Test (2008) 17: 515–530
DOI 10.1007/s11749-006-0041-7

ORIGINAL PAPER

Monitoring shifts in mean: Asymptotic normality of stopping times

Alexander Aue · Lajos Horváth · Piotr Kokoszka · Josef Steinebach

Received: 6 March 2006 / Accepted: 23 October 2006 /
Published online: 27 February 2007

Abstract We consider a sequential procedure designed to detect a possible change in the mean of a random sequence. The procedure is motivated by the problem of detecting an early change in the mean of returns and is based on the CUSUM (cumulative sum) statistic. It terminates when the CUSUM crosses a boundary function or the number of collected observations reaches a prescribed number. We show that the stopping time is asymptotically normal. Simulations using models derived from returns on indexes and individual stocks show that the normal approximation holds in finite samples.

Keywords Asymptotic normality · Change in the mean · CUSUM · Sequential detection

Mathematics Subject Classification (2000) Primary 62L15 · Secondary 62E20

This research has partially been supported by NATO grant PST.EAP.CLG 980599, NSF grants INT-0223262, DMS-0413653, DMS-0604670.

A. Aue
Department of Mathematical Sciences, Clemson University, Clemson, USA

L. Horváth
Department of Mathematics, University of Utah, Salt Lake City, USA

P. Kokoszka (✉)
Department of Mathematics and Statistics, Utah State University, 3900 Old Main Hill, Logan, UT 84322-3900, USA
e-mail: Piotr.Kokoszka@usu.edu

J. Steinebach
Mathematisches Institut, Universität zu Köln, Cologne, Germany

1 Introduction

We study the "change in the mean" model defined as

$$X_i = \begin{cases} \varepsilon_i + \mu, & 1 \le i < m + k^*, \\ \varepsilon_i + \mu + \Delta_m, & m + k^* \le i < \infty, \end{cases}$$

where $k^* \ge 1$ and $\{\varepsilon_i\}$ is an error sequence to be specified below. If $\Delta_m \ne 0$, then a change in mean occurs at time $m + k^*$. The means before and after the possible change are unknown and k^*, the time of the possible change, is also not given.

We are interested in on-line monitoring for a change point k^*. We consider the case of small k^*, i.e., we assume that a change occurs soon after the monitoring has commenced. Such a formulation is, for example, relevant whenever there is a need to quickly detect a change in the mean returns which corresponds to a change in the behavior of the stock price from growth to a sudden decline, or vice versa.

Building on schemes of Chu et al. (1996), a family of monitoring procedures was developed by Horváth et al. (2004). In these procedures the first m observations are used as a training sample and asymptotics are established as m tends to infinity. In this paper we further study a modification proposed in Horváth et al. (2007), which was specifically designed to detect early changes. We show that the detection time is asymptotically normal, as $m \to \infty$. We are not aware of similar results formulated in the context of traditional sequential monitoring; typically only the expected value of the detection time, known as the Average Run Length, can be computed, see Basseville and Nikiforov (1993) for an extensive overview. The paradigm of Chu et al. (1996) and Horváth et al. (2004, 2007) allows however to derive more complex distributional properties of the detection time, the asymptotic normality being an important result.

Our procedure is based on the CUSUM (cumulative sum) statistic because this statistic leads to procedures with good power when k^* is small. Leisch et al. (2000) and Zeileis et al. (2005) proposed MOSUM (moving sum) procedures to detect changes. Their method works better if the change occurs later in the sequence. Leisch et al. (2000) and Zeileis et al. (2005) defined "closed end" stopping rules, i.e., procedures that will terminate after a prescribed number of steps even under the no change null hypothesis. The procedure we study here is a closed end procedure. While Leisch et al. (2000) and Zeileis et al. (2005) concentrated on empirical adjustments, our focus is on rigorous mathematical justification. The details of the practical application of our procedures remain to be developed in a suitable empirical context.

We note that the references discussed above address only the related approaches to detecting changes in the structure of a stochastic process, focusing on those which have gained acceptance in econometric literature. There is a fast and rapidly growing body of research on detecting changes in industrial and engineering contexts which focuses on methods based on tracking, filtering and smoothing, see Gustafsson (2000). These attractive methods are extensively used in digital signal processing, but have not gained a foothold in econometrics yet. A promising contribution in this direction is Gencay et al. (2002).

In the remainder of this section we provide a more detailed background.

Horváth et al. (2004) studied the stopping time

$$\tau_m = \inf\{k \geq 1 : \Gamma(m,k) \geq g(m,k)\},$$

with the understanding $\tau_m = \infty$ if $\Gamma(m,k) < g(m,k)$ for all $k = 1, 2, \ldots$. The detector $\Gamma(m,k)$ and the boundary function $g(m,k)$ were chosen so that under the no change null hypothesis ($\Delta_m = 0$),

$$\lim_{m\to\infty} P\{\tau_m < \infty\} = \alpha, \tag{1.1}$$

where $0 < \alpha < 1$ is a prescribed number, and under the alternative ($\Delta_m \neq 0$),

$$\lim_{m\to\infty} P\{\tau_m < \infty\} = 1. \tag{1.2}$$

By (1.1), as in the Neyman–Pearson test, the probability of a false alarm and, hence, stopping when we should not is asymptotically α.

