Universidade de Aveiro Departamento de Matemática
2017

Cláudia Susana Pereira dos Santos

Análise Estatística de Séries de Contagem com Estrutura Periódica

Statistical Analysis of Count Time Series with Periodic Structure

Universidade de Aveiro Departamento de Matemática
Ano 2017

Cláudia Susana Pereira dos Santos

Análise Estatística de Séries de Contagem com Estrutura Periódica

Statistical Analysis of Count Time Series with Periodic Structure

Tese apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Matemática, realizada sob a orientação científica da Doutora Isabel Maria Simões Pereira, Professora Auxiliar do Departamento de Matemática da Universidade de Aveiro e do Doutor Manuel González Scotto, Professor Associado com Agregação, do Instituto Superior Técnico da Universidade de Lisboa.

To my beautiful triplets: Laura, Glória and Júlia.

O júri

Presidente: **Doutor Artur Manuel Soares da Silva**
Professor Catedrático da Universidade de Aveiro

Vogais: **Doutora Maria Esmeralda Elvas Gonçalves**
Professora Associada da Faculdade de Ciências e Tecnologia, Universidade de Coimbra

Doutora Maria Eduarda da Rocha Pinto Augusto da Silva
Professora Associada da Faculdade de Engenharia, Universidade do Porto

Doutora Patrícia Cortés de Zea Bermudez
Professora Auxiliar da Faculdade de Ciências, Universidade de Lisboa

Doutora Isabel Maria Simões Pereira
Professora Auxiliar da Universidade de Aveiro (Orientadora)

Doutora Clara Maria Henrique Cordeiro
Professora Auxiliar da Faculdade de Ciências e Tecnologia, Universidade do Algarve.

agradecimentos

First of all, I would like to express my gratitude to my academic advisors, Professor Isabel Pereira and Professor Manuel Scotto for their guidance and encouragement. I enjoyed our cooperation and appreciate their kind support.

I would also like to show my appreciation to the Department of Mathematics of the University of Aveiro for providing a spot for my research.

I am grateful to the Escola Superior Agrária de Coimbra for the concentrated schedule allowing me to pursue my PhD at the University of Aveiro.

I thank my friends, in particular Magda Monteiro, for their help and patience.

Last, but not least, I thank my parents Orlando and Natália, Joaquim and our lovely daughters for their support and faith in me.

palavras-chave

Distribuição Skellam, Modelo autoregressivo de valores inteiros, Operadores thinning, Verosimilhança composta.

resumo

Os modelos autoregressivos de valores inteiros multivariados (MINAR) desempenham um papel central na análise estatística de séries temporais de contagem. Dentro do razoavelmente grande espectro de modelos MINAR propostos na literatura, muito poucos focam a análise de séries de contagem com estrutura periódica. A análise dos processos de contagem multivariados apresenta muitos desafios que vão desde a especificação do modelo até à estimação de parâmetros. Esta tese tem como objetivo dar uma contribuição nessa direção. Especificamente, o objetivo deste trabalho é duplo: primeiro, introduzimos o processo multivariado periódico de ordem um, PMINAR(1). As propriedades probabilísticas e estatísticas do modelo são estudadas em detalhe. Para superar as dificuldades computacionais decorrentes da utilização do método da máxima verosimilhança introduzimos uma abordagem baseada na verosimilhança composta. O desempenho do método proposto e outros métodos concorrentes na estimação dos parâmetros é comparado através de um estudo de simulação. A previsão também é abordada. Uma aplicação de dados reais relacionados com a análise de fogos é apresentada. Em segundo lugar, propomos dois modelos INAR (univariado e bivariado) com estrutura periódica, S-PINAR(1) e BS-PINAR(1), respetivamente. Ambos os modelos são baseados no operador signed thinning permitindo contagens de valores positivos e negativos. Apresentamos as propriedades probabilísticas básicas e estatísticas dos modelos periódicos. As inovações são modeladas através das distribuições Skellam univariada e bivariada, respetivamente. Para avaliar o desempenho dos estimadores dos mínimos quadrados condicionais e da máxima verosimilhança condicional, foi realizado um estudo de simulação para o modelo S-PINAR(1).

keywords

Composite likelihood, Integer-valued autoregressive models, Skellam distrubution, Thinning operators.

abstract

Multivariate INteger–valued AutoRegressive (MINAR) processes play a central role in the statistical analysis of integer-valued time series. Within the reasonably large spectrum of MINAR models proposed in the literature, however, only a few focus on the analysis of time series of count data with periodic structure. The analysis of multivariate counting processes presents many challenging problems ranging from model specification to parameter estimation. This thesis aims at giving a contribution towards this direction. Specifically, the purpose of this research is two-fold: first, we introduce the periodic multivariate process of order one (PMINAR(1) in short). The probabilistic and also the statistical properties of the model are studied in detail. To overcome the computational difficulties arising from the use of the maximum likelihood method we introduce a composite likelihood-based approach. The performance of the proposed method and other competitors methods of estimation is compared through a simulation study. Forecasting is also addressed. An application to a real data set related with the analysis of fire activity is presented. Secondly, we propose two INAR (univariate and bivariate) models with periodic structure, S-PINAR(1) and BS-PINAR(1), respectively. Both models are based on the signed thinning operator allowing for positive and negative counts. We examine the basic probabilistic and also the statistical properties of the periodic models. Innovations are modeled by univariate and bivariate Skellam distributions, respectively. To study the performance of the conditional least squares and conditional maximum likelihood estimators, a simulation study is conducted for the S-PINAR(1) model.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Discrete-valued time series are common in many practical situations, often as counts of events or individuals in consecutive intervals or at consecutive points in time. The analysis of non-negative integer-valued time series has become an important area of research in the last decades partially because of its wide applicability, for example, in the fields of public health and medicine (Moriña et al., 2011; Fernández-Fontelo et al., 2016), road safety (Pedeli and Karlis, 2011), economics (Bourguignon, 2016), finance (Barreto-Souza and Bourguignon, 2015), criminology (Nastić and Ristić, 2012; Ilić, 2016) and environment (Pavlopoulos and Karlis, 2008), among others.

The class of linear models with finite variance plays a central role in the analysis of stationary time series. This class includes conventional $\texttt{ARMA}(p, q)$ models of the form

$$X_t = \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j Z_{t-j} + Z_t, \ t \in \mathbb{Z} \tag{1.1}$$

with α_i $(i = 1, \ldots, p)$ and β_j $(j = 1, \ldots, q)$ being constants, and $\{Z_t\}$ constitutes an independent identically distributed (i.i.d.) sequence of random variables. However, such models are unlikely to describe accurately time series of counts due to the discreteness of the process since the multiplication of an integer by a real number usually results in a non-integer value. Addressing this issue various models of discrete time series have been proposed in the literature. Discrete-valued stationary processes have been studied by Jacobs and Lewis (1978a,b, 1983). It was perhaps the first attempt to obtain a general class of simple models for discrete variate time series. These models, referred to as DARMA models, are structurally based on the well-known ARMA processes.

Among the most successful models for integer-valued data we mention the INARMA (INteger AutoRegressive Moving Average) models. INARMA models are the discrete counterparts of the conventional ARMA models, where the scalar multiplication is replaced by an appropriate thinning operator. To ensure the discrete nature of the variates is preserved, $\{Z_t\}$ is a sequence of integer-valued random variables (r.v.'s).

Several models dealing with the discreteness of the data have been proposed in the literature. These models are categorized as either observation-driven or parameter-driven, a nomenclature that is originally due to Cox (1981). In parameter–driven models the serial dependence is induced by a latent variable whose distribution does not depend on the past observations of the outcome variable. In contrast, observation–driven models induce serial dependence by specifying the state variable explicitly as a function of past observations. MacDonald and Zucchini (1997) and McKenzie (2003) provide an overview of the subject. Jung and Tremayne (2011) compare and contrast a variety of time series models for counts. More recently, Davis et al. (2016) address a plethora of diverse topics on modeling discrete-valued time series, and in particular time series of counts. Theoretical, methodological and practical issues are pursued therein.

In this work we will focus on observation-driven models that include models based on the thinning operators, where the multiplication in the common time series models is replaced by an appropriate thinning operator. The remainder of this chapter is organized as follows: the first two sections review univariate and multivariate time series models for count data. In each of the aforementioned sections, we have subdivided the section into three parts: one regarding binomial thinning-based INAR models, another regarding signed thinning-based INAR models and the last subsection covers other related INAR models. Periodic time series are described in a different section. Parameter estimation and forecasting issues are also addressed. At last, we present the outline of the thesis, stating the developed work.

1.1 Univariate time series models for count data - a review

Many models that have been built for count time series data are based on the Steutel and van Harn (1979) thinning operator. A survey on thinning operation for count data was provided by Weiß (2008). Recently, Scotto et al. (2015) reviewed the literature on relevant thinning-based models for the analysis of integer-valued time series with finite/infinite support.

1.1.1 Binomial thinning-based INAR models

The most popular thinning operator is the binomial thinning, introduced by Steutel and van Harn (1979) to adapt the terms of self-decomposability and stability for integer-valued time series.

Definition 1.1. *(Binomial thinning operator)*
Let $(Y_k)_{k\in\mathbb{N}}$ be a sequence of i.i.d. Bernoulli random variables with mean $\alpha \in [0,1]$, independent of X, a non-negative integer-valued random variable with range $\mathbb{N}_0$. The binomial thinning operator $\alpha\circ$ is given by

$$\alpha \circ X := \begin{cases} \sum\limits_{k=1}^{X} Y_k & , \ X > 0 \\ 0 & , \ X = 0 \end{cases} . \tag{1.2}$$

Some elementary properties of the binomial thinning operator, defined above, are summarized in Lemma 1.1. Further properties of the binomial thinning operator can be found in e.g. Silva and Oliveira (2004), Weiß (2008) and more recently, in Turkman et al. (2014).

Lemma 1.1. *(Properties of the binomial thinning operator)*
Let X and Y be two random variables with support in $\mathbb{N}_0$, and $\alpha, \beta \in [0,1]$.

1. $0 \circ X = 0$,

2. $1 \circ X = X$,

3. $\alpha \circ (\beta \circ X) \stackrel{d}{=} (\alpha\beta) \circ X$,

4. $\alpha \circ (X + Y) \stackrel{d}{=} \alpha \circ X + \alpha \circ Y$ *if the two counting sequences are independent,*

5. $E(\alpha \circ X) = \alpha E(X)$,

6. $Var(\alpha \circ X) = \alpha^2 Var(X) + \alpha(1-\alpha)E(X)$,

7. $Cov(\alpha \circ X, X) = \alpha Var(X)$.

Basic properties can easily be derived using known formulas for conditional mean and variance:

$$E[Y] = E_X(E[Y|X]) \quad \text{and} \quad Var[Y] = Var_X(E[Y|X]) + E_X(Var[Y|X]). \tag{1.3}$$

The INARMA(p, q) model (p and q both non-negative) has been defined, on the basis of binomial thinning operator of Steutel and van Harn (1979) in (1.2), through the recursion

$$X_t = \sum_{i=1}^{p} \alpha_i \circ X_{t-i} + \sum_{j=1}^{q} \beta_j \circ Z_{t-j} + Z_t, \; t \in \mathbb{Z}, \tag{1.4}$$

where $\{Z_t\}$ is an i.i.d. sequence of integer-valued r.v.'s with finite mean and variance. It is assumed that all thinning operators are performed independently of each other and of Z_t. The INARMA models directly imitate the classical ARMA recursion. The counterpart to the conventional AR model, in the context of INARMA models, is the INAR model, an important sub–class of the observation–driven models. Hence, when $q = 0$ in equation (1.4), $\{X_t\}$ is called an INAR of order p. If $p = 0$, $\{X_t\}$ is referred to as INteger–valued Moving Average of order q (INMA(q) for short). The INMA models are beyond the scope of this work.

The first-order non-negative integer-valued autoregressive (INAR(1)) process is a particular case of equation (1.4) for $p = 1$ and $q = 0$ and has received considerable attention. This model was introduced independently by McKenzie (1985) and Al-Osh and Alzaid (1987) as a tool for modeling and generating sequences of dependent counting processes. Many authors have studied INAR models extensively. This is partially due to the increasing availability of relevant data sets in various fields of applications (e.g. medicine and finance).

A non-negative integer-valued time series $\{X_t\}$ is said to follow an $\texttt{INAR}(1)$ model if it satisfies a difference equation of the form

$$X_t = \alpha \circ X_{t-1} + Z_t,\ \ t \in \mathbb{Z}, \tag{1.5}$$

where parameter $\alpha \in [0, 1]$ and $\{Z_t\}$ is a sequence of i.i.d. non–negative integer–valued r.v.'s. It is assumed that all thinning operators are performed independently of each other and of Z_t. The term Z_t is referred to as the innovation term and must be independent of $\alpha \circ X_{t-1}$, and follows any discrete distribution with support $\mathbb{N}_0$ (in order for X_t to be counts).
The realization of the process at time t is composed by two parts, the first one clearly relates to the previous observation, while the second one is independent and depends only on the current time point. One can easily see that the binomial thinning operator in equation (1.5) replaces the multiplication used for the standard $\texttt{AR}(1)$ models as to ensure that only integer values will occur. Thus, conditional on X, $\alpha \circ X$ is a binomial r.v., where X denotes the number of trials and α represents the probability of success in every trial. The condition $\alpha < 1$ is necessary and sufficient for equation (1.5) to admit a strictly stationary solution, whose marginal law is uniquely determined by the law of the innovations according to the $\texttt{INAR}(\infty)$ representation, $X_t \overset{d}{=} \sum_{j=0}^{\infty} \alpha^j \circ Z_{t-j}$ (Al-Osh and Alzaid, 1987). The conditions $\alpha = 0$ and $\alpha = 1$ imply independence and non-stationarity for $\{X_t\}$, respectively.
Let μ_Z and σ_Z^2 be the (assumed finite) mean and variance of the i.i.d. innovation Z_t, then the mean and variance of the stationary solution of $\texttt{INAR}(1)$ in (1.5) are

$$\mu_X = E(X_t) = \frac{\mu_Z}{1-\alpha} \ \text{ and } \ \sigma_X^2 = Var(X_t) = \frac{\alpha\mu_Z + \sigma_Z^2}{1-\alpha^2}, \tag{1.6}$$

respectively. The autocovariance and autocorrelation functions of a stationary $\texttt{INAR}(1)$ process $\{X_t\}$ are given by the formulae

$$\gamma(k) = Cov(X_t, X_{t-k}) = \alpha^{|k|}\sigma_X^2 \ \text{ and } \ \rho(k) = Corr(X_t, X_{t-k}) = \alpha^{|k|},\ \ k \in \mathbb{Z}. \tag{1.7}$$

Furthermore, autocorrelation function $\rho(k)$ decays exponentially with lag k and for $k = 1$, the parameter α represents the correlation between successive time points.

`INAR` processes retain some of the properties of the conventional `AR` models while allowing for the discreteness of the data, namely, the fact that for both models the autocorrelation function (ACF) takes the form $\rho(k) = \alpha^k$ for $k \in \mathbb{N}$. Another important property of the `INAR`(1) model in (1.5) is that the discrete self-decomposable (DSD) distributions are possible marginal distributions, since the probability generating function (p.g.f.) of the `INAR`(1) model satisfies

$$G_X(s) = G_X(1-\alpha + \alpha s)G_Z(s), \tag{1.8}$$

where G_Z is the p.g.f. of innovation Z_t. Many important distributions, including Poisson, generalized Poisson and the negative binomial distribution belong to this class of DSD distributions (Zhu and Joe, 2003).

Different distributional forms of the innovation term Z_t have been proposed but main part of the literature have been devoted to the Poisson distribution, the simplest and most common choice. This is partly because of the favoring property that the innovation distribution belongs to the same family as the marginal distribution (Al-Osh and Alzaid, 1987). For more structural and asymptotic properties of an `INAR`(1) process with Poisson marginal, we refer the reader to, e.g. Park and Oh (1997), McKenzie (2003) and Silva and Silva (2006). However, the implied equidispersion (variance equals mean) limits the applicability of the Poisson `INAR` models in real data applications.

The simple Poisson `INAR` model can be extended to a `INAR` Poisson regression model by adding covariates to both the innovation term Z_t and/or the autocorrelation parameter α. The model then takes the form

$$\begin{aligned}
X_t &= \alpha_t \circ X_{t-1} + Z_t, \\
Z_t &\sim Poisson(\lambda_t), \\
log(\lambda_t) &= \mathbf{v}'_t\beta, \\
log\left(\frac{\alpha_t}{1-\alpha_t}\right) &= \mathbf{u}'_t\phi,
\end{aligned}$$

where $\mathbf{v}_t$ and $\mathbf{u}_t$ are vectors of covariates at time t while β and ϕ are the associated regression coefficients. Note that the covariates for the two parts of the model must not necessarily be the same. Using a discrete time series model with this specification, Brijs et al. (2008) studied the effect of weather conditions on daily crash counts, a relevant issue in road safety. For regression models based on count time series see, for instance, the books of Kedem and Fokianos (2005) and Cameron and Trivedi (2013).

Since their introduction, `INAR` processes sustained various generalizations and modifications through the work of several authors. Generalizations of the basic `INAR` model can be based on either other distributional forms for innovation Z_t or by replacing the binomial thinning operator with a different thinning. We postpone details on other thinning operators and concentrate on binomial thinning within this subsection. In practice, some discrete time dependence count data may be overdispersed, i.e., the variance is greater than the mean motivating alternative innovation distributions from the common Poisson distribution.

The generalized Poisson model is a generalization of the Poisson distribution with an extra parameter which reflects overdispersion. Alzaid and Al-Osh (1993) have considered discrete time series with generalized Poisson marginals. Mixed Poisson distributions have been used in a wide range of scientific fields, a thorough review of this family is available in Karlis and Xekalaki (2005). Nikoloulopoulos and Karlis (2008) compared four members of the mixed Poisson family. Only a few of them have been considered in practice, mainly due to computational problems.

`INAR`(1) processes with negative binomial and/or geometric marginal distribution for time series of overdispersed counts have been considered by McKenzie (1985, 1986, 2003), Alzaid and Al-Osh (1988), Al-Osh and Aly (1992), Zhu and Joe (2006) and also by Jazi et al. (2012b).

Other distributions for the innovation term include: zero truncated Poisson (`ZTP`) distribution (Bakouch and Ristić (2010) proposed the `ZTPINAR`(1) process); power series (`PS`) distribution (Bourguignon and Vasconcellos (2015) introduced the `PSINAR`(1) model); Poisson–geometric (`PG`) distribution (Bourguignon (2016) established the `PGINAR`(1) process) and Poisson–negative binomial (`PNB`) distribution (Jose and Mariyamma (2016) proposed the `PNBAR`(1) model). The `PSINAR`(1) model contains, as particular cases, the Poisson `INAR`(1) model (Al-Osh and Alzaid, 1987) and the geometric INAR(1) model (Jazi et al., 2012b). The use of innovations that come from the PS family of distributions has many advantages, this family constitutes a flexible framework for statistical modeling of discrete data in several real-life events (Johnson et al., 2005). The PGINAR(1) model extends the Poisson INAR(1) process (Al-Osh and Alzaid, 1987) and the geometric INAR(1) process (Alzaid and Al-Osh, 1988).

One frequent manifestation of overdispersion is that the incidence of zero counts is greater than expected from a Poisson model. Jazi et al. (2012a) considered an INAR(1) model with zero inflated Poisson innovations (ZINAR(1)). Meanwhile, compound Poisson (CP) distribution for the innovations of an INAR(1) model was considered by Schweer and Weiß (2014) and Weiß and Puig (2015). The CPINAR(1) model for time series of overdispersed counts revealed to be appealing and comprises a number of specialized INAR(1) models within one model.

While models for overdispersed counts have been discussed intensively in the literature by now, the opposite phenomenon, underdispersion, has received little attention. Weiß (2013) gave a detailed survey of distribution models allowing for underdispersion. Properties were derived and possible disadvantages of the model were highlighted.

INAR models contaminated with innovational and additive outliers were introduced and analyzed by Barczy et al. (2010, 2012) and Silva and Pereira (2015). Extensions of the INAR(1) model into the spatial context were considered by Ghodsi et al. (2012). The study of seasonal extensions of the INAR processes has been addressed recently by Bourguignon et al. (2016).

For higher order INAR models, two different specifications of the second-order structure can be distinguished. In Alzaid and Al-Osh (1990), the INAR(p) process has a correlation structure that is similar to that of an ARMA($p, p-1$) model. Du and Li (1991) proposed a process with a correlation structure identical to that of a standard AR(p) process.

1.1.2 Signed thinning-based INAR models

In many real-life events there is a necessity for modeling the data obtained from correlated processes which may deal with positive and negative integer values. Binomial thinning can only be applied to count variables, i.e., to non-negative integer-valued r.v.'s as their range, therefore, cannot account for negative integers. Whilst models for non-negative integer-valued time series are now abundant, there is a shortage of similar models when the time series refer to data defined on $\mathbb{Z}$, i.e., in both the positive and negative integers. Such data occur in certain fields (e.g. finance and sports). The need for such models can also appear when taking differences of positive integer-valued count time series.

The first model for data with range in $\mathbb{Z}$ was introduced by Kim and Park (2008). Their model is based on the signed binomial thinning operator.

Definition 1.2. *(Signed binomial thinning operator)*
Let $(Y_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. Bernoulli random variables with mean $|\alpha|$, independent of X, an integer-valued r.v. with range in $\mathbb{Z}$. The signed binomial thinning operator, represented by $\alpha\oplus$, is defined by

$$\alpha \oplus X := sign(\alpha) sign(X) \sum_{i=1}^{|X|} Y_i, \tag{1.9}$$

where

$$sign(x) = \begin{cases} 1 & , \ x \geq 0 \\ -1 & , \ x < 0 \end{cases}. \tag{1.10}$$

Kim and Park (2008) defined the `INARS`(p) process, an integer-valued autoregressive process of order p with signed binomial thinning operator. When $X \geq 0$ and $\alpha \geq 0$, the signed binomial thinning in (1.9) is reduced to the classic binomial thinning in (1.2). One advantage of the `INARS` model is that it can handle integer-valued time series which allows for negative integer-valued and negative correlated count data unlike the integer-valued time series models in the previous subsection. Those are only appropriate for non-negative integer-valued time series and can only deal with positive autocorrelations. The `INARS` model persists the differences in autocorrelation structure of `INAR`(p) models studied by Alzaid and Al-Osh (1990) and Du and Li (1991). Kim and Park (2008) have proven stationarity and ergodicity of the `INARS`(p) process under the same condition as in the conventional `AR`(p) process.

For a proper time series on $\mathbb{Z}$ we also need to consider a distribution for the innovation term defined on $\mathbb{Z}$. The literature is limited on this subject. However, recently, discrete distributions defined on the set of integers has attracted the attention of several researchers.

Two ways to define distributions on $\mathbb{Z}$ are: the differences between two non-negative discrete r.v.'s and the discrete version of continuous distributions on $\mathbb{R}$. The main distributions on the set $\mathbb{Z}$ are Poisson difference, discrete normal and discrete Laplace. The Poisson difference distribution, also known as the Skellam distribution, is traditionally linked to Skellam (1946) and has found applications in areas such as medicine (Karlis and Ntzoufras, 2006), sports (Karlis and Ntzoufras, 2009) and finance (Alzaid and Omair, 2010). The special case of two independent Poisson distributions for the case of equal means was derived by Irwin (1937) whereas Skellam (1946) and Prékopa (1952) discussed the general case, unequal means.

Definition 1.3. *(Univariate Skellam distribution)*
Let $\theta_1 > 0$ and $\theta_2 > 0$. The r.v. Z has Skellam distribution, denoted by Skellam(θ_1, θ_2) if and only if $Z \stackrel{d}{=} Y_1 - Y_2$ where Y_1 and Y_2 are two independent random variables such that $Y_i \sim Poisson(\theta_i)$ for $i = 1, 2$.

Thus, the probability mass function (p.m.f.) of Z is a discrete distribution, defined on the set of integer numbers $\mathbb{Z}$, given by

$$P(Z = z) = e^{-(\theta_1+\theta_2)} \left(\frac{\theta_1}{\theta_2}\right)^{z/2} I_{|z|}\left(2\sqrt{\theta_1\theta_2}\right), \; z \in \mathbb{Z}, \tag{1.11}$$

where $I_r(x)$ is the modified Bessel function of the first kind of order r defined by

$$I_r(x) = \left(\frac{x}{2}\right)^r \sum_{i=0}^{\infty} \frac{\left(\frac{x^2}{4}\right)^i}{i!\Gamma(r+i+1)}.$$

The definition of the Skellam distribution can be extended to more than the simple difference of two independent Poisson distributions. Indeed, let X_1 and X_2 be two independent Poisson random variables with parameters θ_1 and θ_2 respectively. Let $Y_i = X_i + W$, for $i = 1, 2$, where W is a r.v. independent of X_1 and X_2. Thus, $Z = Y_1 - Y_2 = X_1 - X_2$ also follows a Skellam(θ_1, θ_2) distribution. Alternative formulas for the p.m.f. of the Skellam distribution stem from the work of Alzaid and Omair (2010). For basic properties of the Skellam distribution, see Appendix A.

Some other distributions defined as the difference of two discrete variables are given in Ong et al. (2008). Kemp (1997) introduced a discrete version of normal distribution to cover discrete data on the whole set of integers $\mathbb{Z}$ and, similarly, Inusah and Kozubowski (2006) considered a discrete analogue of Laplace distribution. Kozubowski and Inusah (2006) proposed a discrete version of the skew Laplace distribution as a generalization of discrete Laplace distribution and demonstrated its importance in analysis of climatic episodes such as droughts and floods.

Andersson and Karlis (2014) introduced a first-order model with the signed binomial thinning operator assuming a specific innovation distribution, the Skellam distribution, `SINARS`(1) model (first `S` stands for Skellam). This model is a particular case of the model in Kim and Park (2008). Parametric inference and prediction for the model in Andersson and Karlis (2014) are also addressed. The marginal of `SINARS`(1) process does not have Skellam distribution. An extension of the signed binomial thinning operator given in (1.9) was then established by Zhang et al. (2010) and denoted the signed generalized power series thinning operator. These authors have proposed a generalized version of the `INARS`(p) model in Kim and Park (2008), the `GINARS`(p) process. The counting sequences have a generalized power series as common distribution, which includes the binomial, the negative binomial, the Poisson, among other distributions.

Recent work by Alzaid and Omair (2012) introduced the extended binomial distribution as an alternative to the Skellam distribution. Alzaid and Omair (2014) presented a natural $\mathbb{Z}$-extension of the `INAR` model, originally defined on $\mathbb{N}$, the new `INAR`(1) model has Poisson difference (`PD`) innovations (`PDINAR`(1)). This process can handle negative integer-valued time series and allow for both positive and negative autocorrelation. The `PDINAR`(1) model is based on the extended binomial thinning operator and has Skellam marginal distribution. Special cases of the extended binomial thinning are: binomial thinning in (1.2) and signed binomial thinning in (1.9). We also mention the extended Poisson distribution introduced by Bakouch et al. (2016), the first version of the Poisson distribution over the set of all integers.

Using a slightly different version of the signed thinning operator defined by Kim and Park (2008) in (1.9), Kachour and Truquet (2011) focused on a more general class of $\mathbb{Z}$-valued processes denoted by SINAR(p) (Signed INAR). This modified version of the thinning operator, also called the signed thinning operator, is defined as follows.

Definition 1.4. *(Signed thinning operator)*
Let $(Y_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. integer-valued random variables with distribution F, independent of an integer-valued r.v. X. The signed thinning operator, denoted by $F\odot$, is defined by

$$F \odot X := \begin{cases} sign(X) \sum_{i=1}^{|X|} Y_i & , \; X \neq 0 \\ 0 & , \; otherwise \end{cases} \tag{1.12}$$

with $sign(X)$ as in (1.10). The sequence $(Y_i)_{i\in\mathbb{N}}$ is referred to as a counting sequence.

Some basic properties of the signed thinning operator are listed in Lemma 1.2. These easily follow from the independence assumptions and the obvious identity $x = sign(x)|x|$, $x \in \mathbb{R}$.

Lemma 1.2. *(Properties of the signed thinning operator)*
Let X, W be two random variables and $Y, \tilde{Y}$ two counting sequences with distribution $F, \tilde{F}$, respectively. Assume that (X, W), F and $\tilde{F}$ are independent. Consider α the mean and β the variance of the distribution function F. Then,

1. $E(F \odot X|X) = \alpha X$,

2. $Var(F \odot X|X) = \beta|X|$,

3. $Cov(F \odot X, \tilde{F} \odot W|X, W) = 0$,

4. $Cov(F \odot X, W) = \alpha Cov(X, W)$.

Kachour and Truquet (2011) pointed out that the signed thinning operator is the natural extension of the Steutel and van Harn (1979) operator in (1.2) to $\mathbb{Z}$-valued random variables. Moreover, for a non-negative integer-valued random variable X, the signed thinning operation in (1.12) is the popular binomial thinning operation in (1.2). In Definition 1.4, the notation $F \odot X$ replaces the usual notation $\alpha \circ X$ in binomial thinning, where α denotes the mean of the counting sequence. The choice for the notation $F \odot X$ was motivated by the fact that Kachour and Truquet (2011) did not fix any specific one parameter distribution for the counting sequence $(Y_i)_{i\in\mathbb{N}}$ such as a Bernoulli distribution.

The `SINAR` model allows negative values for both the series and its autocorrelation function. Theoretical results about stationarity and the moments of `SINAR` processes were given in Kachour and Truquet (2011). The authors avoid, however, a parametric assumption for the innovation term. Based on the preceding operator and under a parametric assumption on the common distribution of the counting sequence of the model, Chesneau and Kachour (2012) focus on the simplest `SINAR`(1) model. They also introduced a new class of distribution on $\mathbb{Z}$, denoted by Rademacher(p)-$\mathbb{N}$ with $p \in (0,1)$. This distribution can be interpreted as a natural extension of the Bernoulli distribution from $\{0,1\}$ to $\{-1,1\}$.

Definition 1.5. *(Rademacher(p)-$\mathbb{N}$ distribution)*
Let R and W be two independent random variables such that $R \sim$ Rademacher(p), that is,

$$P(R=1) = p = 1-P(R=-1), \quad p \in (0,1)$$

and support$(W) \subseteq \mathbb{N}$. A r.v. X belongs to the Rademacher(p)-$\mathbb{N}$ class, if and only if $X \overset{d}{=} RW$.

Indeed, the Rademacher distribution is a recoding of the Bernoulli distribution, where 1 still indicates success, but failure is coded as -1. Therefore, if random variable $Y \sim$ Bernoulli(p) then r.v. $R = 2Y-1 \sim$ Rademacher(p). This distribution is also related with the Skellam distribution (Chesneau et al., 2015). Let $R \sim$ Rademacher(p) and $Z \sim$ Skellam(θ_1, θ_2) then the random variable $Z^* = RZ$ is a mixture of two Skellam random variables of the form pSkellam$(\theta_1, \theta_2) + (1-p)$Skellam$(\theta_2, \theta_1)$.

1.1.3 Other univariate thinning-based INAR models

In the previous subsections, emphasis has been given to binomial and signed thinning operators but other generalizations of the `INAR`(1) model are available in the literature. Several authors have proposed modifications of the thinning operation in order to make thinning-based models more flexible for practical purposes. We mention other relevant cases since the variety of counting series demands some modification in terms of the thinning operator and marginal distribution.

The generalized thinning operator introduced by Latour (1998) allowed the counting sequences in (1.2) to be i.i.d. integer-valued r.v.'s with finite mean and variance although not necessarily Bernoulli-distributed. Modifying the `INAR`(1) recursion (1.5) accordingly leads to the generalized `INAR`(1) process denoted by `GINAR`(1). Furthermore, the `GINAR`(p) model in Latour (1998) is the generalized counterpart of the `INAR`(p) model by Du and Li (1991). Special cases of Latour's operator can be found in Zhu and Joe (2003) (extended thinning), in Zhu and Joe (2010) (expectation thinning) and in Weiß (2015) (binomial-Poisson thinning).

Random coefficient `INAR` (`RCINAR`) models were introduced by Zheng et al. (2006, 2007), providing nonparametric as well as parametric methods for parameter estimation. In some situations, the autoregressive parameter α in (1.5) may vary with time and it may be random. For example, let X_t denote the number of terminally ill patients in the t-th month. Here, X_t could potentially satisfy an `INAR` model where $\alpha \circ X_{t-1}$ is the number of surviving patients from the previous month and Z_t stands for the newly admitted patients in the current month. In addition, the survival rate α may be affected by various environmental factors, such as the quality of health care, the state of health of patients, etc. and could vary randomly over time. Another area of application could be unemployment that can be affected by factors such as the state of the economy, productivity growth, among others. The `RCINAR` processes are able to describe overdispersion. Gomes and Canto e Castro (2009) extended the concept of random coefficient thinning in analogy to Latour's generalized thinning operator. For the particular case of the (generalized) binomial thinning, Gomes and Canto e Castro (2009) proved that the necessary and sufficient conditions for weak stationarity are the same as those for continuous-valued `AR`(1) processes.

Wang and Zhang (2011) also extended the signed binomial thinning operator in (1.9) and developed the generalized pth-order random coefficient `INAR` process with signed binomial thinning (`GRCINARS`(p)). Zhang et al. (2012) study the `GRCINARS`(1) model in detail. `INAR` models based on random coefficient thinning operators have been explored by Roitershtein and Zhong (2013), Tang and Wang (2014), Zhao and Hu (2015) and Zhang and Wang (2015). Using the concept of random coefficient thinning for modeling of count data time series with a finite range, Weiß and Kim (2014) introduced a beta-binomial autoregressive model.

Ristić et al. (2009) provided a new stationary INAR(1) process with geometric marginals (NGINAR(1)) based on the negative binomial thinning operator which contains geometric counting series. Further properties of this model were developed by Bakouch (2010). The motivation for time series with geometric marginal distributions is due to their major role in, e.g., the reliability theory, medicine, and precipitation modeling, arising from the number of machines waiting for maintenance, the number of congenital malformations, and the number of thunderstorms in a day. Nastić et al. (2012) considered a new (combined) INAR model of order p with geometric marginal distribution (CGINAR(p)) based on the negative binomial thinning introduced by Ristić et al. (2009). Using the preceding thinning operator but with negative binomial (NB) marginals, Ristić et al. (2012a) established the NBINAR(1) process and Nastić (2014) the combined NBINAR process of order p, CNBINAR(p). Integer-valued time series generated by mixtures of binomial and negative binomial thinning operators are considered in Nastić and Ristić (2012) and Ristić and Nastić (2012). Meanwhile, Li et al. (2015) introduced a first-order mixed INAR processes with zero-inflated generalized power series innovations, denoted by ZIMINAR(1). These innovations contain the commonly used zero-inflated Poisson and geometric distributions. Two thinning operators were mixed, namely the binomial thinning (Al-Osh and Alzaid, 1987) and the negative binomial thinning operator (Ristić et al., 2009). An INAR(1) process with NB thinning and zero-modified geometric (ZMG) marginals was introduced by Barreto-Souza (2015). The ZMGINAR(1) model is also able to capture under/over dispersion, which sometimes is caused by deflation or inflation of zeros.

In a different approach from Kim and Park (2008), Freeland (2010) extended discrete time series models with non-negative values to models over the integers, modifying the binomial thinning operator to produce a stationary AR(1) model with a Skellam marginal distribution. More specifically, the Poisson INAR(1) model is extended to a symmetric model around zero, the true INAR (TINAR(1)). The thinning operator considered by Freeland (2010) is somewhat delicate to work with because it is defined on two latent counts for which only the difference is observed. One of the main features of the TINAR(1) model is it incorporates both positive and negative correlation. Also arising from the difference between two discrete distributions Barreto-Souza and Bourguignon (2015) established a skew INAR(1) process on $\mathbb{Z}$, denoted by SINARZ(1), with skew discrete Laplace (SDL) marginals (Kozubowski and Inusah, 2006). This new model is based on a modified version of the NB thinning operator introduced by Ristić et al. (2009) but in a similar fashion as in Freeland (2010), it acts on two independent but not necessarily identically distributed latent NGINAR(1) processes. While the TINAR(1) process established by Freeland (2010) is symmetric, the skew INAR(1) model on $\mathbb{Z}$ by Barreto-Souza and Bourguignon (2015) can accommodate skewness. The probability function of the SDL distribution has a simple form in contrast with the Skellam distribution which involves the modified Bessel function of the first kind. Following a similar approach in model construction as in Freeland (2010), Bourguignon and Vasconcellos (2016) proposed the new skew INAR(1) process, named NSINAR(1), with geometric–Poisson marginals (which are distributed as a difference between geometric and Poisson r.v.'s) and Nastić et al. (2016a) a process with the discrete Laplace DL marginal distribution (DLINAR(1)). The thinning operator of the model was, once again, based on the negative binomial thinning of Ristić et al. (2009). An extension on INAR models with skew discrete Laplace marginal distributions was introduced by Djordjević (2016), the SDLINAR(1) model, representing a generalization of two mentioned models, SINARZ and DLINAR.

Another contribution in modeling time series that can incorporate negative count and negative values for autocorrelation was made by Kachour and Yao (2009) based on the rounding operator. They have presented the rounded integer-valued time series process of order one (RINAR(1)). A more general setup has been introduced since then by Kachour (2014).

It is important to stress here the fact that all thinning operators previously mentioned depend upon the assumption of independence across the counting variables. Extensions of binomial thinning based on Bernoulli-distributed dependent r.v.'s were proposed by Brännäs and Hellström (2001). Significant contributions on this subject can be found in Ristić et al. (2013) and in recent works of Ilić (2016) and Nastić et al. (2017). The literature for univariate count time series models is now quite mature.

1.2 Multivariate time series models for count data - a review

The study of multivariate `INAR`-type processes for count data has become a topic of special interest during the last years. Multivariate count data can occur in many fields such as finance, criminology, epidemiology, etc. Special attention has been devoted to bivariate integer-valued time series processes. In the situation where two series of counts interact and where the evolution of one series is dependent on the other, bivariate models are the most appropriate. These models maintain the pairing between two count variables that occur over specific times and play a major role in the analysis of the paired correlated count data. Bivariate generalizations of important univariate distributions are also of continuing interest. The generalizations can be constructed in a wide variety of ways like mixing, compounding and trivariate reduction. The former method is a popular method of construction due to its simplicity and ease of computationally generating samples (more details in e.g. Lai (2006)).

1.2.1 Matrix-binomial thinning-based INAR models

As in the univariate case, before defining a multivariate process, we need to define a corresponding thinning operator. The definition of matrix-binomial thinning follows.

Definition 1.6. *(Matrix-binomial thinning operator)*
Let $\mathbf{X} = [X_1\ X_2\ \dots\ X_m]^T$ *be a random vector with values in* $\mathbb{N}_0^m$ *and* $(m \times m)$ *matrix* $A = [a_{ij}]$ *with entries* $a_{ij} \in [0;1]$*. The matrix-binomial thinning* $A \circ \mathbf{X}$ *is a* m*-dimensional random vector whose* i*-th component is given by*

$$[A \circ \mathbf{X}]_i = \sum_{j=1}^{m} a_{ij} \circ X_j,\ i = 1, \dots, m, \tag{1.13}$$

where the operator $\circ$ *represents the binomial thinning in (1.2). Furthermore, the counting series of all* $a_{ij} \circ X_j$*,* $i, j = \dots, m$*, are assumed independent.*

Franke and Subba Rao (1993) introduced a m-variate INAR(1) model (MINAR(1) for short), based on independent binomial thinning operators, given by recursion

$$\mathbf{X}_t = A \circ \mathbf{X}_{t-1} + \mathbf{Z}_t, \ t \in \mathbb{Z}, \tag{1.14}$$

for $(m \times m)$ matrix A with entries in $[0;1]$ and $\mathbf{Z}_t$ an i.i.d. random vector taking values in $\mathbb{N}_0^m$. Some properties of the thinning operator are listed in Lemma 1.3. Proofs of the properties can be found in Franke and Subba Rao (1993).

Lemma 1.3. *(Properties of the matrix-binomial thinning operator)*
Let A and B be $(m \times m)$ matrices; $\mathbf{X}$ and $\mathbf{Y}$ be non-negative integer-valued random m-vectors.

1. $E[A \circ \mathbf{X}] = AE[\mathbf{X}]$;
2. $E[(A \circ \mathbf{X})(A \circ \mathbf{X})^T] = AE[\mathbf{X}\mathbf{X}^T]A^T + diag(DE[\mathbf{X}])$,
   where D is the variance matrix;
3. $E[(A \circ \mathbf{X})(B \circ \mathbf{Y})^T] = AE[\mathbf{X}\mathbf{Y}^T]B^T$,
   if the counting series $A \circ \mathbf{X}$ and $B \circ \mathbf{Y}$ are independent.

Extensions of the univariate INAR processes to the bivariate case have been introduced by several authors. Pedeli and Karlis (2011) extended the INAR(1) model to a bivariate integer-valued autoregressive process of order one (BINAR(1) in short), where the correlation is introduced through innovation components. Let $\mathbf{X}_t$ and $\mathbf{Z}_t$ be non-negative integer-valued random 2-vectors. Let A be a (2×2) diagonal matrix with independent elements $\{\alpha_j\}_{j=1,2}$. The BINAR(1) model is defined as

$$\mathbf{X}_t = A \circ \mathbf{X}_{t-1} + \mathbf{Z}_t \equiv \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \circ \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix}, \ t \in \mathbb{Z}, \tag{1.15}$$

where $A\circ$ is the matrix-binomial thinning defined in (1.13) and $[Z_{1,t} \ Z_{2,t}]^T$ are assumed to be independent $\mathbb{N}_0^2$-valued random pairs. All thinning operations are performed independently of each other and of $\mathbf{Z}_t$.

From the definition of the bivariate model BINAR(1) in (1.15), the j-th component is given by $X_{j,t} = \alpha_j \circ X_{j,t-1} + Z_{j,t}$ for $j = 1, 2$. The marginals behave like the univariate binomial thinning operator. Dependence between the two series that comprise the BINAR(1) process is introduced by allowing for dependence through innovation components $Z_{j,t}$ for $j = 1, 2$. Whatever the underlying joint distribution of $\mathbf{Z}_t$ is, Pedeli and Karlis (2011) have shown that the covariance between the innovations of the two series at time t, totally determines the covariance between the current value of one process and the innovations of the other process at the same point in time t and vice versa. Two specific BINAR(1) models were introduced by Pedeli and Karlis (2011). One model arises from the assumption the innovations follow jointly a bivariate Poisson distribution. For a comprehensive description of the bivariate Poisson distribution, we refer the reader to the books of Kocherlakota and Kocherlakota (1992) and Johnson et al. (1997). Interestingly, the generated BINAR(1) model has a stationary distribution that is itself a bivariate Poisson distribution. Moreover, the univariate processes for each variable are simple INAR(1) processes with Poisson marginals. The BINAR(1) model neatly generalizes the typical univariate model. The disadvantage of this particular model is that it does not allow for over/under dispersion (the marginal distributions are Poisson) or negative correlation, and thus lacks generality. The second model assumes a bivariate negative binomial (BVNB) distribution for the innovations. There are several representations for the BVNB distribution in the literature. Pedeli and Karlis (2011) have considered the distributional form followed by Marshall and Olkin (1990), Boucher et al. (2008) and Cheon et al. (2009). This assumption allows for more flexibility than the Poisson BINAR(1) model due to the involvement of the overdispersion parameter. However, the resulting model is not a BINAR(1) model with negative binomial marginals but a model that effectively accounts for overdispersion. The two distributions seem to be appropriate for modeling equidispersed and overdispersed bivariate time series, respectively. In the bivariate setting, 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. Pedeli and Karlis (2011) addressed forecasting and predictions by means of the conditional forecast distribution. An application concerning road accidents (during day and night) was provided.

Focusing on the specific case of the BINAR(1) model that arises through the assumption of bivariate Poisson innovations, Pedeli and Karlis (2013b) considered some alternative estimators for the unknown parameters of the model and examined their behavior. An extension of the model to a BINAR(1) Poisson regression model is also discussed. Pedeli and Karlis (2013a) then extended the BINAR model proposed in Pedeli and Karlis (2011) to the multi-dimensional space. Again, the authors focused on two parametric cases for the multivariate INAR(1) model: multivariate Poisson distribution and multivariate negative binomial distribution for the innovation processes. However, the classical definition of the multivariate Poisson distribution (Johnson et al., 1997) was not followed, the formulation in Karlis and Meligkotsidou (2005) seemed more convenient. In the multivariate setting, computational issues arise in parameter estimation of the unknown parameters, the complexity of the maximum likelihood approach augments with dimensional increase. To overcome these difficulties, the concept of composite likelihood estimation was suggested by Pedeli and Karlis (2013a) and its performance compared with conditional maximum likelihood estimation. The term composite likelihood originated from Lindsay (1988). Composite likelihood methods based on optimizing sums of log-likelihoods of low-dimensional margins have been considered by many authors in recent years; they are useful for multivariate models in which the likelihood of multivariate data is very time-consuming. In particular, pairwise likelihood or bivariate composite likelihood methods are based on bivariate margins. An excellent overview of composite likelihood methods can be found in Varin et al. (2011), complementing and extending the review in Varin (2008). Other relevant references on this subject are: Cox and Reid (2004), Varin and Vidoni (2005) and Zhao and Joe (2005).

All models proposed in Pedeli and Karlis (2011) and Pedeli and Karlis (2013a,b) rely on a constraint: the matrix A for autocorrelation parameters is diagonal, meaning there is no cross-correlation in the counts. The assumption of diagonality implies that correlation between the innovations is the only source of dependence between the two series $X_{j,t}$, $j = 1, 2$. Removing such an assumption would imply that cross–correlation do not solely arise from the correlation between the innovation series of the multivariate process. The MINAR(1) process in Franke and Subba Rao (1993) relies upon a non-diagonal autoregression matrix A. This framework was also followed by Boudreault and Charpentier (2011) and by Pedeli and Karlis (2013c). By allowing for an additional source of dependence, i.e., relaxing the assumption of diagonality of the matrix A, Pedeli and Karlis (2013c) introduced full multivariate `INAR`(1) process, with special emphasis on the bivariate case with bivariate Poisson innovations. In this context, the joint distribution of $\{X_{1,t}, X_{2,t}\}$ is a 8-parameter bivariate Hermite distribution (Kemp and Papageorgiou, 1982). More insight in bivariate `INAR` models based on binomial thinning operator and Poisson marginals by Nastić et al. (2016b). Due to the growing interest in zero truncated distributions, pioneer work on bivariate `INAR` models with zero truncated Poisson marginal distribution has been introduced by Liu et al. (2016), extending the univariate model in Bakouch and Ristić (2010).

Most of the bivariate `INAR` models investigated in the literature uses constants for the regression coefficients. Popović (2016) developed bivariate models with random coefficients, based on binomial thinning operator with unequal parameters. Innovations are mutually independent and distributed in a way to support the stationarity of the processes. The marginal distribution is assumed to be geometric. Another contribution to models that comprise random coefficients but with dependent innovations was established earlier by Popović (2015).

1.2.2 Signed matrix thinning-based INAR models

An extension of the signed thinning operator in (1.12) to the bivariate case was established by Bulla et al. (2016). The definition of the signed matrix thinning operator follows.

Definition 1.7. *(Signed matrix thinning operator)*
Let $\mathbf{X} = [X_1 \ X_2]^T$ *be an integer-valued random vector. The signed matrix thinning operator is given by*

$$F \odot \mathbf{X} := \begin{bmatrix} F_{11} \odot X_1 + F_{12} \odot X_2 \\ F_{21} \odot X_1 + F_{22} \odot X_2 \end{bmatrix}, \tag{1.16}$$

where F_{ij} *represents the common distribution of the i.i.d. counting sequences* $(Y_k^{ij})_{k\in\mathbb{N}}$ *for any* $(i, j) \in (1, 2) \times (1, 2)$*. It is assumed that all counting sequences associated with* $F_{ij}\odot$ *are mutually independent.*

Bulla et al. (2016) introduced the class of bivariate signed `INAR`(1) processes based on the signed matrix thinning operator in (1.16), denoted by `B-SINAR`(1), which is an extension of the `SINAR`(1) process of Kachour and Truquet (2011) to the bivariate case. Compared to classical bivariate `INAR` models, that cannot fit a time series with negative observations, the `B-SINAR` models have the advantage to allow for negative values for the time series and the autocorrelation functions. A bivariate process $\mathbf{X}_t = [X_{1,t}\ X_{2,t}]^T$ is called a `B-SINAR`(1) process if $\mathbf{X}_t$ admits the following representation

$$\mathbf{X}_t = F \odot \mathbf{X}_{t-1} + \mathbf{Z}_t \equiv F \odot \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix},\ t \in \mathbb{Z}, \tag{1.17}$$

where $\mathbf{Z}_t = [Z_{1,t}\ Z_{2,t}]^T$ are assumed to be independent. All counting sequences associated with $F_{ij}\odot$ are mutually independent for $(i, j) \in (1, 2) \times (1, 2)$.
A particular case is when $F_{12} = F_{21} = 0$ (assumption of diagonal autoregressive matrix) which can be seen as a $\mathbb{Z}^2$-extension of the model presented in Pedeli and Karlis (2011). In contrast to the well-known situation when the paired data are counts, i.e., observed on $\mathbb{N}^2$, sometimes the data take values in $\mathbb{Z}^2$. While bivariate discrete distribution for non-negative paired data are now abundant, there is a shortage of bivariate discrete distribution defined on $\mathbb{Z}^2$. Bulla et al. (2015) contributed to the literature by developing the bivariate Skellam distribution, recasting the interest on the distribution introduced by Skellam (1946).

Definition 1.8. *(Bivariate Skellam distribution)*
Let $\theta_0 \geq 0$, $\theta_1 > 0$ and $\theta_2 > 0$. The bivariate random variable (X_1, X_2) follows a bivariate Skellam distribution, denoted by BSkellam$(\theta_0, \theta_1, \theta_2)$, if and only if

$$X_1 \sim Skellam(\theta_1, \theta_0) \quad and \quad X_2 \sim Skellam(\theta_2, \theta_0).$$

Thus, the joint p.m.f. of (X_1, X_2) is given by

$$P(X_1 = x_1, X_2 = x_2) = e^{-(\theta_1+\theta_2+\theta_0)}\theta_1^{x_1}\theta_2^{x_2} \sum_{i=max(0,-x_1,-x_2)}^{\infty} \frac{(\theta_0\theta_1\theta_2)^i}{(x_1+i)!(x_2+i)!i!},\ (x_1, x_2) \in \mathbb{Z}^2. \tag{1.18}$$

The mean and variance are $E[X_i] = \theta_i - \theta_0$ and $Var[X_i] = \theta_i + \theta_0$ for $i = 1, 2$, respectively. The covariance of X_1 and X_2 is θ_0, and hence, θ_0 is a measure of dependence between the two r.v.'s. However, if $\theta_0 = 0$, then the two variables are independent and the bivariate Skellam distribution reduces to the product of two independent Poisson distributions (referred to as double Poisson). Further details on bivariate Skellam distribution in Bulla et al. (2015).

With the assumption of a diagonal matrix for autocorrelation parameters, Bulla et al. (2016) assumed that innovations $\mathbf{Z}_t$ are modeled through a bivariate Skellam distribution (Bulla et al., 2015). In order to increase the flexibility of the bivariate Skellam distribution, the authors proposed two alternative extensions: the inclusion of a shift parameter $k = (k_1, k_2)$ and mixtures of bivariate Skellam distributions. Many bivariate extensions of Skellam distribution are possible, for example through copulas (Genest and Mesfioui, 2014). For an introduction to the subject, see e.g. Nelsen (2007) and Genest and Nešlehová (2007). A family of distributions on $\mathbb{Z}^2$ based on generalized trivariate reduction technique and the Rademacher distribution (in Definition 1.5) has been explored recently by Chesneau et al. (2015).

1.2.3 Other multivariate INAR models

In the previous two subsections and references therein, focus was placed upon matrix-binomial and signed matrix thinning-based `INAR` models. Due to applications concerning data of different nature and origin, `INAR` models have experienced significant modifications and generalizations over time. In the last two decades, special attention has been devoted to multivariate (mainly bivariate models) integer-valued time series count processes with different thinning operators and different distributions for the underlying innovations.

General discussion on multivariate `INAR` processes has been covered by Latour (1997) where the author introduced the multivariate `INAR` model based on generalized Steutel and van Harn thinning operators (`MGINAR`(p)) as well as proof of the existence of the process. Applications of multivariate `INAR` processes are also presented in Brännäs and Nordström (2000). Several bivariate extensions of the thinned process have been considered by a number of researchers.

Ristić et al. (2012b) discussed the bivariate INAR(1) model based on negative binomial thinning operator while the time series are mutually dependent and have geometric marginal distribution with the same mean parameters. Popović et al. (2016) proposed relaxing the assumption about the equality of the mean parameters. The authors introduced a model with geometric marginal distribution, but with different mean parameters and subsequently derive the distribution of the innovation processes. Although, the structure of their model is similar to the one presented in Ristić et al. (2012b), different marginal distributions significantly influence on the properties of the model and particularly on the definition of the innovation processes in order to achieve stationarity. Meanwhile, a different generalization of the binomial thinning operator in (1.2) for the bivariate case was derived by Scotto et al. (2014), useful to fit count data time series with a finite range of counts. This thinning operator is based on the bivariate binomial distribution of type II ($\texttt{BVB}_{\text{II}}$) (Kocherlakota and Kocherlakota, 1992) and can account for positive or negative cross-correlation. Furthermore, Scotto et al. (2014) introduced the bivariate binomial AR(1) model ($\texttt{BVB}_{\text{II}}$-AR(1)).

Copula-based models can be used in order to define flexible bivariate discrete distributions which can serve as the distribution of the innovations in the bivariate INAR model. Karlis and Pedeli (2013) introduced copulas to create a richer alternative and allow for more flexible bivariate distributions for the innovations making it possible to accommodate both positive and negative correlation. Insight in modeling multivariate count data using copulas can be found in Heinen and Rengifo (2007) and Nikoloulopoulos and Karlis (2010).

1.3 Periodic time series models

There are many applications in which model parameters need to vary periodically to adequately describe the time series. This has lead to the study of the so-called periodic time series models. In this section, we start by reviewing periodic time series in the conventional case, i.e., continuous-valued time series. However, regarding periodically correlated integer-valued time series, very few contributions are known.

1.3.1 Continuous case

The class of periodic autoregressive moving average models of orders p and q with period T, denoted here by $\texttt{PARMA}_T(p, q)$, are an extension of autoregressive moving average ($\texttt{ARMA}(p, q)$) models in recursion (1.1) in the sense that they allow the parameters to vary periodically in time. PARMA model contains an ARMA model for each season. The concept of periodically correlated (also known as cyclostationary) processes was introduced by Gladyshev (1961, 1963) and has received much attention. Franses and Paap (2004) provided a nice overview of the subject and Hurd and Miamee (2007) discussed the procedures to detect periodic correlations in time series. The periodically correlated time series occur in many scientific disciplines where the data may have significant periodic behavior in the mean and covariance structure. Formally, a time series is called periodically correlated (PC) with period T if the mean and covariance of the series remains the same when shifted T units of time. In another words, the fundamental characteristic of a periodic time series $\{Y_n\}$ is the periodic stationarity of the first and second moments, i.e.,

$$E[Y_{n+T}] = E[Y_n] \quad \text{and} \quad Cov(Y_{n+T}, Y_{m+T}) = Cov(Y_n, Y_m), \tag{1.19}$$

for all integers n and m. The period T is the smallest positive integer satisfying (1.19). When $T = 1$, periodically correlated time series are stationary. Since their introduction there have been very extensive developments in the theory and applications of PC processes. A particular example of a periodic series is a monthly time series of air temperature with its annual cycle. The mean is clearly non-stationarity, it varies in a regular pattern depending on the month. One way of handling such a series with ARMA modeling is by applying periodic ARMA models, in which separate parameters are simultaneously estimated for each month of the year. A ubiquitous problem in fitting a PARMA model to a periodic series, however, lies with parsimony. Even very simple PARMA models can have an inordinately large number of parameters. The PARMA model has $(p + q)T$ autoregressive and moving-average parameters and T additional white noise variance parameters. This parameter total can be large for even moderate T, making some PARMA inference matters unwieldy.

Several researchers have dealt with periodic time series models. Contributions have been made in many fields such as: climatology [Jones and Brelsford (1967), Bloomfield et al. (1994), Lund et al. (1995)]; economics [Parzen and Pagano (1979), Franses and Paap (2004)]; hydrology [Vecchia (1985), McLeod (1993), Hipel and McLeod (1994), Tesfaye et al. (2006)]; electrical engineering [Gardner et al. (2006)], among others.

PARMA series differ from seasonal autoregressive moving-average (SARMA) series. McLeod (1993) demonstrated the drawbacks of forecasting a PARMA series with SARMA methods through a real data application. The analysis of basic probabilistic properties of PARMA models as well as statistical inference and forecasting techniques has been addressed by Basawa et al. (2004), Shao and Ni (2004), Shao (2006), Lund et al. (2006) and, more recently, Anderson et al. (2013). Developments on parameter estimation include e.g. Lund and Basawa (2000), Basawa and Lund (2001) and Anderson and Meerschaert (2005). Shao (2008) suggested a robust estimation procedure for the parameters in periodic AR (PAR) models and Sarnaglia et al. (2010) when data contains additive outliers. Using genetic algorithms, Ursu and Turkman (2012) provided PAR model identification.

Although the periodic models have been widely studied, most of the existing studies are confined to the univariate case. Aknouche (2007) established the causality conditions and autocovariance calculations of periodic vector autoregressive models (PVAR). Ursu and Duchesne (2009) derived the asymptotic distribution of the least square estimators of the parameters of the PVAR models. Recent contributions to PVAR models have been made by Duchesne and Lafaye de Micheaux (2013), Ursu and Pereau (2014) and Bentarzi and Djeddou (2014).

1.3.2 Discrete case

In contrast to the continuous case, it is worth to emphasize that the analysis of periodically correlated series of counts has not received much attention in the literature. To our knowledge, the first contribution in the discrete case was introduced by Monteiro et al. (2010). The so-called Periodic INteger-valued AutoRegressive process of order one with period T (PINAR$(1)_T$ for short), based on the binomial thinning operator with periodically varying parameter, is defined by the recursive equation

$$X_t = \phi_t \circ X_{t-1} + Z_t,\ t \in \mathbb{Z}, \tag{1.20}$$

where $\phi_t = \alpha_j \in (0,1)$ for $t = j + kT;\ j = 1,\ldots,T$ and $k \in \mathbb{N}_0$. The thinning operator $\circ$ is given by

$$\phi_t \circ X_{t-1} \overset{d}{=} \sum_{i=1}^{X_{t-1}} U_{i,t}(\phi_t) \tag{1.21}$$

with $(U_{i,t}(\phi_t))_{i\in\mathbb{N}}$ a periodic sequence of independent Bernoulli r.v.'s with success probability $P(U_{i,t}(\phi_t) = 1) = \phi_t$. For this setting, the natural choice for the distribution of the innovation term was Poisson distribution. Monteiro et al. (2010) assumed that innovation term Z_t in recursion (1.20) constitutes a periodic sequence of independent Poisson-distributed random variables with mean $\nu_t = \lambda_j$ for $t = j + kT$ $(j = 1,\ldots,T; k \in \mathbb{N}_0)$, which are assumed to be independent of X_{t-1} and $\phi_t \circ X_{t-1}$. To avoid ambiguity, T is taken as the smallest positive integer satisfying (1.20). Basic probabilistic and statistical properties of the $\mathtt{PINAR}(1)_T$ process with Poisson marginal were established by Monteiro et al. (2010). The existence of an almost surely unique non-negative integer-valued periodically stationary process satisfying equation (1.20) was proven. Furthermore, parameter estimation was addressed through four different methods and their performance compared. An application regarding the number of short-term unemployed people was presented. Recently, Jia et al. (2014) provided several approaches to estimate the parameters of the $\mathtt{PINAR}(1)_T$ model in the presence of missing data, by employing the idea of Andersson and Karlis (2010).

Within the bivariate setting, Monteiro et al. (2015) proposed an extension of the periodic univariate model given in (1.20). The bivariate model is referred to as the periodic bivariate INAR model of order one, denoted by PBINAR(1) with period $T \in \mathbb{N}$, and has the following form

$$\mathbf{X}_t = A_t \circ \mathbf{X}_{t-1} + \mathbf{Z}_t \equiv \begin{bmatrix} \phi_{1,t} & 0 \\ 0 & \phi_{2,t} \end{bmatrix} \circ \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix},\ t \in \mathbb{Z} \tag{1.22}$$

with $\phi_{j,t} = \alpha_{j,i} \in (0,1)$ for $t = i + kT; i = 1,\ldots,T; k \in \mathbb{N}_0$ and $j = 1,2$. The matricial operation $A_t\circ$ follows definition in (1.13) adapted to the periodic case, it acts as the usual matrix multiplication keeping the properties of the binomial thinning operator (Pedeli and Karlis, 2011). The PBINAR(1) model with period T and periodic bivariate Poisson innovations can be viewed as a generalization to the periodic case of the model established in Pedeli and Karlis (2011). The role of the innovations $\mathbf{Z}_t$ is relevant since not only they determine the joint distribution of the two series under consideration but also form the unique source of cross–correlation. Monteiro et al. (2015) derived criteria for the existence and uniqueness of a periodically stationary and causal process. For the bivariate setting, two specific parametric cases for the joint distribution of the innovations of the two series were considered: bivariate Poisson and bivariate negative binomial distributions. However, only the conditional maximum likelihood method was used for parameter estimation of the PBINAR(1) model with period T. The second parametric case revealed better fit and more suitable for series displaying overdispersion. Issues related with forecasting were also provided. Potential application of the proposed periodic bivariate model with period T can be found in the analysis of fire activity (Monteiro et al., 2015).

