XXIII Jornadas de Classificação e Análise de Dados
Universidade de Évora, Évora, 31 de março a 2 de Abril de 2016

Sessão Paralela I – Modelos Estocásticos e Modelos Espaciais – 6ª feira, 1 de abril, Sala 124 (9:30)

Wavelet-based detection of outliers in time series of counts

Isabel Silva[1], **Maria Eduarda Silva**[2]
[1] *Faculdade de Engenharia da Universidade do Porto and CIDMA, ims@fe.up.pt;*
[2] *Faculdade de Economia da Universidade do Porto and CIDMA, mesilva@fep.up.pt*

Abstract: In this work we consider the problem of analysing count time series contaminated with outliers. To address this problem we propose a wavelet-based approach that allows the identification of the time point of occurrence of an outlier in a time series of counts, by using the empirical distribution of the detail coefficient via resampling methods (parametric bootstrap). Results of a simulation study illustrating the effectiveness of the proposed method and a real dataset application are presented.

Keywords: Additive outlier, Discrete wavelet transform, Haar wavelet, INAR(1) models, Parametric resampling.

A problem of interest in time series modelling is to detect outliers which can be seen as discrepant observations. These observations originate biased model estimates and consequently invalid inferences. For conventional ARMA processes estimation and inference in the presence of outliers are well documented. However, the impact of outliers on the parameter estimation for time series of counts has not yet been completely addressed.

Time series of (small) counts are common in practice and appear in a wide variety of fields: social science, biology and environmental processes, economics and finance, telecommunications and insurance, among others. Different types of models that explicitly account for the discreteness of the data have been proposed in the literature. In this work we consider the INAR(1) (INteger-valued AutoRegressive) models, which are constructed by replacing the multiplication in the conventional AR models by an appropriate random operator called thinning operator (for details, see Scotto *et al.* 2015).

In the context of time series analysis we can consider additive outliers (AO), which are external errors or exogenous changes at a certain time point and affect only the observation at the time the disturbance occurs, and innovational outliers (IO), associated with internal changes or endogenous effects of the noise process that affect all the subsequent observations. Silva & Pereira (2015) suggested a Bayesian approach in order to detect additive outliers in Poisson INAR(1) models.

In this work we propose a method based on wavelets, which are a family of basis functions used to localize a given function in both space and scaling (Percival 2006), to address the problem of identifying the time point of the outlier in a time series of counts. The empirical distribution of the detailed coefficient derived from the discrete wavelet transform (DWT), using the Haar wavelet, is obtained by the parametric resampling

method of Tsay (1992). The outliers are identified as the observations outside of the acceptance envelope constructed with quantiles of this empirical distribution.

The proposed procedure is illustrated with simulated data and the results are compared with existing techniques. Furthermore, the method is also applied on an observed dataset.

Acknowledgments: This work is partially supported by Portuguese funds through the CIDMA and the Portuguese Foundation for Science and Technology ("FCT–Fundação para a Ciência e a Tecnologia"), within project UID/MAT/04106/2013.

References

Scotto, M. G., Weiß, C. H. & Gouveia, S. (2015) Thinning-based models in the analysis of integer-valued time series: a review. *Statistical Modelling*, 15, 590-618.

Percival, D., Walden, A. (2006) *Wavelet methods for time series analysis*. Cambridge Series in Statistical and Probabilistic Mathematics. New York, USA: Cambridge University Press.

Silva, M. E. & Pereira, I. (2015) Detection of additive outliers in Poisson INAR(1) time series. In Bourguignon, J. P., Jeltsch, R., Pinto, A. A. & Viana, M. (Eds.) *CIM Series in Mathematical Sciences - Mathematics of Energy and Climate Change*. Springer, 377-388.

Tsay, R. S. (1992) Model checking via parametric bootstraps in time series analysis. *Journal of the Royal Statistical Society: Series C* (*Applied Statistics*), 41, 1-15.