that $k^* = O(m)$, as $m \to \infty$

$$\sup_{1\le k<\infty} \left| \frac{\boldsymbol{J}^{-1/2}(\boldsymbol{\theta}_0)}{\sqrt{m}q(m/k)} \sum_{t=m+k^*+1}^{m+k} \boldsymbol{G}_t(\widehat{\boldsymbol{\theta}}_m) \right| \ge \sup_{k:k=m+k^*} \left| \frac{\boldsymbol{J}^{-1/2}(\boldsymbol{\theta}_0)}{\sqrt{m}q(m/k)} (k-k^*)[\mathsf{E}\,\widetilde{\boldsymbol{G}}_t(\boldsymbol{\theta}_0) + o_p(1)] \right|$$
$$= \sup_{k:k=m+k^*} \left(\frac{m}{k}\right)^{\gamma} \left(1 + \frac{m}{k}\right)^{\gamma-1} \sqrt{m} \left| \boldsymbol{J}^{-1/2}(\boldsymbol{\theta}_0)[\mathsf{E}\,\widetilde{\boldsymbol{G}}_t(\boldsymbol{\theta}_0) + o_p(1)] \right| \xrightarrow{P} \infty$$

which proves (19). The proof of (20) is analogous. □

4. SHORT SIMULATION STUDY

In this short simulation study we demonstrate the performance of the method. First, let us recall that we reject the null hypothesis at the first time k for which $|Q(m,k)| > q(k/m)c_\alpha$ where c_α is the critical value of $\sup_{0\le t\le 1} |\boldsymbol{W}_r(t)|/t^\gamma$ and $\boldsymbol{W}_r$ is the standard r-dimensional Wiener process. The distribution of the limiting process can be computed by using Monte Carlo method or by bootstrap. In Horváth et al. [10], approximate critical values of the limiting process are presented for $r = 1$ and various values of γ.

Since for the standard r-dimensional Wiener process with components $W_j(t)$

$$\sup_{0\le t\le 1} \frac{|\boldsymbol{W}_r(t)|}{t^\gamma} = \sup_{0\le t\le 1} \max_{1\le j\le r} \frac{|W_j(t)|}{t^\gamma} \stackrel{D}{=} \max_{1\le j\le r} \sup_{0\le t\le 1} \frac{|W_j(t)|}{t^\gamma},$$

we can determine the critical value c_α of $\sup_{0\le t\le 1} |\boldsymbol{W}_r(t)|/t^\gamma$ to be the critical value $c_\alpha^{(r)}$ of $\sup_{0\le t\le 1} \frac{|W_j(t)|}{t^\gamma}$ where $\alpha^{(r)} = 1 - (1-\alpha)^{1/r}$. In case of the closed-end procedure, we need to multiply this value by $(T/T+1)^{1/2-\gamma}$.

We have simulated a 2-dimensional VAR(1) process (1) with $\boldsymbol{\mu}_0 = (0.5\ 0.5)'$, and

$$\boldsymbol{\Phi}_0 = \begin{pmatrix} 0.5 & 0.2 \\ 0.2 & 0.1 \end{pmatrix} \tag{41}$$

where the random errors were generated from the 2-dimensional normal distribution with zero mean and variance matrix

$$\boldsymbol{\Omega}_0 = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}. \tag{42}$$

We used historical data of size $m = 100, 200$, and 500, respectively, and monitoring horizon mT for various choices of T. We considered a change in the mean only, here we present the results with $\boldsymbol{\mu}_1 = (2.5, 0.5)'$ based on 500 simulation runs. In Table 1, the empirical sizes of the test are given. In general, they are quite conservative, especially for the values of γ close to 0.5 which should be used when an early change is expected. Here NR means that the null hypothesis was not rejected during the monitoring period. It may be caused by the fact that the used asymptotic critical value is very strict and the monitoring period is short with respect to computational complexity. In Tables 2–3, basic summary statistics for the stopping time (minimum and maximum value, mean, and 25% and 75% percentiles) are displayed for different values of k^*. It is seen that the best results are obtained by using $\gamma = 0.49$ which is recommended in case we expect that a break appears early after the beginning of the monitoring, while $\gamma = 0$ is designed for late changes. Estimated densities of the stopping time in dependence on the historical period are displayed in Figures 1–2.

m	T	$\gamma = 0$	$\gamma = 0.25$	$\gamma = 0.49$
200	2	0.02	0.01	NR
	3	0.07	0.03	0.02
	4	0.06	0.02	0.01
	5	0.02	0.02	0.01
m	T	$\gamma = 0$	$\gamma = 0.25$	$\gamma = 0.49$
500	2	0.02	0.01	0.01
	3	0.02	0.02	0.01
	4	0.03	0.02	0.01
	5	0.03	0.03	0.02

Tab. 1. Empirical level of test statistic, change in the mean, nominal level $\alpha = 5\%$.

