SEQUENTIAL MONITORING FOR CHANGE IN SCALE

Ondřej Chochola

We propose a sequential monitoring scheme for detecting a change in scale. We consider a stable historical period of length m. The goal is to propose a test with asymptotically small probability of false alarm and power 1 as the length of the historical period tends to infinity. The asymptotic distribution under the null hypothesis and consistency under the alternative hypothesis is derived. A small simulation study illustrates the finite sample performance of the monitoring scheme.

Keywords: sequential test for change in scale

AMS Subject Classification: 62J05, 62P20

1. INTRODUCTION

The paper concerns the question of structural stability of a model. Such problems occur in a number of applications, such as in economics and finance, statistical quality control, medical care etc. Precisely speaking the paper is devoted to the detection of a change in scale in a location model when the data arrive sequentially and training (historical) data with no change are available.

We assume that the observations Y_i follow the location model

$$Y_i = \mu + \sigma_i \, e_i, \quad 1 \le i < \infty, \tag{1}$$

where μ is an unknown location parameter, $\{e_i,\ 1 \le i < \infty\}$ are independent identically distributed (i.i.d.) random errors satisfying further conditions specified below and $\{\sigma_i,\ 1 \le i < \infty\}$ are constants determining the variance of the observations $\{Y_i,\ 1 \le i < \infty\}$. Our goal is to monitor the change in variance of the observation, i.e. the change in σ_i.

The observations $Y_1, \ldots, Y_m$ are assumed to represent the training data for which the variance is constant, i.e.

$$\sigma_1^2 = \cdots = \sigma_m^2 = \sigma_0^2,$$

where σ_0^2 is unknown.

Our problem of detection of a change can be formulated as a sequential hypothesis testing problem, where the null hypothesis corresponds to the model without any change:

$$H_0 : \sigma_i^2 = \sigma_0^2, \quad 1 \leq i < \infty \tag{2}$$

against the alternative that the model changes in some unknown time point $m + k^*$:

$$\begin{aligned} &H_1 : \text{there exists } k^* \geq 1 \text{ such that} \\ &\qquad \sigma_i^2 = \sigma_0^2,\ 1 \leq i < m + k^*, \quad \sigma_i^2 = \sigma_*^2,\ m + k^* \leq i < \infty, \quad \sigma_0^2 \neq \sigma_*^2. \end{aligned} \tag{3}$$

The k^*, σ_*^2 and σ_0^2 are unknown.

This represents the so called online monitoring and was originated in [3] where two types of such monitoring procedures in the linear regression settings were studied. The first procedure was based on CUSUM type test statistics calculated from recursive residuals and the second one was the fluctuation test based on differences between estimates of the regression parameters. This test was generalized in [7] to the so called generalized fluctuation test. Similarly to the cumulative sum of residuals one can consider a moving sum. These MOSUM type test statistics were suggested in [10]. All the three described monitoring procedures were compared there through a simulation study which showed that the MOSUM type statistics behave better when the change occurs later in the monitoring period. CUSUM type statistics based on ordinary residuals and on recursive residuals are studied in [4] and also in [6]. A practical side of this testing approach was described in [8], where tests were conducted using the `R` package `strucchange`. This package was presented in [9].

The method used in this paper is related to those introduced in [4] and in [6], however it is adapted to the change in scale.

The rest of the paper is organized as follows: The test procedure is proposed in Section 2, its asymptotic behaviour is given in Section 3. Simulation results are reported in Section 4 and Section 5 contains the proofs.

2. TEST PROCEDURE

Our monitoring scheme is described through the stopping time $\tau(m)$, which is defined as

$$\tau(m) = \inf\{k \geq 1 : \Gamma(m, k) \geq c_m\}, \tag{4}$$

where $\Gamma(m, k)$ (so called detector) depends on observations up to time $m+k$, i.e. on $Y_1, \ldots, Y_{m+k}$. The infimum is defined with a standard understanding that $\inf \emptyset = \infty$. Our test procedure rejects the null hypothesis and we stop observing as soon as $\Gamma(m, k) \geq c_m$, otherwise we continue in observing. The detector $\Gamma(m, k)$ and the constant $c_m = c_m(\alpha)$ are chosen such that the following two conditions are satisfied:

$$\lim_{m\to\infty} \Pr\left(\tau(m) < \infty | H_0\right) \quad \leq \alpha, \tag{5}$$

$$\lim_{m\to\infty} \Pr\left(\tau(m) < \infty | H_1\right) \quad = 1, \tag{6}$$

where $\alpha \in (0, 1)$ is a prescribed number. The standard interpretation of α is in the terms of hypothesis testing, i.e. α represents the level of a test. Then (5) requires the level of the test to be α asymptotically while the condition (6) corresponds to the requirement that the probability of the type II error tends to 0 or, in other words, the power tends to 1, as $m \to \infty$.

We can also consider a situation where the maximal monitoring period is prescribed, i.e. the closed end procedure. Let $N(m)$ be the length of such a period. We define the stopping time as

$$\tau(m, N(m)) = \min\big(N(m)+1, \min\{1 \le k \le N(m) : \Gamma(m,k) \ge c_m\}\big). \tag{7}$$

Here we reject the null hypothesis and stop observing immediately as soon as there exists $k \le N(m)$ such that $\Gamma(m,k) \ge c_m$. Otherwise, i.e. when $\Gamma(m,k) < c_m$ for all $k = 1, \dots, N(m)$, the null hypothesis cannot be rejected. If

$$\lim_{m\to\infty} \frac{N(m)}{m} = \infty \tag{8}$$

then the limit distribution of $\tau(m)$ and $\tau(m, N(m))$ are the same. The analogue of (5) is now

$$\lim_{m\to\infty} \Pr\Big(\tau(m, N(m)) \le N(m) \Big| H_0\Big) \le \alpha,$$

so $N(m)$ can be interpreted as the asymptotic α-quantile of the stopping time $\tau(m, N(m))$. When we use $\alpha = 1/2$ we get a median and due to (8) we require it to be large compared to the length of the historical period. Therefore we have an analogous requirement as in standard sequential analysis. There we are looking for the minimal stopping time that has, under the null hypothesis, the ARL larger than a given bound, i.e. $\mathrm{E}(\tau|H_0) \ge c$, c large enough. However the expectation of the stopping time is not the most convenient criterium due to the asymmetry of the distribution of τ. The idea of using median instead of expectation is more appropriate but not so widespread.

