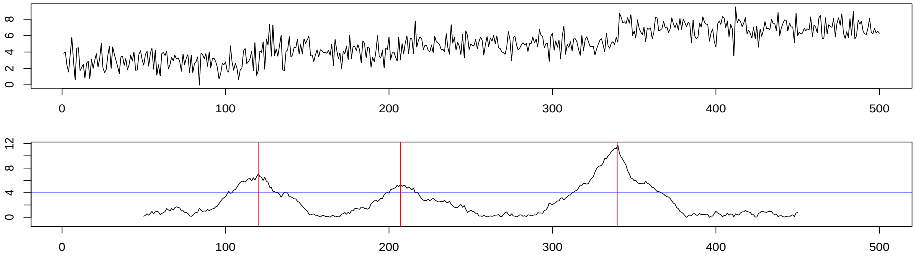
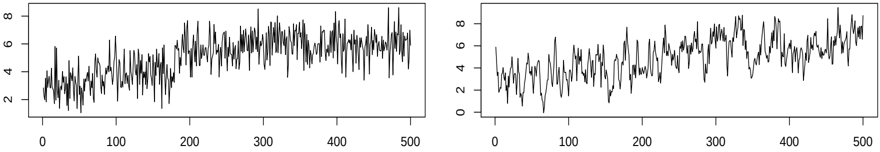
Change-Point Methods for Multiple Structural Breaks and Regime Switching Models

Birte Muhsal, joint work with Claudia Kirch | March 26th, 2012

WORKSHOP ON RECENT ADVANCES IN CHANGEPOINT ANALYSIS AT THE UNIVERSITY OF WARWICK



Multiple Change-Point Problem



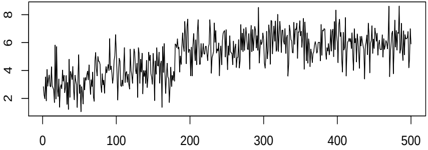
We want to construct:

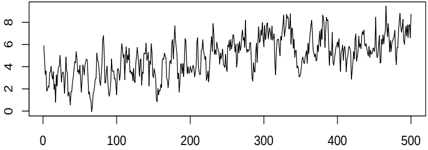
- an asymptotic test with level α

H_0 : no change H_1 : at least one structural change

- a consistent estimator of the number of change-points
- consistent estimators of the location of change-points.

Multiple Change-Point Problem





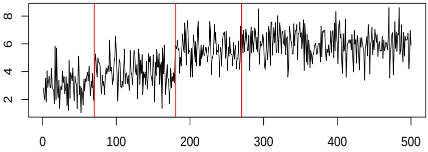
We want to construct:

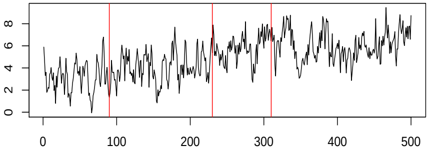
- an asymptotic test with level α

H_0 : no change H_1 : at least one structural change

- a consistent estimator of the number of change-points
- consistent estimators of the location of change-points.

Multiple Change-Point Problem





We want to construct:

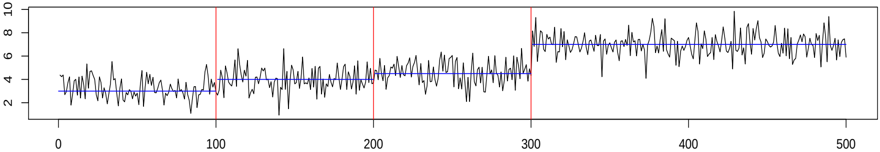
- an asymptotic test with level α

H_0 : no change $\qquad$ H_1 : at least one structural change

- a consistent estimator of the number of change-points
- consistent estimators of the location of change-points.

