Outliers detection in integer-valued time series

Isabel Silva[1,3] and Maria Eduarda Silva[2,3]

Faculdade de Engenharia, Universidade do Porto[1]
Faculdade de Economia, Universidade do Porto[2]
CIDMA[3]

E-mail address: *ims@fe.up.pt; mesilva@fep.up.pt*

Abstract

The presence of outliers or discrepant observations has a negative impact in time series modelling. In this paper, we consider first-order integer-valued autoregressive (INAR(1)) models contaminated with additive and innovational outliers. We propose two methods based on wavelets in order to address the problem of identifying the times of outlier occurrence. The effectiveness of the proposed methods is illustrated and compared through a simulation study. A real dataset application is also presented.

Keywords: discrete wavelet transform, integer-valued autoregressive model, outlier detection, parametric resamplig, threshold.

——⊙——

A problem of interest in time series modelling is detecting outliers to assess data quality, to study the robustness of the statistical inference in the presence of inconsistent observations and to find eventual important phenomena underlying the data. It is well known that the presence of outliers in time series can lead to biased parameter estimation, incorrect inference and inadequate predictions. Several methodologies for detecting and estimating outliers have been established for ARMA models. The emphasis is on iterative procedures and likelihood based statistics [8, 2, 3]. However, the problem of detection and estimation of outliers in time series of counts has received less attention in the literature. Recently, [7] suggested a Bayesian approach in order to detect additive outliers in Poisson first-order INteger-valued AutoRegressive, INAR(1), models. The class of INAR models for time series of counts, first proposed by [1], has been extensively studied in the literature and applied to many real-world problems because of its easiness of interpretation. These models are apparently autoregressive models in which the usual multiplication has been replaced by a random operation, called thinning operation (for details see [6]).

In this work we propose to identify the times of outlier occurrence in INAR(1) time series using wavelets. Wavelets are a family of basis functions

used to localize a given function in both space and scaling [5]. In particular, we propose to apply discrete wavelet transform (DWT), which is a powerful tool for a time-scale multi-resolution analysis. DWT can be considered as filters of different cut-off frequencies used to analyse the signal at different scales. The high-scale (low-frequency) components of the signal are contained in the approximation coefficients. On the other hand, the low-scale (high-frequency) components are represented by the detail coefficients.

In a first approach, similar to that of [4], the detail coefficients derived from DWT, using the Haar wavelet, are compared with a threshold. In a second approach, the parametric resampling method of [9] is used in order to obtain the empirical distribution of these detail coefficients. The proposed procedures are illustrated and compared with synthetic data. Furthermore, the methods are also applied on an observed dataset.

Acknowledgements

References

[1] Al-Osh, M. A. and Alzaid, A. A. (1987) First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis* **8**: 261–275.

[2] Chang, I., Tiao, G. C. and Chen, C. (1988) Estimation of time series parameters in the presence of outliers. *Technometrics* **30**: 193–204.

[3] Chen, C. and Liu, L. M. (1993) Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association* **88**: 284–297.

[4] Grané, A. and Veiga, H. (2010) Wavelet-based detection of outliers in financial time series. *Computational Statistics & Data Analysis* **54**: 2580–2593.

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

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

[7] Silva, M. E. and Pereira, I. (2015) Detection of additive outliers in Poisson INAR(1) time series. In Bourguignon, J. P. *et al.* (Eds.) *CIM Series in Mathematical Sciences - Mathematics of Energy and Climate Change*, Springer: 377–388.

[8] Tsay, R. S. (1986) Time series model specification in the presence of outliers. *Journal of the American Statistical Association* **81**: 132–141.

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