3. ASSUMPTIONS AND MAIN RESULTS

We assume that the random errors $\{e_i,\ 1 \le i < \infty\}$ are

$$\text{i.i.d. with } \mathrm{E}\, e_1 = 0,\ \operatorname{var} e_1 = 1,\ \operatorname{var} e_1^2 = \eta^2 < \infty,\ \mathrm{E}\,|e_1|^{4+\delta} < \infty \ \text{ for some } \delta > 0. \tag{9}$$

We consider the detector $\Gamma(m,k)$ in the form $\Gamma(m,k) = |Q(m,k)|/g(m,k)$, where $Q(m,k)$ is a test statistic based on all observations up to time $m+k$ and $g(m,k)$ is the so called boundary function.

The statistics $Q(m,k)$ are defined as

$$Q(m,k) = \frac{1}{\hat{v}_m}\left\{\sum_{i=m+1}^{m+k}(Y_i - \bar{Y}_m)^2 - \frac{k}{m}\sum_{i=1}^{m}(Y_i - \bar{Y}_m)^2\right\},\ k = 1, 2, \dots$$

where $\bar{Y}_m = \frac{1}{m}\sum_{i=1}^{m} Y_i$ is an average observation in the historical period and

$$\hat{v}_m^2 = \frac{1}{m}\sum_{i=1}^{m}(Y_i - \bar{Y}_m)^4 - \left[\frac{1}{m}\sum_{i=1}^{m}(Y_i - \bar{Y}_m)^2\right]^2$$

is an estimate of $\hat{v}^2 = \mathrm{var}\,(Y_i - \mathrm{E}\,Y_i)^2 = \eta^2\,\sigma_0^4, 1 \le i < \infty$ under the null hypothesis. In other words the test statistic $Q(m,k)$ is a standardized difference of variance estimators from the historical and the monitoring period respectively. Therefore its large values lead to the rejection of the null hypothesis.

The boundary function $g(m,k)$ depends further on a constant γ, so we will write explicitly $g(m,k,\gamma)$. The function is defined as

$$g(m,k,\gamma) = \sqrt{m}\left(\frac{m+k}{m}\right)\left(\frac{k}{m+k}\right)^{\gamma}. \tag{10}$$

This boundary function was for the first time considered in [3]. The constant γ (that is chosen from the interval $[0,1/2)$) plays a role of a tuning parameter that modifies an ability of the test to detect better early or late changes in the following way: $\gamma = 0$ is convenient when a late change is expected, while γ close to $1/2$ is appropriate when an early change is expected.

Another type of detector for a change in scale is proposed and studied in [1]. This detector is based on recursive residuals. However, as the simulations indicate, the performance of this procedure is worse than of the above mentioned one. Therefore we do not report it here.

The question is how to determine the constant $c_m(\alpha)$ to ensure that both (5) and (6) hold. For that we can make use of an asymptotic behaviour of the detector which is formulated in two following theorems, in the first one for the null hypothesis and in the second one for the alternative.

Theorem 1. Assume that $Y_1, Y_2, \ldots$ follow the model (1), assumptions (9) are satisfied and $\gamma \in [0, \frac{1}{2})$. Then, under the null hypothesis (2),

$$\lim_{m\to\infty} \Pr\left(\sup_{1\le k<\infty} \frac{|Q(m,k)|}{g(m,k,\gamma)} \le c\right) = \Pr\left(\sup_{0\le t\le 1} \frac{|W(t)|}{t^{\gamma}} \le c\right)$$

for all $c > 0$, where $\{W(t),\, t \in [0,1]\}$ is a Wiener process.

Theorem 2. Assume that $Y_1, Y_2, \ldots$ follow the model (1), assumptions (9) are satisfied and $\gamma \in [0,1/2)$. Then, under the alternative hypothesis (3),

$$\sup_{1\le k<\infty} \frac{|Q(m,k)|}{g(m,k,\gamma)} \xrightarrow{P} \infty, \quad m \to \infty.$$

Proofs of both theorems are postponed to Section 5. Under the assumption (8), as will be seen from the proofs, both theorems hold also for the stopping time $\tau(m, N(m))$ i.e. when the supremum is taken over $1 \ldots N(m)$, so we can also use the second interpretation described above.

Theorem 1 accounts for the range of constant γ, since for $\gamma \ge 1/2$ the random process $W(t)/t^{\gamma}$ converges to infinity as $t \to 0+$ almost surely. As [5] indicates $\gamma = 1/2$ can be also used in the boundary function but it leads to a different asymptotic distribution of the detector and therefore we will consider only $\gamma \in [0, \frac{1}{2})$.

The limit distribution can be used for approximation of the critical values $c_m(\alpha, \gamma)$, i. e. the values that fulfil

$$\Pr\left(\sup_{1\le k<\infty} \frac{|Q(m,k)|}{g(m,k,\gamma)} \ge c_m(\alpha,\gamma)\right) = \alpha.$$

Due to Theorem 1 the values $c_m(\alpha, \gamma)$ can be approximated by $c(\alpha, \gamma)$, for which

$$\Pr\left(\sup_{0\le t\le 1} \frac{|W(t)|}{t^\gamma} \ge c(\alpha,\gamma)\right) = \alpha.$$

However an explicit formula for the distribution function of this functional of a Wiener process is known only when $\gamma = 0$. Otherwise we have to use simulations so details are provided in the next section.