Outline

Classical Multiple Change-Point Model



$$X_i = \sum_{j=1}^{q+1} d_j I\{k_{j-1} < i \leq k_j\} + \varepsilon_i, \qquad i = 1, ..., n,$$

with random errors $\varepsilon_1, ..., \varepsilon_n$ and unknown

- change points $k_1, \ldots, k_q$ with $0 = k_0 < k_1 \leq \ldots \leq k_q \leq k_{q+1} = n$, $k_j = [\vartheta_j n]$, $j = 1, \ldots, q$, and $0 < \vartheta_1 \leq \ldots \vartheta_q \leq 1$
- number of changes $q \in \mathbb{N}$
- expectations $d_1, \ldots, d_{q+1}$ with $d_i \neq d_{i+1}$ for $i = 1, \ldots, q$.

Regime Switching Model

$$X_i^{(n)} = d_{Q_i^{(n)}} + \varepsilon_i^{(n)}, \quad i = 1, ..., n,$$

with random errors $\varepsilon_1^{(n)}, ..., \varepsilon_n^{(n)}$,

- expectations $d_1, ..., d_K \in \mathbb{R}$ with $d_i \neq d_j$, for $i, j = 1, \ldots, K$
- a non-observable $\{1, ..., K\}$-valued stationary process $\{Q_i^{(n)} : i \in \mathbb{N}\}$.

Key feature of $\{Q_i^{(n)} : i \in \mathbb{N}\}$: long duration times.

Differences to the classical change-point model:

- both number q_n and locations $k_1, ..., k_{q_n}$ of structural breaks are random
- the unbounded number of changes q_n.

Regime Switching Model

$$X_i^{(n)} = d_{Q_i^{(n)}} + \varepsilon_i^{(n)}, \quad i = 1, ..., n,$$

with random errors $\varepsilon_1^{(n)}, ..., \varepsilon_n^{(n)}$,

- expectations $d_1, ..., d_K \in \mathbb{R}$ with $d_i \neq d_j$, for $i, j = 1, \ldots, K$
- a non-observable $\{1, ..., K\}$-valued stationary process $\{Q_i^{(n)} : i \in \mathbb{N}\}$.

Key feature of $\{Q_i^{(n)} : i \in \mathbb{N}\}$: long duration times.

Differences to the classical change-point model:

- both number q_n and locations $k_1, ..., k_{q_n}$ of structural breaks are random
- the unbounded number of changes q_n.

Assumptions on errors

Let the errors $\varepsilon_1, ..., \varepsilon_n$ be a strictly stationary sequence with

(1) $$E\varepsilon_1 = 0, \quad 0 < \sigma^2 = E\varepsilon_1^2 < \infty, \quad E\left|\varepsilon_1\right|^{2+\nu} < \infty \text{ for some } \nu > 0,$$
$$\sum_{h \geq 0} |\gamma(h)| < \infty, \text{ where } \gamma(h) = \operatorname{cov}(\varepsilon_0, \varepsilon_h),$$

and long run variance $\tau^2 = \sigma^2 + 2\sum_{h>0} \gamma(h) < \infty.$

(2) Invariance principle:

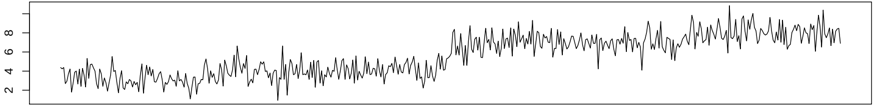
It exists a Wiener process $\{W(k),\ 1 \leq k \leq n\}$ such that

$$\max_{G \leq k \leq n-G} \frac{1}{\sqrt{2G}} \left| \frac{1}{\tau} \sum_{i=k+1}^{k+G} \varepsilon_i - (W(k+G) - W(k)) \right| = o_p\left((\log(n/G))^{-\frac{1}{2}} \right).$$

(3) Hájek-Rényi-type moment condition:

$$E\left|\sum_{k=i}^{j} \varepsilon_k\right|^{\gamma} \leq C|j-i+1|^{\varphi} \text{ for some } \gamma \geq 1,\ \varphi > 1 \text{ and some constant } C > 0.$$

MOSUM (Moving Sum) Statistic



MOSUM statistic (Hušková and Slabý (2001))

