Matheus Bartolo Guerrero

INTEGER-VALUED AUTOREGRESSIVE PROCESSES WITH PRE-ESTABLISHED MARGINALS AND INNOVATIONS: A NEW PERSPECTIVE ON COUNT TIME SERIES MODELING

Brazil

February, 2018

Matheus Bartolo Guerrero

INTEGER-VALUED AUTOREGRESSIVE PROCESSES WITH PRE-ESTABLISHED MARGINALS AND INNOVATIONS: A NEW PERSPECTIVE ON COUNT TIME SERIES MODELING

Master dissertation submitted to the Graduate Program in Statistics of the Universidade Federal de Minas Gerais as part of the requirements for obtaining the Master's Degree in Probability and Statistics.

Universidade Federal de Minas Gerais – UFMG

Departamento de Estatística

Programa de Pós-Graduação

Supervisor: Dr. Wagner Barreto de Souza

Brazil

February, 2018

Ficha catalográfica elaborada pela Biblioteca do ICEx - UFMG

Guerrero, Matheus Bartolo.

G935i Integer-valued autoregressive processes with preestablished marginals and innovations: a new perspective on count time series modeling / Matheus Bartolo Guerrero. — Belo Horizonte, 2018.
96 f.: il.; 29 cm.

Dissertação (mestrado) - Universidade Federal de Minas Gerais – Departamento de Estatística.

Orientador: Wagner Barreto de Souza

1. Estatística - Teses. 2. Análise de séries temporais. 3. Markov, Processos de. I. Orientador. II. Título.

CDU 519.2(043)

Matheus Bartolo Guerrero

INTEGER-VALUED AUTOREGRESSIVE PROCESSES WITH PRE-ESTABLISHED MARGINALS AND INNOVATIONS: A NEW PERSPECTIVE ON COUNT TIME SERIES MODELING

Master dissertation submitted to the Graduate Program in Statistics of the Universidade Federal de Minas Gerais as part of the requirements for obtaining the Master's Degree in Probability and Statistics.

Approved. Brazil, 23 of February of 2018:

Dr. Wagner Barreto de Souza
DEst/ICEx/UFMG (Supervisor)

Drª. Glaura da Conceição Franco
DEst/ICEx/UFMG

Dr. Roger William Câmara Silva
DEst/ICEx/UFMG

Dr. Rodrigo Bernardo da Silva
DE/CCEN/UFPB

Brazil
February, 2018

Este trabalho é dedicado a todos e a todas que lutam,
diuturnamente,
pela Ciência no Brasil.

Agradecimentos

O agradecimento principal é dedicado ao meu orientador Wagner Barreto de Souza. Suas ideias e sugestões formam a base deste trabalho. A liberdade que me concedera foi fundamental para que desempenhasse o meu melhor.

Um terno agradecimento é dedicado à minha família: Domingos, Leia, Lucas e Victória. Vocês são meu porto seguro.

Um doce e singelo agradecimento é dedicado à Tay. Sua presença em minha vida tem sido uma aventura encantadora.

Agradecimentos especiais são dedicados aos professores Enrico Antônio Colosimo, Marcos Oliveira Prates, Rosangela Helena Loschi, Sokol Ndreca e Wagner Barreto de Souza. Muito aprendi durante suas aulas; guardarei vários ensinamentos por toda a vida.

Um agradecimento carinhoso é dedicado à Rogéria F. Figueiredo. Sua dedicação ao trabalho é ímpar. Nossas conversas foram divertidíssimas.

Um agradecimento fraterno é dedicado à minha turma de mestrado: Augusto, Caio, Gabriel, Guilherme, Letícia, Magno e Rodolfo. Não poderia ter pedido melhor companhia para os momentos de estudo e para os de diversão.

Um agradecimento sincero é dedicado aos futuros doutores Rumenick e Uriel. O altruísmo e nobreza em compartilhar os seus conhecimentos serão para sempre lembrados.

Agradeço ainda ao amigo de longa data, Rafael Alves, o Dirla. “Boa gente, que vai à luta e segue sempre em frente”.

Por fim, agradeço ao CNPq pelo suporte financeiro.

"The saddest aspect of life right now
is that science gathers knowledge faster
than society gathers wisdom."
(Isaac Asimov, ⋆1919 - †1992)

Resumo

Séries temporais para dados de contagem é assunto recorrente na literatura científica devido sua aplicabilidade a diversas situações reais. O aprimoramento de métodos consagrados e o desenvolvimento de novas técnicas de modelagem para estas séries temporais são necessários e importantes, não só para a evolução teórica desta área da Estatística, como também para uma melhor representação da realidade enquanto modelo matemático-estatístico. Neste sentido, o presente trabalho propõe uma metodologia inovadora na modelagem dos processos autoregressivos de valores inteiros, conjuntamente pré-especificando em uma mesma família as distribuições marginais e inovações. O processo Autoregressivo de Valores Inteiros Duplo Geométrico de primeira ordem é apresentado e caracterizado através de diversas propriedades estatísticas. A inferência é realizada através dos métodos de mínimos quadrados condicionais, Yule-Walker e máxima verossimilhança; além disso, a consistência e normalidade assintótica dos estimadores são verificadas. Simulação computacional via métodos de Monte Carlo é empregada a fim de verificar a performance dos estimadores propostos. Aplicações a dados reais são exibidas comprovando a relevância prática do modelo desenvolvido; outrossim, a comparação com modelos concorrentes é exibida no intuito de corroborar a competitividade do modelo proposto.

Palavras-chave: Séries temporais de contagem; processo INAR; cadeia de Markov.

Abstract

Count time series is a recurring subject in the scientific literature due to its applicability to several real situations. The improvement of established methods and the development of new modeling techniques for these time series are necessary and important, not only for the theoretical evolution of this area of Statistics but also for a better representation of reality as a mathematical-statistical model. On this direction, the present work proposes an innovative methodology in the modeling of the integer autoregressive processes, jointly prespecifying in the same family the marginal distributions and innovations. The Double Geometric Integer Autoregressive process of first-order is presented and characterized by several statistical properties. The inference is performed through conditional least squares, Yule-Walker, and maximum likelihood. In addition, the consistency and asymptotic normality of the estimators are verified. Computational simulation via Monte Carlo methods is used to verify the performance of the proposed estimators. Applications to real datasets are given proving the practical relevance of the model developed. Moreover, the comparison with competing models is presented in order to corroborate the competitiveness of the proposed model.