Theorem 2 ensures that the requirement (6) is fulfilled i. e. that the true change will be detected with probability tending to 1 as $m \to \infty$.

4. SIMULATIONS

In this section we report the results of a small simulation study we performed in order to check the finite sample performance of the monitoring procedure considered in the previous section. The simulations were performed using the R software.
We used these values of parameters:

- Length of the historical period $m = 100, 500, 1000$,
- Distribution of the errors under the null hypothesis: $N(0, 1)$, Laplace with zero mean and unit variance, i. e. $\sigma_0^2 = 1$,
- Level $\alpha = 0.05,\ 0.1$,
- Tuning constant $\gamma = 0,\ 0.25,\ 0.45, 0.49$,
- True time of the change $k^* = 5, 500, m/2, m, 2m$,
- Variance after the change $\sigma_*^2 = 1.5, 2, 4, 9$.

Different lengths of the historical period are important due to the approximation based on asymptotics. We will be also able to compare the performance for normal data and the data with heavier tails and see an influence of the tuning constant according to the true time of change. The results for both levels are analogous so we will report just results for the 10 % level. More detailed results can be found in [1].

4.1. Null hypothesis

First we check the behaviour under the null hypothesis of no change. For each combination of parameters $20\,m$ variables following the model were generated 10 000 times and in r-multiple of m ($r = \frac{1}{4}, 1, 5, 9, 19$) it was checked whether the detector exceeded the critical value. These values were determined from the asymptotic distribution derived in Theorem 1, i. e. as values $c(\alpha, \gamma)$, such that

$$\Pr\left(\sup_{0\le t\le 1} |W(t)|/t^\gamma \ge c(\alpha,\gamma)\right) = \alpha$$

holds. Unfortunately the explicit form for the distribution of $|W(t)|/t^\gamma$ is known only for $\gamma = 0$ and therefore the critical values have to be simulated. They are reported in [4] (Table 1).

Empirical sizes of the monitoring procedure are reported in Table 1 and in Figure 1. First we focus on the normal data. For $\gamma = 0$ or 0.25 the required level is kept except for the shortest historical period and the increase of rejections is gradual in time. Whereas for γ close to 1/2 the level is exceeded even for $m = 1000$. This is caused by a slow convergence of our detector to the limit distribution. The majority of rejections occurs here at the beginning of the monitoring. A prolongation of the training period visibly reduces the probability of false alarm although the monitoring period remains a multiple of m.

Table 1. Empirical sizes for 10% level, i. e. the percentage of processes stopped until $m.r$, when the asymptotic critical values are used.

		$N(0,1)$					Laplace				
γ	m\r	1/4	1	5	9	19	1/4	1	5	9	19
0	100	0.21	3.76	11.93	13.52	14.91	1.91	8.38	17.90	20.10	21.49
	500	0.02	1.95	7.38	8.81	10.18	0.36	3.47	9.55	11.26	12.31
	1000	0.02	1.43	7.48	9.03	10.46	0.10	2.37	8.13	9.62	10.70
0.25	100	3.42	8.64	14.56	15.71	16.43	6.70	13.61	20.39	21.60	22.64
	500	1.07	4.50	9.09	10.02	10.51	3.37	7.60	12.91	14.13	14.60
	1000	0.77	4.11	8.54	9.39	9.78	2.20	5.91	10.70	11.62	12.40
0.45	100	9.78	13.48	16.77	17.14	17.35	12.10	17.87	21.60	22.23	22.82
	500	7.68	10.26	12.23	12.35	12.64	11.41	14.28	16.30	16.73	17.18
	1000	7.73	10.06	11.81	12.08	12.43	10.96	13.55	16.20	16.61	16.86
0.49	100	10.29	13.19	15.08	15.50	15.71	12.46	16.33	19.22	19.72	20.03
	500	9.74	11.19	12.13	12.26	12.42	12.43	14.70	16.21	16.48	16.57
	1000	9.21	10.74	11.87	11.98	12.09	12.04	13.55	14.50	14.70	14.74

For the Laplace distribution the situation is not satisfactory. It is clear that the asymptotic critical values are too low. Therefore the critical values $c_m(\alpha, \gamma)$ based on the length of the historical period m are needed. These were obtained from the same simulations as quantiles of maximum of our detector. They are reported in Table 2.

Table 2. Simulated critical values $c_m(0.1, \gamma)$ are reported in columns with m, the column with ∞ contains asymptotic critical values $c(0.1, \gamma)$.

	$N(0,1)$				Laplace			
γ\m	100	500	1000	∞	100	500	1000	∞
0	2.17	1.95	1.94	1.95	2.67	2.06	1.97	1.95
0.25	2.41	2.15	2.11	2.11	3.06	2.33	2.19	2.11
0.45	2.98	2.69	2.68	2.54	3.83	3.05	2.89	2.54
0.49	3.24	2.98	2.94	2.83	4.06	3.35	3.19	2.83

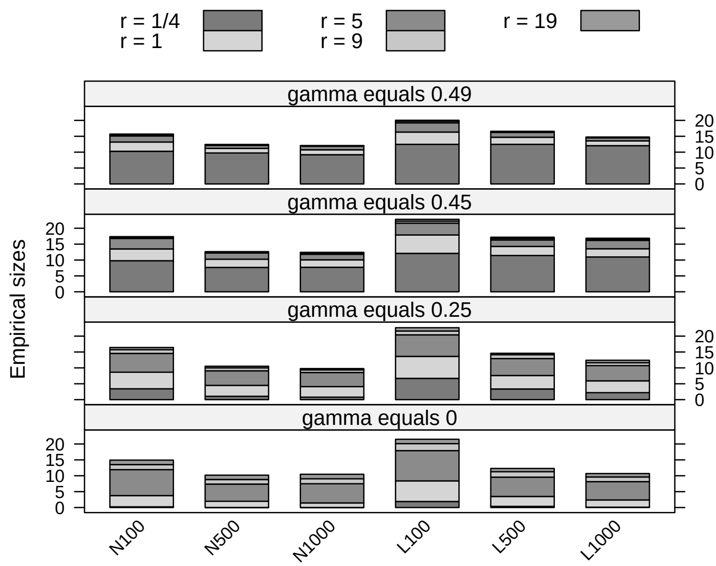

Fig. 1. Empirical sizes for 10 % level (N100 denotes normal errors with $m = 100$ etc.), when the asymptotic critical values are used.

