MR4178193 62M10 62G10 62G20
Dette, Holger (D-BCHMM); **Eckle, Theresa** (D-BCHMM);
Vetter, Mathias (D-KIEL-MS)
Multiscale change point detection for dependent data. (English summary)
Scand. J. Stat. **47** (2020), *no. 4,* 1243–1274.

To study the problem of detecting multiple abrupt changes in time-series data, this article concentrates on the simultaneous multiscale change point estimator (SMUCE) of the signal part in the model

$$Y_i = \vartheta^* \left(\frac{i}{n}\right) + \varepsilon_i, \quad i = 1, \dots, n.$$

The signal $\vartheta^*\colon [0,1] \to \mathbb{R}$ is defined as an unknown piecewise linear function

$$\vartheta^*(t) = \sum_{k=0}^{K^*} \theta_k^* \mathbf{1}_{[\tau_k^*, \tau_{k+1}^*)}(t),$$

where K^* is the unknown number of change points, $0 = \tau_0^* < \tau_1^* < \cdots < \tau_{K^*}^* < \tau_{K^*+1}^* = 1$ are the change point locations, and $\theta_0^*, \dots, \theta_{K^*}^*$ are the function values of ϑ^*. The vector of change points is denoted as $J(\vartheta^*) = (\tau_1^*, \dots, \tau_{K^*}^*)$ with $|J(\vartheta^*)|$ being its dimension.

Consistency of the SMUCE of ϑ^* for an i.i.d. Gaussian error ε_i has already been established. This article aims at demonstrating consistency of the estimator when the error process ε_i is a so-called physically dependent sequence [W. B. Wu, Proc. Natl. Acad. Sci. USA **102** (2005), no. 40, 14150–14154; MR2172215]. The change point location estimates $\widehat{\tau}_k$ are assumed to take only the values $\{0, 1/n, \dots, (n-1)/n\}$. The set of such functions is denoted as $\mathcal{S}_n$, and the estimator of ϑ^* is chosen among the class $\vartheta(\cdot) = \sum_{k=0}^{K} \theta_k \mathbf{1}_{[\tau_k, \tau_{k+1})}(\cdot) \in \mathcal{S}_n$. For this purpose, let the multiscale statistic be

$$V_n(Y, \vartheta) = \max_{0 \le k \le K} \sup_{\substack{n\tau_k \le i \le j \le n\tau_{k+1} \\ j-i+1 \ge nc_n}} \left\{ \frac{1}{\widehat{\sigma}} \sqrt{j-i+1} \left| \overline{Y}_i^j - \theta_k \right| - \sqrt{2 \log \frac{en}{j-i+1}} \right\},$$

where $c_n > 0$ is a sequence converging to zero and $\overline{Y}_i^j = \sum_{\ell=i}^{j} Y_\ell/(j-i+1)$ is a local mean. In addition, $\widehat{\sigma}^2$ is a long-run variance estimator of the error process, and a difference-based one is considered. Let the entire sample $\{Y_i\}_{i=1}^n$ be partitioned into $m_n = \lfloor n/k_n \rfloor$ sub-samples $\{Y_1, \dots, Y_{k_n}\}$, $\{Y_{k_n+1}, \dots, Y_{2k_n}\}, \dots, \{Y_{(m_n-1)k_n+1}, \dots, Y_{m_n k_n}\}$ with equal length k_n. For a local average of the sub-sample $A_i = \sum_{j=1}^{k_n} Y_{ik_n+j}/k_n$, the difference-based variance estimator is given by

$$\widehat{\sigma}^2 = \frac{k_n}{2(m_n-1)} \sum_{i=1}^{m_n-1} |A_i - A_{i-1}|^2.$$

The estimation procedure of the piecewise linear function ϑ^* takes two steps. First, for a threshold q chosen via the limiting null distribution of V_n given below, the number of change points K^* is estimated as

$$\widehat{K} = \widehat{K}(V_n, q) = \inf_{\substack{\vartheta \in \mathcal{S}_n \\ V_n(Y,\vartheta) \le q}} |J(\vartheta)|.$$

Second, ϑ^* is estimated as

$$\widehat{\vartheta} = \underset{\vartheta \in C(V_n, q)}{\arg\min} \sum_{i=1}^{n} \left\{ Y_i - \vartheta\left(\frac{i}{n}\right) \right\}^2$$

for the set $C(V_n, q) = \{\vartheta \in \mathcal{S}_n : |J(\vartheta)| = \widehat{K} \text{ and } V_n(Y, \vartheta) \leq q\}$. The only difference in the estimation procedure from the Gaussian error case is to employ the long-run variance estimator $\widehat{\sigma}^2$ in V_n.