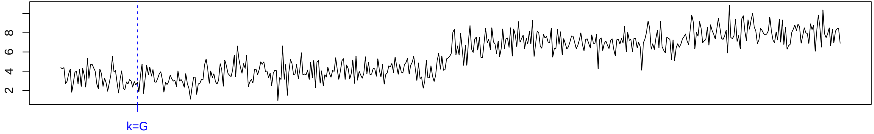
$$T_n(G) = \max_{G \le k \le n-G} \frac{T_{k,n}(G)}{\tau} \text{ with}$$

$$T_{k,n}(G) = \frac{1}{\sqrt{2G}} \left| \sum_{i=k-G+1}^{k} X_i - \sum_{i=k+1}^{k+G} X_i \right|,$$

where $G = G(n)$ is the bandwidth fulfilling

$$\frac{n^{\frac{2}{2+\nu}} \log n}{G} \to 0, \ \frac{n}{G} \to \infty \ .$$

MOSUM (Moving Sum) Statistic



MOSUM statistic (Hušková and Slabý (2001))

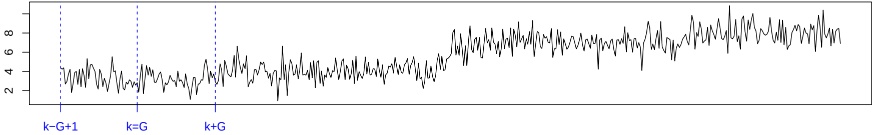
$$T_n(G) = \max_{G \leq k \leq n-G} \frac{T_{k,n}(G)}{\tau} \text{ with}$$

$$T_{k,n}(G) = \frac{1}{\sqrt{2G}} \left| \sum_{i=k-G+1}^{k} X_i - \sum_{i=k+1}^{k+G} X_i \right|,$$

where $G = G(n)$ is the bandwidth fulfilling

$$\frac{n^{\frac{2}{2+\nu}} \log n}{G} \rightarrow 0, \ \frac{n}{G} \rightarrow \infty \ .$$

MOSUM (Moving Sum) Statistic



MOSUM statistic (Hušková and Slabý (2001))

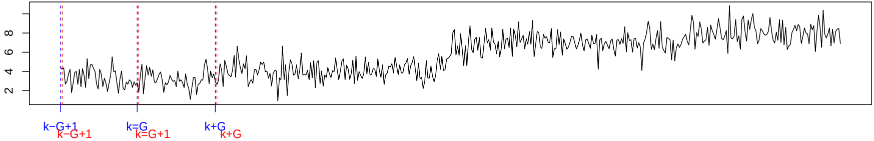
$$T_n(G) = \max_{G \leq k \leq n-G} \frac{T_{k,n}(G)}{\tau} \text{ with}$$

$$T_{k,n}(G) = \frac{1}{\sqrt{2G}} \left| \sum_{i=k-G+1}^{k} X_i - \sum_{i=k+1}^{k+G} X_i \right|,$$

where $G = G(n)$ is the bandwidth fulfilling

$$\frac{n^{\frac{2}{2+\nu}} \log n}{G} \rightarrow 0, \ \frac{n}{G} \rightarrow \infty \ .$$

MOSUM (Moving Sum) Statistic



MOSUM statistic (Hušková and Slabý (2001))

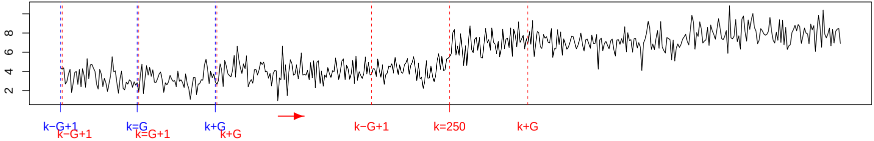
$$T_n(G) = \max_{G \le k \le n-G} \frac{T_{k,n}(G)}{\tau} \quad \text{with}$$

$$T_{k,n}(G) = \frac{1}{\sqrt{2G}} \left| \sum_{i=k-G+1}^{k} X_i - \sum_{i=k+1}^{k+G} X_i \right|,$$

where $G = G(n)$ is the bandwidth fulfilling

$$\frac{n^{\frac{2}{2+\nu}} \log n}{G} \to 0, \; \frac{n}{G} \to \infty \; .$$

MOSUM (Moving Sum) Statistic



MOSUM statistic (Hušková and Slabý (2001))