Keywords: Count time series; INAR process; Markov chain.

List of Figures

List of Tables

List of abbreviations and acronyms

ACF	AutoCorrelation Function
AIC	Akaike Information Criterion
AR	AutoRegressive
ARMA	AutoRegressive Moving Average
BFGS	Broyden–Fletcher–Goldfarb–Shanno
BGD	Bivariate Geometric Distribution
CLS	Conditional Least Square
CV	Coefficient of Variation
DGINAR	Double Geometric INteger AutoRegressive
Eq.	Equation
iid	independent and identically distributed
INAR	INteger AutoRegressive
kurt.	kurtosis
MA	Moving Average
ML	Maximum Likelihood
NGINAR	New Geometric INteger AutoRegressive
PACF	Partial AutoCorrelation Function
pgf	probability generating function
PINAR	Poisson INteger AutoRegressive
pmf	probability mass function
RMS	Root Mean Square

RMSE	Root Mean Square Error
rv	random variable
skew.	skewness
wp	with probability
YW	Yule-Walker
ZMG	Zero-Modified Geometric
ZMNB	Zero-Modified Negative Binomial

Contents

1 Introduction

Integer-valued time series, better known as count time series, are massively present in everyday life, often as counts of events, objects or individuals in consecutive intervals or at consecutive points in time. Series of this kind arise naturally in many areas including queueing systems, finance, insurance theory, medicine, epidemiology, and others. Thus, it is not surprising that there is an increasing interest in modeling such count data.

There are several models to deal with time series, in special, in the class of linear models, we can cite the AutoRegressive Moving Average (ARMA) models (Box et al., 2015). However, these real-valued models are not sufficient to deal with integer-valued time series because, often, the multiplication of an integer by a real number results in a non-integer value. In order to maintain the good properties of the ARMA models, one may replace the usual multiplication by an operator that always lead to integer values. In this context, the pioneering INteger-valued AutoRegressive (INAR) process (McKenzie, 1985, 1988; Alzaid and Al-Osh, 1987) arose from the adoption of the binomial thinning operator (Steutel and van Harn, 1979) as a substitute for the usual multiplication and from the assumption of a Poisson distribution to the marginal distribution of the process. Following, Freeland and McCabe (2004a,b, 2005) discuss the inferential and forecasting aspects of the Poisson INAR process of first-order. Alzaid and Al-Osh (1990) and Du and Li (1991) present INAR processes with time dependence of order p, relating the INAR(p) and the AR(p) processes.

Since Poisson distribution is inadequate to accommodate under or overdispersion, McKenzie (1986) and Al-Osh and Aly (1992) introduced other INAR models, based on different distributions to the marginal of the process, as the negative binomial distribution and the geometric distribution.

All models mentioned above are based on the binomial thinning operator. As Borges et al. (2016) explained, the binomial thinning operator based on a counting series of Bernoulli distributed random variables is appropriate for modeling the number of random events, which may only survive or vanish after a period of observation (sum by 0 or 1). However, when the observed unit is capable of generating more counting objects

or producing more new random events, the Bernoulli random variable is no more the best choice for constructing the counting sequence. In this context, Ristić et al. (2009) introduced the New Geometric INAR process of first-order [in short, NGINAR(1) process] with geometric marginal distribution using the negative binomial thinning (Aly and Bouzar, 1994), which was based on a geometrically distributed counting sequence. Considering the nature of the distribution, this choice is more appropriate for modeling counting processes, which refer to population elements or random events capable of replication (sum by 0, 1, 2 or more).

As some natural phenomena demand that the parameters of the model evolve over time, Zheng et al. (2006), Zheng et al. (2007), Gomes and Canto e Castro (2009), Wang and Zhang (2011), and Zhao and Hu (2015) dealt with INAR processes with random coefficients. While Zheng et al. (2006), Zheng et al. (2007), and Zhao and Hu (2015) used the binomial thinning operator into their models, Gomes and Canto e Castro (2009) used a generalization of this operator. On the other hand, Wang and Zhang (2011) used the signed thinning operator, allowing the presence of negative integers in the count series. Recently, Awale et al. (2017) proposed a hypothesis test for testing if the parameters of the model vary across the time.

In this brief discussion, it is possible to note that there are many particularities in count data models, so it is important to highlight that there is still no unified approach. In order to understand these peculiarities, Weiß (2008) and Scotto et al. (2015) present good reviews. Among recent contributions on INAR models, Barreto-Souza and Bourguignon (2015), Bourguignon and Vasconcellos (2016) and Nastić et al. (2016) dealt with count series that include negative values and that accommodate skew marginals to the process. Barreto-Souza (2015), Maiti et al. (2015), Borges et al. (2016), and Bourguignon and Weiß (2017) faced problems such as overdispersion and as deflation, or inflation, of zeros. Scotto et al. (2017) proposed a max-INAR model to accommodate time series with sudden large counts, often caused by an extreme event, followed by monotone decreasing recovery phase. Barreto-Souza (2017) created a class of INAR(1) processes having mixed Poisson marginal distribution, introducing new INAR(1) processes such as the Poisson Inverse-Guassian INAR(1) model, which deals with overdispersion in a convenient and sophisticated way. The Bayesian approach on INAR models is found at McCabe and Martin (2005) and Bisaglia and Canale (2016), especially in terms of prediction.

Despite all of these different approaches in count series modeling, it does not exist in the literature an approach that jointly prespecifies in the same family the marginals and the innovations of the INAR processes. In this work, we follow this pioneering path and propose an INAR process with pre-established marginals and innovations selected in the same distribution family. This new modeling approach is more natural and intuitive to model real count data; in counting phenomena is easier to identify the distribution of the marginals and innovations than the distribution of the count series behind the thinning operators. Hence, in our approach, the count series is an innate consequence of the marginals and innovations distributions, the thinning operator arises naturally. Among the advantages of our approach are the analytic-mathematical simplifications, the unrestricted parameter space, and good statistical properties as the time reversibility.

