Bernoulli **21**(2), 2015, 1238–1259
DOI: 10.3150/14-BEJ604

Reaction times of monitoring schemes for ARMA time series

ALEXANDER AUE[1,*], CHRISTOPHER DIENES[1,**], STEFAN FREMDT[2] and JOSEF STEINEBACH[3]

[1]*Department of Statistics, University of California, One Shields Avenue, Davis, CA 95616, USA.*
E-mail: [*]aaue@ucdavis.edu; [**]crdienes@ucdavis.edu
[2]*Department of Mathematics, Institute of Statistics, Ruhr-University Bochum, D-44780 Bochum, Germany.*
E-mail: stefan.fremdt@ruhr-uni-bochum.de
[3]*Mathematisches Institut, Universität zu Köln, Weyertal 86-90, D-50931 Köln, Germany.*
E-mail: jost@math.uni-koeln.de

This paper is concerned with deriving the limit distributions of stopping times devised to sequentially uncover structural breaks in the parameters of an autoregressive moving average, ARMA, time series. The stopping rules are defined as the first time lag for which detectors, based on CUSUMs and Page's CUSUMs for residuals, exceed the value of a prescribed threshold function. It is shown that the limit distributions crucially depend on a drift term induced by the underlying ARMA parameters. The precise form of the asymptotic is determined by an interplay between the location of the break point and the size of the change implied by the drift. The theoretical results are accompanied by a simulation study and applications to electroencephalography, EEG, and IBM data. The empirical results indicate a satisfactory behavior in finite samples.

Keywords: CUSUM statistic; on-line monitoring; Page's CUSUM; structural break detection

1. Introduction

Sequential change-point analysis is concerned with uncovering in an on-line fashion what is called structural breaks, deviations from a pre-specified in-control scenario. For the case of time series, relevant for this paper, the natural in-control scenario is the stationarity of the underlying stochastic process. More traditionally, sequential change-point techniques were developed for breaks in the mean and variance in sequences of independent observations. The corresponding literature is reviewed in the monographs Basseville and Nikiforov [6] and Csörgő and Horváth [12]. A more recent survey of both sequential and historical procedures is given in Aue and Horváth [3].

The particular approach to sequential change-point analysis of this paper is grounded in the work of Chu *et al.* [11], who developed procedures using a training sample to estimate an initial model and to monitor for deviations from that model as soon as new observations arrive. This contribution, originally written with applications to econometric data in mind, has been extended in a number of ways. Further sequential procedures covering financial time series were discussed in Andreou and Ghysels [1], and Aue *et al.* [4]. Berkes *et al.* [8] introduced methodology applicable to GARCH processes. Gombay and Serban [18] worked with autoregressive processes, while Gombay and Horváth [17] considered weakly stationary time series. Refinements using bootstrap were considered in Kirch [23] and Hušková and Kirch [20], while resampling schemes were studied by Hušková *et al.* [21].

The basic time series model being utilized in this paper is the class of linear autoregressive moving average, ARMA, processes made popular through the works of Box *et al.* [9]. ARMA processes find widespread applications in a number of fields as evidenced, for example, in the recent text Shumway and Stoffer [29]. As advocated by Brown *et al.* [10]) in a regression setting, the proposed monitoring procedures are based on the residuals obtained from an ARMA model fit to the original data based on a training sample of size m for which stationarity of the underlying process is assumed. If the process remains stationary after the monitoring starts, then residuals of the training period and the monitoring period should possess similar properties. The test procedures to be introduced here are based on traditional cumulative sum, CUSUM, statistics and a modification, Page's CUSUM statistics (see Page [26,27]). The latter tend to react faster to deviations from the in-control scenario and satisfy certain optimality criteria (see Lorden [24]). CUSUMs for residuals of ARMA processes were discussed in a retrospective setting in Bai [5], Yu [34] and Robbins *et al.* [28], and in a sequential framework in Dienes and Aue [14]. Recent work on Page's CUSUMs can be found in Fremdt [15,16].

A stopping rule is then defined as a first crossing time, that is, the time lag for which either the CUSUM or Page's CUSUM statistic exceed a threshold value tolerable for the in-control case. The focus of this paper is on deriving the asymptotic distributions of these stopping rules for the situation that deviations from stationarity of the underlying process occur. The particular deviations of interest are the classic change in mean and general changes in the second-order dynamics, with an emphasis on changes in the variance (or scale) due to the nature of the data examples provided in this paper. Namely, the finite-sample properties of the proposed methods are discussed in two case studies. The first of the applications involves EEG data. Here interest is in detecting the occurrence of an epileptic seizure (see Davis *et al.* [13]). The second application deals with closing prices of IBM stock, a classic data set that has been analyzed with historical procedures for the presence of breaks in variance (see, e.g., Tsay [31]). Accompanying simulation evidence indicates that the procedure works satisfactory for these two examples.

The paper is organized as follows. Section 2 details the ARMA model and states the hypotheses to be tested. Section 3 quantifies the large-sample behavior of the delay times incurred by the CUSUM and Page's CUSUM procedure. Applications to EEG and IBM data are discussed in Section 4. All proofs are given in Section 5.

2. The model

Let $\mathbb{Z}$ denote the set of integers. In what follows, $(Y_t : t \in \mathbb{Z})$ denotes the ARMA(p,q) process specified by the stochastic recurrence equations

$$\phi_t(B)(Y_t - \mu_t) = \theta_t(B)\varepsilon_t, \qquad t \in \mathbb{Z}, \tag{2.1}$$

where μ_t are mean parameters, $\phi_t(z) = 1 - \phi_{t,1}z - \cdots - \phi_{t,p}z^p$ and $\theta_t(z) = 1 + \theta_{t,1}z + \cdots + \theta_{t,q}z^q$ denote respectively the autoregressive and moving average polynomials, and B the backshift operator. The innovations $(\varepsilon_t : t \in \mathbb{Z})$ are assumed to be independent random variables with zero mean and variance σ_t^2. As usual, it is further required that ϕ_t and θ_t have no common zeroes and that the ARMA process is causal and invertible, which means

$$\phi_t(z) \neq 0 \quad \text{and} \quad \theta_t(z) \neq 0 \qquad \text{for all } |z| \leq 1. \tag{2.2}$$

The parametric model in (2.1) depends on the parameter vectors $\boldsymbol{\xi}_t = (\mu_t, \boldsymbol{\phi}_t, \boldsymbol{\theta}_t, \sigma_t)'$, where $\boldsymbol{\phi}_t = (\phi_{t,1}, \ldots, \phi_{t,p})'$ and $\boldsymbol{\theta}_t = (\theta_{t,1}, \ldots, \theta_{t,q})'$, with $'$ denoting transposition. These vectors may be time dependent and interest is in monitoring the constancy of the $\boldsymbol{\xi}_t$ in a sequential fashion. This is important because constancy of the $\boldsymbol{\xi}_t$ would imply stationarity of the underlying ARMA process, so that standard methods are available for estimation and prediction purposes. To set up the monitoring, a training period of size $m + p$ is utilized for which

$$Y_{1-p}, \ldots, Y_m \text{ are governed by } \boldsymbol{\xi}_t = \boldsymbol{\xi}_0 = (\mu_0, \boldsymbol{\phi}_0, \boldsymbol{\theta}_0, \sigma_0)'. \tag{2.3}$$

As Chu *et al.* [11] elaborate, this training period may be used to estimate the parameters of an initial non-contaminated model and to express limit results in the form $m \to \infty$. In particular, let $(X_t : t \in \mathbb{Z})$ be the centered sequence defined by $X_t = Y_t - \mu_t$ and define $\hat{\boldsymbol{\xi}}_m = (\hat{\mu}_m, \hat{\boldsymbol{\phi}}_m, \hat{\boldsymbol{\theta}}_m, \hat{\sigma}_m)'$ to be a $\sqrt{m}$-consistent estimator for $\boldsymbol{\xi}_0$ obtained from the training period data. This gives the model residuals

$$\hat{\varepsilon}_t = \hat{X}_t - \sum_{j=1}^{p} \hat{\phi}_{m,j} \hat{X}_{t-j} - \sum_{j=1}^{q} \hat{\theta}_{m,j} \hat{\varepsilon}_{t-j},$$

with $\hat{X}_t = Y_t - \hat{\mu}_m$ and initializations $\hat{\varepsilon}_{-q+1} = \cdots = \hat{\varepsilon}_0 = 0$ in case $q > 0$.

In the following, two sets of hypotheses will be considered. First, the focus will be on the arguably most studied case for which only mean breaks are permitted. The sequential testing problem then becomes

$$\begin{aligned} H_0 &: Y_{m+1}, Y_{m+2}, \ldots \text{ have mean } \mu_0; \\ H_A^{\mu} &: Y_{m+1}, \ldots, Y_{m+k^*-1} \text{ have mean } \mu_0, \text{ but} \\ &\quad Y_{m+k^*}, Y_{m+k^*+1}, \ldots \text{ have mean } \mu_A \neq \mu_0, \end{aligned}$$

where here the constancy of the remaining model ARMA parameters is required, so that changes may only affect the mean. It may sometimes be of greater importance to test for changes in the underlying second-order dynamics. This can be done via testing the general sequential hypotheses

$$\begin{aligned} H_0 &: Y_{m+1}, Y_{m+2}, \ldots \text{ are governed by } \boldsymbol{\xi}_0; \\ H_A^{\boldsymbol{\xi}} &: Y_{m+1}, \ldots, Y_{m+k^*-1} \text{ are governed by } \boldsymbol{\xi}_0, \text{ but} \\ &\quad Y_{m+k^*}, Y_{m+k^*+1}, \ldots \text{ are governed by } \boldsymbol{\xi}_A \neq \boldsymbol{\xi}_0. \end{aligned}$$

Under $H_A^{\boldsymbol{\xi}}$ the decomposition $\boldsymbol{\xi}_A = \boldsymbol{\xi}_0 + \boldsymbol{\delta}_m^{\boldsymbol{\xi}}$ will be utilized, where $\boldsymbol{\delta}_m^{\boldsymbol{\xi}} = (\delta_m^{\mu}, \boldsymbol{\delta}_m^{\phi}, \boldsymbol{\delta}_m^{\theta}, \delta_m^{\sigma})'$ denotes the difference in parameter values.