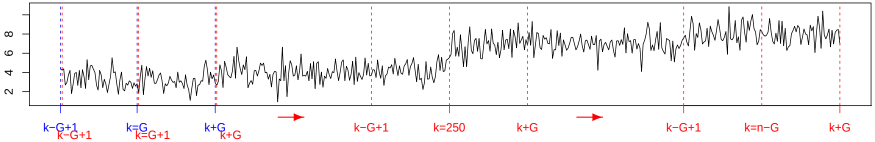
$$T_n(G) = \max_{G \leq k \leq n-G} \frac{T_{k,n}(G)}{\tau} \quad \text{with}$$

$$T_{k,n}(G) = \frac{1}{\sqrt{2G}} \left| \sum_{i=k-G+1}^{k} X_i - \sum_{i=k+1}^{k+G} X_i \right| ,$$

where $G = G(n)$ is the bandwidth fulfilling

$$\frac{n^{\frac{2}{2+\nu}} \log n}{G} \to 0, \ \frac{n}{G} \to \infty \ .$$

MOSUM (Moving Sum) Statistic



MOSUM statistic (Hušková and Slabý (2001))

$$T_n(G) = \max_{G \leq k \leq n-G} \frac{T_{k,n}(G)}{\tau} \text{ with}$$

$$T_{k,n}(G) = \frac{1}{\sqrt{2G}} \left| \sum_{i=k-G+1}^{k} X_i - \sum_{i=k+1}^{k+G} X_i \right|,$$

where $G = G(n)$ is the bandwidth fulfilling

$$\frac{n^{\frac{2}{2+\nu}} \log n}{G} \rightarrow 0, \ \frac{n}{G} \rightarrow \infty \ .$$

Asymptotic Distribution

KIT Karlsruhe Institute of Technology

Theorem (Hušková and Slabý (2001), Kirch and M. (2012))

Let the assumptions on errors (1)-(2) and bandwidth G hold. Then, under H_0,

$$\alpha(n/G) \max_{G \leq k \leq n-G} \frac{T_{k,n}(G)}{\tau} - \beta(n/G) \xrightarrow{\mathcal{D}} \Gamma,$$

where Γ has a Gumbel extreme value distribution, i.e.

$$P(\Gamma \leq x) = exp(-2exp(-x)),$$

and $\alpha(x) = \sqrt{2 \log x}$, $\beta(x) = 2\log(x) + \frac{1}{2} \log\log x - \frac{1}{2} \log \pi$.

- critical value: $D_n(G;\alpha) = \dfrac{\beta(n/G) - \log\log \frac{1}{\sqrt{1-\alpha}}}{\alpha(n/G)}$
- $\alpha_n \to 0$, but not too fast.

Asymptotic Distribution

Theorem (Hušková and Slabý (2001), Kirch and M. (2012))

Let the assumptions on errors (1)-(2) and bandwidth G hold. Then, under H_0,

$$\alpha(n/G) \max_{G \le k \le n-G} \frac{T_{k,n}(G)}{\tau} - \beta(n/G) \xrightarrow{\mathcal{D}} \Gamma,$$

where Γ has a Gumbel extreme value distribution, i.e.

$$P(\Gamma \le x) = exp(-2exp(-x)),$$

and $\alpha(x) = \sqrt{2 \log x}$, $\beta(x) = 2\log(x) + \frac{1}{2} \log\log x - \frac{1}{2} \log \pi$.

- critical value: $D_n(G;\alpha) = \dfrac{\beta(n/G) - \log\log \frac{1}{\sqrt{1-\alpha}}}{\alpha(n/G)}$
- $\alpha_n \to 0$, but not too fast.