Horváth et al. (2004) used the detector

$$\Gamma(m,k) = \hat{Q}(m,k) = \frac{1}{\hat{\sigma}_m}\left|\sum_{m<i\leq m+k}(X_i - \overline{X}_m)\right|, \tag{1.3}$$

where

$$\overline{X}_m = \frac{1}{m}\sum_{1\leq i\leq m} X_i$$

and where $\hat{\sigma}_m^2$ is an asymptotically consistent estimator for

$$\sigma^2 = \lim_{n\to\infty} \mathrm{Var}\left(\sum_{1\leq i\leq n}\varepsilon_i\right)/n.$$

The boundary function was chosen as

$$g(m,k) = cm^{1/2}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^{\gamma}, \quad 0 \leq \gamma < 1/2. \tag{1.4}$$

Under regularity conditions and different assumptions on the errors $\{\varepsilon_i\}$, Horváth et al. (2004) and Aue et al. (2006b) proved that under H_0 ($\Delta_m = 0$),

$$\lim_{m\to\infty} P\{\tau_m < \infty\} = P\left\{\sup_{0<t\leq 1}\frac{|W(t)|}{t^{\gamma}} \geq c\right\}, \tag{1.5}$$

where $\{W(t), 0 \leq t < \infty\}$ is a standard Wiener process (Brownian motion). Aue and Horváth (2004) showed that

$$\lim_{m\to\infty} P\left\{\frac{\tau_m - a_m}{b_m} \leq x\right\} = \Phi(x), \tag{1.6}$$

where Φ denotes the standard normal distribution function,

$$a_m = \left(\frac{cm^{\frac{1}{2}-\gamma}}{|\Delta_m|}\right)^{\frac{1}{1-\gamma}} \quad \text{and} \quad b_m = \frac{\sigma}{1-\gamma}\frac{1}{|\Delta_m|}a_m^{1/2}.$$

Relation (1.6) implies that

$$\tau_m \approx \left(\frac{c}{|\Delta_m|}\right)^{\frac{1}{1-\gamma}} m^{\frac{1-2\gamma}{2(1-\gamma)}},$$

so the reaction time to the change is (asymptotically) small, if γ is close to $1/2$. Naturally, we want to stop as early as possible after the change in the mean, so we might want to use the boundary function $g(m,k)$ with $\gamma = 1/2$. Note, however, that $\gamma = 1/2$ is not allowed in the definition (1.4). This is because, by the law of the iterated logarithm at zero, (1.5) cannot hold when $\gamma = 1/2$. It can be readily verified that, if we used $\gamma = 1/2$ in the definition of τ_m, the procedure would stop with probability one even under H_0. Horváth et al. (2007), therefore, proposed the following modification of τ_m:

$$\tau_m^* = \inf\{k : 1 \le k \le N,\ \hat{Q}(m,k) \ge c(m;t)g^*(m,k)\},$$

with $\tau_m^* = N$ if $\hat{Q}(m,k) < c(m;t)g^*(m,k)$ for all $1 \le k \le N$, where $N = N(m)$ depends on the training sample size m and satisfies condition (1.9). The functions $g^*(m,k)$ and $c(m;t)$ are defined by

$$g^*(m,k) = m^{1/2}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^{1/2} \tag{1.7}$$

and

$$c(m;t) = \frac{t + D(\log m)}{A(\log m)}, \tag{1.8}$$

with

$$A(x) = (2\log x)^{1/2}, \qquad D(x) = 2\log x + \frac{1}{2}\log\log x - \frac{1}{2}\log \pi.$$

Comparing τ_m^* to τ_m, one can observe that, apart from using a finite monitoring horizon and $\gamma = 1/2$ instead of $\gamma < 1/2$, the constant c in (1.5) is replaced with a sequence $c(m;t)$. The asymptotic properties of τ_m^* were studied by Horváth et al. (2007) under the no change null hypothesis and the assumptions of independent, identically distributed errors. For ease of reference, we state here the main result of Horváth et al. (2007).

Theorem 1.1 *If $\{\varepsilon_i, 1 \le i < \infty\}$ are independent, identically distributed random variables with $E\varepsilon_i = 0$, $E\varepsilon_i^2 = \sigma^2 > 0$, $E|\varepsilon_i|^\nu < \infty$ with some $\nu > 2$, $\Delta_m = 0$ and*

$$c_1 m \le N \le c_2 m^\lambda \quad \text{with some } c_1, c_2 > 0 \text{ and } 1 \le \lambda < \infty, \tag{1.9}$$

then

$$\lim_{m\to\infty} P\big\{\hat{Q}(m,k) < c(m;t)g^*(m,k)\ \textit{for all}\ 1 \le k \le N\big\} = \exp\big(-e^{-t}\big). \tag{1.10}$$

The limit result in Theorem 1.1 is a Darling–Erdős type extreme value limit theorem and it is related to the distribution of $W(t)/t^{1/2}$. To apply it we choose t in (1.8) such that $\exp(-e^{-t}) = 1 - \alpha$. This implies that under the null hypothesis

$$\lim_{m\to\infty} P\big\{\tau_m^* < \infty\big\} = \lim_{m\to\infty} P\big\{\text{“stopping in } N \text{ steps”}\big\} = \alpha.$$

In the present paper we focus on the behavior of τ_m^* under the alternative. Moreover, we allow the errors ε_i to be dependent. We show that under weak assumptions on the errors, which permit serial dependence and conditional heteroskedasticity, the stopping time τ_m^* is asymptotically normally distributed. We use this result to conclude that the delay time of τ_m^* is asymptotically smaller than the delay time of any τ_m with $\gamma < 1/2$. This provides theoretical justification for the simulation results reported in Horváth et al. (2007). These asymptotic results are stated in Sect. 2. Section 3 examines the normality of τ_m^* for finite values of m and conditionally heteroskedastic models derived from estimating returns on market indexes and individual stocks. The proof of our main result, Theorem 2.2, is presented in Sect. 4.

2 Main results

In order to accommodate applications, we require much weaker conditions than the independence of the innovations $\varepsilon_1, \varepsilon_2, \ldots$. Namely, we assume that

$$\begin{aligned}&\varepsilon_1, \varepsilon_2, \ldots \text{ form a strictly stationary sequence with}\\ &E\varepsilon_1 = 0 \quad \text{and} \quad E\varepsilon_1^2 < \infty,\end{aligned} \tag{2.1}$$

$$\max_{1\le k\le n} \frac{1}{k^{1/2}} \left|\sum_{i=1}^{k} \varepsilon_i\right| = \mathcal{O}_P\big((\log\log n)^{1/2}\big) \quad (n \to \infty), \tag{2.2}$$

$$\frac{1}{\sigma n^{1/2}} \sum_{1\le i\le nt} \varepsilon_i \xrightarrow{D[0,1]} W(t) \quad (n\to\infty) \text{ with some } \sigma > 0, \tag{2.3}$$

where $\{W(t), 0 \le t < \infty\}$ denotes a standard Wiener process.

