Multivariate INAR processes - Periodic case

Cláudia Santos[a,b], Isabel Pereira[b], and Manuel Scotto[c]

[a]Coimbra College of Agriculture, Polytechnic Institute of Coimbra
[b]CIDMA, University of Aveiro
[c]IST & CEMAT, University of Lisbon
{csps@ua.pt}

Abstract. *A multivariate integer-valued autoregressive model of order one with periodic time-varying parameters, and driven by a periodic innovations sequence of independent random vectors is established. The binomial thinning operator replaces the scalar multiplication in the common time series models. The matricial form of the multivariate model and its basic statistical properties are defined. Emphasis is placed upon models with periodic multivariate negative binomial innovations. Aiming to reduce computational burden arising from the use of the conditional maximum likelihood method a composite likelihood-based approach is adopted and compared with other traditional competitors.*

Keywords: Multivariate models, binomial thinning operator, composite likelihood

1 Introduction

Multivariate count data can occur in many fields. Special attention has been devoted to bivariate integer-valued time series processes (e.g. Pedeli and Karlis [6], [8]) based on the binomial thinning operator introduced by Steutal and van Harn ([9]). The role of the innovations is significant since not only they determine the joint distribution of the two series under consideration but also they form the unique source of cross-correlation. The assumption of a bivariate negative binomial distribution for the innovations allows for more flexibility than the Poisson `BINAR`(1) model due to the involvement of the overdispersion parameter.

Within the reasonably large spectrum of integer-valued models proposed in the literature only a few focus on the modeling of multivariate time series of count data with periodic structure. Our interest in periodic integer-valued autoregressive models was primarily influenced by the work of Monteiro *et al.* whose periodic univariate and bivariate `INAR` models were introduced in [4] and in [5], respectively. We seek to extend `INAR` models to multi-dimensional space, assuming periodic time-varying parameters and periodic sequences of innovations. Apart from the general specification of such models, we also examined their statistical properties and proposed alternative estimation techniques.

2 Periodic MINAR model of order one

Let $\{\mathbf{X}_t\}$ be a periodic m-variate integer-valued autoregressive process of first-order defined by the recursion

$$\mathbf{X}_t = A_t \circ \mathbf{X}_{t-1} + \mathbf{Z}_t,\ t \in \mathbb{Z}, \tag{1}$$

where $\mathbf{X}_t, \mathbf{X}_{t-1}$ and $\mathbf{Z}_t$ are random ms-vectors with $\mathbf{X}_t = [\mathbf{X}_{1,t}\ \mathbf{X}_{2,t}\ \cdots\ \mathbf{X}_{m,t}]'$ for $t = v + ns$, $v = 1, \ldots, s$ and $n \in \mathbb{N}_0$, and $\mathbf{X}_{j,t} = [X_{j,1+ns}\ \cdots\ X_{j,s+ns}]'$, $j = 1, \ldots, m$. The ms-dimensional vector $\mathbf{Z}_t = [\mathbf{Z}_{1,t}\ \mathbf{Z}_{2,t}\ \cdots\ \mathbf{Z}_{m,t}]'$ constitutes a periodic sequence of independent random vectors with

$$\mathbf{Z}_{j,t} = [Z_{j,1+ns}\ Z_{j,2+ns}\ \cdots\ Z_{j,s+ns}]'. \tag{2}$$

The model in (1) will be referred to as the `P`eriodic `M`ultivariate `IN`teger-valued `A`uto`R`egressive model of order one (`PMINAR`(1) in short) with period $s \in \mathbb{N}$. The `PMINAR`(1) model admits the following matricial representation

$$\begin{bmatrix} \mathbf{X}_{1,t} \\ \mathbf{X}_{2,t} \\ \vdots \\ \mathbf{X}_{m,t} \end{bmatrix} = \begin{bmatrix} \phi_{1,t} & 0 & \cdots & 0 \\ 0 & \phi_{2,t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \phi_{m,t} \end{bmatrix} \circ \begin{bmatrix} \mathbf{X}_{1,t-1} \\ \mathbf{X}_{2,t-1} \\ \vdots \\ \mathbf{X}_{m,t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_{1,t} \\ \mathbf{Z}_{2,t} \\ \vdots \\ \mathbf{Z}_{m,t} \end{bmatrix}, \tag{3}$$

where matrix A_t in (1) is a $(ms \times ms)$-diagonal matrix, representing the periodic integer-valued autoregressive coefficients in season v and $\phi_{j,t} = \alpha_{j,v} \in (0,1)$ for $t = v + ns; v = 1, \ldots, s; n \in \mathbb{N}_0$ and $j = 1, \ldots, m$.

For each t, $\mathbf{Z}_{j,t}$ is assumed to be independent of $\mathbf{X}_{j,t-1}$ and $\phi_{j,t} \circ \mathbf{X}_{j,t-1}$. Note that the j-th component in (3) is

$$\mathbf{X}_{j,t} = \phi_{j,t} \circ \mathbf{X}_{j,t-1} + \mathbf{Z}_{j,t}, \tag{4}$$

with $\phi_{j,t} \circ X_{j,t-1} \overset{d}{=} \sum_{r=1}^{X_{j,t-1}} U_{r,t}(\phi_{j,t})$, where $\{U_{r,t}(\phi_{j,t})\}_{r\in\mathbb{N}}$ is a periodic sequence of i.i.d. Bernoulli-distributed random variables with probability of success $P(U_{r,t}(\phi_{j,t}) = 1) = \phi_{j,t}$. Since the autocorrelation matrix A_t is diagonal, the only source of dependence between the series $\mathbf{X}_{1,t}, \ldots, \mathbf{X}_{m,t}$ in (3) is given through $\mathbf{Z}_t$. Therefore, the innovations will play a central role in the specification of the `PMINAR`(1) process.