1.4 Parameter estimation and forecasting

The important issues of estimation and forecasting in INAR-type models are discussed in numerous papers. The most widely used estimators in the literature for the parameters of Poisson INAR(1) processes (Al-Osh and Alzaid, 1987) are Yule-Walker (YW), conditional least squares (CLS) (Klimko and Nelson, 1978) and conditional maximum likelihood (CML) estimators. YW estimation is a traditional way for estimating parameters of an AR(p) model. This method was also used by Du and Li (1991) in an INAR(p) model. For the unknown parameters involved in an INAR(1) model with Poisson marginal, asymptotic distribution of YW-type estimators were derived by Park and Oh (1997) and asymptotic properties of CLS estimators by Freeland and McCabe (2005). However, neither proved to be more efficient than the other to this order. Due to the fact that the conditional variance of the INAR(1) process is not constant over time, weighted conditional least squares (WCLS) estimators seem an attractive alternative to consider (Monteiro et al., 2010).

Jung et al. (2005) study the performance of new types of generalized method of moments estimators. Data generated with different innovations was considered: Poisson innovations and negative binomial (NB) innovations. The former distribution is often used in empirical work to capture overdispersion phenomenon. An estimation approach was proposed by Savani and Zhigljavsky (2007) who studied a family of moment-based estimation methods, called the power method estimators, for estimating parameters of the NB distribution.

Bu et al. (2008) have extended earlier work of Freeland and McCabe (2004a) and developed a general framework for likelihood analysis of GINAR(p) processes with generalized thinning operators and innovation distributions. The likelihood is derived using a recursive formulation of the transition probabilities and, in a similar way as in Freeland and McCabe (2004a), the elements of the score and the Fisher information matrix are expressed in terms of conditional expectations.

The most common technique for constructing forecasts in conventional time series model is to use the conditional expectation because they yield forecasts with minimum mean squared error. However, in the context of count time series, the conditional mean may not be an integer and hence it is not coherent. To preserve the integer-valued nature of data, the median was used as a forecast of an INAR(1) model (Freeland and McCabe, 2004b). Mode forecasting can also be used to obtain h-step ahead coherent forecasting. McCabe and Martin (2005) explored the issue of coherent forecasting with count data models under the Bayesian framework, but they too are concerned only with the first-order case. Jung and Tremayne (2006) proposed the use of second-order INAR models in the context of forecasting low integer values of count data. Bu and McCabe (2008) provided an interesting approach for forecasting based on Markov chains, the forecasts of the distribution of a count series were obtained by means of a transition matrix of the process. In recent work of McCabe et al. (2011), a new method for producing efficient probabilistic forecasts in the INAR(p) class was provided.

To overcome computational difficulties that frequently arise in maximum likelihood (ML) methods, Pedeli and Karlis (2013a) exploited the composite likelihood method, which is based on the idea of constructing lower dimensional score functions that still contain enough information about the structure considered and are computationally less demanding. Pedeli et al. (2015) proposed a simple saddlepoint approximation to the log-likelihood that revealed a good performance concerning INAR(p) models with Poisson and NB innovations. The authors have empirically proven that the estimator that maximizes the saddlepoint approximation behaves very similarly to the ML estimator.

Concerning models with signed binomial thinning, several bootstrap approaches in the literature as distribution free alternatives were used to obtain forecasts and confidence intervals. Kim and Park (2008) employed a modified bootstrap method to incorporate the nature of integer-valued time series. Wang and Zhang (2011) considered three kinds of estimation methods, namely YW, CLS and WCLS. An advantage of these methods is that they do not require specifying the exact family of distribution for the process.

More recently, Bisaglia and Canale (2016) developed a forecasting procedure for count time series, forecasts are produced through a non-parametric Bayesian method, which revealed appealing results. Maiti et al. (2016) explored the usefulness of the standard Box-Jenkins' type AR(p) process for obtaining coherent forecasting from integer-valued time series. To make the forecasting values coherent, they have suggested the rounding operator (Kachour and Yao, 2009) on the forecasting values obtained from the estimated AR(p) model.

In general, detailed studies have been conducted not only on the formulation of models but also on properties, estimation, tests and asymptotic distributions of model estimators for different discrete marginal distributions. Regarding testing serial dependence in count data, a preliminary analysis should first consider independence (Jung and Tremayne, 2003) before fitting INAR models. A general test was derived by Sun and McCabe (2013) for independence in the classic binomial thinning INAR model with a particular feature, the support for the underlying arrivals process is not assumed to be known. Recent developments in tests for time series of counts can be found in Hudecová et al. (2015) and references therein.

1.5 Outline of the thesis

Within the reasonably large spectrum of integer-valued models proposed in the literature only a few focus on the modeling of univariate/multivariate time series of count data with periodic structure. Our aim in this thesis is to give a contribution towards this direction. We develop and study time series models of first-order adequate to describe periodic time series of count data. Focusing on the class of observation–driven models, we seek to extend integer-valued autoregressive 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. Moreover, specific parametric cases that arise from the assumption of a particular joint distribution for the innovation processes were studied in detail. Simulation studies were conducted and forecasting discussed. Applications to modeling time series of counts through the proposed models were also given.

Specifically, in Chapter 2 we generalized the results obtained by Pedeli and Karlis (2013a) to multivariate integer-valued models of first-order with periodic structure in the wide sense, i.e., with periodically varying mean and covariances. Throughout this chapter, the thinning operator considered was the matrix-binomial thinning operator defined in (1.13). Our interest in periodic integer-valued autoregressive models was primarily influenced by the work of Monteiro et al. (2010, 2015) whose periodic (univariate and bivariate) `INAR` models were introduced in recursions (1.20) and (1.22), respectively. We established the periodic m-variate integer-valued autoregressive process, denoted by `PMINAR`(1), with period s in its general matricial form, defined its basic statistical properties and proven its existence. The generalization to a multivariate setting is not straightforward since many computational issues arised especially for the estimation of the parameters. We derived Yule–Walker, conditional maximum likelihood and composite likelihood estimators for the unknown parameters of the proposed model and discussed their asymptotic properties. Particular attention was given to the special case that arises from specifying multivariate negative binomial distribution for the innovations of the `PMINAR`(1) process. This discrete multivariate distribution can account for overdispersion in contrast to the usual multivariate Poisson distribution. An important restriction of the Poisson distribution, as well known, is that its mean and variance are equal. However, in real-life events, the Poisson assumption is often violated, therefore this distribution is not considered in our work. To confront estimation problems due to the complexity of the maximum likelihood, we implemented the recently popular idea of composite likelihood approach (Varin et al., 2011). The basic advantage of this method is the replacement of the full likelihood with a pseudo-likelihood which effectively captures the model properties while at the same time is computationally less demanding, also used by Pedeli and Karlis (2013a). Forecasting is also addressed. The performances of the three aforementioned estimators were compared via a simulation experiment. Simulations were carried out in `R` and suitable parameter transformations were adopted. A real data set related with fire activity in Portugal was used to illustrate the proposed periodic multivariate integer-valued autoregressive model of order one for the trivariate case ($m = 3$) contemplating trivariate negative binomial innovations. However, the models in this chapter have some limitations. Thus, because of binomial thinning operators, all the coefficients of the models must be non-negative. Therefore, the modeling of series with possible negative autocorrelations are excluded. Moreover, these models are defined on $\mathbb{N}$, so they cannot fit a time series with negative observations nor negative correlation.

In Chapter 3, we developed two new integer-valued autoregressive models of first-order introducing time-varying parameters and sequences of innovations with periodic structure in a new framework, regarding the thinning operator and distributions for the innovations. Both proposed `INAR` models (univariate and bivariate) are based on the signed thinning operator defined in subsections 1.1.2 and 1.2.2, adapted to the periodic case accordingly. We provided basic notations and definitions concerning the (periodic) signed thinning operator as well as some of its properties. Before introducing the new models, we also provided a brief description on the (periodic) Skellam distribution for univariate and bivariate distributions defined on the set of integers. Extending the model in Chesneau and Kachour (2012) to the periodic case, we introduced a new univariate signed `INAR`(1) process, by considering a parametric assumption on the common distribution of the periodic counting sequence of the model. In this setting, the new signed periodic model was denoted by `S-PINAR`(1) with Skellam-distributed innovation term. In contrast to traditional `INAR`(1) models, these models are defined in $\mathbb{Z}$ allowing for negative correlation. Due to some limitations of the periodic signed thinning (which lacks the distributive property), only the conditional moments of first and second-order of the process were established. Regarding parameter estimation, two methods were considered: modified conditional least squares and conditional maximum likelihood. The conditional least squares method, first proposed by Klimko and Nelson (1978), was adapted by Alzaid and Omair (2014) with some modifications in order to be able to estimate all parameters integrating the periodic model. In order to study the performance of the proposed methods, an extended simulation experiment was carried out for the `S-PINAR`(1) model with period s. Numerical results from the simulation study suggested that the proposed model is suitable for practical use.

Motivated by the work of Bulla et al. (2016), we then generalized the `S-PINAR`(1) model to the bivariate case. The definition and matrix representation of the bivariate model denoted by `BS-PINAR`(1) with period s was presented and some statistical properties of the model were derived. The assumption of a diagonal autoregressive matrix was made, therefore, the correlation is achieved through their innovation processes, where the distribution of the innovation processes is set a priori which consequently determines the distribution of the underlying time series. Hence, the discrete bivariate distribution on $\mathbb{Z}^2$ assigned to the distribution of the innovations was the bivariate Skellam distribution. Parameter estimation of the unknown parameters was provided through the conditional maximum likelihood method.

Finally, main conclusions of this thesis and some challenges for future work are described in Chapter 4.

Chapter 2

PMINAR(1) model based on the binomial thinning operator

In this chapter, a multivariate first-order integer-valued autoregressive model with time-varying parameters and sequences of innovations having periodic structure in the wide sense, i.e., with periodically varying mean and covariances, is established. The model is based upon the matrix-binomial thinning operator defined in (1.13) and aims to extend the periodic bivariate `INAR`(1) model proposed in Monteiro et al. (2015) to the multi-dimensional space. Therefore, the periodic m-variate integer-valued autoregressive process, denoted by `PMINAR`(1) with period s is presented. The matricial form of the multivariate model and its basic statistical properties are defined. Yule–Walker, conditional maximum likelihood and composite likelihood estimators for the unknown parameters of the `PMINAR`(1) process are derived. Particular attention was given to the special case that arises from specifying multivariate negative binomial distribution for the innovations of the `PMINAR`(1) process. Furthermore, forecasting is also addressed. The performances of the three aforementioned estimators are compared via a simulation study. A real data set related with fire activity in Portugal is used to illustrate the proposed periodic multivariate `INAR`(1) model for the trivariate case ($m = 3$) contemplating trivariate negative binomial innovations.

2.1 Definition of the PMINAR(1) model

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

$$\mathbf{X}_t = A_t \circ \mathbf{X}_{t-1} + \mathbf{Z}_t,\ \ t \in \mathbb{Z}, \tag{2.1}$$

where $\mathbf{X}_t, \mathbf{X}_{t-1}$ and $\mathbf{Z}_t$ are random ms-vectors with $\mathbf{X}_t = [\mathbf{X}_{1,t}\ \mathbf{X}_{2,t}\ \ \ldots\ \ \mathbf{X}_{m,t}]^T$ for $t = v+ns$, $v = 1, \ldots, s$ and $n \in \mathbb{N}_0$, and $\mathbf{X}_{j,t} = [X_{j,1+ns}\ X_{j,2+ns}\ \ \ldots\ \ X_{j,s+ns}]^T, j = 1, \ldots, m$. The ms-dimensional vector $\mathbf{Z}_t = [\mathbf{Z}_{1,t}\ \mathbf{Z}_{2,t}\ \ \ldots\ \ \mathbf{Z}_{m,t}]^T$ constitutes a periodic sequence of independent random vectors with

$$\mathbf{Z}_{j,t} = [Z_{j,1+ns}\ Z_{j,2+ns}\ \ \ldots\ \ Z_{j,s+ns}]^T. \tag{2.2}$$

The model defined in (2.1) will be referred to as the `Periodic Multivariate INteger-valued AutoRegressive` model of order one (`PMINAR`(1) in short) with period $s \in \mathbb{N}$. The `PMINAR`(1) model admits the following matricial representation

$$\begin{bmatrix} \mathbf{X}_{1,t} \\ \mathbf{X}_{2,t} \\ \vdots \\ \mathbf{X}_{m,t} \end{bmatrix} = \begin{bmatrix} \phi_{1,t} & 0 & \cdots & 0 \\ 0 & \phi_{2,t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \phi_{m,t} \end{bmatrix} \circ \begin{bmatrix} \mathbf{X}_{1,t-1} \\ \mathbf{X}_{2,t-1} \\ \vdots \\ \mathbf{X}_{m,t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_{1,t} \\ \mathbf{Z}_{2,t} \\ \vdots \\ \mathbf{Z}_{m,t} \end{bmatrix} \tag{2.3}$$

with $\phi_{j,t} = \alpha_{j,v} \in (0,1)$ for $t = v + ns; v = 1, \ldots, s; n \in \mathbb{N}_0$ and $j = 1, \ldots, m$. The elements $\mathbf{Z}_{j,t}$ joining the system in the interval $(t-1, t]$ are usually referenced to as innovations. For each t, $\mathbf{Z}_{j,t}$ is assumed to be independent of $\mathbf{X}_{j,t-1}$ and $\phi_{j,t} \circ \mathbf{X}_{j,t-1}$. The matrix A_t in equation (2.1) is a $(ms \times ms)$ diagonal matrix, representing the periodic integer-valued autoregressive coefficients in season v $(v = 1, \ldots, s)$:

$$A_t = \begin{bmatrix} \phi_{1,t} & 0 & \cdots & 0 \\ 0 & \phi_{2,t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \phi_{m,t} \end{bmatrix}$$

where

$$\phi_{j,t} = \begin{cases} \alpha_{j,1}, & t = 1 + ns \\ \alpha_{j,2}, & t = 2 + ns \\ \alpha_{j,3}, & t = 3 + ns \\ \vdots & \\ \alpha_{j,s}, & t = s + ns \end{cases} \tag{2.4}$$

for $j = 1, \dots, m$ and $n \in \mathbb{N}_0$. Note that the j-th component $(j = 1, \dots, m)$ is

$$\mathbf{X}_{j,t} = \phi_{j,t} \circ \mathbf{X}_{j,t-1} + \mathbf{Z}_{j,t} \tag{2.5}$$

with

$$\phi_{j,t} \circ X_{j,t-1} \stackrel{d}{=} \sum_{r=1}^{X_{j,t-1}} U_{r,t}\left(\phi_{j,t}\right),$$

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 $P\left(U_{r,t}\left(\phi_{j,t}\right) = 1\right) = \phi_{j,t} = 1 - P\left(U_{r,t}\left(\phi_{j,t}\right) = 0\right)$. The operator $\circ$ corresponds to the binomial thinning operator defined in (1.13).

Since the autocorrelation matrix A_t is diagonal, the only source of dependence between the series $\mathbf{X}_{j,t}$ $(j = 1, \dots, m)$ in (2.3) is given through the vector of innovations $\mathbf{Z}_t$. Therefore, the innovations will play a central role in the specification of the `PMINAR`(1) process.

Considering the j-th component, $\mathbf{X}_{j,t} = [X_{j,1+ns} \ X_{j,2+ns} \ \dots \ X_{j,s+ns}]^T$ of $\mathbf{X}_t$ and by applying the recursive equation in (2.5) with coefficients $\phi_{j,t}$ in (2.4), it follows that

$$X_{j,1+ns} = \alpha_{j,1} \circ X_{j,1+ns-1} + Z_{j,1+ns} = \alpha_{j,1} \circ X_{j,s+(n-1)s} + Z_{j,1+ns}$$

$$\begin{aligned} X_{j,2+ns} &= \alpha_{j,2} \circ X_{j,1+ns} + Z_{j,2+ns} = \alpha_{j,2} \circ (\alpha_{j,1} \circ X_{j,s+(n-1)s} + Z_{j,1+ns}) + Z_{j,2+ns} = \\ &= (\alpha_{j,2}\alpha_{j,1}) \circ X_{j,s+(n-1)s} + \alpha_{j,2} \circ Z_{j,1+ns} + Z_{j,2+ns} = \\ &= \left(\prod_{k=0}^{2-1} \alpha_{j,2-k}\right) \circ X_{j,s+(n-1)s} + \left(\prod_{k=0}^{2-2} \alpha_{j,2-k}\right) \circ Z_{j,1+ns} + Z_{j,2+ns} \\ &\vdots \end{aligned}$$

$$X_{j,s+ns} = \left(\prod_{k=0}^{s-1} \alpha_{j,s-k}\right) \circ X_{j,s+(n-1)s} + \left(\prod_{k=0}^{s-2} \alpha_{j,s-k}\right) \circ Z_{j,1+ns} + \left(\prod_{k=0}^{s-3} \alpha_{j,s-k}\right) \circ Z_{j,2+ns} +$$
$$+ \cdots + \left(\prod_{k=0}^{s-s} \alpha_{j,s-k}\right) \circ Z_{j,s-1+ns} + Z_{j,s+ns} =$$
$$= \left(\prod_{k=0}^{s-1} \alpha_{j,s-k}\right) \circ X_{j,s+(n-1)s} + \sum_{l=1}^{s-1} \left(\prod_{k=0}^{l-1} \alpha_{j,s-k}\right) \circ Z_{j,s-l+ns} + Z_{j,s+ns}.$$

Hence, for $v = 1, \ldots, s$,

$$X_{j,v+ns} = \left(\prod_{k=0}^{v-1} \alpha_{j,v-k}\right) \circ X_{j,s+(n-1)s} + \sum_{l=1}^{v-1} \left(\prod_{k=0}^{l-1} \alpha_{j,v-k}\right) \circ Z_{j,v-l+ns} + Z_{j,v+ns},$$

which implies that $\mathbf{X}_{j,t} = \phi_{j,t} \circ \mathbf{X}_{j,t-1} + \mathbf{Z}_{j,t}$ in (2.5) with $t = v + ns; v = 1, \ldots, s$ and $n \in \mathbb{N}_0$ admits the matricial representation

$$\begin{bmatrix} X_{j,1+ns} \\ X_{j,2+ns} \\ X_{j,3+ns} \\ \vdots \\ X_{j,s+ns} \end{bmatrix} = \begin{bmatrix} 0 & \cdots & 0 & \alpha_{j,1} \\ 0 & \cdots & 0 & \alpha_{j,2}\alpha_{j,1} \\ 0 & \cdots & 0 & \alpha_{j,3}\alpha_{j,2}\alpha_{j,1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \prod_{k=0}^{s-1} \alpha_{j,s-k} \end{bmatrix} \circ \begin{bmatrix} X_{j,1+(n-1)s} \\ X_{j,2+(n-1)s} \\ X_{j,3+(n-1)s} \\ \vdots \\ X_{j,s+(n-1)s} \end{bmatrix} +$$

$$+ \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ \alpha_{j,2} & 1 & 0 & \ldots & 0 \\ \alpha_{j,3}\alpha_{j,2} & \alpha_{j,3} & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \prod_{k=0}^{s-2} \alpha_{j,s-k} & \prod_{k=0}^{s-3} \alpha_{j,s-k} & \prod_{k=0}^{s-4} \alpha_{j,s-k} & \ldots & 1 \end{bmatrix} \circ \begin{bmatrix} Z_{j,1+ns} \\ Z_{j,2+ns} \\ Z_{j,3+ns} \\ \vdots \\ Z_{j,s+ns} \end{bmatrix}.$$

Due to the fact that $t = v + ns$, then $\mathbf{X}_{j,t-s} = \mathbf{X}_{j,v+ns-s} = \mathbf{X}_{j,v+(n-1)s}$ $(v = 1, \ldots, s)$, meaning the j-th component $\mathbf{X}_{j,t}$ in equation (2.5) can be replaced by

$$\mathbf{X}_{j,t} = A_j \circ \mathbf{X}_{j,t-s} + B_j \circ \mathbf{Z}_{j,t}, \tag{2.6}$$

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

$$A_j = \begin{bmatrix} 0 & \cdots & 0 & \alpha_{j,1} \\ 0 & \cdots & 0 & \alpha_{j,2}\alpha_{j,1} \\ 0 & \cdots & 0 & \alpha_{j,3}\alpha_{j,2}\alpha_{j,1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \prod_{k=0}^{s-1} \alpha_{j,s-k} \end{bmatrix} \tag{2.7}$$

and

$$B_j = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ \alpha_{j,2} & 1 & 0 & \ldots & 0 \\ \alpha_{j,3}\alpha_{j,2} & \alpha_{j,3} & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \prod_{k=0}^{s-2} \alpha_{j,s-k} & \prod_{k=0}^{s-3} \alpha_{j,s-k} & \prod_{k=0}^{s-4} \alpha_{j,s-k} & \ldots & 1 \end{bmatrix}, \tag{2.8}$$

respectively, with coefficients $\alpha_{j,v} \in (0, 1)$, $j = 1, \ldots, m$ and $v = 1, \ldots, s$. All columns of matrices A_j, except the last one, are null. The matrices B_j are lower triangular matrices.

Taking all m components, the `PMINAR`(1) model defined in (2.1) can be rewritten in the form

$$\mathbf{X}_t = \widetilde{A} \circ \mathbf{X}_{t-s} + \widetilde{B} \circ \mathbf{Z}_t, \tag{2.9}$$

with matricial representation

$$\begin{bmatrix} \mathbf{X}_{1,t} \\ \mathbf{X}_{2,t} \\ \vdots \\ \mathbf{X}_{m,t} \end{bmatrix} = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_m \end{bmatrix} \circ \begin{bmatrix} \mathbf{X}_{1,t-s} \\ \mathbf{X}_{2,t-s} \\ \vdots \\ \mathbf{X}_{m,t-s} \end{bmatrix} + \begin{bmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_m \end{bmatrix} \circ \begin{bmatrix} \mathbf{Z}_{1,t} \\ \mathbf{Z}_{2,t} \\ \vdots \\ \mathbf{Z}_{m,t} \end{bmatrix}.$$

The $(ms \times ms)$ matrices $\widetilde{A}$ and $\widetilde{B}$ in equation (2.9) are block-diagonal matrices

$$\widetilde{A} = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_m \end{bmatrix} = diag(A_1, A_2, \ldots, A_m) \tag{2.10}$$

and

$$\widetilde{B} = \begin{bmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_m \end{bmatrix} = diag(B_1, B_2, \ldots, B_m) \tag{2.11}$$

with matrices A_j and B_j $(j = 1, \ldots, m)$ in (2.7) and in (2.8), respectively. Generally, matrix $\widetilde{A}$ has entries a^j_{ik} satisfying $0 \leq a^j_{ik} < 1$ and matrix $\widetilde{B}$ has entries b^j_{ik} satisfying $0 \leq b^j_{ik} \leq 1$ with $i, k = 1, \ldots, ms$ and $j = 1, \ldots, m$. Notice that the j-th component in equation (2.6) can also be written as

$$\mathbf{X}^*_{j,n} = A_j \circ \mathbf{X}^*_{j,n-1} + B_j \circ \mathbf{Z}^*_{j,n}, \tag{2.12}$$

where $\mathbf{X}^*_{j,n} = [X_{j,1+ns}\ X_{j,2+ns}\ \ldots\ X_{j,s+ns}]^T$, $\mathbf{Z}^*_{j,n} = [Z_{j,1+ns}\ Z_{j,2+ns}\ \ldots\ Z_{j,s+ns}]^T$ and also $\mathbf{X}^*_{j,n-1} = \left[X_{j,1+(n-1)s}\ X_{j,2+(n-1)s}\ \ldots\ X_{j,s+(n-1)s}\right]^T$. Hence, the corresponding periodic multivariate model is

$$\mathbf{X}^*_n = \widetilde{A} \circ \mathbf{X}^*_{n-1} + \widetilde{B} \circ \mathbf{Z}^*_n, \tag{2.13}$$

where $\mathbf{X}^*_n$, $\mathbf{X}^*_{n-1}$ and $\mathbf{Z}^*_n$ are ms-dimensional random vectors such as

$$\begin{aligned} \mathbf{X}^*_n &= \left[\mathbf{X}^*_{1,n}\ \mathbf{X}^*_{2,n}\ \ldots\ \mathbf{X}^*_{m,n}\right]^T = \\ &= [\overbrace{X_{1,1+ns} \ldots X_{1,s+ns}}^{\mathbf{X}^*_{1,n}}\ \ldots\ \overbrace{X_{m,1+ns} \ldots X_{m,s+ns}}^{\mathbf{X}^*_{m,n}}]^T \\ \mathbf{X}^*_{n-1} &= \left[\mathbf{X}^*_{1,n-1}\ \mathbf{X}^*_{2,n-1}\ \ldots\ \mathbf{X}^*_{m,n-1}\right]^T \\ \mathbf{Z}^*_n &= \left[\mathbf{Z}^*_{1,n}\ \mathbf{Z}^*_{2,n}\ \ldots\ \mathbf{Z}^*_{m,n}\right]^T. \end{aligned}$$

The model present in (2.13) is a periodic multivariate first-order integer-valued autoregressive model regarding the cycle, where $n-1$ represents the cycle preceding n.

Remark: The random vector $\mathbf{X}_t = [\mathbf{X}_{1,t}\ \mathbf{X}_{2,t}\ \ldots\ \mathbf{X}_{m,t}]^T, t = v+ns; v = 1,\ldots,s; n \in \mathbb{N}_0$ defined in (2.1) is the same as $\mathbf{X}_n^*$ in (2.13) because $\mathbf{X}_{j,n}^* = [X_{j,1+ns}\ \ldots\ X_{j,s+ns}]^T = \mathbf{X}_{j,t}$ and also $\mathbf{Z}_{j,n}^* = [Z_{j,1+ns}\ \ldots\ Z_{j,s+ns}]^T = \mathbf{Z}_{j,t}$ which leads to $\mathbf{Z}_t = [\mathbf{Z}_{1,t}\ \mathbf{Z}_{2,t}\ \ldots\ \mathbf{Z}_{m,t}]^T = \mathbf{Z}_n^*$.

As previously mentioned, the ms-dimensional random vector $\mathbf{Z}_t$ with vector $\mathbf{Z}_{j,t}$ in (2.2) for $t = v+ns$; $v = 1,\ldots,s$ and $n \in \mathbb{N}_0$ is a periodic sequence of independent random vectors. The innovations $\mathbf{Z}_t$ have (assumed) finite first and second-order moments:

- Mean vector of $\mathbf{Z}_t$, $E[\mathbf{Z}_t]$:

$$E[\mathbf{Z}_t] = E\begin{bmatrix} \mathbf{Z}_{1,t} \\ \mathbf{Z}_{2,t} \\ \vdots \\ \mathbf{Z}_{m,t} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\delta}_{1,t} \\ \boldsymbol{\delta}_{2,t} \\ \vdots \\ \boldsymbol{\delta}_{m,t} \end{bmatrix} = \boldsymbol{\delta}_t. \tag{2.14}$$

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

$$E[\mathbf{Z}_{j,t}] = \boldsymbol{\delta}_{j,t} = \begin{bmatrix} \lambda_{j,1} \\ \lambda_{j,2} \\ \vdots \\ \lambda_{j,s} \end{bmatrix}, \tag{2.15}$$

for $j = 1,\ldots,m$. For a fixed v, each element of vector (2.15) is

$$E[Z_{j,v+ns}] = \lambda_{j,v}. \tag{2.16}$$

- Variance-covariance matrix of $\mathbf{Z}_t$, $\sum_{\mathbf{Z}_t}$ (symmetric matrix):

$$\sum\nolimits_{\mathbf{Z}_t} = Var[\mathbf{Z}_t] =$$
$$= \begin{bmatrix} Var[\mathbf{Z}_{1,t}] & Cov(\mathbf{Z}_{1,t}, \mathbf{Z}_{2,t}) & \cdots & Cov(\mathbf{Z}_{1,t}, \mathbf{Z}_{m,t}) \\ Cov(\mathbf{Z}_{2,t}, \mathbf{Z}_{1,t}) & Var[\mathbf{Z}_{2,t}] & \cdots & Cov(\mathbf{Z}_{2,t}, \mathbf{Z}_{m,t}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\mathbf{Z}_{m,t}, \mathbf{Z}_{1,t}) & Cov(\mathbf{Z}_{m,t}, \mathbf{Z}_{2,t}) & \cdots & Var[\mathbf{Z}_{m,t}] \end{bmatrix} =$$
$$= \begin{bmatrix} \psi_{11,t} & \psi_{12,t} & \cdots & \psi_{1m,t} \\ & \psi_{22,t} & \cdots & \psi_{2m,t} \\ & & \ddots & \vdots \\ & & & \psi_{mm,t} \end{bmatrix} = \psi_t, \tag{2.17}$$

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

$$\psi_{jk,t} = Cov(\mathbf{Z}_{j,t}, \mathbf{Z}_{k,t}) = \begin{bmatrix} \sigma_{jk,1} & 0 & \ldots & 0 \\ 0 & \sigma_{jk,2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_{jk,s} \end{bmatrix}. \tag{2.18}$$

For a fixed v, each element of the diagonal in matrix (2.18) is given by

$$Cov(Z_{j,v+ns}, Z_{k,v+ns}) = \sigma_{jk,v}. \tag{2.19}$$

For notational simplicity, we use $\sigma^2_{j,t}$ instead of $\sigma_{jj,t}$ when $j = k$ $(j = 1,\ldots,m)$ and for $t = v + ns; v = 1,\ldots,s$:

$$\psi_{jj,t} = Var[\mathbf{Z}_{j,t}] = \begin{bmatrix} \sigma^2_{j,1} & 0 & \ldots & 0 \\ 0 & \sigma^2_{j,2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma^2_{j,s} \end{bmatrix}. \tag{2.20}$$

For a fixed v, each element of the diagonal in matrix (2.20) is given by

$$Var[Z_{j,v+ns}] = \sigma^2_{j,v}. \tag{2.21}$$

The $(ms \times ms)$ matrix ψ_t in (2.17) has m on-diagonal matrices equal to $\psi_{jj,t} = Var[\mathbf{Z}_{j,t}]$ in (2.20) and $(m-1)m$ off-diagonal matrices equal to $\psi_{jk,t} = Cov(\mathbf{Z}_{j,t}, \mathbf{Z}_{k,t})$ in (2.18) with $j \neq k$; $j, k = 1, \ldots, m$.

2.2 Properties of the PMINAR(1) model

2.2.1 Strictly periodically stationary distribution

Let `PMINAR`(1) be the process defined in (2.9). Within this setting, it can be proven that a strictly periodically stationary `INAR` process satisfying (2.9) exists based upon the results provided in Franke and Subba Rao (1993). The existence of a periodically stationary solution of (2.9) depends on the largest eigenvalue of the non-negative matrix $\widetilde{A}$ in (2.10), whose coefficients $\alpha_{j,v} \in (0,1)$ for all components. Take the $(ms \times ms)$ block-diagonal matrix $\lambda I - \widetilde{A}$, where I denotes the identity matrix as usual, then

$$\lambda I - \widetilde{A} = diag(C_1, C_2, \ldots, C_m)$$

with $(s \times s)$ matrix C_j $(j = 1, \ldots, m)$ defined by

$$C_j = \begin{bmatrix} \lambda & 0 & \cdots & 0 & -\alpha_{j,1} \\ 0 & \lambda & \cdots & 0 & -\alpha_{j,2}\alpha_{j,1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda & -\prod_{k=0}^{s-2} \alpha_{j,s-1-k} \\ 0 & 0 & \cdots & 0 & \lambda - \prod_{k=0}^{s-1} \alpha_{j,s-k} \end{bmatrix}. \tag{2.22}$$

The determinant of the matrix $\lambda I - \widetilde{A}$ denoted by $det(\lambda I - \widetilde{A})$ can easily be determined since the matrices C_j ($j = 1, \ldots, m$) are upper triangular matrices (Harville, 2008). The characteristic polynomial of $\widetilde{A}$ is

$$det(\lambda I - \widetilde{A}) = \left(\lambda^{s-1}\right)^m \prod_{j=1}^{m} \left(\lambda - \prod_{k=0}^{s-1} \alpha_{j,s-k}\right).$$

For convenience in notation, let $\prod_{k=0}^{s-1} \alpha_{j,s-k} = T_j$. The polynomial takes the form

$$det(\lambda I - \widetilde{A}) = \lambda^{ms}\lambda^{-m} \prod_{j=1}^{m} (\lambda - T_j) = \lambda^{ms} + \sum_{j=1}^{m} (-1)^j \beta_j \lambda^{ms-j}$$

with coefficients β_j ($j = 1, \ldots, m$) defined as

- $\beta_1 = \sum_{j=1}^{m} T_j$,
- $\beta_2 = \sum_{j=1}^{m-1} \sum_{i=j+1}^{m} T_j T_i$,
- $\beta_3 = \sum_{j=1}^{m-2} \sum_{i=j+1}^{m-1} \sum_{k=i+1}^{m} T_j T_i T_k$,

  $\vdots$
- $\beta_{m-1} = \sum_{j=1}^{m} \prod_{\substack{i=1 \\ i \neq j}}^{m} T_i$,
- $\beta_m = \prod_{j=1}^{m} T_j$.

Let ρ be the maximal eigenvalue of $\widetilde{A}$, then by Proposition B in Dion et al. (1995), $\sum_{j=1}^{m} \beta_j < 1$ if and only if $\rho < 1$.

Lemma 2.1. *For a fixed v ($v = 1, \ldots, s$), $\alpha_{j,v} \in (0, 1)$ with $j = 1, \ldots, m$ and for $t = v + ns$, $0 < P(\mathbf{Z}_t = \underline{0}) < 1$. Then, any solution of process $\{\mathbf{X}_t\}$, $t = v + ns$ and $n \in \mathbb{N}_0$ in (2.9) is an irreducible and aperiodic Markov chain.*

Proof. Let $\underline{r} = [\underline{r}_1\ \underline{r}_2\ \dots\ \underline{r}_m]^T$ with $\underline{r}_j = [r_{j1}\ r_{j2}\ \dots\ r_{js}]$ and $\underline{d} = [\underline{d}_1\ \underline{d}_2\ \dots\ \underline{d}_m]^T$ with $\underline{d}_j = [d_{j1}\ d_{j2}\ \dots\ d_{js}]$ for each $j = 1, \dots, m$.

$$P_{\underline{r},\underline{d}} = P(\mathbf{X}_t = \underline{r}|\mathbf{X}_{t-s} = \underline{d}) = P(\widetilde{A} \circ \mathbf{X}_{t-s} + \widetilde{B} \circ \mathbf{Z}_t = \underline{r}|\mathbf{X}_{t-s} = \underline{d}) =$$

$$= P\left(\begin{bmatrix} A_1 \circ \mathbf{X}_{1,t-s} + B_1 \circ \mathbf{Z}_{1,t} \\ A_2 \circ \mathbf{X}_{2,t-s} + B_2 \circ \mathbf{Z}_{2,t} \\ \dots \\ A_m \circ \mathbf{X}_{m,t-s} + B_m \circ \mathbf{Z}_{m,t} \end{bmatrix} = \underline{r} \middle| \mathbf{X}_{t-s} = \underline{d}\right) =$$

$$= \sum_{j=1}^{m} \left(\sum_{i_{j1}=0}^{d_{j1}} \sum_{i_{j2}=0}^{d_{j2}} \cdots \sum_{i_{js}=0}^{d_{js}} \left[\prod_{v=1}^{s} P(Z_{j,v+ns} = i_{jv}) P\left(\prod_{k=0}^{v-1} \alpha_{j,v-k} \circ X_{m,s+(n-1)s} + \right.\right.\right.$$

$$\left.\left.\left. + \sum_{l=1}^{v-1} \left(\prod_{k=0}^{l-1} \alpha_{j,v-k} \right) \circ Z_{j,v-l+ns} = r_{jv} - i_{jv} | Z_{j,1+ns} = i_{j1}, Z_{j,2+ns} = i_{j2}, \dots, Z_{j,s+ns} = i_{js} \right) \right] \right)$$

$$\geq \sum_{j=1}^{m} \left(\prod_{v=1}^{s} P(Z_{j,v+ns} = r_{jv}) P\left(\prod_{k=0}^{v-1} \alpha_{j,v-k} \circ d_{js} + \right.\right.$$

$$\left.\left. + \sum_{l=1}^{v-1} \left(\prod_{k=0}^{l-1} \alpha_{j,v-k} \right) \circ Z_{j,v-l+ns} = 0 | Z_{j,1+ns} = r_{j1}, Z_{j,2+ns} = r_{j2}, \dots, Z_{j,s+ns} = r_{js} \right) \right)$$

$$\geq \sum_{j=1}^{m} \left(\prod_{v=1}^{s} P(Z_{j,v+ns} = r_{jv}) \left(1 - \prod_{k=0}^{v-1} \alpha_{j,v-k} \right)^{d_{js}} \times \right.$$

$$\left. \times \prod_{k=1}^{v-1} P\left(\left(\prod_{k=0}^{v-1} \alpha_{j,v-k} \right) \circ Z_{j,v-l+ns} = 0 | Z_{j,1+ns} = r_{j1}, Z_{j,2+ns} = r_{j2}, \dots, Z_{j,s+ns} = r_{js} \right) \right)$$

$$\geq \sum_{j=1}^{m} \left(\prod_{v=1}^{s} P(Z_{j,v+ns} = r_{jv}) \left(1 - \prod_{k=0}^{v-1} \alpha_{j,v-k} \right)^{d_{js}} \prod_{k=1}^{v-1} \left(1 - \prod_{k=0}^{v-1} \alpha_{j,v-k} \right)^{r_{jv-k}} \right) > 0.$$

Therefore,

$$P_{\underline{0},\underline{d}} = P(\mathbf{X}_t = \underline{0}|\mathbf{X}_{t-s} = \underline{d}) = P(\widetilde{A} \circ \mathbf{X}_{t-s} + \widetilde{B} \circ \mathbf{Z}_t = \underline{0}|\mathbf{X}_{t-s} = \underline{d}) =$$
$$= \sum_{j=1}^{m} \left(\prod_{v=1}^{s} P(Z_{j,v+ns} = 0) \left(1 - \prod_{k=0}^{v-1} \alpha_{j,v-k} \right)^{d_{js}} \right) > 0$$

and similarly $P_{\underline{r},\underline{0}} = P(\mathbf{X}_t = \underline{r}|\mathbf{X}_{t-s} = \underline{0}) \ > 0$ implying that $\{\mathbf{X}_t\}$ is irreducible.

Moreover,

$$\begin{aligned} P_{\underline{0},\underline{0}} &= P(\mathbf{X}_t = \underline{0}|\mathbf{X}_{t-s} = \underline{0}) = P(\widetilde{A} \circ \mathbf{X}_{t-s} + \widetilde{B} \circ \mathbf{Z}_t = \underline{0}|\mathbf{X}_{t-s} = \underline{0}) = \\ &= \sum_{j=1}^{m} \prod_{v=1}^{s} P(Z_{j,v+ns} = 0) > 0, \end{aligned}$$

which implies that for a fixed v $(v = 1, \ldots, s)$, process $\{\mathbf{X}_t\}$ with $t = v + ns$ and $n \in \mathbb{N}_0$ is an aperiodic Markov chain. □

Theorem 2.1. *(Strictly periodically stationary distribution)*
For a fixed v $(v = 1, \ldots, s)$, let $\{\mathbf{X}_t\}$ with $t = v + ns$ and $n \in \mathbb{N}_0$ satisfying (2.9) 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 (2.9).*

Proof. From Lemma 2.1, $\{\mathbf{X}_t\}$ with $t = v + ns$ and fixed $v = 1, \ldots, s$ is an irreducible and aperiodic Markov chain. The eigenvalues of matrix $\widetilde{A}$ are less than one (Dion et al., 1995). Thus, by Franke and Subba Rao (1993), a strictly periodically stationary m-variate non-negative integer-valued process satisfying the equation (2.9) exists. □

The `PMINAR`(1) model in (2.9) can be expressed as

$$\mathbf{X}_t = \widetilde{A} \circ \mathbf{X}_{t-s} + \mathbf{R}_t, \tag{2.23}$$

where $\mathbf{R}_t = \widetilde{B} \circ \mathbf{Z}_t$ with matrix $\widetilde{B}$ in (2.11). Let

$$\mathbf{R}_t = [\mathbf{R}_{1,t} \ \mathbf{R}_{2,t} \ \ldots \ \mathbf{R}_{m,t}]^T = [B_1 \circ \mathbf{Z}_{1,t} \ B_2 \circ \mathbf{Z}_{2,t} \ \ldots \ B_m \circ \mathbf{Z}_{m,t}]^T \tag{2.24}$$

with $\mathbf{Z}_{j,t}$ in (2.2) for $j = 1, \ldots, m$. The innovation series $\{\mathbf{R}_t\}$ is a sequence of independent non-negative integer-valued random vectors with periodic structure:

$$P\left(\mathbf{R}_t = \underline{k}\right) = P\left(\widetilde{B} \circ \mathbf{Z}_t = \underline{k}\right) =$$
$$= P\left(B_1 \circ \mathbf{Z}_{1,t} = \underline{k_1}, B_2 \circ \mathbf{Z}_{2,t} = \underline{k_2}, \ldots, B_m \circ \mathbf{Z}_{m,t} = \underline{k_m}\right) =$$

$$= P\left(\begin{array}{l} Z_{1,1+ns} = k_{11}; \alpha_{1,2} \circ Z_{1,1+ns} + Z_{1,2+ns} = k_{12}; (\alpha_{1,3}\alpha_{1,2}) \circ Z_{1,1+ns} + \alpha_{1,3} \circ Z_{1,2+ns} + \\ + Z_{1,3+ns} = k_{13}; \cdots ; \sum\limits_{v=1}^{s-1} \left(\prod\limits_{l=0}^{v-1} \alpha_{1,s-l}\right) \circ Z_{1,s-v+ns} + Z_{1,s+ns} = k_{1s}; \cdots ; \\ Z_{m,1+ns} = k_{m1}; \alpha_{m,2} \circ Z_{m,1+ns} + Z_{m,2+ns} = k_{m2}; (\alpha_{m,3}\alpha_{m,2}) \circ Z_{m,1+ns} + \\ + \alpha_{m,3} \circ Z_{m,2+ns} + Z_{m,3+ns} = k_{m3}; \cdots ; \sum\limits_{v=1}^{s-1} \left(\prod\limits_{l=0}^{v-1} \alpha_{m,s-l}\right) \circ Z_{m,s-v+ns} + \\ + Z_{m,s+ns} = k_{ms} \end{array}\right)$$

$$= P\left(\begin{array}{l} Z_{1,1} = k_{11}; \alpha_{1,2} \circ Z_{1,1} + Z_{1,2} = k_{12}; (\alpha_{1,3}\alpha_{1,2}) \circ Z_{1,1} + \alpha_{1,3} \circ Z_{1,2} + \\ + Z_{1,3} = k_{13}; \cdots ; \sum\limits_{v=1}^{s-1} \left(\prod\limits_{l=0}^{v-1} \alpha_{1,s-l}\right) \circ Z_{1,s-v} + Z_{1,s} = k_{1s}; \cdots ; \\ Z_{m,1} = k_{m1}; \alpha_{m,2} \circ Z_{m,1} + Z_{m,2} = k_{m2}; (\alpha_{m,3}\alpha_{m,2}) \circ Z_{m,1} + \\ + \alpha_{m,3} \circ Z_{m,2} + Z_{m,3} = k_{m3}; \cdots ; \sum\limits_{v=1}^{s-1} \left(\prod\limits_{l=0}^{v-1} \alpha_{m,s-l}\right) \circ Z_{m,s-v} + Z_{m,s} = k_{ms} \end{array}\right)$$

$$= P\left(\begin{array}{l} Z_{1,1+hs} = k_{11}; \alpha_{1,2} \circ Z_{1,1+hs} + Z_{1,2+hs} = k_{12}; (\alpha_{1,3}\alpha_{1,2}) \circ Z_{1,1+hs} + \alpha_{1,3} \circ Z_{1,2+hs} + \\ + Z_{1,3+hs} = k_{13}; \cdots ; \sum\limits_{v=1}^{s-1} \left(\prod\limits_{l=0}^{v-1} \alpha_{1,s-l}\right) \circ Z_{1,s-v+hs} + Z_{1,s+hs} = k_{1s}; \cdots ; \\ Z_{m,1+hs} = k_{m1}; \alpha_{m,2} \circ Z_{m,1+hs} + Z_{m,2+hs} = k_{m2}; (\alpha_{m,3}\alpha_{m,2}) \circ Z_{m,1+hs} + \\ \alpha_{m,3} \circ Z_{m,2+hs} + Z_{m,3+hs} = k_{m3}; \cdots ; \sum\limits_{v=1}^{s-1} \left(\prod\limits_{l=0}^{v-1} \alpha_{m,s-l}\right) \circ Z_{m,s-v+hs} + \\ + Z_{m,s+hs} = k_{ms} \end{array}\right)$$

$$= P\left(B_1 \circ \mathbf{Z}_{1,h} = \underline{k_1}, B_2 \circ \mathbf{Z}_{2,h} = \underline{k_2}, \ldots, B_m \circ \mathbf{Z}_{m,h} = \underline{k_m}\right) =$$
$$= P\left(\widetilde{B} \circ \mathbf{Z}_h = \underline{k}\right) = P\left(\mathbf{R}_h = \underline{k}\right).$$

Next we obtain the stationary mean and the variance-covariance matrix of the process $\{\mathbf{X}_t\}$ with $t = v + ns$ for each season v $(v = 1, \ldots, s)$.

2.2.2 Mean vector of cyclostationary PMINAR(1)

The properties of the matrix-binomial thinning operator established in Lemma 1.3 are useful to the derivation of the moments of the PMINAR(1) model in (2.23). Hence, from property 1 (Lemma 1.3), the mean expectation of $\mathbf{R}_t$ is

$$E[\mathbf{R}_t] = E[\widetilde{B} \circ \mathbf{Z}_t] = \widetilde{B} E[\mathbf{Z}_t] = \widetilde{B} \boldsymbol{\delta}_t \tag{2.25}$$

with matrices $\widetilde{B}$ and $\boldsymbol{\delta}_t$ in (2.11) and (2.14), respectively. Furthermore, for each component $j = 1, \ldots, m$, the mean vector of $\mathbf{R}_{j,t}$ takes the form

$$\begin{aligned} E[\mathbf{R}_{j,t}] &= E[B_j \circ \mathbf{Z}_{j,t}] = B_j E[\mathbf{Z}_{j,t}] = B_j \boldsymbol{\delta}_{j,t} = \\ &= \begin{bmatrix} \lambda_{j,1} \\ \lambda_{j,1}\alpha_{j,2} + \lambda_{j,2} \\ \lambda_{j,1}\alpha_{j,3}\alpha_{j,2} + \lambda_{j,2}\alpha_{j,3} + \lambda_{j,3} \\ \vdots \\ \lambda_{j,1} \prod_{k=0}^{s-2} \alpha_{j,s-k} + \lambda_{j,2} \prod_{k=0}^{s-3} \alpha_{j,s-k} + \cdots + \lambda_{j,s-1}\alpha_{j,s} + \lambda_{j,s} \end{bmatrix}, \end{aligned} \tag{2.26}$$

where B_j and $\boldsymbol{\delta}_{j,t}$ are defined in (2.8) and (2.15), respectively.

Let $\boldsymbol{\mu}_t = E[\mathbf{X}_t]$ with $\mathbf{X}_t$ given in equation (2.23), then

$$\boldsymbol{\mu}_t = E[\widetilde{A} \circ \mathbf{X}_{t-s} + \mathbf{R}_t] = \widetilde{A} E[\mathbf{X}_{t-s}] + E[\mathbf{R}_t].$$

Due to the periodically stationary distribution and from (2.25) we can write

$$(I - \widetilde{A})\boldsymbol{\mu}_t = \widetilde{B} \boldsymbol{\delta}_t$$

i.e.,

$$\boldsymbol{\mu}_t = (I - \widetilde{A})^{-1} \widetilde{B} \boldsymbol{\delta}_t \tag{2.27}$$

with I the identity matrix as usual, matrices $\widetilde{A}$ and $\widetilde{B}$, and vector $\boldsymbol{\delta_t}$ in (2.10), (2.11) and (2.14), respectively. Next we prove that $I-\widetilde{A}$ is a regular matrix and therefore, matrix $(I-\widetilde{A})^{-1}$ exists. The matrix $I-\widetilde{A}$ is a $(ms \times ms)$ block-diagonal matrix given by

$$I-\widetilde{A} = diag(C_1, C_2, \ldots, C_m)$$

with $(s \times s)$ matrix C_j $(j = 1, \ldots, m)$ as

$$C_j = \begin{bmatrix} 1 & 0 & \cdots & 0 & -\alpha_{j,1} \\ 0 & 1 & \cdots & 0 & -\alpha_{j,2}\alpha_{j,1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -\prod\limits_{k=0}^{s-2}\alpha_{j,s-1-k} \\ 0 & 0 & \cdots & 0 & 1-\prod\limits_{k=0}^{s-1}\alpha_{j,s-k} \end{bmatrix}.$$

This matrix is the same as C_j with $\lambda = 1$ defined in (2.22). The determinant of the matrix $I-\widetilde{A}$ is easy to determine since the matrices C_j above are $(s \times s)$ upper triangular matrices leading to obtain

$$det(I-\widetilde{A}) = \prod_{j=1}^{m}\left(1-\prod_{k=0}^{s-1}\alpha_{j,s-k}\right). \tag{2.28}$$

The determinant is different from zero because $\prod\limits_{k=0}^{s-1}\alpha_{j,s-k}$ is different from 1 since $\alpha_{j,v} \in (0,1)$ for $j = 1, \ldots, m$ and $v = 1, \ldots, s$. The adjoint matrix of $I-\widetilde{A}$ is

$$adj(I-\widetilde{A}) = \begin{bmatrix} F_1 & 0 & \cdots & 0 \\ 0 & F_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & F_m \end{bmatrix}.$$

Let $det(I-\widetilde{A})$ in (2.28) be d and the product $\prod\limits_{\substack{r=1\\r\neq j}}^{m}\left(1-\prod\limits_{k=0}^{s-1}\alpha_{r,s-k}\right)$ be $d_{(-j)}$ $(j = 1, \ldots, m)$.

The $(s \times s)$ matrix F_j $(j = 1, \ldots, m)$ is given by

$$F_j = \begin{bmatrix} d & 0 & \cdots & 0 & \alpha_{j,1} d_{(-j)} \\ 0 & d & \cdots & 0 & \prod_{k=0}^{2-1} \alpha_{j,2-k} d_{(-j)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & d & \prod_{k=0}^{(s-1)-1} \alpha_{j,s-1-k} d_{(-j)} \\ 0 & 0 & \cdots & 0 & d_{(-j)} \end{bmatrix}.$$

Since by definition, $(I - \widetilde{A})^{-1} = \frac{1}{det(I-\widetilde{A})} adj(I - \widetilde{A})$, the inverse matrix of $I - \widetilde{A}$ is

$$\begin{aligned} (I - \widetilde{A})^{-1} &= \frac{1}{\prod_{j=1}^{m} \left(1 - \prod_{k=0}^{s-1} \alpha_{j,s-k}\right)} \begin{bmatrix} F_1 & 0 & \cdots & 0 \\ 0 & F_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & F_m \end{bmatrix} \\ &= \frac{1}{d} \begin{bmatrix} F_1 & 0 & \cdots & 0 \\ 0 & F_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & F_m \end{bmatrix} = \begin{bmatrix} G_1 & 0 & \cdots & 0 \\ 0 & G_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & G_m \end{bmatrix}, \end{aligned}$$

where

$$G_j = \frac{1}{d} F_j = \begin{bmatrix} 1 & 0 & \cdots & 0 & \frac{d_{(-j)}}{d} \alpha_{j,1} \\ 0 & 1 & \cdots & 0 & \frac{d_{(-j)}}{d} \prod_{k=0}^{2-1} \alpha_{j,2-k} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \frac{d_{(-j)}}{d} \prod_{k=0}^{(s-1)-1} \alpha_{j,s-1-k} \\ 0 & 0 & \cdots & 0 & \frac{d_{(-j)}}{d} \end{bmatrix} \tag{2.29}$$

is an upper triangular matrix with $\dfrac{d_{(-j)}}{d} = \dfrac{1}{1 - \prod_{k=0}^{s-1} \alpha_{j,s-k}}$ for $j = 1, \ldots, m$.

The ms-dimensional mean vector $\boldsymbol{\mu}_t = (I - \widetilde{A})^{-1}\widetilde{B}\boldsymbol{\delta}_t$ with $t = v + ns$; $v = 1, \ldots, s$ and $n \in \mathbb{N}_0$ defined in (2.27) takes the form

$$\boldsymbol{\mu}_t = \begin{bmatrix} \boldsymbol{\mu}_{1,t} \\ \boldsymbol{\mu}_{2,t} \\ \vdots \\ \boldsymbol{\mu}_{m,t} \end{bmatrix} = \begin{bmatrix} G_1 & 0 & \cdots & 0 \\ 0 & G_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & G_m \end{bmatrix} \begin{bmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_m \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta}_{1,t} \\ \boldsymbol{\delta}_{2,t} \\ \vdots \\ \boldsymbol{\delta}_{m,t} \end{bmatrix}$$

and the mean of the j-th component $\mathbf{X}_{j,t}$:

$$\boldsymbol{\mu}_{j,t} = E\left[\mathbf{X}_{j,t}\right] = G_j B_j \boldsymbol{\delta}_{j,t} = \begin{bmatrix} E[X_{j,1+ns}] \\ E[X_{j,2+ns}] \\ \vdots \\ E[X_{j,s+ns}] \end{bmatrix}, \tag{2.30}$$

i.e.,

$$\boldsymbol{\mu}_{j,t} = G_j B_j \boldsymbol{\delta}_{j,t} =$$

$$= \begin{bmatrix} 1 & 0 & \cdots & 0 & \frac{d_{(-j)}}{d}\alpha_{j,1} \\ 0 & 1 & \cdots & 0 & \frac{d_{(-j)}}{d}\prod_{k=0}^{2-1}\alpha_{j,2-k} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \frac{d_{(-j)}}{d}\prod_{k=0}^{(s-1)-1}\alpha_{j,s-1-k} \\ 0 & 0 & \cdots & 0 & \frac{d_{(-j)}}{d} \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ \alpha_{j,2} & 1 & \cdots & 0 & 0 \\ \alpha_{j,3}\alpha_{j,2} & \alpha_{j,3} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \prod_{k=0}^{s-2}\alpha_{j,s-k} & \prod_{k=0}^{s-3}\alpha_{j,s-k} & \cdots & \alpha_{j,s} & 1 \end{bmatrix} \begin{bmatrix} \lambda_{j,1} \\ \lambda_{j,2} \\ \lambda_{j,3} \\ \vdots \\ \lambda_{j,s-1} \\ \lambda_{j,s} \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_{j,1} + \frac{d_{(-j)}}{d}\alpha_{j,1}\left(\lambda_{j,1}\prod_{k=0}^{s-2}\alpha_{j,s-k} + \lambda_{j,2}\prod_{k=0}^{s-3}\alpha_{j,s-k} + \cdots + \lambda_{j,s-1}\alpha_{j,s} + \lambda_{j,s}\right) \\ \lambda_{j,1}\alpha_{j,2} + \lambda_{j,2} + \frac{d_{(-j)}}{d}\prod_{k=0}^{2-1}\alpha_{j,2-k}\left(\lambda_{j,1}\prod_{k=0}^{s-2}\alpha_{j,s-k} + \lambda_{j,2}\prod_{k=0}^{s-3}\alpha_{j,s-k} + \cdots + \lambda_{j,s-1}\alpha_{j,s} + \lambda_{j,s}\right) \\ \vdots \\ \frac{d_{(-j)}}{d}\left(\lambda_{j,1}\prod_{k=0}^{s-2}\alpha_{j,s-k} + \lambda_{j,2}\prod_{k=0}^{s-3}\alpha_{j,s-k} + \cdots + \lambda_{j,s-1}\alpha_{j,s} + \lambda_{j,s}\right) \end{bmatrix}.$$

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

$$\varphi_{l,i}^{(j)} = \begin{cases} \prod_{k=0}^{i-1} \alpha_{j,l-k} & , i \geq 1 \\ 1 & , i = 0 \end{cases} . \tag{2.31}$$

For a fixed v ($v = 1, \ldots, s$) and $j = 1, \ldots, m$, each element of vector (2.30) is given by

$$E[X_{j,v+ns}] = \frac{\sum_{k=0}^{v-1} \varphi_{v,k}^{(j)} \lambda_{j,v-k} + \varphi_{v,v}^{(j)} \sum_{i=0}^{s-(v+1)} \varphi_{s,i}^{(j)} \lambda_{j,s-i}}{1 - \varphi_{s,s}^{(j)}}. \tag{2.32}$$

In the sequel, we adopt the convention $\sum_{i=0}^{s-(s+1)} \varphi_{s,i}^{(j)} \lambda_{j,s-i} = 0$. Moreover, the vector of expectations in (2.30) can be written as

$$\boldsymbol{\mu}_{j,t} = G_j B_j \boldsymbol{\delta}_{j,t} = \frac{1}{1 - \varphi_{s,s}^{(j)}} \begin{bmatrix} \sum_{k=0}^{1-1} \varphi_{1,k}^{(j)} \lambda_{j,1-k} + \varphi_{1,1}^{(j)} \sum_{i=0}^{s-2} \varphi_{s,i}^{(j)} \lambda_{j,s-i} \\ \sum_{k=0}^{2-1} \varphi_{2,k}^{(j)} \lambda_{j,2-k} + \varphi_{2,2}^{(j)} \sum_{i=0}^{s-3} \varphi_{s,i}^{(j)} \lambda_{j,s-i} \\ \sum_{k=0}^{3-1} \varphi_{3,k}^{(j)} \lambda_{j,3-k} + \varphi_{3,3}^{(j)} \sum_{i=0}^{s-4} \varphi_{s,i}^{(j)} \lambda_{j,s-i} \\ \vdots \\ \sum_{k=0}^{s-1} \varphi_{s,k}^{(j)} \lambda_{j,s-k} + \varphi_{s,s}^{(j)} \sum_{i=0}^{s-(s+1)} \varphi_{s,i}^{(j)} \lambda_{j,s-i} \end{bmatrix} \tag{2.33}$$

for $j = 1, \ldots, m;\ t = v + ns;\ v = 1, \ldots, s$ and $n \in \mathbb{N}_0$. Full details in Appendix B.1.