Condition (2.3) means that the innovations satisfy the functional central limit theorem, while (2.2) is a weak version of the law of the iterated logarithm.

If the $\{\varepsilon_i\}$ are independent and identically distributed, then (2.2) and (2.3) hold. If $\{\varepsilon_i\}$ are homoskedastic and serially dependent, e.g., if they satisfy ARMA equations, then (2.2) and (2.3) are also satisfied due to the mixing properties of these sequences. Gorodetskii (1977) and Withers (1981) established the mixing properties of linear processes assuming a smoothness condition on the density of the summands. Mixing properties of augmented GARCH sequences were established by Carrasco and Chen (2002) assuming moment and smoothness conditions on the errors. Aue et al. (2006a) provides another justification for (2.2) and (2.3) in case of augmented GARCH sequences under very mild conditions. The strong approximation result for errors following a random coefficient autoregressive model, established in Aue (2006), implies (2.2) and (2.3), too.

The parameter σ in (2.3) is equal to the square root of the long run variance $\sigma^2 = \lim_{n\to\infty} \operatorname{Var}(\sum_{1\le i\le n} \varepsilon_i)/n$. If the ε_i are uncorrelated, e.g., if they follow a GARCH type model, σ^2 can be estimated by the sample variance of the first m residuals. If the ε_i are serially correlated, different estimators are needed, but discussing them would distract us from the focus of this paper, see Andrews and Monahan (1992), Giraitis et al. (2003), Berkes et al. (2005, 2006) and Sul et al. (2005) for some recent contributions which contain a wealth of references.

We first state an extension of Theorem 1.1

Theorem 2.1 *Suppose* $\{\varepsilon_i, 1 \le i < \infty\}$ *satisfy assumptions* (2.1–2.3). *If* $\Delta_m = 0$ *and the condition* (1.9) *is satisfied then the relation* (1.10) *holds.*

Theorem 2.1 can be established by modifying the technique developed in Horváth et al. (2007). For this reason we omit its proof.

We now turn to the main result of the paper, the asymptotic normality of τ_m^*. Simulations in Horváth et al. (2007) showed that the monitoring scheme works well if k^* is small ("early change") and Δ_m, the size of the change, is not too small. Thus, we assume

$$c_3(\log m)^{-\beta} \le |\Delta_m| \le c_4 \quad \text{for some } c_3, c_4 > 0 \text{ and } \beta > 0 \tag{2.4}$$

and

$$k^* = o\left(\frac{(\log\log m)^{1/2}}{\Delta_m^2}\right). \tag{2.5}$$

Condition (2.5) means that we are trying to detect an early change. This assumption is also justified by the observation in Chu et al. (1996) and Zeileis et al. (2005) that CUSUM based procedures should be used to detect early changes.

Theorem 2.2 *If* (1.9) *and* (2.1–2.5) *are satisfied, then for all* x

$$\lim_{m\to\infty} P\left\{\frac{\Delta_m^2}{2(2\log\log m)^{1/2}}\left(\frac{\tau_m^*}{\sigma^2} - \frac{2\log\log m}{\Delta_m^2}\right) \le x\right\} = \Phi(x). \tag{2.6}$$

If, in addition,

$$\hat\sigma_m - \sigma = o_P(1), \quad \text{as } m \to \infty, \tag{2.7}$$

then

$$\lim_{m\to\infty} P\left\{\frac{\Delta_m^2}{2(2\log\log m)^{1/2}}\left(\frac{\tau_m^*}{\hat\sigma_m^2} - \frac{2\log\log m}{\Delta_m^2}\right) \le x\right\} = \Phi(x). \tag{2.8}$$

Theorem 2.2 is proved in Sect. 4.

Observe that, under the conditions of Theorem 2.2, we immediately have

$$\tau_m^* \approx \frac{2\sigma^2 \log\log m}{\Delta_m^2}. \tag{2.9}$$

This means that we stop in

$$2\sigma^2 \log\log m / \Delta_m^2$$

steps using τ_m^*, while

$$\big(c/|\Delta_m|\big)^{1/(1-\gamma)} m^{(1-2\gamma)/(2(1-\gamma))}$$

steps are needed with τ_m. In particular, if Δ_m is constant, we have an iterated logarithmic order of the detection time using τ_m^* as compared to a polynomial growth of τ_m.

We also note that relations (2.6) and (2.8), and so (2.9), do not involve the change-point k^*, and one would expect the mean detection time to increase with k^*. This is because the goal of Theorem 2.2 is to establish the asymptotic normality of k^* and its rate as a function of the size m of the training sample, rather than the precise parameters of the limiting normal distribution. Observe that, by (2.5), k^* can be added to the mean without changing the asymptotic distribution. For this reason we investigate in Sect. 3 whether the distribution of k^* is normal and how the approximation to normality depends on m, for finite values of m, without using the specific mean and variance of Theorem 2.2.

3 Numerical examples

The goal of the monitoring is to detect a relatively large change in the mean returns which occurs not long after the monitoring has commenced because only a change larger than a certain threshold would necessitate some action, e.g., a reallocation of portfolio assets. The underlying model for asset returns, e.g., a GARCH model considered below, would have to be updated every week or so, and an aposteriori test of the stability of this model would have to be performed. Once such a test has been performed and the stability of the training sample ascertained, the monitoring would start. If a change-point were detected, the approximate normality could, in principle, be used to construct a confidence interval for the time of change. Knowing the range of likely values for the change-point might be useful in reevaluating positions on option-type instruments. Even though these issues are well beyond the focus and scope of the present paper, it is hoped that they will be addressed in future research. The objective of this section is to explore the approximate normality of $\tau^* = \tau_m^*$ for finite values of m emphasizing the empirical coverage of confidence intervals based on normal quantiles. We do not attempt to provide a definite guide for practitioners but merely illustrate the potential applicability of the theory developed in this paper.