Due to the fact that $t = v + ns$, then $\mathbf{X}_{j,t-s} \overset{d}{=} \mathbf{X}_{j,v+(n-1)s}$ $(v = 1, \ldots, s)$, meaning that the j-th component $\mathbf{X}_{j,t}$ in (4) can be expressed as

$$\mathbf{X}_{j,t} = A_j \circ \mathbf{X}_{j,t-s} + B_j \circ \mathbf{Z}_{j,t}, \tag{5}$$

where the $(s \times s)$-matrices A_j and B_j $(j = 1, \ldots, m)$ are given by

$$A_j = \begin{bmatrix} 0 \cdots 0 & \alpha_{j,1} \\ 0 \cdots 0 & \alpha_{j,2}\alpha_{j,1} \\ 0 \cdots 0 & \alpha_{j,3}\alpha_{j,2}\alpha_{j,1} \\ \vdots\ \vdots\ \ddots & \vdots \\ 0 \cdots 0 & \prod_{k=0}^{s-1} \alpha_{j,s-k} \end{bmatrix} \tag{6}$$

and

$$B_j = \begin{bmatrix} 1 & 0 & 0 & \dots 0 \\ \alpha_{j,2} & 1 & 0 & \dots 0 \\ \alpha_{j,3}\alpha_{j,2} & \alpha_{j,3} & 1 & \dots 0 \\ \vdots & \vdots & \vdots & \ddots \vdots \\ \prod_{k=0}^{s-2} \alpha_{j,s-k} & \prod_{k=0}^{s-3} \alpha_{j,s-k} & \prod_{k=0}^{s-4} \alpha_{j,s-k} & \dots 1 \end{bmatrix}, \tag{7}$$

respectively, with coefficients $\alpha_{j,v} \in (0,1)$, $j = 1, \dots, m$ and $v = 1, \dots, s$. Taking all m components, the PMINAR(1) model defined in (1) can be rewritten in the form

$$\mathbf{X}_t = \widetilde{A} \circ \mathbf{X}_{t-s} + \widetilde{B} \circ \mathbf{Z}_t, \tag{8}$$

The $(ms \times ms)$-matrices $\widetilde{A}$ and $\widetilde{B}$ in equation (8) are block-diagonal matrices, that is

$$\widetilde{A} = diag(A_1, A_2, \dots, A_m) \tag{9}$$

and

$$\widetilde{B} = diag(B_1, B_2, \dots, B_m) \tag{10}$$

with matrices A_j and B_j $(j = 1, \dots, m)$ as in (6) and in (7), respectively. Generally, matrix $\widetilde{A}$ has entries a_{ik}^j satisfying $0 \le a_{ik}^j < 1$ and matrix $\widetilde{B}$ has entries b_{ik}^j satisfying $0 \le b_{ik}^j \le 1$ with $i, k = 1, \dots, ms$ and $j = 1, \dots, m$. Furthermore, it will be assumed that the innovations $\mathbf{Z}_t$ have finite first- and second-order moments taking the form

$$E[\mathbf{Z}_t] = E \begin{bmatrix} \mathbf{Z}_{1,t} \\ \mathbf{Z}_{2,t} \\ \vdots \\ \mathbf{Z}_{m,t} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\delta}_{1,t} \\ \boldsymbol{\delta}_{2,t} \\ \vdots \\ \boldsymbol{\delta}_{m,t} \end{bmatrix} =: \boldsymbol{\delta}_t. \tag{11}$$

The ms-mean vector $\boldsymbol{\delta}_t$ with $t = v + ns; v = 1, \dots, s$ and $n \in \mathbb{N}_0$ has m $(s \times 1)$-vectors, i.e.,

$$E[\mathbf{Z}_{j,t}] = \boldsymbol{\delta}_{j,t} = \left[\lambda_{j,1} \ \lambda_{j,2} \ \dots \ \lambda_{j,s} \right]', \tag{12}$$

for $j = 1, \dots, m$. For a fixed v, each element in (12) is

$$E[Z_{j,v+ns}] = \lambda_{j,v}. \tag{13}$$

Turning to the variance-covariance matrix of $\mathbf{Z}_t$, it follows that

$$\sum\nolimits_{\mathbf{Z}_t} = \begin{bmatrix} Var[\mathbf{Z}_{1,t}] & Cov(\mathbf{Z}_{1,t}, \mathbf{Z}_{2,t}) & \cdots & Cov(\mathbf{Z}_{1,t}, \mathbf{Z}_{m,t}) \\ & Var[\mathbf{Z}_{2,t}] & \cdots & Cov(\mathbf{Z}_{2,t}, \mathbf{Z}_{m,t}) \\ & & \ddots & \vdots \\ & & & Var[\mathbf{Z}_{m,t}] \end{bmatrix} = \begin{bmatrix} \psi_{11,t} & \psi_{12,t} & \dots & \psi_{1m,t} \\ & \psi_{22,t} & \dots & \psi_{2m,t} \\ & & \ddots & \vdots \\ & & & \psi_{mm,t} \end{bmatrix} =: \psi_t, \tag{14}$$

where $\psi_{jk,t}$ $(j, k = 1, \dots, m; t = v + ns; v = 1, \dots, s; n \in \mathbb{N}_0)$ are $(s \times s)$- diagonal matrices of the form

$$\psi_{jk,t} = diag(\sigma_{jk,1}, \dots, \sigma_{jk,s}), \tag{15}$$

and for a fixed v, each element of the diagonal in matrix (15) is given by

$$\sigma_{jk,v} = Cov(Z_{j,v+ns}, Z_{k,v+ns}). \tag{16}$$

2.1 Strictly periodically stationary distribution

Let PMINAR(1) be the process in (8). The existence of a periodically stationary solution to (8) depends on the largest eigenvalue of the non-negative matrix $\widetilde{A}$ in (9), whose coefficients $\alpha_{j,v} \in (0,1)$ for all components.

Lemma 1 For a fixed v $(v = 1, \ldots, s)$, $\alpha_{j,v} \in (0,1)$ where $j = 1, \ldots, m$ and for $t = v + ns$, $0 < P(\mathbf{Z}_t = \underline{0}) < 1$. Furthermore, any solution to $\{\mathbf{X}_t\}$, $t = v + ns$ and $n \in \mathbb{N}_0$ in (8) is an irreducible and aperiodic Markov chain.