The dissertation is organized as follows. In Chapter 2, a review of the basic concepts about time series and INAR processes is given. In Chapter 3, a new perspective on count time series modeling is presented. A new process is constructed and its main statistical properties are displayed. Estimators of the model parameters are obtained in Chapter 4, as well as its asymptotic distribution. Moreover, to evaluate the performance of the estimators, a Monte Carlo simulation is driven and discussed. In Chapter 5, real data examples are considered in order to show the relevance of our model. Concluding remarks and future research are discussed in Chapter 6.

2 Preliminaries

The basis for integer-valued time series modeling is the binomial thinning operator proposed by Steutel and van Harn (1979) and the INAR(1) process introduced by McKenzie (1985) and Alzaid and Al-Osh (1987). In this Chapter, these concepts are discussed and also elementary concepts about time series are presented.

2.1 Time Series and Related Concepts

A *time series* is a sequence of observations of a random variable (rv) taken sequentially in time, usually observed at equally spaced time points. Let T be an arbitrary set of indexes and let X be a rv, then we denote a time series by $\{X_t\}_{t\in T}$. Note that a time series process is a stochastic process indexed in time and regard to the nature of T we can have a discrete-time random process, $T \subseteq \mathbb{N}$, or a continuous-time random process, $T \subseteq \mathbb{R}$.

Additionally, unlike many statistical models in which the assumption of independent rvs is required, time series analysis is precisely concerned with describing this dependence among the elements of a sequence of rvs.

Time series can also be classified as discrete-valued (if X is a discrete rv) or continuous-valued (if X is a continuous rv). Examples of discrete-valued time series are the number of accidents occurred on the highways of a determined city during a month or the number of users of a specific website during a day. On the other hand, examples of continuous-valued time series are the price of a share on a stock exchange in a day and the temperature of the earth over the years.

In the present work, we focus on the non-negative integer-valued time series, i.e., time series where the rv X assumes values in a subset of the non-negative integers. In general, this type of time series deals with the count of events, objects or individuals, and this is the reason why we call it a count time series.

The next subsections present elementary concepts regarding time series and is strongly based on Prado and West (2010) and Box et al. (2015).

2.1.1 Stationarity

A time series is said to be *stationary* when it oscillates randomly around a constant mean over time, thus the behavior of a stationary time series process does not depend upon when the process is started. In other words, we say that the time series process is invariant under time translations, keeping the statistical equilibrium and its probabilistic characteristics. Here we provide two widely used definitions of stationarity.

Definition 2.1.1. A time series process $\{X_t\}_{t\in T}$ is *completely* or *strongly stationary* if all finite-dimensional distributions remain the same under time translations, that is,

$$\Pr(X_1 \leq c_1, \ldots, X_n \leq c_n) = \Pr(X_{1+h} \leq c_1, \ldots, X_{n+h} \leq c_n),$$

for any $t = 1, 2, \ldots, n+h$, with $n = 1, 2, \ldots$; for all $c_1, \ldots, c_n$, and for all translations $h = 0, \pm 1, \pm 2, \ldots$.

In particular, if a process is strongly stationary it means that all one-dimensional distributions are invariant under time translations, so the mean $E(X_t)$ and variance $Var(X_t)$ are constants, that is, $E(X_t) = \mu(t) = \mu$ and $Var(X_t) = \sigma^2(t) = \sigma^2 > 0$.

It is very difficult to verify that a process is strongly stationary and so, the notion of *weak* or *second order stationarity* arises as follow:

Definition 2.1.2. A time series process $\{X_t\}_{t\in T}$ is *weakly* or *second-order stationary* if and only if:

i) $E(X_t) = \mu(t) = \mu$.

ii) $E(X_t^2) < \infty$.

iii) $Cov(X_h, X_t) = \gamma_{t-h}$, depending only on the length $|t-h|$.

Note 2.1.3. In this work, when dealing with stationarity we are considering the case of weak stationarity.

2.1.2 Autocorrelation Function

In time series analysis, one of the initial steps is to study the correlation patterns displayed by the data at different time points. This can be done using the sample autocor-

relation function which is an estimator of the autocorrelation function (ACF). In order to present the definition of the autocorrelation, we first introduce the autocovariance.

Definition 2.1.4. Let $\{X_t\}_{t\in\mathbb{Z}}$ be a stationary time series process with mean $E(X_t) = \mu$. The autocovariance is a function that gives the covariance of the process with itself at pairs of time points, i.e.,

$$\gamma_h = Cov(X_t, X_{t+h}) = E[(X_t - \mu)(X_{t+h} - \mu)].$$

Note that $\gamma_0 = Var(X_t) = \sigma^2$.

Now we can define the autocorrelation.

Definition 2.1.5. The ACF is defined by

$$\rho_h = \frac{\gamma_h}{\gamma_0},\ h \in \mathbb{Z}.$$

Note that $\rho_0 = 1$.

The ACF has the following properties:

i) $-1 \leq \rho_h \leq 1$.

ii) $\rho_h = \rho_{-h}$.

So the ACF measures the linear dependence of the time series and it is also used to identify a suitable model for the time series.

2.2 Binomial Thinning Operator

To properly model integer-valued time series, classic ARMA models (Box et al., 2015) should not be used because, often, the multiplication of an integer by a real number results in a non-integer value. To circumvent this problem, we use the binomial thinning operator (Steutel and van Harn, 1979):

Definition 2.2.1. Let X be an arbitrary non-negative integer-valued rv, then for any $\alpha \in [0, 1]$ the binomial thinning operator "$\circ$" is defined as

$$\alpha \circ X = \sum_{i=1}^{X} Y_i,$$

where $\{Y_i\}$, called the count series, is a sequence of independent and identically distributed (iid) Bernoulli(α) rvs. Also, Y_i is independent of X, $\forall i$.