2.2.3 Variance-covariance matrix and auto-covariance function

The variance-covariance matrix of the periodic sequence $\{\mathbf{R}_t\}$ of independent random vectors is

$$\begin{aligned}
\sum\nolimits_{\mathbf{R}_t} &= Var[\mathbf{R}_t] = Var[\widetilde{B} \circ \mathbf{Z}_t] = \\
&= Var[E(\widetilde{B} \circ \mathbf{Z}_t|\mathbf{Z}_t)] + E[Var(\widetilde{B} \circ \mathbf{Z}_t|\mathbf{Z}_t)] = \\
&= Var[\widetilde{B}\mathbf{Z}_t] + diag(QE(\mathbf{Z}_t)) = \\
&= \widetilde{B} \sum\nolimits_{\mathbf{Z}_t} \widetilde{B}^T + diag(Q\boldsymbol{\delta}_t) = \\
&= \widetilde{B}\psi_t\widetilde{B}^T + diag(Q\boldsymbol{\delta}_t)
\end{aligned} \tag{2.34}$$

with matrices $\widetilde{B}$, $\boldsymbol{\delta}_t$ and ψ_t in (2.11), (2.14) and (2.17), respectively. The $(ms \times ms)$ variance matrix $Q = \widetilde{B}(I - \widetilde{B})$ has entries $[q_{ik}^j]_{i,k=1,\ldots,ms}$ for component $j = 1, \ldots, m$ (see property 2 of Lemma 1.3). In this case, $[q_{ik}^j] = [b_{ik}^j(1 - b_{ik}^j)]$ with b_{ik}^j elements of matrix $\widetilde{B}$ in (2.11). Thus matrix Q is also block-diagonal with m $(s \times s)$ matrices Q_j, i.e., $Q = diag(Q_1, Q_2, \ldots, Q_m)$, where

$$Q_j = \begin{bmatrix}
0 & 0 & 0 & \ldots & 0 \\
\alpha_{j,2}(1-\alpha_{j,2}) & 0 & 0 & \ldots & 0 \\
\alpha_{j,3}\alpha_{j,2}(1-\alpha_{j,3}\alpha_{j,2}) & \alpha_{j,3}(1-\alpha_{j,3}) & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\prod_{k=0}^{s-2}\alpha_{j,s-k}\left(1-\prod_{k=0}^{s-2}\alpha_{j,s-k}\right) & \prod_{k=0}^{s-3}\alpha_{j,s-k}\left(1-\prod_{k=0}^{s-3}\alpha_{j,s-k}\right) & \prod_{k=0}^{s-4}\alpha_{j,s-k}\left(1-\prod_{k=0}^{s-4}\alpha_{j,s-k}\right) & \ldots & 0
\end{bmatrix} \tag{2.35}$$

leading to

$$diag(Q\boldsymbol{\delta}_t) = \begin{bmatrix}
Q_1^* & 0 & \cdots & 0 \\
0 & Q_2^* & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & Q_m^*
\end{bmatrix}$$

with $Q_j^* = diag(Q_j\boldsymbol{\delta}_{j,t})$, matrix Q_j in (2.35) and vector $\boldsymbol{\delta}_{j,t}$ in (2.15) for $j = 1, \dots, m$ and $t = v + ns;\ v = 1, \dots, s$ and $n \in \mathbb{N}_0$. Hence

$$Q_j\boldsymbol{\delta}_{j,t} = \begin{bmatrix} 0 \\ \alpha_{j,2}(1-\alpha_{j,2})\lambda_{j,1} \\ \alpha_{j,3}\alpha_{j,2}(1-\alpha_{j,3}\alpha_{j,2})\lambda_{j,1} + \alpha_{j,3}(1-\alpha_{j,3})\lambda_{j,2} \\ \vdots \\ \prod\limits_{k=0}^{s-2}\alpha_{j,s-k}\left(1-\prod\limits_{k=0}^{s-2}\alpha_{j,s-k}\right)\lambda_{j,1} + \prod\limits_{k=0}^{s-3}\alpha_{j,s-k}\left(1-\prod\limits_{k=0}^{s-3}\alpha_{j,s-k}\right)\lambda_{j,2} + \\ \quad + \ldots + \alpha_{j,s}(1-\alpha_{j,s})\lambda_{j,s-1} \end{bmatrix}$$

we can write

$$\begin{aligned} Q_j^* &= diag(Q_j\boldsymbol{\delta}_{j,t}) = \\ &= \begin{bmatrix} 0 & 0 & \ldots & 0 \\ 0 & \alpha_{j,2}(1-\alpha_{j,2})\lambda_{j,1} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sum\limits_{l=1}^{s-1}\prod\limits_{k=0}^{s-(l-1)}\alpha_{j,s-k}\left(1-\prod\limits_{k=0}^{s-(l-1)}\alpha_{j,s-k}\right)\lambda_{j,l} \end{bmatrix}. \end{aligned} \tag{2.36}$$

The variance-covariance matrix of $\mathbf{R}_t$ in (2.34) can be written as

$$\begin{aligned} \sum\nolimits_{\mathbf{R}_t} &= \widetilde{B}\psi_t\widetilde{B}^T + diag(Q\boldsymbol{\delta}_t) = \\ &= \begin{bmatrix} B_1\psi_{11,t}B_1^T + Q_1^* & B_1\psi_{12,t}B_2^T & \ldots & B_1\psi_{1m,t}B_m^T \\ B_2\psi_{12,t}B_1^T & B_2\psi_{22,t}B_2^T + Q_2^* & \ldots & B_2\psi_{2m,t}B_m^T \\ \vdots & \vdots & \ddots & \vdots \\ B_m\psi_{1m,t}B_1^T & B_m\psi_{2m,t}B_2^T & \ldots & B_m\psi_{mm,t}B_m^T + Q_m^* \end{bmatrix}. \end{aligned} \tag{2.37}$$

Furthermore, for each component $j = 1, \dots, m$,

$$Var[\mathbf{R}_{j,t}] = B_j\psi_{jj,t}B_j^T + Q_j^* \tag{2.38}$$

and for $j \neq k, k = 1, \ldots, m$,

$$Cov(\mathbf{R}_{j,t}, \mathbf{R}_{k,t}) = B_j \psi_{jk,t} B_k^T, \tag{2.39}$$

where matrices B_j, $\psi_{jk,t}$ and Q_j^* are given by (2.8), (2.18) and (2.36), respectively. Based on relation (2.31), matrix B_j in (2.11) takes the form

$$B_j = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ \varphi_{2,1}^{(j)} & 1 & 0 & \ldots & 0 \\ \varphi_{3,2}^{(j)} & \varphi_{3,1}^{(j)} & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \varphi_{s,s-1}^{(j)} & \varphi_{s,s-2}^{(j)} & \varphi_{s,s-3}^{(j)} & \ldots & 1 \end{bmatrix}. \tag{2.40}$$

Thus, the matrix product $B_j \psi_{jj,t} B_j^T$ in (2.38) can be simplified to

$$B_j \psi_{jj,t} B_j^T = \\ = \begin{bmatrix} \sigma_{j,1}^2 & \varphi_{2,1}^{(j)} \sigma_{j,1}^2 & \ldots & \varphi_{s,s-1}^{(j)} \sigma_{j,1}^2 \\ \varphi_{2,1}^{(j)} \sigma_{j,1}^2 & (\varphi_{2,1}^{(j)})^2 \sigma_{j,1}^2 + \sigma_{j,2}^2 & \ldots & \varphi_{2,1}^{(j)} \varphi_{s,s-1}^{(j)} \sigma_{j,1}^2 + \varphi_{s,s-2}^{(j)} \sigma_{j,2}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{s,s-1}^{(j)} \sigma_{j,1}^2 & \varphi_{2,1}^{(j)} \varphi_{s,s-1}^{(j)} \sigma_{j,1}^2 + \varphi_{s,s-2}^{(j)} \sigma_{j,2}^2 & \ldots & \sum_{k=1}^{s-1} (\varphi_{s,s-k}^{(j)})^2 \sigma_{j,k}^2 + \sigma_{j,s}^2 \end{bmatrix} \tag{2.41}$$

and the covariance matrix in (2.39) be written as

$$Cov(\mathbf{R}_{j,t}, \mathbf{R}_{k,t}) = B_j \psi_{jk,t} B_k^T \\ = \begin{bmatrix} \sigma_{jk,1} & \sigma_{jk,1} \varphi_{2,1}^{(k)} & \ldots & \sigma_{jk,1} \varphi_{s,s-1}^{(k)} \\ \varphi_{2,1}^{(j)} \sigma_{jk,1} & \varphi_{2,1}^{(j)} \sigma_{jk,1} \varphi_{2,1}^{(k)} + \sigma_{jk,2} & \ldots & \varphi_{2,1}^{(j)} \sigma_{jk,1} \varphi_{s,s-1}^{(k)} + \sigma_{jk,2} \varphi_{s,s-2}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{s,s-1}^{(j)} \sigma_{jk,1} & \varphi_{s,s-1}^{(j)} \sigma_{jk,1} \varphi_{2,1}^{(k)} + \varphi_{s,s-2}^{(j)} \sigma_{jk,2} & \ldots & \varphi_{s,s-1}^{(j)} \sigma_{jk,1} \varphi_{s,s-1}^{(k)} + \varphi_{s,s-2}^{(j)} \sigma_{jk,2} \varphi_{s,s-2}^{(k)} + \ldots + \sigma_{jk,s} \end{bmatrix}. \tag{2.42}$$

Let the variance-covariance matrix of $\mathbf{X}_t$ be $\sum_{\mathbf{X}_t}$ with $\mathbf{X}_t = \widetilde{A} \circ \mathbf{X}_{t-s} + \mathbf{R}_t$ given in recursion (2.23). Recall from (2.23) that $\mathbf{R}_t = \widetilde{B} \circ \mathbf{Z}_t$ and $\mathbf{Z}_t$ are independent of $\mathbf{X}_{t-s}$, thus variance-covariance matrix of $\mathbf{X}_t$ is given by

$$\begin{aligned}
\sum\nolimits_{\mathbf{X}_t} &= Var[\mathbf{X}_t] = Var[\widetilde{A} \circ \mathbf{X}_{t-s}] + Var[\mathbf{R}_t] = \\
&= Var[E(\widetilde{A} \circ \mathbf{X}_{t-s}|\mathbf{X}_{t-s})] + E[Var(\widetilde{A} \circ \mathbf{X}_{t-s}|\mathbf{X}_{t-s})] + Var[\mathbf{R}_t] = \\
&= Var[\widetilde{A}\mathbf{X}_{t-s}] + E[Var(\widetilde{A} \circ \mathbf{X}_{t-s}|\mathbf{X}_{t-s})] + Var[\mathbf{R}_t] = \\
&= \widetilde{A} Var[\mathbf{X}_{t-s}] \widetilde{A}^T + diag\left(DE[\mathbf{X}_{t-s}]\right) + \sum\nolimits_{\mathbf{R}_t}
\end{aligned}$$

and due to cyclostationarity, $\Gamma(0)$ proves to satisfy a difference equation of the form

$$\Gamma(0) = \widetilde{A}\Gamma(0)\widetilde{A}^T + diag(D\boldsymbol{\mu}_t) + \sum\nolimits_{\mathbf{R}_t} \tag{2.43}$$

with matrices $\widetilde{A}$, $\boldsymbol{\mu}_t$ and $\sum_{\mathbf{R}_t}$ defined in (2.10), (2.27) and (2.34), respectively. From Lemma 1.3 (property 2), matrix D in (2.43) is a $(ms \times ms)$ variance matrix, $D_j = [d_{ik}^j] = [a_{ik}^j(1 - a_{ik}^j)]$ for $i, k = 1, \ldots, s$ and $j = 1, \ldots, m$. From matrix A_j in (2.7) and from (2.31), matrix D_j is given by

$$D_j = \begin{bmatrix}
0 & \cdots & 0 & \varphi_{1,1}^{(j)}\left(1 - \varphi_{1,1}^{(j)}\right) \\
0 & \cdots & 0 & \varphi_{2,2}^{(j)}\left(1 - \varphi_{2,2}^{(j)}\right) \\
0 & \cdots & 0 & \varphi_{3,3}^{(j)}\left(1 - \varphi_{3,3}^{(j)}\right) \\
\vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & \varphi_{s,s}^{(j)}\left(1 - \varphi_{s,s}^{(j)}\right)
\end{bmatrix}. \tag{2.44}$$

We then define

$$diag(D\boldsymbol{\mu}_t) = \begin{bmatrix}
D_1^* & 0 & \cdots & 0 \\
0 & D_2^* & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & D_m^*
\end{bmatrix},$$

where $D_j^* = diag(D_j \boldsymbol{\mu}_{j,t})$ with matrix D_j in (2.44) and $\boldsymbol{\mu}_{j,t}$ in (2.33) $(j = 1, \ldots, m)$, yielding

$$D_j \boldsymbol{\mu}_{j,t} = \begin{bmatrix} 0 & \cdots & 0 & \varphi_{1,1}^{(j)}\left(1-\varphi_{1,1}^{(j)}\right) \\ 0 & \cdots & 0 & \varphi_{2,2}^{(j)}\left(1-\varphi_{2,2}^{(j)}\right) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \varphi_{s,s}^{(j)}\left(1-\varphi_{s,s}^{(j)}\right) \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu}_{1,t} \\ \boldsymbol{\mu}_{2,t} \\ \vdots \\ \boldsymbol{\mu}_{m,t} \end{bmatrix} = \begin{bmatrix} \varphi_{1,1}^{(j)}\left(1-\varphi_{1,1}^{(j)}\right)\boldsymbol{\mu}_{m,t} \\ \varphi_{2,2}^{(j)}\left(1-\varphi_{2,2}^{(j)}\right)\boldsymbol{\mu}_{m,t} \\ \vdots \\ \varphi_{s,s}^{(j)}\left(1-\varphi_{s,s}^{(j)}\right)\boldsymbol{\mu}_{m,t} \end{bmatrix}$$

i.e.,

$$D_j \boldsymbol{\mu}_{j,t} = \begin{bmatrix} \varphi_{1,1}^{(j)}\left(1-\varphi_{1,1}^{(j)}\right) \dfrac{\sum_{k=0}^{s-1} \varphi_{s,k}^{(j)} \lambda_{j,s-k}}{1-\varphi_{s,s}^{(j)}} \\ \varphi_{2,2}^{(j)}\left(1-\varphi_{2,2}^{(j)}\right) \dfrac{\sum_{k=0}^{s-1} \varphi_{s,k}^{(j)} \lambda_{j,s-k}}{1-\varphi_{s,s}^{(j)}} \\ \vdots \\ \varphi_{s,s}^{(j)}\left(1-\varphi_{s,s}^{(j)}\right) \dfrac{\sum_{k=0}^{s-1} \varphi_{s,k}^{(j)} \lambda_{j,s-k}}{1-\varphi_{s,s}^{(j)}} \end{bmatrix}.$$

For each component $j = 1, \ldots, m;\ t = v + ns;\ v = 1, \ldots, s$ and $n \in \mathbb{N}_0$, we can write

$$\begin{aligned} D_j^* &= diag(D_j \boldsymbol{\mu}_{j,t}) = \\ &= \frac{\sum_{k=0}^{s-1} \varphi_{s,k}^{(j)} \lambda_{j,s-k}}{1-\varphi_{s,s}^{(j)}} \begin{bmatrix} \varphi_{1,1}^{(j)}\left(1-\varphi_{1,1}^{(j)}\right) & 0 & \cdots & 0 \\ 0 & \varphi_{2,2}^{(j)}\left(1-\varphi_{2,2}^{(j)}\right) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \varphi_{s,s}^{(j)}\left(1-\varphi_{s,s}^{(j)}\right) \end{bmatrix}. \end{aligned} \tag{2.45}$$

For $j = 1, \ldots, m$ and $j \neq k, k = 1, \ldots, m$, the variance of the j-th component $\mathbf{X}_{j,t}$ has the following form:

$$Var[\mathbf{X}_{j,t}] = A_j Var[\mathbf{X}_{j,t}](A_j)^T + diag\left(D_j \boldsymbol{\mu}_{j,t}\right) + Var[\mathbf{R}_{j,t}] \tag{2.46}$$

and covariance between two different components $\mathbf{X}_{j,t}$ and $\mathbf{X}_{k,t}$:

$$Cov(\mathbf{X}_{j,t}, \mathbf{X}_{k,t}) = A_j^s Cov(\mathbf{X}_{j,t}, \mathbf{X}_{k,t})(A_k^s)^T + \sum_{i=0}^{s-1} \left(A_j^i B_j \psi_{jk,t-i} + (A_k^i B_k)^T \right). \tag{2.47}$$

The matricial representation of the variance-covariance matrix, $\sum_{\mathbf{X}_t}$, follows

$$\sum\nolimits_{\mathbf{X}_t} = \begin{bmatrix} Var[\mathbf{X}_{1,t}] & Cov(\mathbf{X}_{1,t}, \mathbf{X}_{2,t}) & \dots & Cov(\mathbf{X}_{1,t}, \mathbf{X}_{m,t}) \\ Cov(\mathbf{X}_{2,t}, \mathbf{X}_{1,t}) & Var[\mathbf{X}_{2,t}] & \dots & Cov(\mathbf{X}_{2,t}, \mathbf{X}_{m,t}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\mathbf{X}_{m,t}, \mathbf{X}_{1,t}) & Cov(\mathbf{X}_{m,t}, \mathbf{X}_{2,t}) & \dots & Var[\mathbf{X}_{m,t}] \end{bmatrix}$$

$$=: \begin{bmatrix} \sum_{1,1} & \sum_{1,2} & \cdots & \sum_{1,m} \\ & \sum_{2,2} & \cdots & \sum_{2,m} \\ & & \ddots & \vdots \\ & & & \sum_{m,m} \end{bmatrix}. \tag{2.48}$$

For $j = 1, \dots, m$, $(s \times s)$ symmetric matrices $\sum_{j,j}$ are given by

$$\sum\nolimits_{j,j} = Var[\mathbf{X}_{j,t}] =$$

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

with diagonal elements

$$
\begin{aligned}
&Var[X_{j,v+ns}] = \\
&= \frac{1}{1-\left(\varphi_{s,s}^{(j)}\right)^2}\left\{\sum_{k=0}^{v-1}\left[\varphi_{s,s}^{(j)}\varphi_{v,k}^{(j)}\lambda_{j,v-k}+\varphi_{v,k}^{(j)}\left(1-\varphi_{v,k}^{(j)}\right)\lambda_{j,v-k}+\left(\varphi_{v,k}^{(j)}\right)^2\sigma_{j,v-k}^2\right]+\right.\\
&\left.+\sum_{m=0}^{s-(v+1)}\left[\varphi_{s,s}^{(j)}\varphi_{v,v}^{(j)}\varphi_{s,m}^{(j)}\lambda_{j,s-m}+\varphi_{v,v}^{(j)}\varphi_{s,m}^{(j)}\left(1-\varphi_{v,v}^{(j)}\varphi_{s,m}^{(j)}\right)\lambda_{j,s-m}+\left(\varphi_{v,v}^{(j)}\varphi_{s,m}^{(j)}\right)^2\sigma_{j,s-k}^2\right]\right\}
\end{aligned}
\tag{2.49}
$$

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

$$
Cov(X_{j,v+ns}, X_{j,v+ns+l}) = \varphi_{v+l,l}^{(j)} Var[X_{j,v+ns}],
\tag{2.50}
$$

where $\lambda_{j,v}$ represents the mean of $Z_{j,v+ns}$ in (2.16) and $\sigma_{j,v}^2$ the variance in (2.21). The $(s \times s)$ non-symmetric matrices $\sum_{j,k}$ $(j \neq k;\ j,k = 1, \ldots, m)$ in (2.48) are given by

$$
\begin{aligned}
&\sum\nolimits_{j,k} = Cov(\mathbf{X}_{j,t}, \mathbf{X}_{k,t}) = \\
&= \begin{bmatrix}
Cov(X_{j,1+ns}, X_{k,1+ns}) & Cov(X_{j,1+ns}, X_{k,2+ns}) & \ldots & Cov(X_{j,1+ns}, X_{k,s+ns}) \\
Cov(X_{j,2+ns}, X_{k,1+ns}) & Cov(X_{j,2+ns}, X_{k,2+ns}) & \ldots & Cov(X_{j,2+ns}, X_{k,s+ns}) \\
\vdots & \vdots & \ddots & \vdots \\
Cov(X_{j,s+ns}, X_{k,1+ns}) & Cov(X_{j,s+ns}, X_{k,2+ns}) & \ldots & Cov(X_{j,s+ns}, X_{k,s+ns})
\end{bmatrix}
\end{aligned}
\tag{2.51}
$$

with diagonal elements

$$
\begin{aligned}
&Cov(X_{j,v+ns}, X_{k,v+ns}) = \\
&= \frac{1}{1-\varphi_{s,s}^{(j)}\varphi_{s,s}^{(k)}}\sum_{i=0}^{v-1}\varphi_{v,i}^{(j)}\varphi_{v,i}^{(k)}\sigma_{jk,v-i} + \frac{\varphi_{v,v}^{(j)}\varphi_{v,v}^{(k)}}{1-\varphi_{s,s}^{(j)}\varphi_{s,s}^{(k)}}\sum_{i=0}^{s-(v+1)}\varphi_{s,i}^{(j)}\varphi_{s,i}^{(k)}\sigma_{jk,s-i}
\end{aligned}
\tag{2.52}
$$

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

$$Cov(X_{j,v+ns+h}, X_{k,v+ns}) =$$
$$= \frac{\varphi^{(j)}_{v+h,h}}{1-\varphi^{(j)}_{s,s}\varphi^{(k)}_{s,s}} \sum_{i=0}^{v-1} \varphi^{(j)}_{v,i}\varphi^{(k)}_{v,i}\sigma_{jk,v-i} + \frac{\varphi^{(j)}_{v+h,h}\varphi^{(j)}_{v,v}\varphi^{(k)}_{v,v}}{1-\varphi^{(j)}_{s,s}\varphi^{(k)}_{s,s}} \sum_{i=0}^{s-(v+1)} \varphi^{(j)}_{s,i}\varphi^{(k)}_{v,i}\sigma_{jk,s-i} \tag{2.53}$$

and

$$Cov(X_{j,v+ns}, X_{k,v+ns+h}) =$$
$$= \frac{\varphi^{(k)}_{v+h,h}}{1-\varphi^{(j)}_{s,s}\varphi^{(k)}_{s,s}} \sum_{i=0}^{v-1} \varphi^{(j)}_{v,i}\varphi^{(k)}_{v,i}\sigma_{jk,v-i} + \frac{\varphi^{(k)}_{v+h,h}\varphi^{(j)}_{v,v}\varphi^{(k)}_{v,v}}{1-\varphi^{(j)}_{s,s}\varphi^{(k)}_{s,s}} \sum_{i=0}^{s-(v+1)} \varphi^{(j)}_{s,i}\varphi^{(k)}_{v,i}\sigma_{jk,s-i}, \tag{2.54}$$

where $\sigma_{jk,v}$ represents the covariance between $Z_{j,v+ns}$ and $Z_{k,v+ns}$ as defined in (2.19).

Auto-covariance function with lag h

For each component $j = 1, \dots, m$ and positive lag h:

$$Cov(\mathbf{X}_{j,t}, \mathbf{X}_{j,t+h}) = Cov\left(\mathbf{X}_{j,t}, A_j^h \circ \mathbf{X}_{j,t} + \sum_{i=0}^{h-1} A_j^i \circ \mathbf{R}_{j,t+h-i}\right) =$$
$$= A_j^h Cov(\mathbf{X}_{j,t}, \mathbf{X}_{j,t}) = A_j^h Var[\mathbf{X}_{j,t}], \tag{2.55}$$
$$Cov(\mathbf{X}_{j,t+h}, \mathbf{X}_{k,t}) = A_j^h Cov(\mathbf{X}_{j,t}, \mathbf{X}_{k,t}), \tag{2.56}$$
$$Cov(\mathbf{X}_{j,t}, \mathbf{X}_{k,t+h}) = A_k^h Cov(\mathbf{X}_{j,t}, \mathbf{X}_{k,t}). \tag{2.57}$$

The matricial form of $Cov(\mathbf{X}_{j,t}, \mathbf{X}_{j,t+h})$ with $t = v + ns;\ v = 1, \dots, s$ and $j = 1, \dots, m$ is

$$Cov(\mathbf{X}_{j,t}, \mathbf{X}_{j,t+h}) =$$
$$= \begin{bmatrix} Cov(X_{j,1+ns}, X_{j,1+ns+h}) & Cov(X_{j,1+ns}, X_{j,2+ns+h}) & \dots & Cov(X_{j,1+ns}, X_{j,s+ns+h}) \\ Cov(X_{j,2+ns}, X_{j,1+ns+h}) & Cov(X_{j,2+ns}, X_{j,2+ns+h}) & \dots & Cov(X_{j,2+ns}, X_{j,s+ns+h}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_{j,s+ns}, X_{j,1+ns+h}) & Cov(X_{j,s+ns}, X_{j,2+ns+h}) & \dots & Cov(X_{j,s+ns}, X_{j,s+ns+h}) \end{bmatrix}. \tag{2.58}$$

2.3 Estimation of the PMINAR(1) parameters

Consider a finite time series $\{\mathbf{X}_{j,ts},\ 1 \leq t \leq N,\ j = 1, \ldots, m\}$ from the PMINAR(1) model in (2.23), where N stands for the number of complete cycles. Without loss of generality it is assumed that $\mathbf{X}_0 = \mathbf{x}_0$. The methods of Yule-Walker, conditional maximum likelihood and composite likelihood are proposed for the estimation of the parameters of the PMINAR(1) model. Let $\boldsymbol{\theta}$ be the vector of unknown parameters

$$\boldsymbol{\theta} := (\boldsymbol{\alpha}_j, \boldsymbol{\lambda}_j, \boldsymbol{\sigma}_j^2, \boldsymbol{\sigma}_{j,k}) \tag{2.59}$$

with s-dimensional vectors $\boldsymbol{\alpha}_j, \boldsymbol{\lambda}_j, \boldsymbol{\sigma}_j^2$ and $\boldsymbol{\sigma}_{jk}$ $(j \neq k; j, k = 1, \ldots, m)$

$$\begin{aligned} &\boldsymbol{\alpha}_j = (\alpha_{j,1}, \ldots, \alpha_{j,s}) \ ; \ \boldsymbol{\lambda}_j = (\lambda_{j,1}, \ldots, \lambda_{j,s}) \ ; \\ &\boldsymbol{\sigma}_j^2 = (\sigma_{j,1}^2, \ldots, \sigma_{j,s}^2) \ ; \ \boldsymbol{\sigma}_{jk} = (\sigma_{jk,1}, \ldots, \sigma_{jk,s}). \end{aligned} \tag{2.60}$$

Alternatively, the vector $\boldsymbol{\theta}$ in (2.59) can be written as

$$\boldsymbol{\theta} := (vec(\boldsymbol{\alpha}_j)^T, vec(\boldsymbol{\lambda}_j)^T, vec(\boldsymbol{\sigma}_j^2)^T, vec(\boldsymbol{\sigma}_{jk})^T),$$

i.e., $vec(U)$ corresponds to the vector obtained by stacking the columns of U (Harville, 2008).

2.3.1 Yule-Walker estimation

Let $\widehat{\boldsymbol{\theta}}^{YW}$ be the vector of the Yule-Walker (YW) estimators for the unknown parameters in (2.59), thus

$$\widehat{\boldsymbol{\theta}}^{YW} := (\widehat{\boldsymbol{\alpha}}_j^{YW}, \widehat{\boldsymbol{\lambda}}_j^{YW}, \widehat{\boldsymbol{\sigma}}_j^{2,YW}, \widehat{\boldsymbol{\sigma}}_{j,k}^{YW}) \tag{2.61}$$

with $(s \times 1)$ vectors $\widehat{\boldsymbol{\alpha}}_j^{YW}, \widehat{\boldsymbol{\lambda}}_j^{YW}, \widehat{\boldsymbol{\sigma}}_j^{2,YW}$ and $\widehat{\boldsymbol{\sigma}}_{jk}^{YW}$ as in (2.60), respectively.

For each j $(j = 1, \ldots, m)$ and for a fixed v $(v = 1, \ldots, s)$, we define:

- sample mean:

$$\overline{X}_{j,v} = \frac{1}{N}\sum_{n=0}^{N-1} X_{j,v+ns}, \tag{2.62}$$

- sample variance:

$$S^2_{j,v} = \frac{1}{N-1}\sum_{n=0}^{N-1}(X_{j,v+ns} - \overline{X}_{j,v})^2, \tag{2.63}$$

- sample autocovariance function at lag 1:

$$\gamma_{j,v}(1) = Cov(X_{j,v+ns}, X_{j,v+1+ns}) = \\ = \begin{cases} \frac{1}{N-1}\sum\limits_{n=0}^{N-1}(X_{j,v+ns} - \overline{X}_{j,v})(X_{j,v+1+ns} - \overline{X}_{j,v+1}) & ,v = 1,\ldots,s-1 \\ \\ \frac{1}{N-1}\sum\limits_{n=0}^{N-1}(X_{j,v+ns} - \overline{X}_{j,v})(X_{j,1+(n+1)s} - \overline{X}^*_{j,1}) & ,v = s \end{cases} \tag{2.64}$$

with $\overline{X}^*_{j,1} = \frac{1}{N}\sum\limits_{n=0}^{N} X_{j,1+ns}$,

- sample cross-covariance function at lag 1:

$$\gamma_{jk,v}(1) = Cov(X_{j,v+ns}, X_{k,v+1+ns}) = \\ = \begin{cases} \frac{1}{N-1}\sum\limits_{n=0}^{N-1}(X_{j,v+ns} - \overline{X}_{j,v})(X_{k,v+1+ns} - \overline{X}_{k,v+1}) & ,v = 1,\ldots,s-1 \\ \\ \frac{1}{N-1}\sum\limits_{n=0}^{N-1}(X_{j,v+ns} - \overline{X}_{j,v})(X_{k,1+(n+1)s} - \overline{X}^*_{k,1}) & ,v = s \end{cases} . \tag{2.65}$$

Take the mean of $X_{j,v+ns}$ in (2.32) and its sample counterpart in (2.62) then

$$E[X_{j,v+ns}] = \overline{X}_{j,v} \;\Leftrightarrow\; \sum_{k=0}^{v-1}\varphi^{(j)}_{v,k}\lambda_{j,v-k} + \varphi^{(j)}_{v,v}\sum_{i=0}^{s-(v+1)}\varphi^{(j)}_{s,i}\lambda_{j,s-i} = \left(1-\varphi^{(j)}_{s,s}\right)\overline{X}_{j,v}.$$

The YW estimators of parameters $\boldsymbol{\lambda}_j$ are calculated through the solution of the following system of s linear equations:

$$\begin{cases} E[X_{1,v+ns}] = \overline{X}_{j,1} \\ E[X_{2,v+ns}] = \overline{X}_{j,2} \\ E[X_{3,v+ns}] = \overline{X}_{j,3} \\ \vdots \\ E[X_{s,v+ns}] = \overline{X}_{j,s} \end{cases} \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} \lambda_{j,1} + \alpha_{j,1}\left(\lambda_{j,2}\prod_{k=0}^{s-3}\alpha_{j,s-k} + \cdots + \lambda_{j,s-1}\alpha_{j,s} + \lambda_{j,s}\right) = \left(1-\varphi_{s,s}^{(j)}\right)\overline{X}_{j,1} \\ \lambda_{j,1}\alpha_{j,2} + \lambda_{j,2} + \alpha_{j,1}\alpha_{j,2}\left(\lambda_{j,3}\prod_{k=0}^{s-4}\alpha_{j,s-k} + \cdots + \lambda_{j,s-1}\alpha_{j,s} + \lambda_{j,s}\right) = \left(1-\varphi_{s,s}^{(j)}\right)\overline{X}_{j,2} \\ \lambda_{j,1}\alpha_{j,3}\alpha_{j,2} + \lambda_{j,2}\alpha_{j,3} + \lambda_{j,3} + \\ \quad + \frac{d_{(-j)}}{d}\prod_{i=1}^{3}\alpha_{j,i}\left(\lambda_{j,1}\prod_{k=0}^{s-2}\alpha_{j,s-k} + \lambda_{j,2}\prod_{k=0}^{s-3}\alpha_{j,s-k} + \cdots + \lambda_{j,s-1}\alpha_{j,s} + \lambda_{j,s}\right) = \left(1-\varphi_{s,s}^{(j)}\right)\overline{X}_{j,3} \\ \vdots \\ \lambda_{j,1}\prod_{k=0}^{s-2}\alpha_{j,s-k} + \lambda_{j,2}\prod_{k=0}^{s-3}\alpha_{j,s-k} + \cdots + \lambda_{j,s-1}\alpha_{j,s} + \lambda_{j,s} = \left(1-\varphi_{s,s}^{(j)}\right)\overline{X}_{j,s} \end{cases}.$$

The matrix representation of the above system of linear equations is

$$\begin{bmatrix} 1 & \alpha_{j,1}\prod_{k=0}^{s-3}\alpha_{j,s-k} & \alpha_{j,1}\prod_{k=0}^{s-4}\alpha_{j,s-k} & \cdots & \alpha_{j,1} \\ \alpha_{j,2} & 1 & \alpha_{j,1}\alpha_{j,2}\prod_{k=0}^{s-4}\alpha_{j,s-k} & \cdots & \alpha_{j,1}\alpha_{j,2} \\ \alpha_{j,3}\alpha_{j,2} & \alpha_{j,3} & 1 & \cdots & \alpha_{j,1}\alpha_{j,2}\alpha_{j,3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \prod_{k=0}^{s-2}\alpha_{j,s-k} & \prod_{k=0}^{s-3}\alpha_{j,s-k} & \prod_{k=0}^{s-4}\alpha_{j,s-k} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \lambda_{j,1} \\ \lambda_{j,2} \\ \lambda_{j,3} \\ \vdots \\ \lambda_{j,s} \end{bmatrix} = \left(1-\varphi_{s,s}^{(j)}\right) \begin{bmatrix} \overline{X}_{j,1} \\ \overline{X}_{j,2} \\ \overline{X}_{j,3} \\ \vdots \\ \overline{X}_{j,s} \end{bmatrix}$$

thus, through equation (2.31), we can write

$$\begin{bmatrix} 1 & \varphi_{1,1}^{(j)}\varphi_{s,s-2}^{(j)} & \varphi_{1,1}^{(j)}\varphi_{s,s-3}^{(j)} & \cdots & \varphi_{1,1}^{(j)} \\ \varphi_{2,1}^{(j)} & 1 & \varphi_{2,2}^{(j)}\varphi_{s,s-3}^{(j)} & \cdots & \varphi_{2,2}^{(j)} \\ \varphi_{3,2}^{(j)} & \varphi_{3,1}^{(j)} & 1 & \cdots & \varphi_{3,3}^{(j)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \varphi_{s,s-1}^{(j)} & \varphi_{s,s-2}^{(j)} & \varphi_{s,s-3}^{(j)} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \lambda_{j,1} \\ \lambda_{j,2} \\ \lambda_{j,3} \\ \vdots \\ \lambda_{j,s} \end{bmatrix} = \left(1-\varphi_{s,s}^{(j)}\right) \begin{bmatrix} \overline{X}_{j,1} \\ \overline{X}_{j,2} \\ \overline{X}_{j,3} \\ \vdots \\ \overline{X}_{j,s} \end{bmatrix},$$

where $W_j \boldsymbol{\delta}_{j,t} = \left(1 - \varphi_{s,s}^{(j)}\right) \overline{X}_{j,v}$, for $j = 1, \ldots, m; v = 1, \ldots, s$, with $\boldsymbol{\delta}_{j,t}$ in (2.15) and

$$W_j = \begin{bmatrix} 1 & \varphi_{1,1}^{(j)}\varphi_{s,s-2}^{(j)} & \varphi_{1,1}^{(j)}\varphi_{s,s-3}^{(j)} & \ldots & \varphi_{1,1}^{(j)} \\ \varphi_{2,1}^{(j)} & 1 & \varphi_{2,2}^{(j)}\varphi_{s,s-3}^{(1)} & \ldots & \varphi_{2,2}^{(j)} \\ \varphi_{3,2}^{(j)} & \varphi_{3,1}^{(j)} & 1 & \ldots & \varphi_{3,3}^{(j)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \varphi_{s,s-1}^{(j)} & \varphi_{s,s-2}^{(j)} & \varphi_{s,s-3}^{(j)} & \ldots & 1 \end{bmatrix}.$$

Taking all m components we obtain

$$\begin{bmatrix} W_1 & 0 & 0 & \ldots & 0 \\ 0 & W_2 & 0 & \ldots & 0 \\ 0 & 0 & W_3 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & W_m \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta}_{1,t} \\ \boldsymbol{\delta}_{2,t} \\ \boldsymbol{\delta}_{3,t} \\ \vdots \\ \boldsymbol{\delta}_{m,t} \end{bmatrix} = \begin{bmatrix} \left(1 - \varphi_{s,s}^{(1)}\right) \overline{X}_{1,v} \\ \left(1 - \varphi_{s,s}^{(2)}\right) \overline{X}_{2,v} \\ \left(1 - \varphi_{s,s}^{(3)}\right) \overline{X}_{3,v} \\ \vdots \\ \left(1 - \varphi_{s,s}^{(m)}\right) \overline{X}_{m,v} \end{bmatrix}.$$

Rewriting the system $W\boldsymbol{\delta}_t = Y$ with $\boldsymbol{\delta}_t$ in (2.14) yields $\widehat{\boldsymbol{\delta}}_t = \widehat{W}^{-1}\widehat{Y}$, i.e.,

$$\begin{bmatrix} \widehat{\boldsymbol{\delta}}_{1,t} \\ \widehat{\boldsymbol{\delta}}_{2,t} \\ \widehat{\boldsymbol{\delta}}_{3,t} \\ \vdots \\ \widehat{\boldsymbol{\delta}}_{m,t} \end{bmatrix} = \begin{bmatrix} \widehat{W_1} & 0 & 0 & \ldots & 0 \\ 0 & \widehat{W_2} & 0 & \ldots & 0 \\ 0 & 0 & \widehat{W_3} & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \widehat{W_m} \end{bmatrix}^{-1} \begin{bmatrix} \left(1 - \widehat{\varphi}_{s,s}^{(1)}\right) \overline{X}_{1,v} \\ \left(1 - \widehat{\varphi}_{s,s}^{(2)}\right) \overline{X}_{2,v} \\ \left(1 - \widehat{\varphi}_{s,s}^{(3)}\right) \overline{X}_{3,v} \\ \vdots \\ \left(1 - \widehat{\varphi}_{s,s}^{(m)}\right) \overline{X}_{m,v} \end{bmatrix}.$$

Let T represent a $(p \times p)$ matrix and S a $(q \times q)$ matrix. The $(p+q) \times (p+q)$ block-diagonal matrix $\begin{bmatrix} T & 0 \\ 0 & S \end{bmatrix}$ is nonsingular if and only if both T and S are nonsingular (Harville, 2008). Moreover, if T and S are nonsingular, then it can be easily verified that

$$\begin{bmatrix} T & 0 \\ 0 & S \end{bmatrix}^{-1} = \begin{bmatrix} T^{-1} & 0 \\ 0 & S^{-1} \end{bmatrix}.$$

In our particular case, the inverse matrix is

$$\begin{bmatrix} \widehat{W}_1 & 0 & 0 & \dots & 0 \\ 0 & \widehat{W}_2 & 0 & \dots & 0 \\ 0 & 0 & \widehat{W}_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \widehat{W}_m \end{bmatrix}^{-1} = \begin{bmatrix} (\widehat{W}_1)^{-1} & 0 & 0 & \dots & 0 \\ 0 & (\widehat{W}_2)^{-1} & 0 & \dots & 0 \\ 0 & 0 & (\widehat{W}_3)^{-1} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & (\widehat{W}_m)^{-1} \end{bmatrix}$$

with

$$(\widehat{W}_j)^{-1} = \frac{1}{1 - \prod_{k=0}^{s-1} \widehat{\alpha}_{j,s-k}} \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & -\widehat{\alpha}_{j,1} \\ -\widehat{\alpha}_{j,2} & 1 & 0 & \dots & 0 & 0 \\ 0 & -\widehat{\alpha}_{j,3} & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\widehat{\alpha}_{j,s} & 1 \end{bmatrix}.$$

Recalling equations (2.15) and (2.31), the estimator for parameters $\boldsymbol{\delta}_{j,t}$ takes the form

$$\widehat{\boldsymbol{\delta}}_{j,t} = (\widehat{W}_j)^{-1} \left(1 - \widehat{\varphi}_{s,s}^{(j)}\right) \overline{X}_{j,v} \Leftrightarrow$$

$$\Leftrightarrow \begin{bmatrix} \widehat{\lambda}_{j,1} \\ \widehat{\lambda}_{j,2} \\ \widehat{\lambda}_{j,3} \\ \vdots \\ \widehat{\lambda}_{j,s} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & -\widehat{\alpha}_{j,1} \\ -\widehat{\alpha}_{j,2} & 1 & 0 & \dots & 0 & 0 \\ 0 & -\widehat{\alpha}_{j,3} & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\widehat{\alpha}_{j,s} & 1 \end{bmatrix} \begin{bmatrix} \overline{X}_{j,1} \\ \overline{X}_{j,2} \\ \overline{X}_{j,3} \\ \vdots \\ \overline{X}_{j,s} \end{bmatrix}.$$

We summarize the YW estimators of $\boldsymbol{\lambda}_j = (\lambda_{j,1}, \lambda_{j,2}, \dots, \lambda_{j,s})$, $\widehat{\lambda}_{j,v}^{YW}$, as

$$\widehat{\lambda}_{j,v}^{YW} = \begin{cases} \overline{X}_{j,v} - \widehat{\alpha}_{j,v}^{YW} \overline{X}_{j,s} & , v = 1 \\ \overline{X}_{j,v} - \widehat{\alpha}_{j,v}^{YW} \overline{X}_{j,v-1} & , v = 2, 3, \dots, s \end{cases}, \tag{2.66}$$

where $\overline{X}_{j,v}$ $(j = 1, \ldots, m)$ is the sample mean defined in (2.62). Notice that estimators $\widehat{\lambda}_{j,v}^{YW}$ depend upon estimators $\widehat{\alpha}_{j,v}^{YW}$. Let lag $h = 1$ and from relation (2.55), we obtain

$$\boldsymbol{\gamma}_{j,t}(1) = A_j \boldsymbol{\gamma}_{j,t}(0). \tag{2.67}$$

This relation is sufficient to derive estimators for parameters $\boldsymbol{\alpha}_j = (\alpha_{j,1}, \alpha_{j,2}, \ldots, \alpha_{j,s})$ for each j $(j = 1, \ldots, m)$ because matrix A_j in (2.7) contemplates all parameters $\alpha_{j,v}$ $(v = 1, \ldots, s)$. The $(s \times s)$ matrices $\boldsymbol{\gamma}_{j,t}(1)$ and $\boldsymbol{\gamma}_{j,t}(0)$ in relation (2.67) can be obtained from (2.58) by replacing lag h with one and zero, respectively. Therefore,

$$\begin{aligned}
&\boldsymbol{\gamma}_{j,v+ns}(1) = Cov(X_{j,v+ns}, X_{j,v+ns+1}) = \\
&= \begin{bmatrix}
Cov(X_{j,1+ns}, X_{j,2+ns}) & Cov(X_{j,1+ns}, X_{j,3+ns}) & \ldots & Cov(X_{j,1+ns}, X_{j,1+ns}) \\
Cov(X_{j,2+ns}, X_{j,2+ns}) & Cov(X_{j,2+ns}, X_{j,3+ns}) & \ldots & Cov(X_{j,2+ns}, X_{j,1+ns}) \\
\vdots & \vdots & \ddots & \vdots \\
Cov(X_{j,s+ns}, X_{j,2+ns}) & Cov(X_{j,s+ns}, X_{j,3+ns}) & \ldots & Cov(X_{j,s+ns}, X_{j,1+ns})
\end{bmatrix}
\end{aligned}$$

and

$$\begin{aligned}
&A_j \gamma_{j,v+ns}(0) = \\
&= \begin{bmatrix}
\alpha_{j,1} Cov(X_{j,1+ns}, X_{j,s+ns}) & \alpha_{j,1} Cov(X_{j,2+ns}, X_{j,s+ns}) & \ldots & \alpha_{j,1} Var[X_{j,s+ns}] \\
\alpha_{j,2}\alpha_{j,1} Cov(X_{j,1+ns}, X_{j,s+ns}) & \alpha_{j,2}\alpha_{j,1} Cov(X_{j,2+ns}, X_{j,s+ns}) & \ldots & \alpha_{j,2}\alpha_{j,1} Var[X_{j,s+ns}] \\
\vdots & \vdots & \ddots & \vdots \\
\alpha_{j,s} \ldots \alpha_{j,1} Cov(X_{j,1+ns}, X_{j,s+ns}) & \alpha_{j,s} \ldots \alpha_{j,1} Cov(X_{j,2+ns}, X_{j,s+ns}) & \ldots & \alpha_{j,s} \ldots \alpha_{j,1} Var[X_{j,s+ns}]
\end{bmatrix}.
\end{aligned}$$

Furthermore, from relation (2.67):

$$\begin{aligned}
&\alpha_{j,1} Var[X_{j,s+ns}] = Cov(X_{j,1+ns}, X_{j,1+ns}) \;\Leftrightarrow\; \alpha_{j,1} = \frac{Var[X_{j,1+ns}]}{Var[X_{j,s+ns}]} \\
&\alpha_{j,2}\alpha_{j,1} Var[X_{j,s+ns}] = Cov(X_{j,2+ns}, X_{j,1+ns}) \;\Leftrightarrow\; \alpha_{j,2} = \frac{Cov(X_{j,2+ns}, X_{j,1+ns})}{Var[X_{j,1+ns}]} \\
&\vdots \\
&\alpha_{j,s} = \frac{Cov(X_{j,s+ns}, X_{j,s-1+ns})}{Var[X_{j,s-1+ns}]}
\end{aligned}$$

Taking the corresponding empirical counterparts, we resume the YW estimators of parameters $\boldsymbol{\alpha}_j = (\alpha_{j,1}, \alpha_{j,2}, \ldots, \alpha_{j,s})$, $\widehat{\alpha}_{j,v}^{YW}$, as

$$\widehat{\alpha}_{j,v}^{YW} = \begin{cases} \dfrac{S_{j,v}^2}{S_{j,s}^2} & , v = 1 \\ \\ \dfrac{\gamma_{j,v-1}(1)}{S_{j,v-1}^2} & , v = 2, 3, \ldots, s \end{cases}, \tag{2.68}$$

where $S_{j,v}^2$ is the sample variance in (2.63) and $\gamma_{j,v}(1)$ the sample auto-covariance function in (2.64) for component $j = 1, \ldots, m$. YW estimators of parameters $\boldsymbol{\sigma}_j^2 = (\sigma_{j,1}^2, \sigma_{j,2}^2, \ldots, \sigma_{j,s}^2)$, $\widehat{\sigma}_{j,v}^{2,YW}$, can be calculated through sample variance in (2.63) and of $\boldsymbol{\sigma}_{jk} = (\sigma_{jk,1}, \sigma_{jk,2}, \ldots, \sigma_{jk,s})$, $\widehat{\sigma}_{jk,v}^{YW}$, through sample cross-covariance in (2.65).

2.3.2 Conditional maximum likelihood estimation

Let $\boldsymbol{\theta}$ be the vector of unknown parameters in (2.59). The joint probability function of the vector of innovations $\mathbf{Z}_{j,t}$ with $j = 1, \ldots, m$; $t = v + ns$ and $v = 1, \ldots, s$ follows the periodic discrete m-variate distribution

$$P(Z_{1,v+ns} = z_1, Z_{2,v+ns} = z_2, \ldots, Z_{m,v+ns} = z_m) = h(z_1, z_2, \ldots, z_m). \tag{2.69}$$

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

$$f_j(r_j) = C_{r_j}^{x_{j,v-1+ns}} \alpha_{j,v}^{r_j} (1 - \alpha_{j,v})^{x_{j,v-1+ns} - r_j},\ j = 1, \ldots, m, \tag{2.70}$$

and the discrete multivariate distribution defined in (2.69). Thus, the conditional density is the multiple sum

$$p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns}) = P(\mathbf{X}_{v+ns} = \mathbf{x}_{v+ns}|\mathbf{X}_{v-1+ns} = \mathbf{x}_{v-1+ns}) =$$
$$= P(X_{1,v+ns} = x_{1,v+ns}, \ldots, X_{m,v+ns} = x_{m,v+ns}|\mathbf{X}_{v-1+ns} = \mathbf{x}_{v-1+ns}) =$$

$$= \sum_{r_1=0}^{g_1} \sum_{r_2=0}^{g_2} \dots \sum_{r_m=0}^{g_m} P(\alpha_{1,v} \circ X_{1,v-1+ns} = r_1, \dots, \alpha_{m,v} \circ X_{m,v-1+ns} = r_m | \mathbf{X}_{v-1+ns} = \mathbf{x}_{v-1+ns}) \times$$

$$\times P(Z_{1,v+ns} = x_{1,v+ns} - r_1, \dots, Z_{m,v+ns} = x_{m,v+ns} - r_m) =$$

$$= \sum_{r_1=0}^{g_1} \sum_{r_2=0}^{g_2} \dots \sum_{r_m=0}^{g_m} f_1(r_1) f_2(r_2) \dots f_m(r_m) \; h(x_{1,v+ns} - r_1, x_{2,v+ns} - r_2, \dots, x_{m,v+ns} - r_m) \tag{2.71}$$

with $g_j = min(x_{j,v+ns}, x_{j,v-1+ns})$, $j = 1, \dots, m;\ v = 1, \dots, s$ and $n \in \mathbb{N}_0$. The conditional likelihood function is given by

$$\begin{aligned} L(\boldsymbol{\theta}|\mathbf{x}) &= \prod_{n=0}^{N-1} \prod_{v=1}^{s} P(\mathbf{X}_{v+ns} = \mathbf{x}_{v+ns} | \mathbf{X}_{v-1+ns} = \mathbf{x}_{v-1+ns}) = \\ &= \prod_{n=0}^{N-1} \prod_{v=1}^{s} p_v(\mathbf{x}_{v+ns} | \mathbf{x}_{v-1+ns}). \end{aligned} \tag{2.72}$$

The conditional maximum likelihood (CML) estimator, $\widehat{\boldsymbol{\theta}}^{CML}$, of the vector of unknown parameters $\boldsymbol{\theta}$ in (2.59) is obtained by maximizing $L(\boldsymbol{\theta}|\mathbf{x})$ which is equivalent to maximizing the conditional log-likelihood

$$C(\boldsymbol{\theta}) = ln(L(\boldsymbol{\theta}|\mathbf{x})) = \sum_{n=0}^{N-1} \sum_{v=1}^{s} ln\left(p_v(\mathbf{x}_{v+ns} | \mathbf{x}_{v-1+ns})\right)$$

with transition probabilities $p_v(\mathbf{x}_{v+ns} | \mathbf{x}_{v-1+ns})$ in equation (2.71). The first-order partial derivatives of function $C(\boldsymbol{\theta})$ are obtained through

$$\frac{\partial}{\partial \boldsymbol{\theta}} C(\boldsymbol{\theta}) = \sum_{n=0}^{N-1} \sum_{v=1}^{s} \frac{\frac{\partial}{\partial \boldsymbol{\theta}} p_v(\mathbf{x}_{v+ns} | \mathbf{x}_{v-1+ns})}{p_v(\mathbf{x}_{v+ns} | \mathbf{x}_{v-1+ns})},$$

when a particular joint distribution for the innovation process in (2.69) is assumed.

2.3.3 Composite likelihood estimation

For multivariate processes, the number of parameters can be quite large and for periodic multivariate processes even larger, the inflation of parameters is 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. To overcome the limitations in computing the exact likelihood, Lindsay (1988) proposed the composite likelihood as a pseudo-likelihood for inference. The pseudo-likelihood may take various forms such as combinations of likelihoods for small subsets of the data or combinations of conditional likelihoods. Pairwise likelihood is one special case of a composite likelihood, in which the pseudo-likelihood is defined as the product of the bivariate likelihood of all possible pairs of observations. A general discussion of pairwise likelihood can be found in Cox and Reid (2004) and Davis and Yau (2011).

Composite likelihood methods based on optimizing sums of log-likelihoods of low-dimensional margins have become popular in recent years; they are 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 (Varin, 2008). An excellent overview of composite likelihood methods can be found in Varin et al. (2011), complementing and extending the review made by Varin (2008). This concept of estimation has also been used by Pedeli and Karlis (2013a). 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. Analogues of the Akaike information criteria for model selection can be derived in the framework of composite likelihoods, having a similar form, see e.g. Varin and Vidoni (2005) and Ng and Joe (2014). Pairwise likelihood or bivariate composite likelihood methods are based on bivariate margins. The bivariate marginal log-likelihood function between two random elements, say X_a and X_b, is defined as

$$l_{ab}(\boldsymbol{\theta}; \mathbf{x}_a, \mathbf{x}_b) = \frac{1}{Ns} \sum_{n=0}^{N-1} \sum_{v=1}^{s} log f_{X_a,X_b}(x_{a,v+ns}, x_{b,v+ns} | x_{a,v-1+ns}, x_{b,v-1+ns}; \boldsymbol{\theta}), \tag{2.73}$$

where $\boldsymbol{\theta}$ is the vector of unknown parameters in (2.59) and function

$$f_{X_a,X_b}(x_{a,v+ns}, x_{b,v+ns}|x_{a,v-1+ns}, x_{b,v-1+ns}; \boldsymbol{\theta}) =$$
$$= \sum_{k_a=0}^{g_1} \sum_{k_b=0}^{g_2} \begin{pmatrix} x_{a,v-1+ns} \\ x_{a,v+ns} - k_a \end{pmatrix} \alpha_{a,v}^{x_{a,v+ns}-k_a} (1-\alpha_{a,v})^{x_{a,v-1+ns}-x_{a,v+ns}+k_a} \times$$
$$\times \begin{pmatrix} x_{b,v-1+ns} \\ x_{b,v+ns} - k_a \end{pmatrix} \alpha_{b,v}^{x_{b,v+ns}-k_b} (1-\alpha_{b,v})^{x_{b,v-1+ns}-x_{b,v+ns}+k_b} \times h_{R_a,R_b}(k_a, k_b) \tag{2.74}$$

with $g_1 = min(x_{a,v+ns}, x_{a,v-1+ns})$ 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 function between the corresponding innovation terms R_a and R_b. 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.,

$$cl(\boldsymbol{\theta}; \mathbf{x}_a, \mathbf{x}_b) = \sum_{a=1}^{m-1} \sum_{b=a+1}^{m} w_{ab}\ l_{ab}(\boldsymbol{\theta}; \mathbf{x}_a, \mathbf{x}_b), \tag{2.75}$$

where w_{ab} is a constant weight for l_{ab}. Typically, the weights are chosen in order to eliminate distant pairs of observations, which should be less informative Varin and Vidoni (2005). For sake of simplicity, it is common to set $w_{ab} = 1$, $1 \leq a \leq b \leq m$. Further details on weighting of bivariate margins in pairwise likelihood in Joe and Lee (2009). Asymptotic results and computational aspects of construction of, and inference from, composite likelihood are available from Varin et al. (2011).

2.4 PMINAR(1) Process with MVNB Innovations

This section is devoted to the `PMINAR`(1) model in (2.23) with a specific multivariate distribution for the innovations. Recall the assumption of diagonality of the autocorrelation matrix, thus correlation between the innovations is the only source of dependence between the series $\mathbf{X}_{j,t}$ ($j = 1, \ldots, m$). Therefore, the choice of the joint distribution for the ms-dimensional random vector of innovations $\mathbf{Z}_{j,t}$ with $t = v + ns; v = 1, \ldots, s$ and $j = 1, \ldots, m$ is quite relevant since it determines the properties of the underlying process. Monteiro et al. (2015) generalized the bivariate model proposed by Pedeli and Karlis (2011) to the periodic case by assuming periodic bivariate sequences of innovations. Two different distributional forms of the innovations have been proposed in both papers: bivariate Poisson and bivariate negative binomial. Much attention has been devoted to the Poisson distribution for the innovation process. However, implying equidispersion (mean equals variance) in real-life events may not reflect the true nature of the data, limiting the applicability of the Poisson distribution. In the periodic bivariate case, Monteiro et al. (2015) has shown the bivariate negative binomial distribution for the underlying innovations series allows for more flexibility, due to the involvement of the overdispersion parameter, than the same model with Poisson innovations. Thus, in the sequel, the distribution of the innovation processes is assumed to be periodic multivariate negative binomial (MVNB) distribution, which can account for overdispersion (variance exceeds mean), a common feature in real data applications.

2.4.1 Multivariate negative binomial distribution and basic properties

For a fixed v $(v = 1, \ldots, s)$, let $\tilde{\boldsymbol{\lambda}}_v = [\lambda_{1,v}\ \lambda_{2,v}\ \ldots\ \lambda_{m,v}]^T$ with positive $\lambda_{j,v}$ $(j = 1, \ldots, m)$ and positive dispersion parameter β_v. Let $\mathbf{Z}_{j,t}$ with $t = v + ns$ be random variables having the Poisson distribution with mean $\eta\lambda_{j,v}$, where η is a r.v. which represents an unobserved heterogeneity that follows a Gamma $(\beta_v^{-1}, \beta_v^{-1})$ distribution. In the aforementioned setting, the innovations $\mathbf{Z}_{j,t}$ follow a multivariate negative binomial distribution, denoted by MVNB$(\tilde{\boldsymbol{\lambda}}_v, \beta_v)$. Hence, the joint probability mass function in (2.69) now takes the form

$$h(z_1, z_2, \ldots, z_m) = P(Z_{1,v+ns} = z_1, Z_{2,v+ns} = z_2, \ldots, Z_{m,v+ns} = z_m) =$$
$$= \frac{\Gamma\left(\beta_v^{-1} + \sum_{j=1}^m z_j\right)}{\Gamma(\beta_v^{-1})} \left(\frac{\beta_v^{-1}}{\beta_v^{-1} + \sum_{j=1}^m \lambda_{j,v}}\right)^{\beta_v^{-1}} \left(\beta_v^{-1} + \sum_{j=1}^{m} \lambda_{j,v}\right)^{-\sum_{j=1}^m z_j} \prod_{j=1}^{m} \frac{\lambda_{j,v}^{z_j}}{z_j!} \qquad (2.76)$$

for $(z_1, z_2, \ldots, z_m) \in \mathbb{N}_0^m$.

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

$$p_{j,v} = \frac{\beta_v^{-1}}{\lambda_{j,v} + \beta_v^{-1}}. \tag{2.77}$$

The multivariate negative binomial distribution defined in (2.76) has also been used by Marshall and Olkin (1990), Boucher et al. (2008), Cheon et al. (2009) and more recently, by Pedeli and Karlis (2011, 2013a).

As previously mentioned, 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 $E[\mathbf{Z}_t] = \boldsymbol{\delta}_t$ in (2.14) and variance-covariance matrix $\sum_{\mathbf{Z}_t} = \psi_t$ in (2.17). Thus, with the specification of a MVNB distribution for the innovation process $\{\mathbf{Z}_t\}$:

$$E[Z_{j,v+ns}] = \beta_v^{-1}\,\frac{1 - p_{j,v}}{p_{j,v}} = \lambda_{j,v}, \tag{2.78}$$

$$Var[Z_{j,v+ns}] = \sigma^2_{j,v} = \beta_v^{-1}\,\frac{1 - p_{j,v}}{p^2_{j,v}} = \lambda_{j,v}(1 + \beta_v \lambda_{j,v}), \tag{2.79}$$

$$Cov(Z_{j,v+ns}, Z_{k,v+ns}) = \sigma_{jk,v} = \beta_v \lambda_{j,v} \lambda_{k,v} \tag{2.80}$$

for a fixed v $(v = 1, \ldots, s)$, $j \neq k;\ j, k = 1, \ldots, m$ and probability $p_{j,v}$ defined in (2.77). The mean of $Z_{j,v+ns}$ in (2.78) is equal to (2.16). The variance and covariance of $Z_{j,v+ns}$ in (2.21) and (2.19), now with MVNB innovations, take the form in (2.79) and (2.80), respectively. Hence, $Var[Z_{j,v+ns}]$ exceeds $E[Z_{j,v+ns}]$, this setting clearly accounts for overdispersion. For each season v, the covariance between two components defined in (2.80) is always positive.

Using the above specification for the joint distribution of the innovation process $\{\mathbf{Z}_t\}$ in (2.76), we can now define a `PMINAR`(1) model with MVNB innovations. The vector of expectations $\boldsymbol{\mu}_{j,t}$ $(j = 1, \ldots, m)$ defined in (2.33) has the same elements as equation (2.32) because $E[Z_{j,v+ns}] = \lambda_{j,v}$ in (2.78), i.e.,

$$E[X_{j,v+ns}] = \frac{\sum\limits_{k=0}^{v-1} \varphi^{(j)}_{v,k} \lambda_{j,v-k} + \varphi^{(j)}_{v,v} \sum\limits_{i=0}^{s-(v+1)} \varphi^{(j)}_{s,i} \lambda_{j,s-i}}{1 - \varphi^{(j)}_{s,s}} \tag{2.81}$$

for a fixed v ($v = 1, \ldots, s$). From (2.78) and (2.79), the variance-covariance matrix $\sum_{\mathbf{X}_t}$ in (2.48) has symmetric matrices $\sum_{j,j}$ ($j = 1, \ldots, m$) with diagonal elements equal to

$$\begin{aligned} &Var[X_{j,v+ns}] = \\ &= \frac{1}{1-\left(\varphi_{s,s}^{(j)}\right)^2}\left\{\sum_{k=0}^{v-1}\left[\varphi_{s,s}^{(j)}\varphi_{v,k}^{(j)}\lambda_{j,v-k} + \varphi_{v,k}^{(j)}\left(1-\varphi_{v,k}^{(j)}\right)\lambda_{j,v-k} + \right.\right. \\ &\left. + \left(\varphi_{v,k}^{(j)}\right)^2 \lambda_{j,v-k}(1+\beta_{v-k}\lambda_{j,v-k})\right] + \sum_{m=0}^{s-(v+1)}\left[\varphi_{s,s}^{(j)}\varphi_{v,v}^{(j)}\varphi_{s,m}^{(j)}\lambda_{j,s-m} + \right. \\ &\left.\left. + \varphi_{v,v}^{(j)}\varphi_{s,m}^{(j)}\left(1-\varphi_{v,v}^{(j)}\varphi_{s,m}^{(j)}\right)\lambda_{j,s-m} + \left(\varphi_{v,v}^{(j)}\varphi_{s,m}^{(j)}\right)^2\lambda_{j,s-k}(1+\beta_{s-k}\lambda_{j,s-k})\right]\right\} \end{aligned} \tag{2.82}$$

and off-diagonal elements equal to

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

2.4.2 Parameter estimation with MVNB innovations

For the `PMINAR`(1) model with multivariate negative binomial innovations, the vector of unknown parameters $\boldsymbol{\theta}$ in (2.59) is a $(2m+1)s$-dimensional vector as

$$\boldsymbol{\theta} := (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m, \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_m, \boldsymbol{\beta}) \tag{2.83}$$

with s-dimensional vectors

$$\boldsymbol{\alpha}_j = (\alpha_{j,1}, \ldots, \alpha_{j,s}) \ ; \ \boldsymbol{\lambda}_j = (\lambda_{j,1}, \ldots, \lambda_{j,s}) \ ; \ \boldsymbol{\beta} = (\beta_1, \ldots, \beta_s), \ j = 1, \ldots, m. \tag{2.84}$$

2.4.2.1 Yule-Walker estimation

The Yule-Walker (YW) estimator of the vector of the $(2m+1)s$ unknown parameters in (2.83) is $\widehat{\boldsymbol{\theta}}_{YW} := (\widehat{\boldsymbol{\alpha}}_1^{YW}, \ldots, \widehat{\boldsymbol{\alpha}}_m^{YW}, \widehat{\boldsymbol{\lambda}}_1^{YW}, \ldots, \widehat{\boldsymbol{\lambda}}_m^{YW}, \widehat{\boldsymbol{\beta}}^{YW})$. The YW estimators $\widehat{\lambda}_{j,v}^{YW}$ and $\widehat{\alpha}_{j,v}^{YW}$ ($j = 1, \ldots, m$) for parameters $\lambda_{j,v}$ in $\boldsymbol{\lambda}_j$ and $\alpha_{j,v}$ in $\boldsymbol{\alpha}_j$ from (2.84), are defined in (2.66) and in (2.68), respectively. The YW estimator for parameter β_v, $\widehat{\beta}_v^{YW}$ $(v = 1, \dots, s)$, follows from equations (2.52) and (2.80),

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

then

$$\beta_v = \frac{\left(1-\varphi_{s,s}^{(j)}\varphi_{s,s}^{(k)}\right) Cov(X_{j,v+ns}, X_{k,v+ns})}{\sum_{i=0}^{v-1} \varphi_{v,i}^{(j)}\varphi_{v,i}^{(k)}\lambda_{j,v-i}\lambda_{k,v-i} + \varphi_{v,v}^{(j)}\varphi_{v,v}^{(k)} \sum_{i=0}^{s-(v+1)} \varphi_{s,i}^{(j)}\varphi_{s,i}^{(k)}\lambda_{j,s-i}\lambda_{k,s-i}}.$$

The sample equivalent of $Cov(X_{j,v+ns}, X_{k,v+ns})$ expressed in (2.65) yields the YW estimator for parameter β_v

$$\widehat{\beta}_v^{YW} = \frac{\left(1-\widehat{\varphi}_{s,s}^{(j)}\widehat{\varphi}_{s,s}^{(k)}\right) \gamma_{jk,v}(0)}{\sum_{i=0}^{v-1} \widehat{\varphi}_{v,i}^{(j)}\widehat{\varphi}_{v,i}^{(k)}\widehat{\lambda}_{j,v-i}\widehat{\lambda}_{k,v-i} + \widehat{\varphi}_{v,v}^{(j)}\widehat{\varphi}_{v,v}^{(k)} \sum_{i=0}^{s-(v+1)} \widehat{\varphi}_{s,i}^{(j)}\widehat{\varphi}_{s,i}^{(k)}\widehat{\lambda}_{j,s-i}\widehat{\lambda}_{k,s-i}}, \tag{2.85}$$

for $v = 1, \dots, s$ and $j \neq k; j, k = 1, \dots, m$.