The main result of this subsection is formalized through the theorem below.

Theorem 1 For a fixed v $(v = 1, \ldots, s)$, let $\{\mathbf{X}_t\}$ with $t = v + ns$ and $n \in \mathbb{N}_0$ as in (8) be an irreducible, aperiodic Markov chain on $\mathbb{N}_0^m$. If $E||\mathbf{Z}_t|| < +\infty$ and if the largest eigenvalue of $\widetilde{A}$ is less than one, then there exists a strictly periodically stationary (or cyclostationary) m-variate INAR(1) process satisfying recursion (8).

Proof. From Lemma 1, $\{\mathbf{X}_t\}$ with $t = v + ns$ and fixed $v = 1, \ldots, s$ is an irreducible and aperiodic Markov chain being the eigenvalues of matrix $\widetilde{A}$ are less than one. Thus by Franke and Subba Rao ([2]) a strictly periodically stationary m-variate non-negative integer-valued process satisfying the equation (8) exists.

The PMINAR(1) model in (8) can be expressed as

$$\mathbf{X}_t = \widetilde{A} \circ \mathbf{X}_{t-s} + \mathbf{R}_t, \tag{17}$$

where $\mathbf{R}_t = \widetilde{B} \circ \mathbf{Z}_t$ with matrix $\widetilde{B}$ in (10).

2.2 Mean vector of cyclostationary PMINAR(1)

In this subsection we obtain the periodic mean and autocovariance function of the PMINAR(1) model. First note that the expectation of $\mathbf{R}_t$ is

$$E[\mathbf{R}_t] = E[\widetilde{B} \circ \mathbf{Z}_t] = \widetilde{B}E[\mathbf{Z}_t] = \widetilde{B}\boldsymbol{\delta}_t \tag{18}$$

with matrices $\widetilde{B}$ and $\boldsymbol{\delta}_t$ as in (10) and (11), respectively. Furthermore, for each component $j = 1, \ldots, m$, the mean vector of $\mathbf{R}_{j,t}$ takes the form

$$E[\mathbf{R}_{j,t}] = \begin{bmatrix} \lambda_{j,1} \\ \lambda_{j,1}\alpha_{j,2} + \lambda_{j,2} \\ \lambda_{j,1}\alpha_{j,3}\alpha_{j,2} + \lambda_{j,2}\alpha_{j,3} + \lambda_{j,3} \\ \vdots \\ \lambda_{j,1}\prod_{k=0}^{s-2}\alpha_{j,s-k} + \lambda_{j,2}\prod_{k=0}^{s-3}\alpha_{j,s-k} + \cdots + \lambda_{j,s-1}\alpha_{j,s} + \lambda_{j,s} \end{bmatrix}.$$

Moreover,

$$\boldsymbol{\mu}_t = E[\mathbf{X}_t] = E[\widetilde{A} \circ \mathbf{X}_{t-s} + \mathbf{R}_t] = (I - \widetilde{A})^{-1}\widetilde{B}\boldsymbol{\delta}_t \tag{19}$$

with matrices $\widetilde{A}$ and $\widetilde{B}$, and vector $\boldsymbol{\delta}_t$ as in (9), (10) and (11), respectively. Note that the ms-dimensional mean vector $\boldsymbol{\mu}_t$ for $t = v+ns;\ v = 1, \ldots, s$ and $n \in \mathbb{N}_0$ in (19) takes the form $\boldsymbol{\mu}_t = \left[\boldsymbol{\mu}_{1,t}\ \boldsymbol{\mu}_{2,t} \ldots \boldsymbol{\mu}_{m,t}\right]'$.

For each $j = 1, \ldots, m$ and $l \geq i$, let

$$\varphi_{l,i}^{(j)} = \begin{cases} \prod_{k=0}^{i-1} \alpha_{j,l-k}, & i \geq 1 \\ 1, & i = 0 \end{cases}. \tag{20}$$

It follows by tedious (although straightforward) calculations that for a fixed v and j, each entry in $\boldsymbol{\mu}_{j,t} = [E(X_{j,1+ns})\ E(X_{j,2+ns}) \cdots E(X_{j,s+ns})]'$ is given by

$$E(X_{j,v+ns}) = \frac{\sum_{k=0}^{v-1} \varphi_{v,k}^{(j)} \lambda_{j,v-k} + \varphi_{v,v}^{(j)} \sum_{i=0}^{s-(v+1)} \varphi_{s,i}^{(j)} \lambda_{j,s-i}}{1 - \varphi_{s,s}^{(j)}}, \tag{21}$$

for $j = 1, \ldots, m;\ t = v + ns;\ v = 1, \ldots, s$ and $n \in \mathbb{N}_0$. We adopt the convention $\sum_{i=0}^{s-(s+1)} \varphi_{s,i}^{(j)} \lambda_{j,s-i} = 0$.

2.3 Variance-covariance matrix

To derive the variance-covariance matrix of $\{\mathbf{X}_t\}$ we start by calculation the variance-covariance matrix, $\sum_{\mathbf{R}_t}$, of $\{\mathbf{R}_t\}$. To this extent note that,

$$\sum\nolimits_{\mathbf{R}_t} = \widetilde{B} \sum\nolimits_{\mathbf{Z}_t} \widetilde{B}' + diag(Q\boldsymbol{\delta}_t) = \widetilde{B}\psi_t\widetilde{B}' + diag(Q\boldsymbol{\delta}_t)$$

with matrices $\widetilde{B}$, $\boldsymbol{\delta}_t$ and ψ_t in (10), (11) and (14), respectively. Moreover the $(ms \times ms)$-matrix $Q = \widetilde{B}(I - \widetilde{B})$ has entries $[q_{ik}^j] = [b_{ik}^j(1 - b_{ik}^j)]_{i,k=1,\ldots,ms}$ for $j = 1, \ldots, m$, being b_{ik}^j the elements of matrix $\widetilde{B}$ in (10). Thus, matrix Q is also block-diagonal with m $(s \times s)$-matrices Q_j.