4.2. Alternative hypothesis

Now we focus on the performance under various alternatives represented by the time of change k^* and the amount of change σ_*^2. For each combination of parameters 5 000 variables following the model were generated. It means that e. g. for $m = 100$ the length of monitoring period equals 4 900. The first time when the detector exceeded the critical value was saved. If this did not happen, the length of monitoring period was used. Results are based on 2 500 replications.

First we will discuss the simulations with normally distributed errors. As indicated above, the asymptotic critical values can almost ensure the required level. Therefore we present results of the monitoring when these values were used.

Characteristics of the stopping time τ are given in Table 3. The detection delay then equals $\tau - k^*$. Two points of change are reported using the proper value of γ as will be discussed later. Naturally we can see improvement in detection delay with a growing amount of change. However from a certain amount the benefit is not so large. Also the prolongation of the training period has a positive effect on detection delay. Particularly an increase of m from 100 to 500 results in significant drop in detection delay (especially for small changes). Additional prolongation is not so influential, only the extremes are further restricted, that results in decrease in the mean.

Sometimes the change was not detected (the maximum of τ is equal to the length of the monitoring period $5\,000 - m$). This concerns the shortest training period and small amounts of change. Here we can also see that the distribution of the stopping time is asymmetric. On the other hand procedures with longer training period are

Table 3. Influence of m and σ_*^2 on the stopping time τ.

		$k^* = 5,\ \gamma = 0.45$			$k^* = 500,\ \gamma = 0.25$		
	$\backslash m$	100	500	1000	100	500	1000
$\sigma_*^2 = 1.5$	1st Qu.	25	36	36	676	665	668
	Median	72	73	76	1034	783	761
	Mean	859	106	98	1716	819	779
	3rd Qu.	330	134	135	2231	944	875
	Max.	4900	1569	643	4900	3825	1880
$\sigma_*^2 = 2$	1st Qu.	13	15	16	576	574	578
	Median	27	27	29	691	620	619
	Mean	47	34	35	712	615	620
	3rd Qu.	50	45	48	849	676	660
	Max.	4900	209	206	4900	986	933
$\sigma_*^2 = 4$	1st Qu.	8	8	8	526	525	525
	Median	10	11	12	554	539	537
	Mean	12	12	13	508	531	534
	3rd Qu.	15	15	16	587	556	550
	Max.	69	45	55	799	628	625
$\sigma_*^2 = 9$	1st Qu.	6	6	6	509	509	509
	Median	7	7	8	519	514	514
	Mean	8	8	8	466	505	511
	3rd Qu.	9	9	9	530	521	520
	Max.	25	24	27	589	559	552

able to detect even small changes with reasonable delay.

Kernel density estimates of the stopping time are shown in Figures 2 and 3. The former one contains densities for different amount of change. With a growing amount of change one can see a shift of a mass of the distribution closer to the true time of change. Also the asymmetry is visible. The latter figure contains densities for different training periods. Small change was chosen since the difference there is the largest. One can see a quite similar shape of densities for $m = 500$ and $m = 1000$, which is different from that for $m = 100$, as was already indicated in Table 3. The above described results hold for an early as well as a late change.

Influence of the true time of change on the detection delay is shown in Table 4. Unlike the previous tables, where the values were detection times, here they represent already the detection delays. For an early change the medians of delay are quite the same for all training periods, however substantial improvement with a prolongation of the historical period can be observed in mean value, especially for a small change. For $k^* = 500$ is this improvement visible also in medians. However the total delay is much larger than that for an early change. This is a typical behaviour of procedures based on cumulative sums. This also relates to the last part of the table, where k^* is a constant multiple of m and therefore takes value from 200 up to 2000. Longer training period normally brings shorter delay however this is here beaten by a substantially later time of change.

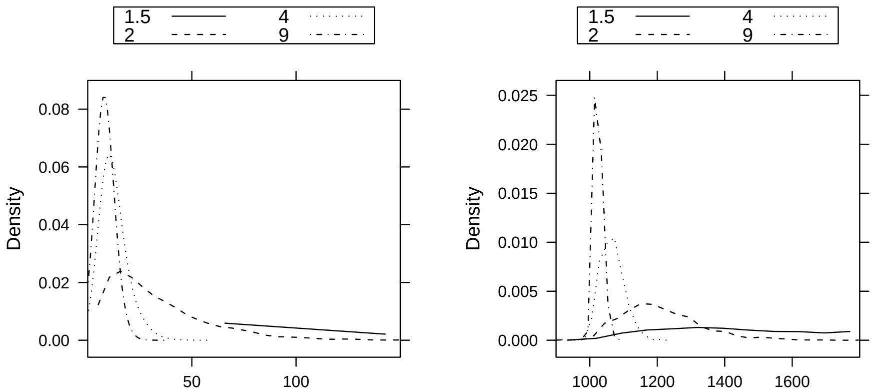

Fig. 2. Kernel density estimates of τ for different amounts of change. Left picture is for $k^* = 5$, right one for $k^* = 1000$, $m = 1000$ for both.