2.4.2.2 Conditional maximum likelihood estimation

The conditional maximum likelihood (CML) estimator of the vector of the $(2m+1)s$ unknown parameters in (2.83) is $\widehat{\boldsymbol{\theta}}_{CML} := (\widehat{\boldsymbol{\alpha}}_1^{CML}, \dots, \widehat{\boldsymbol{\alpha}}_m^{CML}, \widehat{\boldsymbol{\lambda}}_1^{CML}, \dots, \widehat{\boldsymbol{\lambda}}_m^{CML}, \widehat{\boldsymbol{\beta}}^{CML})$. Thus assuming MVNB innovations, the conditional density defined in (2.71) takes the form

$$p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns}) =$$
$$= \sum_{r_1=0}^{g_1} \sum_{r_2=0}^{g_2} \dots \sum_{r_m=0}^{g_m} \left(\prod_{j=1}^{m} f_j(r_j) \right) h(x_{1,v+ns} - r_1, x_{2,v+ns} - r_2, \dots, x_{m,v+ns} - r_m) =$$

$$= \sum_{r_1=0}^{g_1} \sum_{r_2=0}^{g_2} \dots \sum_{r_m=0}^{g_m} \left(\prod_{j=1}^{m} f_j(r_j) \right) \frac{\Gamma\left(\beta_v^{-1} + \sum_{j=1}^{m}(x_{j,v+ns} - r_j)\right)}{\Gamma(\beta_v^{-1})} \left(\frac{\beta_v^{-1}}{\beta_v^{-1} + \sum_{j=1}^{m} \lambda_{j,v}} \right)^{\beta_v^{-1}} \times$$

$$\times \left(\beta_v^{-1} + \sum_{j=1}^{m} \lambda_{j,v} \right)^{-\sum_{j=1}^{m}(x_{j,v+ns}-r_j)} \prod_{j=1}^{m} \frac{\lambda_{j,v}^{(x_{j,v+ns}-r_j)}}{(x_{j,v+ns} - r_j)!} \tag{2.86}$$

with $g_j = min(x_{j,v-1+ns}, x_{j,v+ns})$ and $f_j(r_j)$ in (2.70) for $j = 1, \dots, m$. The conditional log-likelihood function is given by

$$C(\boldsymbol{\theta}) = ln(L(\boldsymbol{\theta}|\mathbf{x})) = \sum_{n=0}^{N-1} \sum_{v=1}^{s} ln\left(p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})\right) \tag{2.87}$$

with transition probabilities $p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})$ defined in equation (2.86). Hence, the first-order partial derivatives of the conditional log-likelihood $C(\boldsymbol{\theta})$ in (2.87) are obtained through

$$\frac{\partial}{\partial \boldsymbol{\theta}} C(\boldsymbol{\theta}) = \sum_{n=0}^{N-1} \sum_{v=1}^{s} \frac{\frac{\partial}{\partial \boldsymbol{\theta}} p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})}{p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})}. \tag{2.88}$$

For a fixed v $(v = 1, \dots, s)$, let the vector of unknown parameters be

$$\eta_v = (\alpha_{1,v}, \dots, \alpha_{m,v}, \lambda_{1,v}, \dots, \lambda_{m,v}, \beta_v). \tag{2.89}$$

The first-order partial derivatives of function $p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})$ (p_v for short) in (2.86) with respect to the autocorrelation coefficients $\alpha_{j,v}$ $(j = 1, \dots, m)$ are:

$$\frac{\partial p_v}{\partial \alpha_{1,v}} = \frac{x_{1,v-1+ns}}{1 - \alpha_{1,v}} \left[p_v(\mathbf{x}_{v+ns} - (1, 0, \dots, 0)|\mathbf{x}_{v-1+ns} - (1, 0, \dots, 0)) - p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns}) \right],$$

$$\frac{\partial p_v}{\partial \alpha_{2,v}} = \frac{x_{2,v-1+ns}}{1 - \alpha_{2,v}} \left[p_v(\mathbf{x}_{v+ns} - (0, 1, \dots, 0)|\mathbf{x}_{v-1+ns} - (0, 1, \dots, 0)) - p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns}) \right],$$

$$\vdots$$

$$\frac{\partial p_v}{\partial \alpha_{m,v}} = \frac{x_{m,v-1+ns}}{1 - \alpha_{m,v}} \left[p_v(\mathbf{x}_{v+ns} - (0, 0, \dots, 1)|\mathbf{x}_{v-1+ns} - (0, 0, \dots, 1)) - p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns}) \right].$$

Partial derivatives of function p_v with respect to the parameters $\lambda_{j,v}$ $(j = 1, \ldots, m)$ and β_v, regarding the multivariate function $h(z_1, z_2, \ldots, z_m)$ defined in (2.76) are:

$$\frac{\partial p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})}{\partial \lambda_{j,v}} = \sum_{r_1=0}^{g_1}\sum_{r_2=0}^{g_2}\cdots\sum_{r_m=0}^{g_m}\left(\prod_{j=1}^{m} f_j(r_j)\right)\left\{\frac{z_j}{\lambda_{j,v}} - \left(\beta_v^{-1} + \sum_{j=1}^{m} z_j\right) \times \right.$$
$$\left. \times \left(\beta_v^{-1} + \sum_{j=1}^{m}\lambda_{j,v}\right)^{-1}\right\} h(z_1, z_2, \ldots, z_m) \tag{2.90}$$

and

$$\frac{\partial p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})}{\partial \beta_v} = \sum_{r_1=0}^{g_1}\sum_{r_2=0}^{g_2}\cdots\sum_{r_m=0}^{g_m}\left(\prod_{j=1}^{m} f_j(r_j)\right)\beta_v^{-2}\left[\psi\left(\beta_v^{-1}\right) - \psi\left(\beta_v^{-1} + \sum_{j=1}^{m} z_j\right) + \right.$$
$$\left. + \, ln\left(1 + \beta_v \sum_{j=1}^{m}\lambda_{j,v}\right) + \frac{1}{\beta_v^{-1} + \sum_{j=1}^{m}\lambda_{j,v}} - 1\right] h(z_1, z_2, \ldots, z_m) \tag{2.91}$$

with $z_j = x_{j,v+ns} - r_j$, $j = 1, \ldots, m$. For $w \neq v$, $v = 1, \ldots, s$: $\dfrac{\partial p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})}{\partial \eta_w} = 0$. First-order partial derivatives of the transition probability function $p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})$ are available in Appendix B.2.

Differentiating the conditional log-likelihood function $C(\boldsymbol{\theta})$ in (2.87) partially with respect to all $(2m+1)s$ parameters and setting the derivatives in (2.88) to zero, we obtain the following system of first-order partial derivatives:

$$\begin{cases} \dfrac{\partial C(\boldsymbol{\theta})}{\partial \alpha_{j,v}} = 0, \ j = 1, 2, \ldots, m \\[2ex] \dfrac{\partial C(\boldsymbol{\theta})}{\partial \lambda_{j,v}} = 0, \ j = 1, 2, \ldots, m \qquad (v = 1, \ldots, s), \\[2ex] \dfrac{\partial C(\boldsymbol{\theta})}{\partial \beta_v} = 0 \end{cases}$$

i.e.,

$$\begin{cases} \prod_{n=0}^{N-1} \frac{x_{1,v-1+ns}}{1-\alpha_{1,v}} \left(\frac{p_v(\mathbf{x}_{v+ns} - (1,0,\ldots,0)|\mathbf{x}_{v-1+ns} - (1,0,\ldots,0))}{p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})} - 1 \right) = 0 \\ \prod_{n=0}^{N-1} \frac{x_{2,v-1+ns}}{1-\alpha_{2,v}} \left(\frac{p_v(\mathbf{x}_{v+ns} - (0,1,\ldots,0)|\mathbf{x}_{v-1+ns} - (0,1,\ldots,0))}{p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})} - 1 \right) = 0 \\ \vdots \\ \prod_{n=0}^{N-1} \frac{x_{m,v-1+ns}}{1-\alpha_{m,v}} \left(\frac{p_v(\mathbf{x}_{v+ns} - (0,0,\ldots,1)|\mathbf{x}_{v-1+ns} - (0,0,\ldots,1))}{p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})} - 1 \right) = 0 \\ \prod_{n=0}^{N-1} \frac{1}{p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})} \frac{\partial p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})}{\partial \lambda_{1,v}} = 0 \\ \prod_{n=0}^{N-1} \frac{1}{p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})} \frac{\partial p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})}{\partial \lambda_{2,v}} = 0 \\ \vdots \\ \prod_{n=0}^{N-1} \frac{1}{p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})} \frac{\partial p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})}{\partial \lambda_{m,v}} = 0 \\ \prod_{n=0}^{N-1} \frac{1}{p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})} \frac{\partial p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})}{\partial \beta_v} = 0 \end{cases}$$

with $\frac{\partial p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})}{\partial \lambda_{j,v}}$ $(j = 1, 2, \ldots, m)$ in (2.90) and $\frac{\partial p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})}{\partial \beta_v}$ in (2.91). The above system of equations does not provide explicit CML estimators for the parameters. However, they can be numerically obtained by using common statistical packages in R. Asymptotic properties of the CML estimator $\widehat{\boldsymbol{\theta}}_{CML}$ of $\boldsymbol{\theta}$ are given below. Results from Billingsley (1961) are applied.

Theorem 2.2. *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

$$I = \begin{bmatrix} M_1 & 0 & \cdots & 0 \\ 0 & M_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & M_s \end{bmatrix}$$

with matrices M_v $(v = 1, \ldots, s)$ *given by*

$$\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 theorem is a particular case of theorem 2.2 in Billingsley (1961). For each season v $(v = 1, \ldots, s)$, $p_v(\cdot|\cdot)$ is the transition probabilities in (2.86) of the PMINAR(1) model, therefore the regularity conditions in Billingsley's Theorem 2.2 are satisfied. We postpone those assumptions to the Appendix B.3. □

2.4.2.3 Composite likelihood estimation

The composite likelihood (CL) estimator of the vector of the $(2m+1)s$ unknown parameters in (2.83) is $\widehat{\boldsymbol{\theta}}_{CL} := (\widehat{\boldsymbol{\alpha}}_1^{CL}, \ldots, \widehat{\boldsymbol{\alpha}}_m^{CL}, \widehat{\boldsymbol{\lambda}}_1^{CL}, \ldots, \widehat{\boldsymbol{\lambda}}_m^{CL}, \widehat{\boldsymbol{\beta}}^{CL})$. The bivariate marginal log-likelihood function between two random elements X_a and X_b can be defined as

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

where the corresponding bivariate marginal probability density with bivariate negative binomial innovations is given by (2.74). The bivariate distribution between the innovation terms Z_a and Z_b, $h_{Z_a,Z_b}(k_a, k_b)$, is a particular case $(m = 2)$ of the multivariate negative binomial distribution in (2.76), therefore,

$$\begin{aligned}
&f_{X_a,X_b}(x_{a,v+ns}, x_{b,v+ns}|x_{a,v-1+ns}, x_{b,v-1+ns}; \boldsymbol{\theta}) = \\
&= \sum_{k_a=0}^{g_1}\sum_{k_b=0}^{g_2} \begin{pmatrix} x_{a,v-1+ns} \\ x_{a,v+ns}-k_a \end{pmatrix} \alpha_{a,v}^{x_{a,v+ns}-k_a}(1-\alpha_{a,v})^{x_{a,v-1+ns}-x_{a,v+ns}+k_a} \times \\
&\times \begin{pmatrix} x_{b,v-1+ns} \\ x_{b,v+ns}-k_b \end{pmatrix} \alpha_{b,v}^{x_{b,v+ns}-k_b}(1-\alpha_{b,v})^{x_{b,v-1+ns}-x_{b,v+ns}+k_b} \times \frac{\Gamma\left(\beta_v^{-1}+k_a+k_b\right)}{\Gamma(\beta_v^{-1})} \times \\
&\times \left(\frac{\beta_v^{-1}}{\beta_v^{-1}+\lambda_{a,v}+\lambda_{b,v}}\right)^{\beta_v^{-1}} \left(\beta_v^{-1}+\lambda_{a,v}+\lambda_{b,v}\right)^{-(k_a+k_b)} \frac{\lambda_{a,v}^{k_a}}{k_a!}\frac{\lambda_{b,v}^{k_b}}{k_b!} \qquad (2.92)
\end{aligned}$$

with $g_1 = min(x_{a,v+ns}, x_{a,v-1+ns})$ and $g_2 = min(x_{b,v+ns}, x_{b,v-1+ns})$. The composite log-likelihood function $cl(\boldsymbol{\theta}; \mathbf{x}_a, \mathbf{x}_b)$ is comprised of all the bivariate log-likelihood functions

$$cl(\boldsymbol{\theta}; \mathbf{x}_a, \mathbf{x}_b) = \sum_{a=1}^{m-1}\sum_{b=a+1}^{m} w_{ab}\ l_{ab}(\boldsymbol{\theta}; \mathbf{x}_a, \mathbf{x}_b), \qquad (2.93)$$

where $w_{a,b}$ is a constant weight for l_{ab}.

2.5 Forecasting

We consider the forecasting of future values $\mathbf{X}_{t+h}$ $(t = v+ns;\ v = 1, \ldots, s)$ of the periodic `MINAR`(1) process, given past observations through time $t = v + ns$ for $v = 1, \ldots, s$. Let $h = u + ls$ for $u = 1, \ldots, s$ throughout this section. Due to the definition of the model and by iterating equation (2.5), the j-th component $\mathbf{X}_{j,t}$ can be expressed as

$$\mathbf{X}_{j,t} \overset{d}{=} \left(\prod_{i=0}^{n-1}\phi_{j,t-i}\right) \circ \mathbf{X}_{j,t-n} + \sum_{k=1}^{n-1}\left(\prod_{i=0}^{k-1}\phi_{j,t-i}\right) \circ \mathbf{Z}_{j,t-k} + \mathbf{Z}_{j,t}$$

with $\phi_{j,t}$ defined in (2.4) and $\mathbf{Z}_{j,t}$ in (2.2). Then

$$\mathbf{X}_{j,t} \overset{d}{=} \zeta_{t,n}^{(j)} \circ \mathbf{X}_{j,t-n} + \sum_{k=0}^{n-1} \zeta_{t,k}^{(j)} \circ \mathbf{Z}_{j,t-k}, \qquad (2.94)$$

where, for $t \geq i$

$$\zeta_{t,i} := \begin{cases} \prod_{k=0}^{i-1} \phi_{t-k} & ,i > 0 \\ 1 & ,i = 0 \end{cases}$$

and also $\zeta_{t,i} := \zeta_{t,v}(\zeta_{s,s})^l$ for $i = v + ls;\ v = 1, \ldots, s$, leading to

$$X_{j,v+ns+h} \stackrel{d}{=} \zeta^{(j)}_{v+ns+h,h} \circ X_{j,v+ns} + \sum_{k=0}^{h-1} \zeta^{(j)}_{v+ns+h,k} \circ Z_{j,v+ns+h-k}.$$

Since $h = u + ls$ for $u = 1, \ldots, s$, it follows that

$$\begin{aligned} X_{j,v+ns+h} &\stackrel{d}{=} \zeta^{(j)}_{v+u+(n+l)s,u+ls} \circ X_{j,v+ns} + \sum_{k=1}^{u+ls-1} \zeta^{(j)}_{v+u+(n+l)s,k} \circ Z_{j,v+u+(n+l)s-k} \\ &\stackrel{d}{=} \zeta^{(j)}_{v+u,u} \left(\zeta^{(j)}_{s,s}\right)^l \circ X_{j,v+ns} + Y_{j,v+u+ls} \end{aligned}$$

with

$$Y_{j,v+u+ls} = \sum_{k=0}^{v-1} \zeta^{(j)}_{v+u,k} \circ Z_{j,v+u+ns-k} + \sum_{w=0}^{l-1} \sum_{k=0}^{s-1} \zeta^{(j)}_{v+u+(n+l)s,k+u+ws} \circ Z_{j,v+(n+l-w)s-k}.$$

One way to generate the h-step ahead prediction is to employ the mean, median or mode of the predictive distribution of $\mathbf{X}_{v+ns+h}|\mathbf{X}_{v+ns}$ as a point forecast. The median and mode are considered as coherent predictions (integer-valued) but the mean is not. The h-step ahead point predictor that minimizes the mean square error (MSE) is given by

$$\begin{aligned} \widehat{X}_{j,v+ns+h} &= E[X_{j,v+ns+h}|X_{j,v+ns}] \\ &= E\left[\zeta^{(j)}_{v+u,u} \left(\zeta^{(j)}_{s,s}\right)^l \circ X_{j,v+ns}|X_{j,v+ns}\right] + E[Y_{j,v+u+ls}], \end{aligned} \tag{2.95}$$

where

$$\begin{aligned} E[Y_{j,v+u+ls}] &= E\left[\sum_{k=0}^{v-1} \zeta^{(j)}_{v+u,k} \circ Z_{j,v+u+ns-k} + \sum_{w=0}^{l-1} \sum_{k=0}^{s-1} \zeta^{(j)}_{v+u+(n+l)s,k+u+ws} \circ Z_{j,v+(n+l-w)s-k}\right] = \\ &= \sum_{k=0}^{v-1} \zeta^{(j)}_{v+u,k} \lambda_{j,v+u-k} + \sum_{w=0}^{l-1} \sum_{k=0}^{s-1} \zeta^{(j)}_{v+u+(n+l)s,k+u+ws} \lambda_{j,v+(n+l-w)s-k} \end{aligned} \tag{2.96}$$

with $E[Z_{j,v+ns}] = \lambda_{j,v}$ in (2.16). For the particular case, $h = 1$, the one-step ahead predictive function is

$$p_v(\mathbf{x}_{v+1+ns}|\mathbf{x}_{v+ns}) = \\ = \sum_{r_1=0}^{g_1}\sum_{r_2=0}^{g_2}\cdots\sum_{r_m=0}^{g_m}\left(\prod_{j=1}^{m} f_j(r_j)\right) h(x_{1,v+1+ns}-r_1, x_{2,v+1+ns}-r_2,\ldots,x_{m,v+1+ns}-r_m) \tag{2.97}$$

with $g_j = min(x_{j,v+ns}, x_{j,v+1+ns})$, $j = 1,\ldots,m$ and MVNB distribution defined in (2.76) takes the form

$$h(x_{1,v+1+ns}-r_1, x_{2,v+1+ns}-r_2,\ldots,x_{m,v+1+ns}-r_m) = \\ = \frac{\Gamma\left(\beta_v^{-1}+\sum_{j=1}^{m}(x_{j,v+1+ns}-r_j)\right)}{\Gamma(\beta_v^{-1})}\left(\frac{\beta_v^{-1}}{\beta_v^{-1}+\sum_{j=1}^{m}\lambda_{j,v}}\right)^{\beta_v^{-1}}\times \\ \times\left(\beta_v^{-1}+\sum_{j=1}^{m}\lambda_{j,v}\right)^{-\sum_{j=1}^{m}(x_{j,v+1+ns}-r_j)}\prod_{j=1}^{m}\frac{\lambda_{j,v}^{(x_{j,v+1+ns}-r_j)}}{(x_{j,v+1+ns}-r_j)!}.$$

Furthermore, from equations (2.81) and (2.96), the one-step ahead predictor of $X_{j,v+ns+1}$ takes the form

$$\widehat{X}_{j,v+1+ns} = E[X_{j,v+1+ns}|X_{j,v+ns}] \\ = \frac{\sum_{k=0}^{v}\varphi_{v+1,k}^{(j)}\lambda_{j,v+1-k}+\varphi_{v+1,v+1}^{(j)}\sum_{i=0}^{s-(v+2)}\varphi_{s,i}^{(j)}\lambda_{j,s-i}}{1-\varphi_{s,s}^{(j)}}+ \\ +\sum_{k=0}^{v-1}\zeta_{v+1,k}^{(j)}\lambda_{j,v+1-k}+\sum_{w=0}^{l-1}\sum_{k=0}^{s-1}\zeta_{v+1+(n+l)s,k+1+ws}^{(j)}\lambda_{j,v+(n+l-w)s-k}. \tag{2.98}$$

In order to evaluate the prediction performance given by the mean, median or mode of the predictive distribution, Monteiro et al. (2015) has considered the square root of the mean squared error (RMSE), the mean absolute error (MAE) or the loss function everything or nothing (LFEN).

2.6 Simulation study

The Yule-Walker (YW), conditional maximum likelihood (CML) and composite likelihood (CL) estimators of the `PMINAR`(1) model were compared through a simulation experiment for $m = 3$ (trivariate) and with periodic trivariate negative binomial innovations. The choice for this dimension is due to the complexity of the model and also because a real data application with three series is presented in the following section. Hence, a simulation study in the trivariate context is suitable. The simulation study was carried out in `R` using the `optim` function for the optimization of the likelihood functions and adopting convenient parameter transformations. See Appendix D.1 for `R` functions concerning the data generation and estimation in the present scenario.

Count series were generated assuming the innovation process $\{\mathbf{Z}_t\}$ follows jointly a periodic trivariate negative binomial distribution with parameters $(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3, \boldsymbol{\beta})$. We have set period $s = 4$, thus the vector of unknown parameters $\boldsymbol{\theta}$ in (2.83) is $\boldsymbol{\theta} := (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3, \boldsymbol{\beta})$ with $\boldsymbol{\alpha}_j = (\alpha_{j,1}, \alpha_{j,2}, \alpha_{j,3}, \alpha_{j,4})$, $\boldsymbol{\lambda}_j = (\lambda_{j,1}, \lambda_{j,2}, \lambda_{j,3}, \lambda_{j,4})$ for $j = 1, 2, 3$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)$, leading to a total of 28 parameters. This simulation 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, in particular, $n = 400, 1000, 2000$. Since $n = sN$ then we have $N = 100, 250, 500$ complete cycles. For each experiment we conducted 200 independent replications.

The simulated data sets that produced YW estimates in an inadmissible range were disregarded and iterations were continued till reaching the specified number of 200 replications per experiment. 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. Comparison of the YW, CML and CL estimators was made in terms of the mean square error (MSE) and the biases of the produced estimates. Tables 2.1-2.3 summarize the estimates of the parameters of the periodic trivariate `INAR`(1) model with trivariate negative binomial innovations and includes MSE in parenthesis.

Table 2.1: 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		
$(\alpha_{j,1}, \alpha_{j,2}, \alpha_{j,3}, \alpha_{j,4})$	YW	CML	CL	YW	CML	CL	YW	CML	CL
$(0.53, 0.75, 0.62, 0.83)$									
$\hat{\alpha}_{1,1}$	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)
$(0.72, 0.85, 0.56, 0.91)$									
$\hat{\alpha}_{2,1}$	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)
$(0.83, 0.60, 0.41, 0.58)$									
$\hat{\alpha}_{3,1}$	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)

Table 2.2: YW, CML and CL estimates for $\boldsymbol{\lambda}_j = (\lambda_{j,1}, \lambda_{j,2}, \lambda_{j,3}, \lambda_{j,4})$ with $j = 1, 2, 3$. Mean square error in parenthesis.

	n =400			n =1000			n =2000		
$(\lambda_{j,1}, \lambda_{j,2}, \lambda_{j,3}, \lambda_{j,4})$	YW	CML	CL	YW	CML	CL	YW	CML	CL
$(4, 2, 3, 5)$									
$\hat{\lambda}_{1,1}$	4.182	3.909	3.927	3.986	3.981	3.988	3.995	3.999	4.033
	(0.5357)	(0.7701)	(0.1028)	(0.0556)	(0.0012)	(0.0059)	(0.0651)	(0.0164)	(0.0179)
$\hat{\lambda}_{1,2}$	2.011	2.008	2.012	2.001	1.997	2.005	1.977	2.009	2.007
	(0.0109)	(0.0829)	(0.2645)	(0.0396)	(0.1528)	(0.1499)	(0.1510)	(0.0042)	(0.0057)
$\hat{\lambda}_{1,3}$	3.146	3.024	3.027	3.078	2.980	2.984	3.063	2.978	2.986
	(0.3970)	(1.1090)	(0.0103)	(0.7042)	(0.0499)	(0.0576)	(0.0712)	(0.0520)	(0.0502)
$\hat{\lambda}_{1,4}$	5.289	5.059	5.067	5.011	5.027	5.031	5.031	4.983	4.997
	(2.2418)	(0.3154)	(0.0003)	(0.0153)	(0.1633)	(0.1761)	(0.2178)	(0.0196)	(0.0157)
$(5, 3, 1.2, 2)$									
$\hat{\lambda}_{2,1}$	5.243	4.924	4.954	4.965	5.010	5.022	4.965	5.005	5.042
	(0.2664)	(1.0982)	(0.2708)	(0.0958)	(0.0391)	(0.0247)	(0.6790)	(0.0912)	(0.1081)
$\hat{\lambda}_{2,2}$	3.071	3.005	3.017	3.040	2.970	2.986	3.009	3.005	3.008
	(0.0299)	(0.2824)	(0.0124)	(0.0196)	(0.4746)	(0.4608)	(0.1941)	(0.00002)	(0.00003)
$\hat{\lambda}_{2,3}$	1.324	1.220	1.215	1.221	1.196	1.201	1.203	1.188	1.184
	(0.0097)	(0.2301)	(0.0309)	(0.0902)	(0.0027)	(0.0021)	(0.0170)	(0.0161)	(0.0156)
$\hat{\lambda}_{2,4}$	2.110	2.026	2.029	1.986	1.992	1.993	2.027	1.998	2.001
	(1.0020)	(0.1524)	(0.0143)	(0.1429)	(0.1044)	(0.0899)	(0.0002)	(0.0004)	(0.0007)
$(3, 1.6, 2, 4)$									
$\hat{\lambda}_{3,1}$	3.061	2.929	2.942	2.950	2.986	2.994	2.989	2.992	3.017
	(0.5619)	(1.0935)	(0.0318)	(0.0083)	(0.0073)	(0.0019)	(0.0397)	(0.0444)	(0.0507)
$\hat{\lambda}_{3,2}$	1.619	1.587	1.590	1.591	1.587	1.592	1.598	1.606	1.601
	(0.0149)	(0.0922)	(0.0391)	(0.0788)	(0.0901)	(0.0764)	(0.0521)	(0.0797)	(0.0956)
$\hat{\lambda}_{3,3}$	2.150	2.004	2.008	2.027	2.006	2.003	2.036	1.987	1.989
	(0.0057)	(0.4347)	(0.2044)	(0.0127)	(0.0425)	(0.0500)	(0.0002)	(0.0250)	(0.0204)
$\hat{\lambda}_{3,4}$	4.167	4.009	4.015	4.039	3.981	3.978	4.001	3.997	4.001
	(0.1478)	(0.3033)	(0.2086)	(0.1328)	(0.1158)	(0.0882)	(0.0289)	(0.0594)	(0.0582)

Table 2.3: YW, CML and CL estimates for $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)$. Mean square error in parenthesis.

	n =400			n =1000			n =2000		
$(\beta_1, \beta_2, \beta_3, \beta_4)$	YW	CML	CL	YW	CML	CL	YW	CML	CL
$(1.6, 0.9, 1.8, 1.2)$									
$\hat{\beta}_1$	1.085	1.607	1.609	1.201	1.607	1.614	1.175	1.599	1.611
	(0.0128)	(0.0646)	(0.2239)	(0.0984)	(0.0085)	(0.02105)	(0.0001)	(0.0026)	(0.0007)
$\hat{\beta}_2$	1.481	0.915	0.902	1.529	0.903	0.903	1.554	0.895	0.897
	(0.3959)	(0.0055)	(0.1054)	(0.4486)	(0.0106)	(0.0137)	(0.1550)	(0.0007)	(1.2×10^{-7})
$\hat{\beta}_3$	2.668	1.844	1.814	2.826	1.839	1.832	2.880	1.793	1.798
	(0.1454)	(0.9795)	(0.5293)	(0.3356)	(0.5068)	(0.6481)	(0.0224)	(0.0025)	(0.0031)
$\hat{\beta}_4$	1.045	1.227	1.231	1.139	1.196	1.205	1.128	1.203	1.202
	(0.9726)	(0.0607)	(0.1194)	(0.5961)	(0.0083)	(0.0051)	(0.3896)	(0.0085)	(0.0115)

Specifically, Table 2.1 reports the estimates for autocorrelation parameters $\boldsymbol{\alpha}_j$ $(j = 1, 2, 3)$, where small MSE's characterize all estimates of $(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3)$. The performance of the estimators $\hat{\boldsymbol{\lambda}}_j$ $(j = 1, 2, 3)$ in Table 2.2 and estimator $\hat{\boldsymbol{\beta}}$ in Table 2.3, is slightly worse. The YW estimator does not perform well for all the parameters involved in the model, revealing to be a not so good estimator for the dispersion parameter $\boldsymbol{\beta}$. 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 $(n = 1000$ and $n = 2000)$, both estimators seem to perform well and in a similar way.

Graphical inspection is given through the boxplots of the biases of the produced estimates. Figures 2.1-2.3 display boxplots of the biases of the estimates for $\boldsymbol{\alpha}_j = (\alpha_{j,1}, \alpha_{j,2}, \alpha_{j,3}, \alpha_{j,4})$, with $j = 1, 2, 3$. Figures 2.4-2.7 refer to the boxplots of the biases of the estimates for the parameters regarding the trivariate negative binomial distributed innovation process $\boldsymbol{\lambda}_j = (\lambda_{j,1}, \lambda_{j,2}, \lambda_{j,3}, \lambda_{j,4})$, $j = 1, 2, 3$, and $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)$, respectively. The effect of sample size on the behavior of the estimators can be seen in Figures 2.1-2.7. 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.

Closing this section, it is worth mentioning that numerical maximization of the conditional maximum likelihood was very time consuming. The composite likelihood method was suggested in order to overcome the computational difficulties of the conditional maximum likelihood approach in multivariate models. The CL method requires significantly less time for the optimization of the likelihood function without obvious losses in precision.

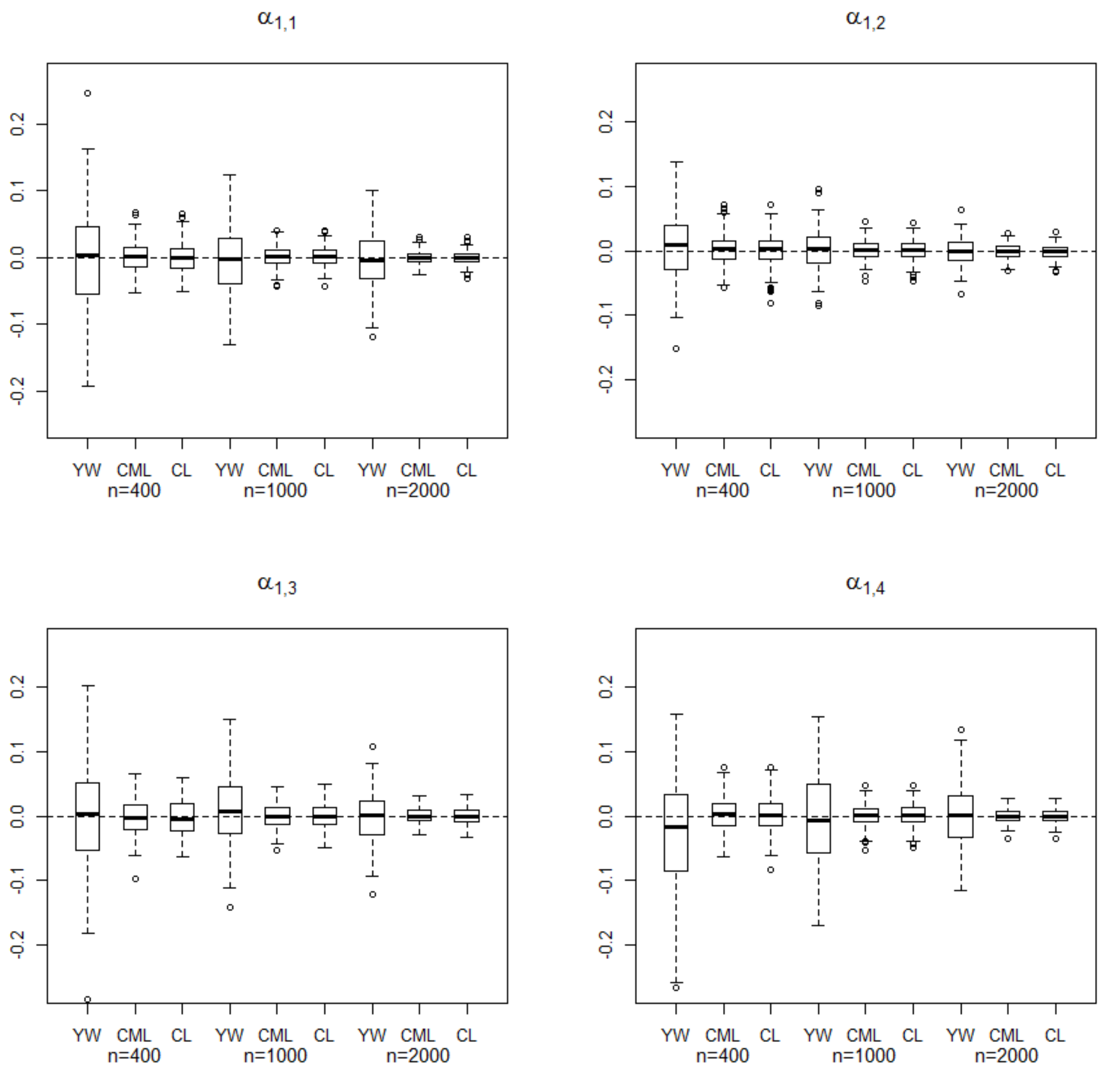

Figure 2.1: Boxplots for the biases of the YW, CML and CL estimates of the parameter $\boldsymbol{\alpha}_1 = (\alpha_{1,1}, \alpha_{1,2}, \alpha_{1,3}, \alpha_{1,4})$. From left to right, the first three boxplots display the biases of $\hat{\alpha}_{1,1}$ for the three methods with $n = 400, 1000, 2000$. The same information follows for $\hat{\alpha}_{1,2}$, $\hat{\alpha}_{1,3}$ and $\hat{\alpha}_{1,4}$, respectively.

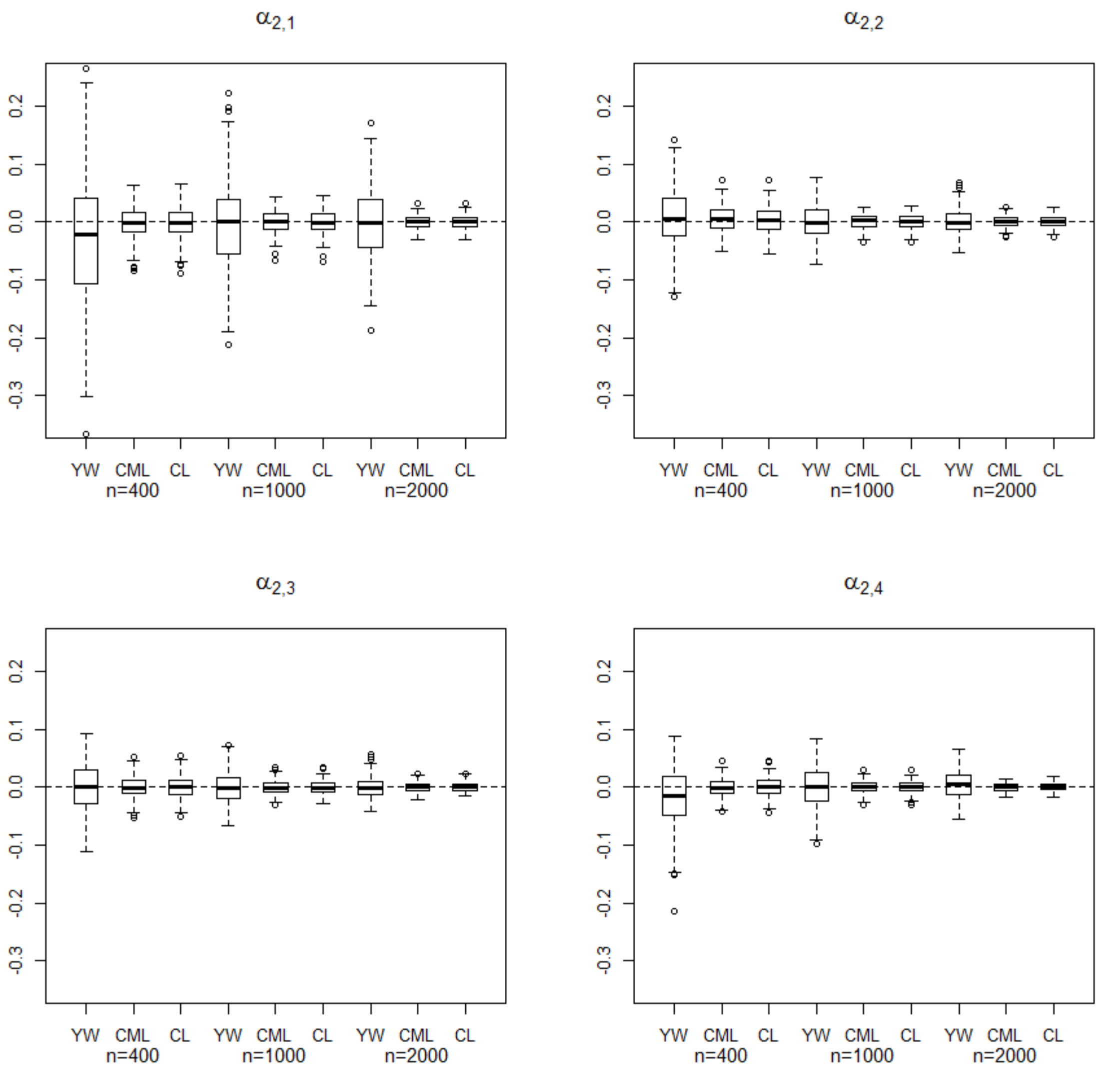

Figure 2.2: Boxplots for the biases of the YW, CML and CL estimates of the parameter $\boldsymbol{\alpha}_2 = (\alpha_{2,1}, \alpha_{2,2}, \alpha_{2,3}, \alpha_{2,4})$. From left to right, the first three boxplots display the biases of $\hat{\alpha}_{2,1}$ for the three methods with $n = 400, 1000, 2000$. The same information follows for $\hat{\alpha}_{2,2}$, $\hat{\alpha}_{2,3}$ and $\hat{\alpha}_{2,4}$, respectively.

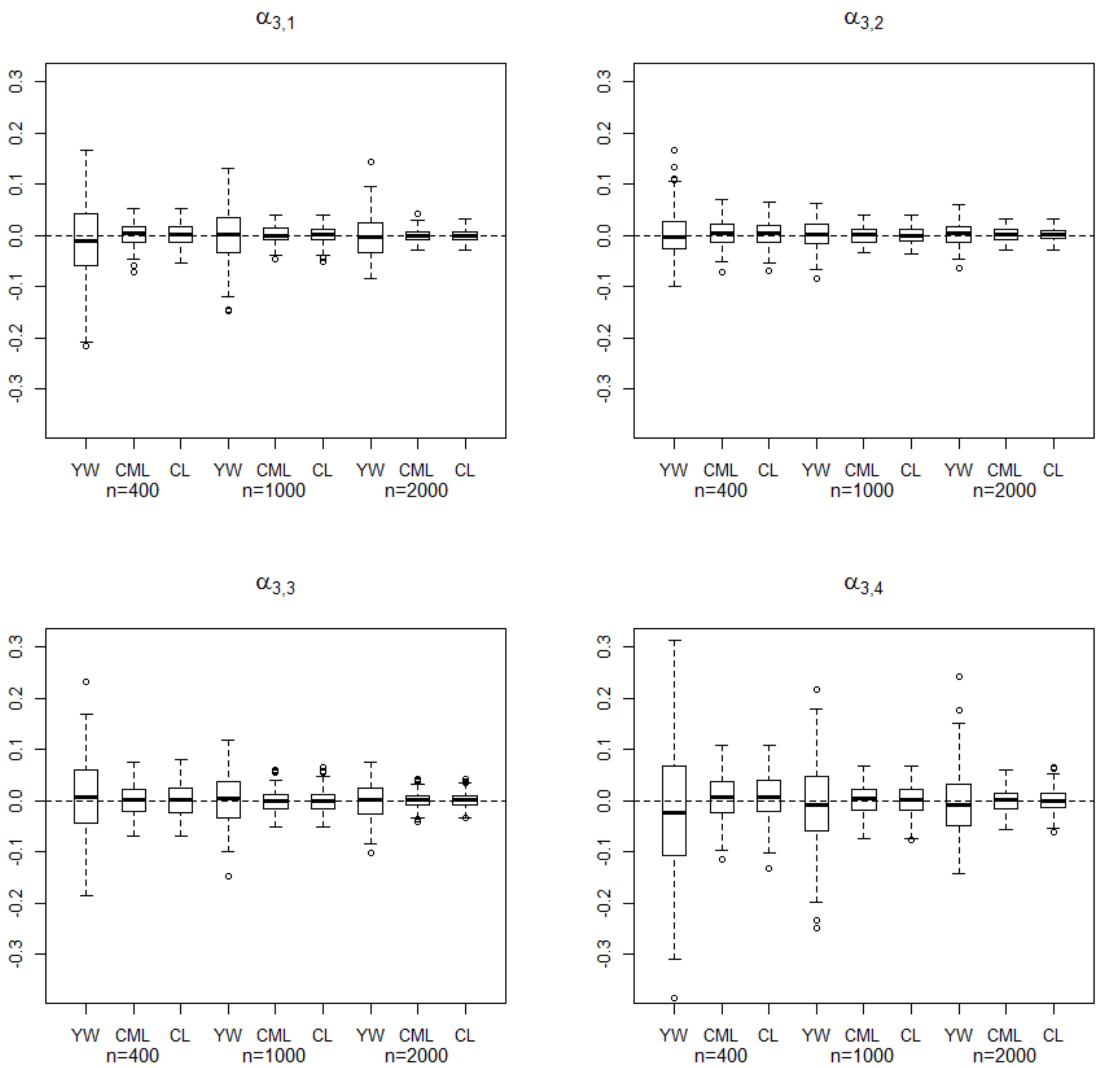

Figure 2.3: Boxplots for the biases of the YW, CML and CL estimates of the parameter $\boldsymbol{\alpha}_3 = (\alpha_{3,1}, \alpha_{3,2}, \alpha_{3,3}, \alpha_{3,4})$. From left to right, the first three boxplots display the biases of $\hat{\alpha}_{3,1}$ for the three methods with $n = 400, 1000, 2000$. The same information follows for $\hat{\alpha}_{3,2}$, $\hat{\alpha}_{3,3}$ and $\hat{\alpha}_{3,4}$, respectively.

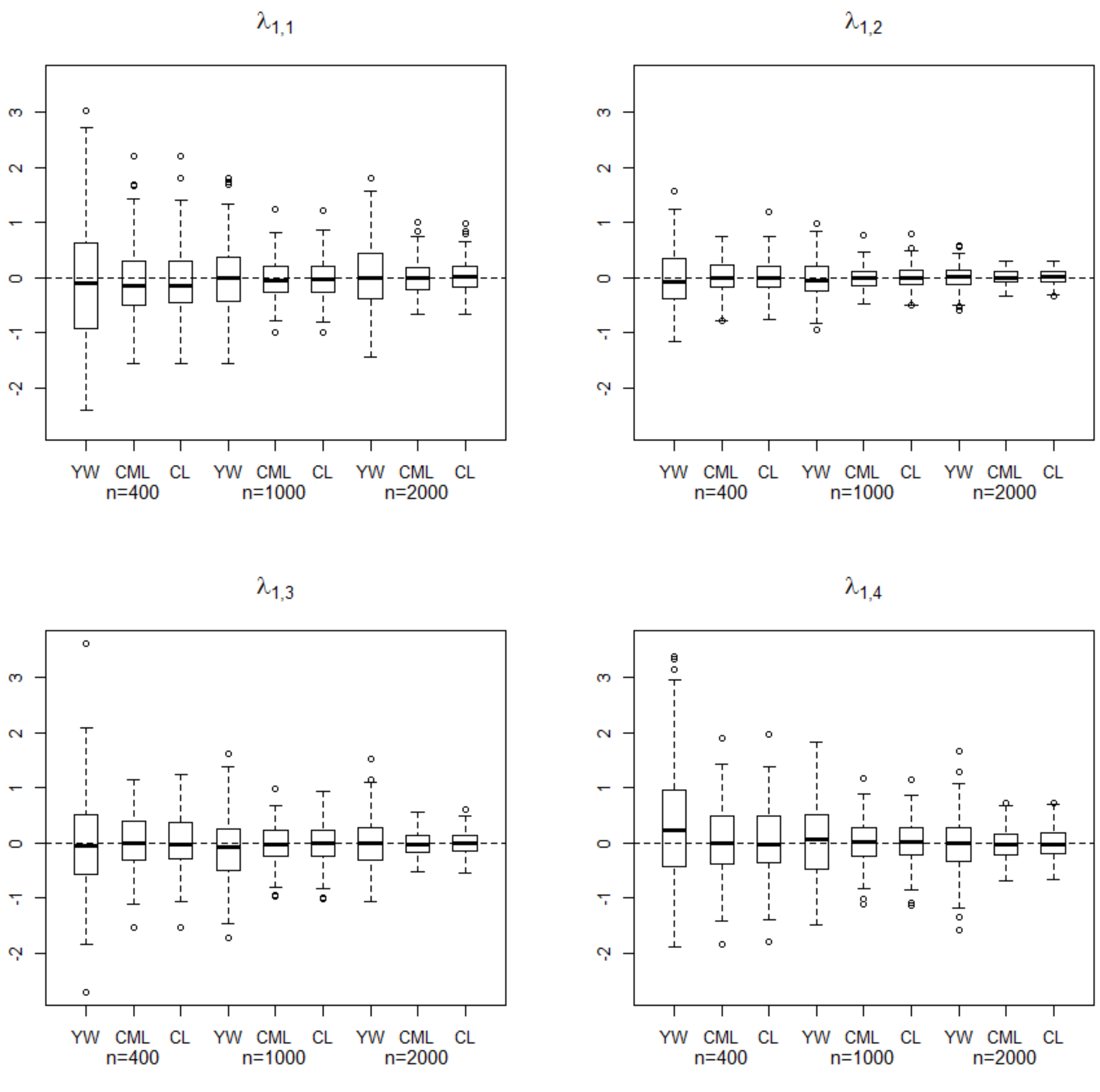

Figure 2.4: Boxplots for the biases of the YW, CML and CL estimates of the parameter $\boldsymbol{\lambda}_1 = (\lambda_{1,1}, \lambda_{1,2}, \lambda_{1,3}, \lambda_{1,4})$. From left to right, the first three boxplots display the biases of $\hat{\lambda}_{1,1}$ for the three methods with $n = 400, 1000, 2000$. The same information follows for $\hat{\lambda}_{1,2}$, $\hat{\lambda}_{1,3}$ and $\hat{\lambda}_{1,4}$, respectively.

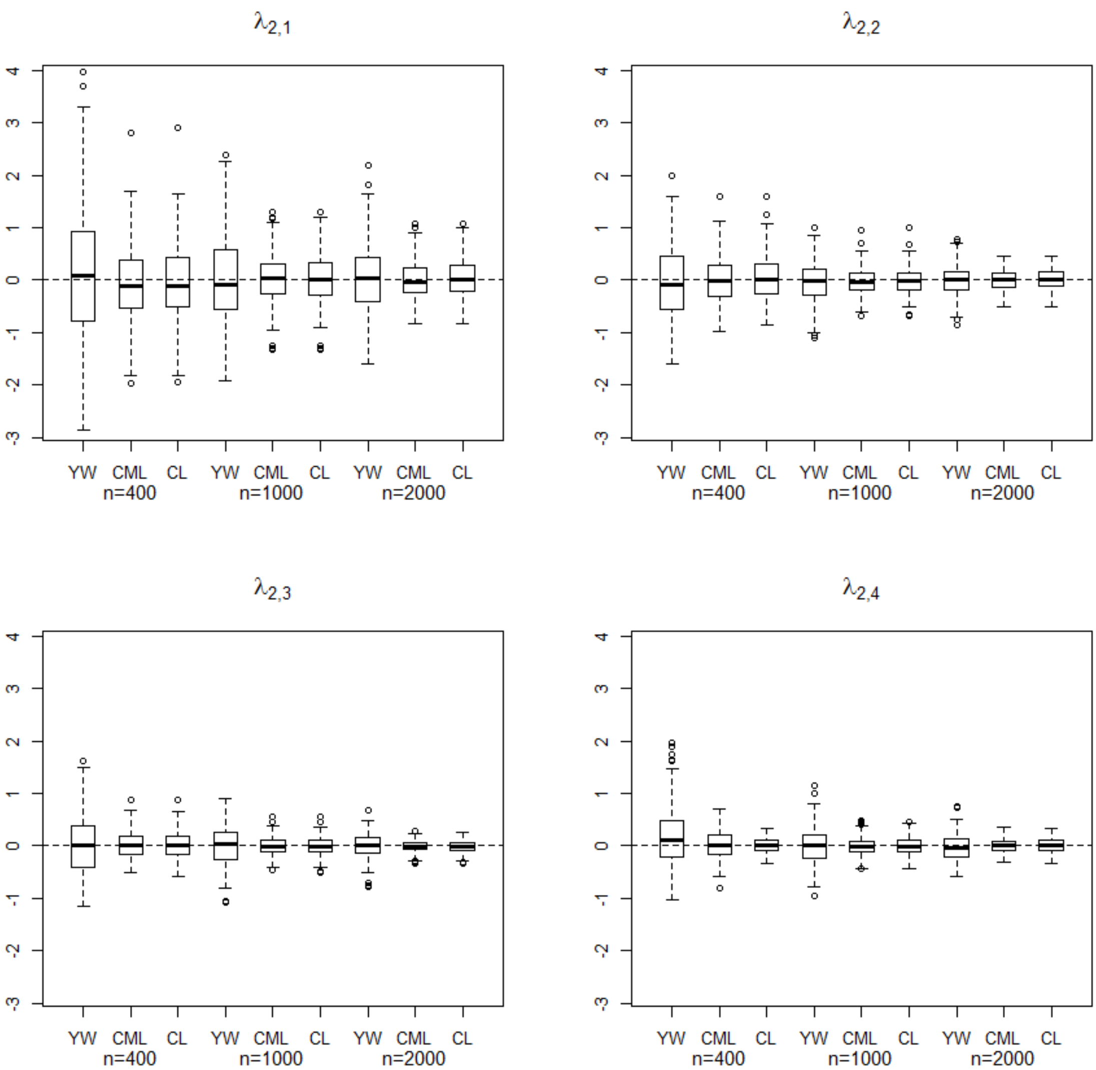

Figure 2.5: Boxplots for the biases of the YW, CML and CL estimates of the parameter $\boldsymbol{\lambda}_2 = (\lambda_{2,1}, \lambda_{2,2}, \lambda_{2,3}, \lambda_{2,4})$. From left to right, the first three boxplots display the biases of $\hat{\lambda}_{2,1}$ for the three methods with $n = 400, 1000, 2000$. The same information follows for $\hat{\lambda}_{2,2}$, $\hat{\lambda}_{2,3}$ and $\hat{\lambda}_{2,4}$, respectively.

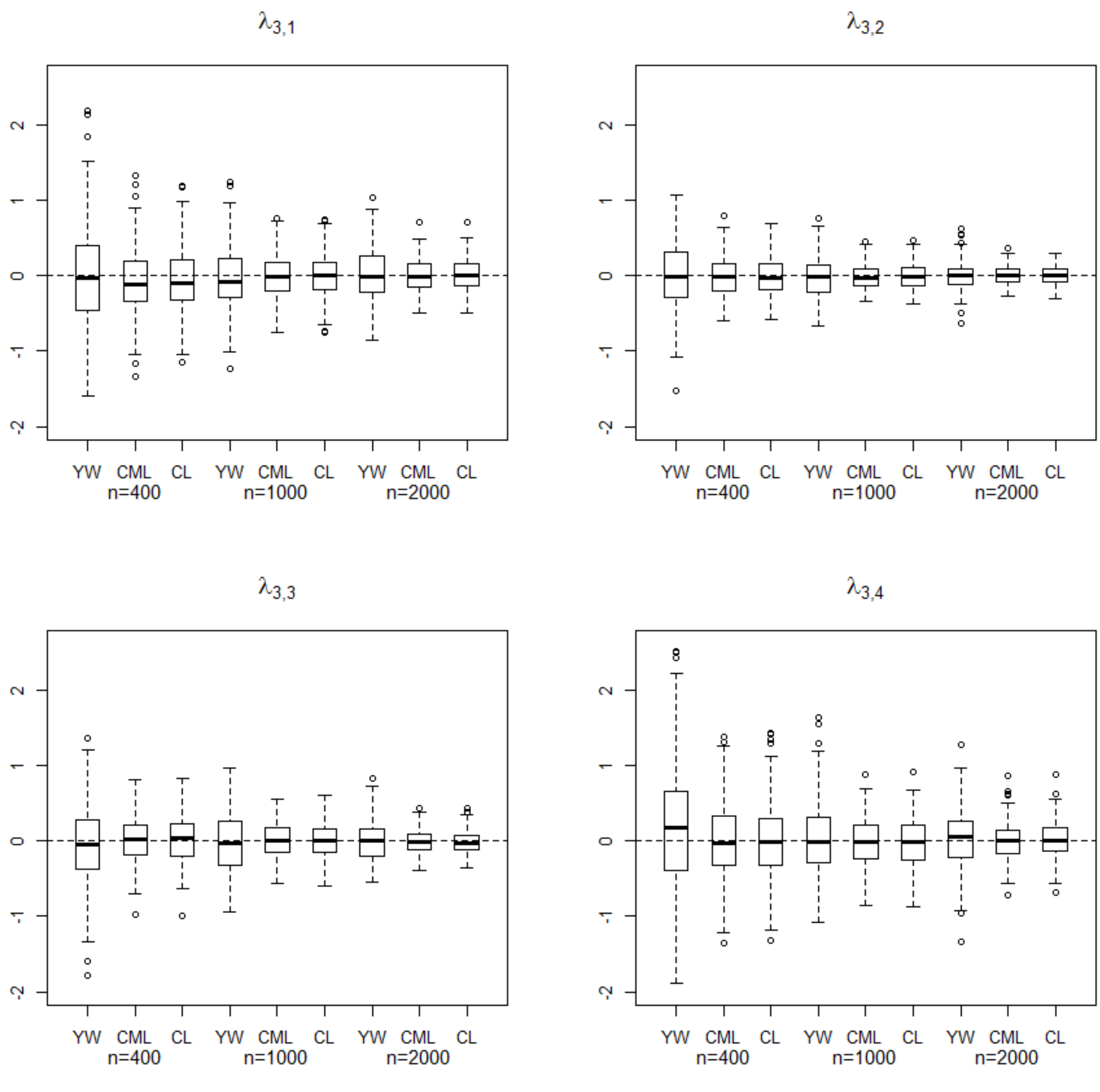

Figure 2.6: Boxplots for the biases of the YW, CML and CL estimates of the parameter $\boldsymbol{\lambda}_3 = (\lambda_{3,1}, \lambda_{3,2}, \lambda_{3,3}, \lambda_{3,4})$. From left to right, the first three boxplots display the biases of $\hat{\lambda}_{3,1}$ for the three methods with $n = 400, 1000, 2000$. The same information follows for $\hat{\lambda}_{3,2}$, $\hat{\lambda}_{3,3}$ and $\hat{\lambda}_{3,4}$, respectively.

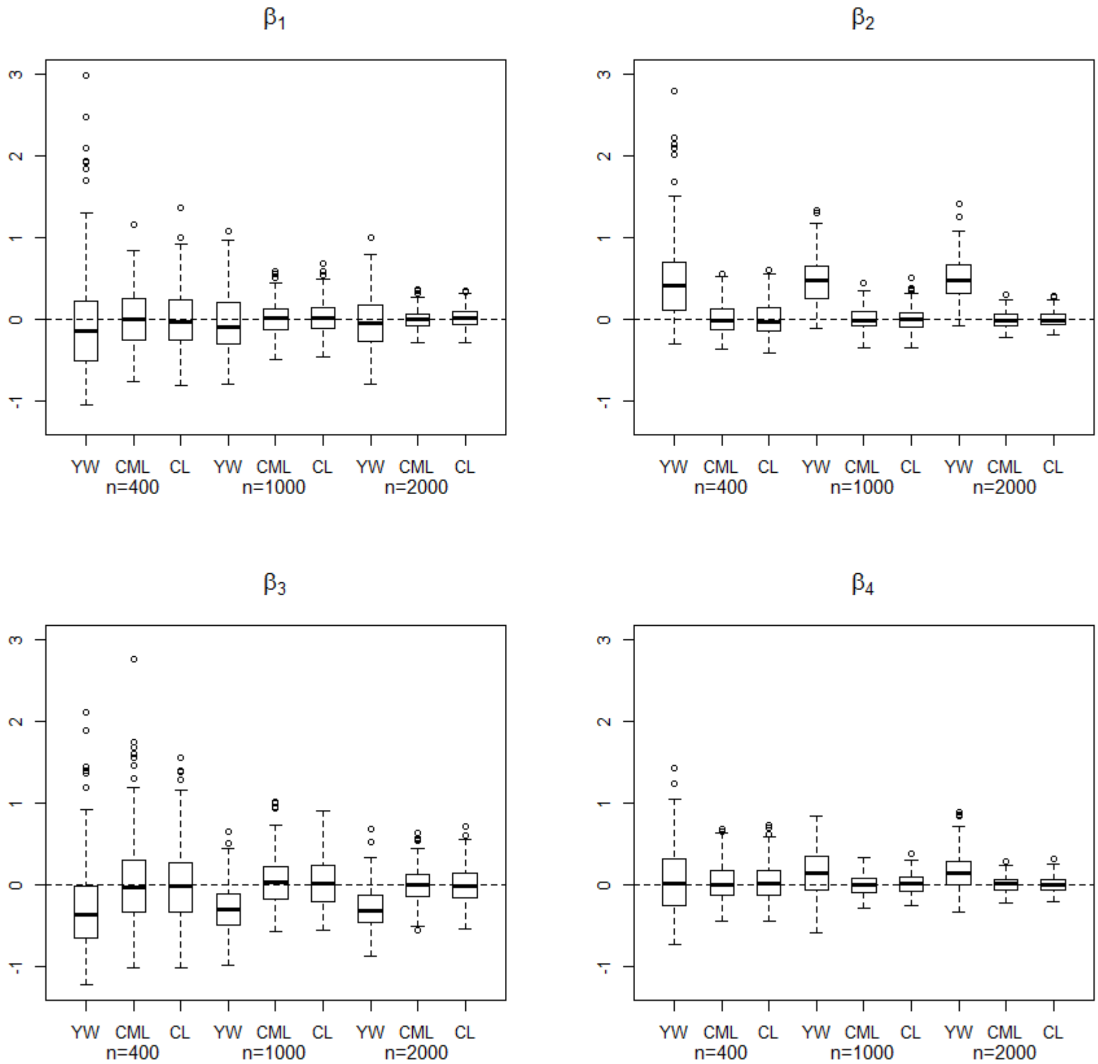

Figure 2.7: Boxplots for the biases of the YW, CML and CL estimates of the parameter $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)$. From left to right, the first three boxplots display the biases of $\hat{\beta}_1$ for the three methods with $n = 400, 1000, 2000$. The same information follows for $\hat{\beta}_2$, $\hat{\beta}_3$ and $\hat{\beta}_4$, respectively.