For both sets of hypotheses, one can now proceed as follows. If the respective null scenarios hold, then the residuals $\hat{\varepsilon}_t$ should roughly resemble the corresponding innovations ε_t and suitably constructed statistics should therefore behave similarly on the training period and after monitoring commences. Under the alternatives, this should not be the case. This approach will be detailed in the next section.

3. Monitoring schemes and their large-sample properties

3.1. CUSUM and Page's CUSUM procedures under the null

Testing procedures for the set of hypotheses introduced in the previous section are commonly defined as stopping times that reject the null if a detector crosses the boundary prescribed by a threshold function. Popular choices for the detector are based on cumulative sum, CUSUM, statistics and on its variant, called Page's CUSUM. Let $\mathbb{N}$ denote the positive integers. To introduce the CUSUM of (squared) residual procedures, define for $k \in \mathbb{N}$ the detectors

$$\hat{D}_\mu(m,k) = \sum_{t=m+1}^{m+k} \hat{\varepsilon}_t - \frac{k}{m}\sum_{t=1}^{m} \hat{\varepsilon}_t \quad \text{and} \quad \hat{D}_{\boldsymbol{\xi}}(m,k) = \sum_{t=m+1}^{m+k} \hat{\varepsilon}_t^2 - \frac{k}{m}\sum_{t=1}^{m} \hat{\varepsilon}_t^2. \tag{3.1}$$

The detector $\hat{D}_\mu(m,k)$ is built from the residuals $\hat{\varepsilon}_t$ and used to test H_0 against H_A^μ, while the detector $\hat{D}_{\boldsymbol{\xi}}$ is built from the squared residuals $\hat{\varepsilon}_t^2$ and used to test H_0 against $H_A^{\boldsymbol{\xi}}$. Using the class of weight functions

$$g_\gamma(m,k) = \sqrt{m}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^\gamma, \tag{3.2}$$

indexed by a sensitivity parameter $\gamma \in [0, 1/2)$, a stopping time corresponding to the detector $\hat{D}_\mu(m,k)$ can be defined by

$$\tau_\mu(m) = \min\{k \in \mathbb{N} : |\hat{D}_\mu(m,k)| \geq c_\alpha \hat{\sigma}_m g_\gamma(m,k)\}, \tag{3.3}$$

where $c_\alpha = c_\alpha(\gamma)$ is a critical constant, derived from the limit distribution of the detector under H_0 (see Theorem 3.1 below), ensuring that $P(\tau_\mu(m) < \infty) = \alpha$ for a given level $\alpha \in (0,1)$.

The stopping time $\tau_{\boldsymbol{\xi}}(m)$ for the detector $\hat{D}_{\boldsymbol{\xi}}(m,k)$ is defined analogously: Let $\hat{\eta}_m^2$ denote a weakly consistent estimator of the quantity $\eta^2 = \mathrm{E}[(\varepsilon_1^2 - \sigma^2)^2]$. Then $\tau_{\boldsymbol{\xi}}(m)$ is given by replacing $\hat{D}_\mu(m,k)$ and $\hat{\sigma}_m$ with $\hat{D}_{\boldsymbol{\xi}}(m,k)$ and $\hat{\eta}_m$, respectively.

Page's CUSUM procedure is a modification of the CUSUM detectors in (3.1) based on the adjusted detectors

$$\begin{aligned} \hat{D}_\mu^P(m,k) &= \max_{0\leq k'\leq k} |\hat{D}_\mu(m,k) - \hat{D}_\mu(m,k')|, \\ \hat{D}_{\boldsymbol{\xi}}^P(m,k) &= \max_{0\leq k'\leq k} |\hat{D}_{\boldsymbol{\xi}}(m,k) - \hat{D}_{\boldsymbol{\xi}}(m,k')|, \end{aligned} \tag{3.4}$$

setting $\hat{D}_\mu(m,0) = \hat{D}_{\boldsymbol{\xi}}(m,0) = 0$. Utilizing the same class of weight functions in (3.2) as before gives rise to the Page-type stopping time

$$\tau_\mu^P(m) = \min\{k \in \mathbb{N} : \hat{D}_\mu^P(m,k) \geq c_\alpha^P \hat{\sigma}_m g_\gamma(m,k)\}, \tag{3.5}$$

where $c_\alpha^P = c_\alpha^P(\gamma)$ controls again the level of the sequential procedure. The stopping time $\tau_{\boldsymbol{\xi}}^P(m)$ is defined in a similar fashion. These sequential detectors were introduced in the seminal papers (Page [26,27]).

All procedures are based on residuals instead of directly on the observations. This has the advantage that the notoriously difficult estimation of long-run variances of the dependent observations can be completely avoided. Better size and power properties are expected from this approach as pointed out in Robbins *et al.* [28], who confirmed these statements in an extensive simulation study.

The large-sample behavior under the null hypotheses for the four detectors is quantified in the following two theorems, the first one of which states the results for the mean only procedures.

Theorem 3.1. *Let $(Y_t : t \in \mathbb{Z})$ follow the ARMA equations (2.1) and assume that $\mathrm{E}[|\varepsilon_1|^\nu] < \infty$ for some $\nu > 2$. Then it holds under H_0 and for all real c that*

(a) $$\lim_{m\to\infty} P\left(\frac{1}{\hat{\sigma}_m} \sup_{k\geq 1} \frac{|\hat{D}_\mu(m,k)|}{g_\gamma(m,k)} \leq c\right) = P\left(\sup_{0<x<1} \frac{|W(x)|}{x^\gamma} \leq c\right),$$

(b) $$\lim_{m\to\infty} P\left(\frac{1}{\hat{\sigma}_m} \sup_{k\geq 1} \frac{\hat{D}_\mu^P(m,k)}{g_\gamma(m,k)} \leq c\right) = P\left(\sup_{0<x<1} \sup_{0\leq y\leq x} \frac{1}{x^\gamma}\left|W(x) - \frac{1-x}{1-y}W(y)\right| \leq c\right),$$

where $(W(x) : x \in [0,1])$ denotes a standard Brownian motion.

Theorem 3.2. *Let $(Y_t : t \in \mathbb{Z})$ follow the ARMA equations (2.1) and assume that $\mathrm{E}[|\varepsilon_1|^\nu] < \infty$ for some $\nu > 4$. Then, under H_0 and for all real c, the limit results of Theorem 3.1 are retained if $\hat{D}_\mu(m,k)$, $\hat{D}_\mu^P(m,k)$ and $\hat{\sigma}_m$ are replaced with the respective objects $\hat{D}_{\boldsymbol{\xi}}(m,k)$, $\hat{D}_{\boldsymbol{\xi}}^P(m,k)$ and $\hat{\eta}_m$.*

The proofs of the theorems follow from the results in Dienes and Aue [14] for the CUSUM procedure, and from a combination of the latter with the proofs in Fremdt [16] for Page's CUSUM procedure. Tables containing simulated critical values for a selection of sensitivity parameters γ and test levels α can be found in Horváth *et al.* [19] for the limit in Theorem 3.1, part (a) and in Fremdt [16] for the limit in part (b).

3.2. Limiting delay times for mean breaks

The quality of monitoring procedures is often quantified via the mean delay time which measures how long, on average, one has to wait before the structural break in the underlying processes is detected. For example, certain optimality criteria for Page's CUSUM were developed in Lorden [24]. The monograph by Basseville and Nikiforov [6] gives an account of the subsequent contributions in this area. The main theoretical contribution of this paper is the derivation of the complete limit distribution of the stopping times under consideration. Taking the mean of this distribution, one obtains in particular also the information on the average delay time. Related results in the literature are Aue and Horváth [2], Aue *et al.* [4] and Fremdt [16]. To account for the ARMA time series character, modifications of the methodology in these papers become necessary. These will be developed in the following.

It is subsequently assumed that H_A^μ holds and that thus changes in the second-order structure of the ARMA process do not occur. Notice that assumption (2.2) implies that the reciprocals of $\phi_t(z)$ and $\theta_t(z)$ admit, for $|z| \le 1$, the power series expansions

$$\frac{1}{\phi_t(z)} = \sum_{\ell=0}^{\infty} \pi_\ell(\boldsymbol{\phi}_t) z^\ell \quad \text{and} \quad \frac{1}{\theta_t(z)} = \sum_{\ell=0}^{\infty} \psi_\ell(\boldsymbol{\theta}_t) z^\ell. \tag{3.6}$$

Denoting the training period estimates of the autoregressive and moving average polynomials by $\hat{\phi}_m(z)$ and $\hat{\theta}_m(z)$, for large enough m, one finds analogously power series expansions for their reciprocals. These will be written as

$$\frac{1}{\hat{\phi}_m(z)} = \sum_{\ell=0}^{\infty} \pi_\ell(\hat{\boldsymbol{\phi}}_m) z^\ell \quad \text{and} \quad \frac{1}{\hat{\theta}_m(z)} = \sum_{\ell=0}^{\infty} \psi_\ell(\hat{\boldsymbol{\theta}}_m) z^\ell. \tag{3.7}$$

Under H_A^μ, the asymptotic behavior of the delay time will depend on the size of the mean change $\delta_m^\mu = \mu_A - \mu_0$ which in turn induces the drift term

$$\Delta_m^\mu = \delta_m^\mu \left(1 - \sum_{j=1}^{p} \phi_{0,j}\right) \sum_{\ell=0}^{\infty} \psi_\ell(\boldsymbol{\theta}_0) = \delta_m^\mu \frac{\phi_0(1)}{\theta_0(1)}. \tag{3.8}$$

Note that the difference of pre-mean and post-mean is allowed to depend on m, so that one could more explicitly write $\mu_{A,m}$. The precise limit distribution will crucially depend on the interplay between the behavior of the drift term Δ_m^μ and the location of the mean change k^*. This leads to the following set of assumptions which, in view of the theorems to come, are formulated for a general sequence Δ_m and not directly for Δ_m^μ. Superscripts, such as μ here, will indicate which drift term is being used.