By convention, $\sum_{i=1}^{0} Y_i = 0$.

Due to the operator definition, we have the following properties:

i) $0 \circ X \overset{a.s.}{=} 0$ and $1 \circ X \overset{a.s.}{=} X$.

ii) $E(\alpha \circ X) = \alpha E(X)$.

iii) Associative Law: for any $\beta \in [0,1]$, $\beta \circ (\alpha \circ X) \overset{d}{=} (\beta\alpha) \circ X$.

iv) Distributive Law: if Z, independent of X, is a non-negative integer-valued rv, then

$$\alpha \circ (X + Z) \overset{d}{=} \alpha \circ X + \alpha \circ Z.$$

Note 2.2.2. "$\overset{a.s.}{=}$" means almost surely and "$\overset{d}{=}$" means equality in distribution.

The proofs of these four properties are provided in Appendix A.

The next result shows the reason the operator "$\circ$" is named a "binomial" operator.

Proposition 2.2.3. *The conditional distribution of $\alpha \circ X$ given $X = x$, is Binomial with parameters x and α.*

Proof. Since the rv X is independent of the sequence $\{Y_i\} \overset{iid}{\sim}$ Bernoulli(α), it follows that

$$\begin{aligned}
\Pr(\alpha \circ X = k | X = x) &= \Pr\left(\sum_{i=1}^{X} Y_i = k \middle| X = x\right) \\
&= \Pr\left(\sum_{i=1}^{x} Y_i = k\right) \left[\binom{x}{k} \text{ ways to obtain this summation}\right] \\
&= \binom{x}{k} \Pr(Y_1 = 1, \ldots, Y_k = 1, Y_{k+1} = 0, \ldots, Y_x = 0) \\
&= \binom{x}{k} \prod_{i=1}^{k} \Pr(Y_i = 1) \prod_{i=k+1}^{x} \Pr(Y_i = 0) \text{ [due to independence]} \\
&= \binom{x}{k} \alpha^k (1-\alpha)^{x-k} \text{ [Due to equality in distribution]},
\end{aligned}$$

which is the probability mass function (pmf) of a Binomial distribution with parameters x and α. □

To realize why the operator "$\circ$" is a "thinning" operator, consider a population of size X_t at a particular time t, e.g., the number of patients in a hospital. If we observe the size population at time $t+1$, then the population may have **shrunk**(in the meaning of the **thin** verb: make or become less dense, crowded, or numerous), because some of the patients may have received a discharge from the hospital or have died between times t and $t+1$. If the patients receive a discharge or die independently of each other, and if the probability of receiving a discharge or dying in between t and $t+1$ is equal to $1-\alpha$ for all patients, then the number of patients still hospitalized, at time $t+1$, is given by $\alpha \circ X_t$. Note that we disregard admissions of new patients to the hospital between times t and $t+1$.

2.3 Negative Binomial Thinning Operator

In spite of innovating the modeling of count time series, the binomial thinning operator has some limitations, for example, it is inadequate for count series with full range in $\mathbb{N}$ and to negative integers. Thus, some types of processes, such as those in which there is a replication of the phenomena modeled by the count series, are not attended by the binomial thinning operator. Some alternatives, reviewed and compiled by Weiß (2008), are the generalized thinning operator, the signed thinning operator, and others.

In the class of the generalized thinning operators, we can mention the negative binomial thinning operator, whose initial idea was presented by Aly and Bouzar (1994) and incorporated by Ristić et al. (2009) in its NGINAR model.

Definition 2.3.1. Let X be an arbitrary non-negative integer-valued rv. For any $\alpha \in [0,1)$, the negative binomial thinning operator "$*$" is defined by

$$\alpha * X = \sum_{i=1}^{X} W_i,$$

where the count series $\{W_i\}$ is a sequence of iid rvs with Geometric(α) distribution with parametrization

$$\Pr(W = w) = \frac{\alpha^w}{(1+\alpha)^{w+1}}, \quad w = 0, 1, 2, \ldots.$$

In this case, $E(W) = \alpha$ and $Var(W) = \alpha(1+\alpha)$.

Also, W_i is independent of X, $\forall i$.

Note 2.3.2. The operator "$*$" is not actually a "thinning" operator because $\alpha * X \leq X$ is not always true.

In the case where $X \sim \text{Geo}(\mu)$, $\mu > 0$, it is possible to show (Ristić et al., 2009) that the main differences between the binomial thinning operator $\circ$ and the negative binomial thinning operator $*$ are

- $0 \circ X = 0$ and $0 * X = 0$, but $1 \circ X = X$, while

$$1 * X \stackrel{d}{=} \begin{cases} 0 & \text{wp} \quad 1/(1+\mu) \\ X & \text{wp} \quad \mu^2/(1+\mu)^2, \\ X + Y & \text{wp} \quad \mu/(1+\mu)^2 \end{cases}$$

  where $Y \sim \text{Geo}\left((1+\mu)/(2+\mu)\right)$ is independent of X and where "wp" means with probability.

- $\beta \circ (\alpha \circ X) \stackrel{d}{=} (\beta\alpha) \circ X$, where counting sequences of "$\alpha\circ$" and "$\beta\circ$" are independent, but

$$\beta * (\alpha * X) \stackrel{d}{=} \begin{cases} 0 & \text{wp} \quad \frac{1+\alpha}{1+\alpha+\alpha\mu} \\ (\beta\alpha) * X + Y_1 & \text{wp} \quad \frac{\alpha^2\mu^2}{(1+\alpha+\alpha\mu)(1+\alpha\mu)}, \\ (\beta\alpha) * X + Y_2 & \text{wp} \quad \frac{\alpha\mu}{(1+\alpha+\alpha\mu)(1+\alpha\mu)} \end{cases}$$

  where Y_1 and Y_2 are independent and geometrically distributed with mean parameters $\beta\alpha$ and $\beta(1+\alpha+\alpha\mu)$, respectively, and are independent of X.