2.7 Application

This section illustrates the PMINAR(1) model with a trivariate real environmental data set. The data refers to the number of fires collected in three counties in Portugal, namely Aveiro, Coimbra and Faro, during 30 years, from 1981 to 2010. The data are monthly observations based on the mean of daily fires in those counties. This collection of fires can be seen in Figure 2.8. The number of fires in Faro is higher than in the other two counties. Creating appropriate time series models for handling multiple time series together is of great interest. In fact, forest fires is a major problem in many countries, as they are a threat not only to forests but also to people and their surroundings. In Europe, Portugal is the country with

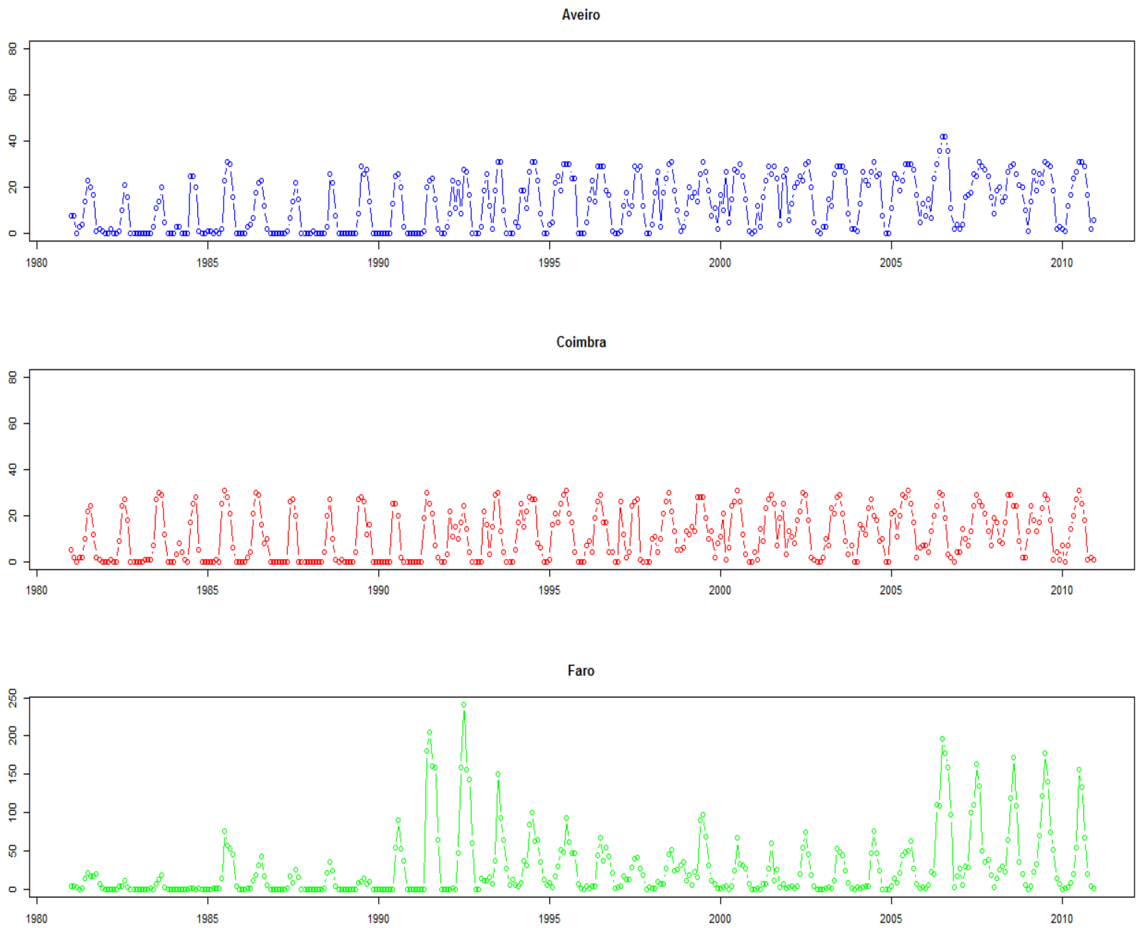

Figure 2.8: Number of monthly fires in Aveiro, Coimbra and Faro counties in Portugal.

the highest number of forest fires per unit surface and per number of inhabitants (San-Miguel and Camia (2009)). Fire frequency is markedly different from north to south and from east to west (Nunes, 2012). The distribution of fires across the year follows a regular pattern, strongly influenced by seasonal variations of temperature and rainfall. Hence, it is expected to find the highest number of fires in the summer season, with a peak in July/August and the lowest number of fires in the rainy season. The sample autocorrelation function (ACF) in Figure 2.9 reveals a periodic pattern of 12 months.

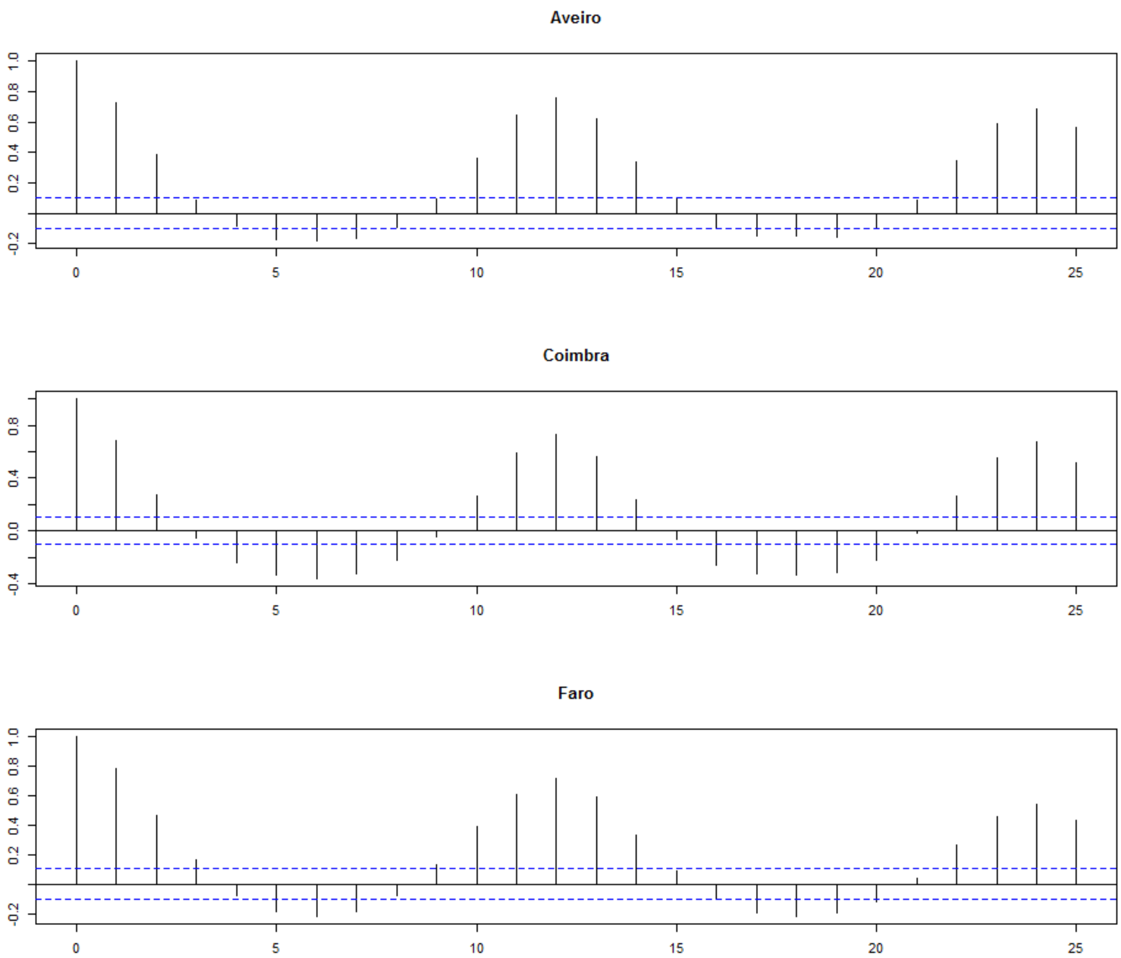

Figure 2.9: Sample ACF for the number of monthly fires in Aveiro, Coimbra and Faro counties in Portugal.

The mean values and standard deviation (sd) of the number of fires per month are shown in Figure 2.10 and cross-correlations in Figure 2.11. In the three counties, most months have variance greater than the mean, implying overdispersion. The innovation series plays an important role in the specification of the periodic trivariate INAR(1) process being responsible for both the introduction of dependence and the determination of the joint distribution of the three series. The distribution for the innovations was assumed to be trivariate negative

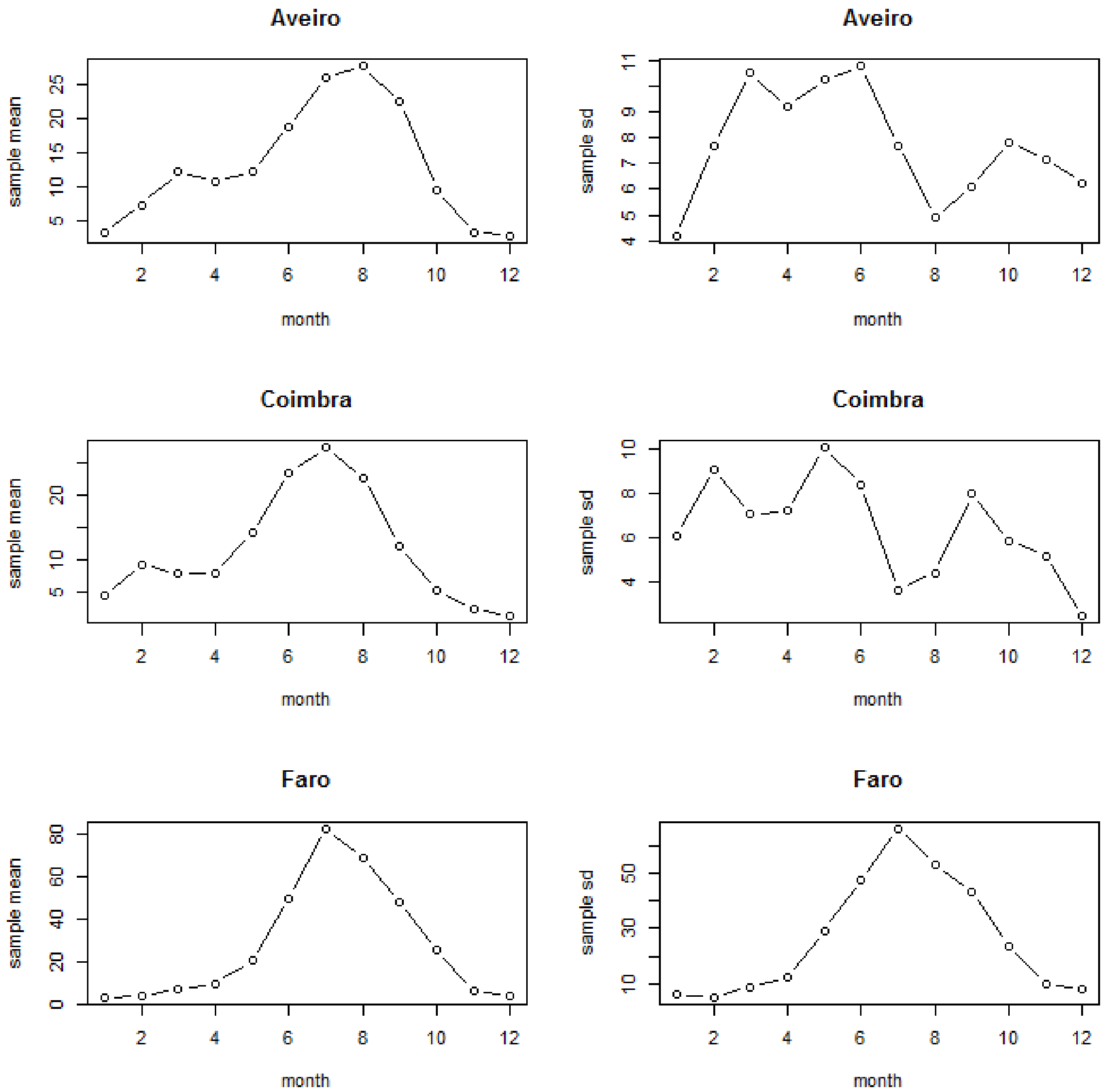

Figure 2.10: Sample mean and standard deviation for the number of monthly fires in the Aveiro, Coimbra and Faro counties in Portugal.

binomial, which only allows for non-negative correlation as established in (2.80). However, Figure 2.11 displays a slight negative cross-correlation (-0.18) in August between Aveiro and Coimbra. The counties of Coimbra and Faro also have a small negative cross-correlation in August (-0.25), September (-0.17) and October (-0.22). The significance of these correlations was tested and the null hypothesis was not rejected for the usual significance levels. The periodic trivariate `INAR`(1) model with period $s = 12$ and trivariate negative binomial innovations is appropriate for series displaying overdispersion. For this particular application, the Yule-Walker estimates are non-admissible for some months, hence are not presented. We were aware this could happen. Table 2.4 summarizes the CML and CL estimates and the corresponding standard errors (SE) obtained by fitting the periodic trivariate `INAR`(1) model with period $s = 12$ and trivariate negative binomial innovations. The SE were calculated numerically from the Hessian matrix during the optimization procedure in `R`. For some months, the estimates from both methods (CML and CL) are very close, however this does not always happen. Some loss of efficiency is noticed when the CL method is employed but we have to remember that the CL is an approximate likelihood, leading to inevitable losses. The CL method could be regarded as a satisfactory approach for the estimation of the unknown parameters of the `PMINAR`(1) process, especially when other alternatives are not available. The CL estimates could also be used to initialize the CML method. Some estimates of the autocorrelation parameters in Table 2.4 are not significant, namely for the months of February, March and November, suggesting that in those months the number of fires is being mainly modeled through the innovation process.

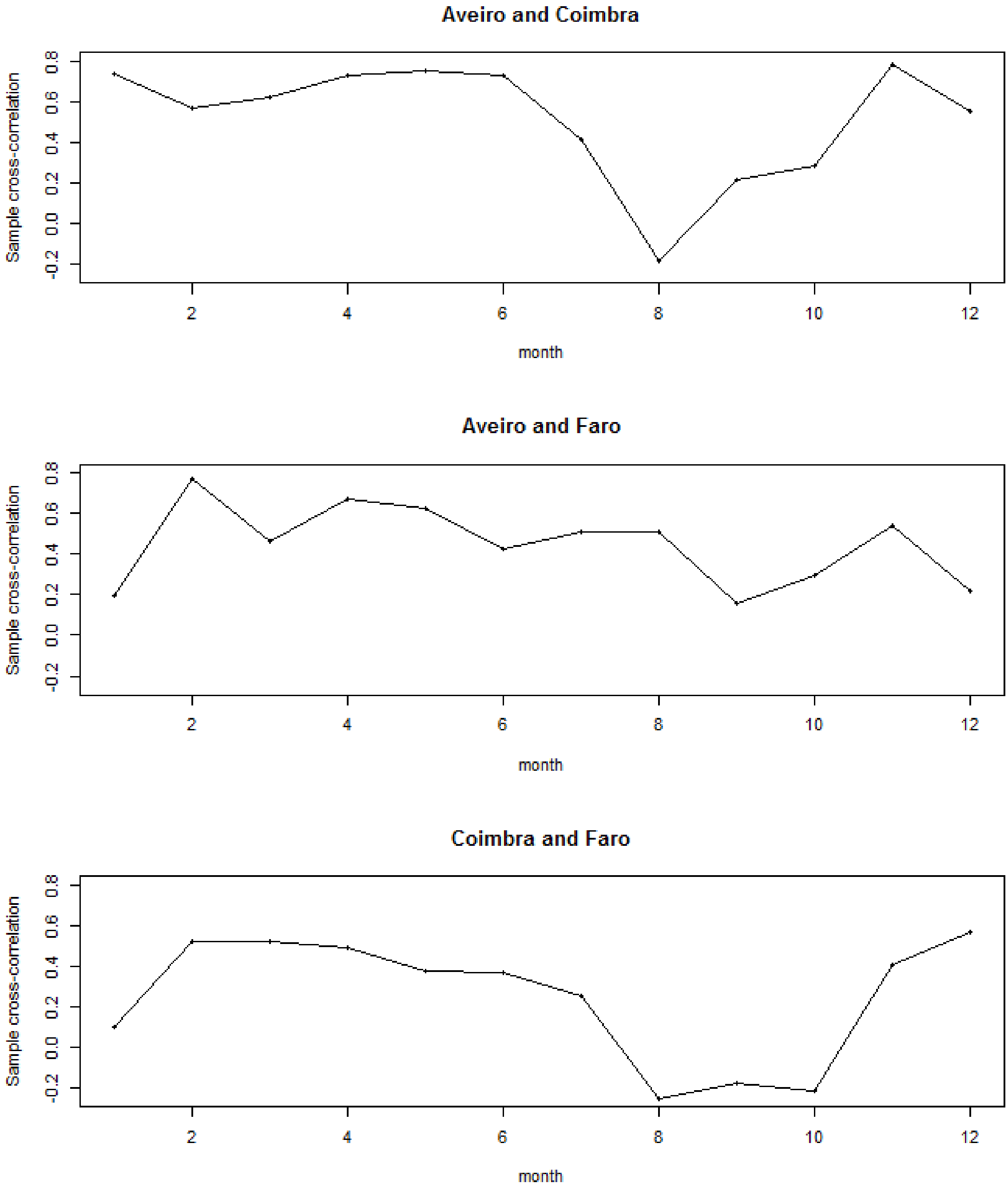

Figure 2.11: Sample cross-correlations for the number of monthly fires in the Aveiro, Coimbra and Faro counties in Portugal.

Table 2.4: CML and CL estimates from fitting the periodic trivariate INAR(1) model with trivariate negative binomial innovations. Standard errors in parenthesis.

	Conditonal maximum likelihood (CML)							Composite likelihood (CL)						
	Aveiro		Coimbra		Faro			Aveiro		Coimbra		Faro		
	$\boldsymbol{\alpha}_1$	$\boldsymbol{\lambda}_1$	$\boldsymbol{\alpha}_2$	$\boldsymbol{\lambda}_2$	$\boldsymbol{\alpha}_3$	$\boldsymbol{\lambda}_3$	$\boldsymbol{\beta}$	$\boldsymbol{\alpha}_1$	$\boldsymbol{\lambda}_1$	$\boldsymbol{\alpha}_2$	$\boldsymbol{\lambda}_2$	$\boldsymbol{\alpha}_3$	$\boldsymbol{\lambda}_3$	$\boldsymbol{\beta}$
January	0.0313	3.0149	0.3408	4.1279	0.1502	2.4818	2.5052	0.1861	2.6684	0.6528	3.8223	0.1065	2.7870	2.4862
	(0.0987)	(0.9819)	(0.2676)	(1.3166)	(0.0747)	(0.8337)	(0.7762)	(0.0610)	(0.5379)	(0.1516)	(0.7472)	(0.0629)	(0.5944)	(0.4834)
February	0.4226	6.0551	4.48×10^{-06}	9.2665	0.2298	3.0056	2.1378	0.4866	5.9225	4.49×10^{-06}	9.4738	0.2549	2.9024	2.1950
	(0.1467)	(1.7365)	(0.1598)	(2.6106)	(0.0936)	(0.9017)	(0.6091)	(0.0998)	(1.0455)	(0.0764)	(1.5841)	(0.0627)	(0.5427)	(0.3848)
March	0.1903	10.7851	0.0002	7.8996	0.2748	5.8071	1.7431	0.3822	9.7927	0.0001	8.1526	0.2233	6.1628	1.6039
	(0.0666)	(1.9262)	(0.0605)	(1.9016)	(0.0394)	(4.6350)	(0.0738)	(0.0823)	(1.5034)	(0.0764)	(1.2357)	(0.0923)	(0.9605)	(0.2695)
April	0.0640	10.0523	0.1181	7.1002	0.6095	5.7017	1.9305	0.2057	8.4738	0.1689	6.8685	0.6720	5.5386	2.0476
	(0.0810)	(2.7908)	(0.0866)	(1.9802)	(0.0761)	(1.5806)	(0.5550)	(0.0531)	(1.5144)	(0.0647)	(1.2316)	(0.0512)	(0.9673)	(0.3673)
May	0.3971	7.9989	0.3601	11.2745	0.3379	17.1677	1.6314	0.4920	7.1231	0.3962	11.2488	0.3836	17.3825	1.7389
	(0.0805)	(2.1025)	(0.1091)	(2.8277)	(0.0957)	(4.1738)	(0.4200)	(0.0489)	(1.1816)	(0.0832)	(1.7882)	(0.0749)	(2.6518)	(0.2761)
June	0.6404	11.0332	0.3038	19.0123	0.6179	37.2652	0.5579	0.6682	10.6394	0.3994	17.9422	0.6269	37.8833	0.5826
	(0.0672)	(1.7962)	(0.1027)	(3.0644)	(0.0675)	(5.3732)	(0.1474)	(0.0465)	(1.1180)	(0.0795)	(1.9484)	(0.0582)	(3.4789)	(0.0919)
July	0.6279	14.0669	0.6355	12.3922	0.8027	42.5129	0.2732	0.7129	12.4433	0.7159	10.3999	0.7947	43.9054	0.3377
	(0.0666)	(1.9262)	(0.0605)	(1.9016)	(0.0394)	(4.6350)	(0.0738)	(0.0453)	(1.2294)	(0.0401)	(1.1979)	(0.0344)	(3.4417)	(0.0557)
August	0.8420	5.8303	0.6361	5.2994	0.5572	22.4300	0.4406	0.8720	5.2137	0.6372	5.1134	0.5210	26.1427	0.5385
	(0.0341)	(1.1656)	(0.0371)	(1.1825)	(0.0278)	(3.5681)	(0.1377)	(0.0489)	(0.0832)	(2.6518)	(0.2761)	(1.1816)	(1.7882)	(0.0749)
September	0.7054	2.9515	0.4135	2.7881	0.3676	22.8409	0.8308	0.7024	2.9245	0.3240	4.6880	0.3795	24.0514	1.1132
	(0.0290)	(0.8889)	(0.0324)	(0.8136)	(0.0424)	(4.8063)	(0.2584)	(0.0202)	(0.6069)	(0.0288)	(0.8400)	(0.0291)	(3.6009)	(0.2401)
October	0.0627	8.1248	0.1199	3.8122	0.3735	7.3728	1.0098	0.1848	5.5172	0.1266	3.8764	0.3625	7.6962	1.3280
	(0.0816)	(2.4112)	(0.0433)	(0.9217)	(0.0226)	(1.6954)	(0.3349)	(0.0314)	(1.0664)	(0.0316)	(0.7093)	(0.0176)	(1.4996)	(0.3525)
November	0.0031	3.3325	1.82×10^{-06}	2.3665	0.0598	5.1057	4.6627	0.0039	3.3398	2.54×10^{-07}	2.4160	0.0165	5.1300	4.4599
	(0.0271)	(1.3696)	(0.0952)	(1.0082)	(0.0206)	(2.1023)	(2.0003)	(0.0095)	(0.7971)	(0.0029)	(0.5784)	(0.0031)	(1.5148)	(0.8463)
December	0.5496	1.1012	0.0262	1.2712	0.1625	2.9611	5.7266	0.5275	1.2056	0.1065	1.0983	0.1738	3.0610	6.6444
	(0.0706)	(0.5441)	(0.0856)	(0.6244)	(0.0577)	(1.3738)	(2.0903)	(0.0537)	(0.3905)	(0.0515)	(0.3518)	(0.0462)	(0.9420)	(1.6083)

Chapter 3

Periodic INAR(1) models based on the signed thinning operator

The class of `INAR` models, based on the binomial thinning operator introduced by Steutel and van Harn (1979), only applies to non-negative integer-valued time series. The binomial thinning operator defined in (1.2) has been generalized in a number of different ways. Kim and Park (2008) introduced the signed binomial thinning operator given in (1.9), allowing time series with negative values, the so-called $\mathbb{Z}$-valued time series. Kachour and Truquet (2011) established a slightly different signed thinning operator in (1.12) also allowing for negative values both for the series and its autocorrelation function. Recently, Bulla et al. (2016) proposed an extension of the preceding signed thinning operator to the bivariate case defined in (1.16).

In this chapter, we introduce two new first-order integer-valued autoregressive models with time-varying parameters and sequences of innovations with periodic structure. Both models are based on the signed thinning operator defined in the univariate case by Kachour and Truquet (2011) in (1.12) and in the bivariate case by Bulla et al. (2016) in (1.16) adapted to the periodic case, accordingly. Basic notations and definitions concerning the periodic signed thinning operator are established as well as some of its properties. Emphasis will be placed on models with innovations following Skellam distribution (Skellam, 1946) and bivariate Skellam distribution (Bulla et al., 2015), respectively. Therefore, a brief description of the periodic Skellam distribution for both univariate and bivariate distributions defined on the whole set of integers is also provided.

In extending the model proposed by Chesneau and Kachour (2012) to the periodic case, we introduce a univariate signed periodic `INAR`(1) process (`S-PINAR`(1) for short) with period s, by considering a parametric assumption on the common distribution of the periodic counting sequence of the model. The properties of the `S-PINAR`(1) model with period s are discussed. We focus on a specific parametric case which arises under the assumption of periodic Skellam-distributed innovation. Regarding parameter estimation, two methods are considered: conditional least squares and conditional maximum likelihood. The performance of the proposed estimation methods for the `S-PINAR`(1) model is accomplished through a simulation study.

Within the bivariate setting, the work of Bulla et al. (2016) has motivated a new periodic bivariate model. The generalization of the previous signed model with period s to the bivariate case is denoted by `BS-PINAR`(1). Several statistical properties of this periodic model are derived. The assumption of a diagonal autoregressive matrix is made, thus the correlation is achieved through their innovation processes, where the distribution of the innovation processes is set a priori which consequently determines the distribution of the underlying time series. Hence, the discrete bivariate distribution on $\mathbb{Z}^2$ assigned to the distribution of the innovations is the periodic bivariate Skellam distribution. Parameter estimation of the unknown parameters of the `BS-PINAR`(1) model with period s is provided through conditional maximum likelihood method.

3.1 The periodic signed thinning operator

Basic notations and definitions concerning the periodic signed thinning operator for both univariate and bivariate cases are established.

3.1.1 Univariate case

The definition of the signed thinning operator introduced by Kachour and Truquet (2011) is given in (1.12) and its properties in Lemma 1.2. In the periodic case, the signed thinning operator is defined by

$$F_t \odot X := \begin{cases} sign(X) \sum_{i=1}^{|X|} U_{i,t}(\phi_t) & , X \neq 0 \\ 0 & , otherwise \end{cases} \tag{3.1}$$

with $sign(X)$ as in (1.10) and where F_t represents the common distribution of the periodic sequence of i.i.d. counting sequences $(U_{i,t}(\phi_t))_{i\in\mathbb{N}}$. All counting sequences associated to the operator $F_t\odot$ are mutually independent.

We consider that F_t, the distribution of the periodic sequence of i.i.d. random variables $(U_{i,t}(\phi_t))_{i\in\mathbb{N}}$, has probability mass function given by

$$P(U_{1,t}(\phi_t) = a) = \begin{cases} (1-\phi_t)^2, & a = -1 \\ 2\phi_t(1-\phi_t), & a = 0 \\ \phi_t^2, & a = 1 \end{cases}, \tag{3.2}$$

with $\phi_t = \alpha_v \in (0,1)$ for $t = v+ns;\ v = 1,\ldots,s$ and $n \in \mathbb{N}_0$. Without the periodic structure, Chesneau and Kachour (2012) have also made use of this common distribution. Note that, for a fixed v, the random variable

$$U_{i,t}(\phi_t) = U_t(\phi_t) \stackrel{d}{=} R_t(\phi_t) - 1, \qquad R_t(\phi_t) \sim Bin(2,\phi_t) \tag{3.3}$$

and

$$P\left(\sum_{i=1}^{k} U_{i,t}(\phi_t) = l\right) = P\left(R_t^{(k)}(\phi_t) = k + l\right), \qquad l \in \{-k,\ldots,k\},$$

where

$$R_t^{(k)}(\phi_t) = \sum_{i=1}^{k} R_{i,t}(\phi_t), \qquad R_t^{(k)}(\phi_t) \sim Bin(2k,\phi_t),\ k \in \mathbb{N}. \tag{3.4}$$

Then, for $x \in \mathbb{Z}\backslash\{0\}$ and $y \in \mathbb{Z}$, the conditional probability function of the periodic signed thinning operator $F_t \odot$ defined in (3.1) is

$$P\left(F_t \odot X = y | X = x\right) = P\left(sign(x)\sum_{i=1}^{|x|} U_{i,t}(\phi_t) = y\right) =$$

$$= P\left(R_t^{|x|}(\phi_t) - |x| = sign(x)\cdot y\right) = P\left(R_t^{|x|}(\phi_t) = |x| + sign(x)\cdot y\right) =$$

$$= C_{|x|+sign(x)\cdot y}^{2|x|}\alpha_v^{|x|+sign(x)\cdot y}(1-\alpha_v)^{|x|-sign(x)\cdot y}, \qquad y \in \{-|x|,\ldots,|x|\} \tag{3.5}$$

with mean value

$$E[F_t \odot X|X] = (2\alpha_v - 1)X \tag{3.6}$$

and variance

$$Var[F_t \odot X|X] = 2\alpha_v(1-\alpha_v)|X| \tag{3.7}$$

for $t = v + ns$, $v = 1, \ldots, s$ and $n \in \mathbb{N}_0$.

Probability generating function

For sake of simplicity, let $U_{i,t}(\phi_t) = U_t(\phi_t) = U_t$ and $R_{i,t}(\phi_t) = R_t(\phi_t) = R_t$ then $U_t \stackrel{d}{=} R_t - 1$ where $R_t \sim Bin(2, \phi_t)$ for $\phi_t = \alpha_v$, $t = v + ns$; $v = 1, \ldots, s$ and $n \in \mathbb{N}_0$. We denote by $G_{R_t}(r)$ the probability generating function (p.g.f.) of the well known Binomial-distributed random variable

$$G_{R_t}(r) = (1 - \alpha_v + \alpha_v r)^2,$$

thus

$$G_{U_t}(r) = E\left[r^{U_t}\right] = E\left[r^{R_t - 1}\right] = \frac{1}{r}G_{R_t}(r) = \frac{(1-\alpha_v+\alpha_v r)^2}{r}. \tag{3.8}$$

Let $W_1 = F_t \odot X$. From (3.1), the p.g.f. takes the form

$$\begin{aligned} G_{W_1}(r) &= E\left[r^{W_1}\right] = E\left[r^{F_t \odot X}\right] = E\left[E\left[r^{F_t \odot X}|X\right]\right] = \\ &= E\left[r^{\sum\limits_{i=1}^{X} U_{i,t}} \mathbb{1}_{\{X>0\}}\right] + E\left[r^{-\sum\limits_{i=1}^{-X} U_{i,t}} \mathbb{1}_{\{X<0\}}\right] = \\ &= E\left[\mathbb{1}_{\{X>0\}}(G_{U_t}(r))^X\right] + E\left[\mathbb{1}_{\{X<0\}}(G_{-U_t}(r))^{-X}\right] = \\ &= G_X(G_{U_t}(r))\mathbb{1}_{\{X>0\}} + G_{-X}(G_{-U_t}(r))\mathbb{1}_{\{X<0\}} = \\ &= G_X(G_{U_t}(r))\mathbb{1}_{\{X>0\}} + G_X\left(G_{U_t}^{-1}\left(r^{-1}\right)\right)\mathbb{1}_{\{X<0\}} \end{aligned} \tag{3.9}$$

with $G_{U_t}(r)$ defined above in equation (3.8).

Let $W_2 = F_t \odot W_1 = F_t \odot (F_t \odot X)$, the probability generating function is

$$\begin{aligned} G_{W_2}(r) &= E\left[r^{W_2}\right] = E\left[r^{F_t \odot W_1}\right] = E\left[E\left[r^{F_t \odot W_1}|W_1\right]\right] = \\ &= G_{W_1}(G_{U_t}(r))\mathbb{1}_{\{W_1>0\}} + G_{W_1}\left(G_{U_t}^{-1}\left(r^{-1}\right)\right)\mathbb{1}_{\{W_1<0\}} \end{aligned}$$

and from equation (3.9), it follows that

$$\begin{aligned} G_{W_2}(r) = &\left[G_X(G_{U_t}(G_{U_t}(r)))\mathbb{1}_{\{X>0\}} + G_X\left(G_{U_t}^{-1}\left(G_{U_t}^{-1}(r)\right)\right)\mathbb{1}_{\{X<0\}}\right]\mathbb{1}_{\{W_1>0\}} + \\ &+ \left[G_X\left(G_{U_t}\left(G_{U_t}^{-1}\left(r^{-1}\right)\right)\right)\mathbb{1}_{\{X>0\}} + G_X\left(G_{U_t}^{-1}\left(G_{U_t}\left(r^{-1}\right)\right)\right)\mathbb{1}_{\{X<0\}}\right]\mathbb{1}_{\{W_1<0\}}. \end{aligned}$$

The generalization to p consecutive signed operators depends on whether p is odd or even. However, the correspondent p.g.f. of W_p will have 2^p $(p \in \mathbb{N})$ terms, where

$$W_p = F_t \odot W_{p-1} = F_t \odot (F_t \odot W_{p-2}) = F_t \odot (F_t \odot (\ldots (F_t \odot X))).$$

Remark: The periodic signed thinning operator $F_t\odot$ lacks the distributive property, i.e.,

$$F_t \odot (X_1 + X_2) \overset{d}{\neq} F_t \odot X_1 + F_t \odot X_2. \tag{3.10}$$

3.1.2 Bivariate case

Bulla et al. (2016) introduced the so-called signed matrix thinning operator as an extension of the signed thinning operator in (1.12) for the bivariate case. For the periodic bivariate case, the signed matrix thinning operator is defined by

$$F_t \odot \mathbf{X} := \begin{bmatrix} F_{11,t} \odot X_1 + F_{12,t} \odot X_2 \\ F_{21,t} \odot X_1 + F_{22,t} \odot X_2 \end{bmatrix}, \tag{3.11}$$

where $\mathbf{X} = [X_1 \ X_2]^T$ is an integer-valued random vector and $F_{ij,t}$ represents the common distribution of the periodic sequence of i.i.d. counting sequences for $(i, j) \in (1, 2) \times (1, 2)$. It is assumed that all counting sequences associated with $F_{ij,t}\odot$ are mutually independent.

In this work, the particular case $F_{12,t}\odot = F_{21,t}\odot = 0$ (assumption of diagonal matrix) will be of interest. Similarly to the univariate case in (3.2), we consider that $F_{j,t}$, the distribution of the periodic sequence of i.i.d. r.v.'s $(U_{k,t}(\phi_{j,t}))_{k\in\mathbb{N}}$, has probability mass function given by

$$P(U_{1,t}(\phi_{j,t}) = a) = \begin{cases} (1-\phi_{j,t})^2 \ , & a = -1 \\ 2\phi_{j,t}(1-\phi_{j,t}) \ , & a = 0 \\ \phi_{j,t}^2 \ , & a = 1 \end{cases}, \tag{3.12}$$

with $\phi_{j,t} = \alpha_{j,v} \in (0, 1)$ for $j = 1, 2;\ t = v + ns;\ v = 1, \ldots, s$ and $n \in \mathbb{N}_0$. Note that, for a fixed v $(v = 1, \ldots, s)$, the random variables

$$U_{i,t}(\phi_{j,t}) = U_t(\phi_{j,t}) \overset{d}{=} R_t(\phi_{j,t}) - 1, \qquad R_t(\phi_{j,t}) \sim Bin(2, \phi_{j,t}),\ j = 1, 2$$

and

$$P\left(\sum_{i=1}^{k} U_{i,t}(\phi_{j,t}) = l\right) = P\left(R_t^{(k)}(\phi_{j,t}) = k + l\right), \qquad l \in \{-k, \ldots, k\},$$

where

$$R_t^{(k)}(\phi_{j,t}) = \sum_{i=1}^{k} R_{i,t}(\phi_{j,t}), \qquad R_t^{(k)}(\phi_{j,t}) \sim Bin(2k, \phi_{j,t}),\ k \in \mathbb{N}. \tag{3.13}$$

Let $x_j \in \mathbb{Z}\backslash\{0\}$ and $y_j \in \mathbb{Z}$ for $j = 1, 2$, the conditional probability function takes the form

$$P\left(F_{j,t} \odot X_j = y_j | X_j = x_j\right) = P\left(sign(x_j)\sum_{i=1}^{|x_j|} U_{i,t}(\phi_{j,t}) = y_j\right) =$$

$$= P\left(R_t^{(|x_j|)}(\phi_{j,t}) = |x_j| + sign(x_j)\cdot y_j\right) =$$

$$= C_{|x_j|+sign(x_j)\cdot y_j}^{2|x_j|}\alpha_{j,v}^{|x_j|+sign(x_j)\cdot y_j}(1-\alpha_{j,v})^{|x_j|-sign(x_j)\cdot y_j},\ y_j \in \{-|x_j|, \ldots, |x_j|\}.$$

Moreover, for $t = v + ns$, $v = 1, \ldots, s$ and $n \in \mathbb{N}_0$, mean value and variance are, respectively,

$$E[F_{j,t} \odot X_j | X_j] = (2\alpha_{j,v} - 1)X_j \tag{3.14}$$

and

$$Var[F_{j,t} \odot X_j | X_j] = 2\alpha_{j,v}(1-\alpha_{j,v})|X_j|,\ j = 1, 2. \tag{3.15}$$

3.2 The periodic Skellam distribution

The Skellam distribution is traditionally linked to Skellam (1946). A brief description of the Skellam distribution and the bivariate Skellam distribution adapted to the periodic case is presented. Basic properties of these periodic distributions are also given, namely, finite first and second-order moments as well as the probability generating function (p.g.f.).

3.2.1 Univariate case

The univariate Skellam distribution, without periodic structure, was given in Definition 1.3. For the periodic case, the definition follows.

Definition 3.1. *(Periodic univariate Skellam distribution)*
Let $\{Z_t\},\ t = v + ns;\ v = 1, \ldots, s$ *and* $n \in \mathbb{N}_0$ *be a periodic sequence of random variables. For a fixed* v $(v = 1, \ldots, s)$*, let* $\lambda_v > 0$ *and* $\tau_v > 0$*. The periodic s-dimensional r.v.* Z_t *follows a periodic Skellam distribution, denoted by* $Skellam(\lambda_v, \tau_v)$*, if and only if*

$$Z_{v+ns} \overset{d}{=} Y_{v+ns} - W_{v+ns},$$

where Y_{v+ns} *and* W_{v+ns} *are two independent random variables such that* $Y_{v+ns} \sim Poisson(\lambda_v)$ *and* $W_{v+ns} \sim Poisson(\tau_v)$*.*

Thus, the probability mass function is given by

$$P(Z_{v+ns} = z) = e^{-(\lambda_v+\tau_v)}\lambda_v^z \sum_{i=max(0,-z)}^{\infty} \frac{(\lambda_v\tau_v)^i}{i!(i+z)!},\ z \in \mathbb{Z}. \tag{3.16}$$

The random vector Z_t has finite first and second-order moments. The mean of Z_t, $t = v+ns$ for a fixed v $(v = 1, \ldots, s)$, is

$$\xi_v = E[Z_{v+ns}] = E[Y_{v+ns} - W_{v+ns}] = \lambda_v - \tau_v. \tag{3.17}$$

Due to the independence of the r.v.'s Y_{v+ns} and W_{v+ns}, the variance of Z_t for $t = v+ns$ with a fixed v is

$$\sigma_v^2 = Var[Z_{v+ns}] = Var[Y_{v+ns} - W_{v+ns}] = \lambda_v + \tau_v. \tag{3.18}$$

The p.g.f. of Z_{v+ns} is $G_{Z_{v+ns}}(r) = exp\{-(\lambda_v + \tau_v) + \lambda_v r + \tau_v/r\}$, $v = 1, \ldots, s$.

3.2.2 Bivariate case

The bivariate Skellam distribution, without periodic structure, has been proposed by Bulla et al. (2015) and is given in Definition 1.8. For the periodic case, the definition of the bivariate distribution follows.

Definition 3.2. *(Periodic bivariate Skellam distribution)*
Let $\mathbf{Z}_t = [Z_{1,t}\ Z_{2,t}]^T$*,* $t = v+ns; v = 1, \ldots, s$ *and* $n \in \mathbb{N}_0$ *be a periodic* $2s$*-dimensional random vector. For a fixed* v $(v = 1, \ldots, s)$*, let*

$$\begin{cases} Z_{1,v+ns} = Y_{1,v+ns} - B_{v+ns} \\ Z_{2,v+ns} = Y_{2,v+ns} - B_{v+ns} \end{cases},$$

where $Y_{j,v+ns}$ $(j = 1, 2)$ *and* B_{v+ns} *are three independent Poisson-distributed variables with parameters* $\lambda_{j,v} > 0$ $(j = 1, 2)$ *and* $\tau_v \geq 0$*, respectively. The random vector* $\mathbf{Z}_t$ *follows a periodic bivariate Skellam distribution, denoted* $BiSkellam(\tau_v, \lambda_{1,v}, \lambda_{2,v})$*, if and only if*

$$Z_{1,v+ns} \sim Skellam(\lambda_{1,v}, \tau_v) \quad and \quad Z_{2,v+ns} \sim Skellam(\lambda_{2,v}, \tau_v).$$

Thus, the joint probability mass function is given by

$$\begin{aligned} &P(Z_{1,v+ns} = z_1, Z_{2,v+ns} = z_2) = \\ &= e^{-(\lambda_{1,v}+\lambda_{2,v}+\tau_v)} \lambda_{1,v}^{z_1} \lambda_{2,v}^{z_2} \sum_{i=max(0,-z_1,-z_2)}^{\infty} \frac{(\lambda_{1,v}\lambda_{2,v}\tau_v)^i}{i!(i+z_1)!(i+z_2)!}, \ (z_1, z_2) \in \mathbb{Z}^2. \end{aligned} \tag{3.19}$$

The mean vector of $\mathbf{Z}_t$ is

$$E[\mathbf{Z}_t] = E \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix} = \begin{bmatrix} \delta_{1,t} \\ \delta_{2,t} \end{bmatrix} = \boldsymbol{\delta}_t. \tag{3.20}$$

Each s-vector $\delta_{j,t}$ $(j = 1, 2)$ with $t = v + ns;\ v = 1, \ldots, s$ and $n \in \mathbb{N}_0$ is given by

$$E[Z_{j,t}] = \boldsymbol{\delta}_{j,t} = \begin{bmatrix} \xi_{j,1} & \xi_{j,2} & \ldots & \xi_{j,s} \end{bmatrix}^T. \tag{3.21}$$

For a fixed v, each element of vector (3.21) is

$$\xi_{j,v} = E[Z_{j,v+ns}] = \lambda_{j,v} - \tau_v. \tag{3.22}$$

The variance-covariance matrix of $\mathbf{Z}_t$ is given by

$$\sum\nolimits_{\mathbf{Z}_t} = Var[\mathbf{Z}_t] = \begin{bmatrix} Var[Z_{1,t}] & Cov(Z_{1,t}, Z_{2,t}) \\ Cov(Z_{2,t}, Z_{1,t}) & Var[Z_{2,t}] \end{bmatrix} = \begin{bmatrix} \psi_{11,t} & \psi_{12,t} \\ \psi_{21,t} & \psi_{22,t} \end{bmatrix} = \psi_t, \tag{3.23}$$

where $\psi_{jj,t}$ for $j = 1, 2;\ t = v + ns;\ v = 1, \ldots, s$ and $n \in \mathbb{N}_0$ are $(s \times s)$ diagonal matrices

$$\psi_{jj,t} = Var[Z_t] = \begin{bmatrix} \sigma^2_{j,1} & 0 & \ldots & 0 \\ 0 & \sigma^2_{j,2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma^2_{j,s} \end{bmatrix} \tag{3.24}$$

and for $j \neq k$ $(j, k = 1, 2)$, the matrix $\psi_{jk,t}$ takes the form

$$\psi_{jk,t} = Cov(Z_{j,t}, Z_{k,t}) = \begin{bmatrix} \sigma_{jk,1} & 0 & \ldots & 0 \\ 0 & \sigma_{jk,2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_{jk,s} \end{bmatrix}. \tag{3.25}$$

For a fixed v, each element of the diagonal in matrix (3.24) is given by

$$\sigma^2_{j,v} = Var[Z_{j,v+ns}] = \lambda_{j,v} + \tau_v. \tag{3.26}$$

and for matrix (3.25) by

$$\sigma_{jk,v} = Cov(Z_{j,v+ns}, Z_{k,v+ns}) = Cov(B_{v+ns}, B_{v+ns}) = Var[B_{v+ns}] = \tau_v. \tag{3.27}$$

3.3 The univariate periodic model: S-PINAR(1)

The integer-valued autoregressive models with binomial thinning operators have non-negative coefficients. Thus modeling of series with possible negative autocorrelations are excluded. Moreover, those models defined on $\mathbb{N}$ cannot fit a time series with negative observations. Motivated by the work of Chesneau and Kachour (2012), we extend their univariate model with signed thinning operator to the periodic case, introducing the signed periodic INAR(1) process (S-PINAR(1) for short) with period s. A parametric assumption on the common distribution of the periodic counting sequence of the model is made. Emphasis is placed upon a specific parametric case that arises under the assumption of periodic Skellam-distributed innovation. In contrast to traditional INAR(1) models, these models are defined in $\mathbb{Z}$ allowing for negative integer values and negative correlation. The properties of the S-PINAR(1) model with period s are discussed. Regarding parameter estimation, two methods are considered: conditional least squares and conditional maximum likelihood. The performance of the proposed estimation methods for the S-PINAR(1) model with period s is accomplished and compared through a simulation study.

3.3.1 Definition and basic properties

Let $\{X_t\}$ be a periodic integer-valued autoregressive process of first-order defined by the recursion

$$X_t = F_t \odot X_{t-1} + Z_t,\ t \in \mathbb{Z}, \tag{3.28}$$

where X_t, X_{t-1} and Z_t are random s-vectors for $t = v + ns$; $v = 1, \dots, s$ and $n \in \mathbb{N}_0$. The random vector $Z_t = [Z_{1+ns}\ Z_{2+ns}\ \dots\ Z_{s+ns}]^T$ represents a periodic sequence of independent random variables. The model defined in equation (3.28) will be referred to as S-PINAR(1) for Signed Periodic INteger-valued AutoRegressive model of order one with period $s \in \mathbb{N}$. For each t, the innovation term Z_t in recursion (3.28) is assumed to be independent of X_{t-1} and

$F_t \odot X_{t-1}$. Writing the periodic signed thinning operator in (3.1) as

$$F_t\odot = \begin{cases} f_1\odot, & t = 1+ns \\ f_2\odot, & t = 2+ns \\ \vdots & \\ f_s\odot, & t = s+ns \end{cases}, \tag{3.29}$$

the periodic model in (3.28) can have the form

$$X_{v+ns} = f_v \odot X_{v-1+ns} + Z_{v+ns}, \tag{3.30}$$

where $f_v \odot X_{v-1+ns} = sign(X_{v-1+ns}) \sum_{i=1}^{|X_{v-1+ns}|} U_{i,t}(\phi_t)$ with $U_{i,t}(\phi_t)$ as defined in (3.3).

We assume the innovation term Z_t in the `S-PINAR`(1) model proposed in (3.28) follows the periodic Skellam distribution with parameters λ_v and τ_v established in Definition 3.1 with p.m.f. given by equation (3.16). Therefore, for a fixed v ($v = 1, \ldots, s$) the first and second-order moments of Z_{v+ns} are defined in (3.17) and (3.18), respectively.

Some distributional properties of the `S-PINAR`(1) process in recursion (3.28) with Skellam-distributed innovation are derived, namely the conditional moments of first and second-order of the model. Hence, from (3.6) and (3.17)

$$\begin{aligned} E[X_{v+ns}|X_{v-1+ns}] &= E[f_v \odot X_{v-1+ns} + Z_{v+ns}|X_{v-1+ns}] = \\ &= (2\alpha_v - 1)X_{v-1+ns} + \lambda_v - \tau_v \end{aligned} \tag{3.31}$$

and from equations (3.7) and (3.18),

$$\begin{aligned} Var[X_{v+ns}|X_{v-1+ns}] &= Var[f_v \odot X_{v-1+ns} + Z_{v+ns}|X_{v-1+ns}] = \\ &= 2\alpha_v(1-\alpha_v)|X_{v-1+ns}| + \lambda_v + \tau_v. \end{aligned} \tag{3.32}$$

Recall the periodic signed thinning operator given in (3.1) lacks the distributive property in (3.10) which limits the development of other properties concerning the S-PINAR(1) process with period s. For a fixed value of $v = 1, \ldots, s$, the process $\{X_t\}$ with $t = v + ns$ is a Markov chain with transition probability function

$$
\begin{aligned}
p_v(b|a) &= P(X_{v+ns} = b | X_{v-1+ns} = a) = \\
&= \sum_{l=-|a|}^{|a|} P\left(sign(a) \sum_{i=1}^{|a|} U_{i,t}(\phi_t) = l \right) \times P(Z_{v+ns} = b - l) = \\
&= \sum_{l=-|a|}^{|a|} P\left(R_t^{(|a|)}(\phi_t) = |a| + sign(a) \cdot l \right) \times P(Z_{v+ns} = b - l) = \\
&= \sum_{l=-|a|}^{|a|} \left\{ C^{2|a|}_{|a|+sign(a)\cdot l} \alpha_v^{|a|+sign(a)\cdot l} (1-\alpha_v)^{|a|-sign(a)\cdot l} \times \right. \\
&\left. \times\ e^{-(\lambda_v+\tau_v)} \lambda_v^{b-l} \sum_{i=max(0,-(b-l))}^{\infty} \frac{(\lambda_v \tau_v)^i}{i!(i+b-l)!} \right\},
\end{aligned} \tag{3.33}
$$

where the f.m.p. of $R_t^{(|a|)}$ and Z_{v+ns} can be found in (3.4) and (3.16), respectively.

3.3.2 Parameter estimation of the S-PINAR(1) model

This subsection is devoted to parameter estimation of the S-PINAR(1) process with period s under the parametric assumption previously mentioned. Lets us assume we have $(X_0, X_1, \ldots, X_{Ns})$ observations from the S-PINAR(1) process with Skellam-distributed innovations. Two estimation methods are proposed to estimate the parameters of the model: conditional least squares and conditional maximum likelihood. For the S-PINAR(1) model with period s, the vector of unknown parameters $\boldsymbol{\theta}$ has $3s$ parameters, i.e.,

$$\boldsymbol{\theta} := (\boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\tau}) \tag{3.34}$$

with $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_s)$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_s)$ and $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_s)$.

3.3.2.1 Conditional least squares estimation

The conditional least squares (CLS) estimator of the vector of the unknown parameters in (3.34) is $\widehat{\boldsymbol{\theta}}_{CLS} := (\widehat{\boldsymbol{\alpha}}^{CLS}, \widehat{\boldsymbol{\lambda}}^{CLS}, \widehat{\boldsymbol{\tau}}^{CLS})$. The estimation procedure that follows was proposed by Klimko and Nelson (1978). The CLS estimators of $\boldsymbol{\theta}$ are obtained by minimizing the criterion function $S_1(\boldsymbol{\theta})$ given by

$$\begin{aligned} S_1(\boldsymbol{\theta}) &= \sum_{n=0}^{N-1} \sum_{v=1}^{s} \left(X_{v+ns} - E[X_{v+ns}|X_{v-1+ns}]\right)^2 = \\ &= \sum_{n=0}^{N-1} \sum_{v=1}^{s} \left(X_{v+ns} - (2\alpha_v - 1)X_{v-1+ns} - \lambda_v + \tau_v\right)^2 . \end{aligned}$$

It is clear that differentiating $S_1(\boldsymbol{\theta})$ with respect to λ_v and τ_v and equating the resulting expressions to zero, the same equation is obtained. For these parameters, direct CLS estimators are not available. The conditional least squares method was adapted by Alzaid and Omair (2014) with some modifications in order to be able to estimate all parameters integrating the model. Hence, in order to estimate λ_v and τ_v using the CLS method, the following reparametrization is needed

$$\begin{cases} \xi_v = \lambda_v - \tau_v \\ \sigma_v^2 = \lambda_v + \tau_v \end{cases}, \quad v = 1, \ldots, s. \tag{3.35}$$

Estimators for all parameters of the `S-PINAR`(1) process, i.e., α_v, ξ_v and σ_v^2 are obtained in a two step procedure as described below.

First step - estimates for α_v and ξ_v $(v = 1, \ldots, s)$:

Consider the conditional mean prediction error

$$\begin{aligned} e_{1,v+ns} &= X_{v+ns} - E[X_{v+ns}|X_{v-1+ns}] = \\ &= X_{v+ns} - (2\alpha_v - 1)X_{v-1+ns} - \xi_v, \end{aligned} \tag{3.36}$$

where conditional first-order moment $E[X_{v+ns}|X_{v-1+ns}]$ is defined in (3.31). The CLS estimators of α_v and ξ_v are derived by minimizing the criterion function

$$S_2(\boldsymbol{\theta}) = \sum_{n=0}^{N-1}\sum_{v=1}^{s} e^2_{1,v+ns} = \sum_{n=0}^{N-1}\sum_{v=1}^{s} \left(X_{v+ns} - (2\alpha_v - 1)X_{v-1+ns} - \xi_v\right)^2.$$

After differentiating $S_2(\boldsymbol{\theta})$ with respect to parameters α_v and ξ_v, the following system of equations arises

$$\begin{cases} \dfrac{\partial S_2(\boldsymbol{\theta})}{\partial \alpha_v} = \sum\limits_{n=0}^{N-1} \left(X_{v+ns} - (2\alpha_v - 1)X_{v-1+ns} - \xi_v\right) X_{v-1+ns} = 0 \\ \\ \dfrac{\partial S_2(\boldsymbol{\theta})}{\partial \xi_v} = \sum\limits_{n=0}^{N-1} \left(X_{v+ns} - (2\alpha_v - 1)X_{v-1+ns} - \xi_v\right) = 0 \end{cases}$$

and consequently, for $v = 1, \ldots, s$, the CLS estimators are

$$\begin{cases} \widehat{\alpha}_v^{CLS} = \dfrac{1}{2}\left(\dfrac{N\sum\limits_{n=0}^{N-1} X_{v+ns}X_{v-1+ns} - \sum\limits_{n=0}^{N-1} X_{v+ns}\sum\limits_{n=0}^{N-1} X_{v-1+ns}}{N\sum\limits_{n=0}^{N-1} X^2_{v-1+ns} - \left(\sum\limits_{n=0}^{N-1} X_{v-1+ns}\right)^2} + 1\right) \\ \\ \widehat{\xi}_v^{CLS} = \dfrac{1}{N}\left(\sum\limits_{n=0}^{N-1} X_{v+ns} - (2\widehat{\alpha}_v^{CLS} - 1)\sum\limits_{n=0}^{N-1} X_{v-1+ns}\right) \end{cases}. \tag{3.37}$$

Second step - estimate for σ_v^2 $(v = 1, \ldots, s)$:

The conditional variance prediction error has been used by Alzaid and Omair (2014) to obtain the CLS estimator for the variance parameter. Thus in the periodic case, the conditional variance prediction error is defined by

$$\begin{aligned} e_{2,v+ns} &= (X_{v+ns} - E[X_{v+ns}|X_{v-1+ns}])^2 - Var[X_{v+ns}|X_{v-1+ns}] = \\ &= e_{1,v+ns}^2 - 2\alpha_v(1-\alpha_v)|X_{v-1+ns}| - \sigma_v^2 \end{aligned} \tag{3.38}$$

with conditional moments $E[X_{v+ns}|X_{v-1+ns}]$ and $Var[X_{v+ns}|X_{v-1+ns}]$ in equations (3.31) and (3.32), respectively. The conditional mean prediction error $(e_{1,v+ns})$ is derived in the first step of the estimation procedure from (3.36). The equation $\sum_{n=0}^{N-1} e_{2,v+ns} = 0$ yields a direct estimator for σ_v^2 by solving the nonlinear equation

$$\sum_{n=0}^{N-1} \left(\widehat{e}_{1,v+ns}^2 - 2\widehat{\alpha}_v^{CLS}(1-\widehat{\alpha}_v^{CLS})|X_{v-1+ns}| - \sigma_v^2\right) = 0,$$

i.e.,

$$\widehat{\sigma}_v^2 = \frac{1}{N}\sum_{n=0}^{N-1} \left(\widehat{e}_{1,v+ns}^2 - 2\widehat{\alpha}_v^{CLS}(1-\widehat{\alpha}_v^{CLS})|X_{v-1+ns}|\right), \tag{3.39}$$

where $\widehat{e}_{1,v+ns} = X_{v+ns} - (2\widehat{\alpha}_v^{CLS} - 1)X_{v-1+ns} - \widehat{\xi}_v^{CLS}$ with CLS estimators $\widehat{\alpha}_v^{CLS}$ and $\widehat{\xi}_v^{CLS}$ in (3.37). After estimating σ_v^2 through (3.39), the CLS estimators of λ_v and τ_v from reparametrization (3.35) take the form

$$\begin{cases} \widehat{\lambda}_v^{CLS} = \dfrac{1}{2}\left(\widehat{\sigma}_v^{2,CLS} + \widehat{\xi}_v^{CLS}\right) \\ \\ \widehat{\tau}_v^{CLS} = \dfrac{1}{2}\left(\widehat{\sigma}_v^{2,CLS} - \widehat{\xi}_v^{CLS}\right) \end{cases}, \quad v = 1, \ldots, s. \tag{3.40}$$

Alzaid and Omair (2014) have also considered an alternative method for estimating the variance in the second step and compared both estimators.

3.3.2.2 Conditional maximum likelihood estimation

The conditional maximum likelihood (CML) estimator of the vector of the unknown parameters in (3.34) is $\widehat{\boldsymbol{\theta}}_{CML} := (\widehat{\boldsymbol{\alpha}}^{CML}, \widehat{\boldsymbol{\lambda}}^{CML}, \widehat{\boldsymbol{\tau}}^{CML})$. The conditional log-likelihood function is given by

$$C(\boldsymbol{\theta}) = ln(L(\boldsymbol{\theta}|\mathbf{x})) = \sum_{n=0}^{N-1}\sum_{v=1}^{s} ln\left(p_v(x_{v+ns}|x_{v-1+ns})\right), \tag{3.41}$$

where $p_v(b|a)$ has the expression given in (3.33) by replacing $a = x_{v-1+ns}$ and $b = x_{v+ns}$. Differentiating the conditional log-likelihood function in equation (3.41) with respect to the parameters α_v, λ_v and τ_v $(v = 1, \ldots, s)$ in (3.34), the system of first-order partial derivatives follows

$$\begin{cases} \dfrac{\partial C(\boldsymbol{\theta})}{\partial \alpha_v} = 0 \\[2ex] \dfrac{\partial C(\boldsymbol{\theta})}{\partial \lambda_v} = 0 \\[2ex] \dfrac{\partial C(\boldsymbol{\theta})}{\partial \tau_v} = 0 \end{cases} \Leftrightarrow \begin{cases} \sum\limits_{n=0}^{N-1} \dfrac{\frac{\partial}{\partial \alpha_v} p_v(x_{v+ns}|x_{v-1+ns})}{p_v(x_{v+ns}|x_{v-1+ns})} = 0 \\[2ex] \sum\limits_{n=0}^{N-1} \dfrac{\frac{\partial}{\partial \lambda_v} p_v(x_{v+ns}|x_{v-1+ns})}{p_v(x_{v+ns}|x_{v-1+ns})} = 0 \\[2ex] \sum\limits_{n=0}^{N-1} \dfrac{\frac{\partial}{\partial \tau_v} p_v(x_{v+ns}|x_{v-1+ns})}{p_v(x_{v+ns}|x_{v-1+ns})} = 0 \end{cases}, \quad v = 1, \ldots, s,$$

i.e.,

$$\begin{cases} \sum\limits_{n=0}^{N-1} \dfrac{2|x_{v-1+ns}|}{1-\alpha_v} \left(\dfrac{p_v(x_{v+ns}-1|x_{v-1+ns}-1)}{p_v(x_{v+ns}|x_{v-1+ns})} - 1 \right) = 0 \\[2ex] \sum\limits_{n=0}^{N-1} \dfrac{p_v(x_{v+ns}-1|x_{v-1+ns})}{p_v(x_{v+ns}|x_{v-1+ns})} = N \\[2ex] \sum\limits_{n=0}^{N-1} \dfrac{p_v(x_{v+ns}+1|x_{v-1+ns})}{p_v(x_{v+ns}|x_{v-1+ns})} = N \end{cases}, \quad v = 1, \ldots, s.$$

First-order partial derivatives of transition probability function $p_v(x_{v+ns}|x_{v-1+ns})$ are available in Appendix C.1. Numerical maximization can be obtained with standard statistical packages in `R`.