Assumption 3.1. *It is required that*

(a) *there is* $\theta > 0$ *such that* $k^* = \lfloor \theta m^\beta \rfloor$ *with* $\beta \in [0, 1)$, *where* $\lfloor \cdot \rfloor$ *denotes integer part*;
(b) $\sqrt{m}|\Delta_m| \to \infty$;
(c) $|\Delta_m| = \mathrm{O}(1)$.

Part (a) of Assumption 3.1 specifies the order of the change-point k^* as a power of m. It is a standard assumption in the change-point literature. However, it should be noted that the expression $k^* = \lfloor \theta m^\beta \rfloor$ is not unique for fixed m and k^*, and different specifications of θ and β may lead to different limit distributions. A discussion of this matter can be found in Section 3 of Fremdt [16]. Note also that parts (b) and (c) implicitly allow for the decay of the sequence $|\Delta_m|$. The proofs show that the form of the limit distribution of the stopping times depends then on the asymptotic behavior of the sequence $|\Delta_m| m^{\gamma-1/2} k^{*1-\gamma}$ of scaled drift terms. Due to part (a) of Assumption 3.1 which allows for the re-expression of k^* in terms of m, they depend consequently on the asymptotic behavior of the scaled terms

$$\tilde{\Delta}_m = |\Delta_m| m^{\beta(1-\gamma)-1/2+\gamma},$$

which do not explicitly contain k^* anymore. We distinguish between the three cases

$$\text{(i) } \tilde{\Delta}_m \to 0, \qquad \text{(ii) } \tilde{\Delta}_m \to \tilde{C}_1 \in (0,\infty), \qquad \text{(iii) } \tilde{\Delta}_m \to \infty.$$

In case (ii), it follows from part (a) of Assumption 3.1 that $|\Delta_m| m^{\gamma-1/2} k^{*1-\gamma} \to \theta^{1-\gamma}\tilde{C}_1 = C_1 \in (0,\infty)$. For this scenario and any real c define $d_1 = d_1(c)$ to be the unique solution of

$$d_1 = 1 - \frac{c}{C_1} d_1^{1-\gamma}. \tag{3.9}$$

In order to exhibit the asymptotic distribution of the stopping times, introduce first the case-dependent distribution function Ψ by setting, for all real arguments u,

$$\Psi(u) = \begin{cases} \Phi(u), & \text{in case (i)},\\ P\Big(\sup\limits_{d_1<x<1} W(x) \le u\Big), & \text{in case (ii)},\\ P\Big(\sup\limits_{0<x<1} W(x) \le u\Big) = \begin{cases} 0, & u<0,\\ 2\Phi(u)-1, & u \ge 0,\end{cases} & \text{in case (iii)},\end{cases}$$

where Φ denotes the standard normal distribution function. The next theorem gives the large-sample behavior of $\tau_\mu(m)$ and $\tau_\mu^P(m)$.

Theorem 3.3. *Let $(Y_t : t \in \mathbb{Z})$ follow the ARMA equations (2.1) so that (2.2) and (2.3) hold, and suppose that Assumption 3.1 is satisfied for $\Delta_m = \Delta_m^\mu$. Then it holds under H_A^μ for all real u that*

$$\text{(a) } \lim_{m\to\infty} P\left(\frac{\tau_\mu^P(m) - a_m(c_\alpha^P)}{b_m(c_\alpha^P)} \le u \right) = 1 - \Psi(-u).$$

Additionally,

$$\text{(b) } \lim_{m\to\infty} P\left(\frac{\tau_\mu(m) - a_m(c_\alpha)}{b_m(c_\alpha)} \le u \right) = \Phi(u),$$

where $a_m(c)$ is the unique positive solution of

$$a_m(c) = \left(\frac{c m^{1/2-\gamma}}{|\Delta_m^\mu|} + \frac{k^*}{(a_m(c))^\gamma} \right)^{1/(1-\gamma)} \tag{3.10}$$

and

$$b_m(c) = \frac{\sigma\sqrt{a_m(c)}}{|\Delta_m^\mu|} \left(1 - \gamma\left(1 - \frac{k^*}{a_m(c)}\right)\right)^{-1}.$$

The proof of Theorem 3.3 is in Section 5.2. Note that the uniqueness of $a_m(c)$ follows from a rewriting of equation (3.10) to

$$a_m(c) = \frac{cm^{1/2-\gamma}}{|\Delta_m^\mu|}\big(a_m(c)\big)^\gamma + k^*.$$

Now it can be seen that $a_m(c)$ solves an equation of the form $x = ax^\gamma + b$ for appropriately chosen $a > 0$, $b > 0$ and $\gamma \in [0, 1/2)$. Since $a_m(c) > 0$, it is unique as the intersection of the identity with a transformed power function whose exponent is smaller than one.

A similar result can be obtained for the squared-residual procedures $\tau_\xi(m)$ and $\tau_\xi^P(m)$ after appropriate modification. The proof of the following theorem may also be found in Section 5.2 below.

Theorem 3.4. *Let* $(Y_t : t \in \mathbb{Z})$ *follow the ARMA equations* (2.1) *so that* (2.2) *and* (2.3) *hold, and suppose that Assumption* 3.1 *is satisfied for* $\Delta_m = (\Delta_m^\mu)^2$. *Then, under* H_A^μ *for all real* u, *the limit results of Theorem* 3.3 *are retained if* $\tau_\mu(m)$, $\tau_\mu^P(m)$ *and* σ *are replaced with the respective objects* $\tau_\xi(m)$, $\tau_\xi^P(m)$ *and* η.

Some discussion is in order. First, the limit distributions for Page's CUSUM and the traditional CUSUM coincide for the early change scenario (i). Therefore, all procedures work similar in a large-sample setting. The critical values for the traditional CUSUM are somewhat smaller than those for Page's CUSUM (comparing the tables in Horváth *et al.* [19] with those of Fremdt [16]), giving it a slight edge for this case. However, limiting distributions are different for the intermediate and late change scenarios (ii) and (iii), respectively. Here, Page's CUSUM outperforms the traditional CUSUM. This can be explained by the fact that, unlike Page's CUSUM, the traditional CUSUM is not resetting and so becomes less sensitive to a change the later it occurs after the onset of monitoring.

Second, in view of the last paragraph, Page's CUSUM is generally preferred for applications unless the changes happen early. For the early change scenario (i) both procedures perform alike in finite samples (based on simulations not reported in the paper), but as the theoretical results indicate, the performance of the traditional CUSUM decays noticeably for (ii) and (iii). In fact, this stopping rule often exhibits significant non-zero probabilities of non-detection in intermediate and late changes scenarios if the monitoring period is not sufficiently long.

Third, the sensitivity of the test can be adjusted by the statistician through the choice of γ. For example, it has been pointed out by Aue and Horváth [2] that the term $a_m(c)$ can be interpreted as the average delay time $E[\tau]$, where τ stands for any of the stopping times under consideration. For the early change scenario (i), it follows then that $E[\tau] \approx (c/|\Delta_m|)^{1/(1-\gamma)} m^{1-2\gamma/[2(1-\gamma)]}$. This quantity becomes small if γ is chosen close to $1/2$, thus ensuring a quicker detection. However, there is an obvious trade-off between detection time and false alarm rates, with the latter increasing with increasing γ. Similar computations can be obtained for cases (ii) and (iii) as well.

3.3. Limiting delay times for scale breaks

In view of the applications, for which only changes in the scale are considered, presentation in this section is focused on the case of a break in the scale parameter σ only. All other parameters are assumed to remain the same before and after the change occurs. The section closes with remarks for the general case, but a more in-depth analysis is beyond the scope of the present paper. The special case of the general alternative $H_A^{\boldsymbol{\xi}}$, for which only the scale parameter is subject to change, will be called H_A^{σ} in the following. A change of scale will induce the drift term

$$\Delta_m^{\sigma} = \left(\delta_m^{\sigma}\right)^2 + 2\sigma_0\delta_m^{\sigma}$$

into the squared-residual procedures. If this drift term satisfies the regularity conditions imposed through Assumption 3.1, then the asymptotic delay time distribution can be quantified accordingly.

Theorem 3.5. *Let* $(Y_t : t \in \mathbb{Z})$ *follow the ARMA equations* (2.1) *so that* (2.2) *and* (2.3) *hold, and suppose that Assumption* 3.1 *is satisfied for* $\Delta_m = \Delta_m^{\sigma} \neq 0$ *and* $\delta_m^{\sigma} = \mathrm{O}(1)$. *Then the results of Theorem* 3.4 *remain valid under* H_A^{σ}.

