Educational Administration: Theory and Practice
2024, 30(5), 5283-5286
ISSN: 2148-2403
https://kuey.net/ **Research Article**

Modelling Educational Data Series using a Novel High-Ordered BINAR Model

Yuvraj Sunecher[1]*, Naushad Mamode khan[2]

[1]*University of Technology Mauritius, Email: ysunecher@utm.ac.mu
[2]University of Mauritius, Email: n.mamodekhan@uom.ac.mu

***Corresponding Author:** Yuvraj Sunecher
*University of Technology Mauritius, Email: ysunecher@utm.ac.mu

Citation: Yuvraj Sunecher, Naushad Mamode khan (2024), Modelling Educational Data Series using a Novel High-Ordered BINAR Model, *Educational Administration: Theory and Practice*, *30*(5), 5283-5286
Doi: 10.53555/kuey.v30i5.3777

ARTICLE INFO

ABSTRACT

This paper treats the modelling of a bivariate integer-valued autoregressive of order p (BINAR(p)) model under non-stationary moment conditions where the inter-relation between the series is induced by pair of related Poisson innovations. This novel model is more convenient to model time series data on Education. The model parameters constituting of the mean or regression effects and the correlation effects are estimated via a generalized quasi-likelihood (GQL) equation since the joint and marginal distribution of the counting series under the non-stationary conditions is difficult to specify. The proposed model is used to analyse a series of educational data.

Keywords: Poisson, BINAR(p), GQL, Time Series Data, Education, Indiscipline.

I. INTRODUCTION

In literature, several extensions of the classical integer-valued autoregressive of order 1 (INAR(1)) model (Pedeli and Karlis (2011), Mamode Khan et al. (2016)) have been made. This paper focusses on extending the simple INAR(1) to the bivariate INAR(1) (BINAR(1)) model where the cross-correlation is induced by the related pair of innovations. In this context, Pedeli and Karlis (2011) proposed the first BINAR(1) model under the bivariate Poisson innovations but this BI- NAR(1) model was restricted to stationary condition only. Mamode Khan et al. (2016) developed a non-stationary BINAR(1) model under same distributional in-novation assumptions with same cross structure design but additionally assumed that the link predictor function of the counting series is influenced by time-variant covariates that ultimately induces non-stationary correlations. In these two major model development, the research focus was on extending the McKenzie's (1986) INAR(1) to the BINAR(1) model.
On the other hand from an application perspective, several time series models have been used in literature to model time series data on education (Vanitha and Jayashree (2022), Bloom (1999), Yang and Cheng (2020)). However, these models cannot be used to model time series data which exhibit non-stationary and different level of dispersion. Hence, this paper considers a further extension to a bivariate integer-valued autoregressive of p^{th} order (BINAR(p)) under some cross-relation with paired Poisson Innovations.

II. MODEL DEVELOPMENT

The BINAR(p) model is presented below:

$$Y_t^{[1]} = \sum_{i=1}^{p} \rho_{11}^{[i]} o\, Y_{t-i}^{[1]} + R_t^{[1]} \quad (1)$$

$$Y_t^{[2]} = \sum_{i=1}^{p} \rho_{22}^{[i]} o\, Y_{t-i}^{[2]} + R_t^{[2]} \quad (2)$$

Equations (1) and (2) may be combined in vector form as follows:

$$Y_t = A^{(i)} o\, Y_{t-1} + R_t \qquad (3)$$

Where $A^{(i)}$ represent a (2x2) matrix of autocorrelation coefficients for k is an element of {1,2} and the superscript (i) indicates the autocorrelation coefficients at lag-*i*. The model (3) is governed by some important assumptions:

(a)"o" indicates the binomial thinning operator (Steutel and Van, 1979) such that

$$\{\rho_{kk} o\, Y_{t-i}^{[k]}\} = \sum_{l=1}^{Y_{t-i}^{[k]}} b_l(\rho_{kk}),\ Y_{t-i}^{[k]} > 0,$$
$$= 0, \qquad Y_{t-i}^{[k]} = 0.$$

(b) The pair of innovations follows the bivariate Poisson distribution (Kocherlakota and Kocherlakota, 2001), such that the correlation between them is $k_{12,t}$.