For simplicity in notation we define

$$\sum\nolimits_{\mathbf{X}_t} =: \begin{bmatrix} \sum_{1,1} & \sum_{1,2} & \cdots & \sum_{1,m} \\ & \sum_{2,2} & \cdots & \sum_{2,m} \\ & & \ddots & \vdots \\ & & & \sum_{m,m} \end{bmatrix}.$$

Note that

$$\sum\nolimits_{j,j} = \begin{bmatrix} Var[X_{j,1+ns}] & Cov(X_{j,1+ns}, X_{j,2+ns}) & \ldots & Cov(X_{j,1+ns}, X_{j,s+ns}) \\ & Var[X_{j,2+ns}] & \ldots & Cov(X_{j,2+ns}, X_{j,s+ns}) \\ & & \ddots & \vdots \\ & & & Var[X_{j,s+ns}] \end{bmatrix}$$

with diagonal elements

$$Var[X_{j,v+ns}] = \frac{g(\varphi,\lambda)}{1-\left(\varphi_{s,s}^{(j)}\right)^2}, \tag{22}$$

where

$$\begin{aligned} g(\varphi,\lambda) := & \sum_{k=0}^{v-1} \varphi_{s,s}^{(j)}\varphi_{v,k}^{(j)}\lambda_{j,v-k} + \varphi_{v,k}^{(j)}\left(1-\varphi_{v,k}^{(j)}\right)\lambda_{j,v-k} + \left(\varphi_{v,k}^{(j)}\right)^2\sigma_{j,v-k}^2 + \\ & + \sum_{m=0}^{s-(v+1)} \varphi_{s,s}^{(j)}\varphi_{v,v}^{(j)}\varphi_{s,m}^{(j)}\lambda_{j,s-m} + \varphi_{v,v}^{(j)}\varphi_{s,m}^{(j)}\left(1-\varphi_{v,v}^{(j)}\varphi_{s,m}^{(j)}\right)\lambda_{j,s-m} + \\ & + \sum_{m=0}^{s-(v+1)} \left(\varphi_{v,v}^{(j)}\varphi_{s,m}^{(j)}\right)^2\sigma_{j,s-k}^2, \end{aligned}$$

for a fixed v $(v = 1,\ldots,s)$ and off-diagonal elements

$$Cov(X_{j,v+ns}, X_{j,v+ns+l}) = \varphi_{v+l,l}^{(j)} Var[X_{j,v+ns}]. \tag{23}$$

2.4 PMINAR(1) process with MVNB Innovations

In this subsection we derive the first- and second-order moment structure of the `PMINAR`(1) process driven by periodic multivariate negative binomial (MVBN) innovations. Hence, the joint probability mass function

$$\begin{aligned} & h(z_1,\ldots,z_m) = P(Z_{1,v+ns} = z_1,\ldots,Z_{m,v+ns} = z_m) = \\ & = \frac{\Gamma\left(\beta_v^{-1} + \sum_{j=1}^m z_j\right)}{\Gamma(\beta_v^{-1})}\left(\frac{\beta_v^{-1}}{\beta_v^{-1}+\sum_{j=1}^m \lambda_{j,v}}\right)^{\beta_v^{-1}}\left(\beta_v^{-1} + \sum_{j=1}^m \lambda_{j,v}\right)^{-\sum_{j=1}^m z_j} \cdot \\ & \cdot \prod_{j=1}^m \frac{\lambda_{j,v}^{z_j}}{z_j!},\ (z_1,\ldots,z_m) \in \mathbb{N}_0^m. \end{aligned} \tag{24}$$

Notice the marginal distribution of $\mathbf{Z}_{j,t}$ is univariate negative binomial with parameters β_v^{-1} and $p_{j,v}$ $(j = 1,\ldots,m; v = 1,\ldots,s)$ with

$$p_{j,v} = \frac{\beta_v^{-1}}{\lambda_{j,v} + \beta_v^{-1}}. \tag{25}$$

The innovation process $\{\mathbf{Z}_t\}$, $t = v+ns;\ v = 1,\ldots,s$ and $n \in \mathbb{N}_0$ is generally defined as a periodic sequence of independent random vectors with mean as in (11) and variance-covariance matrix as in (14), respectively. Thus,

$$\lambda_{j,v} = E[Z_{j,v+ns}] = \beta_v^{-1}\ \frac{1-p_{j,v}}{p_{j,v}}, \tag{26}$$

$$\sigma_{j,v}^2 = Var[Z_{j,v+ns}] = \beta_v^{-1}\ \frac{1-p_{j,v}}{p_{j,v}^2} = \lambda_{j,v}(1+\beta_v\lambda_{j,v}), \tag{27}$$

$$\sigma_{jk,v} = Cov(Z_{j,v+ns}, Z_{k,v+ns}) = \beta_v\lambda_{j,v}\lambda_{k,v}, \tag{28}$$

for a fixed v ($v = 1, \ldots, s$), $j \neq k$; $j, k = 1, \ldots, m$. Note that $Var[Z_{j,v+ns}] > E[Z_{j,v+ns}]$. Thus, the first-order moment and the auto- and cross-covariance structure PMINAR(1) process are obtained from (21), (22) and (23) by plugging-in the values of $\lambda_{j,v}$, $\sigma^2_{j,v}$ and $\sigma_{jk,v}$ as in (26)-(28).