The proof of Theorem 3.5 is given in Section 5.3. The general case is much more difficult to handle. The induced drift term will be a complicated function of the pre-break parameters $\boldsymbol{\xi}_0$ and post-break parameters $\boldsymbol{\xi}_A$. In principle, the arguments developed in order to verify the theorems of Sections 3.2 and 3.3 could be adjusted to this case. However, one has to keep track of additional terms, the number of which may be growing exponentially in the number of parameters. Given the complexity of the proofs, we refrain from pursuing this direction further for this paper.

4. Applications

In order to demonstrate the proposed methodology in the finite sample setting, two case studies are provided in this section. The first involves an EEG data set considered in Davis *et al.* [13], the second is a classic data set on IBM stock given in Box *et al.* [9], previously analyzed for breaks in variance with retrospective methods.

4.1. EEG data

In this section, the proposed methodology is applied to two snapshots of a longer series of 32 768 EEG measurements observed from a female patient diagnosed with left temporal lobe epilepsy.[1] This is the "T3 channel" data of Davis *et al.* [13]. Measurements were taken at a sampling rate of 100 Hz (i.e., 100 observations per second), so that the recording took place over a time period of 5 minutes and 28 seconds. As explained in Davis *et al.* [13], expert analysis suggests the

[1] We thank Dr. Beth Malow (formerly Department of Neurology, University of Michigan) for providing the data.

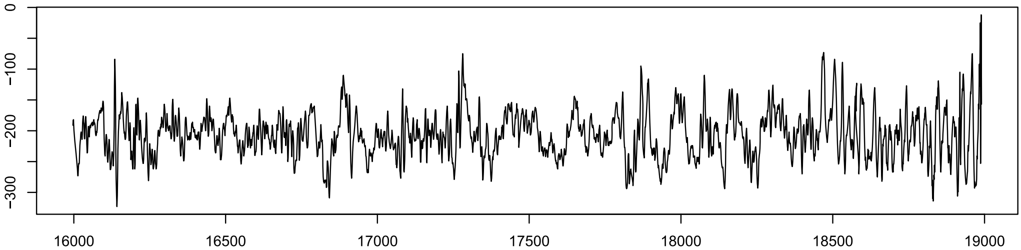

Figure 1. EEG data set.

onset of an epileptic seizure at observation 18 500. Their (retrospective) segmentation procedure estimates the seizure onset at observation 18 580. A similar analysis is reported in Ombao *et al.* [25]. Particular interest here is in two segments of the original data focusing on the interval 16 000–19 000 before and immediately after the suspected seizure onset. These observations are plotted in Figure 1. A visual inspection of the time series plot indicates that the level of the observations remains roughly the same. There is, however, an apparent increase in the amplitude around time 18 500, perhaps indicating a scale break.

To test for this possibility, the following two scenarios for the training period, both of size $m = 1000$, were considered:

(TP1) Observations 16 001–17 000,
(TP2) Observations 17 001–18 000.

The training periods predate the epileptic seizure, with (TP1) implying a longer monitoring period before the break occurrence than (TP2). The choices of training periods enable to examine the effect of the change-point location on the monitoring procedures. In each case, model selection procedures suggest nearly identical AR(4) models. Table 1 contains estimated parameter values for both training cases as well as the monitoring observations immediately following the change-point suggested by the experts. To be precise, the post-change period is

(PC) Observations 18 501–18 580.

All models were fit conditionally on four additional observations in the respective windows. (E.g., in the case of (TP1), the $m + p = 1004$ observations 15 997–17 000 were used for the estimation.) The tabulated estimates suggest the primary change occurs in the innovation variance, while the dynamics of the series remains largely intact.

Table 1. Summary of EEG modeling. Standard errors in parenthesis

Case	$\hat{\phi}_{m,1}$	$\hat{\phi}_{m,2}$	$\hat{\phi}_{m,3}$	$\hat{\phi}_{m,4}$	$\hat{\mu}_m$	$\hat{\sigma}_m^2$
(TP1)	1.66 (0.03)	−0.79 (0.06)	−0.12 (0.06)	0.20 (0.03)	−207.2 (3.85)	63.1
(TP2)	1.64 (0.03)	−0.74 (0.06)	−0.13 (0.06)	0.18 (0.03)	−206.6 (4.90)	61.9
(PC)	1.46 (0.15)	−0.61 (0.27)	0.20 (0.27)	−0.18 (0.16)	−194.8 (12.53)	227.9

Table 2. Summary of EEG stopping times and empirical values based on simulations from the estimated model with 2500 iterations

				Simulated empirical values					
		Stopping times		95% upper limits		Medians		FRR	
Case	γ	Page	CUSUM	Page	CUSUM	Page	CUSUM	Page	CUSUM
	0	18 637	18 676	18 728	18 808	18 623	18 643	0.0240	0.0200
(TP1)	0.25	18 609	18 661	18 718	18 798	18 614	18 634	0.0580	0.0484
	0.49	18 673	18 691	18 768	18 830	18 641	18 650	0.1344	0.1296
	0	18 580	18 581	18 648	18 674	18 576	18 588	0.0036	0.0028
(TP2)	0.25	18 580	18 580	18 626	18 657	18 561	18 573	0.0288	0.0228
	0.49	18 580	18 580	18 633	18 660	18 563	18 569	0.1140	0.1104

The proposed testing procedures were applied to the two training sets at the $\alpha = 0.05$ level. Critical values for the CUSUM procedure were obtained from Horváth *et al.* [19] and critical values for Page's CUSUM procedure from Fremdt [16]. No changes were found by the mean-only procedures $\tau_\mu(m)$ and $\tau_\mu^P(m)$ given in (3.3) and (3.5) when truncating the tests at monitoring time point $10m$. The results for the general procedures $\tau_{\boldsymbol{\xi}}(m)$ and $\tau_{\boldsymbol{\xi}}^P(m)$ are summarized for three choices of γ in the column labeled "Stopping Times" of Table 2. For (TP1), both procedures terminate within two seconds after the suspected onset of the change, for (TP2) within one second. Stopping times for (TP1) generally lag behind stopping times for (TP2). Page's CUSUM detector displays faster detection for both training periods.

It should be noted that a sequential procedure does not provide an estimator for the time of change. In general, it is a difficult problem to estimate the change-point after a sequential procedure has terminated because the post-change sample is typically (much) smaller than the pre-change sample. In the literature, Srivastava and Wu [30] and Wu [33] have discussed options for this problem. It would be worthwhile to follow up on their work elsewhere in the future.

Motivated by the EEG data, several simulations were conducted to further elaborate on the distribution of the stopping times when a change occurs only in the innovation variance. The simulations utilized an AR(4) model with $\mu = -207$ and $\boldsymbol{\phi} = (1.65, -0.75, -0.12, 0.18)'$. The innovations were distributed Laplace$(0, b_0 = 5.6)$ since this closely described the behavior of the residuals from the EEG training models. Mimicking the two cases from the EEG application, we used training sizes of $m = 1000$ and induced changes in the variance by adjusting the scale parameter to $b_A = 10.7$ at time point 18 500 (i.e., monitoring time points 500 and 1500 for (TP1) and (TP2), respectively). The choice of scale parameters imply the difference $\delta^\sigma = 7.21$. Table 2 provides simulated empirical confidence limits, empirical median rejection times and false rejection rates (FRR). The reported values have been adjusted to fit the time locations observed in the EEG example. The reported stopping times for the EEG example all fall within the empirical upper bounds from the simulation study. The large false rejection rates for $\gamma = 0.49$ display the delay in convergence to the asymptotic levels suggested by Horváth *et al.* [19] and Fremdt [16] when the sensitivity parameter is close to the upper boundary.

4.2. IBM data

The second application is a study of a classic retrospective data set which has been previously studied for changes in the variance, albeit in a retrospective setting. The observations are on the IBM common stock daily closing prices from May 17, 1961 to November 2, 1962. This is Series B as reported in Box *et al.* [9]. The data set contains 369 observations and has been examined in several retrospective studies which focused primarily on changes in the variance. Several authors have detected two change-points. Inclán and Tiao [22] detected change-points at observations 235 and 279 using their ICSS algorithm, Baufays and Rasson [7] proposed 235 and 280, Wichern *et al.* [32] gave 180 and 235, while Tsay [31] reported only one change at observation 237. As previously suggested in order to stabilize the variance, the first difference of the log transformed series will be analyzed. Figure 2 displays the corresponding time series plot. It can be seen that fluctuations appear to be around a constant level, while amplitudes are larger for roughly the last third of the observations.

To estimate an initial model, the training period is selected to consist of the first $m = 200$ observations. Two competing models were identified based on AIC and model selection diagnostic plots. The competing fits are the ARMA(2, 2) and AR(4) estimated models summarized in Table 3, with the AIC value being slightly smaller for the ARMA(2, 2) model.