- $E(\alpha \circ X)^2 = \alpha^2 E(X^2) + \alpha(1-\alpha)E(X)$, similarly

$$E(\alpha * X)^2 = \alpha^2 E(X^2) + \alpha(1+\alpha)E(X).$$

2.4 Discrete Self-decomposability

Despite the introduction of the binomial thinning operator, the main goal of Steutel and van Harn (1979) was to propose analogues for the (continuous) concepts of self-decomposability and stability for distributions on the non-negative integers. In order to illustrate these concepts, we present a glimpse of discrete self-decomposable distributions.

Let $\varphi_Z(s) := E(s^Z) = \sum_{k=0}^{\infty} p_k s^k$ be the probability generating function (pgf) of a non-negative integer-valued rv Z, where $p_k = \Pr(Z = k)$, $k = 0, 1, 2, \ldots$. Note that the pgf

of Z is a power series representation of its pmf. Besides, $\varphi_Z(1) = 1$ so the series converges absolutely for $|s| \leq 1$. Also, $\varphi_Z(0) = p_0$. Finally, the most important property of the pgf is its uniqueness: if Z_1 and Z_2 have pgfs φ_{Z_1} and φ_{Z_2} respectively, then $\varphi_{Z_1}(s) = \varphi_{Z_2}(s)$, $\forall s$, if, and only if, $\Pr(Z_1 = k) = \Pr(Z_2 = k)$, for $k = 0, 1, 2, \ldots$. In addition, as the name of the pgf itself says, by taking derivatives of $\varphi_Z(s)$ it is possible to generate values from the pmf of the rv Z: $p_k = \varphi_Z(0)^{(k)}/k!$.

Example 2.4.1. Let $Y \sim \text{Bernoulli}(\alpha)$, $\alpha \in [0, 1]$. Then, the pgf of Y is given by

$$\varphi_Y(s) = E(s^Y) = s^0 \Pr(Y = 0) + s^1 \Pr(Y = 1) = 1 - \alpha + \alpha s, \quad |s| \leq 1.$$

Example 2.4.2. Let $X \sim \text{Geo}(\mu)$, $\mu > 0$. Hence, the pgf of X is

$$\varphi_X(s) = E(s^X) = \frac{1}{1+\mu} \sum_{x=0}^{\infty} \left(\frac{s\mu}{1+\mu} \right)^x = \frac{1}{1+\mu(1-s)}, \quad |s| < \frac{1+\mu}{\mu}.$$

The following definition is found in Steutel and van Harn (1979).

Definition 2.4.3. A distribution X on $\mathbb{N}$ with pgf φ_X is called discrete self-decomposable if

$$\varphi_X(s) = \varphi_X(1 - \alpha + \alpha s)\varphi_Y(s), \ |s| \leq 1, \ \forall \alpha \in (0, 1), \tag{2.1}$$

with φ_Y a pgf, and Y a rv with parameter(s) depending on α.

Next, we state a proposition in order to ensure self-decomposability to a discrete rv X.

Proposition 2.4.4. *Let X be a rv on $\mathbb{N}$ with the following stochastic representation:*

$$X \stackrel{d}{=} \alpha \circ X' + Y,$$

where $\alpha \circ X'$ and Y are independent, and X' is distributed as X. Then, X has a discrete self-decomposable distribution.

Proof. Just apply the pgf calculation in both sides of equality and use the fact that X and X' are identically distributed. □

Discrete self-decomposable distributions have remarkable and useful properties such as unimodality, closure under convolution and under weak convergence, and infinite divisibility. For more details, the reader may consult Steutel and van Harn (2003).

2.5 INAR(1) Process

The INAR(1) process was introduced by McKenzie (1985) and Alzaid and Al-Osh (1987) in order to model and generate sequences of dependent counting processes. In an autoregressive process of order p, what happens in a certain time t depends upon what happened in the p previous times. The INAR(1) process focuses on the time immediately preceding t and satisfies the following definition:

Definition 2.5.1. A INAR(1) process is a discrete time non-negative integer-valued stochastic process $\{X_t\}$, satisfying

$$X_t = \alpha \circ X_{t-1} + \varepsilon_t,\ t > 1, \tag{2.2}$$

where $\alpha \in [0, 1]$, $\circ$ represents the binomial thinning operation in Definition 2.2.1, and $\{\varepsilon_t\}$, called innovation, is a sequence of non-negative integer-valued iid rvs having mean μ_ε and finite variance σ^2_ε. Also, X_{t-h} is independent of ε_t, $\forall h \geq 1$.

In particular, the INAR(1) process is a discrete version of the linear AR process, copycatting its structure and correlation, with $\circ$ replacing the usual multiplication.

To illustrate the practical meaning behind Eq. (2.2), remember the exemplification of the binomial thinning operator regard to the population of patients at a hospital. We have at time t a population of size X_t and probability $1-\alpha$ of a patient leaves the hospital (by receiving a hospital discharge or dying). Now, consider that ε_t is the number of new patients arriving at the hospital at time t to be hospitalized. Then, clearly, at time t, X_t is a contribution of $\alpha \circ X_{t-1}$ (the number of remaining hospitalized patients from time $t-1$ to t) and a immigration/innovation (new patients being hospitalized) of size ε_t.

Alzaid and Al-Osh (1987) showed that the marginal distribution of the model in Eq. (2.2) can be expressed in terms of the arrival process, $\{\varepsilon_t\}$, as

$$X_t \overset{d}{=} \sum_{i=1}^{\infty} \alpha^i \varepsilon_{t-i},$$

where the dependence of $\{X_t\}$ on the sequence of innovations decays exponentially with the time lag. Note that there is a parallel with the Box and Jenkins' models, in which an AR(1) model can be represented by a moving average model MA(∞), satisfied some stationarity conditions.