3 Parameter estimation

Consider a finite time series $\{\mathbf{X}_{j,ts}\}$ with $1 \leq t \leq N$, $j = 1, \ldots, m$ (N-number of complete cycles) from the PMINAR(1) model in (17) with MVNB innovations. Without loss of generality it will be assumed that $\mathbf{X}_0 = \mathbf{x}_0$. The vector of unknown parameters $\boldsymbol{\theta}$ is a $(2m+1)s$-dimensional vector

$$\boldsymbol{\theta} := (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m, \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_m, \boldsymbol{\beta}) \tag{29}$$

with s-vectors ($j = 1, \ldots, m$):

$$\boldsymbol{\alpha}_j = (\alpha_{j,1}, \ldots, \alpha_{j,s}) \ ; \ \boldsymbol{\lambda}_j = (\lambda_{j,1}, \ldots, \lambda_{j,s}) \ ; \ \boldsymbol{\beta} = (\beta_1, \ldots, \beta_s). \tag{30}$$

In order to estimate the unknown parameters in $\boldsymbol{\theta}$, three estimation methods are proposed, namely: Yule-Walker, conditional maximum likelihood and composite likelihood.

3.1 Yule-Walker estimation

The YW estimator of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}_{YW} := (\widehat{\boldsymbol{\alpha}}_1^{YW}, \ldots, \widehat{\boldsymbol{\alpha}}_m^{YW}, \widehat{\boldsymbol{\lambda}}_1^{YW}, \ldots, \widehat{\boldsymbol{\lambda}}_m^{YW}, \widehat{\boldsymbol{\beta}}^{YW})$ are calculated as follows: first, the YW estimators of parameters $\boldsymbol{\lambda}_j$ are calculated through the solution of the system of s linear equations yielding

$$\widehat{\lambda}_{j,v}^{YW} = \begin{cases} \overline{X}_{j,v} - \widehat{\alpha}_{j,v}^{YW} \overline{X}_{j,s} & , v = 1 \\ \overline{X}_{j,v} - \widehat{\alpha}_{j,v}^{YW} \overline{X}_{j,v-1} & , v = 2, 3, \ldots, s \end{cases}, \tag{31}$$

where $\overline{X}_{j,v} = \frac{1}{N} \sum_{n=0}^{N-1} X_{j,v+ns}$, $j = 1, \ldots, m$, is the sample mean. The YW estimators of parameters $\boldsymbol{\alpha}_j$ are given by

$$\widehat{\alpha}_{j,v}^{YW} = \begin{cases} \frac{S^2_{j,v}}{S^2_{j,s}} & , v = 1 \\ \frac{\gamma_{j,v-1}(1)}{S^2_{j,v-1}} & , v = 2, 3, \ldots, s \end{cases}, \tag{32}$$

where $S^2_{j,v}$ is the sample variance and $\gamma_{j,v}(1)$ ($j = 1, \ldots, m$) is defined as

$$\gamma_{j,v}(1) = Cov(X_{j,v+ns}, X_{j,v+1+ns}) =$$
$$= \begin{cases} \frac{1}{N-1} \sum_{n=0}^{N-1} (X_{j,v+ns} - \overline{X}_{j,v})(X_{j,v+1+ns} - \overline{X}_{j,v+1}) & , v = 1, \ldots, s-1 \\ \frac{1}{N-1} \sum_{n=0}^{N-1} (X_{j,v+ns} - \overline{X}_{j,v})(X_{j,1+(n+1)s} - \overline{X}^*_{j,1}) & , v = s \end{cases},$$

with $\overline{X}^*_{j,1} = \frac{1}{N}\sum_{n=0}^{N} X_{j,1+ns}$. Finally, the YW estimator of β_v is

$$\widehat{\beta}_v^{YW} = \frac{\left(1 - \widehat{\varphi}_{s,s}^{(j)}\widehat{\varphi}_{s,s}^{(k)}\right)\gamma_{jk,v}(0)}{\sum_{i=0}^{v-1}\widehat{\varphi}_{v,i}^{(j)}\widehat{\varphi}_{v,i}^{(k)}\widehat{\lambda}_{j,v-i}\widehat{\lambda}_{k,v-i} + \widehat{\varphi}_{v,v}^{(j)}\widehat{\varphi}_{v,v}^{(k)}\sum_{i=0}^{s-(v+1)}\widehat{\varphi}_{s,i}^{(j)}\widehat{\varphi}_{s,i}^{(k)}\widehat{\lambda}_{j,s-i}\widehat{\lambda}_{k,s-i}}, \quad (33)$$

for $v = 1, \ldots, s$ and $j, k = 1, \ldots, m$ $(j \neq k)$.

3.2 Conditional maximum likelihood estimation

In order to obtain the CML estimator $\widehat{\boldsymbol{\theta}}_{CML}$ of $\boldsymbol{\theta}$ we proceed as follows. First note that the transition probabilities for the `PMINAR`(1) model can be expressed as the convolution of m binomials with parameters $(x_{j,v-1+ns}, \alpha_{j,v})$ for $v = 1, \ldots, s$ with probability mass function

$$f_j(r_j) = C_{r_j}^{x_{j,v-1+ns}}\alpha_{j,v}^{r_j}(1-\alpha_{j,v})^{x_{j,v-1+ns}-r_j},\ j = 1, \ldots, m, \quad (34)$$

and the periodic discrete m-variate distribution defined as in (24). Thus, the conditional density is the multiple sum

$$p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns}) = P(\mathbf{X}_{v+ns} = \mathbf{x}_{v+ns}|\mathbf{X}_{v-1+ns} = \mathbf{x}_{v-1+ns}) =$$
$$= \sum_{r_1=0}^{g_1}\sum_{r_2=0}^{g_2}\cdots\sum_{r_m=0}^{g_m}\left(\prod_{j=1}^{m} f_j(r_j)\right) h(x_{1,v+ns} - r_1, \ldots, x_{m,v+ns} - r_m) \quad (35)$$

with $g_j := min(x_{j,v-1+ns}, x_{j,v+ns})$. Hence, the CML estimator is obtained by maximizing the conditional log-likelihood $C(\boldsymbol{\theta}) = ln(L(\boldsymbol{\theta}|\mathbf{x}))$. Explicit CML estimators are not available so numerical procedures have to be employed. The asymptotic properties of $\widehat{\boldsymbol{\theta}}_{CML}$ are given through the following result.