The proposed procedures were applied at the $\alpha = 0.05$ level, utilizing both model fits. Monitoring commences at observation 201. The mean-only procedures $\tau_\mu(m)$ and $\tau_\mu^P(m)$ do not detect deviations from a constant level. Table 4 provides the observed values for the general stopping rules $\tau_\xi(m)$ and $\tau_\xi^P(m)$. Depending on the choice of γ, both procedures report a change has occurred at or near observation 238. For comparison purposes, a simulation study was conducted and is also summarized in Table 4. The simulations generated training data from the observed ARMA(2, 2) model. A change was induced at time point 235 to reflect the observed instability in the IBM example. Based on observations 235–279 (retrospective studies suggest stability over this period), the best fitting model was a white noise process with innovation variance given by 0.00135. The empirical measures from the simulation study are similar when assuming the correct ARMA model orders, as well as when the AR(4) is assumed. This highlights an important feature. Models with nearly identical MA(∞) representations exhibit similar behavior with re-

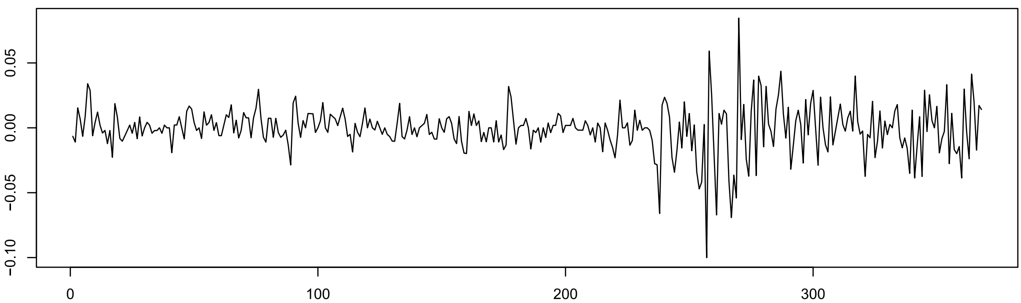

Figure 2. Plot of transformed IBM data set.

Table 3. Summary of IBM modeling. Standard errors in parenthesis

Model	AIC	$\hat{\phi}_{m,1}$	$\hat{\phi}_{m,2}$	$\hat{\phi}_{m,3}$	$\hat{\phi}_{m,4}$	$\hat{\theta}_{m,1}$	$\hat{\theta}_{m,2}$	$\hat{\sigma}_m^2$
ARMA(2, 2)	−1296	−0.40 (0.13)	−0.68 (0.11)	–	–	0.67 (0.12)	0.76 (0.10)	8.5e−05
AR(4)	−1292	0.26 (0.07)	−0.12 (0.07)	−0.10 (0.07)	0.16 (0.07)	–	–	8.7e−05

spect to our proposed methodology. For our observed ARMA(2, 2) and AR(4) models, Figure 3 displays the differences in the initial MA(∞) coefficients.

5. Proofs

5.1. Preliminaries

The following auxiliary result will be used frequently. It establishes the behavior of the coefficients $\pi_j(\mathbf{v})$ and $\psi_j(\mathbf{u})$ in (3.6) if instead of the true parameter vectors $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$, generic elements $\mathbf{v} \in \mathbb{R}^p$ and $\mathbf{u} \in \mathbb{R}^q$ in their vicinity are used for the respective power series expansions. Let $|\cdot|$ denote the maximum norm of vectors.

Proposition 5.1. *Let $(Y_t : t \in \mathbb{Z})$ follow the ARMA equations (2.1) so that (2.2) holds. Let $\mathbf{v} \in \mathbb{R}^p$ and $\mathbf{u}, \mathbf{u}_1, \mathbf{u}_2 \in \mathbb{R}^q$. Then there are $\varepsilon > 0$, $c \in (0, 1)$ and $K > 0$ such that, for all $j \geq 0$,*

(a) $|\pi_j(\mathbf{v})| \leq Kc^j$, *if* $|\mathbf{v} - \boldsymbol{\phi}| \leq \varepsilon$;
(b) $|\psi_j(\mathbf{u})| \leq Kc^j$, *if* $|\mathbf{u} - \boldsymbol{\theta}| \leq \varepsilon$;
(c) $|\psi_j(\mathbf{u}_1) - \psi_j(\mathbf{u}_2)| \leq K|\mathbf{u}_1 - \mathbf{u}_2| j c^{j-1}$, *if* $|\mathbf{u}_1 - \boldsymbol{\theta}| \leq \varepsilon$ *and* $|\mathbf{u}_2 - \boldsymbol{\theta}| \leq \varepsilon$.

Proof. The proof of these statements can be found in Bai [5]. □

Table 4. Summary of IBM stopping times and empirical values based on simulations from the estimated model with 2500 iterations

				Simulated empirical values					
		Stopping times		95% upper limits		Medians		FRR	
Case	γ	Page	CUSUM	Page	CUSUM	Page	CUSUM	Page	CUSUM
	0	238	239	244	244	238	238	0.0024	0.0024
ARMA(2, 2)	0.25	238	238	242	242	237	237	0.0228	0.0216
	0.49	238	238	241	242	236	237	0.1008	0.1072
	0	239	242	244	244	238	239	0.0004	0.0004
AR(4)	0.25	238	238	242	243	237	237	0.0120	0.0100
	0.49	238	238	241	242	237	237	0.0848	0.0904

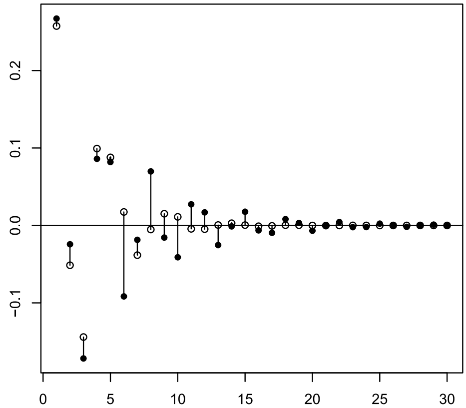

Figure 3. Comparing MA(∞) coefficients from the observed ARMA(2, 2) (filled) and AR(4) (opened) models.

Throughout the proofs, let λ_t denote the difference between the residuals $\hat{\varepsilon}_t$ and the innovations ε_t if the null hypothesis H_0 is valid. Since none of the parameters is subject to change, it holds then that

$$\lambda_t = \lambda_t\big(\sqrt{m}[\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0], \sqrt{m}[\hat{\boldsymbol{\phi}}_m - \boldsymbol{\phi}_0], \sqrt{m}[\hat{\mu}_m - \mu_0]\big), \tag{5.1}$$

where

$$\lambda_t(\mathbf{u}, \mathbf{v}, w) = \zeta_t(\mathbf{u}) + \frac{\beta_t(\mathbf{u}, \mathbf{v})}{\sqrt{m}} + \frac{\rho_t(\mathbf{u}, \mathbf{v}, w)}{\sqrt{m}}.$$

To define the quantities on the right-hand side of the latter equality, let first $\mathbf{u}^* = \boldsymbol{\theta} + \mathbf{u}/\sqrt{m}$ and $\mathbf{v}^* = \boldsymbol{\phi} + \mathbf{v}/\sqrt{m}$, and set $u_0^* = 0$. Then

$$\zeta_t(\mathbf{u}) = -\sum_{j=1}^{q}\left(\sum_{\ell=1}^{j}\psi_{t-1+\ell}\big(\mathbf{u}^*\big)u^*_{j-\ell}\right)\varepsilon_{1-j},$$

$$\beta_t(\mathbf{u}, \mathbf{v}) = -\sum_{j=1}^{p} v_j \sum_{\ell=0}^{t-1}\psi_\ell\big(\mathbf{u}^*\big)X_{t-j-\ell} - \sum_{j=1}^{q} u_j \sum_{\ell=0}^{t-1}\psi_\ell\big(\mathbf{u}^*\big)\varepsilon_{t-j-\ell},$$

$$\rho_t(\mathbf{u}, \mathbf{v}, w) = -\left(1 - \sum_{j=1}^{p} v_j^*\right) w \sum_{\ell=0}^{t-1}\psi_\ell\big(\mathbf{u}^*\big).$$

This is the decomposition given in Yu [34], which is useful to derive the limit distributions of the various test procedures under the null hypotheses as given in Theorems 3.1 and 3.2. This was done for the CUSUM-type procedure in Dienes and Aue [14], but the same approach works also for the procedure based on Page's CUSUM using the work of Fremdt [15,16].

To prove the new results on the asymptotic delay time distribution of the stopping times one may modify methodology developed in Aue and Horváth [2]: It is subsequently shown that sequences $N = N(m, x)$ can be found such that, for the stopping time τ with corresponding detector $D(m, k)$, it holds that

$$P(\tau > N) = P\left(\max_{1 \le k \le N} \frac{\hat{D}(m,k)}{g_\gamma(m,k)} \le c\right)$$

converges to the appropriate limit distribution. The standardizations for τ in the various theorems are then implied by the definition of N. The next section contains the verification for the mean break case.