ACKNOWLEDGEMENT

This work was partially supported by the Czech Science Foundation project No. P402/12/G097 DYME – Dynamic Models in Economics. The author would like to thank the referees for their suggestions to improve the paper.

(Received December 10, 2017)

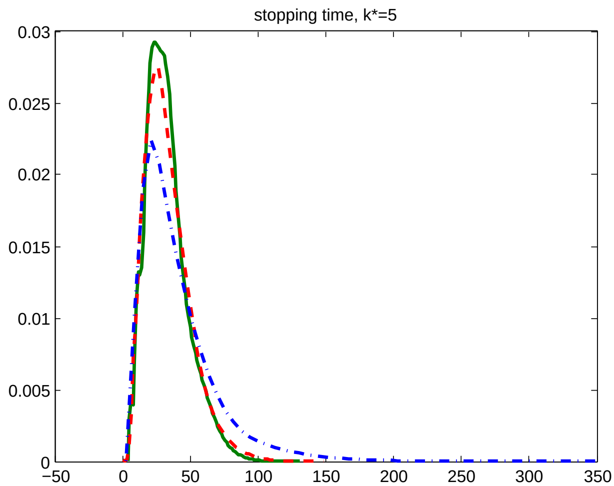

Fig. 1. Densities of stopping times, $k^* = 5, \gamma = 0.49, T = 5$, $m = 500$ (solid), $m = 200$ (dash), $m = 100$ (dash-dot).

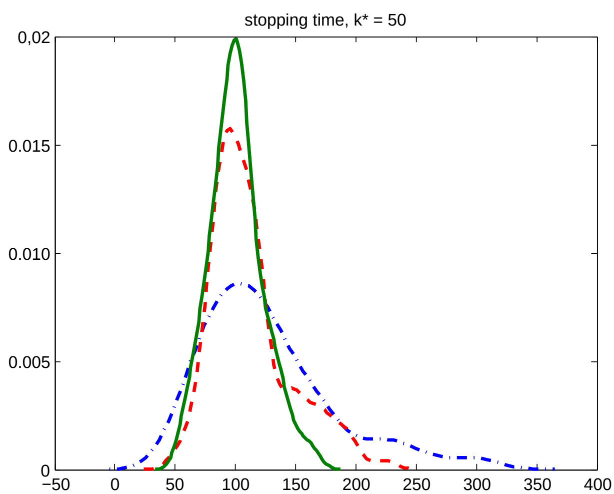

Fig. 2. Densities of stopping times, $k^* = 50, \gamma = 0.49, T = 5$, $m = 500$ (solid), $m = 200$ (dash), $m = 100$ (dash-dot).

γ	T	min	25%	med	75%	max	mean
0	2	53.00	70.50	77.50	89.00	118.00	79.85
	3	51.00	72.50	82.00	96.50	122.00	84.14
	4	59.00	78.50	88.00	96.50	140.00	87.92
	5	61.00	77.00	90.00	103.50	131.00	91.19
0.25	2	22.00	43.00	52.00	61.00	78.00	51.92
	3	23.00	43.00	49.00	60.50	110.00	51.65
	4	25.00	45.00	53.00	61.00	90.00	53.71
	5	27.00	46.00	54.00	59.50	109.00	52.74
0.49	2	11.00	22.50	31.50	38.00	67.00	32.27
	3	6.00	22.50	30.50	39.00	68.00	31.61
	4	10.00	21.00	29.00	39.50	77.00	32.40
	5	10.00	22.00	30.50	39.50	114.00	32.52

Tab. 2. distribution of stopping time, change in the mean, $m = 500, k^* = 5$.

γ	T	min	25%	med	75%	max	mean
0	2	90.00	114.00	126.50	140.50	169.00	127.76
	3	90.00	121.00	131.50	142.50	185.00	132.54
	4	99.00	127.50	136.00	151.00	189.00	139.04
	5	90.00	124.00	137.00	156.50	195.00	139.27
0.25	2	72.00	94.50	105.00	121.50	177.00	107.65
	3	71.00	99.50	108.50	119.00	157.00	109.46
	4	71.00	95.50	107.00	125.50	161.00	110.02
	5	76.00	94.50	108.50	121.00	169.00	109.41
0.49	2	54.00	83.00	92.00	109.50	189.00	96.25
	3	57.00	81.00	94.00	109.00	151.00	96.75
	4	60.00	86.50	99.50	118.50	152.00	103.00
	5	57.00	85.00	95.50	110.00	165.00	98.26

Tab. 3. distribution of stopping time, change in the mean, $m = 500, k^* = 50$.