Assumptions on errors

Let the errors $\varepsilon_1, ..., \varepsilon_n$ be a strictly stationary sequence with

(1) $$E\varepsilon_1 = 0, \quad 0 < \sigma^2 = E\varepsilon_1^2 < \infty, \quad E|\varepsilon_1|^{2+\nu} < \infty \text{ for some } \nu > 0,$$
$$\sum_{h \geq 0} |\gamma(h)| < \infty, \text{ where } \gamma(h) = \text{cov}(\varepsilon_0, \varepsilon_h),$$

and long run variance $\tau^2 = \sigma^2 + 2\sum_{h>0} \gamma(h) < \infty.$

(2) Invariance principle:

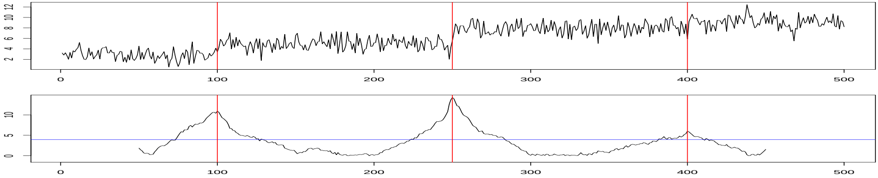
It exists a Wiener process $\{W(k),\ 1 \leq k \leq n\}$ such that

$$\max_{G \leq k \leq n-G} \frac{1}{\sqrt{2G}} \left| \frac{1}{\tau} \sum_{i=k+1}^{k+G} \varepsilon_i - (W(k+G) - W(k)) \right| = o_p\left((\log(n/G))^{-\frac{1}{2}} \right).$$

(3) Hájek-Rényi-type moment condition:

$$E\left| \sum_{k=i}^{j} \varepsilon_k \right|^{\gamma} \leq C|j-i+1|^{\varphi} \text{ for some } \gamma \geq 1, \varphi > 1 \text{ and some constant } C > 0.$$

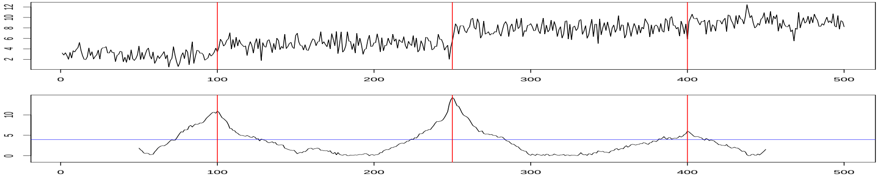
Change-Point Estimators (Antoch et al. (2000))



Test statistic: $T_n(G) := \max_{G \leq k \leq n-G} T_{k,n}(G)/\tau$

— critical value $D_n(G; \alpha_n)$

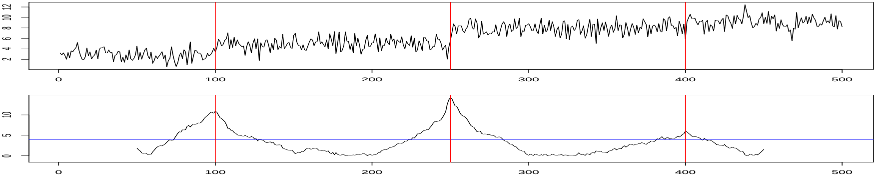
Change-Point Estimators (Antoch et al. (2000))



Test statistic: $T_n(G) := \max_{G \le k \le n-G} T_{k,n}(G)/\tau$

— critical value $D_n(G; \alpha_n)$

Change-Point Estimators (Antoch et al. (2000))



All pairs of indices v_j, w_j are chosen such that

$$\begin{aligned} T_{k,n}(G)/\tau &\geq D_n(G;\alpha_n) \quad k = v_j, \dots, w_j, \\ T_{k,n}(G)/\tau &< D_n(G;\alpha_n) \quad k = v_j - 1, w_j + 1 \\ w_j - v_j &> \varepsilon G. \end{aligned}$$

- The number of change-points q can be estimated by $\hat{q}_n$, the number of pairs (v_j, w_j).
- The estimator of change-point k_j is defined as $\hat{k}_j := \arg \max_{v_j \leq k \leq w_j} T_{k,n}(G)/\tau$.