5.2. Proofs of the results in Section 3.2

For the mean break case, changes in the second order parameters $\boldsymbol{\phi}$, $\boldsymbol{\theta}$ and σ^2 are precluded. To determine the effect of the mean break on the differences $\hat{\varepsilon}_t - \varepsilon_t$, one consequently needs to check only the terms including the μ_{t-j}. It can be seen from (5.1) that these terms only enter through ρ_t. To determine the drift induced by the change in mean under H_A^μ, a similar decomposition to (5.1) is needed. Following equation (14) in Yu [34], it follows that

$$\begin{aligned}\hat{\varepsilon}_t - \varepsilon_t &= \tilde{\zeta}_t(\hat{\boldsymbol{\theta}}_m) + \frac{\tilde{\beta}_t(\hat{\boldsymbol{\theta}}_m, \hat{\boldsymbol{\phi}}_m)}{\sqrt{m}} \\ &\quad - \sum_{\ell=0}^{t-1} \psi_\ell(\hat{\boldsymbol{\theta}}_m)\left[(\hat{\mu}_m - \mu_{t-k}) - \sum_{j=1}^{p} \hat{\phi}_{m,j}(\hat{\mu}_m - \mu_{t-j-\ell})\right],\end{aligned} \tag{5.2}$$

where $\tilde{\zeta}_t(\hat{\boldsymbol{\theta}}_m) = \zeta_t(\sqrt{m}[\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0])$ and $\tilde{\beta}_t(\hat{\boldsymbol{\theta}}_m, \hat{\boldsymbol{\phi}}_m) = \beta_t(\sqrt{m}[\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0], \sqrt{m}[\hat{\boldsymbol{\phi}}_m - \boldsymbol{\phi}_0])$ are respectively the terms of initialization effects and the partial sums of centered observations and innovations. To derive (5.2), one uses the recursiveness of the difference $\hat{\varepsilon}_t - \varepsilon_t$ and the invertibility of the underlying ARMA process. Now, as under H_A^μ a change occurs only in μ_t for $t \ge m + k^*$, it suffices to investigate the term

$$\begin{aligned}&- \sum_{\ell=0}^{t-1} \psi_\ell(\hat{\boldsymbol{\theta}}_m)\left[(\hat{\mu}_m - \mu_{t-k}) - \sum_{j=1}^{p} \hat{\phi}_{m,j}(\hat{\mu}_m - \mu_{t-j-\ell})\right] \\ &\quad = -\left(1 - \sum_{j=1}^{p} \hat{\phi}_{m,j}\right)(\hat{\mu}_m - \mu_0)\sum_{\ell=0}^{t-1} \psi_\ell(\hat{\boldsymbol{\theta}}_m) + \delta_m^\mu \sum_{\ell=0}^{t-1} \psi_\ell(\hat{\boldsymbol{\theta}}_m)\left[I_{t,0,\ell} - \sum_{j=1}^{p} \hat{\phi}_{m,j} I_{t,j,\ell}\right] \\ &\quad = \frac{\tilde{\rho}_t(\hat{\boldsymbol{\theta}}_m, \hat{\boldsymbol{\phi}}_m, \hat{\mu}_m)}{\sqrt{m}} + \Lambda_{t-m-k^*}^\mu,\end{aligned}$$

where $\tilde{\rho}_t(\hat{\boldsymbol{\theta}}_m, \hat{\boldsymbol{\phi}}_m, \hat{\mu}_m) = \rho_t(\sqrt{m}[\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0], \sqrt{m}[\hat{\boldsymbol{\phi}}_m - \boldsymbol{\phi}_0], \sqrt{m}[\hat{\mu}_m - \mu_0])$ and $I_{t,j,\ell}$ is short for $I_{\{t-j-\ell \ge k^*+m\}}(t, j, \ell)$. Here, I_A denotes the indicator function of a set A. Letting $t \ge m + k^*$ and

$s = t - m - k^*$, the drift term can be written as

$$\Lambda_s^\mu = \delta_m^\mu \sum_{\ell=0}^{t-1} \psi_\ell(\hat{\boldsymbol{\theta}}_m) \left[I_{t,0,\ell} - \sum_{j=1}^{p} \hat{\phi}_{m,j} I_{t,j,\ell} \right]$$

$$= \begin{cases} 0, & s < 0, \\ \delta_m^\mu \sum_{\ell=0}^{s} \psi_{s-\ell}(\hat{\boldsymbol{\theta}}_m) \left(1 - \sum_{j=1}^{\ell} \hat{\phi}_{m,j} \right), & 0 \le s < p, \\ \delta_m^\mu \left[\left(1 - \sum_{j=1}^{p} \hat{\phi}_{m,j} \right) \sum_{\ell=0}^{s-p} \psi_\ell(\hat{\boldsymbol{\theta}}_m) + \sum_{\ell=0}^{p-1} \psi_{s-\ell}(\hat{\boldsymbol{\theta}}_m) \left(1 - \sum_{j=1}^{\ell} \hat{\phi}_{m,j} \right) \right], & s \ge p. \end{cases} \tag{5.3}$$

Note that the drift has been rescaled, so that $s < 0$ indicates that the change has not yet occurred. The further distinction into the cases $0 \le s < p$ and $s \ge p$ takes into account the autoregressive order. It follows that

$$\hat{\varepsilon}_t - \varepsilon_t = \lambda_t + \Lambda_{t-m-k^*}^\mu, \tag{5.4}$$

with λ_t from (5.1). To prove the theorems of Section 3.2, it remains to analyze partial sums of the $\hat{\varepsilon}_t - \varepsilon_t$ and compare them to the growth of the threshold $g_\gamma(m, k)$.

Proof of Theorem 3.3. Let $k \ge k^*$ and $M = k - k^*$. Utilizing λ_t from (5.1) and display (5.4), it follows that

$$\sum_{t=m+1}^{m+k} (\hat{\varepsilon}_t - \varepsilon_t) = \sum_{t=m+1}^{m+k} \left(\lambda_t + \Lambda_{t-m-k^*}^\mu \right) = \sum_{t=m+1}^{m+k} \lambda_t + \sum_{s=0}^{M} \Lambda_s^\mu.$$

The first term on the right-hand side can be treated as under the null hypothesis, see Dienes and Aue [14]. The drift of the cumulative sum procedure can be determined as follows. First, for $M < p$, (5.3) implies directly that

$$\sum_{s=0}^{M} \Lambda_s^\mu = \delta_m^\mu \sum_{s=0}^{M} \sum_{\ell=0}^{s} \psi_{s-\ell}(\hat{\boldsymbol{\theta}}_m) \left(1 - \sum_{j=1}^{\ell} \hat{\phi}_{m,j} \right).$$

Second, for $M \ge p$, another application of (5.3) using the cases for $0 \le s < p$ and $p \le s \le M$ to split up the sum and subsequently combining the terms involving the incomplete sums $1 - \sum_{j=1}^{\ell} \hat{\phi}_{m,j}$ of estimated autoregressive coefficients, yields

$$\begin{aligned} \sum_{s=0}^{M} \Lambda_s^\mu &= \delta_m^\mu \Bigg[\left(1 - \sum_{j=1}^{p} \hat{\phi}_{m,j} \right) \sum_{\ell=0}^{M-p} \psi_\ell(\hat{\boldsymbol{\theta}}_m) \big[(M - p + 1) - \ell \big] \\ &\qquad + \sum_{s=0}^{p-1} \left(1 - \sum_{j=1}^{s} \hat{\phi}_{m,j} \right) \sum_{\ell=0}^{M-s} \psi_\ell(\hat{\boldsymbol{\theta}}_m) \Bigg] \\ &= \hat{\Delta}_m^\mu (M - p + 1) + \delta_m^\mu \big[A_1(M) - A_2(M) + A_3(M) \big], \end{aligned}$$

where $\hat{\Delta}_m^\mu = \delta_m^\mu \hat{\phi}_m(1)/\hat{\theta}_m(1)$ with $\hat{\phi}_m(1) = 1 - \hat{\phi}_{m,1}z - \cdots - \hat{\phi}_{m,p}z^p$ and $\hat{\theta}_m(1) = 1 + \hat{\theta}_{m,1}z + \cdots + \hat{\theta}_{m,q}z^q$, and

$$A_1(M) = (M-p+1)\left(1 - \sum_{j=1}^{p} \hat{\phi}_{m,j}\right) \sum_{\ell=M-p+1}^{\infty} \psi_\ell(\hat{\boldsymbol{\theta}}_m),$$

$$A_2(M) = \left(1 - \sum_{j=1}^{p} \hat{\phi}_{m,j}\right) \sum_{\ell=0}^{M-p} \ell\psi_\ell(\hat{\boldsymbol{\theta}}_m),$$

$$A_3(M) = \sum_{s=0}^{p-1}\left(1 - \sum_{j=1}^{s} \hat{\phi}_{m,j}\right) \sum_{\ell=0}^{M-s} \psi_\ell(\hat{\boldsymbol{\theta}}_m).$$

It is clear that $\hat{\Delta}_m^\mu$ will be close to its deterministic equivalent Δ_m^μ if m is large. The terms $A_1(M)$, $A_2(M)$ and $A_3(M)$ are stochastically bounded, so that Proposition 5.1 implies that, as $m \to \infty$,

$$\left(\frac{N}{m}\right)^{\gamma-1/2} \max_{k^* \le k \le N} \frac{\delta_m^\mu A_i(k-k^*)}{g_\gamma(m,k)} = o_P(1), \qquad i = 1,2,3.$$

If the sequence $N = N(m,x)$ given in Fremdt [16] is used as the upper bound for the maximum. The rest of the proof of part (a) of the theorem follows now analogously to the proof of Theorem 2.2 in Fremdt [16].