3.3.3 Simulation study

In order to provide an idea about the relative merits of each method (CLS and CML) used in parameter estimation of the `S-PINAR`(1) model with period s and Skellam-distributed innovation term, a simulation study is conducted. To generate count data from the periodic univariate model proposed in (3.28), we have set period $s = 4$, thus the vector of unknown parameters in (3.34) is $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\tau}) = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \tau_1, \tau_2, \tau_3 \tau_4)$. Several combinations of values for parameters $\boldsymbol{\alpha}, \boldsymbol{\lambda}$ and $\boldsymbol{\tau}$ are available in Table 3.1. Three sets: Set 1, Set 2 and Set 3 are displayed. Each set has been subdivided into settings A and B, where parameter $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ is fixed. Hence in Table 3.1, the different scenarios will be referred to as Set 1A, Set 1B, Set 2A, Set 2B, Set 3A and Set 3B. For Set 1, values for α_v ($v = 1, 2, 3, 4$) are above and below 0.5. For both settings (A and B), different values for $\boldsymbol{\lambda}$ are considered while parameter $\boldsymbol{\tau}$ remains the same. Regarding Set 2, values for α_v are all below 0.5 and both parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\tau}$ take different values. For Set 3, values for α_v are all above 0.5, parameter $\boldsymbol{\lambda}$ is fixed but parameter $\boldsymbol{\tau}$ assumes different values. The choice for certain values of parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\tau}$ arise from the fact that $\lambda_v - \tau_v$ represents the mean of Z_{v+ns} given in (3.17).

Table 3.1: Parameters: $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ and $\boldsymbol{\tau} = (\tau_1, \tau_2, \tau_3, \tau_4)$.

Set 1	A:	$\boldsymbol{\alpha} = (0.60, 0.40, 0.75, 0.30)$;	$\boldsymbol{\lambda} = (2, 1, 6, 5)$;	$\boldsymbol{\tau} = (4, 5, 3, 1)$
	B:	$\boldsymbol{\alpha} = (0.60, 0.40, 0.75, 0.30)$;	$\boldsymbol{\lambda} = (5, 2, 1, 6)$;	$\boldsymbol{\tau} = (4, 5, 3, 1)$
Set 2	A:	$\boldsymbol{\alpha} = (0.20, 0.45, 0.10, 0.30)$;	$\boldsymbol{\lambda} = (2, 1, 6, 5)$;	$\boldsymbol{\tau} = (4, 5, 3, 1)$
	B:	$\boldsymbol{\alpha} = (0.20, 0.45, 0.10, 0.30)$;	$\boldsymbol{\lambda} = (5, 2, 1, 6)$;	$\boldsymbol{\tau} = (2, 1, 4, 3)$
Set 3	A:	$\boldsymbol{\alpha} = (0.75, 0.62, 0.51, 0.86)$;	$\boldsymbol{\lambda} = (4, 5, 3, 1)$;	$\boldsymbol{\tau} = (1, 3, 2, 4)$
	B:	$\boldsymbol{\alpha} = (0.75, 0.62, 0.51, 0.86)$;	$\boldsymbol{\lambda} = (4, 5, 3, 1)$;	$\boldsymbol{\tau} = (2, 1, 4, 3)$

Three sample sizes are contemplated in this simulation study: $n = 4N = 200, 800, 2000$, i.e., $N = 50, 200, 500$ cycles. For a fixed set of parameters in Table 3.1, 1000 independent replications of the `S-PINAR`(1) process have been generated. The results from the simulation study for three scenarios, Set 1A, Set 2A and Set 3A, are summarized through Tables 3.2-3.4 and Figures 3.1-3.6. The results (tables and figures) for the remaining scenarios (Set 1B, Set 2B and Set 3B) are displayed in Appendix C.2. All simulation and estimation procedures were realized through functions written in `R` and available in Appendix D.2.

Tables 3.2-3.4 report the average parameter estimates for the three mentioned sets. To facilitate comparison between the CLS and CML methods and the aforementioned sample sizes, the mean square error (MSE) was computed and included in parenthesis below each estimate. According to Tables 3.2-3.4, parameter estimates in both cases are very close, because both methods give consistent estimates of the parameters. Nevertheless, the autoregressive parameters $\boldsymbol{\alpha}$ appear to be less biased. For smaller samples, the CLS method seems to have a better performance in estimating the parameters. Computationally, there is extra work with the CML method. The accuracy of all estimation improves as the length of the time series increases. When length increases from $N = 50$ to $N = 200$, the improvement of accuracy is more obvious than when length increases from $N = 200$ to $N = 500$.

The bias of the produced estimates were used to quantify their quality. The boxplots of the bias for different combinations of parameters in Set 1A, Set 2A and Set 3A are in Figures 3.1-3.6. These figures also show the effect of sample size on the behavior of CLS and CML estimators. No matter the sample size, the difference between CLS and CML is small and becomes even smaller when the length of time series increases. The estimates for parameter $\boldsymbol{\lambda}$ seem slightly worse when parameter $\boldsymbol{\alpha}$ has all values above 0.5 (Set 3A). The estimates for parameter $\boldsymbol{\tau}$ seem slightly worse when parameter $\boldsymbol{\alpha}$ has all values below 0.5 (Set 2A). Furthermore, Figures 3.1-3.6 reveal that estimates of $\boldsymbol{\lambda}$ and $\boldsymbol{\tau}$ componentwise tend to be biased to the left which implies that both estimation methods have a tendency to underestimate $\boldsymbol{\lambda}$ and $\boldsymbol{\tau}$, mainly in the case of small sample sizes. Regarding parameter $\boldsymbol{\alpha}$, where α_v are below 0.5 (Set 2A), it can also be observed that both methods produce slightly overestimated estimates, componentwise. As expected, both bias and skewness approach zero as sample size increases. Overall, the difference between the two approaches will vanish when the length of time series increases.

Table 3.2: CLS and CML estimates for $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\tau})$ in Set 1A. MSE in parenthesis.

	$N = 50$		$N = 200$		$N = 500$	
	CLS	CML	CLS	CML	CLS	CML
$\boldsymbol{\alpha} = (0.60, 0.40, 0.75, 0.30)$						
$\hat{\alpha}_1$	0.599	0.600	0.601	0.600	0.600	0.601
	(0.0003)	(0.0006)	(0.0002)	(0.0015)	(0.0006)	(0.0004)
$\hat{\alpha}_2$	0.407	0.406	0.399	0.402	0.401	0.402
	(0.0113)	(0.0001)	(0.0007)	(0.0036)	(0.0002)	(0.0001)
$\hat{\alpha}_3$	0.749	0.750	0.747	0.751	0.751	0.750
	(0.0160)	(0.0116)	(0.0001)	(0.0028)	(0.0007)	(0.0001)
$\hat{\alpha}_4$	0.302	0.297	0.301	0.300	0.300	0.299
	(0.0003)	(0.0095)	(0.0010)	(0.0002)	(0.0001)	(0.0002)
$\boldsymbol{\lambda} = (2, 1, 6, 5)$						
$\hat{\lambda}_1$	1.915	1.880	1.963	1.982	1.994	2.000
	(0.1013)	(0.8744)	(0.0048)	(0.0774)	(0.0001)	(0.0319)
$\hat{\lambda}_2$	1.001	0.895	0.963	0.975	0.989	0.998
	(0.6518)	(0.5694)	(0.0099)	(0.0357)	(0.0358)	(0.0542)
$\hat{\lambda}_3$	5.806	5.810	5.910	5.987	5.985	5.999
	(0.0948)	(0.1168)	(0.1250)	(0.0944)	(0.1330)	(0.1804)
$\hat{\lambda}_4$	4.964	4.936	4.952	4.960	4.975	4.977
	(0.4467)	(0.3806)	(0.0203)	(0.2504)	(0.0012)	(0.0220)
$\boldsymbol{\tau} = (4, 5, 3, 1)$						
$\hat{\tau}_1$	3.893	3.880	3.965	3.970	3.991	3.998
	(1.0386)	(0.9610)	(0.0930)	(0.0106)	(0.1715)	(0.0027)
$\hat{\tau}_2$	4.977	4.868	4.967	4.968	4.987	4.995
	(0.7353)	(1.2057)	(0.0203)	(0.1526)	(0.0397)	(0.0037)
$\hat{\tau}_3$	2.812	2.834	2.937	2.970	2.984	2.987
	(1.0748)	(0.3680)	(0.0517)	(0.0001)	(0.0200)	(0.0295)
$\hat{\tau}_4$	1.002	0.915	0.959	0.961	0.981	0.977
	(0.4951)	(0.6673)	(0.0111)	(0.2119)	(0.0225)	(0.0019)

Table 3.3: CLS and CML estimates for $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\tau})$ in Set 2A. MSE in parenthesis.

	$N = 50$		$N = 200$		$N = 500$	
	CLS	CML	CLS	CML	CLS	CML
$\boldsymbol{\alpha} = (0.20, 0.45, 0.10, 0.30)$						
$\hat{\alpha}_1$	0.199	0.201	0.201	0.200	0.199	0.200
	(0.0047)	(0.0031)	(0.0020)	(0.0003)	(0.0014)	(0.0001)
$\hat{\alpha}_2$	0.455	0.453	0.451	0.451	0.449	0.450
	(0.0002)	(0.0014)	(0.0035)	(0.0021)	(0.0011)	(0.0004)
$\hat{\alpha}_3$	0.115	0.108	0.100	0.101	0.099	0.101
	(0.0054)	(0.0004)	(0.0040)	(0.0002)	(0.0017)	(0.0006)
$\hat{\alpha}_4$	0.305	0.304	0.300	0.300	0.300	0.300
	(0.0121)	(0.0001)	(0.0001)	(0.0013)	(0.0007)	(0.0001)
$\boldsymbol{\lambda} = (2, 1, 6, 5)$						
$\hat{\lambda}_1$	1.883	1.855	1.973	1.959	1.987	1.985
	(0.2336)	(0.0185)	(0.0705)	(0.0135)	(0.1523)	(0.0421)
$\hat{\lambda}_2$	1.047	0.886	0.970	0.964	0.989	0.983
	(0.6237)	(1.0923)	(0.0235)	(0.2988)	(0.0252)	(0.0511)
$\hat{\lambda}_3$	5.814	5.909	5.936	5.959	5.964	5.992
	(0.0004)	(0.0588)	(0.0523)	(0.5441)	(0.0001)	(0.0281)
$\hat{\lambda}_4$	4.981	4.803	4.962	4.934	4.974	4.963
	(1.7878)	(1.6886)	(0.1190)	(0.5135)	(0.0652)	(0.1936)
$\boldsymbol{\tau} = (4, 5, 3, 1)$						
$\hat{\tau}_1$	3.890	3.853	3.976	3.961	3.986	3.987
	(0.6893)	(0.5806)	(0.0014)	(0.0439)	(0.1394)	(0.0361)
$\hat{\tau}_2$	4.984	4.854	4.971	4.955	5.000	4.978
	(0.9576)	(2.0269)	(0.2442)	(0.0650)	(0.1503)	(0.0064)
$\hat{\tau}_3$	2.708	2.841	2.933	2.948	2.973	2.984
	(0.2179)	(0.0098)	(0.5292)	(0.4945)	(0.0160)	(0.0002)
$\hat{\tau}_4$	1.054	0.832	0.978	0.930	0.975	0.961
	(0.0616)	(0.4011)	(0.2080)	(0.0057)	(0.2973)	(0.0700)

Table 3.4: CLS and CML estimates for $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\tau})$ in Set 3A. MSE in parenthesis.

	$N = 50$		$N = 200$		$N = 500$	
	CLS	CML	CLS	CML	CLS	CML
$\boldsymbol{\alpha} = (0.75, 0.62, 0.51, 0.86)$						
$\hat{\alpha}_1$	0.749	0.752	0.750	0.750	0.750	0.750
	(0.0005)	(0.0023)	(0.0003)	(0.0015)	(0.0003)	(0.0001)
$\hat{\alpha}_2$	0.620	0.620	0.620	0.621	0.619	0.620
	(0.0008)	(0.0146)	(0.0017)	(0.0024)	(0.0013)	(0.0002)
$\hat{\alpha}_3$	0.507	0.510	0.510	0.511	0.511	0.511
	(0.0011)	(0.0126)	(0.0003)	(0.0001)	(0.0019)	(0.0001)
$\hat{\alpha}_4$	0.857	0.858	0.860	0.860	0.861	0.859
	(0.0003)	(0.0058)	(0.0002)	(0.0003)	(0.0002)	(0.0004)
$\boldsymbol{\lambda} = (4, 5, 3, 1)$						
$\hat{\lambda}_1$	3.990	3.922	3.993	3.980	3.996	3.990
	(0.0367)	(0.0687)	(0.0102)	(0.0042)	(0.0201)	(0.0585)
$\hat{\lambda}_2$	4.801	4.840	4.963	4.949	4.983	4.971
	(1.3727)	(0.5495)	(0.0098)	(0.0311)	(0.0340)	(0.0925)
$\hat{\lambda}_3$	2.852	2.932	2.972	2.982	2.982	2.994
	(0.4910)	(0.0041)	(0.0086)	(0.0026)	(0.0223)	(0.0521)
$\hat{\lambda}_4$	0.956	0.907	0.969	0.963	0.984	0.981
	(0.0891)	(0.0723)	(0.0019)	(0.0982)	(0.0002)	(0.0867)
$\boldsymbol{\tau} = (1, 3, 2, 4)$						
$\hat{\tau}_1$	1.000	0.896	1.001	0.969	0.997	0.986
	(0.2271)	(0.0954)	(0.0128)	(0.0003)	(0.1242)	(0.0113)
$\hat{\tau}_2$	2.803	2.845	2.957	2.956	2.979	2.975
	(0.5236)	(0.4166)	(0.0197)	(0.0427)	(0.0248)	(0.2250)
$\hat{\tau}_3$	1.843	1.925	1.958	1.989	1.979	2.005
	(0.2197)	(0.0357)	(0.0213)	(0.0062)	(0.1071)	(0.0328)
$\hat{\tau}_4$	3.938	3.904	3.964	3.960	3.988	3.980
	(0.0365)	(0.0559)	(0.0017)	(0.0193)	(0.0156)	(0.0165)

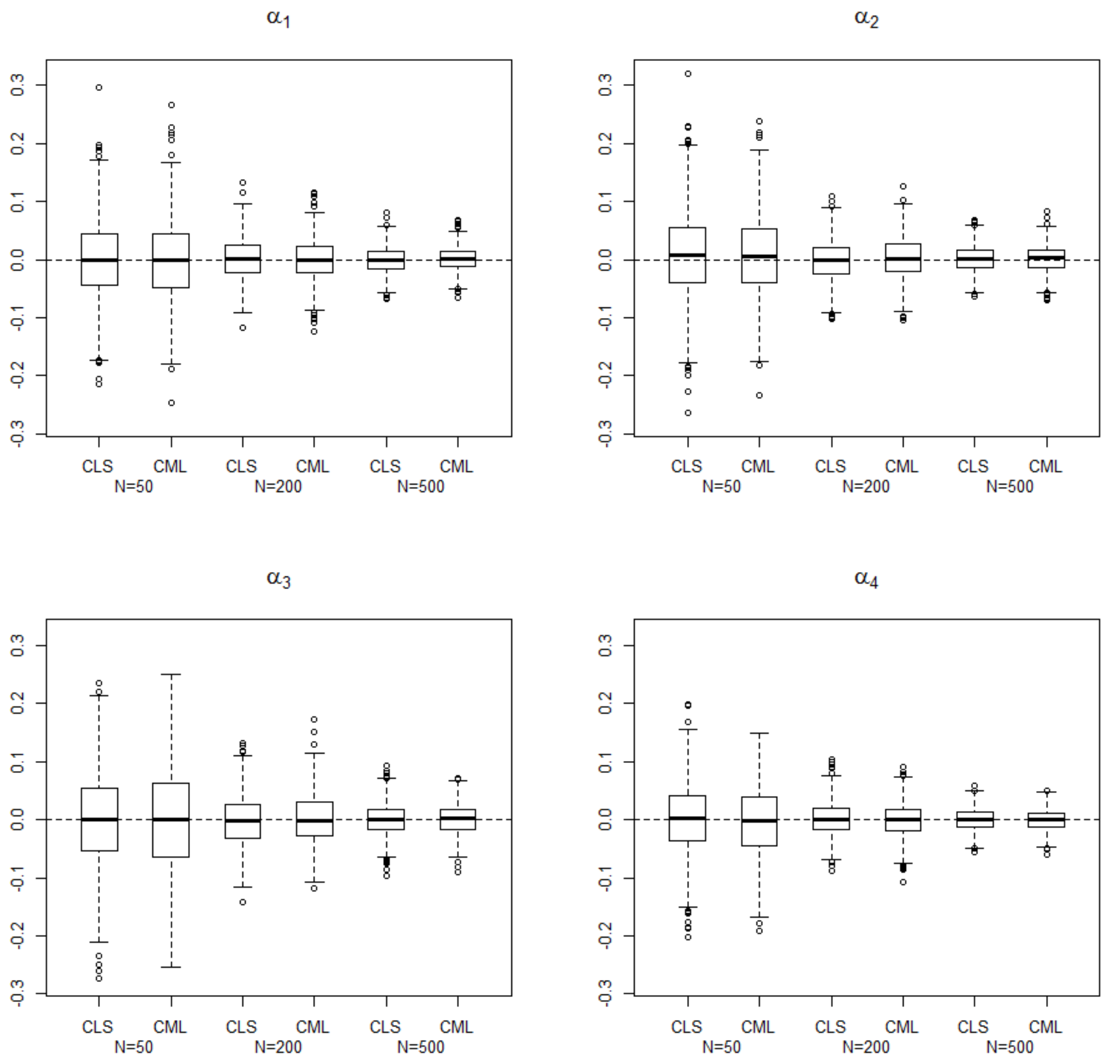

Figure 3.1: Boxplots for the biases of the CLS and CML estimates of parameter $\boldsymbol{\alpha}$ in Set 1A for $n = 4N = 200, 800, 2000$.

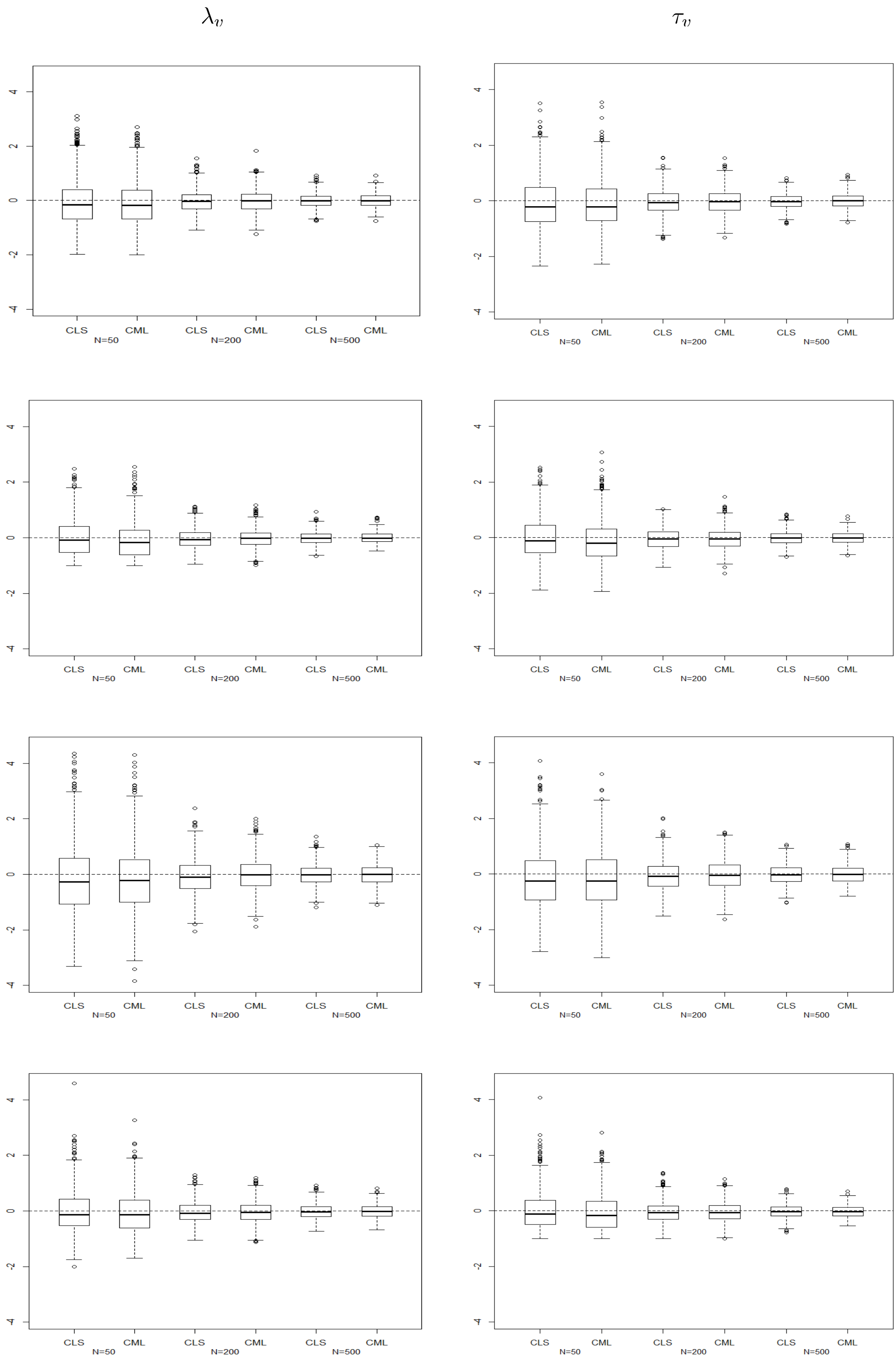

Figure 3.2: Boxplots for the biases of the CLS and CML estimates of parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\tau}$ in Set 1A for $n = 4N = 200, 800, 2000$.

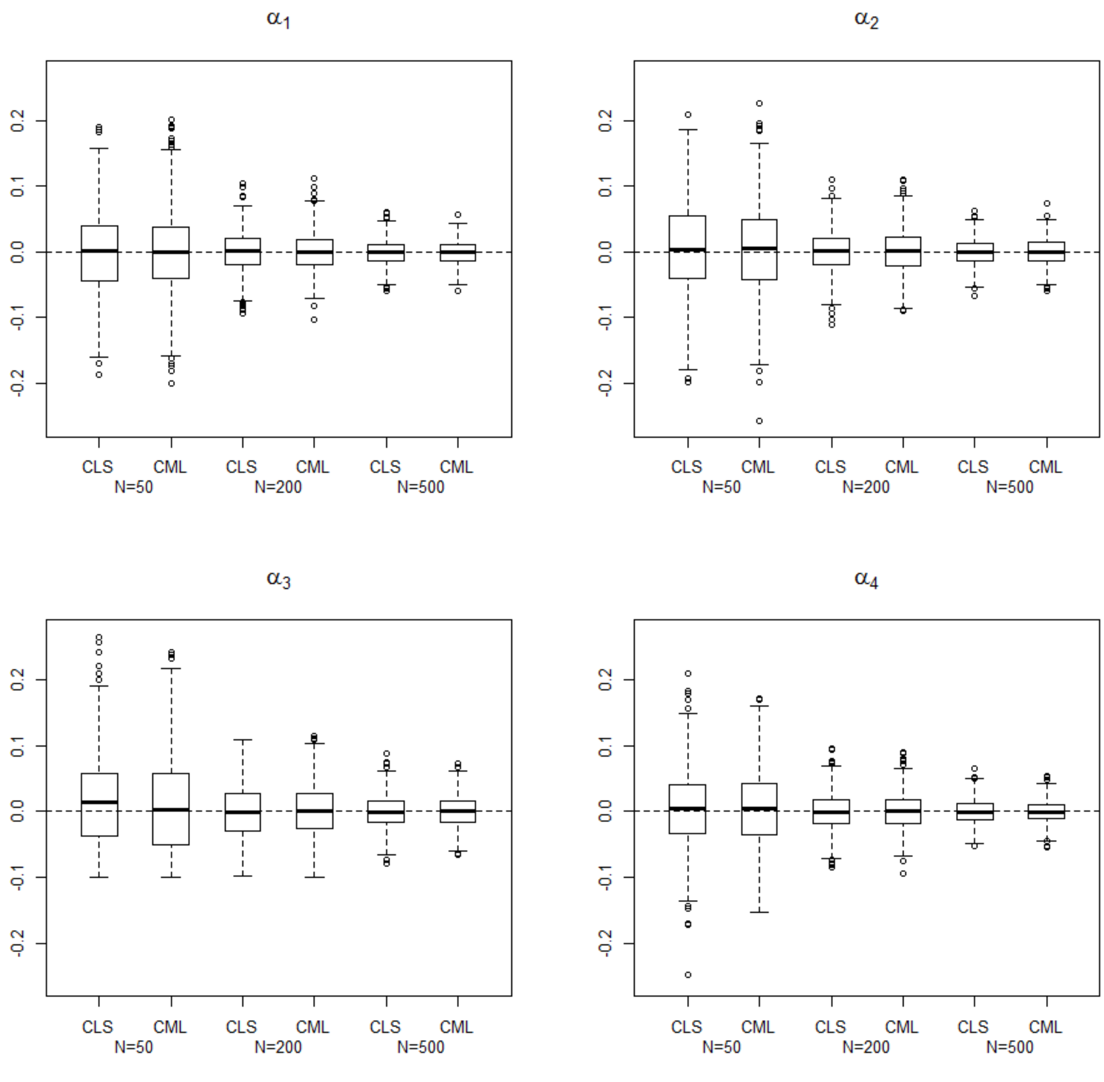

Figure 3.3: Boxplots for the biases of the CLS and CML estimates of parameter $\boldsymbol{\alpha}$ in Set 2A for $n = 4N = 200, 800, 2000$.

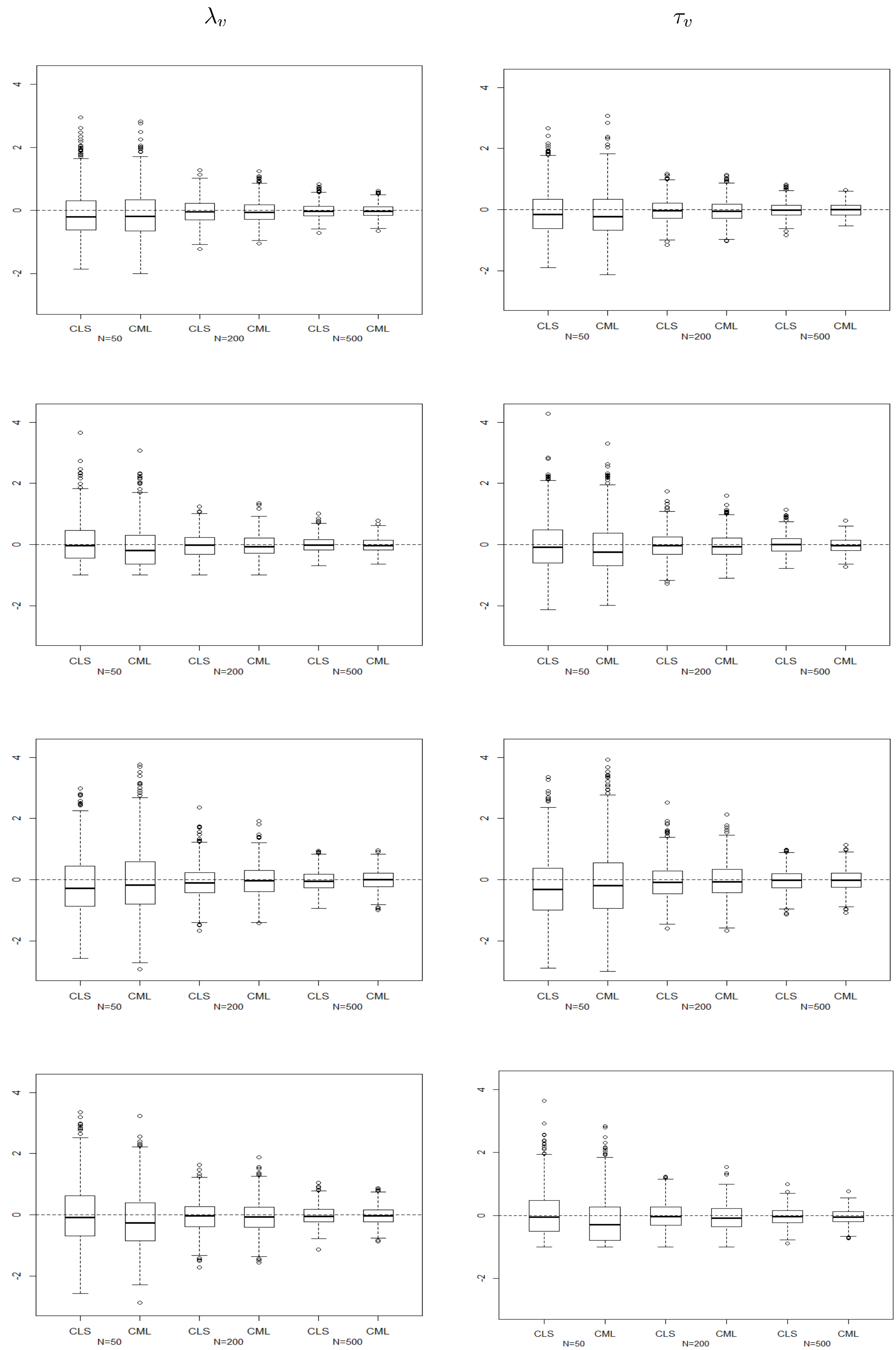

Figure 3.4: Boxplots for the biases of the CLS and CML estimates of parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\tau}$ in Set 2A for $n = 4N = 200, 800, 2000$.

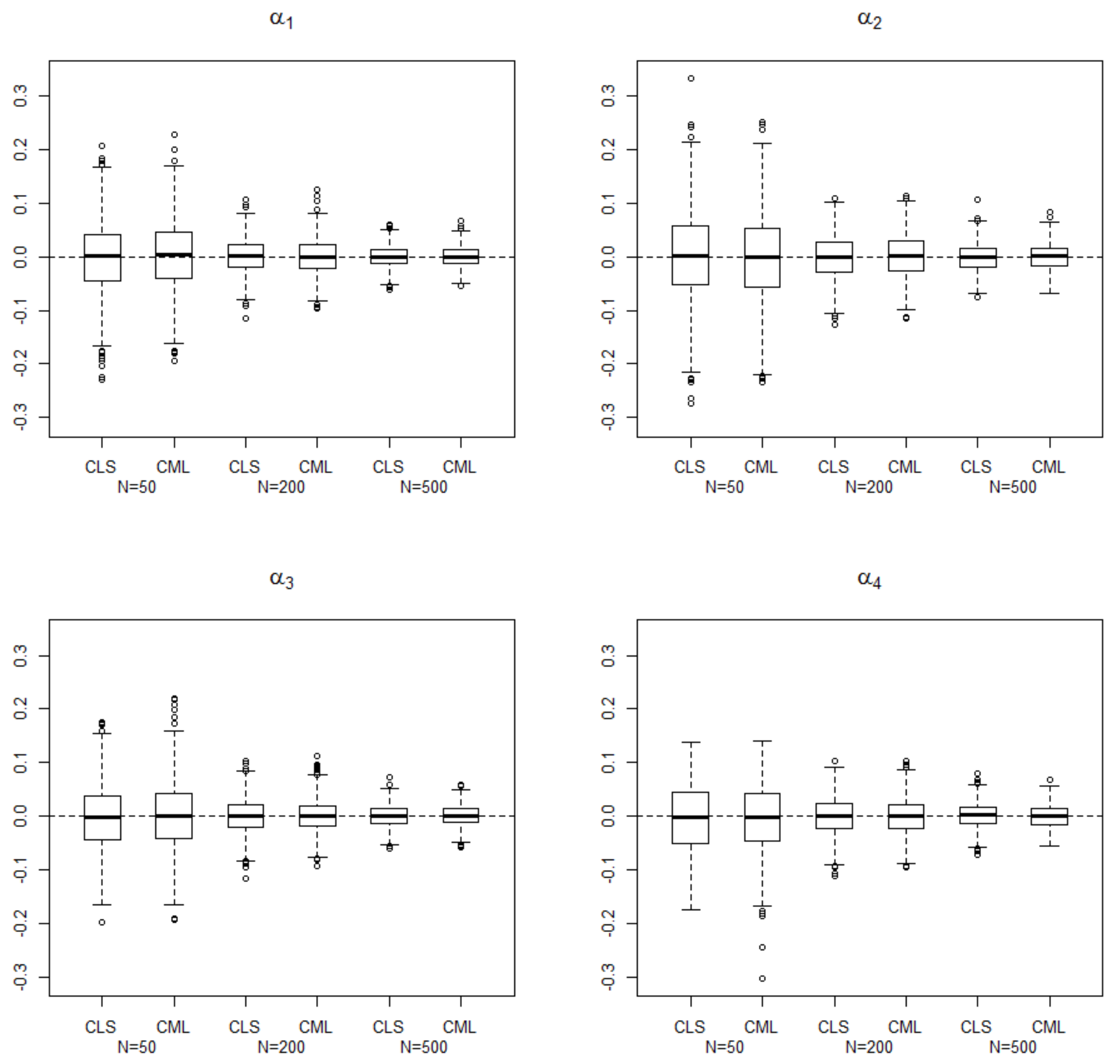

Figure 3.5: Boxplots for the biases of the CLS and CML estimates of parameter $\boldsymbol{\alpha}$ in Set 3A for $n = 4N = 200, 800, 2000$.

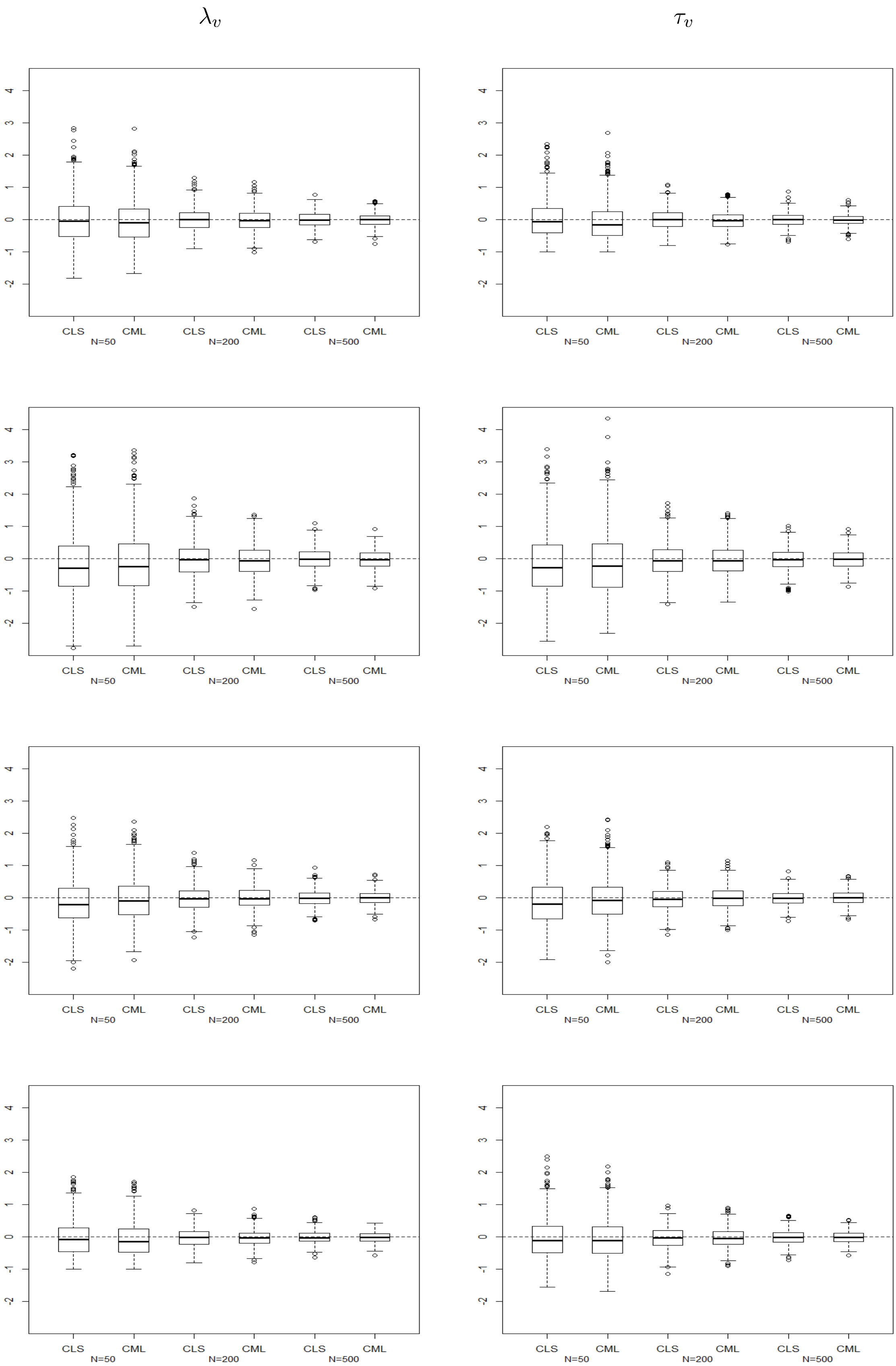

Figure 3.6: Boxplots for the biases of the CLS and CML estimates of parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\tau}$ in Set 3A for $n = 4N = 200, 800, 2000$.

3.4 The bivariate periodic model: BS-PINAR(1)

Bulla et al. (2016) introduced the class of bivariate signed INAR(1) processes, which is an extension of the SINAR(1) process of Kachour and Truquet (2011) to the bivariate case, based on the signed matrix thinning operator in (1.12). Therefore, motivated by Bulla et al. (2016), we generalize the S-PINAR(1) model with period s to the bivariate case. The definition and matrix representation of the bivariate model, denoted by BS-PINAR(1) with period s, is presented and some statistical properties of the model are derived. The assumption of a diagonal autoregressive matrix is made, which can be seen as a $\mathbb{Z}^2$-extension of the model presented in Pedeli and Karlis (2011), here with periodic structure. The correlation is achieved through their innovation processes. The discrete bivariate distribution on $\mathbb{Z}^2$ considered for the distribution of the innovations is the periodic bivariate Skellam distribution established previously in Definition 3.2. Parameter estimation of the unknown parameters is provided through the conditional maximum likelihood method.

3.4.1 Definition and basic properties

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

$$\mathbf{X}_t = F_t \odot \mathbf{X}_{t-1} + \mathbf{Z}_t,\ t \in \mathbb{Z}, \tag{3.42}$$

where $\mathbf{X}_t, \mathbf{X}_{t-1}$ and $\mathbf{Z}_t$ are random $2s$-vectors. The vector $\mathbf{X}_t = [X_{1,t}\ X_{2,t}]^T$ for $t = v + ns$, $v = 1, \dots, s$ and $n \in \mathbb{N}_0$ has components $X_{j,t} = [X_{j,1+ns}\ X_{j,2+ns}\ \dots\ X_{j,s+ns}]^T$ for $j = 1, 2$. The vector $\mathbf{Z}_t = [Z_{1,t}\ Z_{2,t}]^T$ represents a periodic sequence of independent random vectors. The model defined by recursion (3.42) will be referred to as BS-PINAR(1) for Bivariate Signed Periodic INteger-valued AutoRegressive model of order one with period $s \in \mathbb{N}$ and is based on the periodic signed matrix thinning operator in (3.11). The BS-PINAR(1) model admits the following matricial representation

$$\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} F_{1,t} & 0 \\ 0 & F_{2,t} \end{bmatrix} \odot \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix}, \ t \in \mathbb{Z}, \tag{3.43}$$

where $t = v + ns; \ v = 1, \ldots, s$ and $n \in \mathbb{N}_0$. Each component of the bivariate model in (3.43) admits the representation of a periodic univariate S-PINAR(1) process as in equation (3.28), i.e.,

$$X_{j,t} = F_{j,t} \odot X_{j,t-1} + Z_{j,t}, \ j = 1, 2 \tag{3.44}$$

with

$$F_{j,t} \odot X_{j,t-1} \stackrel{d}{=} sign(X_{j,t-1}) \sum_{i=1}^{|X_{j,t-1}|} U_{i,t}(\phi_{j,t}),$$

where $F_{j,t}$ represents the common distribution of the periodic sequence of i.i.d. counting sequences defined in (3.12) for any $j \in \{1, 2\}$. All counting sequences associated to the operators $F_{j,t}\odot$ are mutually independent. Furthermore, for each t, $Z_{j,t}$ is assumed to be independent of $X_{j,t-1}$ and $F_{j,t} \odot X_{j,t-1}$, $j = 1, 2$.

We assume the innovations series of the BS-PINAR(1) model in (3.43) jointly follow the periodic bivariate Skellam distribution established in Definition 3.2 with p.m.f. given by equation (3.19). For a fixed v ($v = 1, \ldots, s$) the first and second-order moments of $Z_{j,v+ns}$ ($j = 1, 2$) are defined in (3.22) and (3.26), respectively. The covariance between $Z_{j,v+ns}$ and $Z_{k,v+ns}$ ($j \neq k$) is given in (3.27).

Expressions for conditional mean and variance of the BS-PINAR(1) model with period s are derived. From equations (3.14) and (3.22)

$$E[X_{j,v+ns}|X_{j,v-1+ns}] = (2\alpha_{j,v} - 1)X_{j,v-1+ns} + \lambda_{j,v} - \tau_v \tag{3.45}$$

and from (3.15) and (3.26),

$$Var[X_{j,v+ns}|X_{j,v-1+ns}] = 2\alpha_{j,v}(1 - \alpha_{j,v})|X_{j,v-1+ns}| + \lambda_{j,v} + \tau_v. \tag{3.46}$$

3.4.2 Parameter estimation of the BS-PINAR(1) model

Consider a finite time series $(\mathbf{X}_1, \ldots, \mathbf{X}_{Ns})$ from the BS-PINAR(1) process with periodic bivariate Skellam-distributed innovations, where N represents the number of complete cycles. Without loss of generality it is assumed $\mathbf{X}_0 = \mathbf{x}_0$. The conditional maximum likelihood method is proposed to estimate the parameters of this bivariate model. The vector of unknown parameters $\boldsymbol{\theta}$ has $5s$ parameters, i.e.,

$$\boldsymbol{\theta} := (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\tau}) \tag{3.47}$$

with $\boldsymbol{\alpha}_j = (\alpha_{j,1}, \ldots, \alpha_{j,s})$, $\boldsymbol{\lambda}_j = (\lambda_{j,1}, \ldots, \lambda_{j,s})$ and $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_s)$, $j = 1, 2$. Hence, the conditional maximum likelihood (CML) estimator of the vector of the $(5s)$ unknown parameters in (3.47) is $\widehat{\boldsymbol{\theta}}_{CML} := (\widehat{\boldsymbol{\alpha}}_1^{CML}, \widehat{\boldsymbol{\alpha}}_2^{CML}, \widehat{\boldsymbol{\lambda}}_1^{CML}, \widehat{\boldsymbol{\lambda}}_2^{CML}, \widehat{\boldsymbol{\tau}}^{CML})$. The conditional log-likelihood function is given by

$$C(\boldsymbol{\theta}) = ln(L(\boldsymbol{\theta}|\mathbf{x})) = \sum_{n=0}^{N-1} \sum_{v=1}^{s} ln\left(p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})\right) \tag{3.48}$$

with conditional density

$$\begin{aligned} &p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns}) = P(\mathbf{X}_{v+ns} = \mathbf{x}_{v+ns}|\mathbf{X}_{v-1+ns} = \mathbf{x}_{v-1+ns}) = \\ &= P(X_{1,v+ns} = x_{1,v+ns}, X_{2,v+ns} = x_{2,v+ns}|X_{1,v-1+ns} = x_{1,v-1+ns}, X_{2,v-1+ns} = x_{2,v-1+ns}). \end{aligned} \tag{3.49}$$

For simplicity, let $(x_{1,v-1+ns}, x_{2,v-1+ns}) = (a, b)$ and $(x_{1,v+ns}, x_{2,v+ns}) = (c, d)$. Then the transition probability function in (3.49) takes the form

$$\begin{aligned} &p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns}) = \\ &= P(X_{1,v+ns} = c, X_{2,v+ns} = d|X_{1,v-1+ns} = a, X_{2,v-1+ns} = b) = \\ &= \sum_{k_1=-|a|}^{|a|} \sum_{k_2=-|b|}^{|b|} P\left(sign(a)\sum_{i=1}^{|a|} U_{i,t}(\phi_{1,t}) = k_1\right) P\left(sign(b)\sum_{i=1}^{|b|} U_{i,t}(\phi_{2,t}) = k_2\right) \times \end{aligned}$$

$$\times \ P(Z_{1,v+ns} = c - k_1, Z_{2,v+ns} = d - k_2) =$$

$$= \sum_{k_1=-|a|}^{|a|} \sum_{k_2=-|b|}^{|b|} P\left(R_t^{(|a|)}(\phi_{1,t}) = |a| + sign(a) \cdot k_1\right) P\left(R_t^{(|b|)}(\phi_{2,t}) = |b| + sign(b) \cdot k_2\right) \times$$

$$\times \ P(Z_{1,v+ns} = c - k_1, Z_{2,v+ns} = d - k_2) =$$

$$= \sum_{k_1=-|a|}^{|a|} \sum_{k_2=-|b|}^{|b|} \left\{ C^{2|a|}_{|a|+sign(a)\cdot k_1} \alpha_{1,v}^{|a|+sign(a)\cdot k_1} (1-\alpha_{1,v})^{|a|-sign(a)\cdot k_1} \times \right.$$
$$\times \ C^{2|b|}_{|b|+sign(b)\cdot k_2} \alpha_{2,v}^{|b|+sign(b)\cdot k_2} (1-\alpha_{2,v})^{|b|-sign(b)\cdot k_2} \times$$
$$\left. \times \ e^{-(\lambda_{1,v}+\lambda_{2,v}+\tau_v)} \lambda_{1,v}^{c-k_1} \lambda_{2,v}^{d-k_2} \sum_{i=max(0,-(c-k_1),-(d-k_2)}^{\infty} \frac{(\lambda_{1,v}\lambda_{2,v}\tau_v)^i}{i!(i+c-k_1)!(i+d-k_2)!} \right\}. \tag{3.50}$$

Differentiating the conditional log-likelihood function in (3.48) with respect to the 5s parameters, the system of first-order partial derivatives follows

$$\begin{cases} \dfrac{\partial C(\boldsymbol{\theta})}{\partial \alpha_{j,v}} = 0, \ \ j = 1, 2 \\ \\ \dfrac{\partial C(\boldsymbol{\theta})}{\partial \lambda_{j,v}} = 0, \ \ j = 1, 2; \ \ v = 1, \ldots, s, \\ \\ \dfrac{\partial C(\boldsymbol{\theta})}{\partial \tau_v} = 0 \end{cases}$$

i.e.,

$$
\begin{cases}
\prod_{n=0}^{N-1} \dfrac{2|x_{1,v-1+ns}|}{1-\alpha_{1,v}} \left(\dfrac{p_v(\mathbf{x}_{v+ns}-(1,0)|\mathbf{x}_{v-1+ns}-(1,0))}{p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})} - 1 \right) = 0 \\[2ex]
\prod_{n=0}^{N-1} \dfrac{2|x_{2,v-1+ns}|}{1-\alpha_{2,v}} \left(\dfrac{p_v(\mathbf{x}_{v+ns}-(0,1)|\mathbf{x}_{v-1+ns}-(0,1))}{p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})} - 1 \right) = 0 \\[2ex]
\prod_{n=0}^{N-1} \dfrac{p_v(\mathbf{x}_{v+ns}-(1,0)|\mathbf{x}_{v-1+ns})}{p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})} = N \\[2ex]
\prod_{n=0}^{N-1} \dfrac{p_v(\mathbf{x}_{v+ns}-(0,1)|\mathbf{x}_{v-1+ns})}{p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})} = N \\[2ex]
\prod_{n=0}^{N-1} \dfrac{p_v(\mathbf{x}_{v+ns}+(1,1)|\mathbf{x}_{v-1+ns})}{p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})} = N
\end{cases}
$$

for $v = 1, \dots, s$. First-order partial derivatives are omitted, however they are calculated in the similar way as in the univariate case (see Appendix C.1). Numerical maximization is straightforward with standard statistical packages in `R`.

Chapter 4

Conclusions and future challenges

The aim of this thesis is to provide contributions to the analysis of count time series with periodic structure. The main focus is on the definition and study of time series for count data with periodic time-varying parameters and periodic sequences of innovations. For this purpose, we focused on a particular type of processes for count time series, namely the integer–valued autoregressive (`INAR`) process of order one.

In Chapter 2, we introduced the periodic multivariate integer-valued process of order one (`PMINAR`(1) for short) with period s based on the matrix-binomial thinning operator. Apart from the general specification of the periodic multivariate process, the probabilistic and also the statistical properties of the model were studied in detail. Furthermore, the constraint of diagonality of the matrix of autocorrelation parameters was considered. Thus, the correlation between the innovation series of the periodic multivariate process was the only source of cross–correlation. A specific parametric case that arises under the assumption of a multivariate negative binomial distribution for the innovations of the process was assumed. The former specification of the `PMINAR`(1) process has the useful property that it can effectively account for overdispersion (variance exceeds mean). Deviations from the equidispersed settings often occur in real-life events. Concerning parameter estimation of the `PMINAR`(1) process, three methods were proposed, namely, Yule–Walker, conditional maximum likelihood and

composite likelihood. The computational complexity of the maximum likelihood approach augments with dimensional increase. To overcome the computational difficulties arising from that method, composite likelihood-based approach was suggested. The loss of efficiency due to the replacement of the full likelihood with a pseudo–likelihood was investigated. Hence, the performance of the proposed method and other competitors methods of estimation was compared through a simulation study. Although very demanding computationally, the conditional maximum likelihood method proved to outperform the other methods, thus the differences to the composite likelihood method were small. The composite likelihood method revealed to be computationally more convenient and impressively less time–consuming than the maximum likelihood method. After addressing one-step ahead forecasts, the proposed multivariate model with periodic structure and multivariate negative binomial distribution for the innovations series was applied to a real data set related with the analysis of fire activity. This application was made to a particular trivariate real data series regarding the number of monthly fires (period $s = 12$) in three counties in Portugal, namely Aveiro, Coimbra and Faro, during 30 years $(1981 - 2010)$. Additionally, the composite likelihood approach seemed satisfactory although some loss of efficiency was noticed but considered acceptable.
One topic for future work regarding the specification of a PMINAR(1) process could be removing the constraint of diagonality of the matrix of autocorrelation parameters. However, similar to what happens with conventional PAR models, PMINAR models can have an extremely large number of parameters increased with period s. The development of procedures for dimensionality reduction continues to be an interesting subject to be studied in this context. A common feature in real data applications is times series exhibiting overdispersion, therefore other distributions for the innovations series might also be of interest.

In Chapter 3, our attention was turned to periodic INAR(1) models based on a different type of thinning operator, the signed thinning operator, adapted accordingly to the periodic case. These models can handle integer-valued time series which allow for negative integer-valued and negative correlated count data unlike the integer-valued time series models in Chapter 2. Those models were only appropriate for non-negative integer-valued time series and could only deal with positive autocorrelations. Pursing our goal, two first-order `INAR` (univariate and bivariate) models with periodic structure were introduced, allowing for positive and negative counts, `S-PINAR`(1) and `BS-PINAR`(1), respectively. Basic probabilistic and also statistical properties of the periodic models were provided. A drawback of the signed thinning operator was the fact that the distributive property did not hold. This enabled us from writing the periodic process recursively as in Chapter 2 and therefore, obtaining the cycle-stationary distribution. This issue, however, is worth further exploration.

Particular emphasis was given to innovations modeled by univariate and bivariate Skellam distributions defined on the set of integers, respectively. The interest in the Skellam distribution or Poisson difference distribution has been recast. There are few discrete distributions defined in $\mathbb{Z}$. On the other hand, bivariate Skellam distribution is quite recent and appealing for models with innovations series defined in the $\mathbb{Z}^2$ context. To study the performance of the conditional least squares and conditional maximum likelihood estimators, a simulation study was conducted for the `S-PINAR`(1) model with period s. A modification of the traditional conditional least squares method was made through a two step procedure in order to provide estimators for all parameters involved in the periodic univariate model. The proposed estimation methods were compared through an extended simulation experiment contemplating six different combinations of the parameters. For each set of parameters and for each sample size, 1000 independent replicates were simulated from the `S-PINAR`(1) model. Numerical results from the simulation study suggested that the proposed model is suitable for practical use. However, this is an issue we would like to explore in future work considering the application of the univariate model to real data time series exhibiting periodic structure.

Regarding periodic models based on the signed thinning operator, an important subject to investigate in further research, is the forecasting distribution of these models.

Appendix A

Auxiliary results of Chapter 1

Univariate Skellam distribution

In Definition 1.3, if $Z \sim Skellam(\theta_1, \theta_2)$ then the probability mass function (p.m.f.) is

$$P(Z = z) = e^{-(\theta_1+\theta_2)} \left(\frac{\theta_1}{\theta_2}\right)^{z/2} I_{|z|} \left(2\sqrt{\theta_1\theta_2}\right), \;\; z \in \mathbb{Z},$$

where $I_r(x)$ is the modified Bessel function of the first kind of order r defined by

$$I_r(x) = \left(\frac{x}{2}\right)^r \sum_{i=0}^{\infty} \frac{\left(\frac{x^2}{4}\right)^i}{i!\Gamma(r+i+1)}. \tag{A.1}$$

The mean and the variance are, respectively, $E[Z] = \theta_1 - \theta_2$ and $Var[Z] = \theta_1 + \theta_2$. Clearly, the variance exceeds the mean, i.e., $Var[Z] \geq |E[Z]|$. The distribution is symmetric only when $\theta_1 = \theta_2$ (case discussed by Irwin (1937)).

The probability generating function is given by

$$G_Z(s) = E\left[s^Z\right] = e^{-(\theta_1+\theta_2)+\theta_1 s+\theta_2/s}.$$

A new representation of the Skellam (Poisson difference) distribution by replacing the Bessel function in (A.1) was established in Alzaid and Omair (2010). Hence, an alternative formula for the p.m.f. of the Skellam distribution is

$$P(Z = z) = e^{-(\theta_1+\theta_2)}\theta_1^z \sum_{i=0}^{\infty} \frac{(\theta_1\theta_2)^i}{i!(i+z)!}, \ z \in \mathbb{Z}. \tag{A.2}$$

For large values of the sum $\theta_1 + \theta_2$, the distribution can be sufficiently approximated by the normal distribution. If $\theta_2 = 0$, the distribution tends to a Poisson distribution and if $\theta_1 = 0$, tends to the negative of a Poisson distribution. The Skellam distribution is unimodal. The sum and the difference of two Skellam r.v.'s also follow the same distribution.

Note that Skellam distribution is not necessarily the distribution of the difference of two uncorrelated Poisson r.v.'s (Karlis and Ntzoufras, 2006). This implies that we can derive the Skellam distribution as the difference of other distributions as well. Further details in Alzaid and Omair (2010).