Consistency of $\hat{q}_n$

Classical model:

Theorem (Kirch and M. (2012))

Let the assumptions on errors (1)-(2), bandwidth G and level $\{\alpha_n\}$ hold. Furthermore assume

$$\limsup_{n\to\infty} d_0(n)/G = C > 2 \qquad \text{with} \qquad d_0(n) := \min_{0\leq j\leq q} |k_{j+1} - k_j|.$$

Then, under H_1,

$$P(\hat{q}_n = q \ ) \longrightarrow 1 \text{ as } n \to \infty.$$

Consistency of $\hat{q}_n$

Regime switching model:

Theorem (Kirch and M. (2012))

Let the assumptions on errors (1)-(2), bandwidth G and level $\{\alpha_n\}$ hold. Furthermore assume

$$\limsup_{n\to\infty} d_0(n)/G = C > 2 \qquad \text{with} \qquad d_0(n) := \min_{0\leq j\leq q} |k_{j+1} - k_j|.$$

Then, under H_1,

$$P(\hat{q}_n = q\ ) \longrightarrow 1 \text{ as } n \to \infty.$$

Consistency of $\hat{q}_n$

Karlsruhe Institute of Technology

Regime switching model:

Theorem (Kirch and M. (2012))

Let the assumptions on errors (1)-(2), bandwidth G and level $\{\alpha_n\}$ hold. Furthermore assume

$$\lim_{n\to\infty} P(d_0(n) > 2G) = 1 \qquad \text{with} \qquad d_0(n) := \min_{0\leq j\leq q_n} |k_{j+1} - k_j|.$$

Then, under H_1,

$$P(\hat{q}_n = q_n) \longrightarrow 1 \text{ as } n \to \infty.$$

Assumptions on errors

Let the errors $\varepsilon_1, ..., \varepsilon_n$ be a strictly stationary sequence with

(1) $$E\varepsilon_1 = 0, \quad 0 < \sigma^2 = E\varepsilon_1^2 < \infty, \quad E\left|\varepsilon_1\right|^{2+\nu} < \infty \text{ for some } \nu > 0,$$
$$\sum_{h \geq 0} |\gamma(h)| < \infty, \text{ where } \gamma(h) = \text{cov}(\varepsilon_0, \varepsilon_h),$$

and long run variance $\tau^2 = \sigma^2 + 2\sum_{h>0} \gamma(h) < \infty.$

(2) Invariance principle:

It exists a Wiener process $\{W(k),\ 1 \leq k \leq n\}$ such that

$$\max_{G \leq k \leq n-G} \frac{1}{\sqrt{2G}} \left| \frac{1}{\tau} \sum_{i=k+1}^{k+G} \varepsilon_i - (W(k+G) - W(k)) \right| = o_p\left((\log(n/G))^{-\frac{1}{2}} \right).$$

(3) Hájek-Rényi-type moment condition:

$$E\left|\sum_{k=i}^{j} \varepsilon_k\right|^{\gamma} \leq C|j-i+1|^{\varphi} \text{ for some } \gamma \geq 1,\ \varphi > 1 \text{ and some constant } C > 0.$$

Consistency of Change-Point Estimators

Karlsruhe Institute of Technology

Classical model:

Theorem (Kirch and M. (2012))

Let the assumptions on errors (1)-(3), bandwidth G and level $\{\alpha_n\}$ hold. Furthermore assume

$$\limsup_{n\to\infty} d_0(n)/G = C > 2 \qquad \text{with} \qquad d_0(n) := \min_{0\le j\le q} |k_{j+1} - k_j|.$$

With $\overline{q}_n := \min(q\ ,\hat{q}_n)$ we have, under H_1,

$$\max_{1\le j\le \overline{q}_n} |\hat{k}_j - k_j| \;=\; O_p\,(1)\,.$$

Consistency of Change-Point Estimators

Karlsruhe Institute of Technology

Regime switching model:

Theorem (Kirch and M. (2012))

Let the assumptions on errors (1)-(3), bandwidth G and level $\{\alpha_n\}$ hold. Furthermore assume

$$\limsup_{n\to\infty} d_0(n)/G = C > 2 \qquad \text{with} \qquad d_0(n) := \min_{0\le j\le q} |k_{j+1} - k_j|.$$