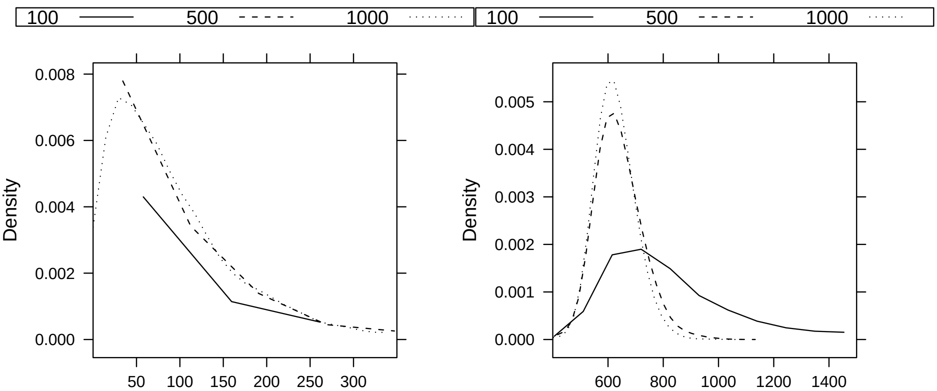

Fig. 3. Kernel density estimates of τ for different training periods. Left picture is for $k^* = 5$ and $\sigma_*^2 = 1.5$, right one for $k^* = 500$ and $\sigma_*^2 = 2$.

An influence of the tuning constant γ is illustrated in Table 5. For an early change ($k^* = 5$) the best results are obtained for $\gamma = 0.45$ or $\gamma = 0.49$ as was already mentioned. For a moderate change $k^* \sim m$ the best results are obtained by $\gamma = 0.25$ closely followed by $\gamma = 0$, whereas for a later change the situation is reversed. For a very late change ($k^* = 500$ a $m = 100$) $\gamma = 0$ clearly prevails. The above described pattern is more distinct for a small change.

Kernel density estimates of the stopping time for different γ's are shown in Figure 4. Smaller change was chosen to better indicate the pattern mentioned earlier.

Now we turn to the performance of the procedure for data with heavier tails, represented here by Laplace distribution.

We will not present the results obtained by asymptotic critical values since the

Table 4. Median and mean of detection delay.

		median			mean		
	$m \setminus \sigma_*^2$	1,5	2	4	1,5	2	4
	100	50	18	5	636	33	6
$k^* = 5$	500	55	19	6	83	25	7
$\gamma = 0.45$	1000	59	21	6	78	26	7
	100	534	191	54	1216	212	38
$k^* = 500$	500	283	120	39	319	115	31
$\gamma = 0.25$	1000	261	119	37	279	120	34
	100	260	96	28	803	123	24
$k^* = 2m$	500	441	196	62	501	191	46
$\gamma = 0$	1000	590	274	86	601	259	59

Table 5. Influence of γ on the stopping time τ according to k^*.

		$m = 500$, $\sigma_*^2 = 4$				$m = 100$, $\sigma_*^2 = 1.5$				
k^*	$\setminus\gamma$	0	0.25	0.45	0.49	0	0.25	0.45	0.49	k^*
	1st Qu.	22	11	8	7	53	32	20	20	
	Median	28	15	11	10	94	67	55	64	
5	Mean	29	17	12	11	454	470	641	838	5
	3rd Qu.	35	21	15	14	210	176	197	300	
	Max.	83	54	45	45	4900	4900	4900	4900	
	1st Qu.	272	267	265	266	177	165	162	177	
	Median	282	277	275	277	274	268	317	385	
250	Mean	284	276	258	256	804	867	1206	1481	100
	3rd Qu.	294	288	288	289	548	582	1009	2228	
	Max.	353	345	345	348	4900	4900	4900	4900	
	1st Qu.	528	524	523	525	304	298	306	332	
	Median	542	539	540	543	460	468	564	686	
500	Mean	542	529	495	491	1003	1080	1443	1712	200
	3rd Qu.	558	554	556	560	852	928	1693	3838	
	Max.	626	623	626	640	4900	4900	4900	4900	
	1st Qu.	1039	1037	1039	1045	674	676	724	807	
	Median	1062	1059	1065	1072	990	1034	1254	1536	
1000	Mean	1046	1022	966	970	1609	1716	2058	2350	500
	3rd Qu.	1086	1084	1092	1099	1909	2231	4058	4900	
	Max.	1196	1196	1199	1236	4900	4900	4900	4900	

procedure does not keep the required level even for the long training period (cf. Table 1). Therefore we use the simulated critical values for given length of training period $c_m(0.1, \gamma)$. Results follow the same pattern as those for normally distributed errors so we will not go into details. The comparison of both errors is given in Table 6. The results for normal errors are also obtained using the simulated critical values $c_m(0.1, \gamma)$. We report only the results for an early change, for a late change are analogous.

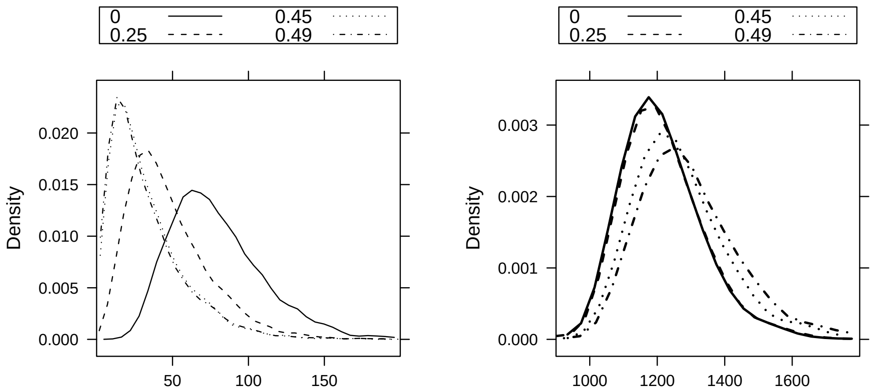

Fig. 4. Kernel density estimates of τ for different γ when $m = 500$ and $\sigma_*^2 = 2$. Left picture is for $k^* = 5$, right one for $k^* = 1000$.