We first consider the simplest case of independent identically distributed standard normal errors ε_i. Figure 1 shows the histograms of 2500 simulated values of τ^* for $\Delta = 1.2$, $m = 200, 400, 600$ and $k^* = 3$ and $k^* = 5$. The choice of the values of m is motivated by reasonable sizes of training samples encountered in financial time series

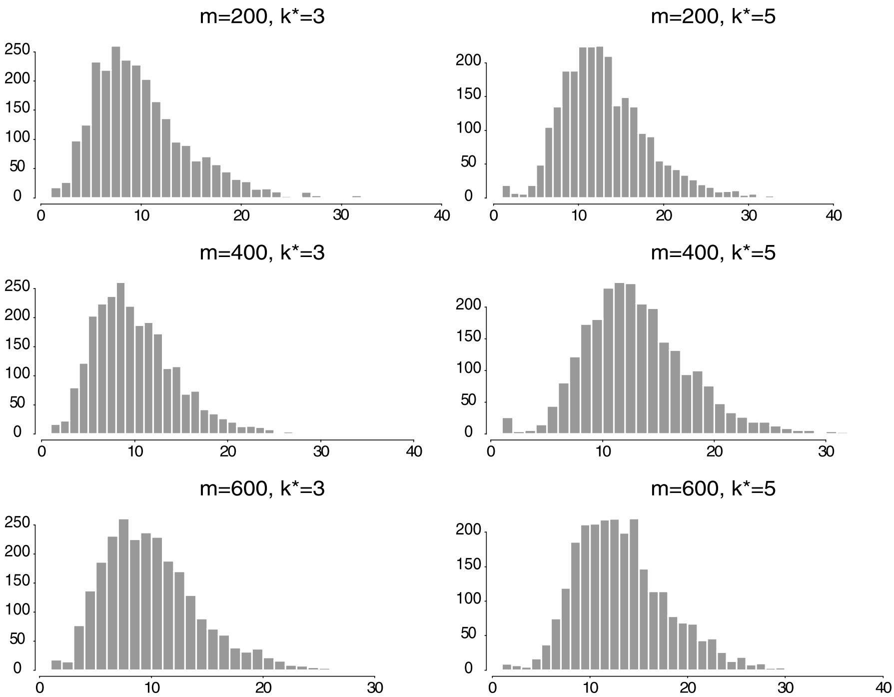

Fig. 1 Histograms of the detection time τ^*

to be discussed below. The values of $k^* = 3$ and $k^* = 5$ were chosen as reasonable early changes. For series of daily returns, they correspond to a change in mean within the first week after the monitoring has commenced.

The histograms look roughly normal but are visibly skewed to the right and lack some probability mass in the left tail. This is to be expected, as the detection delay can, in principle, be arbitrarily long, whereas a detection before the change point is unlikely and τ^* cannot be smaller than 1. These observations are illustrated in a different way in Fig. 2 which shows QQ-plots against the quantiles of the standard normal distribution. If the distribution of τ^* were normal, these plots would approximately be straight lines. Figure 2 shows that the central part of the QQ-plots becomes closer to a straight line, as m increases. There is a pronounced "step" at the left tail due the aforementioned boundedness of τ^*.

Asymptotic normality of the distribution of a statistic is typically used to construct confidence intervals and to test statistical hypotheses. In such applications it is only important that the empirical quantiles are close to the corresponding critical values derived from the normal approximation. It is, thus, useful to know if

$$P\Big[\big|\tau^* - E\tau^*\big| \le z_{\alpha/2}\sqrt{\mathrm{Var}\big[\tau^*\big]}\Big] \approx 1 - \alpha,$$

where the standard normal critical value $z_{\alpha/2}$ is defined by $\Phi(z_{\alpha/2}) = 1 - \alpha/2$. Table 1 shows the empirical probabilities that $|\tau^* - E\tau^*| \le z_{\alpha/2}\sqrt{\mathrm{Var}[\tau^*]}$ for $\alpha = 0.05$ and $\alpha = 0.01$, for the simulated values of τ^* used to construct Figs. 1 and 2. The ex-

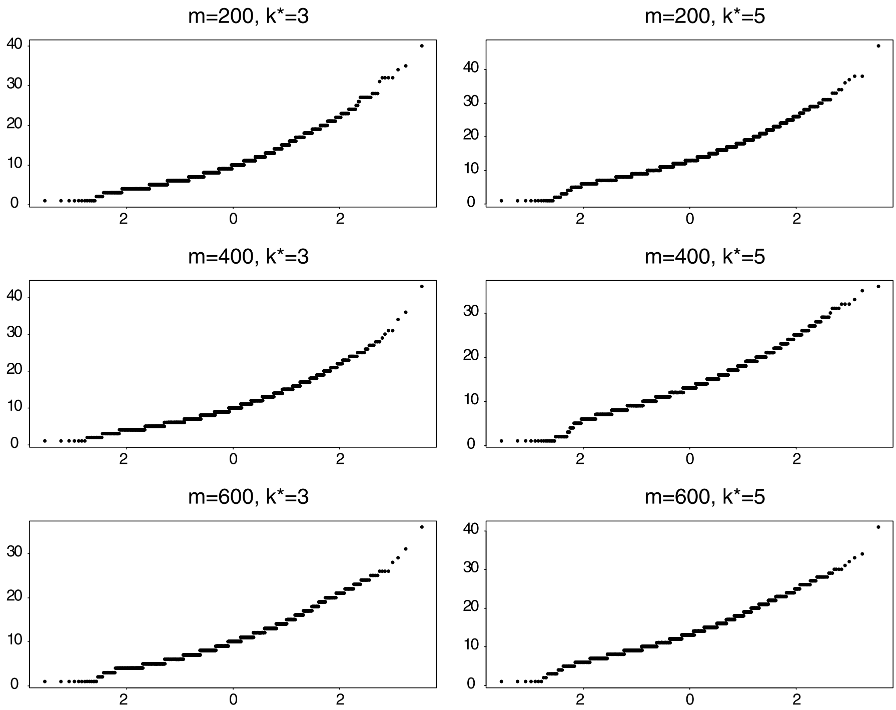

Fig. 2 QQ plots of the detection time τ^*

Table 1 Percentage of stopping times τ^* within 1.960 and 2.576 standard deviations away from the sample mean. Standard normal errors. $\Delta = 1.2$

