# Ratio Type Statistics for Detection of Changes in Mean and the Bootstrap Method

**Barbora Madurkayová**
madurka@karlin.mff.cuni.cz
**Department of Probability and Statistics, Charles University in Prague, Czech Republic**

**Summary** The poster presents procedures for detection of changes in mean. In particular test procedures based on ratio type test statistics that are functionals of partial sums of $M$-residuals are studied. We assume to have data obtained in ordered time points and study the null hypothesis of no change against the alternative of a change occurring at some unknown time point. We explore the possibility of applying the bootstrap method for obtaining critical values of the proposed test statistics.

## Model description

Let us consider observations $Y_1, \dots, Y_n$ that were obtained at $n$ time-ordered points. We study the location model with at most one abrupt change in the mean:

$$\boxed{Y_k = \mu + \delta \mathbf{I}\{k > k^*\} + e_k, \qquad k = 1, \dots, n,} \tag{1}$$

where $\mu$, $\delta = \delta_n$ and $k^* = k_n^*$ are unknown parameters. $k^*$ is called the *change-point*. By $e_1, \dots, e_n$, we denote the random error terms.

We are going to test the null hypothesis that no change occurred

$$H_0 : k^* = n \tag{2}$$

against the alternative that change occurred at some unknown time-point $k^*$

$$H_1 : k^* < n, \; \delta \neq 0. \tag{3}$$

## Test statistic

Following the ideas described in [1] and [2], a test statistic based on $M$-residuals is considered:

$$\boxed{W_n = \max_{n\gamma \le k \le n - n\gamma} \frac{\max\limits_{1 \le i \le k} \left| \sum\limits_{1 \le j \le i} \psi(Y_j - \hat{\mu}_{1k}(\psi)) \right|}{\max\limits_{k \le i \le n} \left| \sum\limits_{i+1 \le j \le n} \psi(Y_j - \hat{\mu}_{2k}(\psi)) \right|},} \tag{4}$$

where $0 < \gamma < 1/2$ is a given constant, $\psi$ is a score function, $\hat{\mu}_{1k}(\psi)$ is an $M$-estimate of parameter $\mu$ based on observations $Y_1, \dots, Y_k$ and $\hat{\mu}_{2k}(\psi)$ is an $M$-estimate of $\mu$ based on observations $Y_{k+1}, \dots, Y_n$.

## Limit distribution under $H_0$

### Assumption 1

The random error terms $e_i$, $i = 1, 2, \dots$ form a strictly stationary $\alpha$-*mixing* sequence with distribution function $F$, that is symmetric around zero and for $\delta > 0$, $\Delta > 0$ there exists a constant $C(\delta, \Delta) > 0$ such that

$$\sum_{h=0}^{\infty} (h+1)^{\delta/2} \alpha(h)^{\Delta/(2+\delta+\Delta)} \le C(\delta, \Delta). \tag{5}$$

### Assumption 2

The score function $\psi$ is non-decreasing and antisymmetric function.

### Assumption 3

$$\int |\psi(x)|^{2+\delta+\Delta} dF(x) < \infty \tag{6}$$

and

$$\int |\psi(x + t_2) - \psi(x + t_1)|^{2+\delta+\Delta} dF(x) \le C_1(\delta, \Delta) |t_2 - t_1|^a,$$

$$|t_j| \le C_2(\delta, \Delta), \; j = 1, 2 \tag{7}$$

for some constants $1 \le a \le 2 + \delta + \Delta$, $\delta > 0$, $\Delta > 0$ as in (5) and constants $C1$, $C2$ depending on $\delta$ and $\Delta$.

### Assumption 4

Let us denote $\lambda(t) = -\int \psi(e - t) dF(e)$, for $t \in \mathbb{R}$. We assume that $\lambda(0) = 0$ and that there exists a first derivative $\lambda'$ for all $t \in \mathbb{R}$ such that it is Lipschitz in the neighborhood of $0$ and such that $\lambda'(0) > 0$.

**Remark 1** The above conditions regarding $\psi$ are satisfied for example for $\psi(x) = x$ ($L_2$-method), $\psi(x) = \mathrm{sgn}(x)$ ($L_1$-method) or for the Huber function.

**Theorem 1** *Let us assume that the above described conditions 1-4 hold. Then, under the null hypothesis (2)*

$$\boxed{W_n \xrightarrow{\mathcal{D}} \sup_{\gamma \le t \le 1-\gamma} \left( \frac{t}{1-t} \right)^{1/2} \frac{\sup\limits_{0 \le u \le 1} |B_1(u)|}{\sup\limits_{0 \le u \le 1} |B_2(u)|} \quad \text{as } n \to \infty,} \tag{8}$$

*where $\{B_1(u), 0 \le u \le 1\}$ and $\{B_2(u), 0 \le u \le 1\}$ are independent Brownian bridges.*

**Remark 2** The null hypothesis is rejected for large values of $W_n$. Explicit form of the limit distribution (9) under the null hypothesis is not known. Therefore, to obtain critical values, we have to use either simulation from the limit distribution or resampling methods.

## Permutation version of the test statistic

Let $Y_{1,b}^*, \dots, Y_{n,b}^*$, $b \in B$ be samples obtained by sampling without replacement from the original set of observations $Y_1, \dots, Y_n$. Let $W_{n,b}^*$ be the statistic created by replacing $Y_1, \dots, Y_n$ by $Y_{1,b}^*, \dots, Y_{n,b}^*$ in (4). The following theorem states that in the case of i.i.d. random errors, the conditional distribution of the resampling statistic may be used to obtain critical values for the studied test statistics.