With $\overline{q}_n := \min(q\ , \hat{q}_n)$ we have, under H_1,

$$\max_{1\le j\le \overline{q}_n} |\hat{k}_j - k_j| \;=\; O_p(1)\,.$$

Consistency of Change-Point Estimators

Karlsruhe Institute of Technology

Regime switching model:

Theorem (Kirch and M. (2012))

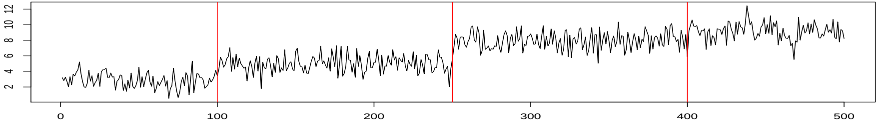
Let the assumptions on errors (1)-(3), bandwidth G and level $\{\alpha_n\}$ hold. Furthermore assume

$$\lim_{n\to\infty} P(d_0(n) > 2G) = 1 \qquad \text{with} \qquad d_0(n) := \min_{0\le j\le q_n} |k_{j+1} - k_j|$$

and let $P(q_n > \gamma_n) \to 0$, where $\{\gamma_n\}$ satisfies $\gamma_n \log(G) G^{-\frac{\gamma}{2}} \to 0$.
With $\overline{q}_n := \min(q_n, \hat{q}_n)$ we have, under H_1,

$$\max_{1\le j\le \overline{q}_n} |\hat{k}_j - k_j| \quad = \quad O_p\left(\gamma_n^{\frac{2}{\gamma}}\right).$$

Variance Estimators (i.i.d. case)



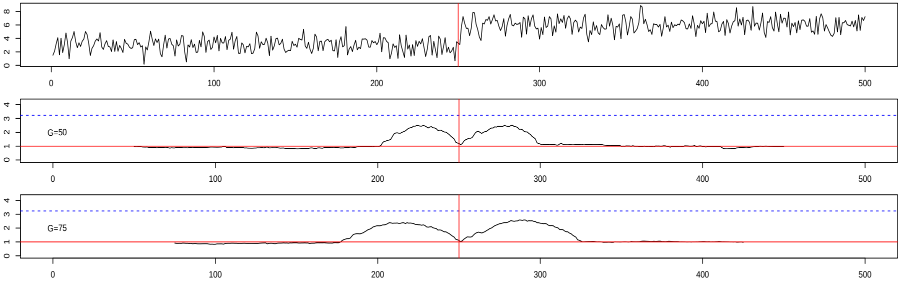
- Under H_1 the standard variance estimator overestimates the variance.

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

- Solution: Variance estimator depends on time point k.

$$\hat{\sigma}_{k,n}^2 := \frac{1}{2G}\left(\sum_{i=k-G+1}^{k}(X_i - \overline{X}_{k-G+1,k})^2 + \sum_{i=k+1}^{k+G}(X_i - \overline{X}_{k+1,k+G})^2\right).$$

Performance of $\hat{\sigma}^2_{k,n}$



— σ^2

- - - $\hat{\sigma}^2_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$

— $\hat{\sigma}^2_{k,n} = \frac{1}{2G}\left(\sum_{i=k-G+1}^{k}(X_i - \overline{X}_{k-G+1,k})^2 + \sum_{i=k+1}^{k+G}(X_i - \overline{X}_{k+1,k+G})^2\right)$

Asymptotic Results with $\hat{\tau}_{k,n}$

Lemma (Kirch and M. (2012))

If the long run variance estimator $\hat{\tau}^2_{k,n}$ fulfills

$$\max_{G \le k \le n-G} |\hat{\tau}_{k,n} - \tau| = o_p\left((\log(n/G))^{-\frac{1}{2}}\right) \text{ under } H_0$$

and

$$\max_{G \le k \le n-G} \hat{\tau}_{k,n} = O_p(1) \text{ under } H_1$$

all of the above results remain true.

Example: $\hat{\sigma}^2_{k,n}$ in the case of i.i.d. random variables.