Table 6. Influence of the random errors. $\gamma = 0.45$ is used.

		Laplace				normal			
	$\backslash\sigma_*^2$	1.5	2	4	9	1.5	2	4	9
$m = 100$	1st Qu.	106	29	11	8	33	15	8	6
	Median	4900	87	19	10	95	30	11	8
$k^* = 5$	Mean	2832	1081	36	12	1043	61	13	8
	3rd Qu.	4900	529	34	15	505	56	16	9
	Max.	4900	4900	4900	67	4900	4900	93	26
$m = 500$	1st Qu.	76	27	11	7	31	15	8	6
	Median	216	61	18	10	67	26	11	8
$k^* = 5$	Mean	703	97	22	11	98	33	12	8
	3rd Qu.	576	115	29	14	124	44	15	10
	Max.	4500	4500	108	39	3744	239	47	27
$m = 1000$	1st Qu.	65	28	11	7	35	16	8	7
	Median	171	58	17	10	72	27	11	8
$k^* = 5$	Mean	316	78	21	11	92	33	13	8
	3rd Qu.	361	105	28	14	125	45	15	9
	Max.	4000	658	102	40	630	215	53	25

The monitoring procedure gives worse results for the Laplace distribution. In case of a short training period and small amounts of change is the worsening significant. For some combinations even more than a quarter of monitoring does not detect the change. In case of longer training period or bigger change the difference is not so large, however it is still clearly visible.

Above described worsening is in line with our expectation since our monitoring procedure uses sample variance as an estimator of the observations' variance. This is appropriate for data not deviating too much from normality, however for data with heavier tails it does not give a reliable estimate. Some robust estimator should be used instead.

5. PROOFS

The proof of Theorem 1 is divided into several lemmas. If not otherwise stated the convergence is for $m \to \infty$.

Lemma 1. Let the assumptions of Theorem 1 be satisfied. Then

$$\hat{v}_m^2 \xrightarrow{P} v^2 = \eta^2 \, \sigma_0^4.$$

Proof. The lemma is an easy consequence of law of large numbers (LLN). □

Lemma 2. Let the assumptions of Theorem 1 be satisfied. Then

$$\sup_{1\le k<\infty} \frac{\left|\sum_{i=m+1}^{m+k}(Y_i-\bar{Y}_m)^2 - \frac{k}{m}\sum_{i=1}^{m}(Y_i-\bar{Y}_m)^2 - \left[\sigma_0^2\sum_{i=m+1}^{m+k} e_i^2 - \sigma_0^2\frac{k}{m}\sum_{i=1}^{m} e_i^2\right]\right|}{\sqrt{m}(1+\frac{k}{m})(\frac{k}{m+k})^\gamma} \xrightarrow{P} 0.$$

Proof. Due to stability in the historical period and the null hypothesis the numerator can be rewritten as

$$\begin{aligned}
&\left\{\sum_{i=m+1}^{m+k}(Y_i-\bar{Y}_m)^2 - \frac{k}{m}\sum_{i=1}^{m}(Y_i-\bar{Y}_m)^2 - \left[\sigma_0^2\sum_{i=m+1}^{m+k} e_i^2 - \sigma_0^2\frac{k}{m}\sum_{i=1}^{m} e_i^2\right]\right\}\\
&=\left\{\sigma_0^2\sum_{i=m+1}^{m+k}(e_i-\bar{e}_m)^2 - \frac{k}{m}\sigma_0^2\sum_{i=1}^{m}(e_i-\bar{e}_m)^2 - \left[\sigma_0^2\sum_{i=m+1}^{m+k} e_i^2 - \sigma_0^2\frac{k}{m}\sum_{i=1}^{m} e_i^2\right]\right\}\\
&=2\sigma_0^2\bar{e}_m\left\{\frac{k}{m}\sum_{i=1}^{m} e_i - \sum_{i=m+1}^{m+k} e_i\right\}.
\end{aligned}$$

We can split the supremum as follows

$$\begin{aligned}
&\sup_{1\le k<\infty}\frac{|\bar{e}_m\frac{k}{m}\sum_{i=1}^{m} e_i - \bar{e}_m\sum_{i=m+1}^{m+k} e_i|}{g(m,k,\gamma)}\\
&\le \sup_{1\le k<\infty}\frac{|\bar{e}_m\frac{k}{m}\sum_{i=1}^{m} e_i|}{g(m,k,\gamma)} + \sup_{1\le k<\infty}\frac{|\bar{e}_m\sum_{i=m+1}^{m+k} e_i|}{g(m,k,\gamma)}, \qquad (11)
\end{aligned}$$

and investigate both terms separately.

For $k \le m$ the boundary function fulfils

$$g(m,k,\gamma) = \sqrt{m}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^\gamma = m^{-1/2}(m+k)^{1-\gamma}k^\gamma > m^{1/2-\gamma}k^\gamma$$

Since $\sum_{i=1}^{m} e_i = O_P(m^{1/2})$ the first term on r.h.s. of (11) is then

$$\max_{1\le k\le m}\frac{|\bar{e}_m\frac{k}{m}\sum_{i=1}^{m} e_i|}{g(m,k,\gamma)} = O_P(1)\max_{1\le k\le m}\frac{\frac{k}{m}}{m^{1/2-\gamma}k^\gamma} = O_P(1)\max_{1\le k\le m}\frac{k^{1-\gamma}}{m^{1-\gamma}}m^{-1/2} \xrightarrow{P} 0.$$