	$m = 200$		$m = 400$		$m = 600$	
Nominal	0.9500	0.9900	0.9500	0.9900	0.9500	0.9900
Empirical $k^* = 3$	0.9612	0.9848	0.9524	0.9796	0.9528	0.9816
Empirical $k^* = 5$	0.9444	0.9800	0.9524	0.9764	0.9564	0.9852

pected value $E\tau^*$ was replaced by the sample mean and $\mathrm{Var}[\tau^*]$ by the sample variance. It is seen that in this sense the normal approximation is excellent for $\alpha = 0.05$ and reasonably good for $\alpha = 0.01$ in which case the tails of the empirical distribution are slightly heavier (by about 1% of the total probability mass) than implied by the normal approximation. The empirical probabilities in Table 1 (and in Table 2 below) are based on $R = 2500$ replications, so they are unlikely to differ from the probabilities based on $R = \infty$ replications by more than 0.01 (standard errors are about 0.005).

As explained in Sect. 2, the theory developed in this paper allows conditionally heteroskedastic errors ε_i. We now investigate if the normal approximation holds in case of GARCH(1, 1) errors and values of m relevant for the series of daily returns. To conserve space we focus on the analysis analogous to that presented in Table 1; the histograms and QQ-plots are very similar to those in Figs. 1 and 2. We present the results only for $k^* = 5$ what corresponds to a change a week after the monitoring has commenced. We, however, consider four different GARCH(1, 1) models esti-

Table 2 Percentage of stopping times τ^* within 1.645, 1.960 and 2.576 standard deviations away from the sample mean. GARCH(1, 1) errors. The shift Δ is expressed in units equal to the implied standard deviations of the GARCH(1.1) errors, i.e., to $\sqrt{\omega/(1-\alpha-\beta)}$

	DJIA $m = 1061$			NASDAQ $m = 775$		
Nominal	0.9000	0.9500	0.9900	0.9000	0.9500	0.9900
Empirical $\Delta = 0.4$	0.9352	0.9532	0.9780	0.9284	0.9560	0.9764
Empirical $\Delta = 0.8$	0.9184	0.9580	0.9816	0.9064	0.9432	0.9852
Empirical $\Delta = 1.2$	0.9032	0.9536	0.9832	0.9112	0.9428	0.9768
Empirical $\Delta = 1.6$	0.9164	0.9492	0.9764	0.9204	0.9528	0.9736
Empirical $\Delta = 2.0$	0.9196	0.9656	0.9788	0.9280	0.9412	0.9788
	GE $m = 473$			WALMART $m = 457$		
Nominal	0.9000	0.9500	0.9900	0.9000	0.9500	0.9900
Empirical $\Delta = 0.4$	0.9424	0.9620	0.9756	0.9468	0.9640	0.9792
Empirical $\Delta = 0.8$	0.9212	0.9492	0.9832	0.9184	0.9500	0.9840
Empirical $\Delta = 1.2$	0.9092	0.9488	0.9804	0.9140	0.9488	0.9792
Empirical $\Delta = 1.6$	0.9144	0.9520	0.9764	0.9180	0.9500	0.9772
Empirical $\Delta = 2.0$	0.9216	0.9380	0.9800	0.9072	0.9508	0.9692

mated on index and stock returns with different dynamics and five values of the shift $\Delta = \Delta_m$. The values of m are equal to those determined in Zhang et al. (2006) and reflect periods of time over which the models are believed to be stable. In the simulations we used the following four models which well represent the various dynamics of return series:

DJIA, $m = 1061$ trading days starting 1/1/1992:
$\mu = 0.076$; $\omega = 0.025$; $\alpha = 0.064$; $\beta = 0.879$. Implied variance 0.439;

NASDAQ, $m = 775$ trading days starting 7/1/1994:
$\mu = 0.129$; $\omega = 0.133$; $\alpha = 0.132$; $\beta = 0.750$. Implied variance 1.127;

GE, $m = 473$ trading days starting 1/1/2001:
$\mu = -0.130$; $\omega = 4.345$; $\alpha = 0.094$; $\beta = 0.281$. Implied variance 6.952;

WALMART, $m = 457$ trading days starting 1/1/2001:
$\mu = 0.027$; $\omega = 1.407$; $\alpha = 0.172$; $\beta = 0.499$. Implied variance 4.227.

The parameter μ is equal to the mean return over the indicated period. The parameters ω, α and β are the usual GARCH(1, 1) parameters in the specification $\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta\sigma_{t-1}^2$ estimated on the demeaned returns. The implied variance is equal to $\omega/(1-\alpha-\beta)$.

Table 2 shows results similar to those presented in Table 1. We also included nominal coverage of 0.90, in which case the empirical distribution tends to have slightly too much mass (about 2%). This is also the case for independent errors. For the nominal coverage of 0.99, some mass is missing (about 1%). As for independent errors, this is due to slightly too heavy right tails. The empirical coverage corresponding to the nominal coverage of 0.95 continues to be remarkably good except in the cases of very small or very large changes in mean returns in models corresponding to individual stocks. This is in accordance with the condition (2.4) which requires that the shift cannot be too small or two large. For all models considered in this simulation study, the empirical coverage is excellent provided Δ is larger than 0.8 and smaller than 1.6 standard deviations of the returns (for all models the implied standard deviation is practically equal to the sample standard deviation). In most cases the quantiles computed using the normal approximations are quite close to the empirical quantiles.