Furthermore, it is possible to show that $\{X_t\}$ is markovian and a stationary process and hence if $\varphi_{X_t}(s)$ and $\varphi_{\varepsilon_t}(s)$ are the pgfs of the rvs X_t and ε_t, respectively, then

$$\begin{aligned}
\varphi_{X_t}(s) &= E\left(s^{X_t}\right) = E\left(s^{\alpha \circ X_{t-1}+\varepsilon_t}\right) \\
&= E\left(s^{\sum_{i=1}^{X_{t-1}} Y_i}\right) E\left(s^{\varepsilon_t}\right) \\
&= E\left[E\left(s^{\sum_{i=1}^{X_{t-1}} Y_i}\middle| X_{t-1}\right)\right] \varphi_{\varepsilon_t}(s) \\
&= E\left[\left(\varphi_Y(s)\right)^{X_{t-1}}\right] \varphi_{\varepsilon_t}(s) \\
&= \varphi_{X_{t-1}}\left(\varphi_Y(s)\right) \varphi_{\varepsilon_t}(s) \\
&= \varphi_{X_t}\left(\varphi_Y(s)\right) \varphi_{\varepsilon_t}(s) \\
&= \varphi_{X_t}\left(1-\alpha-\alpha s\right) \varphi_{\varepsilon_t}(s). \qquad (2.3)
\end{aligned}$$

Thus, the pgf of an INAR(1) process satisfies Eq. (2.1), i.e., the definition of a discrete self-decomposable distribution. Consequently, the marginal distribution of an INAR(1) process can be chosen from the class of discrete self-decomposable distributions. Among many others, this class contains the Poisson and Negative Binomial distributions as special cases.

Example 2.5.2 (PINAR(1))**.** An INAR(1) process given by Eq. (2.2), where the marginal distributions are Poisson(μ), $\mu > 0$, and where the $\circ$ operator is the binomial thinning operator, is called a Poisson INAR(1) process [in short, PINAR(1) process]. Its innovations are such that $\{\varepsilon_t\} \overset{iid}{\sim} \text{Poisson}\left(\mu(1-\alpha)\right)$, $0 < \alpha < 1$.

Example 2.5.3 (NGINAR(1))**.** In Eq. (2.2), if we replace the $\circ$ operator by the $*$ operator, given by Definition 2.3.1, and set $\{X_t\} \overset{iid}{\sim} \text{Geo}(\mu)$, $\mu > 0$, we have that the innovations, $\{\varepsilon_t\}$, are a mixture of two rvs with Geo(μ) and Geo(α) distributions, where $\alpha \in [0, \mu/(1+\mu)]$. Its pmf is given by

$$\Pr(\varepsilon_t = k) = \left(1 - \frac{\alpha\mu}{\mu-\alpha}\right) \frac{\mu^k}{(1+\mu)^{k+1}} + \frac{\alpha\mu}{\mu-\alpha} \cdot \frac{\alpha^k}{(1+\alpha)^{k+1}}, k = 0, 1, 2, \ldots, \forall t \geq 1.$$

This process is the NGINAR(1) process presented by Ristić et al. (2009).

3 A New Look at INAR Processes

In this Chapter, we present a new approach in the modeling of INAR processes by jointly prespecifying the marginals and innovations of the process in the same distribution family. From this perspective, a new type of INAR process is constructed and characterized, as well as its statistical properties are investigated.

3.1 Pre-established Marginals and Innovations

In general, autoregressive integer-valued time series, at a time t, can be represented by its stochastic form as

$$X_t = \boldsymbol{\theta} \star X_{t-1} + \varepsilon_t, \quad t > 1, \tag{3.1}$$

where

- $\{X_t\}$ is a stationary process with a non-negative discrete marginal distribution.
- $\star$ is a compounding operator, such as $\boldsymbol{\theta} \star X_{t-1} = \sum_{i=1}^{X_{t-1}} G_i$, where $\{G_i\}$ is a sequence of iid non-negative integer rvs with common distribution G, whose parameters depend upon the vector $\boldsymbol{\theta}$.
- $\{\varepsilon_t\}$ is a sequence of iid integer rvs, called innovations, independent of $\{G_i\}$.

Furthermore, X_{t-h} is independent of ε_t, $\forall h \geq 1$.

In the modeling of this kind of integer-valued time series, one usually chooses between the two most common paths. (1) Specify the marginal distribution of the process and the distribution of G, then obtain the distribution of the innovations. Or, (2) specify the distribution of the innovations and the distribution of G, then derive the marginal distribution of the process. In both cases, the calculations are made in order to keep the stationarity of the process.

In this work, what we propose is to choose a third path: jointly specify, in the same discrete distribution family, both the marginals and innovations of the process, then verify if there is a proper G distribution that keeps the stationarity of the process.

Note that, if the process in Eq. (3.1) is stationary, the pgf of $\{X_t\}$ is such that

$$\varphi_X(s) = \varphi_X\left(\varphi_G(s)\right)\varphi_\varepsilon(s). \tag{3.2}$$

Hence, if there exists $\varphi_X^{-1}(s)$, the inverse function of $\varphi_X(s)$, we have from Eq. (3.2) that

$$\varphi_G(s) = \varphi_X^{-1}\left(\frac{\varphi_X(s)}{\varphi_\varepsilon(s)}\right). \tag{3.3}$$

Therefore, since the distributions of X_t and ε_t are specified, Eq. (3.3) represents a mechanism to obtain the distribution of G through its pgf, as long as $\varphi_G(s)$ is a pgf from a proper discrete distribution.

Example 3.1.1. Let, $\forall t \geq 1$, $X_t \sim$ Poisson(μ) and $\varepsilon_t \sim$ Poisson$((1-\alpha)\mu)$, $\mu > 0$, $0 < \alpha < 1$. So,

$$\varphi_X(s) = e^{-\mu(1-s)},\ s \in \mathbb{R} \quad \Rightarrow \quad \varphi_X^{-1}(s) = 1 + \frac{1}{\mu}\ln(s),\ s > 0.$$

And, $\varphi_\varepsilon(s) = e^{-(1-\alpha)\mu(1-s)},\ s \in \mathbb{R}$.

Replacing these quantities in Eq. (3.3), we have that $\varphi_G(s) = 1 - \alpha + \alpha s$, $s \in \mathbb{R}$, which is the pgf of a Bernoulli distribution with parameter $\alpha \in (0, 1)$.

Note that in this case, the binomial thinning operator arises naturally.