Appendix B

Auxiliary results of Chapter 2

B.1 Proof of equation (2.32)

Let $t = 1 + ns$, then

$$
\begin{aligned}
& E[X_{j,1+ns}] = \\
& = \lambda_{j,1} + \frac{d_{(-1)}}{d}\alpha_{j,1}\left(\lambda_{j,1}\prod_{k=0}^{s-2}\alpha_{j,s-k} + \lambda_{j,2}\prod_{k=0}^{s-3}\alpha_{j,s-k} + \cdots + \lambda_{j,s-1}\alpha_{j,s} + \lambda_{j,s}\right) = \\
& = \lambda_{j,1} + \frac{\alpha_{j,1}}{1 - \prod_{k=0}^{s-1}\alpha_{j,s-k}}\left(\lambda_{j,1}\prod_{k=0}^{s-2}\alpha_{j,s-k} + \lambda_{j,2}\prod_{k=0}^{s-3}\alpha_{j,s-k} + \cdots + \lambda_{j,s-1}\alpha_{j,s} + \lambda_{j,s}\right) = \\
& = \frac{\lambda_{j,1}\left(1 - \prod_{k=0}^{s-1}\alpha_{j,s-k}\right) + \alpha_{j,1}\left(\lambda_{j,1}\prod_{k=0}^{s-2}\alpha_{j,s-k} + \lambda_{j,2}\prod_{k=0}^{s-3}\alpha_{j,s-k} + \cdots + \lambda_{j,s-1}\alpha_{j,s} + \lambda_{j,s}\right)}{1 - \prod_{k=0}^{s-1}\alpha_{j,s-k}} = \\
& = \frac{\lambda_{j,1} - \lambda_{j,1}\prod_{k=0}^{s-1}\alpha_{j,s-k} + \lambda_{j,1}\prod_{k=0}^{s-1}\alpha_{j,s-k} + \alpha_{j,1}\left(\lambda_{j,2}\prod_{k=0}^{s-3}\alpha_{j,s-k} + \cdots + \lambda_{j,s-1}\alpha_{j,s} + \lambda_{j,s}\right)}{1 - \prod_{k=0}^{s-1}\alpha_{j,s-k}} = \\
& = \frac{\lambda_{j,1} + \alpha_{j,1}\left(\lambda_{j,2}\prod_{k=0}^{s-3}\alpha_{j,s-k} + \cdots + \lambda_{j,s-1}\alpha_{j,s} + \lambda_{j,s}\right)}{1 - \prod_{k=0}^{s-1}\alpha_{j,s-k}}.
\end{aligned}
$$

Attending to relation (2.31), it follows that

$$\lambda_{j,s-1}\alpha_{j,s}+\cdots+\lambda_{j,2}\prod_{k=0}^{s-3}\alpha_{j,s-k}=\sum_{i=1}^{s-2}\left(\lambda_{j,s-i}\prod_{k=0}^{i-1}\alpha_{j,s-k}\right)=\sum_{i=1}^{s-2}\lambda_{j,s-i}\varphi_{s,i}^{(1)}$$

thus

$$E[X_{j,1+ns}]=\frac{\lambda_{j,1}+\alpha_{j,1}\left(\lambda_{j,s}+\lambda_{j,s-1}\alpha_{j,s}+\cdots+\lambda_{j,2}\prod_{k=0}^{s-3}\alpha_{j,s-k}\right)}{1-\prod_{k=0}^{s-1}\alpha_{j,s-k}}=$$
$$=\frac{\lambda_{j,1}+\alpha_{j,1}\left(\lambda_{j,s}+\sum_{i=1}^{s-2}\varphi_{s,i}^{(j)}\lambda_{j,s-i}\right)}{1-\prod_{k=0}^{s-1}\alpha_{j,s-k}}=$$
$$=\frac{\sum_{k=0}^{1-1}\varphi_{1,k}^{(j)}\lambda_{j,1-k}+\varphi_{1,1}^{(j)}\sum_{i=0}^{s-2}\varphi_{s,i}^{(j)}\lambda_{j,s-i}}{1-\varphi_{s,s}^{(j)}}.$$

Let $t=2+ns$, then

$$E[X_{j,2+ns}]=$$
$$=\lambda_{j,1}\alpha_{j,2}+\lambda_{j,2}+\frac{d_{(-j)}}{d}\prod_{k=0}^{2-1}\alpha_{j,2-k}\left(\lambda_{j,1}\prod_{k=0}^{s-2}\alpha_{j,s-k}+\lambda_{j,2}\prod_{k=0}^{s-3}\alpha_{j,s-k}+\cdots+\lambda_{j,s-1}\alpha_{j,s}+\lambda_{j,s}\right)$$
$$=\frac{(\lambda_{j,1}\alpha_{j,2}+\lambda_{j,2})\left(1-\prod_{k=0}^{s-1}\alpha_{j,s-k}\right)+\alpha_{j,2}\alpha_{j,1}\left(\lambda_{j,1}\prod_{k=0}^{s-2}\alpha_{j,s-k}+\lambda_{j,2}\prod_{k=0}^{s-3}\alpha_{j,s-k}+\cdots+\lambda_{j,s}\right)}{1-\prod_{k=0}^{s-1}\alpha_{j,s-k}}$$
$$=\frac{\lambda_{j,1}\alpha_{j,2}+\lambda_{j,2}-\lambda_{j,1}\alpha_{j,2}\prod_{k=0}^{s-1}\alpha_{j,s-k}-\lambda_{j,2}\prod_{k=0}^{s-1}\alpha_{j,s-k}}{1-\prod_{k=0}^{s-1}\alpha_{j,s-k}}+$$
$$+\frac{\alpha_{j,2}\alpha_{j,1}\left(\lambda_{j,1}\prod_{k=0}^{s-2}\alpha_{j,s-k}+\lambda_{j,2}\prod_{k=0}^{s-3}\alpha_{j,s-k}+\cdots+\lambda_{j,s-1}\alpha_{j,s}+\lambda_{j,s}\right)}{1-\prod_{k=0}^{s-1}\alpha_{j,s-k}}=$$

$$= \frac{\lambda_{j,1}\alpha_{j,2} + \lambda_{j,2} - \lambda_{j,1}\alpha_{j,2}\prod_{k=0}^{s-1}\alpha_{j,s-k} - \lambda_{j,2}\prod_{k=0}^{s-1}\alpha_{j,s-k} + \lambda_{j,1}\alpha_{j,2}\prod_{k=0}^{s-1}\alpha_{j,s-k} + \lambda_{j,2}\prod_{k=0}^{s-1}\alpha_{j,s-k}}{1-\prod_{k=0}^{s-1}\alpha_{j,s-k}} +$$

$$+ \frac{\alpha_{j,2}\alpha_{j,1}\left(\lambda_{j,3}\prod_{k=0}^{s-4}\alpha_{j,s-k} + \cdots + \lambda_{j,s-1}\alpha_{j,s} + \lambda_{j,s}\right)}{1-\prod_{k=0}^{s-1}\alpha_{j,s-k}} =$$

$$= \frac{\lambda_{j,1}\alpha_{j,2} + \lambda_{j,2} + \alpha_{j,2}\alpha_{j,1}\left(\lambda_{j,3}\prod_{k=0}^{s-4}\alpha_{j,s-k} + \cdots + \lambda_{j,s-1}\alpha_{j,s} + \lambda_{j,s}\right)}{1-\prod_{k=0}^{s-1}\alpha_{j,s-k}},$$

yielding

$$E[X_{j,2+ns}] = \frac{\lambda_{j,2} + \alpha_{j,2}\lambda_{j,1} + \alpha_{j,1}\alpha_{j,2}\sum_{i=0}^{s-3}\varphi_{s,i}^{(j)}\lambda_{j,s-i}}{1-\prod_{k=0}^{s-1}\alpha_{j,s-k}} =$$

$$= \frac{\sum_{k=0}^{2-1}\varphi_{2,k}^{(j)}\lambda_{j,2-k} + \varphi_{2,2}^{(j)}\sum_{i=0}^{s-3}\varphi_{s,i}^{(j)}\lambda_{j,s-i}}{1-\varphi_{s,s}^{(j)}}.$$

Let $t = 3 + ns$, then

$$E[X_{j,3+ns}] = \lambda_{j,1}\alpha_{j,3}\alpha_{j,2} + \lambda_{j,2}\alpha_{j,3} + \lambda_{j,3} + \frac{d_{(-j)}}{d}\prod_{k=0}^{3-1}\alpha_{j,3-k} \times$$

$$\times \left(\lambda_{j,1}\prod_{k=0}^{s-2}\alpha_{j,s-k} + \lambda_{j,2}\prod_{k=0}^{s-3}\alpha_{j,s-k} + \cdots + \lambda_{j,s-1}\alpha_{j,s} + \lambda_{j,s}\right) =$$

$$= \frac{\sum_{k=0}^{3-1}\varphi_{3,k}^{(j)}\lambda_{j,3-k} + \varphi_{3,3}^{(j)}\sum_{i=0}^{s-4}\varphi_{s,i}^{(j)}\lambda_{j,s-i}}{1-\varphi_{s,s}^{(j)}},$$

and likewise until $t = s + ns$:

$$E[X_{j,s+ns}] = \frac{d_{(-j)}}{d}\left(\lambda_{j,1}\prod_{k=0}^{s-2}\alpha_{j,s-k} + \lambda_{j,2}\prod_{k=0}^{s-3}\alpha_{j,s-k} + \cdots + \lambda_{j,s-1}\alpha_{j,s} + \lambda_{j,s}\right) =$$

$$= \frac{\lambda_{j,1}\prod_{k=0}^{s-2}\alpha_{j,s-k} + \lambda_{j,2}\prod_{k=0}^{s-3}\alpha_{j,s-k} + \cdots + \lambda_{j,s-1}\alpha_{j,s} + \lambda_{j,s}}{1 - \prod_{k=0}^{s-1}\alpha_{j,s-k}} =$$

$$= \frac{\lambda_{j,s} + \lambda_{j,s-1}\alpha_{j,s} + \cdots + \lambda_{j,2}\prod_{k=0}^{s-3}\alpha_{j,s-k} + \lambda_{j,1}\prod_{k=0}^{s-2}\alpha_{j,s-k}}{1 - \varphi_{s,s}^{(j)}} =$$

$$= \frac{\sum_{k=0}^{s-1}\varphi_{s,k}^{(j)}\lambda_{j,s-k} + \varphi_{s,s}^{(j)}\sum_{i=0}^{s-(s+1)}\varphi_{s,i}^{(j)}\lambda_{j,s-i}}{1 - \varphi_{s,s}^{(j)}}.$$

Hence, for $j = 1, \ldots, m$ and $v = 1, \ldots, s$:

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

with convention $\sum_{i=0}^{s-(s+1)}\varphi_{s,i}^{(j)}\lambda_{j,s-i} = 0$.

B.2 First-order partial derivatives of the transition probability function

For convenience, let $x_{j,v-1+ns} = a_j$ and $x_{j,v+ns} = b_j$ for $j = 1, \dots, m$, hence $\mathbf{a} = (a_1, \dots, a_m)$ and $\mathbf{b} = (b_1, \dots, b_m)$. The binomial distribution in (2.70) can be written as

$$f_j(r_j) = C_{r_j}^{a_j} \alpha_{j,v}^{r_j} (1 - \alpha_{j,v})^{a_j - r_j} \quad (v = 1, \dots, s).$$

Moreover, the transition probability function $p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns})$ in (2.86) takes the form

$$\begin{aligned} & p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns}) = p_v(\mathbf{b}|\mathbf{a}) = \\ & = \sum_{r_1=0}^{g_1} \sum_{r_2=0}^{g_2} \dots \sum_{r_m=0}^{g_m} \left(\prod_{j=1}^{m} f_j(r_j) \right) h(b_1 - r_1, b_2 - r_2, \dots, b_m - r_m). \end{aligned} \tag{B.1}$$

Recall the vector of unknown parameters in (2.89), i.e., $\eta_v = (\alpha_{1,v}, \dots, \alpha_{m,v}, \lambda_{1,v}, \dots, \lambda_{m,v}, \beta_v)$. For a fixed v $(v = 1, \dots, s)$, the first-order partial derivative of function $p_v(\mathbf{b}|\mathbf{a})$ in (B.1) with respect to parameter $\alpha_{1,v}$ is

$$\begin{aligned} & \frac{\partial}{\partial \alpha_{1,v}} p_v(\mathbf{b}|\mathbf{a}) = \\ & = \sum_{r_1=0}^{g_1} \sum_{r_2=0}^{g_2} \dots \sum_{r_m=0}^{g_m} C_{r_1}^{a_1} \left(r_1 \alpha_{1,v}^{r_1-1} (1 - \alpha_{1,v})^{a_1 - r_1} - (a_1 - r_1) \alpha_{1,v}^{r_1} (1 - \alpha_{1,v})^{a_1 - r_1 - 1} \right) \times \\ & \times f_2(r_2) \dots f_m(r_m) \; h(b_1 - r_1, b_2 - r_2, \dots, b_m - r_m) = \\ & = \sum_{r_1=0}^{g_1} \sum_{r_2=0}^{g_2} \dots \sum_{r_m=0}^{g_m} \left(\frac{r_1}{\alpha_{1,v}} - \frac{a_1 - r_1}{1 - \alpha_{1,v}} \right) \prod_{j=1}^{m} f_j(r_j) \; h(b_1 - r_1, b_2 - r_2, \dots, b_m - r_m) = \\ & = \sum_{r_1=1}^{g_1} \sum_{r_2=0}^{g_2} \dots \sum_{r_m=0}^{g_m} \left(\frac{r_1}{\alpha_{1,v}(1 - \alpha_{1,v})} - \frac{a_1}{1 - \alpha_{1,v}} \right) \prod_{j=1}^{m} f_j(r_j) h(b_1 - r_1, b_2 - r_2, \dots, b_m - r_m) = \\ & = \underbrace{\sum_{r_1=1}^{g_1} \sum_{r_2=0}^{g_2} \dots \sum_{r_m=0}^{g_m} \frac{r_1}{\alpha_{1,v}(1 - \alpha_{1,v})} \prod_{j=1}^{m} f_j(r_j) \; h(b_1 - r_1, b_2 - r_2, \dots, b_m - r_m)}_{\mathbf{I}} - \\ & - \frac{a_1}{1 - \alpha_{1,v}} \; p_v(\mathbf{b}|\mathbf{a}), \end{aligned}$$

where

$$\begin{aligned}
\mathbf{I} &= \sum_{r_1=1}^{g_1}\sum_{r_2=0}^{g_2}\dots\sum_{r_m=0}^{g_m} r_1 C_{r_1}^{a_1}\alpha_{1,v}^{r_1-1}(1-\alpha_{1,v})^{a_1-r_1-1} f_2(r_2)\dots f_m(r_m)\times \\
&\times h(b_1-r_1, b_2-r_2,\dots,b_m-r_m) = \\
&= \sum_{r_1=1}^{g_1}\sum_{r_2=0}^{g_2}\dots\sum_{r_m=0}^{g_m} \frac{a_1!}{(r_1-1)!(a_1-r_1)!}\ \alpha_{1,v}^{r_1-1}(1-\alpha_{1,v})^{a_1-r_1-1} f_2(r_2)\dots f_m(r_m)\times \\
&\times h(b_1-r_1, b_2-r_2,\dots,b_m-r_m) = \\
&= \sum_{r_1=1}^{g_1}\sum_{r_2=0}^{g_2}\dots\sum_{r_m=0}^{g_m} \frac{a_1(a_1-1)!}{(r_1-1)!(a_1-1-(r_1-1))!}\left(\alpha_{1,v}^{r_1-1}(1-\alpha_{1,v})^{a_1-1-(r_1-1)}\ + \right. \\
&\left. +\ \alpha_{1,v}^{r_1}(1-\alpha_{1,v})^{a_1-r_1-1}\right) f_2(r_2)\dots f_m(r_m)\ h(b_1-r_1, b_2-r_2,\dots,b_m-r_m).
\end{aligned}$$

Notice that

$$\begin{aligned}
\alpha_{1,v}^{r_1-1}(1-\alpha_{1,v})^{a_1-r_1-1} &= \alpha_{1,v}^{r_1}(1-\alpha_{1,v})^{a_1-r_1}\left(\alpha_{1,v}^{-1}+(1-\alpha_{1,v})^{-1}\right) = \\
&= \alpha_{1,v}^{r_1-1}(1-\alpha_{1,v})^{a_1-1-(r_1-1)} + \alpha_{1,v}^{r_1}(1-\alpha_{1,v})^{a_1-r_1-1}
\end{aligned}$$

because

$$\alpha_{1,v}^{-1}(1-\alpha_{1,v})^{-1} = \frac{1}{\alpha_{1,v}(1-\alpha_{1,v})} = \frac{1-\alpha_{1,v}+\alpha_{1,v}}{\alpha_{1,v}(1-\alpha_{1,v})} = \frac{1}{\alpha_{1,v}} + \frac{1}{1-\alpha_{1,v}} = \alpha_{1,v}^{-1} + (1-\alpha_{1,v})^{-1}.$$

Then

$$\begin{aligned}
\mathbf{I} &= \sum_{r_1=1}^{g_1}\sum_{r_2=0}^{g_2}\dots\sum_{r_m=0}^{g_m} a_1 C_{r_1-1}^{a_1-1}\left(\alpha_{1,v}^{r_1-1}(1-\alpha_{1,v})^{a_1-1-(r_1-1)}\ + \right. \\
&\left. +\ \alpha_{1,v}^{r_1-1+1}(1-\alpha_{1,v})^{a_1-1-(r_1-1)-1}\right) f_2(r_2)\dots f_m(r_m)\times \\
&\times h(b_1-1-(r_1-1), b_2-r_2,\dots,b_m-r_m) = \\
&= \sum_{i=0}^{g_1}\sum_{r_2=0}^{g_2}\dots\sum_{r_m=0}^{g_m} a_1 C_{i}^{a_1-1}\left(\alpha_{1,v}^{i}(1-\alpha_{1,v})^{a_1-1-i}\ + \alpha_{1,v}^{i+1}(1-\alpha_{1,v})^{a_1-1-i-1}\right)\times \\
&\times f_2(r_2)\dots f_m(r_m)\ h(b_1-1-i, b_2-r_2,\dots,b_m-r_m) =
\end{aligned}$$

$$= \sum_{r_1=0}^{g_1} \sum_{r_2=0}^{g_2} \cdots \sum_{r_m=0}^{g_m} a_1 C_{r_1}^{a_1-1} \left(\alpha_{1,v}^{r_1} (1-\alpha_{1,v})^{a_1-1-r_1} \ + \ \alpha_{1,v}^{r_1+1} (1-\alpha_{1,v})^{a_1-1-r_1-1} \right) \times$$
$$\times f_2(r_2) \dots f_m(r_m) \ h(b_1 - 1 - r_1, b_2 - r_2, \dots, b_m - r_m) =$$
$$= \sum_{r_1=0}^{g_1} \sum_{r_2=0}^{g_2} \cdots \sum_{r_m=0}^{g_m} \frac{a_1}{1-\alpha_{1,v}} C_{r_1}^{a_1-1} \alpha_{1,v}^{r_1} (1-\alpha_{1,v})^{a_1-1-r_1} \ f_2(r_2) \dots f_m(r_m) \times$$
$$\times h(b_1 - 1 - r_1, b_2 - r_2, \dots, b_m - r_m) =$$
$$= \frac{a_1}{1-\alpha_{1,v}} \ p_v(\mathbf{b} - (1,0,\dots,0) | \mathbf{a} - (1,0,\dots,0))$$

leading to

$$\frac{\partial}{\partial \alpha_{1,v}} p_v(\mathbf{b}|\mathbf{a}) = \frac{a_1}{1-\alpha_{1,v}} \ \left[p_v(\mathbf{b} - (1,0,\dots,0) | \mathbf{a} - (1,0,\dots,0)) - p_v(\mathbf{b}|\mathbf{a}) \right].$$

Replacing $\mathbf{a} = (a_1, \dots, a_m)$ with $a_j = x_{j,v-1+ns}$ and $\mathbf{b} = (b_1, \dots, b_m)$ with $b_j = x_{j,v+ns}$ for $j = 1, \dots, m$, we obtain

$$\frac{\partial}{\partial \alpha_{1,v}} p_v(\mathbf{x}_{j,v+ns} | \mathbf{x}_{j,v-1+ns}) =$$
$$= \frac{x_{1,v-1+ns}}{1-\alpha_{1,v}} \ \left[p_v(\mathbf{x}_{j,v+ns} - (1,0,\dots,0) | \mathbf{x}_{j,v-1+ns} - (1,0,\dots,0)) - p_v(\mathbf{x}_{j,v+ns} | \mathbf{x}_{j,v-1+ns}) \right],$$

and in a similar way regarding the other partial derivatives with respect to $\alpha_{2,v}, \dots, \alpha_{m,v}$:

$$\frac{\partial}{\partial \alpha_{2,v}} p_v(\mathbf{x}_{j,v+ns} | \mathbf{x}_{j,v-1+ns}) =$$
$$= \frac{x_{2,v-1+ns}}{1-\alpha_{2,v}} \ \left[p_v(\mathbf{x}_{j,v+ns} - (0,1,\dots,0) | \mathbf{x}_{j,v-1+ns} - (0,1,\dots,0)) - p_v(\mathbf{x}_{j,v+ns} | \mathbf{x}_{j,v-1+ns}) \right],$$
$$\vdots$$
$$\frac{\partial}{\partial \alpha_{m,v}} p_v(\mathbf{x}_{j,v+ns} | \mathbf{x}_{j,v-1+ns}) =$$
$$= \frac{x_{m,v-1+ns}}{1-\alpha_{m,v}} \ \left[p_v(\mathbf{x}_{j,v+ns} - (0,0,\dots,1) | \mathbf{x}_{j,v-1+ns} - (0,0,\dots,1)) - p_v(\mathbf{x}_{j,v+ns} | \mathbf{x}_{j,v-1+ns}) \right]$$

for $v = 1, \dots, s$. The first-order partial derivatives of function $p_v(\mathbf{b}|\mathbf{a})$ concerning the remaining parameters $(\lambda_{1,v}, \dots, \lambda_{m,v}, \beta_v)$ integrating vector (2.89) follow shortly. Those parameters are from the MVNB distribution $h(z_1, z_2, \dots, z_m)$ established in (2.76). Taking advantage of the well-known property $e^{ln[h(z_1,z_2,\ldots,z_m)]}$, the log-function $ln[h(z_1,z_2,\ldots,z_m)]$ takes the form

$$\begin{aligned} & ln[h(z_1,z_2,\ldots,z_m)] = \\ & = ln\ \Gamma\left(\beta_v^{-1}+\sum_{j=1}^{m} z_j\right) - ln\ \Gamma(\beta_v^{-1}) + \sum_{j=1}^{m} z_j\ ln(\lambda_{j,v}) - \sum_{j=1}^{m} ln(z_j!) - \\ & - \beta_v^{-1} ln(\beta_v) - \left(\beta_v^{-1}+\sum_{j=1}^{m} z_j\right) ln\left(\beta_v^{-1}+\sum_{j=1}^{m}\lambda_{j,v}\right). \end{aligned} \tag{B.2}$$

Therefore, the first-order partial derivatives of function $h(z_1,z_2,\ldots,z_m)$ with respect to the parameters $(\lambda_{1,v},\ldots,\lambda_{m,v},\beta_v)$ can be obtained through

$$\begin{aligned} & \frac{\partial}{\partial\lambda_{j,v}} h(z_1,z_2,\ldots,z_m) = \frac{\partial}{\partial\lambda_{j,v}} e^{ln[h(z_1,z_2,\ldots,z_m)]} = \\ & = e^{ln[h(z_1,z_2,\ldots,z_m)]}\frac{\partial}{\partial\lambda_{j,v}} ln[h(z_1,z_2,\ldots,z_m)] = \\ & = h(z_1,z_2,\ldots,z_m)\frac{\partial}{\partial\lambda_{j,v}} ln[h(z_1,z_2,\ldots,z_m)] \end{aligned}$$

with $v = 1,\ldots,s$ and $j = 1,\ldots,m$, and likewise for the dispersion parameter, β_v. On differentiating the function $ln[h(z_1,z_2,\ldots,z_m)]$ in (B.2), the partial derivatives are

$$\frac{\partial}{\partial\lambda_{j,v}} ln[h(z_1,z_2,\ldots,z_m)] = \frac{z_j}{\lambda_{j,v}} - \left(\beta_v^{-1}+\sum_{j=1}^{m} z_j\right)\left(\beta_v^{-1}+\sum_{j=1}^{m}\lambda_{j,v}\right)^{-1}$$

and

$$\begin{aligned} & \frac{\partial}{\partial\beta_v} ln[h(z_1,z_2,\ldots,z_m)] = -\beta_v^{-2}\psi\left(\beta_v^{-1}+\sum_{j=1}^{m} z_j\right) + \beta_v^{-2}\psi\left(\beta_v^{-1}\right) + \\ & + \beta_v^{-2} ln(\beta_v) - \beta_v^{-2} + \beta_v^{-2} ln\left(\beta_v^{-1}+\sum_{j=1}^{m}\lambda_{j,v}\right) + \frac{\beta_v^{-2}}{\beta_v^{-1}+\sum_{j=1}^{m}\lambda_{j,v}} = \\ & = \beta_v^{-2}\psi\left(\beta_v^{-1}\right) - \beta_v^{-2}\psi\left(\beta_v^{-1}+\sum_{j=1}^{m} z_j\right) + \end{aligned}$$

$$+ \beta_v^{-2} \left[ln(\beta_v) - 1 + ln \left(\frac{1 + \beta_v \sum_{j=1}^m \lambda_{j,v}}{\beta_v} \right) + \frac{1}{\beta_v^{-1} + \sum_{j=1}^m \lambda_{j,v}} \right] =$$

$$= \beta_v^{-2} \left[\psi\left(\beta_v^{-1}\right) - \psi\left(\beta_v^{-1} + \sum_{j=1}^m z_j\right) + ln\left(1 + \beta_v \sum_{j=1}^m \lambda_{j,v}\right) + \frac{1}{\beta_v^{-1} + \sum\limits_{j=1}^m \lambda_{j,v}} - 1 \right]$$

i.e.,

$$\frac{\partial}{\partial \lambda_{j,v}} h(z_1, \ldots, z_m) = \left[\frac{z_j}{\lambda_{j,v}} - \left(\beta_v^{-1} + \sum_{j=1}^m z_j\right)\left(\beta_v^{-1} + \sum_{j=1}^m \lambda_{j,v}\right)^{-1} \right] h(z_1, \ldots, z_m) \quad \text{(B.3)}$$

and

$$\frac{\partial}{\partial \beta_v} h(z_1, \ldots, z_m) = \beta_v^{-2} \left[\psi\left(\beta_v^{-1}\right) - \psi\left(\beta_v^{-1} + \sum_{j=1}^m z_j\right) + ln\left(1 + \beta_v \sum_{j=1}^m \lambda_{j,v}\right) + \right.$$
$$\left. + \frac{1}{\beta_v^{-1} + \sum_{j=1}^m \lambda_{j,v}} - 1 \right] h(z_1, \ldots, z_m). \quad \text{(B.4)}$$

Furthermore,

$$\frac{\partial}{\partial \lambda_{j,v}} p_v(\mathbf{x}_{j,v+ns} | \mathbf{x}_{j,v-1+ns}) = \sum_{r_1=0}^{g_1} \sum_{r_2=0}^{g_2} \cdots \sum_{r_m=0}^{g_m} \left(\prod_{j=1}^m f_j(r_j) \right) \frac{\partial}{\partial \lambda_{j,v}} h(z_1, z_2, \ldots, z_m)$$

and

$$\frac{\partial}{\partial \beta_v} p_v(\mathbf{x}_{j,v+ns} | \mathbf{x}_{j,v-1+ns}) = \sum_{r_1=0}^{g_1} \sum_{r_2=0}^{g_2} \cdots \sum_{r_m=0}^{g_m} \left(\prod_{j=1}^m f_j(r_j) \right) \frac{\partial}{\partial \beta_v} h(z_1, z_2, \ldots, z_m),$$

where $z_j = x_{j,v+ns} - r_j$ $(j = 1, \ldots, m)$ and the first-order partial derivatives of function $h(z_1, z_2, \ldots, z_m)$ are expressed in (B.3) and (B.4), respectively.

Remark: The digamma function, ψ, is defined as the logarithmic derivative of the gamma function $\psi(x) = \frac{d}{dx} \ln[\Gamma(x)] = \frac{\Gamma'(x)}{\Gamma(x)}$.

B.3 Assumptions of Billingsley's theorem

For a fixed v ($v = 1, \dots, s$), let the vector of parameters from the innovation process be

$$\xi_v = (\lambda_{1,v}, \lambda_{2,v}, \dots, \lambda_{m,v}, \beta_v) = (\xi_{1,v}, \xi_{2,v}, \dots, \xi_{m,v}, \xi_{m+1,v}) \in B.$$

(**C1**) The set $\{a : P(\mathbf{Z}_{v+ns} = a) = f(a, \xi_v)\}$ does not depend on ξ_v;

(**C2**) $E[\mathbf{Z}^3_{v+ns}] < \infty$;

(**C3**) $f(a, \xi_v)$ is three times continuously differentiable on the set of parameters B;

(**C4**) For any $\xi_v \in B$, there exists a neighbourhood U of ξ_v such that

$$\sum_{a=0}^{\infty} \sup_{\xi_v \in U} f(a, \xi_v) < \infty \ ,$$

$$\sum_{a=0}^{\infty} \sup_{\xi_v \in U} \left| \frac{\partial}{\partial \xi_{u,v}} f(a, \xi_v) \right| < \infty, \ u = 1, \dots, m+1,$$

$$\sum_{a=0}^{\infty} \sup_{\xi_v \in U} \left| \frac{\partial^2}{\partial \xi_{u,v} \partial \xi_{w,v}} f(a, \xi_v) \right| < \infty, \ u, w = 1, \dots, m+1;$$

(**C5**) For any $\xi_v \in B$ there exists a neighbourhood U of ξ_v and increasing sequences $\psi_u(n)$, $\psi_{u,w}(n)$, $\psi_{u,w,y}(n), n \geq 0$ such that for all $\xi_v \in B$ and all $a \leq n$ with nonvanishing $f(a, \xi_v)$

$$\left| \frac{\partial}{\partial \xi_{u,v}} f(a, \xi_v) \right| \leq \psi_u(n) \ f(a, \xi_v),$$

$$\left| \frac{\partial^2}{\partial \xi_{u,v} \partial \xi_{w,v}} f(a, \xi_v) \right| \leq \psi_{u,w}(n) \ f(a, \xi_v),$$

$$\left| \frac{\partial^3}{\partial \xi_{u,v} \partial \xi_{w,v} \partial \xi_{y,v}} f(a, \xi_v) \right| \leq \psi_{u,w,y}(n) \ f(a, \xi_v), \ u, w, y = 1, \dots, m+1;$$

and also concerning the cyclostationary distribution of $\mathbf{X}_t$, with $t = v + ns$:

$E[\psi_u^3(\mathbf{X}_v)] < \infty$, $\quad E[\mathbf{X}_v \psi_{u,w}(\mathbf{X}_{v+1})] < \infty$,

$E[\psi_u(\mathbf{X}_v) \psi_{u,w}(\mathbf{X}_{v+1})] < \infty$, $\quad E[\psi_{u,w,y}(\mathbf{X}_v)] < \infty$;

(**C6**) The Fisher information matrix, $I(\boldsymbol{\theta})$, is nonsingular.

Appendix C

Auxiliary results of Chapter 3

C.1 First-order partial derivatives of the transition probability function

The transition probability function in (3.33) has the expression

$$
\begin{aligned}
p_v(b|a) &= \sum_{l=-|a|}^{|a|} P\left(R_t^{(|a|)}(\phi_t) = |a| + sign(a)\cdot l\right)\ P(Z_{v+ns} = b-l) = \\
&= \sum_{l=-|a|}^{|a|} \left\{ Ke^{-(\lambda_v+\tau_v)}\lambda_v^{b-l} \sum_{k=max(0,-(b-l))}^{\infty} \frac{\lambda_v^k \tau_v^k}{k!(k+b-l)!} \right\} = \\
&= \sum_{l=-|a|}^{|a|} K\ e^{-(\lambda_v+\tau_v)}\ Q(\eta_v, b-l), \qquad \text{(C.1)}
\end{aligned}
$$

where $a = x_{v-1+ns}$, $b = x_{v+ns}$ and $K = C^{2|a|}_{|a|+sign(a)\cdot l}\alpha_v^{|a|+sign(a)\cdot l}(1-\alpha_v)^{|a|-sign(a)\cdot l}$. For any $c \in \mathbb{Z}$ and $\eta_v = (\lambda_v, \tau_v) \in [0,\infty[\times[0,\infty[$, the auxiliary function Q is defined as

$$
Q(\eta_v, c) = \sum_{k=max(0,-c)}^{\infty} \frac{\lambda_v^{k+c}\tau_v^k}{k!(k+c)!}. \qquad \text{(C.2)}
$$

First-order partial derivatives of auxiliary function $Q(\eta_v, c)$ in (C.2):

$$\frac{\partial}{\partial \lambda_v} Q(\eta_v, c) = \sum_{k=0}^{\infty} \frac{\tau_v^k (k+c) \lambda_v^{k+c-1}}{k!(k+c)(k+c-1)!} = \sum_{k=0}^{\infty} \frac{\tau_v^k \lambda_v^{k+c-1}}{k!(k+c-1)!} = Q(\eta_v, c-1)$$

and

$$\frac{\partial}{\partial \tau_v} Q(\eta_v, c) = \sum_{k=1}^{\infty} \frac{k \tau_v^{k-1} \lambda_v^{k+c}}{k(k-1)!(k+c)!} = \sum_{i=0}^{\infty} \frac{\tau_v^i \lambda_v^{i+c+1}}{i!(i+c+1)!} = Q(\eta_v, c+1).$$

First-order partial derivatives of transition probability function $p_v(b|a)$ in (C.1) with respect to parameters λ_v and τ_v:

$$\begin{aligned}
\frac{\partial}{\partial \lambda_v} p_v(b|a) &= -p_v(b|a) + \sum_{l=-|a|}^{|a|} K \ e^{-(\lambda_v+\tau_v)} \ \frac{\partial}{\partial \lambda_v} Q(\eta_v, b-l) \\
&= -p_v(b|a) + \sum_{l=-|a|}^{|a|} K \ e^{-(\lambda_v+\tau_v)} \ Q(\eta_v, b-l-1) \\
&= -p_v(b|a) + p_v(b-1|a), \ \ v = 1, \ldots, s;
\end{aligned}$$

$$\frac{\partial}{\partial \lambda_w} p_v(b|a) = 0, \ \ w \neq v, v = 1, \ldots, s$$

and

$$\begin{aligned}
\frac{\partial}{\partial \tau_v} p_v(b|a) &= -p_v(b|a) + \sum_{l=-|a|}^{|a|} K \ e^{-(\lambda_v+\tau_v)} \ \frac{\partial Q(\eta_v, b-l)}{\partial \tau_v} = \\
&= -p_v(b|a) + \sum_{l=-|a|}^{|a|} K \ e^{-(\lambda_v+\tau_v)} \ Q(\eta_v, b-l+1) = \\
&= -p_v(b|a) + p_v(b+1|a), \ \ v = 1, \ldots, s;
\end{aligned}$$

$$\frac{\partial}{\partial \tau_w} p_v(b|a) = 0, \ \ w \neq v, v = 1, \ldots, s.$$

First-order partial derivatives of transition probability function $p_v(b|a)$ in (C.1) with respect to parameter α_v:

$$\frac{\partial}{\partial \alpha_v} p_v(b|a) = \sum_{l=-|a|}^{|a|} C_{|a|+sign(a)\cdot l}^{2|a|} \Big((|a| + sign(a) \cdot l) \alpha_v^{|a|+sign(a)\cdot l-1} (1-\alpha_v)^{|a|-sign(a)\cdot l} -$$

$$- (|a| - sign(a) \cdot l)(1-\alpha_v)^{|a|-sign(a)\cdot l-1} \alpha_v^{|a|+sign(a)\cdot l} \Big) P(Z_{v+ns} = b-l) =$$

$$= \sum_{l=-|a|}^{|a|} K \left(\frac{|a| + sign(a)\cdot l}{\alpha_v} - \frac{|a| - sign(a)\cdot l}{1-\alpha_v} \right) P(Z_{v+ns} = b-l) =$$

$$= \sum_{l=-|a|}^{|a|} K \left(\frac{|a| + sign(a)\cdot l}{\alpha_v(1-\alpha_v)} - \frac{2|a|}{1-\alpha_v} \right) P(Z_{v+ns} = b-l) =$$

$$= \sum_{l=-|a|}^{|a|} K \frac{|a| + sign(a)\cdot l}{\alpha_v(1-\alpha_v)} \; P(Z_{v+ns} = b-l) - \frac{2|a|}{1-\alpha_v} \sum_{l=-|a|}^{|a|} K P(Z_{v+ns} = b-l) =$$

$$= \underbrace{\sum_{l=-|a|}^{|a|} K \frac{|a| + sign(a)\cdot l}{\alpha_v(1-\alpha_v)} \; P(Z_{v+ns} = b-l)}_{\mathbf{A}} - \frac{2|a|}{1-\alpha_v} \; p_v(b|a),$$

where

$$\mathbf{A} = \sum_{l=-|a|}^{|a|} K \frac{|a| + sign(a)\cdot l}{\alpha_v(1-\alpha_v)} \; P(Z_{v+ns} = b-l) =$$

$$= \sum_{l=-|a|}^{|a|} C_{|a|+sign(a)\cdot l}^{2|a|} \alpha_v^{|a|+sign(a)\cdot l-1} (1-\alpha_v)^{|a|-sign(a)\cdot l-1} (|a| + sign(a)\cdot l) \; P(Z_{v+ns} = b-l) =$$

$$= \sum_{l=-|a|}^{|a|} \frac{(2|a|)! \; \alpha_v^{|a|+sign(a)\cdot l-1} (1-\alpha_v)^{|a|-sign(a)\cdot l-1}}{(|a| + sign(a)\cdot l - 1)!(|a| - sign(a)\cdot l)!} \; P(Z_{v+ns} = b-l) =$$

$$= \sum_{l=-|a|}^{|a|} \frac{2|a|(2|a|-1)!}{(|a| + sign(a)\cdot l - 1)!(2|a| - 1 - (|a| + sign(a)\cdot l - 1))!} \times$$

$$\times \left(\alpha_v^{|a|+sign(a)\cdot l-1} (1-\alpha_v)^{2|a|-1-(|a|+sign(a)\cdot l-1)} \;\; + \alpha_v^{|a|+sign(a)\cdot l} (1-\alpha_v)^{2|a|-|a|+sign(a)\cdot l-1} \right) \times$$

$$\times \;\; P(Z_{v+ns} = b-l) =$$

$$= \sum_{l=-|a|}^{|a|} 2|a| C_{|a|+sign(a)\cdot l-1}^{2|a|-1} \Big(\alpha_v^{|a|+sign(a)\cdot l-1} (1-\alpha_v)^{2|a|-1-(|a|+sign(a)\cdot l-1)} +$$

$$+ \alpha_v^{|a|+sign(a)\cdot l-1+1} (1-\alpha_v)^{2|a|-1-(|a|+sign(a)\cdot l-1)-1} \Big) P(Z_{v+ns} = b-1-(l-1)) =$$

$$= \sum_{l=-|a|}^{|a|} 2|a| C_{|a|+sign(a)\cdot l}^{2|a|-1} \Big(\alpha_v^{|a|+sign(a)\cdot l} (1-\alpha_v)^{2|a|-1-|a|+sign(a)\cdot l} +$$

$$+ \alpha_v^{|a|+sign(a)\cdot l+1} (1-\alpha_v)^{2|a|-1-|a|+sign(a)\cdot l-1} \Big) P(Z_{v+ns} = b-1-l) =$$

$$= \sum_{l=-|a|}^{|a|} \frac{2|a|}{1-\alpha_v} C_{|a|+sign(a)\cdot l}^{2|a|-1} \alpha_v^{|a|+sign(a)\cdot l} (1-\alpha_v)^{2|a|-1-|a|+sign(a)\cdot l} P(Z_{v+ns} = b-1-l) =$$

$$= \frac{2|a|}{1-\alpha_v}\ p_v(b-1|a-1).$$

Hence,

$$\frac{\partial}{\partial \alpha_v} p_v(b|a) = \frac{2|a|}{1-\alpha_v}\ \left[p_v(b-1|a-1) - p_v(b|a)\right],\ v = 1, \ldots, s;$$

$$\frac{\partial}{\partial \alpha_w} p_v(b|a) = 0,\ w \neq v, v = 1, \ldots, s.$$

C.2 Simulation study - Tables and Figures for Set 1B, Set 2B and Set 3B

Table C.1: CLS and CML estimates for $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\tau})$ in Set 1B. MSE in parenthesis.

	$N = 50$		$N = 200$		$N = 500$	
	CLS	CML	CLS	CML	CLS	CML
$\boldsymbol{\alpha} = (0.60, 0.40, 0.75, 0.30)$						
$\hat{\alpha}_1$	0.596	0.599	0.602	0.600	0.598	0.600
	(0.0221)	(0.0278)	(0.0010)	(0.0001)	(0.0002)	(0.0002)
$\hat{\alpha}_2$	0.402	0.404	0.400	0.401	0.400	0.401
	(0.0045)	(0.0020)	(0.0022)	(0.0010)	(0.0001)	(0.0002)
$\hat{\alpha}_3$	0.752	0.752	0.748	0.750	0.751	0.750
	(0.0055)	(0.0002)	(0.0003)	(0.0011)	(0.0003)	(0.0002)
$\hat{\alpha}_4$	0.291	0.302	0.300	0.300	0.300	0.300
	(0.0133)	(0.0010)	(0.0009)	(0.0002)	(0.0026)	(0.0001)
$\boldsymbol{\lambda} = (5, 2, 1, 6)$						
$\hat{\lambda}_1$	4.890	4.851	4.961	4.960	5.013	4.999
	(1.2441)	(0.0203)	(0.0015)	(0.4656)	(0.1746)	(0.0104)
$\hat{\lambda}_2$	1.879	1.827	1.969	1.961	1.993	1.978
	(0.2489)	(0.4645)	(0.1178)	(0.1671)	(0.2582)	(0.0618)
$\hat{\lambda}_3$	0.969	0.944	0.953	0.970	0.994	0.989
	(0.0504)	(0.2910)	(0.0344)	(0.2480)	(0.0006)	(0.0005)
$\hat{\lambda}_4$	6.062	5.972	5.994	5.967	5.971	5.983
	(0.7703)	(0.2740)	(0.3578)	(0.2704)	(0.0002)	(0.0006)
$\boldsymbol{\tau} = (4, 5, 3, 1)$						
$\hat{\tau}_1$	3.849	3.822	3.989	3.956	3.996	4.006
	(0.1385)	(1.7209)	(0.0965)	(1.1438)	(0.3113)	(0.0099)
$\hat{\tau}_2$	4.882	4.878	4.972	4.972	4.990	4.983
	(0.1419)	(0.0579)	(0.0426)	(0.1829)	(0.1108)	(0.0084)
$\hat{\tau}_3$	2.944	2.930	2.968	2.970	2.995	2.987
	(0.1213)	(1.0097)	(0.0001)	(0.0152)	(0.0050)	(0.0053)
$\hat{\tau}_4$	1.134	0.922	0.980	0.958	0.970	0.981
	(0.1173)	(0.0522)	(0.0209)	(0.0380)	(0.0535)	(0.0184)

Table C.2: CLS and CML estimates for $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\tau})$ in Set 2B. MSE in parenthesis.

	$N = 50$		$N = 200$		$N = 500$	
	CLS	CML	CLS	CML	CLS	CML
$\boldsymbol{\alpha} = (0.20, 0.45, 0.10, 0.30)$						
$\hat{\alpha}_1$	0.201	0.198	0.200	0.199	0.200	0.200
	(0.0010)	(0.0007)	(0.0044)	(0.0001)	(0.0002)	(0.0004)
$\hat{\alpha}_2$	0.451	0.451	0.450	0.449	0.451	0.449
	(0.0064)	(0.0040)	(0.0007)	(0.0009)	(0.0001)	(0.0005)
$\hat{\alpha}_3$	0.110	0.105	0.100	0.099	0.100	0.100
	(0.0022)	(0.0011)	(0.0017)	(0.0003)	(0.0001)	(0.0002)
$\hat{\alpha}_4$	0.302	0.301	0.301	0.300	0.302	0.301
	(0.0202)	(0.0028)	(0.0014)	(0.0014)	(0.0003)	(0.0006)
$\boldsymbol{\lambda} = (5, 2, 1, 6)$						
$\hat{\lambda}_1$	4.966	4.868	4.976	5.001	5.003	5.012
	(1.4272)	(0.4976)	(0.4260)	(0.0528)	(0.0095)	(0.4522)
$\hat{\lambda}_2$	1.917	1.937	1.979	1.974	1.981	1.981
	(0.2015)	(0.2534)	(0.0710)	(0.2369)	(0.0834)	(0.0286)
$\hat{\lambda}_3$	0.942	0.894	0.978	0.974	0.994	0.992
	(0.0393)	(0.3031)	(0.0604)	(0.1032)	(0.0512)	(0.0034)
$\hat{\lambda}_4$	5.799	5.828	5.965	5.944	5.979	5.978
	(0.9201)	(0.1717)	(0.3392)	(0.0079)	(0.1061)	(0.0019)
$\boldsymbol{\tau} = (2, 1, 4, 3)$						
$\hat{\tau}_1$	1.966	1.849	1.985	1.988	2.007	2.002
	(1.4475)	(0.3664)	(0.0453)	(0.0042)	(0.0261)	(0.1342)
$\hat{\tau}_2$	0.932	0.929	0.976	0.970	0.985	0.983
	(0.0052)	(0.3645)	(0.0008)	(0.1674)	(0.1286)	(0.0728)
$\hat{\tau}_3$	3.945	3.896	3.988	3.981	3.997	3.996
	(0.1780)	(0.6659)	(0.1507)	(0.1789)	(0.0033)	(0.0031)
$\hat{\tau}_4$	2.796	2.826	2.953	2.945	2.957	2.970
	(0.0266)	(0.8391)	(0.1803)	(0.1891)	(0.1254)	(0.0103)

Table C.3: CLS and CML estimates for $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\tau})$ in Set 3B. MSE in parenthesis.

	$N = 50$		$N = 200$		$N = 500$	
	CLS	CML	CLS	CML	CLS	CML
$\boldsymbol{\alpha} = (0.75, 0.62, 0.51, 0.86)$						
$\hat{\alpha}_1$	0.747	0.750	0.750	0.751	0.750	0.751
	(0.0059)	(0.0025)	(0.0047)	(0.0007)	(0.0001)	(0.0009)
$\hat{\alpha}_2$	0.620	0.619	0.618	0.620	0.620	0.620
	(0.0005)	(0.0064)	(0.0004)	(0.0021)	(0.0002)	(0.0006)
$\hat{\alpha}_3$	0.502	0.505	0.508	0.508	0.510	0.508
	(0.0109)	(0.0037)	(0.0001)	(0.0008)	(0.0005)	(0.0009)
$\hat{\alpha}_4$	0.860	0.858	0.859	0.860	0.860	0.860
	(0.0003)	(0.0015)	(0.0027)	(0.0001)	(0.0003)	(0.0002)
$\boldsymbol{\lambda} = (4, 5, 3, 1)$						
$\hat{\lambda}_1$	3.908	3.874	3.963	3.980	3.986	3.999
	(0.0877)	(0.7006)	(0.0417)	(0.0002)	(0.0072)	(0.0012)
$\hat{\lambda}_2$	4.991	4.898	4.953	4.950	4.981	4.967
	(0.8363)	(0.0014)	(0.0622)	(0.0181)	(0.0303)	(0.0780)
$\hat{\lambda}_3$	2.893	2.858	2.985	2.959	2.983	2.974
	(0.3084)	(1.9967)	(0.0191)	(0.6133)	(0.0214)	(0.0066)
$\hat{\lambda}_4$	0.932	0.941	0.976	0.987	0.996	0.999
	(0.2119)	(0.9387)	(0.0636)	(0.0019)	(0.0030)	(0.0009)
$\boldsymbol{\tau} = (2, 1, 4, 3)$						
$\hat{\tau}_1$	1.922	1.879	1.961	1.982	1.981	1.997
	(0.0288)	(0.0695)	(0.4861)	(0.0074)	(0.0216)	(0.1217)
$\hat{\tau}_2$	0.967	0.912	0.952	0.953	0.985	0.973
	(0.2418)	(0.3462)	(0.2008)	(0.0015)	(0.0299)	(0.1177)
$\hat{\tau}_3$	3.819	3.838	3.978	3.943	3.988	3.958
	(0.2147)	(1.0893)	(0.1076)	(0.1036)	(0.0808)	(0.0905)
$\hat{\tau}_4$	2.923	2.947	2.980	2.993	2.998	3.003
	(0.4280)	(0.3441)	(0.0246)	(0.0018)	(0.0001)	(0.0109)

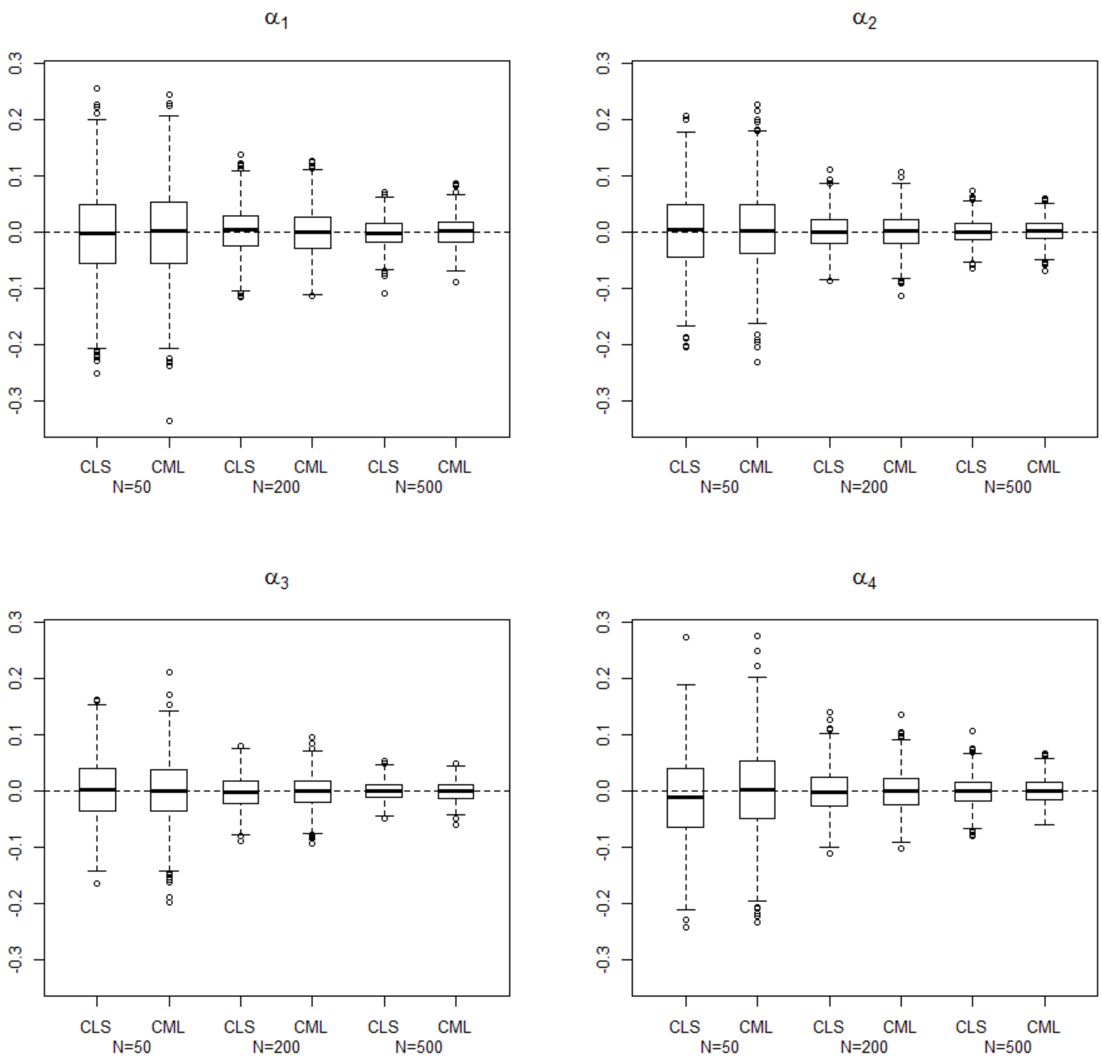

Figure C.1: Boxplots for the biases of the CLS and CML estimates of parameter $\boldsymbol{\alpha}$ in Set 1B for $n = 4N = 200, 800, 2000$.

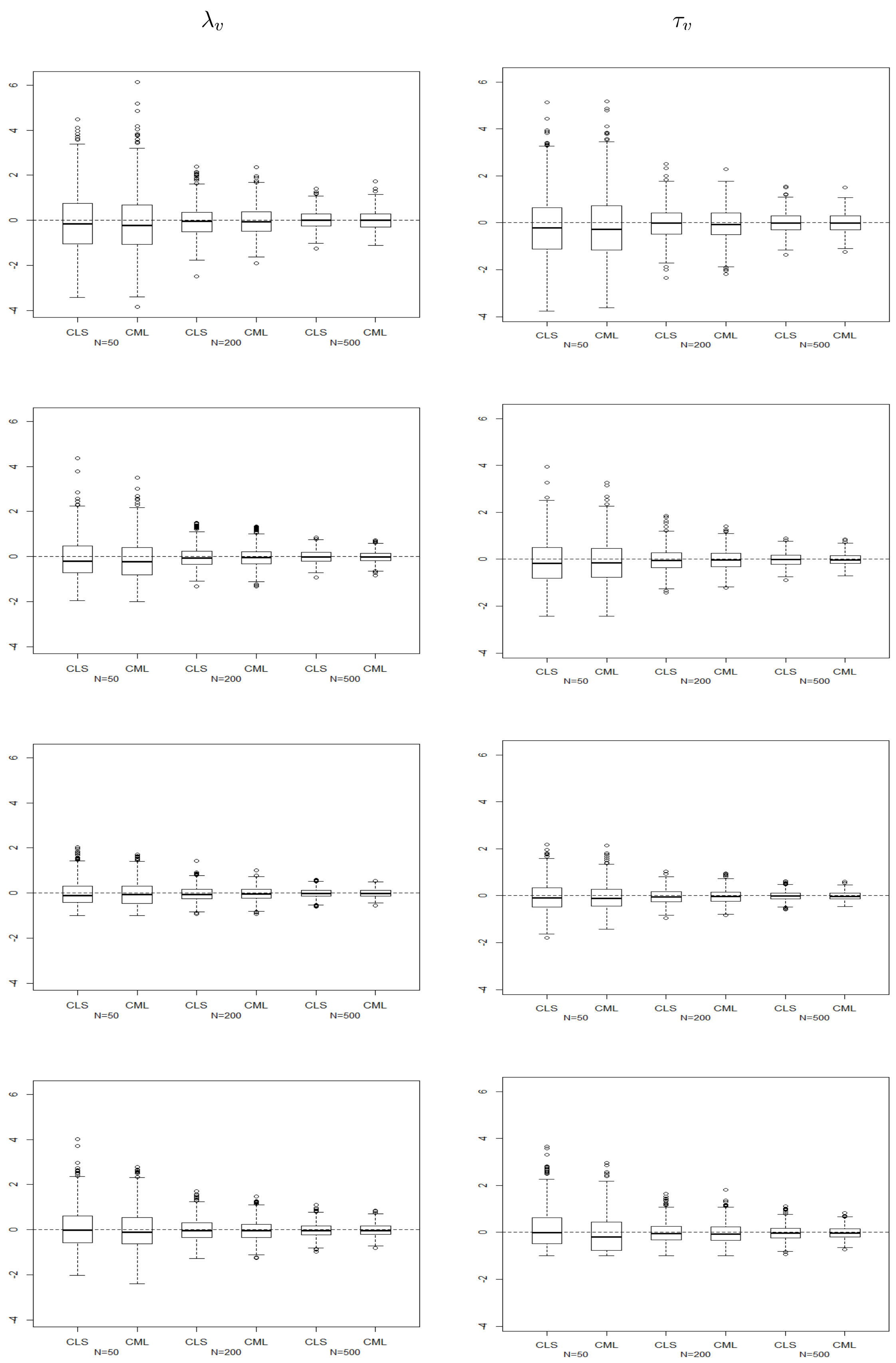

Figure C.2: Boxplots for the biases of the CLS and CML estimates of parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\tau}$ in Set 1B for $n = 4N = 200, 800, 2000$.

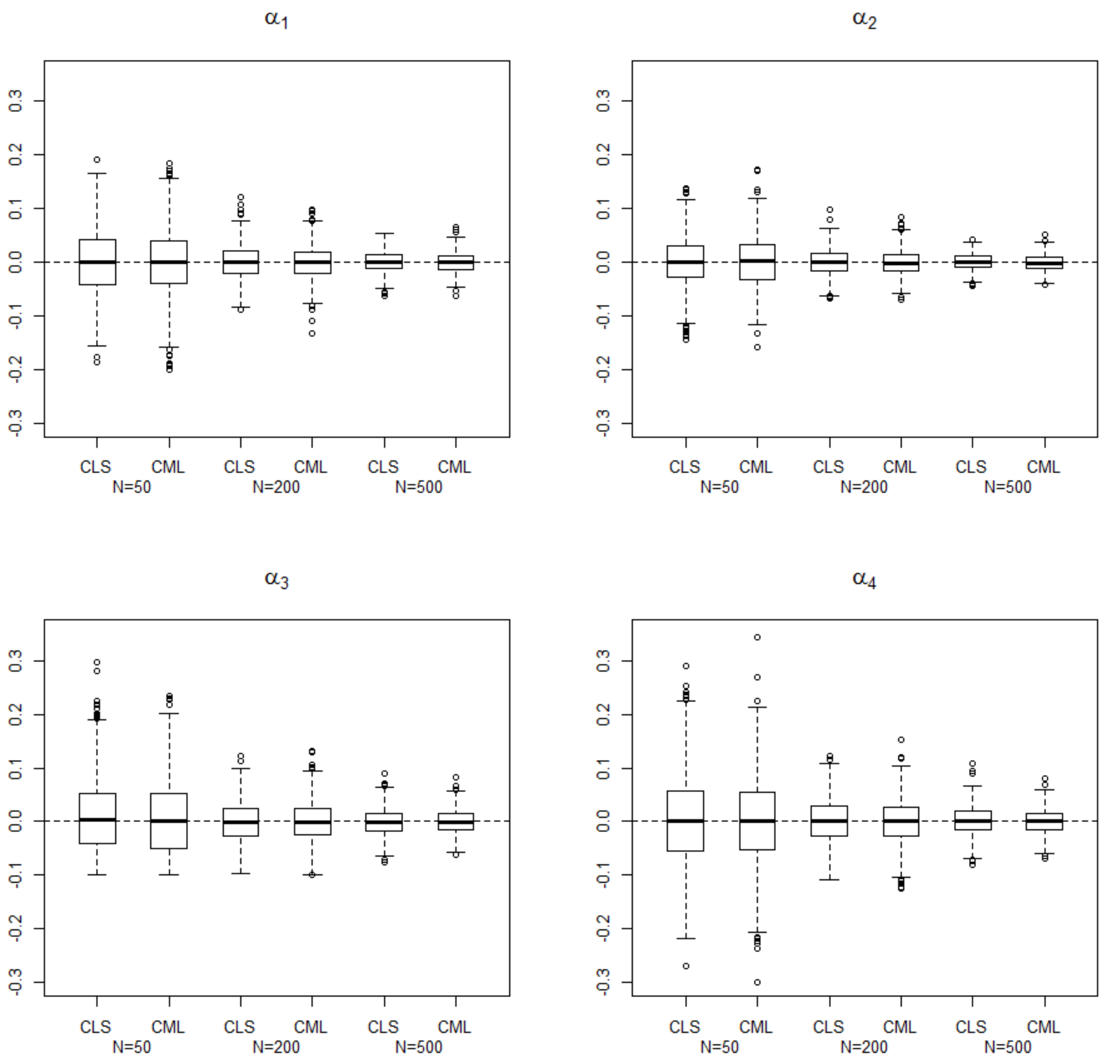

Figure C.3: Boxplots for the biases of the CLS and CML estimates of parameter $\boldsymbol{\alpha}$ in Set 2B for $n = 4N = 200, 800, 2000$.

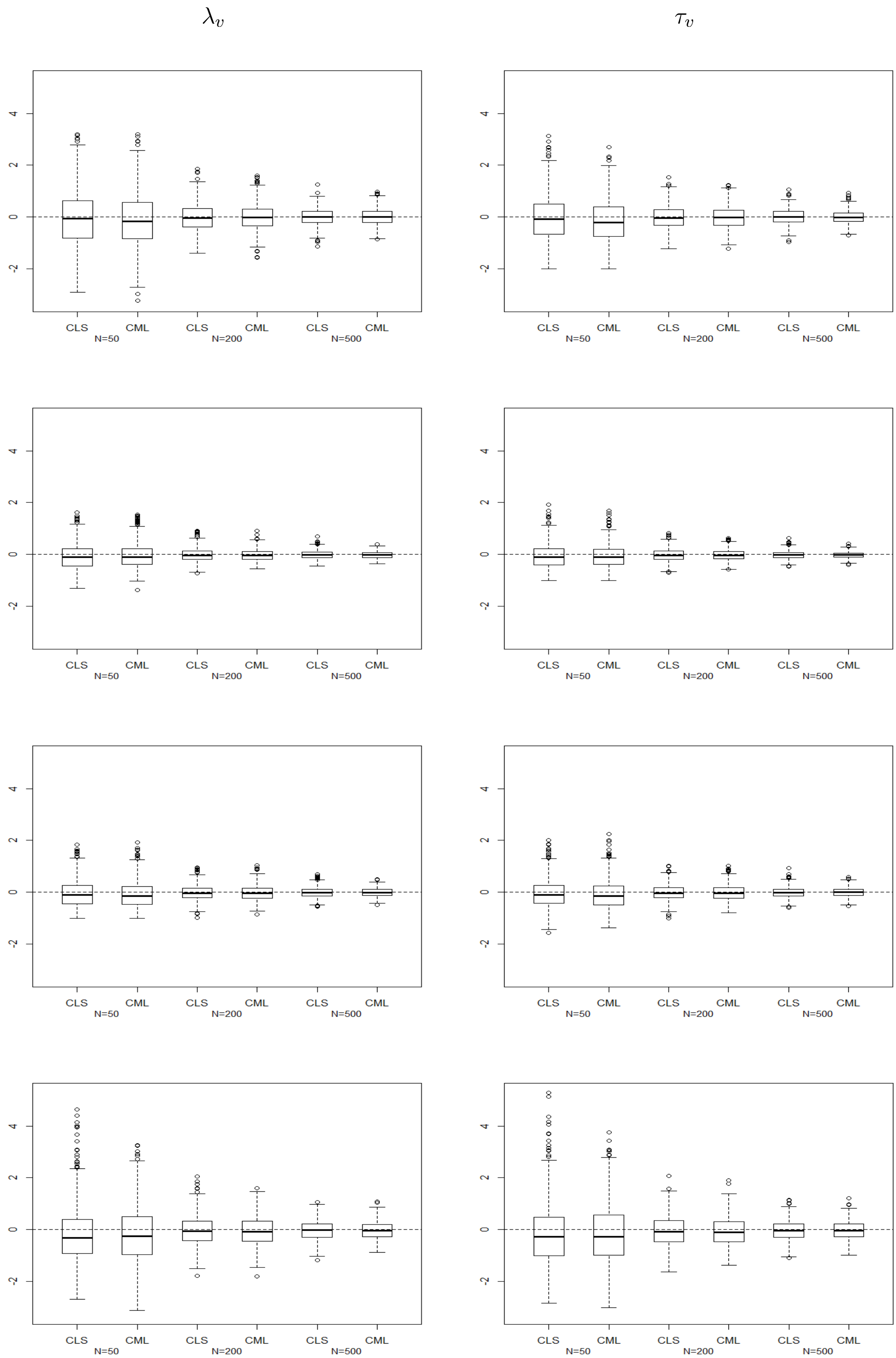

Figure C.4: Boxplots for the biases of the CLS and CML estimates of parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\tau}$ in Set 2B for $n = 4N = 200, 800, 2000$.

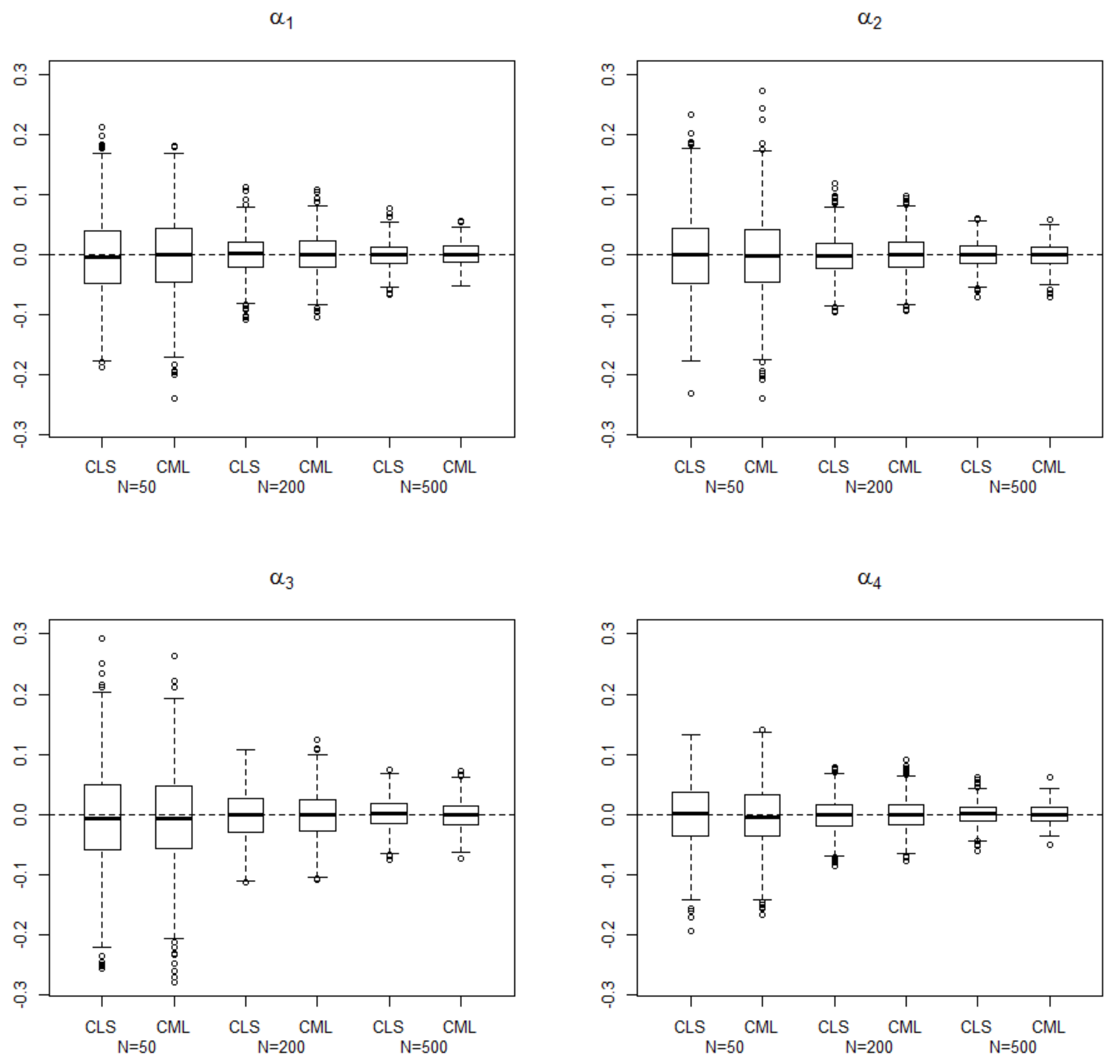

Figure C.5: Boxplots for the biases of the CLS and CML estimates of parameter $\boldsymbol{\alpha}$ in Set 3B for $n = 4N = 200, 800, 2000$.