Theorem 2 The conditional maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_{CML}$ of $\boldsymbol{\theta}$ is asymptotically normal

$$\sqrt{N}(\widehat{\boldsymbol{\theta}}_{CML} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, I^{-1}(\boldsymbol{\theta}))$$

where $I(\boldsymbol{\theta})$ represents the Fisher information matrix, i.e., $I = diag(M_1, \ldots, M_s)$ with matrices M_v $(v = 1, \ldots, s)$ given by

$$(-1)\times\begin{bmatrix}
E\left[\frac{\partial^2 C(\theta)}{\partial\alpha_{1,v}^2}\right] & \cdots E\left[\frac{\partial^2 C(\theta)}{\partial\alpha_{1,v}\partial\alpha_{m,v}}\right] & E\left[\frac{\partial^2 C(\theta)}{\partial\alpha_{1,v}\partial\lambda_{1,v}}\right] & \cdots E\left[\frac{\partial^2 C(\theta)}{\partial\alpha_{1,v}\partial\lambda_{m,v}}\right] & E\left[\frac{\partial^2 C(\theta)}{\partial\alpha_{1,v}\partial\beta_v}\right] \\
\vdots & \ddots\vdots & \vdots & \vdots\ \vdots & \vdots \\
E\left[\frac{\partial^2 C(\theta)}{\partial\alpha_{m,v}\partial\alpha_{1,v}}\right] & \cdots E\left[\frac{\partial^2 C(\theta)}{\partial\alpha_{m,v}^2}\right] & E\left[\frac{\partial^2 C(\theta)}{\partial\alpha_{m,v}\partial\lambda_{1,v}}\right] & \cdots E\left[\frac{\partial^2 C(\theta)}{\partial\alpha_{m,v}\partial\lambda_{m,v}}\right] & E\left[\frac{\partial^2 C(\theta)}{\partial\alpha_{m,v}\partial\beta_v}\right] \\
E\left[\frac{\partial^2 C(\theta)}{\partial\lambda_{1,v}\partial\alpha_{1,v}}\right] & \cdots E\left[\frac{\partial^2 C(\theta)}{\partial\lambda_{1,v}\partial\alpha_{m,v}}\right] & E\left[\frac{\partial^2 C(\theta)}{\partial\lambda_{1,v}^2}\right] & \cdots E\left[\frac{\partial^2 C(\theta)}{\partial\lambda_{1,v}\partial\lambda_{m,v}}\right] & E\left[\frac{\partial^2 C(\theta)}{\partial\lambda_{1,v}\partial\beta_v}\right] \\
\vdots & \vdots\ \vdots & \vdots & \ddots\vdots & \vdots \\
E\left[\frac{\partial^2 C(\theta)}{\partial\lambda_{m,v}\partial\alpha_{1,v}}\right] & \cdots E\left[\frac{\partial^2 C(\theta)}{\partial\lambda_{m,v}\partial\alpha_{m,v}}\right] & E\left[\frac{\partial^2 C(\theta)}{\partial\lambda_{m,v}\partial\lambda_{1,v}}\right] & \cdots E\left[\frac{\partial^2 C(\theta)}{\partial\lambda_{m,v}^2}\right] & E\left[\frac{\partial^2 C(\theta)}{\partial\lambda_{m,v}\partial\beta_v}\right] \\
E\left[\frac{\partial^2 C(\theta)}{\partial\beta_v\partial\alpha_{1,v}}\right] & \cdots E\left[\frac{\partial^2 C(\theta)}{\partial\beta_v\partial\alpha_{m,v}}\right] & E\left[\frac{\partial^2 C(\theta)}{\partial\beta_v\partial\lambda_{1,v}}\right] & \cdots E\left[\frac{\partial^2 C(\theta)}{\partial\beta_v\partial\lambda_{m,v}}\right] & E\left[\frac{\partial^2 C(\theta)}{\partial\beta_v^2}\right]
\end{bmatrix}.$$

Proof. This result is a particular case of Theorem 2.2 in [1]. For each v ($v = 1, \ldots, s$), $p_v(\cdot|\cdot)$ is the transition probabilities in (35) of the PMINAR(1) model, therefore the regularity conditions in Billingsley's Theorem 2.2 are satisfied.

3.3 Composite likelihood estimation

For periodic multivariate processes, the number of parameters can be quite large due to season v ($v = 1, \ldots, s$) with s representing the period. Computational issues often arise when applying the conditional maximum likelihood approach, the complexity of the method augments with dimensional increase. Composite likelihood inherits many of the good properties of inference based on the full likelihood function, but is more easily implemented with high-dimensional data sets. These methods based on optimizing sums of log-likelihoods of low-dimensional margins have become popular in recent years (e.g. [7]); being useful for multivariate models in which the likelihood of multivariate data is very time-consuming. The methodology has drawn considerable attention in a broad range of applied disciplines in which complex data structures arise (e.g. [10]).

Note that the bivariate marginal log-likelihood function between two random elements, say X_a and X_b, can be defined as

$$l_{ab}(\boldsymbol{\theta}; \mathbf{x}_a, \mathbf{x}_b) = \frac{1}{Ns} \sum_{n=0}^{N-1} \sum_{v=1}^{s} log f_{X_a,X_b}(x_{a,v+ns}, x_{b,v+ns} | x_{a,v-1+ns}, x_{b,v-1+ns}; \boldsymbol{\theta}), \tag{36}$$

where

$$\begin{aligned} & f_{X_a,X_b}(x_{a,v+ns}, x_{b,v+ns} | x_{a,v-1+ns}, x_{b,v-1+ns}; \boldsymbol{\theta}) = \\ & = \sum_{k_a=0}^{g_1} \sum_{k_b=0}^{g_2} \binom{x_{a,v-1+ns}}{x_{a,v+ns} - k_a} \alpha_{a,v}^{x_{a,v+ns}-k_a} (1-\alpha_{a,v})^{x_{a,v-1+ns}-x_{a,v+ns}+k_a} \cdot \\ & \cdot \binom{x_{b,v-1+ns}}{x_{b,v+ns} - k_a} \alpha_{b,v}^{x_{b,v+ns}-k_b} (1-\alpha_{b,v})^{x_{b,v-1+ns}-x_{b,v+ns}+k_b} \cdot h_{R_a,R_b}(k_a, k_b) \end{aligned}$$

for $g_1 = min(x_{a,v+ns}, x_{a,v-1+ns})$ and $g_2 = min(x_{b,v+ns}, x_{b,v-1+ns})$. The bivariate function $h_{R_a,R_b}(k_a, k_b)$ represents the bivariate marginal probability density with bivariate negative binomial innovations being a particular case ($m = 2$) of the multivariate negative binomial distribution in (24).