If $k_n \asymp n^{1/3}$ and other regularity conditions hold, then V_n has the limiting null distribution

$$V_n(Y, \vartheta^*) \xrightarrow{d} \max_{0 \leq k \leq K^*} \sup_{\tau_k^* \leq s < t \leq \tau_{k+1}^*} \left\{ \frac{|B(t) - B(s)|}{\sqrt{t-s}} - \sqrt{2 \log \frac{e}{t-s}} \right\},$$

where $\{B(t)\}_{t \in [0,1]}$ is the standard Brownian motion. This limiting distribution coincides with the one for the Gaussian error case. Subsequently, the following two results on consistency of SMUCE are established:

$$\lim_{n \to \infty} \Pr\left\{ \widehat{K}(V_n, q_n) = K^* \right\} = 1$$

for a sequence q_n satisfying $1/q_n + q_n/\sqrt{n} \to 0$; and

$$\lim_{n \to \infty} \Pr\left\{ \sup_{\vartheta \in C(V_n, q_n)} |\tau_k^* - \tau_k| > c_n \right\} = 0$$

for $k = 1, \ldots, K^*$. The first and second statements demonstrate consistencies of model selection and change point locations in ϑ^*, respectively. *Masayuki Hirukawa*

References

1. Anastasiou, A. & Fryzlewicz, P. (2019). *Detecting multiple generalized change-points by isolating single ones.* arXiv:1901.10852v1.
2. Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica, 66*, 47–78. MR1616121
3. Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics, 18*, 1–22.
4. Baranowski, R., Chen, Y., & Fryzlewicz, P. (2019a). Narrowest-over-threshold detection of multiple change-points and change-point-like features. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 81*, 649–672. MR3961502
5. Baranowski, R., Chen, Y. & Fryzlewicz, P. (2019b). Not: Narrowest-over-threshold change-point detection. R package version 1.2.
6. Braun, J. V., Braun, R. K., & Muller, H.-G. (2000). Multiple changepoint fitting via quasilikelihood, with application to DNA sequence segmentation. *Biometrika, 87*, 301–314. MR1782480
7. Chakar, S., Lebarbier, E., Lévy-Leduc, C., & Robin, S. (2017). A robust approach for estimating change-points in the mean of an AR(1) process. *Bernoulli, 23*, 1408–1447. MR3606770
8. Cho, H., & Fryzlewicz, P. (2015). Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 77*, 475–507. MR3310536
9. Ciuperca, G. (2011). A general criterion to determine the number of change-points. *Statistics & Probability Letters, 81*, 1267–1275. MR2803773
10. Ciuperca, G. (2014). Model selection by LASSO methods in a change-point model. *Statistics Papers, 55*, 349–374. MR3188408

11. Davis, R. A., Lee, T. C. M., & Rodriguez-Yam, G. A. (2006). Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association, 101*, 223–239. MR2268041
12. Dette, H., & Gösmann, J. (2018). Relevant change points in high dimensional time series. *Electronic Journal of Statistics, 12*, 2578–2636. MR3849896
13. Dette, H., Munk, A., & Wagner, T. (1998). Estimating the variance in nonparametric regression - what is a reasonable choice? *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 60*, 751–764. MR1649480
14. Eichinger, B., & Kirch, C. (2018). A MOSUM procedure for the estimation of multiple random change points. *Bernoulli, 24*, 526–564. MR3706768
15. Frick, K., Munk, A., & Sieling, H. (2014). Multiscale change point inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76*, 495–580. MR3210728
16. Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *The Annals of Statistics, 42*, 2243–2281. MR3269979
17. Fryzlewicz, P. (2018a). Supplement to "Tail-greedy bottom-up data decompositions and fast multiple change-point detection". *DOI:*https://doi.org/10.1214/17-AOS1662SUPP.
18. Fryzlewicz, P. (2018b). Tail-greedy bottom-up data decompositions and fast multiple change-point detection. *The Annals of Statistics, 46*, 3390–3421. MR3852656
19. Hall, P., Kay, J. W., & Titterington, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika, 77*, 521–528. MR1087842
20. Harchaoui, Z., & Lévy-Leduc, C. (2010). Multiple change-point estimation with a total variation penalty. *Journal of the American Statistical Association, 105*, 1480–1493. MR2796565
21. Haynes, K., Fearnhead, P., & Eckley, I. A. (2017). A computationally efficient nonparametric approach for changepoint detection. *Statistics and Computing, 27*, 1293–1305. MR3647098
22. Jirak, M. (2015). Uniform change point tests in high dimension. *The Annals of Statistics, 43*, 2451–2483. MR3405600
23. Killick, R., Fearnhead, P., & Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association, 107*, 1590–1598. MR3036418
24. Kolaczyk, E. D., & Nowak, R. D. (2005). Multiscale generalised linear models for nonparametric function estimation. *Biometrika, 92*, 119–133. MR2158614
25. Korkas, K., & Fryzlewicz, P. (2017). Multiple change-point detection for non-stationary time series using wild binary segmentation. *Statistica Sinica, 27*, 287–311. MR3618170
26. Lavielle, M., & Moulines, E. (2000). Least-squares estimation of an unknown number of shifts in a time series. *Journal of Time Series Analysis, 21*, 33–59. MR1766173
27. Li, H., Guo, Q., & Munk, A. (2019). Multiscale change-point segmentation: Beyond step functions. *Electronic Journal of Statistics, 13*, 3254–3296. MR4010980
28. Li, H., Munk, A., & Sieling, H. (2016). FDR-control in multiscale change-point segmentation. *Electronic Journal of Statistics, 10*, 918–959. MR3486421
29. Liu, W., Xiao, H., & Wu, W. B. (2013). Probability and moment inequalities under dependence. *Statistica Sinica, 23*, 1257–1272. MR3114713
30. Matteson, D. S., & James, N. A. (2014). A nonparametric approach for multiple change point analysis of multivariate data. *Journal of the American Statistical Associations, 109*, 334–345. MR3180567
31. McMurry, T. L., & Politis, D. N. (2010). Banded and tapered estimates for autoco-