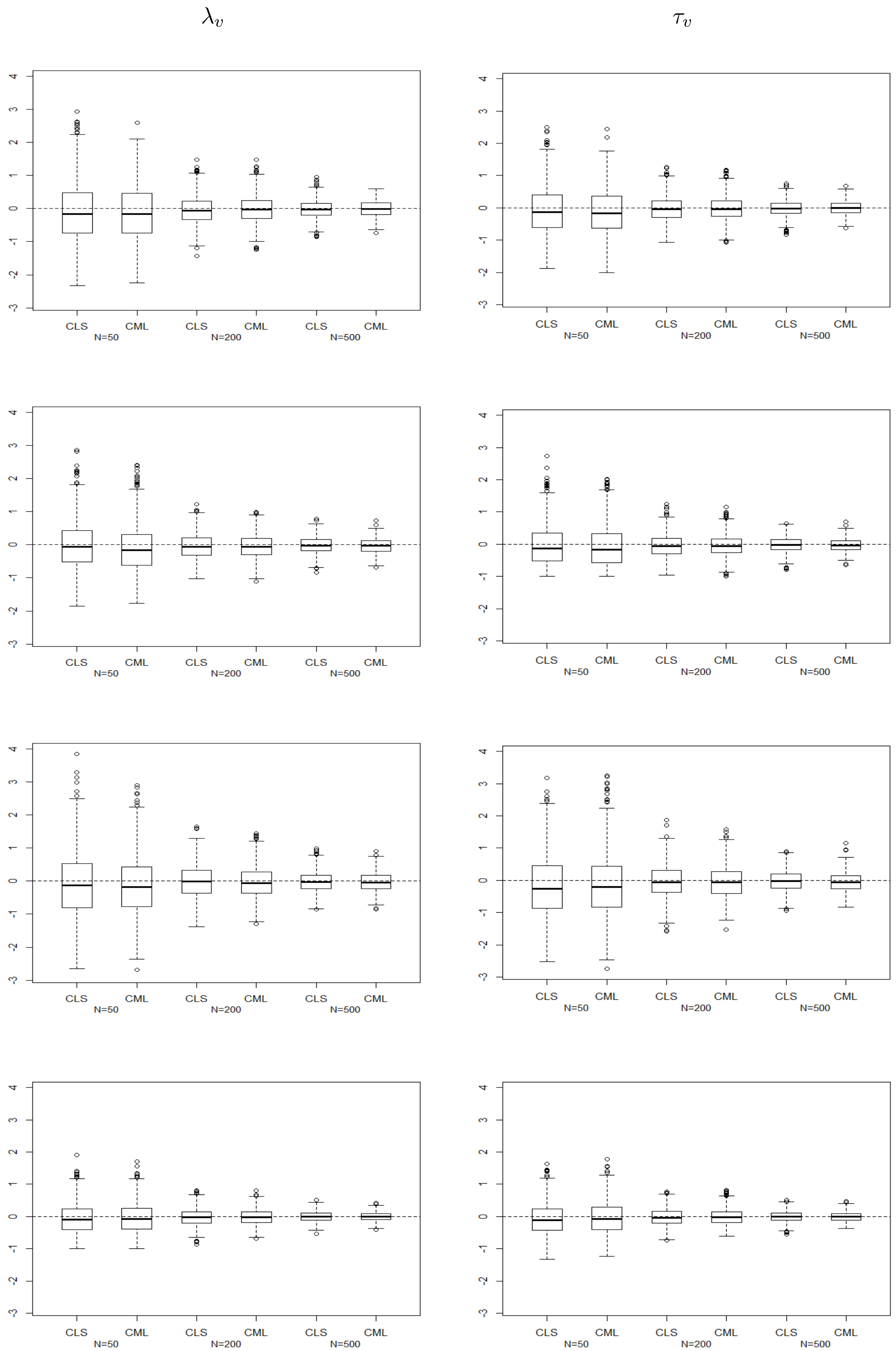

Figure C.6: Boxplots for the biases of the CLS and CML estimates of parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\tau}$ in Set 3B for $n = 4N = 200, 800, 2000$.

Appendix D

R codes

D.1 R functions related to Chapter 2

```
##########################################################
## Generate trivariate negative binomial innovations ##
##########################################################
gera_binomNeg <- function(num, bet, lam1, lam2, lam3){
niu <- rgamma(num, shape=1/bet, rate=1/bet);
L1 <- niu*lam1;
L2 <- niu*lam2;
L3 <- niu*lam3;
z1 <- rpois(num, L1);
z2 <- rpois(num, L2);
z3 <- rpois(num, L3);
return(array(cbind(z1, z2, z3), dim=c(num,3)))
}        ## end function
```

```
######################################
## Generate PMINAR(1) - trivariate ##
######################################
gera_inarTri <- function(n, s, N, alfa, lamb, betta){
# s - nº periods
# N - nº cycles
# n - nº observ total, n=sN
# alfa, lamb - matrix s by 3 ; betta - vector
nobs <- 3*(n+1);
x <- array(rep(0, nobs), dim=c(n+1,3));
x[1,1] <- 3;          ## initial observ.: x0=c( , , )
x[1,2] <- 6;          ## 1º v (season v=1,...,s)
x[1,3] <- 4;          ## 2º j (component)
alfa_matx <- array(0, dim=c((N)*s,3));
lamb_matx <- array(0, dim=c((N)*s,3));
for(j in 1:3) {
        alfa_matx[,j] <- rep(alfa[,j],N);
        lamb_matx[,j] <- rep(lamb[,j],N)
}       ## end for
alfa_aux <- array(0, dim=c(s*N+1,3));
alfa_aux[2:(s*N+1),] <- alfa_matx;
lamb_aux <- array(0, dim=c(s*N+1,3));
lamb_aux[2:(N*s+1),] <- lamb_matx;
betaa_vec <- rep(betta, N);
betaa_aux <- array(0, dim=c(s*N+1,1));
betaa_aux[2:(N*s+1)] <- betaa_vec;
for(v in 2:(n+1)){
        inov_NBtri <- gera_binomNeg(1, betaa_aux[v],
                lamb_aux[v,1], lamb_aux[v,2], lamb_aux[v,3]);
```

```
          for(j in 1:3){
                    binom <- rbinom(1, x[v-1,j], alfa_aux[v,j]);
                    x[v,j] <- binom + inov_NBtri[1,j]
          }         ## end for
}         ## end for
return(x[(2:(n+1)),])
}         ## end function

#######################################
## Function for admissible values ##
#######################################
fun_alfa <- function(alfaC){
all((alfaC > 0) & (alfaC < 1))
}         ## end function

fun_lamb <- function(lambC){
all(lambC > 0)
}         ## end function

######################
## Product - alphas ##
######################
mult_alfa <- function(alfas, m, i){
if(i==0){                    ## i - nº factors
          phi <- 1
          } else {
                    alf <- alfas[(m-i+1):m];
                    phi <- prod(alf)
}         ## end if
```

```
return(phi)
}          ## end function

##############################
## Estimation: Yule-Walker ##
##############################
estim_YW <- function(n, s, N, X){
s_mu0 <- array(0, dim=c(s,3));
s_var0 <- array(0, dim=c(s,3));
for(j in 1:3){
          for(v in 1:s){
                    s_mu0[v,j] <- mean(X[v+s*(0:(N-1)),j]);
                    s_var0[v,j] <- var(X[v+s*(0:(N-1)),j])
          }          ## end for
}          ## end for
s_gama0 <- array(0, dim=c(s,3));
for(j in 1:3){
          for(v in 1:(s-1)){
          s_gama0[v,j] <- cov(X[v+s*(0:(N-1)),j],X[v+1+s*(0:(N-1)),j])
          }          ## end for
          s_gama0[s,j] <- cov(X[s+s*(0:(N-2)),j],X[1+s+s*(0:(N-2)),j])
}          ## end for
alfaYW0 <- array(0, dim=c(s,3));
for(j in 1:3){
          alfaYW0[1,j] <- s_gama0[s,j]/s_var0[s,j];
          for(v in 2:s){
                    alfaYW0[v,j] <- s_gama0[(v-1),j]/s_var0[(v-1),j]
          }          ## end for
}          ## end for
```

```
lambdaYW0 <- array(0, dim=c(s,3));
for(j in 1:3){
          lambdaYW0[1,j] <- s_mu0[1,j]- alfaYW0[1,j]*s_mu0[s,j];
          for(v in 2:s){
          lambdaYW0[v,j] <- s_mu0[v,j]- alfaYW0[v,j]*s_mu0[(v-1),j]
          }          ## end for
}          ## end for
phi1 <- mult_alfa(alfaYW0[,1],s,s);
phi2 <- mult_alfa(alfaYW0[,2],s,s);
phi3 <- mult_alfa(alfaYW0[,3],s,s);
betaYW0 <- rep(0,s);
numer1 <- (1-phi1*phi2)*cov(X[1+s*(0:(N-1)),1],X[1+s*(0:(N-1)),2]);
d1a <- mult_alfa(alfaYW0[,1],1,0)*mult_alfa(alfaYW0[,2],1,0)*
                              lambdaYW0[1,1]*lambdaYW0[1,2];
const1 <- mult_alfa(alfaYW0[,1],1,1)*mult_alfa(alfaYW0[,2],1,1);
soma11 <- 0;
for(i in 0:2){
          d1b <- mult_alfa(alfaYW0[,1],4,i)*mult_alfa(alfaYW0[,2],4,i)*
                              lambdaYW0[4-i,1]*lambdaYW0[4-i,2];
          soma11 <- soma11+d1b
}          ## end for
betaYW0[1] <- numer1/(d1a+(const1*soma11));
numer2 <- (1-phi1*phi2)*cov(X[2+s*(0:(N-1)),1],X[2+s*(0:(N-1)),2]);
soma21 <- 0;
for(i in 0:1){
          d2a<-mult_alfa(alfaYW0[,1],2,i)*mult_alfa(alfaYW0[,2],2,i)*
                              lambdaYW0[2-i,1]*lambdaYW0[2-i,2];
          soma21<-soma21+d2a
}          ## end for
```

```
const2 <- mult_alfa(alfaYW0[,1],2,2)*mult_alfa(alfaYW0[,2],2,2);
soma22 <- 0;
for(i in 0:1){
        d2b <- mult_alfa(alfaYW0[,1],4,i)*mult_alfa(alfaYW0[,2],4,i)*
                        lambdaYW0[4-i,1]*lambdaYW0[4-i,2];
        soma22 <- soma22+d2b
}       ## end for
betaYW0[2] <- numer2/(soma21+(const2*soma22));
numer3 <- (1-phi1*phi2)*cov(X[3+s*(0:(N-1)),1],X[3+s*(0:(N-1)),2]);
soma31 <- 0;
for(i in 0:2){
        d3a <- mult_alfa(alfaYW0[,1],3,i)*mult_alfa(alfaYW0[,2],3,i)*
                        lambdaYW0[3-i,1]*lambdaYW0[3-i,2];
        soma31 <- soma31+d3a
}       ## end for
const3 <- mult_alfa(alfaYW0[,1],3,3)*mult_alfa(alfaYW0[,2],3,3);
d3b <- mult_alfa(alfaYW0[,1],4,0)*mult_alfa(alfaYW0[,2],4,0)*
                        lambdaYW0[4,1]*lambdaYW0[4,2];
betaYW0[3] <- numer3/(soma31+(const3*d3b));
numer4 <- (1-phi1*phi2)*cov(X[4+s*(0:(N-1)),1],X[4+s*(0:(N-1)),2]);
soma4 <- 0;
for(i in 0:3){
        d4 <- mult_alfa(alfaYW0[,1],4,i)*mult_alfa(alfaYW0[,2],4,i)*
                        lambdaYW0[4-i,1]*lambdaYW0[4-i,2];
        soma4 <- soma4+d4
}       ## end for
betaYW0[4] <- numer4/soma4;
if(fun_alfa(alfaYW0)& fun_lamb(lambdaYW0)){
        alfaYW0 <- alfaYW0;
```

```
            lambdaYW0 <- lambdaYW0;
            betaYW0 <- betaYW0
            }else {
                        alfaYW0 <- array(NA, dim=c(s,3));
                        lambdaYW0 <- array(NA, dim=c(s,3));
                        betaYW0 <- rep(NA, s)
}           ## end if
param_est <- array(cbind(alfaYW0, lambdaYW0, betaYW0), dim=c(s,7))
return(param_est)
}           ## end function

##########################################################
## Functions for conditional maximum likelihood (CML) ##
##########################################################

#################################
## Trivariate NB distribution ##
#################################
ptri_NB <- function(Z, L1, L2, L3, bb){
b <- 1/(bb);                    ## b=1/beta
z1 <- Z[1];
z2 <- Z[2];
z3 <- Z[3];
n <- length(z1);
logbivNB <- vector(length=n);
for(k in 1:n){
            sumpar <- L1+L2+L3+b;
            parc_tau <- lgamma(z1[k]+z2[k]+z3[k]+b)-lgamma(b)-
                        lgamma(z1[k]+1)-lgamma(z2[k]+1)-lgamma(z3[k]+1);
```

```
        logbivNB[k] <- parc_tau+z1[k]*log(L1)+z2[k]*log(L2)+z3[k]*
                log(L3)+b*log(b)-(z1[k]+z2[k]+z3[k]+b)*log(sumpar)
}       ## fim do for
return(exp(logbivNB))
}       ## end function

######################
## Transition prob. ##
######################
prob_trans <- function(xt_1, xt, pars_v){
# xt_1, xt matrices with 3 columns each
# pars_v =(alf1, alf2, alf3, lam1, lam2, lam2, bet)  one season
dimen <- dim(xt);
d1 <- dimen[1];
prob <- rep(0, d1);
if(fun_alfa(pars_v[1:3])& fun_lamb(pars_v[4:7])){
for(v in 1:d1){
        soma <- 0;
        for(r1 in 0:min(xt_1[v,1], xt[v,1])){
                bin1 <- dbinom(r1, xt_1[v,1], pars_v[1]);
                for(r2 in 0:min(xt_1[v,2], xt[v,2])){
                        bin2 <- dbinom(r2, xt_1[v,2], pars_v[2]);
                        for(r3 in 0:min(xt_1[v,3], xt[v,3])) {
                                bin3 <- dbinom(r3, xt_1[v,3], pars_v[3]);
                                t1 <- xt[v,1]-r1;
                                t2 <- xt[v,2]-r2;
                                t3 <- xt[v,3]-r3;
                                negbin <- ptri_NB(c(t1,t2,t3), pars_v[4],
                                        pars_v[5], pars_v[6], pars_v[7]);
```

```
                                        soma <- soma + bin1*bin2*bin3*negbin
                                } ## end for
                        } ## end for
                } ## end for
                prob[v] <- soma;
        } ## end for
} ## end if
return(prob)
} ## end function

#######################################
## CM Log-likelihood for v=1,...,s ##
#######################################
loglik_v <- function(pars_v, v, s, N, X){
# pars_v =(alf1, alf2, alf3, lam1, lam2, lam2, bet)
logk <- 0;
if(v==1){
        xt_1 <- X[v-1+s*(1:(N-1)),];
        xt <- X[v+s*(1:(N-1)),]
        } else{
                xt_1 <- X[v-1+s*(0:(N-1)),];
                xt <- X[v+s*(0:(N-1)),]
} ## end if
logk <- logk + sum(log(prob_trans(xt_1, xt, pars_v)));
logk <- -logk
return(logk)
} ## end function
```

```
######################
## Estimation: CML ##
######################
estimCML_N1 <- function(X, param7){
pars_CML_N1 <- array(0,dim=c(4,7));
for(v in 1:s){
        resN1 <- optim(par=param7[v,], f=loglik_v, v=v, X=X,
                        method="BFGS");
        pars_CML_N1[v,] <- resN1$par
}       ## end for
return(pars_CML_N1)
}       ## end function

#################################################
## Functions for composite likelihood (CL) ##
#################################################
################################
## Bivariate NB distribution ##
################################
pbiv_NB <- function(Z, L1, L2, bb){
b <- 1/(bb);                 ## b=1/beta
z1 <- Z[,1];
z2 <- Z[,2];
n <- length(z1);
logbivNB <- vector(length=n);
for(k in 1:n){
        sumpar <- L1+L2+b;
        parc_tau <- lgamma(z1[k]+z2[k]+b)-lgamma(b)-
                        lgamma(z1[k]+1)-lgamma(z2[k]+1);
```

```
          logbivNB[k] <- parc_tau+z1[k]*log(L1)+z2[k]*log(L2)+
                                  b*log(b)-(z1[k]+z2[k]+b)*log(sumpar)
}         ## fim do for
return(exp(logbivNB))
}         ## end function

##############################################
## Composite Log-likelihood for v=1,...,s ##
##############################################
cloglik_v <- function(theta, v, X, s, N){
if(v==1){
          xtminus1 <- X[v-1+s*(1:(N-1)),];
          xt <- X[v+s*(1:(N-1)),]
          }else{
                    xtminus1 <- X[v-1+s*(0:(N-1)),];
                    xt <- X[v+s*(0:(N-1)),]
}         ## end if
xtminus1_1 <- xtminus1[,1];
xtminus1_2 <- xtminus1[,2];
xtminus1_3 <- xtminus1[,3];
xt1<- xt[,1];
d1 <- length(xt1);
xt2 <- xt[,2];
xt3 <- xt[,3];
alp1 <- theta[1];
alp2 <- theta[2];
alp3 <- theta[3];
lm1 <- theta[4];
lm2 <- theta[5];
```

```
lm3 <- theta[6];
betaa <- theta[7];
p1 <- NULL; p2 <- NULL; p3 <- NULL;
for (v in 2:d1){
        k12 <- rep(0:xt1[v], each=xt2[v]+1);
        s12 <- rep(0:xt2[v], xt1[v]+1);
        k13 <- rep(0:xt1[v], each=xt3[v]+1);
        s13 <- rep(0:xt3[v], xt1[v]+1);
        k23 <- rep(0:xt2[v], each=xt3[v]+1);
        s23 <- rep(0:xt3[v], xt2[v]+1);
        f1 <- dbinom(xt1[v]-k12, xtminus1_1[v], alp1);
        f2 <- dbinom(xt2[v]-s12, xtminus1_2[v], alp2);
        z12 <- matrix(c(k12,s12), ncol=2);
        f12 <- pbiv_NB(z12, lm1, lm2, betaa);
        f3 <- dbinom(xt1[v]-k13, xtminus1_1[v], alp1);
        f4 <- dbinom(xt3[v]-s13, xtminus1_3[v], alp3);
        z13 <- matrix(c(k13,s13), ncol=2);
        f13 <- pbiv_NB(z13, lm1, lm3, betaa);
        f5 <- dbinom(xt2[v]-k23, xtminus1_2[v], alp2);
        f6 <- dbinom(xt3[v]-s23, xtminus1_3[v], alp3);
        z23 <- matrix(c(k23,s23), ncol=2);
        f23 <- pbiv_NB(z23, lm2, lm3, betaa);
        p1 <- c(p1,sum(f1*f2*f12));
        p2 <- c(p2,sum(f3*f4*f13));
        p3 <- c(p3,sum(f5*f6*f23))
}       ## end for
soma <- sum(log(p1)+log(p2)+log(p3), na.rm=T);
return(-soma)
}       ## end function
```

```
###################################
## Parameters: unconst to const ##
###################################
param_cl <- function(parvect){
A1 <- exp(parvect[1])/(1+exp(parvect[1]));
A2 <- exp(parvect[2])/(1+exp(parvect[2]));
A3 <- exp(parvect[3])/(1+exp(parvect[3]));
L1 <- exp(parvect[4]);
L2 <- exp(parvect[5]);
L3 <- exp(parvect[6]);
beti <- exp(parvect[7]);
params <- c(A1,A2,A3,L1,L2,L3,beti);
return(params)
}        ## end function

###################################
## Parameters: const to unconst ##
###################################
param_cl_inv <- function(params){
A1i <- log(params[1]/(1-params[1]));
A2i <- log(params[2]/(1-params[2]));
A3i <- log(params[3]/(1-params[3]));
L1i <- log(params[4]);
L2i <- log(params[5]);
L3i <- log(params[6]);
betii <- log(params[7]);
parvect <- c(A1i,A2i,A3i,L1i,L2i,L3i,betii);
return(parvect)
}        ## end function
```

```
###############################################
## Auxiliary function for composite logL ##
###############################################
cloglik_unconst <- function(parvect, v, X, N){
pars <- param_cl(parvect);
resi <- cloglik_v(pars, v, X, s, N);
return(resi)
}           ## end function

######################
## Estimation: CL ##
######################
estimComp_N1 <- function(X, par7){
pars_Comp_N1 <- array(0,dim=c(4,7));
parvect <- array(0,dim=c(4,7));
for(v in 1:s){
           parvect[v,] <- param_cl_inv(par7[v,]);
           resC_N1 <- optim(par=parvect[v,], f=cloglik_unconst, v=v,
                            X=X, method = "BFGS");
           paramvect <- resC_N1$par;
           pars_Comp_N1[v,] <- param_cl(paramvect)
}           ## end for
return(pars_Comp_N1)
}           ## end function

########################################################################
```

D.2 R functions related to Chapter 3

```
#######################################
## Generate S-PINAR(1) (univariate) ##
#######################################
gera_inarSign <- function(n, s, N, alfaS, lambS, tauS){
# s - nº periods
# N - nº cycles
# n - nº observ total ; n=sN
# alfaS, lambS, tauS - vectors
nobs <- n+1;
x <- rep(0,nobs);
part1_thin <- rep(0,nobs);
part2_inov <- rep(0,nobs);
alfa_matx <- array(0,dim=c(n,1));
lamb_matx <- array(0,dim=c(n,1));
for(j in 1:1) {
        alfa_matx[,j] <- rep(alfaS[,j],N);
        lamb_matx[,j] <- rep(lambS[,j],N)
}       ## end for
alfa_aux <- array(0,dim=c(n+1,1));
lamb_aux <- array(0,dim=c(n+1,1));
alfa_aux[2:(n+1),] <- alfa_matx;
lamb_aux[2:(n+1),] <- lamb_matx;
tau_vec <- rep(tauS, N);
tau_aux <- array(0,dim=c(n+1,1));
tau_aux[2:(n+1),1] <- tau_vec;
x[1] <- -3;
```

```
for(v in 2:(n+1)){
        part1_thin[v] <- sign(x[v-1])*(rbinom(1, 2*abs(x[v-1]),
                        alfa_aux[v,1])-abs(x[v-1]));
        part2_inov[v] <- rskellam(1, lamb_aux[v,1], tau_aux[v,1]);
        x[v] <- part1_thin[v] + part2_inov[v]
}       ## end for
return(x[2:(n+1)])
}       ## end function

####################################
## Function for admissible values ##
####################################
fun_alfa <- function(alfaC){
all((alfaC > 0) & (alfaC < 1))
}       ## end function

fun_lamb <- function(lambC){
all(lambC > 0)
}       ## end function

##################################################
## Estimation: conditional least squares (CLS) ##
##################################################
estimCLS_Skellam <- function(X, s, N){
d <- length(X);
aux <- rep(0,d+1);
aux[2:(d+1)] <- X;
aux1 <- rep(0,d);
aux1[2:d] <- X[2:d];
```

```
Ni <- rep(N,s);
Ni[1] <- N-1;
## Step 1: parameters alfa_LS and ksi_LS
#########################################
alfa_LS <- rep(0,s);
ksi_LS <- rep(0,s);
part <- array(0,dim=c(s,4));
for(v in 1:s){
        part[v,2] <- sum(aux1[v+s*(0:(N-1))]);
        part[v,4] <- sum(aux[v+s*(0:(N-1))]);
        part[v,3] <- sum(aux[v+s*(0:(N-1))]^2);
        part[v,1] <- sum(aux[v+1+s*(0:(N-1))]*aux[v+s*(0:(N-1))])
}       ## end for
for(v in 1:s){
        num <- Ni[v]*part[v,1]-part[v,2]*part[v,4];
        den <- Ni[v]*part[v,3]-(part[v,4])^2;
        alfa_LS[v] <- (num/den+1)/2;
        ksi_LS[v] <- (part[v,2]-(2*alfa_LS[v]-1)*part[v,4])/Ni[v]
}       ## end for
## Step 2: parameters sigma2_LS, lamb_LS and tau_LS
####################################################
sigma2_LS <- rep(0,s);
lamb_LS <- rep(0,s);
tau_LS <- rep(0,s);
meanpred_error <- rep(0,d);    ## mean prediction error
aux_ksi <- rep(ksi_LS,N);
for(v in 1:s){
        meanpred_error[v+s*(0:(N-1))] <- aux1[v+s*(0:(N-1))]-(2*alfa_LS[v]-1)*
                aux[v+s*(0:(N-1))] - aux_ksi[v+s*(0:(N-1))];
```

```
          sigma2_LS[v] <- (sum((meanpred_error[v+s*(0:(N-1))]^2))-2*alfa_LS[v]*
                    (1-alfa_LS[v])*sum(abs(aux[v+s*(0:(N-1))]))))/Ni[v];
          lamb_LS[v] <- (sigma2_LS[v]+ksi_LS[v])/2;
          tau_LS[v] <- (sigma2_LS[v]-ksi_LS[v])/2
}         ## end for
if(fun_alfa(alfa_LS) & fun_lamb(lamb_LS) & fun_lamb(tau_LS)){
          alfa_LS <- alfa_LS;
          lamb_LS <- lamb_LS;
          tau_LS <- tau_LS
          }else{
                    alfa_LS <- rep(NA,s);
                    lamb_LS <- rep(NA,s);
                    tau_LS <- rep(NA,s)
}         ## end if
parvect <- cbind(alfa_LS, lamb_LS, tau_LS);
return(parvect)
}         ## end function

###########################################################
## Functions for conditional maximum likelihood (CML) ##
###########################################################

#######################
## Transition prob. ##
#######################
prob_trans<-function(xt_1, xt, pars_v){
# xt_1, xt - vectors ; pars_v =(alfaS, lambS, tauS)  one season
d1 <- length(xt);
prob <- rep(0,d1);
```

```
a <- 0;
b <- 0;
if(fun_alfa(pars_v[1])& fun_lamb(pars_v[2:3])){
for(v in 1:d1){
        soma <- 0;
        a <- xt_1[v];
        b <- xt[v];
        for(r in (-abs(a)):abs(a)){
                binMod <- dbinom(abs(a)+sign(a)*r, 2*abs(a), pars_v[1]);
                skel <- dskellam(b-r, pars_v[2], pars_v[3]);
                soma <- soma + binMod*skel
        }       ## end for
        prob[v] <- soma
}       ## end for
}       ## end if
return(prob)
}       ## end function

######################################
## CM Log-likelihood for v=1,...,s ##
######################################
loglik_S <- function(pars_v, v, s, N, X){
# pars_v =(alfaS, lambS, tauS)
logk <- 0;
if(v==1){
        xt_1 <- X[v-1+s*(1:(N-1))];
        xt <- X[v+s*(1:(N-1))]
        } else{
                xt_1<-X[v-1+s*(0:(N-1))];
```

```
            xt<-X[v+s*(0:(N-1))]
}       ## end if
logk <- logk + sum(log(prob_trans(xt_1, xt, pars_v)));
logk <- -logk;
return(logk)
}       ## end function

#####################
## Estimation: CML ##
#####################
estimCML_S1 <- function(X, parvec){
pars_CML_S1 <- array(0,dim=c(4,3));
for(v in 1:s){
        result_S1 <- optim(par=parvec[v,], f=loglik_S, v=v,
                X=X, method = "BFGS");
        pars_CML_S1[v,] <- result_S1$par
}       ## end for
return(pars_CML_S1)
}       ## end function

#################################################################
```

Bibliography

Aknouche, A. (2007). Causality conditions and autocovariance calculations in PVAR models. *Journal of Statistical Computation and Simulation*, 77(9):769–780.

Al-Osh, M. A. and Aly, E.-E. A. (1992). First–order autoregressive time series with negative binomial and geometric marginals. *Communications in Statistics–Theory and Methods*, 21(9):2483–2492.

Al-Osh, M. A. and Alzaid, A. A. (1987). First–order integer–valued autogressive (INAR(1)) process. *Journal of Time Series Analysis*, 8(3):261–275.

Alzaid, A. A. and Al-Osh, M. A. (1988). First–order integer–valued autoregressive (INAR(1)) process: distributional and regression properties. *Statistica Neerlandica*, 42(1):53–61.

Alzaid, A. A. and Al-Osh, M. A. (1990). An integer–valued pth–order autoregressive structure (INAR(p)) process. *Journal of Applied Probability*, 27(2):314–324.

Alzaid, A. A. and Al-Osh, M. A. (1993). Some autoregressive moving average processes with generalized Poisson marginal distributions. *Annals of the Institute of Statistical Mathematics*, 45(2):223–232.

Alzaid, A. A. and Omair, M. A. (2010). On the Poisson difference distribution inference and applications. *Bulletin of the Malaysian Mathematical Sciences Society*, 8(33):17–45.

Alzaid, A. A. and Omair, M. A. (2012). An extended binomial distribution with applications. *Communications in Statistics–Theory and Methods*, 41(19):3511–3527.

Alzaid, A. A. and Omair, M. A. (2014). Poisson difference integer–valued autoregressive model of order one. *Bulletin of the Malaysian Mathematical Sciences Society*, 2(37):465–485.

Anderson, P. L. and Meerschaert, M. M. (2005). Parameter estimation for periodically stationary time series. *Journal of Time Series Analysis*, 26(4):489–518.

Anderson, P. L., Meerschaert, M. M., and Zhang, K. (2013). Forecasting with prediction intervals for periodic autoregressive moving average models. *Journal of Time Series Analysis*, 34(2):187–193.

Andersson, J. and Karlis, D. (2010). Treating missing values in INAR(1) models: An application to syndromic surveillance data. *Journal of Time Series Analysis*, 31(1):12–19.

Andersson, J. and Karlis, D. (2014). A parametric time series model with covariates for integers in Z. *Statistical Modelling*, 14(2):135–156.

Bakouch, H. S. (2010). Higher–order moments, cumulants and spectral densities of the NGINAR(1) process. *Statistical Methodology*, 7(1):1–21.

Bakouch, H. S., Kachour, M., and Nadarajah, S. (2016). An extended Poisson distribution. *Communications in Statistics–Theory and Methods*, 45(22):6746–6764.

Bakouch, H. S. and Ristić, M. M. (2010). Zero truncated Poisson integer–valued AR(1) model. *Metrika*, 72(2):265–280.

Barczy, M., Ispány, M., Pap, G., Scotto, M., and Silva, M. E. (2010). Innovational outliers in INAR(1) models. *Communications in Statistics–Theory and Methods*, 39(18):3343–3362.

Barczy, M., Ispány, M., Pap, G., Scotto, M., and Silva, M. E. (2012). Additive outliers in INAR(1) models. *Statistical Papers*, 53(4):935–949.

Barreto-Souza, W. (2015). Zero–modified geometric INAR(1) process for modelling count time series with deflation or inflation of zeros. *Journal of Time Series Analysis*, 36(6):839–852.

Barreto-Souza, W. and Bourguignon, M. (2015). A skew INAR(1) process on Z. *AStA–Advances in Statistical Analysis*, 99(2):189–208.

Basawa, I. V. and Lund, R. (2001). Large sample properties of parameter estimates for periodic ARMA models. *Journal of Time Series Analysis*, 22(6):651–663.

Basawa, I. V., Lund, R., and Shao, Q. (2004). First–order seasonal autoregressive processes with periodically varying parameters. *Statistics & Probability Letters*, 67(4):299–306.

Bentarzi, M. and Djeddou, L. (2014). On mixture periodic vector autoregressive models. *Communications in Statistics–Simulation and Computation*, 43(10):2325–2352.

Billingsley, P. (1961). *Statistical Inference for Markov Processes*. Statistical Research Monographs. University of Chicago Press.

Bisaglia, L. and Canale, A. (2016). Bayesian nonparametric forecasting for INAR models. *Computational Statistics & Data Analysis*, 100:70–78.

Bloomfield, P., Hurd, H. L., and Lund, R. B. (1994). Periodic correlation in stratospheric ozone data. *Journal of Time Series Analysis*, 15(2):127–150.

Boucher, J.-P., Denuit, M., and Guillen, M. (2008). Models of insurance claim counts with time dependence based on generalization of Poisson and negative binomial distributions. *Variance*, 2(1):135–162.

Boudreault, M. and Charpentier, A. (2011). Multivariate integer–valued autoregressive models applied to earthquake counts. *Submitted*.

Bourguignon, M. (2016). Poisson–geometric INAR(1) process for modeling count time series with overdispersion. *Statistica Neerlandica*, 70(3):176–192.

Bourguignon, M. and Vasconcellos, K. L. P. (2015). First–order non–negative integer–valued autoregressive processes with power series innovations. *Brazilian Journal of Probability and Statistics*, 29(1):71–93.

Bourguignon, M. and Vasconcellos, K. L. P. (2016). A new skew integer valued time series process. *Statistical Methodology*, 31:8–19.

Bourguignon, M., Vasconcellos, K. L. P., Reisen, V. A., and Ispány, M. (2016). A Poisson INAR(1) process with a seasonal structure. *Journal of Statistical Computation and Simulation*, 86(2):373–387.

Brännäs, K. and Hellström, J. (2001). Generalized integer–valued autoregression. *Econometric Reviews*, 20(4):425–443.

Brännäs, K. and Nordström, J. (2000). A bivariate integer valued allocation model for guest nights in hotels and cottages. *Umea Economic Studies 547*.

Brijs, T., Karlis, D., and Wets, G. (2008). Studying the effect of weather conditions on daily crash counts using a discrete time series model. *Accident Analysis & Prevention*, 40(3):1180–1190.

Bu, R. and McCabe, B. (2008). Model selection, estimation and forecasting in INAR(p) models: a likelihood–based Markov chain approach. *International Journal of Forecasting*, 24(1):151–162.

Bu, R., McCabe, B., and Hadri, K. (2008). Maximum likelihood estimation of higher–order integer–valued autoregressive processes. *Journal of Time Series Analysis*, 29(6):973–994.

Bulla, J., Chesneau, C., and Kachour, M. (2015). On the bivariate Skellam distribution. *Communications in Statistics–Theory and Methods*, 44(21):4552–4567.

Bulla, J., Chesneau, C., and Kachour, M. (2016). A bivariate first–order signed integer–valued autoregressive process. *Communications in Statistics–Theory and Methods*, (accepted).

Cameron, A. C. and Trivedi, P. K. (2013). *Regression Analysis of Count Data*, volume 53. Cambridge University Press.

Cheon, S., Song, S. H., and Jung, B. C. (2009). Tests for independence in a bivariate negative binomial model. *Journal of the Korean Statistical Society*, 38(2):185–190.

Chesneau, C. and Kachour, M. (2012). A parametric study for the first–order signed integer–valued autoregressive process. *Journal of Statistical Theory and Practice*, 6(4):760–782.

Chesneau, C., Kachour, M., and Karlis, D. (2015). On some distributions arising from a generalized trivariate reduction scheme. *Statistical Methodology*, 25:36–50.

Cox, D. R. (1981). Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics*, 8:93–115.

Cox, D. R. and Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika*, 91(3):729–737.

Davis, R. A., Holan, S. H., Lund, R., and Ravishanker, N. (2016). *Handbook of Discrete-Valued Time Series*. CRC Press.

Davis, R. A. and Yau, C. Y. (2011). Comments on pairwise likelihood in time series models. *Statistica Sinica*, pages 255–277.

Dion, J., Gauthier, G., and Latour, A. (1995). Branching processes with immigration and integer–valued time series. *Serdica Mathematical Journal*, 21(2):123–136.

Djordjević, M. S. (2016). An extension on INAR models with discrete Laplace marginal distributions. *Communications in Statistics–Theory and Methods*, (accepted).

Du, J.-G. and Li, Y. (1991). The integer–valued autoregressive (INAR(p)) model. *Journal of Time Series Analysis*, 12(2):129–142.

Duchesne, P. and Lafaye de Micheaux, P. (2013). Distributions for residual autocovariances in parsimonious periodic vector autoregressive models with applications. *Journal of Time Series Analysis*, 34(4):496–507.

Fernández-Fontelo, A., Cabaña, A., Puig, P., and Moriña, D. (2016). Under–reported data analysis with INAR–hidden Markov chains. *Statistics in Medicine*, 35(26):4875–4890.

Franke, J. and Subba Rao, T. (1993). Multivariate first–order integer–valued autoregression. *Technical report, Universität Kaiserslautern.*

Franses, P. H. and Paap, R. (2004). *Periodic Time Series Models*. Oxford University Press.

Freeland, R. K. (2010). True integer value time series. *AStA–Advances in Statistical Analysis*, 94(3):217–229.

Freeland, R. K. and McCabe, B. (2005). Asymptotic properties of CLS estimators in the Poisson AR(1) model. *Statistics & Probability Letters*, 73(2):147–153.

Freeland, R. K. and McCabe, B. P. M. (2004a). Analysis of low count time series data by Poisson autoregression. *Journal of Time Series Analysis*, 25(5):701–722.

Freeland, R. K. and McCabe, B. P. M. (2004b). Forecasting discrete valued low count time series. *International Journal of Forecasting*, 20(3):427–434.

Gardner, W. A., Napolitano, A., and Paura, L. (2006). Cyclostationarity: Half a century of research. *Signal Processing*, 86(4):639–697.

Genest, C. and Mesfioui, M. (2014). Bivariate extensions of Skellam's distribution. *Probability in the Engineering and Informational Sciences*, 28:401–417.

Genest, C. and Nešlehová, J. (2007). A primer on copulas for count data. *Astin Bulletin*, 37(2):475–515.

Ghodsi, A., Shitan, M., and Bakouch, H. S. (2012). A first–order spatial integer–valued autoregressive SINAR(1,1) model. *Communications in Statistics–Theory and Methods*, 41(15):2773–2787.

Gladyshev, E. G. (1961). Periodically correlated random sequences. *Soviet Mathematics*, 2:385–388.

Gladyshev, E. G. (1963). Periodically and almost–periodically correlated random processes with a continuous time parameter. *Theory of Probability & Its Applications*, 8(2):173–177.

Gomes, D. and Canto e Castro, L. (2009). Generalized integer–valued random coefficient for a first–order structure autoregressive (RCINAR) process. *Journal of Statistical Planning and Inference*, 139(12):4088–4097.

Harville, D. A. (2008). *Matrix Algebra from a Statistician's Perspective*. Springer.

Heinen, A. and Rengifo, E. (2007). Multivariate autoregressive modeling of time series count data using copulas. *Journal of Empirical Finance*, 14(4):564–583.

Hipel, K. W. and McLeod, A. I. (1994). *Time Series Modelling of Water Resources and Environmental Systems*, volume 45. Elsevier.

Hudecová, Š., Hušková, M., and Meintanis, S. G. (2015). Tests for time series of counts based on the probability–generating function. *Statistics*, 49(2):316–337.

Hurd, H. L. and Miamee, A. (2007). *Periodically Correlated Random Sequences: Spectral Theory and Practice*, volume 355. John Wiley & Sons.

Ilić, A. V. M. (2016). A geometric time series model with a new dependent Bernoulli counting series. *Communications in Statistics–Theory and Methods*, 45(21):6400–6415.

Inusah, S. and Kozubowski, T. J. (2006). A discrete analogue of the Laplace distribution. *Journal of Statistical Planning and Inference*, 136(3):1090–1102.

Irwin, J. O. (1937). The frequency distribution of the difference between two independent variates following the same Poisson distribution. *Journal of the Royal Statistical Society*, 100(3):415–416.

Jacobs, P. A. and Lewis, P. A. W. (1978a). Discrete time series generated by mixtures I: Correlational and runs properties. *Journal of the Royal Statistical Society: Series B*, 40(1):94–105.

Jacobs, P. A. and Lewis, P. A. W. (1978b). Discrete time series generated by mixtures II: Asymptotic properties. *Journal of the Royal Statistical Society: Series B*, 40(2):222–228.

Jacobs, P. A. and Lewis, P. A. W. (1983). Stationary discrete autoregressive moving average time series generated by mixtures. *Journal of Time Series Analysis*, 4(1):19–36.

Jazi, M. A., Jones, G., and Lai, C.-D. (2012a). First–order integer–valued AR processes with zero inflated Poisson innovations. *Journal of Time Series Analysis*, 33(6):954–963.

Jazi, M. A., Jones, G., and Lai, C.-D. (2012b). Integer valued AR(1) with geometric innovations. *Journal of Iranian Statistical Society*, 11(2):173–190.

Jia, B., Wang, D., and Zhang, H. (2014). A study for missing values in PINAR$(1)_T$ processes. *Communications in Statistics–Theory and Methods*, 43(22):4780–4789.

Joe, H. and Lee, Y. (2009). On weighting of bivariate margins in pairwise likelihood. *Journal of Multivariate Analysis*, 100(4):670–685.

Johnson, N. L., Kemp, A. W., and Kotz, S. (2005). *Univariate Discrete Distributions*, volume 444. John Wiley & Sons.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1997). *Discrete Multivariate Distributions*. New York: Wiley.

Jones, R. H. and Brelsford, W. M. (1967). Time series with periodic structure. *Biometrika*, 54(3-4):403–408.

Jose, K. K. and Mariyamma, K. D. (2016). A note on an integer–valued time series model with Poisson–negative binomial marginal distribution. *Communications in Statistics–Theory and Methods*, 45(1):123–131.

Jung, R. C., Ronning, G., and Tremayne, A. R. (2005). Estimation in conditional first-order autoregression with discrete support. *Statistical Papers*, 46(2):195–224.

Jung, R. C. and Tremayne, A. R. (2003). Testing for serial dependence in time series models of counts. *Journal of Time Series Analysis*, 24(1):65–84.

Jung, R. C. and Tremayne, A. R. (2006). Coherent forecasting in integer time series models. *International Journal of Forecasting*, 22(2):223–238.

Jung, R. C. and Tremayne, A. R. (2011). Useful models for time series of counts or simply wrong ones? *AStA–Advances in Statistical Analysis*, 95(1):59–91.

Kachour, M. (2014). On the rounded integer–valued autoregressive process. *Communications in Statistics–Theory and Methods*, 43(2):355–376.

Kachour, M. and Truquet, L. (2011). A p–order signed integer–valued autoregressive (SINAR(p)) model. *Journal of Time Series Analysis*, 32(3):223–236.

Kachour, M. and Yao, J.-F. (2009). First–order rounded integer–valued autoregressive (RINAR(1)) process. *Journal of Time Series Analysis*, 30(4):417–448.

Karlis, D. and Meligkotsidou, L. (2005). Multivariate Poisson regression with covariance structure. *Statistics and Computing*, 15(4):255–265.

Karlis, D. and Ntzoufras, I. (2006). Bayesian analysis of the differences of count data. *Statistics in Medicine*, 25(11):1885–1905.

Karlis, D. and Ntzoufras, I. (2009). Bayesian modelling of football outcomes: using the Skellam's distribution for the goal difference. *IMA Journal of Management Mathematics*, 20(2):133–145.

Karlis, D. and Pedeli, X. (2013). Flexible bivariate INAR(1) processes using copulas. *Communications in Statistics–Theory and Methods*, 42(4):723–740.

Karlis, D. and Xekalaki, E. (2005). Mixed Poisson distributions. *International Statistical Review*, 73(1):35–58.

Kedem, B. and Fokianos, K. (2005). *Regression Models for Time Series Analysis*, volume 488. John Wiley & Sons.

Kemp, A. W. (1997). Characterizations of a discrete normal distribution. *Journal of Statistical Planning and Inference*, 63(2):223–229.

Kemp, C. D. and Papageorgiou, H. (1982). Bivariate Hermite distributions. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 269–280.

Kim, H.-Y. and Park, Y. (2008). A non–stationary integer–valued autoregressive model. *Statistical Papers*, 49(3):485–502.

Klimko, L. A. and Nelson, P. I. (1978). On conditional least squares estimation for stochastic processes. *The Annals of Statistics*, 6(3):629–642.

Kocherlakota, S. and Kocherlakota, K. (1992). *Bivariate Discrete Distributions*. Statistics: Textbooks and Monographs, volume 132. Marcel Dekker, Inc.

Kozubowski, T. J. and Inusah, S. (2006). A skew Laplace distribution on integers. *Annals of the Institute of Statistical Mathematics*, 58(3):555–571.

Lai, C.-D. (2006). Constructions of discrete bivariate distributions. In *Advances in Distribution Theory, Order Statistics, and Inference*, pages 29–58. Springer.

Latour, A. (1997). The multivariate GINAR(p) process. *Advances in Applied Probability*, 29(1):228–248.

Latour, A. (1998). Existence and stochastic structure of a non–negative integer–valued autoregressive process. *Journal of Time Series Analysis*, 19(4):439–455.

Li, C., Wang, D., and Zhang, H. (2015). First–order mixed integer–valued autoregressive processes with zero–inflated generalized power series innovations. *Journal of the Korean Statistical Society*, 44(2):232–246.

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary Mathematics*, 80(1):221–39.

Liu, Y., Wang, D., Zhang, H., and Shi, N. (2016). Bivariate zero truncated Poisson INAR(1) process. *Journal of the Korean Statistical Society*, 45(2):260–275.

Lund, R. and Basawa, I. V. (2000). Recursive prediction and likelihood evaluation for periodic ARMA models. *Journal of Time Series Analysis*, 21(1):75–93.

Lund, R., Hurd, H., Bloomfield, P., and Smith, R. (1995). Climatological time series with periodic correlation. *Journal of Climate*, 8(11):2787–2809.

Lund, R., Shao, Q., and Basawa, I. (2006). Parsimonious periodic time series modeling. *Australian & New Zealand Journal of Statistics*, 48(1):33–47.

MacDonald, I. L. and Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete–Valued Time Series*, volume 110. CRC Press.

Maiti, R., Biswas, A., and Das, S. (2016). Coherent forecasting for count time series using Box–Jenkins's AR(p) model. *Statistica Neerlandica*, 70(2):123–145.

Marshall, A. W. and Olkin, I. (1990). *Multivariate distributions generated from mixtures of convolution and product families*, volume 16 of *Lecture Notes–Monograph Series*, pages 371–393. Institute of Mathematical Statistics.

McCabe, B. P. and Martin, G. M. (2005). Bayesian predictions of low count time series. *International Journal of Forecasting*, 21(2):315–330.

McCabe, B. P., Martin, G. M., and Harris, D. (2011). Efficient probabilistic forecasts for counts. *Journal of the Royal Statistical Society: Series B*, 73(2):253–272.

McKenzie, E. (1985). Some simple models for discrete time series. *Journal of the American Water Resources Association*, 21(4):645–650.

McKenzie, E. (1986). Autoregressive moving–average processes with negative–binomial and geometric marginal distributions. *Advances in Applied Probability*, 18(3):679–705.

McKenzie, E. (2003). Discrete Variate Time Series, Chapter 16. *Handbook of Statistics*, 21:573–606.

McLeod, A. I. (1993). Parsimony, model adequacy and periodic correlation in time series forecasting. *International Statistical Review*, 61(3):387–393.

Monteiro, M., Scotto, M. G., and Pereira, I. (2010). Integer–valued autoregressive processes with periodic structure. *Journal of Statistical Planning and Inference*, 140(6):1529–1541.

Monteiro, M., Scotto, M. G., and Pereira, I. (2015). *Dynamics, Games and Science: International Conference and Advanced School Planet Earth, DGS II, Portugal, August 28–September 6, 2013*, chapter A Periodic Bivariate Integer–Valued Autoregressive Model, pages 455–477. Springer International Publishing.

Moriña, D., Puig, P., Ríos, J., Vilella, A., and Trilla, A. (2011). A statistical model for hospital admissions caused by seasonal diseases. *Statistics in Medicine*, 30(26):3125–3136.

Nastić, A. S. (2014). On suitability of negative binomial marginals and geometric counting sequence in some applications of combined INAR(p) model. *Facta Universitatis, Series: Mathematics and Informatics*, 29(1):25–42.

Nastić, A. S. and Ristić, M. M. (2012). Some geometric mixed integer–valued autoregressive (INAR) models. *Statistics & Probability Letters*, 82(4):805–811.

Nastić, A. S., Ristić, M. M., and Bakouch, H. S. (2012). A combined geometric INAR(p) model based on negative binomial thinning. *Mathematical and Computer Modelling*, 55(5):1665–1672.

Nastić, A. S., Ristić, M. M., and Djordjević, M. S. (2016a). An INAR model with discrete Laplace marginal distributions. *Brazilian Journal of Probability and Statistics*, 30(1):107–126.

Nastić, A. S., Ristić, M. M., and Ilić, A. V. M. (2017). A geometric time series model with an alternative dependent Bernoulli counting series. *Communications in Statistics–Theory and Methods*, 46(2):770–785.

Nastić, A. S., Ristić, M. M., and Popović, P. M. (2016b). Estimation in a bivariate integer–valued autoregressive process. *Communications in Statistics–Theory and Methods*, 45(19):5660–5678.

Nelsen, R. B. (2007). *An Introduction to Copulas*. Springer Science & Business Media.

Ng, C. T. and Joe, H. (2014). Model comparison with composite likelihood information criteria. *Bernoulli*, 20(4):1738–1764.

Nikoloulopoulos, A. K. and Karlis, D. (2008). On modeling count data: a comparison of some well–known discrete distributions. *Journal of Statistical Computation and Simulation*, 78(3):437–457.

Nikoloulopoulos, A. K. and Karlis, D. (2010). Modeling multivariate count data using copulas. *Communications in Statistics–Simulation and Computation*, 39(1):172–187.

Nunes, A. N. (2012). Regional variability and driving forces behind forest fires in Portugal an overview of the last three decades (1980–2009). *Applied Geography*, 34:576–586.

Ong, S. H., Shimizu, K., and Ng, C. M. (2008). A class of discrete distributions arising from difference of two random variables. *Computational Statistics & Data Analysis*, 52(3):1490–1499.

Park, Y. and Oh, C. W. (1997). Some asymptotic properties in INAR(1) processes with Poisson marginals. *Statistical Papers*, 38(3):287–302.

Parzen, E. and Pagano, M. (1979). An approach to modeling seasonally stationary time series. *Journal of Econometrics*, 9(1):137–153.

Pavlopoulos, H. and Karlis, D. (2008). INAR(1) modeling of overdispersed count series with an environmental application. *Environmetrics*, 19(4):369–393.

Pedeli, X., Davison, A. C., and Fokianos, K. (2015). Likelihood estimation for the INAR(p) model by saddlepoint approximation. *Journal of the American Statistical Association*, 110(511):1229–1238.

Pedeli, X. and Karlis, D. (2011). A bivariate INAR(1) process with application. *Statistical Modelling*, 11(4):325–349.

Pedeli, X. and Karlis, D. (2013a). On composite likelihood estimation of a multivariate INAR(1) model. *Journal of Time Series Analysis*, 34(2):206–220.

Pedeli, X. and Karlis, D. (2013b). On estimation of the bivariate Poisson INAR process. *Communications in Statistics–Simulation and Computation*, 42(3):514–533.

Pedeli, X. and Karlis, D. (2013c). Some properties of multivariate INAR(1) processes. *Computational Statistics and Data Analysis*, 67(C):213–225.

Popović, P. M. (2015). Random coefficient bivariate INAR(1) process. *Facta Universitatis, Series: Mathematics and Informatics*, 30(3):263–280.

Popović, P. M. (2016). A bivariate INAR(1) model with different thinning parameters. *Statistical Papers*, 57(2):517–538.

Popović, P. M., Ristić, M. M., and Nastić, A. S. (2016). A geometric bivariate time series with different marginal parameters. *Statistical Papers*, 57(3):731–753.

Prékopa, A. (1952). On composed Poisson distributions, IV: Remarks on the theory of differential processes. *Acta Mathematica Hungarica*, 3(4):317–325.

Ristić, M. M., Bakouch, H. S., and Nastić, A. S. (2009). A new geometric first–order integer–valued autoregressive (NGINAR(1)) process. *Journal of Statistical Planning and Inference*, 139(7):2218–2226.

Ristić, M. M. and Nastić, A. S. (2012). A mixed INAR(p) model. *Journal of Time Series Analysis*, 33(6):903–915.

Ristić, M. M., Nastić, A. S., and Bakouch, H. S. (2012a). Estimation in an integer–valued autoregressive process with negative binomial marginals (NBINAR(1)). *Communications in Statistics–Theory and Methods*, 41(4):606–618.

Ristić, M. M., Nastić, A. S., and Ilić, A. V. M. (2013). A geometric time series model with dependent Bernoulli counting series. *Journal of Time Series Analysis*, 34(4):466–476.

Ristić, M. M., Nastić, A. S., Jayakumar, K., and Bakouch, H. S. (2012b). A bivariate INAR(1) time series model with geometric marginals. *Applied Mathematics Letters*, 25(3):481–485.

Roitershtein, A. and Zhong, Z. (2013). On random coefficient INAR(1) processes. *Science China Mathematics*, 56(1):177–200.

San-Miguel, J. and Camia, A. (2009). Forest fires at a glance: facts, figures and trends in the EU. *Living with wildfires: what science can tell us. A Contribution to the Science-Policy Dialogue, Joensuu: European Forest Institute.*

Sarnaglia, A. J. Q., Reisen, V. A., and Lévy-Leduc, C. (2010). Robust estimation of periodic autoregressive processes in the presence of additive outliers. *Journal of Multivariate Analysis*, 101(9):2168–2183.

Savani, V. and Zhigljavsky, A. A. (2007). Efficient parameter estimation for independent and INAR(1) negative binomial samples. *Metrika*, 65(2):207–225.

Schweer, S. and Weiß, C. H. (2014). Compound Poisson INAR(1) processes: Stochastic properties and testing for overdispersion. *Computational Statistics & Data Analysis*, 77:267–284.

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(6):590–618.

Scotto, M. G., Weiß, C. H., Silva, M. E., and Pereira, I. (2014). Bivariate binomial autoregressive models. *Journal of Multivariate Analysis*, 125:233–251.

Shao, Q. (2006). Mixture periodic autoregressive time series models. *Statistics & Probability Letters*, 76(6):609–618.

Shao, Q. (2008). Robust estimation for periodic autoregressive time series. *Journal of Time Series Analysis*, 29(2):251–263.

Shao, Q. and Ni, P. (2004). Least–squares estimation and ANOVA for periodic autoregressive time series. *Statistics & Probability Letters*, 69(3):287–297.

Silva, I. and Silva, M. E. (2006). Asymptotic distribution of the Yule–Walker estimator for INAR(p) processes. *Statistics & Probability Letters*, 76(15):1655–1663.

Silva, M. E. and Oliveira, V. L. (2004). Difference equations for the higher–order moments and cumulants of the INAR(1) model. *Journal of Time Series Analysis*, 25(3):317–333.

Silva, M. E. and Pereira, I. (2015). Detection of additive outliers in Poisson INAR(1) time series. In *Mathematics of Energy and Climate Change*, pages 377–388. Springer.

Skellam, J. G. (1946). The frequency distribution of the difference between two Poisson variates belonging to different populations. *Journal of the Royal Statistical Society*, 109(3):296.

Steutel, F. W. and van Harn, K. (1979). Discrete analogues of self–decomposability and stability. *The Annals of Probability*, 7(5):893–899.

Sun, J. and McCabe, B. P. (2013). Score statistics for testing serial dependence in count data. *Journal of Time Series Analysis*, 34(3):315–329.

Tang, M. and Wang, Y. (2014). Asymptotic behavior of random coefficient INAR model under random environment defined by difference equation. *Advances in Difference Equations*, 2014(1):1–9.

Tesfaye, Y. G., Meerschaert, M. M., and Anderson, P. L. (2006). Identification of periodic autoregressive moving average models and their application to the modeling of river flows. *Water Resources Research*, 42(W01419).

Turkman, K., Scotto, M. G., and de Zea Bermudez, P. (2014). *Non–Linear Time Series: Extreme Events and Integer Value Problems*. Springer.

Ursu, E. and Duchesne, P. (2009). On modelling and diagnostic checking of vector periodic autoregressive time series models. *Journal of Time Series Analysis*, 30(1):70–96.

Ursu, E. and Pereau, J.-C. (2014). Robust modelling of periodic vector autoregressive time series. *Journal of Statistical Planning and Inference*, 155:93–106.

Ursu, E. and Turkman, K. F. (2012). Periodic autoregressive model identification using genetic algorithms. *Journal of Time Series Analysis*, 33(3):398–405.

Varin, C. (2008). On composite marginal likelihoods. *AStA–Advances in Statistical Analysis*, 92(1):1–28.

Varin, C., Reid, N. M., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21(1):5–42.

Varin, C. and Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika*, 92(3):519–528.

Vecchia, A. V. (1985). Periodic autoregressive–moving average (PARMA) modeling with applications to water resources. *JAWRA-Journal of the American Water Resources Association*, 21(5):721–730.

Wang, D. and Zhang, H. (2011). Generalized RCINAR(p) process with signed thinning operator. *Communications in Statistics–Simulation and Computation®*, 40(1):13–44.

Weiß, C. H. (2008). Thinning operations for modeling time series of counts–a survey. *AStA–Advances in Statistical Analysis*, 92(3):319–341.

Weiß, C. H. (2013). Integer–valued autoregressive models for counts showing underdispersion. *Journal of Applied Statistics*, 40(9):1931–1948.

Weiß, C. H. (2015). A Poisson INAR(1) model with serially dependent innovations. *Metrika*, 78(7):829–851.

Weiß, C. H. and Kim, H.-Y. (2014). Diagnosing and modeling extra–binomial variation for time–dependent counts. *Applied Stochastic Models in Business and Industry*, 30(5):588–608.

Weiß, C. H. and Puig, P. (2015). The Marginal Distribution of Compound Poisson INAR(1) Processes. In *Stochastic Models, Statistics and Their Applications*, pages 351–359. Springer.

Zhang, H. and Wang, D. (2015). Inference for random coefficient INAR(1) process based on frequency domain analysis. *Communications in Statistics–Simulation and Computation*, 44(4):1078–1100.

Zhang, H., Wang, D., and Zhu, F. (2010). Inference for INAR(p) processes with signed generalized power series thinning operator. *Journal of Statistical Planning and Inference*, 140(3):667–683.

Zhang, H., Wang, D., and Zhu, F. (2012). Generalized RCINAR(1) process with signed thinning operator. *Communications in Statistics–Theory and Methods*, 41(10):1750–1770.

Zhao, Y. and Joe, H. (2005). Composite likelihood estimation in multivariate data analysis. *Canadian Journal of Statistics*, 33(3):335–356.

Zhao, Z. and Hu, Y. (2015). Statistical inference for first–order random coefficient integer–valued autoregressive processes. *Journal of Inequalities and Applications*, 2015(1):1–12.

Zheng, H., Basawa, I. V., and Datta, S. (2006). Inference for pth–order random coefficient integer–valued autoregressive processes. *Journal of Time Series Analysis*, 27(3):411–440.

Zheng, H., Basawa, I. V., and Datta, S. (2007). First–order random coefficient integer–valued autoregressive processes. *Journal of Statistical Planning and Inference*, 137(1):212–229.

Zhu, R. and Joe, H. (2003). A new type of discrete self–decomposability and its application to continuous–time Markov processes for modeling count data time series. *Stochastic Models*, 19(2):235–254.

Zhu, R. and Joe, H. (2006). Modelling count data time series with Markov processes based on binomial thinning. *Journal of Time Series Analysis*, 27(5):725–738.

Zhu, R. and Joe, H. (2010). Negative binomial time series models based on expectation thinning operators. *Journal of Statistical Planning and Inference*, 140(7):1874–1888.