Simulation Study

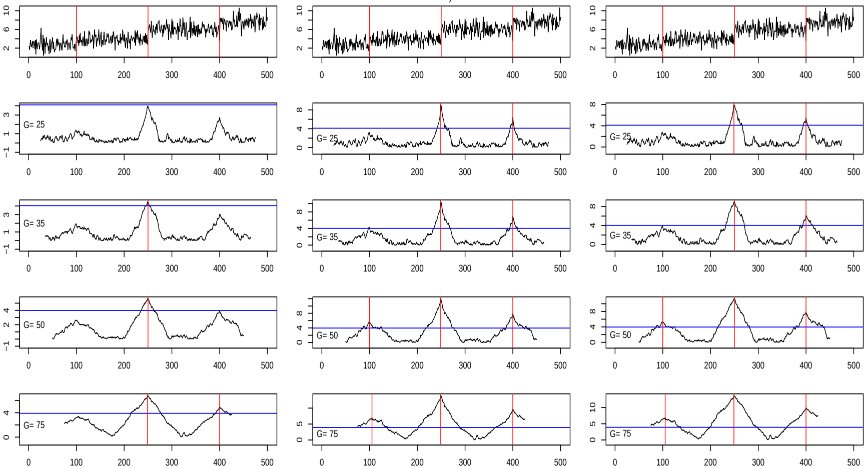
Questions:

- How good is the general performance of the MOSUM procedure?
- How does the choice of the variance estimator influence the performance of the MOSUM procedure?
- How does the bandwidth selection influence the performance?

Simulation Study

$$\varepsilon_1, \ldots, \varepsilon_n \text{ i.i.d. with } \varepsilon_1 \sim \mathcal{N}(0,1)$$

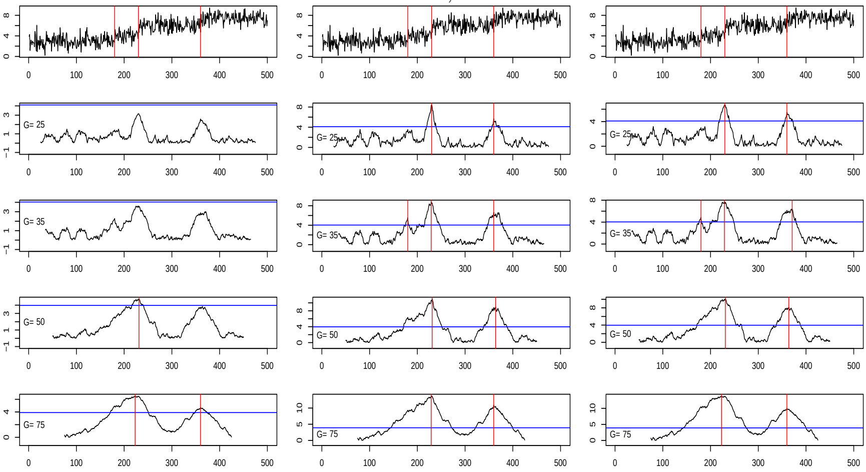
σ_n^2 | $\sigma_{k,n}^2$ | σ^2



Simulation Study

$$\varepsilon_1, \ldots, \varepsilon_n \text{ i.i.d. with } \varepsilon_1 \sim \mathcal{N}(0,1)$$

σ_n^2 | $\sigma_{k,n}^2$ | σ^2



Conclusion

Karlsruhe Institute of Technology

- We have theoretically justified the use of the MOSUM procedure for both the classical model as well as the regime switching model by analysing the consistency of the change point estimators.
- In the simulation study the procedure gives good estimates for the change points (as long as the bandwidth G is appropriate) and is additionally easy to implement.

Future research:

- MOSUM procedure for change detection in more general models, i.e. autoregressive or ARMA models.

References

- Antoch, J. and Hušková, M. and Jarušková, D. (2000). Change point detection. *5th ERS IASC Summer School*.
- Hušková, M. and Slabý, A. (2001). Permutation tests for multiple changes. *Kybernetica*, **37**, 605 – 622.
- Kirch, C. and Muhsal, B. (2012). Change-point methods for multiple structural breaks and regime switching models. *In Preparation*.

Thank you for your attention!