Using the Hájek–Rényi inequality (cf. [2]) for the second term on r.h.s. of (11) we have

$$\Pr\left(\max_{1\le k\le m}\frac{|\sum_{i=m+1}^{m+k}e_i|}{g(m,k,\gamma)}\ge A\right)\le\Pr\left(\max_{1\le k\le m}\frac{|\sum_{i=m+1}^{m+k}e_i|}{k^\gamma}\ge Am^{1/2-\gamma}\right)$$
$$\le\left(Am^{1/2-\gamma}\right)^{-2}\sum_{i=m+1}^{m+m}\frac{\mathrm{E}e_i^2}{i^{2\gamma}}=A^{-2}m^{2\gamma-1}O(m^{1-2\gamma})\sim A^{-2},$$

where $\sim$ denotes the asymptotic behaviour i. e. $a_n\sim b_n\Leftrightarrow a_n=O(b_n)\;\&\;b_n=O(a_n)$. The approximation $\sum_{i=m+1}^{2m}i^{-2\gamma}\sim\int_{m+1}^{2m}z^{-2\gamma}dz=O(m^{1-2\gamma})$ was used. Since A can be chosen arbitrary large this part is bounded in probability. Therefore since $\bar{e}_m=O_P(m^{-1/2})$ we have

$$\max_{1\le k\le m}\frac{|\bar{e}_m\sum_{i=m+1}^{m+k}e_i|}{g(m,k,\gamma)}\xrightarrow{P}0.$$

For $k\ge m$ the boundary function fulfils

$$g(m,k,\gamma)=\sqrt{m}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^\gamma=m^{-1/2}(m+k)^{1-\gamma}k^\gamma>m^{-1/2}k.$$

Hence for the first term on r.h.s. of (11) we have

$$\sup_{k\ge m}\frac{|\bar{e}_m\frac{k}{m}\sum_{i=1}^{m}e_i|}{g(m,k,\gamma)}=O_P(1)\sup_{k\ge m}\frac{\frac{k}{m}}{m^{-1/2}k}=O_P(1)\sup_{k\ge m}\frac{1}{m^{1/2}}\xrightarrow{P}0.$$

For the second term we use again the Hájek–Rényi inequality

$$\Pr\left(\max_{m\le k\le n}\frac{|\sum_{i=m+1}^{m+k}e_i|}{g(m,k,\gamma)}\ge A\right)\le\Pr\left(\max_{m\le k\le n}\frac{|\sum_{i=m+1}^{m+k}e_i|}{m^{-1/2}k}\ge A\right)$$
$$\le\left(Am^{-1/2}\right)^{-2}\sum_{i=m+1}^{m+n}\frac{1}{i^2}=A^{-2}m\,O(m^{-1})\sim A^{-2},$$

uniformly in n. Therefore again we have

$$\sup_{k\ge m}\frac{|\bar{e}_m\sum_{i=m+1}^{m+k}e_i|}{g(m,k,\gamma)}\xrightarrow{P}0,$$

which was the last part needed to complete the proof. □

Lemma 3. Let the assumptions of Theorem 1 be satisfied. Then there for each m exist two independent Wiener processes $\{W_{1,m}(t),\,0\le t<\infty\}$ a $\{W_{2,m}(t),\,0\le t<\infty\}$ such that

$$\sup_{1\le k<\infty}\frac{\left|\sum_{i=m+1}^{m+k}\varepsilon_i-\frac{k}{m}\sum_{i=1}^{m}\varepsilon_i-\eta\left(W_{1,m}(k)-\frac{k}{m}W_{2,m}(m)\right)\right|}{\sqrt{m}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^\gamma}\xrightarrow{P}0,$$

where $\varepsilon_i = e_i^2 - \mathrm{E}\, e_i^2 = e_i^2 - 1$ and $\eta^2 = \operatorname{var} \varepsilon_i = \operatorname{var} e_i^2$.

Proof. It follows from Lemma 5.3 in [4]. □

Lemma 4. Let $\{W_{1,m}(t),\, 0 \le t < \infty\}$ and $\{W_{2,m}(t),\, 0 \le t < \infty\}$ be two independent Wiener processes and $\{W(t),\, 0 \le t < \infty\}$ denote a Wiener process as well. Then we have

$$\sup_{1\le k<\infty} \frac{\left|W_{1,m}(k) - \frac{k}{m} W_{2,m}(m)\right|}{\sqrt{m}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^{\gamma}} \xrightarrow{D} \sup_{0\le t\le 1} \frac{|W(t)|}{t^{\gamma}}.$$

Proof. The proof can be found in [4] (beginning of the proof of Theorem 2.1).□

Proof of Theorem 1. Using the previous lemmas we can write

$$\begin{aligned}
\sup_{1\le k<\infty} \frac{|Q(m,k)|}{g(m,k,\gamma)} &= \frac{1}{\hat{v}_m} \sup_{1\le k<\infty} \frac{\left|\sum_{i=m+1}^{m+k}(Y_i-\bar{Y}_m)^2 - \frac{k}{m}\sum_{i=1}^{m}(Y_i-\bar{Y}_m)^2\right|}{g(m,k,\gamma)} \\
&\stackrel{1}{=} \frac{1}{\hat{v}_m} \sup_{1\le k<\infty} \frac{\left|\sigma_0^2 \sum_{i=m+1}^{m+k} e_i^2 - \sigma_0^2 \frac{k}{m}\sum_{i=1}^{m} e_i^2\right|}{g(m,k,\gamma)} + o_P(1) \\
&= \frac{\sigma_0^2}{\hat{v}_m} \sup_{1\le k<\infty} \frac{\left|\sum_{i=m+1}^{m+k} \varepsilon_i - \frac{k}{m}\sum_{i=1}^{m} \varepsilon_i\right|}{g(m,k,\gamma)} + o_P(1) \\
&\stackrel{2}{=} \frac{\sigma_0^2\eta}{\hat{v}_m} \sup_{1\le k<\infty} \frac{\left|W_{1,m}(k) - \frac{k}{m} W_{2,m}(m)\right|}{g(m,k,\gamma)} + o_P(1) \\
&\xrightarrow{D} \sup_{0\le t\le 1} \frac{|W(t)|}{t^{\gamma}}.
\end{aligned}$$

Equality 1 follows from Lemma 2, equality 2 from Lemma 3 and the last convergence is a consequence of Lemmas 4 and 1. □