Example 3.1.2. Consider that $\{X_t\} \overset{iid}{\sim}$ Geo(μ), $\mu > 0$, and that $\{\varepsilon_t\} \overset{iid}{\sim}$ Geo$((1-\alpha)\mu)$, $0 < \alpha < 1$. Remember that,

$$\Pr(X = x) = \frac{\mu^x}{(1+\mu)^{x+1}},\ x = 0, 1, 2, \ldots,$$

with $E(X_t) = \mu$ and $Var(X_t) := \sigma^2 = \mu(1+\mu)$.

Consider $E(\varepsilon_t) := \mu_\varepsilon = (1-\alpha)\mu$ and $Var(\varepsilon_t) := \sigma_\varepsilon^2 = \mu_\varepsilon(1+\mu_\varepsilon)$.

Note that,

$$\varphi_X(s) = \frac{1}{1+\mu(1-s)},\ |s| < \frac{1+\mu}{\mu} \quad \Rightarrow \quad \varphi_X^{-1}(s) = 1 - \frac{1}{\mu}\left(\frac{1-s}{s}\right). \tag{3.4}$$

Analogously,

$$\varphi_\varepsilon(s) = \frac{1}{1+\mu_\varepsilon(1-s)},\ |s| < \frac{1+\mu_\varepsilon}{\mu_\varepsilon}. \tag{3.5}$$

By replacing Eq. (3.4) and Eq. (3.5) in Eq. (3.3) we have

$$\varphi_G(s) = 1 - \frac{1}{\mu}\left(\frac{\varphi_\varepsilon(s)}{\varphi_X(s)} - 1\right) = 1 - \frac{1}{\mu}\left[\frac{1+\mu(1-s)}{1+(1-\alpha)\mu(1-s)} - 1\right] \Rightarrow$$

$$\varphi_G(s) = 1 - \frac{1}{\mu}\left[\frac{\alpha\mu(1-s)}{1+(1-\alpha)\mu(1-s)}\right] = \frac{1+[(1-\alpha)\mu-\alpha](1-s)}{1+(1-\alpha)\mu(1-s)}$$

$$= \frac{1+\left[1-\dfrac{\alpha}{(1-\alpha)\mu}\right](1-\alpha)\mu(1-s)}{1+(1-\alpha)\mu(1-s)}, \ |s| < \frac{1+\mu_\varepsilon}{\mu_\varepsilon}. \tag{3.6}$$

Therefore, Eq. (3.6) shows that $\varphi_G(s)$ is a pgf of a Zero-modified Geometric (ZMG) distribution with parameters

$$\pi = 1 - \frac{\alpha}{(1-\alpha)\mu} = 1 - \frac{\alpha}{\mu_\varepsilon} \quad \text{and} \quad \mu_\varepsilon,$$

both depending on the parameter vector $\boldsymbol{\theta} = (\mu, \alpha)$. We denote $G \sim \text{ZMG}(\pi, \mu_\varepsilon)$ and its pmf is given by

$$\Pr(G=k) = \begin{cases} \pi + (1-\pi)\left(\dfrac{1}{1+\mu_\varepsilon}\right), & \text{for } k = 0, \\ (1-\pi)\dfrac{\mu_\varepsilon^k}{(1+\mu_\varepsilon)^{k+1}}, & \text{for } k = 1, 2, \ldots. \end{cases}$$

The mean and variance of G are, respectively,

$$E(G) = \mu_\varepsilon(1-\pi) = \alpha,$$

and

$$Var(G) = \mu_\varepsilon(1-\pi)\left[1+\mu_\varepsilon(1+\pi)\right] = (1+2\mu)(1-\alpha)\alpha.$$

Note that for $\pi \in (-1/\mu_\varepsilon, 0)$ and $\pi \in (0,1)$, we have a zero-deflated model and a zero-inflated model with respect to the geometric distribution, respectively. The ZMG distribution has the geometric distribution as a special case by taking $\pi = 0$.

3.2 The DGINAR(1) Process

To build our new INAR process, we jointly prespecify the marginal distribution and innovations of the process as geometric distributions as stated in Example 3.1.2.

In this set up, we call the process $\{X_t\}$ given by Eq. (3.1) a Double Geometric INAR Process of first-order [in short, DGINAR(1) process].

Although this is just the specification phase of our model, we already can highlight an advantage when comparing with many models, specially with the model proposed by Ristić et al. (2009). In the DGINAR(1) process, the autocorrelation parameter α varies

freely in the interval $(0, 1)$, not being limited by any imposed constraint dependent on the mean parameter μ.

In order to characterize the DGINAR process, some properties of the $\star$ operator, associated with a ZMG distribution, are required. Proposition 3.2.1 lists those most useful and its proof is presented in Appendix B.

Proposition 3.2.1. *Let* $\{G_i\} \overset{iid}{\sim} G \sim ZMG(\pi, \mu_\varepsilon)$. *Let* X *and* Y *be discrete non-negative rvs, not necessarily independent of each other, but independent of* G_i, $\forall i$. *Then,*

i) $E(\boldsymbol{\theta} \star X) = E(G)E(X)$.

ii) $E((\boldsymbol{\theta} \star X)^2) = Var(G)E(X) + E^2(G)E(X^2)$.

iii) $E\left((\boldsymbol{\theta} \star X)Y\right) = E(G)E(XY)$.

iv) $Var(\boldsymbol{\theta} \star X) = Var(G)E(X) + E^2(G)Var(X)$.

v) $Cov(\boldsymbol{\theta} \star X, X) = E(G)Var(X)$.

In the context of INAR processes, one interest is to determine the conditional distribution of $\boldsymbol{\theta} \star X$ given $X = x$. Proposition 3.2.2 gives this distribution.