$$\mu_t^{[k]} = \Lambda_t^{[k]} + \sum_{i=1}^{p} \rho_{11}^{[i]}\, \mu_{t-i}^{[k]}, \qquad (4)$$

$$Var(Y_t^{[k]}) = \Lambda_t^{[k]} + \sum_{i=1}^{p} [\rho_{kk}^{[i]}(1-\rho_{kk}^{[i]})\, \mu_{t-i}^{[k]} + (\rho_{kk}^{[i]})^2\, Var(Y_{t-i}^{[k]})]$$

$$+ \sum_{j=1}^{p-1} \sum_{m=j+1}^{p} \rho_{kk}^{[j]}\, \rho_{kk}^{[m]}\, Cov(Y_{t-j,}^{[k]}, Y_{t-m,}^{[k]}) \qquad (5)$$

$$Cov(Y_{t,}^{[k]}\, Y_{t+h}^{[k]}) = \sum_{i=1}^{p} \rho_{kk}^{[i]}\, Cov(Y_t^{[k]}, Y_{t+h-1,}^{[k]}) \qquad (6)$$

And

$$Cov(Y_{t,}^{[1]}\, Y_t^{[2]}) = k_{12,t} \sqrt{\Lambda_t^{[1]}} \sqrt{\Lambda_t^{[2]}} + \sum_{j=1}^{p} \sum_{i=1}^{p} \rho_{11}^{[j]}\, \rho_{22}^{[i]}\, Cov(Y_{t-j}^{[1]}, Y_{t-i,}^{[2]}) \qquad (7)$$

Under the special case of p=1,

$$\mu_t^{[k]} = \Lambda_t^{[k]} + \sum_{i=1}^{p} \rho_{11}^{[i]}\, \mu_{t-i}^{[k]}, \qquad (8)$$

$$Var(Y_t^{[k]}) = \Lambda_t^{[k]} + \rho_{kk}^{[1]} \left(1 - \rho_{kk}^{[1]}\right) \mu_{t-i}^{[k]} + (\rho_{kk}^{[1]})^2\, Var(Y_{t-i}^{[k]}) \qquad (9)$$

$$Cov(Y_{t,}^{[1]}\, Y_{t}^{[2]}) = \rho_{11}^{[1]}\, \rho_{22}^{[1]} Cov(Y_{t-1}^{[1]}, Y_{t-1,}^{[2]}) + k_{12,t} \sqrt{\Lambda_t^{[1]}} \sqrt{\Lambda_t^{[2]}} \quad (10)$$

III. ESTIMATION METHODOLOGY: PARAMETER ESTIMATION UNDER P=1

From the special case demonstrated in Section 2 and under the assumption

$$\Lambda_t^{[k]} = \exp\left[\sum_{j=1}^{q} [\beta_j^{[k]}\ x_{ij}]\right.$$

Where x is the j[th] effect covariate with β the corresponding regression coefficients, the parameters of the model

$$\theta = [\beta_1^{[k]}, \beta_2^{[k]}, \ldots, \beta_q^{[k]}, \rho_{11}^{[1]}, \rho_{22}^{[1]}\ ]$$

are estimated via a generalized quasi-likelihood (GQL) approach (Mamode Khan et al., 2016). In fact, it is shown in Mamode Khan et al. (2016) that the GQL approach yields asymptotically equally efficient estimates as likelihood-based approach with significantly lesser non-convergent simulations. The GQL equation is specified as:

$$D \sum\ (f - \mu) = 0 \quad (11)$$

Under p=1, the derivative entries with respect to the model parameters are obtained iteratively as follows:

$$\frac{\delta \mu_t^{[k]}}{\delta \beta_j^{[k]}} = \frac{\Lambda_t^{[k]} x_{tj} x}{1-\rho_{kk}^{[1]}} \quad t=1$$

$$= \rho_{kk}^{[1]} \frac{\delta \mu_{t-1}^{[k]}}{\delta \beta_j^{[k]}}\ \Lambda_t^{[k]}\ x_{tj} \quad t=2,\ldots,T \quad (12)$$

$$\frac{\delta \mu_t^{[k]}}{\delta \rho_{kk}^{[1]}} = \frac{\Lambda_t^{[k]}}{(1-\rho_{kk}^{[1]})} \quad t=1$$

$$\rho_{kk}^{[1]} \frac{\delta \mu_{t-1}^{[k]}}{\delta \rho_{kk}^{[1]}} + \mu_{t-1}^{[k]} \quad t=2,\ldots,T. \quad (13)$$