Part (b) can be verified by an extension of the proof in Aue and Horváth [2], relaxing their assumption on the order of the change-point to the requirement of part (a) in Assumption 3.1. This can be done with the sharper estimates developed in Fremdt [16]. Further details are omitted to conserve space. □

Proof of Theorem 3.4. To investigate the behavior of the general detectors under H_A^μ, the previous proof needs to be adjusted for the squared residuals. From (5.4) it follows that, for $t \ge m + k^*$,

$$\hat{\varepsilon}_t^2 - \varepsilon_t^2 = \lambda_t^2 + 2\lambda_t\varepsilon_t + \left(\Lambda_{t-m-k^*}^\mu\right)^2 + 2\Lambda_{t-m-k^*}^\mu(\varepsilon_t + \lambda_t). \tag{5.5}$$

The first two terms on the right-hand side can again be treated as under the null hypothesis. The relevant drift term for the sequential procedures consists then of the partial sums of $(\Lambda_{t-m-k^*}^\mu)^2$ and $2\Lambda_{t-m-k^*}^\mu(\varepsilon_t + \lambda_t)$, of which the latter will be negligible. To verify this claim, observe first that

$$\max_{k^* \le k < \infty} \frac{1}{k} \sum_{t=0}^{k} |\varepsilon_t| = O_P(1)$$

since, on account of the strong law of large numbers, $\frac{1}{k}\sum_{t=0}^{k} |\varepsilon_t|$ converges almost surely as $k \to \infty$. Because

$$\max_{k^* \le k \le N} \left(\frac{N}{m}\right)^{\gamma-1/2} \frac{k}{g_\gamma(m,k)} = o(1) \qquad (m \to \infty),$$

it follows from Proposition 5.1 that

$$\left(\frac{N}{m}\right)^{\gamma-1/2} \max_{k^*\le k\le N} \sum_{t=m+k^*}^{m+k} \frac{\Lambda^{\mu}_{t-m-k^*}\varepsilon_t}{g_\gamma(m,k)} = o_P(1) \qquad (m\to\infty).$$

Utilizing the definition of Λ_s^{μ} in (5.3) and another application of Proposition 5.1 in combination with Lemmas 6.1–6.3 of Dienes and Aue [14] yield also that

$$\left(\frac{N}{m}\right)^{\gamma-1/2} \max_{k^*\le k\le N} \sum_{t=m+k^*}^{m+k} \frac{\Lambda^{\mu}_{t-m-k^*}\lambda_t}{g_\gamma(m,k)} = o_P(1) \qquad (m\to\infty).$$

It therefore remains to extract the dominating term from the partial sums of $(\Lambda^{\mu}_{t-m-k^*})^2$. To facilitate notation, the abbreviations $\psi_\ell=\psi_\ell(\hat\theta_m)$, $\hat\phi_m^{(\ell)}(z)=1-\hat\phi_{m,1}z-\cdots-\hat\phi_{m,\ell}z^\ell$, $\ell=1,\dots,p-1$, and $k'=k^*+m+p$ are used. Then,

$$\begin{aligned}\sum_{t=k'}^{m+k}\big(\Lambda^{\mu}_{t-m-k^*}\big)^2 &= \big(\delta_m^{\mu}\big)^2\Bigg[\hat\phi_m^2(1)\sum_{t=k'}^{m+k}\Bigg(\sum_{\ell=0}^{t-k'}\psi_\ell\Bigg)^2+\sum_{t=k'}^{m+k}\Bigg(\sum_{\ell=0}^{p-1}\hat\phi_m^{(\ell)}(1)\psi_{t-m-k^*-\ell}\Bigg)^2\\ &\quad -2\hat\phi_m^2(1)\sum_{t=k'}^{m+k}\Bigg(\sum_{\ell=0}^{t-k'}\psi_\ell\Bigg)\Bigg(\sum_{\ell'=0}^{p-1}\hat\phi_m^{(\ell')}(1)\psi_{t-m-k^*-\ell'}\Bigg)\Bigg].\end{aligned}$$

Similar arguments to those used in the proof of Theorem 3.3 yield that only the first term needs to be investigated. Since

$$\begin{aligned}\sum_{s=0}^{t}\Bigg(\sum_{\ell=0}^{s}\psi_\ell\Bigg)^2 &= (t+1)\Bigg(\sum_{\ell=0}^{\infty}\psi_\ell\Bigg)^2-2\Bigg(\sum_{\ell=0}^{\infty}\psi_\ell\Bigg)\sum_{s=0}^{t}\Bigg(\sum_{\ell=s+1}^{\infty}\psi_\ell\Bigg)+\sum_{s=0}^{t}\Bigg(\sum_{\ell=s+1}^{\infty}\psi_\ell\Bigg)^2\\ &= (t+1)\Bigg(\sum_{\ell=0}^{\infty}\psi_\ell\Bigg)^2-2\Bigg(\sum_{\ell=0}^{\infty}\psi_\ell\Bigg)\Bigg[\sum_{s=1}^{t}s\psi_s+(t+1)\sum_{s=t+1}^{\infty}\psi_s\Bigg]\\ &\quad+\sum_{s=0}^{t}\Bigg(\sum_{\ell=s+1}^{\infty}\psi_\ell\Bigg)^2,\end{aligned}$$

following the arguments of the proof of Theorem 3.3 implies that $(t+1)\hat\theta_m^{-2}(1)$ is the dominating term in this expression. The rest follows analogously to the proof of Theorem 2.2 in Fremdt [16]. □

5.3. Proofs of the results in Section 3.3

Denote by $(z_t : t\in\mathbb{Z})$ the sequence of independent, identically distributed and standardized random variables given by the requirement $\varepsilon_t=\sigma_t z_t$ for all $t\in\mathbb{Z}$. Therefore changes in scale do not affect the z_t's. In the following, the subscript 0 in the quantities $y_{0,t}$, $x_{0,t}$ and $\varepsilon_{0,t}$ will indicate that the corresponding random variables are generated according to the null parameter vector $\boldsymbol{\xi}_0$.

Precluding a break in the mean, autoregressive and moving average parameters and only allowing breaks in the scale parameter, leads to the decomposition

$$\hat{\varepsilon}_t^2 - \varepsilon_{0,t}^2 = \lambda_t^2 + 2\lambda_t\varepsilon_{0,t} + \left(\Lambda_{t-m-k^*}^{\sigma}\right)^2 + 2\Lambda_{t-m-k^*}^{\sigma}(\varepsilon_{0,t} + \lambda_t), \tag{5.6}$$

for $t \geq m + k^*$, which is analogous to (5.5). Now $\Lambda_{t-m-k^*}^{\sigma}$ can be further decomposed into

$$\Lambda_{t-m-k^*}^{\sigma} = \delta_m^{\sigma}(z_t + B_t).$$

Using that, for $t \geq m + k^*$,

$$x_t - x_{0,t} = \delta_m^{\sigma}\left(\sum_{k=0}^{t-m-k^*} \pi_k(\boldsymbol{\phi}_A)\left[z_{t-k} + \sum_{j=1}^{\min(q,t-m-k^*-k)} \theta_{0,j} z_{t-j-k}\right]\right),$$

and setting again $s = t - m - k^*$ gives

$$\begin{aligned}
B_t &= \sum_{j=1}^{q}(\theta_{0,j} - \hat{\theta}_{m,j})\sum_{k=0}^{s-q}\psi_k z_{t-j-k} + \sum_{k=1}^{q-1}\sum_{j=1}^{k}\psi_{s-k}(\theta_{0,j} - \hat{\theta}_{m,j})z_{m+k^*-j+k} \\
&\quad + \sum_{j=1}^{p}(\phi_{0,j} - \hat{\phi}_{m,j})\sum_{k=0}^{s-j}\psi_k \sum_{n=0}^{s-j-k}\pi_n(\boldsymbol{\phi}_0)z_{t-j-k-n} \\
&\quad + \sum_{j=1}^{p}(\phi_{0,j} - \hat{\phi}_{m,j})\sum_{\ell=1}^{q}\theta_{0,j}\sum_{k=0}^{s-j}\sum_{n=0}^{s-j-k-q}\psi_k\pi_n(\boldsymbol{\phi}_0)z_{t-j-k-\ell-n} \\
&\quad + \sum_{j=1}^{p}(\phi_{0,j} - \hat{\phi}_{m,j})\sum_{k=0}^{s-j}\psi_k\sum_{n=1}^{q-1}\pi_{s-j-k-n}(\boldsymbol{\phi}_0)\sum_{\ell=1}^{n}\theta_{0,\ell}z_{m+k^*+n-\ell} \\
&= B_{1,t} + \cdots + B_{5,t},
\end{aligned}$$

where $\psi_k = \pi_k(\boldsymbol{\phi}_0) = 0$ for $k < 0$. The following lemma identifies the dominating term in the partial sums of $\hat{\varepsilon}_t^2 - \varepsilon_{0,t}^2$.

Lemma 5.1. *Under the assumptions of Theorem 3.5,*

$$\left(\frac{N}{m}\right)^{\gamma-1/2}\max_{k^*\leq k\leq N}\frac{1}{g_\gamma(m,k)}\left|\sum_{t=m+k^*}^{m+k}\left(\hat{\varepsilon}_t^2 - \varepsilon_{0,t}^2\right) - \Delta_m^{\sigma}\sum_{t=m+k^*}^{m+k} z_t^2\right| = o_P(1)$$

as $m \to \infty$.