The composite log-likelihood function, $cl(\boldsymbol{\theta}; \mathbf{x}_a, \mathbf{x}_b)$, then arises as the sum of all bivariate log-likelihood functions, i.e.,

$$cl(\boldsymbol{\theta}; \mathbf{x}_a, \mathbf{x}_b) = \sum_{a=1}^{m-1} \sum_{b=a+1}^{m} w_{ab}\ l_{ab}(\boldsymbol{\theta}; \mathbf{x}_a, \mathbf{x}_b), \tag{37}$$

where w_{ab} is a constant weight for l_{ab}. For sake of simplicity, it is common to set $w_{ab} = 1$, $1 \leq a \leq b \leq m$ (e.g. [3]).

4 Simulation study

The performance of the three estimation methods of the `PMINAR`(1) model driven by multivariate negative binomial innovations is compared through a simulation experiment for $m = 3$ (trivariate). We have set period $s = 4$ therefore $\boldsymbol{\theta} := (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3, \boldsymbol{\beta})$. This study contemplates the following set of parameters: $\boldsymbol{\alpha}_1 = (0.53, 0.75, 0.62, 0.83)$, $\boldsymbol{\alpha}_2 = (0.72, 0.85, 0.56, 0.91)$, $\boldsymbol{\alpha}_3 = (0.83, 0.60, 0.41, 0.58)$ and $\boldsymbol{\lambda}_1 = (4, 2, 3, 5)$, $\boldsymbol{\lambda}_2 = (5, 3, 1.2, 2)$, $\boldsymbol{\lambda}_3 = (3, 1.6, 2, 4)$ and $\boldsymbol{\beta} = (1.6, 0.9, 1.8, 1.2)$. Three alternative samples sizes where considered: $n = 400, 1000, 2000$. Thus $n = sN$, $N = 100, 250, 500$ complete cycles. For each experiment we conducted 200 independent replications.

Comparison of the YW, CML and CL estimators was made in terms of the mean square error and the biases of the produced estimates. Table 1 reports the estimates for autocorrelation parameters $\boldsymbol{\alpha}_j$ $(j = 1, 2, 3)$, where small MSE characterize all estimates of $(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3)$.

Table 1. YW, CML and CL estimates for $\boldsymbol{\alpha}_j = (\alpha_{j,1}, \alpha_{j,2}, \alpha_{j,3}, \alpha_{j,4})$ with $j = 1, 2, 3$. Mean square error in parenthesis.

	n =400			n =1000			n =2000		
	YW	CML	CL	YW	CML	CL	YW	CML	CL
$\hat{\alpha}_{1,1}$	0.521	0.531	0.531	0.528	0.531	0.531	0.528	0.529	0.529
	(0.0018)	(5.1×10^{-7})	(0.0002)	(0.0001)	(0.00002)	(0.00004)	(0.0001)	(0.00002)	(0.00003)
$\hat{\alpha}_{1,2}$	0.746	0.752	0.751	0.750	0.751	0.751	0.752	0.749	0.748
	(0.0001)	(0.00005)	(0.0008)	(0.0006)	(0.0009)	(0.0009)	(0.0001)	(0.00002)	(0.00003)
$\hat{\alpha}_{1,3}$	0.608	0.618	0.617	0.617	0.621	0.620	0.615	0.620	0.619
	(0.0004)	(0.0026)	(0.00002)	(0.0074)	(0.0011)	(0.0013)	(0.0005)	(0.00006)	(0.00006)
$\hat{\alpha}_{1,4}$	0.789	0.833	0.832	0.826	0.830	0.830	0.825	0.830	0.830
	(0.0111)	(0.0007)	(0.0006)	(0.0020)	(0.00002)	(0.00001)	(0.0011)	(0.00007)	(0.00006)
$\hat{\alpha}_{2,1}$	0.717	0.718	0.717	0.739	0.719	0.719	0.740	0.720	0.720
	(0.0027)	(0.00002)	(0.0038)	(0.0001)	(0.00003)	(0.00002)	(0.0010)	(7.5×10^{-6})	(0.00006)
$\hat{\alpha}_{2,2}$	0.845	0.854	0.852	0.845	0.851	0.851	0.849	0.851	0.850
	(0.0001)	(0.0008)	(0.00003)	(0.0002)	(0.00002)	(0.00002)	(0.0004)	(0.00003)	(0.00003)
$\hat{\alpha}_{2,3}$	0.552	0.559	0.560	0.559	0.559	0.559	0.561	0.560	0.560
	(0.0002)	(0.0025)	(0.00009)	(0.0006)	(0.00007)	(0.00007)	(0.0002)	(0.00004)	(0.00002)
$\hat{\alpha}_{2,4}$	0.894	0.910	0.910	0.910	0.910	0.911	0.906	0.909	0.910
	(0.0105)	(0.0001)	(0.0002)	(0.0003)	(0.0003)	(0.0003)	(0.0001)	(0.00001)	(0.00002)
$\hat{\alpha}_{3,1}$	0.823	0.832	0.832	0.832	0.831	0.831	0.830	0.830	0.830
	(0.0071)	(0.0005)	(0.0013)	(0.0001)	(0.00006)	(0.0001)	(0.0003)	(0.00001)	(0.00003)
$\hat{\alpha}_{3,2}$	0.596	0.603	0.603	0.601	0.600	0.600	0.599	0.601	0.602
	(0.0011)	(0.0008)	(0.0020)	(0.0020)	(0.00002)	(0.0004)	(0.0001)	(0.0003)	(0.0004)
$\hat{\alpha}_{3,3}$	0.391	0.411	0.410	0.411	0.409	0.409	0.407	0.411	0.411
	(0.0002)	(0.0001)	(0.0009)	(0.0004)	(0.0037)	(0.0040)	(0.0001)	(8.2×10^{-6})	(0.00002)
$\hat{\alpha}_{3,4}$	0.545	0.587	0.588	0.566	0.580	0.580	0.578	0.580	0.580
	(0.0002)	(0.0020)	(0.0046)	(0.0155)	(0.0003)	(0.0003)	(0.0028)	(0.0027)	(0.0028)

