Robust modelling data series with abrupt structure changes at unknown places

Kang-Ping Lu
Department of Applied Statistics
National Taichung University of Science and Technology
Taichung, Taiwan
kplu@nutc.edu.tw, kplu5208@gmail.com

Shao-Tung Chang
Department of Mathematics
National Taiwan Normal University
Taipei, Taiwan
schang4809@gmail.com

Abstract—In many real problems, the data structure of a data series changes abruptly at some places called break-points or switch-points. Estimation for break-points is important for knowing how the structure of data changes, the information which is critical for decision-making. In particular, the importance of break-points can be revealed through break-point model analysis but this important information is discarded in smoothing model analysis. Such data can be encountered in many applications such as medical research, industrial engineering, climatology and finance studies. Break-point problem has become an important subject in recent studies, especially the explosion in big data applications. In this study, we consider incorporating latent class analysis with the M-estimation technique to make robust estimations for break-point models. The break-points are regarded as latent class variables and hence the break-point model can be transferred into a latent class model. Consequently, the latent class analysis is attainable for estimating break-points and model parameters. Moreover, consider real datasets are often contaminated by outliers or have heavy-tailed distributions, but clustering methods are not resistant to atypical observations. Consequently, estimators can be distorted seriously due to the existence of anomalous data. In order to derive robust estimators, the M-estimation associated with a robust criterion is embedded into the latent class analysis. Furthermore, because of the vagueness of boundary between two adjacent segments, the concept of soft partitioning is considered in the latent class analysis for better precision of estimations. The robust criteria frequently used in statistics such as the Tukey's criterion and the absolute deviations, are feasible to construct a robust method for solving break-point problems. The effectiveness and practicability of the proposed method are shown through experiments with numerical examples.

Keywords—break-point model, latent class analysis, robust

[1] R. E. Quandt, The estimation of the parameters of a linear regression system obeying two separate regimes, J. Am. Stat. Assoc. 53 (1958) 873–880.
[2] [2] S. T. Chang, K. P. Lu, and M. S. Yang, Fuzzy Change-Point Algorithms for Regression Models, IEEE Trans. Fuzzy Syst. 23 (2015) 2343-2357.
[3] G. Ciuperca The M-estimation in a multi-phase random nonlinear model. Statistics and Probability Letters 79 (2009) 573-580.
[4] G. Ciuperca Estimating nonlinear regression with and without change-points by the LAD method. Ann Inst Stat Math (2011) 63:717–743.
[5] P. J. Huber, E. M. Ronchetti, Robust Statistics, Second Edition. John Wiley & Sons, New York, 2009.
[6] X. Bai, W. Yao, and J. E. Boyer, Robust fitting of mixture regression models, Comput. Stat. Data. An. 56 (2012) 2347-2359.