Diagnostic tools for bivariate time series of counts

Isabel Silva[1], Maria Eduarda Silva[2], Cristina Torres [3]

[1]*Faculdade de Engenharia da Universidade do Porto, ims@fe.up.pt;*

[2] *Faculdade de Economia da Universidade do Porto and CIDMA, mesilva@fep.up.pt;*

[3] *ISCAP-IPP and Universidade do Porto, cmptorres@gmail.com*

Abstract

An important step in any statistical investigation is the assessment of the adequacy of the model proposed and fitted to the data under analysis. In this work, methods for model diagnostic and validation based on residual analysis, predictive distributions and parametric resampling methods are presented and illustrated with a real dataset modelled by a Bivariate INteger-valued Moving Average (BINMA(1, 1)) model.

Keywords: BINMA models, Model diagnostic, Parametric bootstrap, Predictive distributions, Residual analysis.

Introduction

The evaluation of the quality of a model fit is an important part of any statistical data analysis. For linear models there exist a wide variety of methods for this purpose. However, for discrete-valued time series, and particularly for count time series, this is not the case. As noted by Jung *et al.* (2015), some of the existing model assessment methods for continuous models can be adapted to the discrete-valued framework. In this work, the following tools will be considered: residual analysis through the serial correlation of (standardized) Pearson residuals; parametric bootstrap or resampling methods (Tsay, 1992) and predictive distributions by using a nonrandomized probability integral transform, PIT, histogram (Czado *et al.* 2009).

Results

These procedures will be illustrated on a bivariate dataset consisting of the number of rainy days per week at Bremen and Cuxhaven, which will be denoted by $(X_{1,t})$ and $(X_{2,t})$, $t=1,...,n$, respectively (n=574 observations, from the year 2000 to 2010), see Scotto *et al.* (2014) for details. A preliminary analysis of the sample mean, variance and autocorrelation function (ACF) indicates that this dataset can be modelled by a BINMA(1, 1) model with Bivariate Poisson distribution for the innovation process, as defined in Silva *et al.* (2014):

$$X_{1,t} = \varepsilon_{1,t} + \beta_{1,1} \circ \varepsilon_{1,t-1}; \qquad X_{2,t} = \varepsilon_{2,t} + \beta_{2,1} \circ \varepsilon_{2,t-1},$$

where $(\varepsilon_t) = (\varepsilon_{1,t}, \varepsilon_{2,t})$ follow a Bivariate Poisson distribution, $BP(\lambda_1, \lambda_2, \phi)$, and $\circ$ denotes the

binomial thinking operation ($\alpha \circ Y = \sum_{j=0}^{Y} B_j$, where Y is a non-negative random variable, $\alpha \in [0,1]$ and $B_j \sim \text{Ber}(\alpha)$, independent of Y). Generalized Method of Moment estimates and their corresponding standard errors (in brackets) are $\hat{\beta}_{1,1} = 0.010(0.054)$, $\hat{\lambda}_1 = 0.970(0.326)$, $\hat{\beta}_{2,1} = 0.161(0.100)$, $\hat{\lambda}_2 = 0.596(0.393)$ and $\hat{\phi} = 2.799(0.155)$.

Although the Pearson residuals exhibit the expected sample mean and variance (zero and one, respectively), there is evidence of serial correlation in some lags. Moreover, the parametric resampling exercise for the ACF also indicates that the fitted BP BINMA(1,1) model does not represent completely the dependence structure in the data. Additionally, the PIT histograms present an U-form, indicating underdispersed distribution. Thus, a BINMA model with a different order for each equation besides a different bivariate distribution for the innovation process (for instance bivariate negative binomial distribution) must be considered in order to improve the model fitting.

Conclusions

The diagnostic and validation of a fitted model must be performed on any time series analysis. In this work, different tools suggest that the BP BINMA(1, 1) do not completely capture the dependence structure of the data. However, as stated by Box: "*All models are wrong, but some are useful.*"

References

CZADO, C., GNEITING, T. & HELD, L. (2009) Predictive model assessment for count data. *Biometrics*, 65, 1254-1261.

JUNG, R.C., McCABE, B.P.M & TREMAYNE, A.R. (2015) Model validation and diagnostics. IN *Handbook of Discrete-Valued Time Series*. Chapman & Hall/CRC Press (forthcoming).

SCOTTO, M., WEIß, C., SILVA, M.E. & PEREIRA, I. (2014) Bivariate binomial autoregressive models. *Journal of Multivariate Analysis,* 125, 233 - 251.

SILVA, I., TORRES, C. & SILVA, M.E. (2014) Estimating bivariate integer-valued moving average models with the generalized method of moments. IN *Livro de resumos das XXI Jornadas de Classificação e Análise de Dados* (*JOCLAD 2014*), 111-114.

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.