The tendency of the YW method to produce inadmissible estimates was greater for smaller sample sizes. YW estimates were used as initial values in numerical routines for the optimization procedure of CML and CL methods. The performance of the estimators $\hat{\boldsymbol{\lambda}}_j$ $(j = 1, 2, 3)$ and estimator $\hat{\boldsymbol{\beta}}$ (not shown here) is slightly worse. The estimates obtained by adopting either the CML or the CL method are very close to the real parameter values, even in the case of a moderate sample size ($n = 400$). For larger samples ($n = 1000, 2000$), both estimators seem to perform well and in a similar way.

Graphical inspection is provided through the boxplots of the biases of the produced estimates. Fig. 1 displays the boxplots of the biases of the estimates for $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)$, the dispersion parameter, where the effect of sample size on the behavior of the estimators can be seen. As expected, increasing the sample size improves the performance of all estimators in terms of both location (median closer to zero) and dispersion (narrower interquartile ranges). Small and not definite differences are observed between CML and CL methods, regarding both location and dispersion. Therefore, this indicates the superiority of CML and CL estimators over the YW estimator.

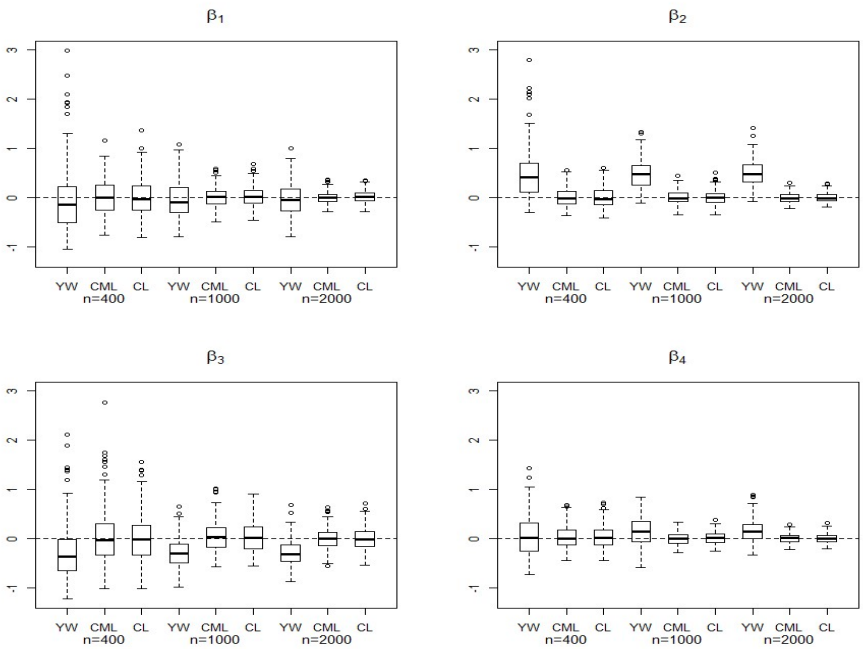

Fig. 1. Boxplots for the biases of the YW, CML and CL estimates of the parameter $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)$. From left to right, the first three boxplots display the biases of $\hat{\beta}_1$ for the three methods with $n = 400, 1000, 2000$. The same information follows for $\hat{\beta}_2$, $\hat{\beta}_3$ and $\hat{\beta}_4$, respectively.

Acknowledgments. Portuguese Foundation for Science and Technology (FCT-Fundacão para a Ciência e a Tecnologia) through CIDMA-Center for Research and Development in Mathematics and Applications (University of Aveiro), within project UID/MAT/04106/2013.

References

1. Billingsley, P.: Statistical Inference for Markov Processes. Statistical Research Monographs, University of Chicago Press (1961)
2. Franke, J., Subba Rao, T.: Multivariate first–order integer–valued autoregression, Technical report, Universität Kaiserslautern (1993)
3. Joe, H., Lee, Y.: On weighting of bivariate margins in pairwise likelihood. J. Mult. Anal. 100, 670–685 (2009)
4. Monteiro, M., Scotto, M.G., Pereira, I.: Integer–valued autoregressive processes with periodic structure. J. Stat. Plan. Inf. 140, 1529–1541 (2010)
5. Monteiro, M., Scotto, M.G., Pereira, I.: A periodic bivariate integer-valued autoregressive model. In: Bourguignon J.P., Jelstch R., Pinto A., Viana M. (eds). Dynamics, Games and Science - International Conference and Advanced School Planet Earth, DGS II, pp. 455–477. Springer International Publishing (2015)
6. Pedeli X., Karlis, D.: A bivariate INAR(1) process with application. Stat. Modelling 11, 325–349 (2011)
7. Pedeli, X., Karlis, D.: On composite likelihood estimation of a multivariate INAR(1) model. J. Time S. Anal. 34, 206–220 (2013a)
8. Pedeli, X., Karlis, D.: On estimation of the bivariate Poisson INAR process. Commun. in Stat: Simulation and Comput. 42, 514–533 (2013b)
9. Steutel, F.W., van Harn, K.: Discrete analogues of self–decomposability and stability. The Annals of Prob. 7, 893–899 (1979)
10. Varin, C., Reid, N.M., Firth, D.: An overview of composite likelihood methods. Stat. Sinica 21, 5–42 (2011)