variance matrices and the linear process bootstrap. *Journal of Time Series Analysis, 31*, 471–482. MR2732601
32. Meier, A., Cho, H. & Kirch, C. (2019). *Mosum: Moving sum based procedures for changes in the mean*. R package version 1.2.1.
33. Padilla, O. H. M., Yu, Y., Wang, D. & Rinaldo, A. (2019). *Optimal nonparametric change point detection and localization*. arXiv:1905.10019v1.
34. Pein, F., Hotz, T., Sieling, H. & Aspelmeier, T. (2017). *StepR: Multiscale change-point inference*. R package version 2.0-1.
35. Pein, F., Sieling, H., & Munk, A. (2017). Heterogeneous change point inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 79*, 1207–1227. MR3689315
36. Preuss, P., Puchstein, R., & Dette, H. (2015). Detection of multiple structural breaks in multivariate time series. *Journal of the American Statistical Association, 110*, 654–668. MR3367255
37. Shao, Q.-M. (1995). On a conjecture of Révész. *Proceedings of the American Mathematical Society, 123*, 575–582. MR1231304
38. Tecuapetla-Gómez, I. (2015). *dbacf: Autocovariance estimation via difference-based methods*. R package version 0.0.0.9000.
39. Tecuapetla-Gómez, I., & Munk, A. (2017). Autocovariance estimation in regression with a discontinuous signal and m-dependent errors: A difference-based approach. *Scandinavian Journal of Statistics, 44*, 346–368. MR3658518
40. Wang, D., Yu, Y. & Rinaldo, A. (2019). *Univariate mean change point detection: Penalization, CUSUM and optimality*. arXiv:1810.09498v4. MR4091859
41. Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences, 102*, 14150–14154. MR2172215
42. Wu, W. B. (2011). Asymptotic theory for stationary processes. *Statistics and its Interface, 4*, 207–226. MR2812816
43. Wu, W. B., & Pourahmadi, M. (2009). Banding sample autocovariance matrices of stationary processes. *Statistica Sinica, 19*, 1755–1768. MR2589209
44. Wu, W. B., & Zhao, Z. (2007). Inference of trends in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 69*, 391–410. MR2323759
45. Wu, W. B., & Zhou, Z. (2011). Gaussian approximations for non-stationary multiple time series. *Statistica Sinica, 21*, 1397–1413. MR2827528
46. Yao, Y. (1988). Estimating the number of change-points via Schwarz' criterion. *Statistics & Probability Letters, 6*, 181–189. MR0919373
47. Yau, C. Y., & Zhao, Z. (2016). Inference for multiple change points in time series via likelihood ratio scan statistics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 78*, 895–916. MR3534355
48. Zou, C., Yin, G., Feng, L., & Wang, Z. (2014). Nonparametric maximum likelihood approach to multiple change-point problems. *The Annals of Statistics, 42*, 970–1002. MR3210993

Note: This list reflects references listed in the original paper as accurately as possible with no attempt to correct errors.