Journal of Multivariate Analysis 125 (2014) 233–251

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

# Bivariate binomial autoregressive models

Manuel G. Scotto[a,*], Christian H. Weiß[b], Maria Eduarda Silva[c], Isabel Pereira[a]

[a] CIDMA - Center for Research and Development in Mathematics and Applications, Department of Mathematics, University of Aveiro, Portugal

[b] Department of Mathematics and Statistics, Helmut Schmidt University Hamburg, Germany

[c] CIDMA - Center for Research and Development in Mathematics and Applications, Faculty of Economics, University of Porto, Portugal

ARTICLE INFO

*Article history:*
Received 31 May 2013
Available online 13 January 2014

*AMS subject classifications:*
60G10
62M10
60H99

*Keywords:*
Bivariate binomial distribution
Binomial AR(1) model
INAR(1) model
INGARCH model
Thinning operation

ABSTRACT

This paper introduces new classes of bivariate time series models being useful to fit count data time series with a finite range of counts. Motivation comes mainly from the comparison of schemes for monitoring tourism demand, stock data, production and environmental processes. All models are based on the bivariate binomial distribution of Type II. First, a new family of bivariate integer-valued GARCH models is proposed. Then, a new bivariate thinning operation is introduced and explained in detail. The new thinning operation has a number of advantages including the fact that marginally it behaves as the usual binomial thinning operation and also that allows for both positive and negative cross-correlations. Based upon this new thinning operation, a bivariate extension of the binomial autoregressive model of order one is introduced. Basic probabilistic and statistical properties of the model are discussed. Parameter estimation and forecasting are also covered. The performance of these models is illustrated through an empirical application to a set of rainy days time series collected from 2000 up to 2010 in the German cities of Bremen and Cuxhaven.



## 1. Introduction

Time series of (low) counts play an important role in the analysis of data sets ranging from economy and finance [20,9,13] to medicine [27,34,3,1] and biology [45]. It is worth to mention that a large part of the literature on this topic is devoted to the analysis of time series having an infinite range of counts. In particular, INteger-valued AutoRegressive-type (INAR) models based on the binomial thinning operation of Steutel and van Harn [36], defined as $\alpha \circ X := Y_1 + \cdots + Y_X$ if $X > 0$, and 0 otherwise, where the $Y_i$'s are independent and identically distributed (i.i.d.) Bernoulli random variables with success probability $\alpha \in (0; 1)$, play a central role. For example, the INAR model of order one [26] is defined by the recursion

$$X_t = \alpha \circ X_{t-1} + \varepsilon_t \equiv \sum_{i=1}^{X_{t-1}} Y_{t,i} + \varepsilon_t, \quad t \in \mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}, \tag{1}$$

where $(\varepsilon_t)$ is an i.i.d. process with range $\mathbb{N}_0 = \{0, 1, \ldots\}$, and where all thinning operations are performed independently of each other and of $(\varepsilon_t)$. Furthermore, the thinning operations at each time $t$ and $\varepsilon_t$ are independent of $(X_s)_{s<t}$. Note that the thinning operation ensures the integer discreteness of the process. More general INAR processes of order $p > 1$ were introduced by Alzaid and Al-Osh [2].

* Correspondence to: Department of Mathematics, University of Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal.
*E-mail address:* mscotto@ua.pt (M.G. Scotto).

0047-259X/$ – see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmva.2013.12.014

A different approach to handle time series of counts is to consider Generalized AutoRegressive Conditional Heteroscedastic (GARCH) models, where the autoregressive structure is incorporated via a link function. A commonly used model is the INteger-valued GARCH (INGARCH) process of order $(p, q)$ of Heinen [17], defined as

$$\begin{cases} X_t|\mathscr{F}_{t-1} : P(\lambda_t); \quad \forall t \in \mathbb{Z} \\ \lambda_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j}, \end{cases} \tag{2}$$

where $\mathscr{F}_{t-1} := \sigma(X_s,\ s \le t-1)$, $\alpha_0 > 0$, $\alpha_i \ge 0$, and $\beta_j > 0$. Ferland et al. [12] showed that $(X_t)$ is strictly stationary with finite first- and second-order moments provided that $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$. Weiß [39] derived the variance and autocorrelation function for the INGARCH models with $p, q > 1$. Further properties have been obtained by Zhu and Wang [47,46]. The particular case $p = q = 1$ was analyzed by Fokianos and Tjøstheim [14] and Fokianos et al. [13] under the designation of Poisson Autoregression. We refer the reader to the survey of Tjøstheim [37] and the references therein for further details.