4 Proof of Theorem 2.2

Recall the definitions of the detector $\hat{Q}(m,k)$ and the threshold $g^*(m,k)$ in (1.3) and (1.7), respectively. It is easy to see that

$$\sum_{m<i\le m+k}(X_i-\overline{X}_m)=\sum_{m<i\le m+k}(\varepsilon_i-\overline{\varepsilon}_m)+\big(k-k^*+1\big)\Delta_m I\big\{k\ge k^*\big\},\qquad(4.1)$$

where

$$\overline{\varepsilon}_m=\frac{1}{m}\sum_{1\le i\le m}\varepsilon_i.$$

Furthermore, we can assume without loss of generality that $\Delta_m>0$.

The proof of Theorem 2.2 is based on a series of lemmas, where the following quantity is frequently used. For any x and t, we introduce

$$K(m)=K(m,x,t)=\left(\frac{\hat{\sigma}_m c(m;t)-\sigma x}{\Delta_m}\right)^2.$$

In the following we focus on the verification of (2.8). The verification of (2.6) follows exactly the same lines except that $\hat{\sigma}_m$, in the definition of $K(m)$, is replaced by σ and the condition (2.7) is then not needed.

It follows from the definition of $c(m;t)$ in (1.8) and from the condition (2.7) that

$$K(m)\left(\frac{2\sigma^2\log\log m}{\Delta_m^2}\right)^{-1}\xrightarrow{P}1\quad(m\to\infty).\qquad(4.2)$$

In particular, $K(m)$ is of lower order than N whose range is determined by (1.9).

The proof steps can be outlined as follows. In Lemma 4.1 we show that the threshold function $g^*(m,k)$ can be replaced asymptotically by the simpler expression $k^{1/2}$. Lemma 4.2 simplifies the CUSUM-type expression in (4.1) by dropping the average $\overline{\varepsilon}_m$. Then Lemma 4.3 determines the range of those k which contribute asymptotically, while the normal limit distribution is eventually established in Lemmas 4.4 and 4.5. Theorem 2.2 follows readily from these auxiliary results.

Lemma 4.1 *If the conditions of Theorem 2.2 are satisfied then, as $m\to\infty$,*

$$\max_{1\le k\le K(m)}\frac{1}{g^*(m,k)}\Bigg|\sum_{m<i\le m+k}(X_i-\overline{X}_m)\Bigg|$$
$$=\max_{1\le k\le K(m)}\frac{1}{k^{1/2}}\Bigg|\sum_{m<i\le m+k}(X_i-\overline{X}_m)\Bigg|+o_P(1).$$

Proof By applying (4.1) and the mean-value theorem, it is clear that

$$\max_{1\le k\le K(m)} \frac{1}{k^{1/2}}\left|\sum_{m<i\le m+k}(X_i-\overline{X}_m)\right|\left|\left(1+\frac{k}{m}\right)^{-1/2}-1\right|$$
$$=\mathcal{O}(1)\max_{1\le k\le K(m)}\frac{k^{1/2}}{m}\left|\sum_{m<i\le m+k}(\varepsilon_i-\overline{\varepsilon}_m)+\big(k-k^*+1\big)\Delta_m I\big\{k\ge k^*\big\}\right|.$$

It follows from the stationarity of $\{\varepsilon_i\}$ and the weak convergence in (2.3) that, as $m\to\infty$,

$$\begin{aligned}\max_{1\le k\le K(m)}\frac{k^{1/2}}{m}\left|\sum_{m<i\le m+k}\varepsilon_i\right| &\stackrel{D}{=}\max_{1\le k\le K(m)}\frac{k^{1/2}}{m}\left|\sum_{1\le i\le k}\varepsilon_i\right|\\ &\le\frac{K(m)}{m}\frac{1}{K(m)^{1/2}}\max_{1\le k\le K(m)}\left|\sum_{1\le i\le k}\varepsilon_i\right|\\ &=\mathcal{O}_P(1)\frac{\log\log m}{m\Delta_m^2}=o_P(1),\end{aligned}$$

due to (4.2), since the order of magnitude of Δ_m is controlled by (2.4). The same arguments imply similarly, as $m\to\infty$,

$$\max_{1\le k\le K(m)}\frac{k^{1/2}}{m}k|\overline{\varepsilon}_m|=\mathcal{O}_P(1)\frac{K(m)^{3/2}}{m^{3/2}}=o_P(1)$$

and

$$\begin{aligned}\max_{1\le k\le K(m)}\frac{k^{1/2}}{m}\big(k-k^*+1\big)\Delta_m I\big\{k\ge k^*\big\}&=\mathcal{O}(1)\frac{K(m)^{3/2}}{m}\Delta_m\\ &=\mathcal{O}_P(1)\frac{(\log\log m)^{3/2}}{m\Delta_m^2}=o_P(1),\end{aligned}$$

completing the proof of Lemma 4.1. □

Lemma 4.2 *If the conditions of Theorem 2.2 are satisfied then, as $m\to\infty$,*

$$\max_{1\le k\le K(m)}\frac{1}{k^{1/2}}\left|\sum_{m<i\le m+k}(X_i-\overline{X}_m)\right|$$
$$=\max_{1\le k\le K(m)}\frac{1}{k^{1/2}}\left|\sum_{m<i\le m+k}\varepsilon_i+\big(k-k^*+1\big)\Delta_m I\big\{k\ge k^*\big\}\right|+o_P(1).$$

Proof Conditions (2.3) and (2.4) give

$$\max_{1\le k\le K(m)}\frac{1}{k^{1/2}}k|\overline{\varepsilon}_m|=\mathcal{O}_P\left(\left(\frac{K(m)}{m}\right)^{1/2}\right)=o_P(1)\quad(m\to\infty),$$

so that the result is readily proved. □

Lemma 4.3 *If the conditions of Theorem 2.2 are satisfied then, for any* $0 < \delta < 1$, *as* $m \to \infty$,

$$\max_{1 \le k \le (1-\delta)K(m)} \frac{1}{k^{1/2}} \left| \sum_{m < i \le m+k} \varepsilon_i + \big(k - k^* + 1\big)\Delta_m I\big\{k \ge k^*\big\} \right| - K(m)^{1/2}\Delta_m$$

diverges in probability to $-\infty$.