Proof. It suffices to examine the quantities on the right-hand side of (5.6). Notice first that $\lambda_t^2 + 2\Delta_t \varepsilon_{0,t}$ contains only terms related to the behavior under the null hypothesis. For the next two terms on the right-hand side of (5.6), write

$$\begin{aligned}&\big(\Lambda_s^\sigma\big)^2 + 2\Lambda_s^\sigma(\varepsilon_{0,t} + \lambda_t)\\&\quad= \big[\big(\delta_m^\sigma\big)^2 + 2\sigma_0\delta_m^\sigma\big]z_t^2 + 2\delta_m^\sigma\big(\delta_m^\sigma + \sigma_0\big)z_t B_t + \big(\delta_m^\sigma\big)^2 B_t^2 + 2\lambda_t\Lambda_s^\sigma.\end{aligned} \tag{5.7}$$

The first term is the dominating term. Since $\Delta_m^\sigma = (\delta_m^\sigma)^2 + 2\sigma_0\delta_m^2$, the assertion of the lemma will follow if the remaining terms can be shown to be negligible. For the second term notice that

$$\left(\frac{N}{m}\right)^{\gamma-1/2} \max_{k^*\le k\le N} \sum_{t=m+k^*}^{m+k} \frac{z_t B_t}{g_\gamma(m,k)} \le \left(\frac{N}{m}\right)^{\gamma-1/2} \sum_{t=m+k^*}^{m+N} \frac{|z_t||B_t|}{g_\gamma(m,k^*)} = o_P(1),$$

since z_t and B_t are independent and $\sqrt{m}E[B_t] < \infty$, following the arguments used in Dienes and Aue [14]. For the third term in (5.6), observe that

$$\left(\frac{N}{m}\right)^{\gamma-1/2} \max_{k^*\le k\le N} \sum_{t=m+k^*}^{m+k} \frac{B_t^2}{g_\gamma(m,k)} \le \left(\frac{N}{m}\right)^{\gamma-1/2} \sum_{t=m+k^*}^{m+k} \sum_{\ell=1}^{5} \frac{B_{\ell,t}^2}{g_\gamma(m,k^*)}.$$

The proof is only detailed for $\ell = 1$, since all other terms can be handled in a similar fashion. For this case,

$$\begin{aligned}&\left(\frac{N}{m}\right)^{\gamma-1/2} \frac{1}{g_\gamma(m,k^*)} \sum_{t=m+k^*}^{m+N} \left(\sum_{j=1}^{q}(\theta_{0,j} - \hat\theta_{m,j}) \sum_{k=0}^{s-q} \psi_k z_{t-j-k}\right)^2\\&\quad\le \left(\frac{N}{m}\right)^{\gamma-1/2} \frac{q}{g_\gamma(m,k^*)} \sum_{t=m+k^*}^{m+N} \sum_{j=1}^{q}(\theta_{0,j} - \hat\theta_{m,j})^2 \left(\sum_{k=0}^{s-q} \psi_k z_{t-j-k}\right)^2.\end{aligned}$$

Now the arguments of Lemma 5.2 in Dienes and Aue [14] apply and yield the $o_P(1)$ rate. For the last term in (5.6) there is nothing to show, since

$$\Lambda_s^\sigma \lambda_t = \delta_m^\sigma(\lambda_t z_t + \lambda_t B_t) \le \delta_m^\sigma\big(\lambda_t z_t + \lambda_t^2 + B_t^2\big)$$

and all these terms have already been shown to be negligible. The proof is complete. □

Proof of Theorem 3.5. The relevant drift term has been identified in Lemma 5.1. Noticing that the law of the iterated logarithm implies that, for all $\delta \in (0, 1/2)$ and as $m \to \infty$,

$$\begin{aligned}&\left(\frac{N}{m}\right)^{\gamma-1/2} \max_{k^*\le k\le N} \frac{1}{g_\gamma(m,k)} \sum_{t=m+k^*}^{m+k} \big(z_t^2 - 1\big)\\&\quad= O_P(1)\frac{m^{1-\gamma}}{N^{1/2-\gamma}} \max_{1\le k\le N-k^*} \frac{k^{1/2-\gamma+\delta}}{(m+k^*+m)^{1-\gamma}} = o_P(1),\end{aligned}$$

the assertion of the theorem follows. □

Acknowledgements

This research was partially supported by Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823) of the German Research Foundation (DFG), NSF grants DMS 0905400, DMS 1209226 and DMS 1305858, and DFG grant STE 306/22-1.

References

[1] Andreou, E. and Ghysels, E. (2006). Monitoring disruptions in financial markets. *J. Econometrics* **135** 77–124. MR2328397

[2] Aue, A. and Horváth, L. (2004). Delay time in sequential detection of change. *Statist. Probab. Lett.* **67** 221–231. MR2053524

[3] Aue, A. and Horváth, L. (2013). Structural breaks in time series. *J. Time Series Anal.* **34** 1–16. MR3008012

[4] Aue, A., Horváth, L. and Reimherr, M.L. (2009). Delay times of sequential procedures for multiple time series regression models. *J. Econometrics* **149** 174–190. MR2518505

[5] Bai, J. (1993). On the partial sums of residuals in autoregressive and moving average models. *J. Time Ser. Anal.* **14** 247–260. MR1221937

[6] Basseville, M. and Nikiforov, I.V. (1993). *Detection of Abrupt Changes: Theory and Application. Prentice Hall Information and System Sciences Series*. Englewood Cliffs, NJ: Prentice Hall. MR1210954

[7] Baufays, P. and Rasson, J.-P. (1985). Variance changes in autoregressive model. In *Time Series Analysis: Theory and Practice* **7** (Toronto, Ont., 1983) (O.D. Anderson, ed.) 119–128. Amsterdam: North-Holland. MR0787816

[8] Berkes, I., Gombay, E., Horváth, L. and Kokoszka, P. (2004). Sequential change-point detection in GARCH(p, q) models. *Econometric Theory* **20** 1140–1167. MR2101953

[9] Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (2008). *Time Series Analysis. Forecasting and Control*, 4th ed. *Wiley Series in Probability and Statistics*. Hoboken, NJ: Wiley. MR2419724

[10] Brown, R.L., Durbin, J. and Evans, J.M. (1975). Techniques for testing the constancy of regression relationships over time. *J. Roy. Statist. Soc. Ser. B* **37** 149–192. With discussion by D. R. Cox, P. R. Fisk, Maurice Kendall, M. B. Priestley, Peter C. Young, G. Phillips, T. W. Anderson, A. F. M. Smith, M. R. B. Clarke, A. C. Harvey, Agnes M. Herzberg, M. C. Hutchison, Mohsin S. Khan, J. A. Nelder, Richard E. Quant, T. Subba Rao, H. Tong and W. G. Gilchrist and with reply by J. Durbin and J. M. Evans. MR0378310

[11] Chu, C.-S., Stinchcombe, J. and White, H. (1996). Monitoring structural change. *Econometrica* **64** 1045–1065.

[12] Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis. Wiley Series in Probability and Statistics*. Chichester: Wiley. With a foreword by David Kendall. MR2743035

[13] Davis, R.A., Lee, T.C.M. and Rodriguez-Yam, G.A. (2006). Structural break estimation for nonstationary time series models. *J. Amer. Statist. Assoc.* **101** 223–239. MR2268041

[14] Dienes, C. and Aue, A. (2014). On-line monitoring of pollution concentrations with autoregressive moving average time series. *J. Time Ser. Anal.* To appear.

[15] Fremdt, S. (2014). Page's sequential procedure for change-point detection in time series regression. *Statistics*. To appear. DOI:10.1080/02331888.2013.870568

[16] Fremdt, S. (2014). Asymptotic distribution of the delay time in Page's sequential procedure. *J. Statist. Plann. Inference* **145** 74–91. MR3125350
[17] Gombay, E. and Horváth, L. (2009). Sequential tests and change detection in the covariance structure of weakly stationary time series. *Comm. Statist. Theory Methods* **38** 2872–2883. MR2568192
[18] Gombay, E. and Serban, D. (2009). Monitoring parameter change in AR(p) time series models. *J. Multivariate Anal.* **100** 715–725. MR2478193
[19] Horváth, L., Hušková, M., Kokoszka, P. and Steinebach, J. (2004). Monitoring changes in linear models. *J. Statist. Plann. Inference* **126** 225–251. MR2090695
[20] Hušková, M. and Kirch, C. (2012). Bootstrapping sequential change-point tests for linear regression. *Metrika* **75** 673–708. MR2946654
[21] Hušková, M., Kirch, C., Prášková, Z. and Steinebach, J. (2008). On the detection of changes in autoregressive time series. II. Resampling procedures. *J. Statist. Plann. Inference* **138** 1697–1721. MR2400475
[22] Inclán, C. and Tiao, G.C. (1994). Use of cumulative sums of squares for retrospective detection of changes of variance. *J. Amer. Statist. Assoc.* **89** 913–923. MR1294735
[23] Kirch, C. (2008). Bootstrapping sequential change-point tests. *Sequential Anal.* **27** 330–349. MR2446906
[24] Lorden, G. (1971). Procedures for reacting to a change in distribution. *Ann. Math. Statist.* **42** 1897–1908. MR0309251
[25] Ombao, H.C., Raz, J.A., von Sachs, R. and Malow, B.A. (2001). Automatic statistical analysis of bivariate nonstationary time series. *J. Amer. Statist. Assoc.* **96** 543–560. MR1946424
[26] Page, E.S. (1954). Continuous inspection schemes. *Biometrika* **41** 100–115. MR0088850
[27] Page, E.S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika* **42** 523–527. MR0072412
[28] Robbins, M., Gallagher, C., Lund, R. and Aue, A. (2011). Mean shift testing in correlated data. *J. Time Series Anal.* **32** 498–511. MR2835683
[29] Shumway, R.H. and Stoffer, D.S. (2011). *Time Series Analysis and Its Applications. With R Examples*, 3rd ed. *Springer Texts in Statistics*. New York: Springer. MR2721825
[30] Srivastava, M.S. and Wu, Y. (1999). Quasi-stationary biases of change point and change magnitude estimation after sequential CUSUM test. *Sequential Anal.* **18** 203–216. MR1737186
[31] Tsay, R.S. (1988). Outliers, level shifts, and variance changes in time series. *J. Forecast.* **7** 1–20.
[32] Wichern, D.W., Miller, R.B. and Hsu, D.A. (1976). Changes of variance in first order autoregressive time series models – With an application. *Appl. Statist.* **25** 248–256.
[33] Wu, Y. (2005). *Inference for Change-Point and Post-Change Means After a CUSUM Test. Lecture Notes in Statistics* **180**. New York: Springer. MR2142337
[34] Yu, H. (2007). High moment partial sum processes of residuals in ARMA models and their applications. *J. Time Ser. Anal.* **28** 72–91. MR2332851

Received April 2013 and revised July 2013