In contrast, however, the analysis of integer-valued time series with a finite range of counts has not received much attention in the literature. The origins of the use of models based on thinning operations applied to time series with a finite range of counts, say $\{0, 1, \ldots, n\}$, can be traced back to McKenzie [26] who gave a remarkable contribution by suggesting to replace the INAR(1) recursion in (1) by

$$X_t = \alpha \circ X_{t-1} + \beta \circ (n - X_{t-1}), \quad t \in \mathbb{Z} \tag{3}$$

with $\alpha = \beta + \rho$, $\beta = \pi(1-\rho)$ for $\pi \in (0; 1)$ and $\rho \in [\max(-\pi/(1-\pi), -(1-\pi)/\pi);\ 1]$, where all thinnings are performed independently of each other, and being the thinnings at time $t$ independent of $(X_s)_{s<t}$. Note that the representation for $X_t$ in (3) guarantees that the range of $X_t$ is given by $\{0, 1, \ldots, n\}$. Furthermore, the condition on $\rho$ guarantees that $\alpha, \beta \in (0; 1)$. The process in (3) used to be referred to as *binomial* AR(1) *process* and is a stationary Markov chain with $n+1$ states and binomial marginal distribution $\mathrm{Bi}(n, \pi)$. The binomial AR(1) process shares some properties with the conventional AR(1) process, namely $\rho(k) := \rho(X_t, X_{t-k}) = \rho^k$, where $\rho(Y, Z)$ abbreviates the correlation between $Y$ and $Z$. Other important features of the binomial AR(1) process are that both the conditional mean and variance of $X_t$ given $X_{t-1}$ are linear in $X_{t-1}$, and the fact that is time-reversible. For further properties see [41,43,11,40]. For binomial AR($p$) processes with order $p > 1$ see [38]. Further enhancements of the basic binomial AR(1) model are proposed by Weiß and Kim [42] and Weiß and Pollett [44].

The literature on bivariate (and also multivariate) time series with finite or infinite range of counts is still in its infancy. There have been only few attempts to model bivariate/multivariate time series of counts via multivariate INGARCH models. A notable exception is the work of Heinen and Rengifo [18] who introduced the multivariate autoregressive conditional double Poisson model generalizing previous results by Heinen [17] for the univariate case. Another generalization being based on the bivariate Poisson distribution is considered by Liu [24]. Also multivariate models based upon thinning ideas have received little attention in the literature. An important contribution was made by Franke and Rao [15] who introduced the multivariate integer-valued autoregressive (MINAR, in short) model of order one based upon independent binomial thinning operations. Extensions of MINAR models with order $p > 1$ were introduced in [23] in which matrices operate on vectors using the generalized thinning operation. More recently, Pedeli and Karlis [29] introduced the bivariate INAR model of order one with Poisson and negative binomial innovations. The authors illustrated the performance of the model through an empirical application to the joint modeling of the number of daytime and nighttime road accidents in the Netherlands for the year 2001. It is important to refer that in Pedeli and Karlis' model the autoregression matrix is diagonal which means that it causes no cross-correlation in the counts; see also [30,31] for further details. The bivariate INAR model considered by Boudreault and Charpentier [7], in contrast, is the one of Franke and Rao [15], and therefore accounts for cross-correlation in the counts. An important limitation of Pedeli and Karlis' model and also Boudreault and Charpentier's model is that they only allow for positive correlations between the two time series. In order to also account for negative correlation between the time series, Karlis and Pedeli [21] introduced a family of bivariate INAR(1) processes where negative cross-correlation is introduced through the innovations, by defining the distribution of the innovations in terms of appropriate bivariate copulas. Extensions for bivariate INAR(1) models with positively correlated geometric marginals can be found in [33]. Bivariate INMA models based on the binomial thinning operation and contemporaneous only cross-correlation in the counts was proposed by Quoreshi [32] who reports an application to the number of transactions in intra-day data of stocks.

Applied to the bivariate case with $\boldsymbol{X} := [X_1\ X_2]'$, the thinning concept of Franke and Rao [15] and Boudreault and Charpentier [7] leads to the operation

$$\mathbf{A} \circ \boldsymbol{X} = \begin{bmatrix} a_{11} \circ X_1 + a_{12} \circ X_2 \\ a_{21} \circ X_1 + a_{22} \circ X_2 \end{bmatrix} \quad \text{with } \mathbf{A} \in [0; 1]^{2\times 2}, \tag{4}$$

where the thinnings are performed independently of each other. Karlis and Pedeli [21] and Pedeli and Karlis [30,31,29] restrict to the case where $a_{12} = a_{21} = 0$ such that $(\mathbf{A} \circ \boldsymbol{X})_i$ has the same distribution as $a_{ii} \circ X_i$, i.e., the marginals behave like the univariate thinning operation. However, this nice feature is obtained at the cost of no additional cross-correlation between $(\mathbf{A} \circ \boldsymbol{X})_1$ and $(\mathbf{A} \circ \boldsymbol{X})_2$, in the sense that

$$\mathrm{Cov}\big((\mathbf{A} \circ \boldsymbol{X})_1,\ (\mathbf{A} \circ \boldsymbol{X})_2\big) = \mathrm{Cov}\big(\mathrm{E}(a_{11} \circ X_1 \mid \boldsymbol{X}),\ \mathrm{E}(a_{22} \circ X_2 \mid \boldsymbol{X})\big) = a_{11}a_{22} \cdot \mathrm{Cov}(X_1, X_2). \tag{5}$$