Proof Using (2.2) and the stationarity of $\{\varepsilon_i\}$ we conclude

$$\max_{1 \le k \le K(m)} \frac{1}{k^{1/2}} \left| \sum_{m < i \le m+k} \varepsilon_i \right| = \mathcal{O}_P\big((\log\log\log m)^{1/2}\big),$$

where relation (2.4) has been used to obtain the last estimate. Also,

$$\max_{1 \le k \le (1-\delta)K(m)} \frac{1}{k^{1/2}} \big(k - k^* + 1\big)\Delta_m I\big\{k \ge k^*\big\} \le \big((1-\delta)K(m)\big)^{1/2}\Delta_m.$$

On observing that, by (4.2), $K(m)^{1/2}\Delta_m \xrightarrow{P} \infty$, as $m \to \infty$, the proof of Lemma 4.3 is complete. □

Lemma 4.4 *If the conditions of Theorem 2.2 hold then, for all* x,

$$\lim_{\delta \to 0} \lim_{m \to \infty} P\big\{A_{\delta,m}(x)\big\} = \Phi(x),$$

where

$$A_{\delta,m}(x) = \left\{ \max_{(1-\delta)K(m) \le k \le K(m)} \frac{1}{k^{1/2}} \left(\sum_{m < i \le m+k} \varepsilon_i + k\Delta_m \right) - K(m)^{1/2}\Delta_m \le \sigma x \right\}.$$

Proof By the stationarity of $\{\varepsilon_i\}$ it is enough to prove

$$\lim_{\delta \to 0} \lim_{m \to \infty} P\big\{A^*_{\delta,m}(x)\big\} = \Phi(x),$$

where

$$A^*_{\delta,m}(x) = \left\{ \max_{(1-\delta)K(m) \le k \le K(m)} \frac{1}{k^{1/2}} \left(\sum_{1 \le i \le k} \varepsilon_i + k\Delta_m \right) - K(m)^{1/2}\Delta_m \le \sigma x \right\}.$$

Let $\lfloor \cdot \rfloor$ denote the integer part. Clearly, if $K(m) \ge 1$,

$$\max_{(1-\delta)K(m) \le k \le K(m)} \frac{1}{k^{1/2}} \left(\sum_{1 \le i \le k} \varepsilon_i + k\Delta_m \right) - K(m)^{1/2}\Delta_m$$

$$\ge \frac{1}{\lfloor K(m) \rfloor^{1/2}} \sum_{1 \le i \le \lfloor K(m) \rfloor} \varepsilon_i - \big(K(m)^{1/2} - \lfloor K(m) \rfloor^{1/2}\big)\Delta_m.$$

Since, by the mean value theorem,

$$K(m)^{1/2} - \lfloor K(m) \rfloor^{1/2} = \mathcal{O}_P\big(K(m)^{-1/2}\big) = o_P(1),$$

we see from (2.3) and (2.4) that

$$P\left\{\frac{1}{\sigma \lfloor K(m) \rfloor^{1/2}} \sum_{1 \le i \le \lfloor K(m) \rfloor} \varepsilon_i - \big(K(m)^{1/2} - \lfloor K(m) \rfloor^{1/2}\big)\Delta_m \le x\right\}$$

converges to $\Phi(x)$, as $m \to \infty$. On the other hand,

$$\begin{aligned}
&\max_{(1-\delta)K(m) \le k \le K(m)} \frac{1}{k^{1/2}} \left(\sum_{1 \le i \le k} \varepsilon_i + k\Delta_m\right) - K(m)^{1/2}\Delta_m \\
&\quad \le \max_{(1-\delta)K(m) \le k \le K(m)} \frac{1}{k^{1/2}} \sum_{1 \le i \le k} \varepsilon_i \\
&\quad \le \frac{1}{(1-\delta)^{1/2}} \max_{(1-\delta)K(m) \le k \le K(m)} \frac{1}{K(m)^{1/2}} \sum_{1 \le i \le k} \varepsilon_i .
\end{aligned}$$

Condition (2.3) yields that, for any $0 < \delta < 1$,

$$\max_{(1-\delta)K(m) \le k \le K(m)} \frac{1}{K(m)^{1/2}} \sum_{1 \le i \le k} \varepsilon_i \xrightarrow{D} \sigma \sup_{(1-\delta) \le t \le 1} W(t) \quad (m \to \infty).$$

Since $\sup_{(1-\delta) \le t \le 1} W(t) \xrightarrow{D} W(1)$ as $\delta \to 0$, we get that

$$\lim_{\delta \to 0} \lim_{m \to \infty} P\left\{\frac{1}{(1-\delta)^{1/2}} \max_{(1-\delta)K(m) \le k \le K(m)} \frac{1}{K(m)^{1/2}} \sum_{1 \le i \le k} \varepsilon_i \le \sigma x\right\} = \Phi(x).$$

The proof of Lemma 4.4 is complete. □

Lemma 4.5 *If the conditions of Theorem 2.2 are satisfied then*

$$\lim_{m \to \infty} P\big\{\tau_m^* \ge K(m, x, t)\big\} = \Phi(x),$$

for all x and t.