Proof of Theorem 2. It is sufficient to find a certain $\tilde{k}$ such that the detector diverges in probability i. e.

$$\frac{|Q(m,\tilde{k})|}{g(m,\tilde{k},\gamma)} \xrightarrow{P} \infty.$$

We will show that the previous holds for $\tilde{k} = k^* - 1 + \max(m, k^*)$.
We denote the parts of the test statistics before and after the change of variance as

$$\begin{aligned}
Q(m,\tilde{k}) &= A_{1,m} + A_{2,m} \\
&:= \sum_{i=m+1}^{m+k^*-1} (Y_i - \bar{Y}_m)^2 - \frac{k^*-1}{m}\sum_{i=1}^{m}(Y_i-\bar{Y}_m)^2 \\
&+ \sum_{i=m+k^*}^{m+\tilde{k}} (Y_i - \bar{Y}_m)^2 - \frac{\tilde{k}-k^*+1}{m}\sum_{i=1}^{m}(Y_i-\bar{Y}_m)^2.
\end{aligned}$$

By Theorem 1 we know that $A_{1,m}$ (the part before the change) divided by $g(m, k^*-1, \gamma)$ is bounded in probability. Since $g(m,k,\gamma)$ is monotonically increasing in k, this also holds for $g(m,\tilde{k},\gamma)$ instead of $g(m,k^*-1,\gamma)$.
We concentrate therefore on $A_{2,m}$

$$\sum_{i=m+k^*}^{m+\tilde{k}} (Y_i - \bar{Y}_m)^2 - \frac{\max(m,k^*)}{m}\sum_{i=1}^{m}(Y_i - \bar{Y}_m)^2$$

$$= \sum_{i=m+k^*}^{m+\tilde{k}} \sigma_*^2 e_i^2 - \sigma_0^2 \frac{\max(m,k^*)}{m}\sum_{i=1}^{m} e_i^2 - 2\sigma_0 \bar{e}_m \left[\sigma_* \sum_{i=m+k^*}^{m+\tilde{k}} e_i - \sigma_0 \frac{\max(m,k^*)}{m}\sum_{i=1}^{m} e_i\right].$$

The third term as a linear combination of e_i divided by $g(m,\tilde{k},\gamma)$ is again bounded in probability. The first two terms yield

$$\sigma_*^2 \sum_{i=m+k^*}^{m+\tilde{k}} e_i^2 - \sigma_0^2 \frac{\max(m,k^*)}{m}\sum_{i=1}^{m} e_i^2$$

$$= \sigma_*^2 \sum_{i=m+k^*}^{m+\tilde{k}} (e_i^2 - 1) - \sigma_0^2 \frac{\max(m,k^*)}{m}\sum_{i=1}^{m}(e_i^2 - 1) + \sigma_*^2 \max(m,k^*) - \sigma_0^2 \max(m,k^*). \tag{12}$$

Since $\sum_{i=m+k^*}^{m+\tilde{k}} (e_i^2 - 1) = O_P\left((m+\tilde{k})^{1/2}\right)$, the first two terms in (12) divided by $g(m,\tilde{k},\gamma)$ are again bounded in probability. Since

$$\frac{\max(m,k^*)}{\sqrt{m}\left(1+\frac{\tilde{k}}{m}\right)\left(\frac{\tilde{k}}{m+\tilde{k}}\right)^{\gamma}} \geq \frac{m}{\sqrt{m}}\frac{\max(m,k^*)}{m+k^*+\max(m,k^*)} \geq \frac{m}{3\sqrt{m}} \to \infty$$

and $\sigma_0^2 \neq \sigma_*^2$, the theorem is proved. □

ACKNOWLEDGEMENT

This work was partially supported by the Czech Science Foundation under Grants 201/05/H007 and 201/06/0186. The author acknowledges Professor Marie Hušková for useful advice and recommendations.

(Received April 23, 2008.)

REFERENCES

[1] O. Chochola: Rekurzivní postupy pro detekci změny rozdělení (Recursive Procedures for Detection of Changes). Master's Thesis. Faculty of Mathematics and Physics, Charles University, Prague 2007.

[2] Y. S. Chow and H. Teicher: Probability Theory: Independence, Interchangeability, Martingales. Third edition. Springer, New York 2003.

[3] C.-S. J. Chu, M. Stinchcombe, and H. White: Monitoring structural change. Econometrica *64* (1996), 1045–1065.
[4] L. Horváth, M. Hušková, M. Kokoszka, and J. Steinebach: Monitoring changes in linear models. J. Stat. Plann. Inference *126* (2004), 225–251.
[5] L. Horváth, P. Kokoszka, and J. Steinebach: On sequential detection of parameter changes in linear regression. Statist. Probab. Lett. *77* (2007), 885–895.
[6] A. Koubková: Sequential Change-Point Analysis. Ph.D. Thesis. Faculty of Mathematics and Physics, Charles University, Prague 2006.
[7] F. Leisch, K. Hornik, and E. M. Kuan: Monitoring structural changes with the generalized fluctuation test. Econometric Theory *16* (2000), 835–854.
[8] A. Zeileis, C. Kleiber, W. Krämer, and K. Hornik: Testing and dating of structural changes in practice. Comput. Statist. Data Anal. *44* (2003), 1–2, 109–123.
[9] A. Zeileis, F. Leisch, K. Hornik, and C. Kleiber: Strucchange: An r package for testing for structural change in linear regression models. J. Statist. Software *7* (2002), 2.
[10] A. Zeileis, F. Leisch, C. Kleiber, and K. Hornik: Monitoring structural change in dynamic econometric models. J. Appl. Econometrics *20* (1005), 99–121.

Ondřej Chochola, Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics — Charles University, Sokolovská 83, 186 75 Praha 8. Czech Republic.
e-mail: chochola@karlin.mff.cuni.cz