**Theorem 2** *Let us assume that the random error terms in (1) are i.i.d. with* $\mathsf{E}\, e_1 = 0$ *and* $\mathrm{var}\, e_i = \sigma^2 < \infty$. *Then*

$$\boxed{\mathsf{P}(W_{n,b}^* \le x | Y_1, \dots, Y_n) \xrightarrow{a.s.} \mathsf{P}\left( \sup_{\gamma \le t \le 1-\gamma} \left( \frac{t}{1-t} \right)^{1/2} \frac{\sup\limits_{0 \le u \le 1} |B_1(u)|}{\sup\limits_{0 \le u \le 1} |B_2(u)|} \le x \right),} \tag{9}$$

*as $b \to \infty$, where $\{B_1(u), 0 \le u \le 1\}$ and $\{B_2(u), 0 \le u \le 1\}$ are independent Brownian bridges.*

## Simulation

Even though in the above theorem, i.i.d. random errors are considered, in the simulations presented below, we also simulated data with random errors that are AR(1) sequences (with standard normal and $t_5$-distributed innovations). Since there were dependent observations, a *block permutation* method was used. For more details on this method we refer to [3]. In all of the below presented simulations, we chose $\gamma$=0.2 and block length $K$=5.

### Critical values

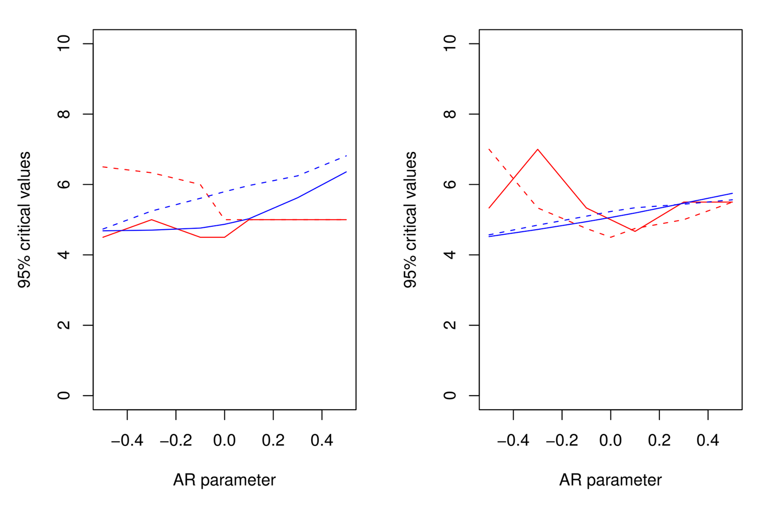


Figure 1: Comparison of $L_1$ (red color) and $L_2$ method (blue color) 95% critical values under $H_0$. Solid lines correspond to $N(0,1)$ innovations, dashed lines correspond to $t_5$ innovations. Results for $n$=100 (on the left) and $n$=200 (on the right). $\gamma$=0.2 block length $K$=5.

### Simulation of rejection probability

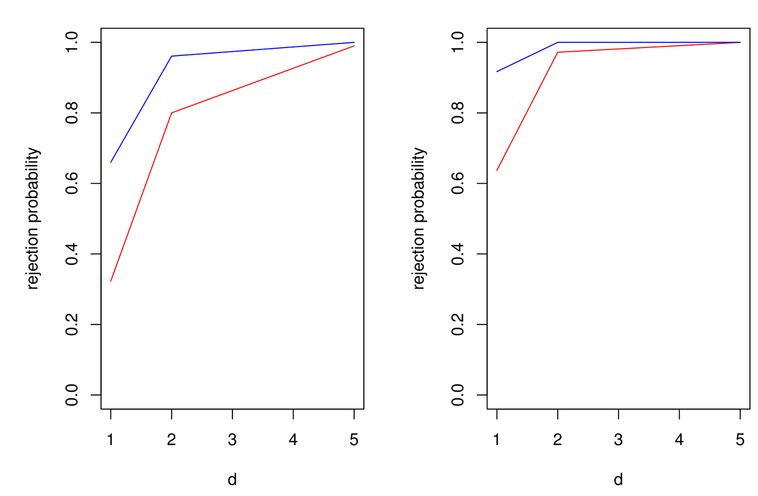


Figure 2: Comparison of $L_1$ (red color) and $L_2$ method (blue color) test performance under $H_1$. Results for $n$=100 (on the left) and $n$=200 (on the right). We simulated standard normal random errors and used the 95% critical values.

## Summary

Comparing to the critical values obtained by simulation of the limit distribution (see [4]), resampling methods seem to be more suitable to obtain the critical values. Simulation showed that the block resampling method gives reasonable results for AR(1) sequences with values of the autoregression coefficient between -0.5 and 0.5. In future work, we may focus on studying the optimal choice of block length, as well as deriving theoretical results for the permutation version of the test statistics under more general set of assumptions.

**Acknowledgement** Author would like to express her thanks to Prof. RNDr. Marie Hušková, DrSc. for her valuable comments and suggestions. The work was supported by grant GAUK B-MAT 116209.

### References

[1] Horváth, L., Horváth, Z., and Hušková, M. (2008). Ratio tests for change point detection. Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor P. K. Sen, 1:293–304.

[2] Hušková, M. and Marušiaková, M. (2009). M-procedures for detection of changes for dependent observations. Proceedings of the 6th St. Petersburg Workshop on Simulation, 2:673–678.

[3] Kirch, C. (2006). Resampling Methods for the Change Analysis of Dependent Data. PhD thesis, University of Cologne, Cologne. http://kups.ub.uni- koeln.de/volltexte/2006/1795/.

[4] Madurkayová, B. (2009). Ratio Type Statistics for Detection of Changes in Location. Proceedings of the 16th European Young Statisticians Meeting, 144–148