REFERENCES

[1] J. Bai and P. Perron: Estimating and testing linear models with multiple structural changes. Econometrica *66* (1998), 47–78. DOI:10.2307/2998540

[2] I. Berkes, E. Gombay, L. Horváth, and P. Kokoszka: Sequential change-point detection in GARCH(p,q) models. Econometr. Theory *20* (2004), 1140–1167. DOI:10.1017/s0266466604206041

[3] Y. Boubacar Mainassara and C. Francq: Estimating structural VARMA models with un-

correlated but non-independent error terms. J. Multivariate Anal. *102* (2011), 496–505. DOI:10.1016/j.jmva.2010.10.009

[4] F. Carsoule and P. H. Franses: A note on monitoring time-varying parameters in an autoregression. Metrika *57* (2003), 51-62. DOI:10.1007/s001840200198

[5] C.-S. J. Chu, M. Stinchcombe, and H. White: Monitoring structural change. Econometrica *64* (1996), 1045-1065. DOI:10.2307/2171955

[6] J. Davidson: Stochastic Limit Theory. Oxford University Press, Oxford 1994. DOI:10.1093/0198774036.001.0001

[7] M. Dvořák: Stability in Autoregressive Time Series Models. Ph.D. Dissertation, Charles University, Prague 2015.

[8] E. Gombay and D. Serban: Monitoring parameter change in AR(p) time series models. J. Multivariate Anal. *100* (2009), 715-725. DOI:10.1016/j.jmva.2008.08.005

[9] J. D. Hamilton: Time Series Analysis. Princeton University Press, Princeton 1994. DOI:10.1017/s0266466600009440

[10] L. Horváth, M. Hušková, P. Kokoszka, and J. Steinebach: Monitoring changes in linear models. J. Statist. Plann. Inference *126* (2004), 225-251. DOI:10.1016/j.jspi.2003.07.014

[11] Z. Hlávka, M. Hušková, C. Kirch, and S. G. Meintanis: Monitoring changes in the error distribution of autoregressive models based on Fourier methods. Test *21* (2012), 605–634. DOI:10.1007/s11749-011-0265-z

[12] M. Hušková and A. Koubková: Sequential procedures for detection of changes in autoregressive sequences. In: Proc. Prague Stochastics (M. Hušková and M. Janžura, eds.), MATFYZPRESS, Charles University, Prague 2006, pp. 437-447.

[13] J. Kuelbs and W. Philipp: Almost sure invariance principles for partial sums of mixing $B-$valued random variables. Ann. Probab. *8* (1980), 1003-1036. DOI:10.1214/aop/1176994565

[14] F. Li, Z. Tian, and P. Qi: Structural change monitoring for random coefficient autoregressive time series. Comm. Statist. Simulation Comput. *44* (2015), 996–1009. DOI:10.1080/03610918.2013.800205

[15] S. Lee, Y. Lee, and O. Na: Monitoring distributional changes in autoregressive models Comm. Statist. Theory Methods *38* (2009), 2969-2982. DOI:10.1080/03610920902947261

[16] H. Lütkepohl: New Introduction to Multiple Time Series Analysis. Springer-Verlag, Berlin 2005. DOI:10.1007/978-3-540-27752-1

[17] O. Na, J. Lee, and S. Lee: Monitoring parameter changes for random coefficient autoregressive models. J. Korean Statist. Soc. *39* (2010), 281-288. DOI:10.1016/j.jkss.2010.03.006

[18] Z. Prášková : Monitoring changes in RCA models. In: Stochastic Models, Statistics and Their Applications, Springer Proceedings in Mathematics and Statistics *122* (A. Steland, E. Rafajlowicz and K. Szajowski, eds.), Springer, Berlin 2015, pp. 129–137. DOI:10.1007/978-3-319-13881-7_15

[19] D. Straumann: Estimation in Conditionally Heteroscedastic Time Series Models. Lecture Notes in Statistics *181*, Springer, Berlin 2005. DOI:10.1007/b138400

Zuzana Prášková, Charles University, Faculty of Mathematics and Physics, Sokolovská 83, 186 75 Praha 8. Czech Republic.
e-mail: praskova@karlin.mff.cuni.cz