Proof Let A_m be the event $\{K(m, x, t) \ge N\}$. By (1.9), (2.3) and (4.2), we have

$$\lim_{m \to \infty} P(A_m) = 1. \tag{4.3}$$

Clearly,

$$\begin{aligned}
&P\big\{\tau_m^* \ge K(m), A_m\big\} \\
&\quad = P\left\{\max_{1 \le k \le K(m)} \frac{1}{g^*(m,k)} \left|\sum_{m < i \le m+k} (X_i - \overline{X}_m)\right| \le \hat{\sigma}_m c(m; t), A_m\right\}
\end{aligned}$$

$$= P\left\{\max_{1\le k\le K(m)} \frac{1}{\sigma g^*(m,k)}\left|\sum_{m<i\le m+k}(X_i-\overline{X}_m)\right| - \frac{K(m)^{1/2}\Delta_m}{\sigma}\right.$$
$$\left.\le \frac{\hat{\sigma}_m}{\sigma}c(m;t) - \frac{K(m)^{1/2}\Delta_m}{\sigma}, A_m\right\}.$$

It follows from the definition of $K(m)$ that

$$\frac{\hat{\sigma}_m}{\sigma}c(m;t) - \frac{K(m)^{1/2}\Delta_m}{\sigma} = x,$$

so that, in view of (4.3), it is enough to prove

$$\lim_{m\to\infty} P\left\{\max_{1\le k\le K(m)} \frac{1}{g^*(m,k)}\left|\sum_{m<i\le m+k}(X_i-\overline{X}_m)\right| - K(m)^{1/2}\Delta_m \le \sigma x\right\} = \Phi(x). \tag{4.4}$$

Next observe that, according to Lemmas 4.1–4.3, we get (4.4) if

$$\lim_{\delta\to 0}\lim_{m\to\infty} P\left\{\max_{(1-\delta)K(m)\le k\le K(m)} \frac{1}{k^{1/2}}\left|\sum_{m<i\le m+k}\varepsilon_i + \big(k-k^*+1\big)\Delta_m\right|\right.$$
$$\left. - K(m)^{1/2}\Delta_m \le \sigma x\right\} = \Phi(x).$$

Condition (2.5) together with (4.2) gives $k^*\Delta_m K(m)^{-1/2} = o_P(1)$ and

$$\lim_{m\to\infty} P\left\{\sum_{m<i\le m+k}\varepsilon_i + k\Delta_m > 0 \text{ for all } (1-\delta)K(m)\le k\le K(m)\right\} = 1.$$

Hence, Lemma 4.4 implies Lemma 4.5 □

Proof of Theorem 2.2 The result is an immediate consequence of Lemma 4.5. First we write

$$P\left\{\tau_m^* \ge \left(\frac{\hat{\sigma}_m c(m;t) - \sigma x}{\Delta_m}\right)^2\right\}$$
$$= P\left\{\frac{\hat{\sigma}_m}{\sigma}\frac{\Delta_m^2}{2c(m;t)}\left(\frac{\tau_m^*}{\hat{\sigma}_m^2} - \frac{c(m;t)^2}{\Delta_m^2}\right) \ge -x + \frac{x^2}{2\hat{\sigma}_m\sigma c(m;t)}\right\}.$$

By (2.7) and the definition of $c(m;t)$, we have, as $m\to\infty$,

$$\frac{\hat{\sigma}_m}{\sigma}\xrightarrow{P} 1, \qquad \frac{x^2}{2\hat{\sigma}_m\sigma c(m;t)}\xrightarrow{P} 0, \qquad \frac{c(m;t)}{(2\log\log m)^{1/2}}\longrightarrow 1,$$

and

$$\frac{1}{(2\log\log m)^{1/2}}\left|c(m;t)^2 - 2\log\log m\right| = o(1).$$

Therefore,

$$\lim_{m\to\infty} P\left\{\frac{\Delta_m^2}{2(2\log\log m)^{1/2}}\left(\frac{\tau_m^*}{\hat{\sigma}_m^2} - \frac{2\log\log m}{\Delta_m^2}\right) \geq -x\right\} = \Phi(x),$$

which completes the proof of Theorem 2.2. □

References

Andrews DWK, Monahan JC (1992) An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. Econometrica 60:953–966

Aue A (2006) Testing for parameter stability in RCA(1) time series. J Stat Plan Inference 136:3070–3089

Aue A, Horváth L (2004) Delay time in sequential detection of change. Stat Probab Lett 67:221–231

Aue A, Berkes I, Horváth L (2006a) Strong approximation for the sums of squares of augmented GARCH sequences. Bernoulli 12:583–608

Aue A, Horváth L, Hušková M, Kokoszka P (2006b) Change-point monitoring in linear models. Econom J 9:373–403

Basseville M, Nikiforov IV (1993) Detection of abrupt changes: theory and applications. Prentice Hall, Englewood Cliffs

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

Berkes I, Horváth L, Kokoszka P, Shao Q-M (2006). On discriminating between long-range dependence and changes in mean. Ann Stat 34:1140–1165

Carrasco M, Chen X (2002) Mixing and moment properties of various GARCH and stochastic volatility models. Econom Theory 18:17–39

Chu C-SJ, Stinchcombe M, White H (1996) Monitoring structural change. Econometrica 64:1045–1065

Gencay R, Selcuk F, Whitcher B (2002) An introduction to wavelets and other filtering methods in finance and economics. Academic, San Diego

Giraitis L, Kokoszka P, Leipus R, Teyssiere G (2003) Rescaled variance and related tests for long memory in volatility and levels. J Econom 112:265–294

Gorodetskii VV (1977) On the strong mixing property for linear processes. Theory Probab Appl 22:411–413

Gustafsson F (2000) Adaptive filtering and change detection. Wiley, New York

Horváth L, Hušková M, Kokoszka P, Steinebach J (2004) Monitoring changes in linear models. J Stat Plan Inference 126:225–251

Horváth L, Kokoszka P, Steinebach J (2007) On sequential detection of parameter changes in linear regression. Stat Probab Lett (forthcoming)

Leisch F, Hornik K, Kuan C-M (2000) Monitoring structural changes with the generalized fluctuation test. Econom Theory 16:835–854

Sul D, Phillips PCB, Choi C-Y (2005) Prewhitening bias in HAC estimation. Oxf Bull Econ Stat 67:517–546

Withers CS (1981) Conditions for linear processes to be strong-mixing. Z Wahrsch Verwandte Geb 57:477–480

Zeileis A, Leisch F, Kleiber C, Hornik K (2005) Monitoring structural change in dynamic econometric models. J Appl Econom 20:99–121

Zhang A, Gabrys R, Kokoszka P (2006) Discriminating between long memory and volatility shifts. Preprint