The entries of the covariance structure Σ are computed iteratively using Equations (9)-(10). The cross-correlation κ12,t is estimated using Equation (10), that is,

$$k_{12,t} = \frac{Cov(Y_{t,}^{[1]}\, Y_{t}^{[2]}) - \rho_{11}^{[1]}\, \rho_{22}^{[1]} Cov(Y_{t-1}^{[1]}, Y_{t-1,}^{[2]})}{\sqrt{\Lambda_t^{[1]}} \sqrt{\Lambda_t^{[2]}}} \quad (14)$$

IV. APPLICATION

The grading of computer and Maths for students of Grade 9 have been collected. The series were collected from secondary schools in Mauritius for the year 2023 and totaled 355 observations. The grading for

computer and Maths are as follows: Grading 1 – 85-100 marks, Grading 2 – 70-85 marks, Grading 3 – 60-70 marks, Grading 4 – 50-60 marks, Grading 5 – 40-50 marks and Grading 6 – 0-40 marks. The following table includes some illustrative figures for the grading:

Descriptive Data	**Computer**	**Maths**
Mean	3.5776	3.5481
Variance	2.1455	2.0559

Table 1: Descriptive data for the grading of Computer and Maths.

The attendance of students (A), their IQ, motivation level (ML) and social support (SS) were taken into account as covariates. Therefore, the grading of students for computer and maths series are subjected to the BINAR(2) model, with the following results:

Series	A	IQ	ML	SS
Computer ($Y_t^{[1]}$)	0.4881 (0.005) (0.0155)		0.5213 (0.002)	0.2692 (0.073)
Maths ($Y_t^{[2]}$)	0.5101 (0.007) (0.0139)	(0.003)	0.5218 (0.063)	0.3163

Table 2: Estimates of Regression Coefficients

Based on the findings in Table 2, we can see that all the covariates significantly impact the grading of the Grade 9 students for computer and Maths in Mauritius.

Conclusion

This paper introduces a non-stationary BINAR(p) model with correlated paired Poisson innovations. From the marginal moment expression, such BINAR(p) model is suited to analyze some over-dispersed series. Under the special case p = 1, the model reduces to a simple BINAR(1) model wherein the parameter estimation via the GQL approach can be handled. However, for a general pth order, solving for the unknown parameters via the GQL entails some numerical problems. A robust estimating algorithm for the above parameters have been established and a time series data on education has been modelled using the proposed high- ordered BINAR model.

REFERENCES

[1] Bloom H.S (1999). MDRC Working Papers on Research Methodology Estimating Program Impacts on Student Achievement Using ”Short” Interrupted Time Series

[2] Kocherlakota, S. and Kocherlakota, K. (2001). Regression in the Bivariate Pois- son Distribution. *Communications in Statistics-Theory and Methods*, 30(5), 815 – 825.

[3] Mamode Khan, N.A., Sunecher, Y., and Jowaheer, V. (2016). Modelling a Non- Stationary BINAR(1) Poisson Process. *Journal of Statistical Computation and Simulation*, 86, 3106 – 3126.

[4] McKenzie, E. (1986). Autoregressive moving average processes with Negative Bi- nomial and geometric marginal distributions. *Advanced Applied Probability*, 18, 679 – 705.

[5] Pedeli, X. and Karlis, D. (2011) A bivariate INAR(1) process with application.

[6] Statistical Modelling: An International Journal, 11, 325 – 349.

[7] Steutel F.W. and Van Harn K. (1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability*, 7, 3893 – 3899.

[8] Vanitha S. and Jayashree R. (2022). A Prediction on Educational Time Series Data Using Statistical Machine Learning Model -An Experimental Analysis *Journal of Theoretical and Applied Information Technology*, 100,14.

[9] Yang S. and Chen H.C. (2020). Student Enrollment and Teacher Statistics Fore- casting Based on Time-Series Analysis. *https://doi.org/10.1155/2020/1246920*.