Proposition 3.2.2. *The conditional distribution of* $\boldsymbol{\theta} \star X$ *given* $X = x$ *is a Zero-modified Negative Binomial (ZMNB) distribution with parameters* π, μ_ε *and* $x \geq 1$. *Its pmf is given by*

$$\Pr\left(\boldsymbol{\theta} \star X = k | X = x\right) = \begin{cases} \pi_\star^x, \ k = 0, \\ \displaystyle\sum_{i=1}^{k} \binom{i+x-1}{i} \pi_\star^x (1-\pi_\star)^i \binom{k-1}{i-1} p^{k-i}(1-p)^i, \ k \geq 1, \end{cases}$$

where

$$\pi_\star = \pi + (1-\pi)\frac{1}{1+\mu_\varepsilon} = 1 - \frac{\alpha}{1+\mu_\varepsilon} \quad \textit{and} \quad p = \frac{\mu_\varepsilon - \alpha}{1+\mu_\varepsilon - \alpha}.$$

Furthermore,

$$E(\boldsymbol{\theta} \star X = k | X = x) = xE(G),$$

and

$$Var(\boldsymbol{\theta} \star X = k | X = x) = xVar(G).$$

We denote $(\boldsymbol{\theta} \star X | X = x) \sim ZMNB(\pi, \mu_\varepsilon, x)$.

Remark 3.2.3. The ZMNB distribution has the Negative Binomial distribution as a special case by taking $\pi = 0$.

Proof. Note that

$$\Pr\left(\boldsymbol{\theta} \star X = k | X = x\right) = \Pr\left(\sum_{i=1}^{X} G_i = k \middle| X = x\right) = \Pr\left(\sum_{i=1}^{x} G_i = k\right) = \Pr\left(S_x = k\right),$$

where $S_x = \sum_{i=1}^{x} G_i$.

We want to find the pmf of S_x which is the sum of x independent and equally distributed ZMG distributions. One way to do this is to compute the pgf of S_x.

$$\begin{aligned}\varphi_{S_x}(s) &= E(s^{S_x}) = E\left(s^{\sum_{i=1}^{x} G_i}\right) = E\left(\prod_{i=1}^{x} s^{G_i}\right) \\ &= \prod_{i=1}^{x} E(s^G) = (\varphi_G(s))^x = \left(\frac{1+\pi\mu_\varepsilon(1-s)}{1+\mu_\varepsilon(1-s)}\right)^x. \end{aligned} \tag{3.7}$$

Now, rearranging $\varphi_{S_x}(s)$ given by Eq. (3.7) in a convenient way and following the steps of Proposition 4.1 from Kolev et al. (2000) (after an adequate reparametrization), we obtain that

$$\varphi_{S_x}(s) = \left(\frac{1+\pi\mu_\varepsilon(1-s)}{1+\mu_\varepsilon(1-s)}\right)^x = \left(\frac{1+\pi\mu_\varepsilon}{1+\mu_\varepsilon}\right)^x \left(1 - \frac{(1-\pi)\mu_\varepsilon}{(1+\mu_\varepsilon)(1+\pi\mu_\varepsilon)} \frac{s}{1 - s\frac{\pi\mu_\varepsilon}{1+\pi\mu_\varepsilon}}\right)^{-x}.$$

Then, using the fact that $(1-z)^{-x} = \sum_{i=0}^{\infty} \binom{i+x-1}{i} z^i$, $|z| < 1$ and $x \geq 1$, we have that

$$\begin{aligned}\varphi_{S_x}(s) &= \left(\frac{1+\pi\mu_\varepsilon}{1+\mu_\varepsilon}\right)^x \sum_{i=0}^{\infty} \binom{i+x-1}{i} \left(\frac{(1-\pi)\mu_\varepsilon}{(1+\mu_\varepsilon)(1+\pi\mu_\varepsilon)} \frac{s}{1 - s\frac{\pi\mu_\varepsilon}{1+\pi\mu_\varepsilon}}\right)^i \\ &= \left(\frac{1+\pi\mu_\varepsilon}{1+\mu_\varepsilon}\right)^x + \left(\frac{1+\pi\mu_\varepsilon}{1+\mu_\varepsilon}\right)^x \times \\ &\quad \sum_{i=1}^{\infty} s^i \binom{i+x-1}{i} \left(\frac{(1-\pi)\mu_\varepsilon}{1+\mu_\varepsilon} \frac{1}{1+\pi\mu_\varepsilon}\right)^i \left(1 - s\frac{\pi\mu_\varepsilon}{1+\pi\mu_\varepsilon}\right)^{-i} \\ &= \left(\frac{1+\pi\mu_\varepsilon}{1+\mu_\varepsilon}\right)^x + \left(\frac{1+\pi\mu_\varepsilon}{1+\mu_\varepsilon}\right)^x \times \\ &\quad \sum_{i=1}^{\infty} s^i \binom{i+x-1}{i} \left(\frac{(1-\pi)\mu_\varepsilon}{1+\mu_\varepsilon} \frac{1}{1+\pi\mu_\varepsilon}\right)^i \sum_{j=0}^{\infty} \binom{j+i-1}{j} \left(s\frac{\pi\mu_\varepsilon}{1+\pi\mu_\varepsilon}\right)^j \\ &= \left(\frac{1+\pi\mu_\varepsilon}{1+\mu_\varepsilon}\right)^x + \left(\frac{1+\pi\mu_\varepsilon}{1+\mu_\varepsilon}\right)^x \times \end{aligned}$$

$$\sum_{i=1}^{\infty}\sum_{j=0}^{\infty} s^{i+j}\binom{i+x-1}{i}\binom{j+i-1}{j}\left(\frac{(1-\pi)\mu_\varepsilon}{1+\mu_\varepsilon}\frac{1}{1+\pi\mu_\varepsilon}\right)^i\left(\frac{\pi\mu_\varepsilon}{1+\pi\mu_\varepsilon}\right)^j$$

$$\overset{(I)}{=}\left(\frac{1+\pi\mu_\varepsilon}{1+\mu_\varepsilon}\right)^x+\left(\frac{1+\pi\mu_\varepsilon}{1+\mu_\varepsilon}\right)^x\times$$
$$\sum_{i=1}^{\infty}\sum_{j=0}^{\infty} s^{i+j}\binom{i+x-1}{i}\binom{j+i-1}{i-1}\left(\frac{(1-\pi)\mu_\varepsilon}{1+\mu_\varepsilon}\frac{1}{1+\pi\mu_\varepsilon}\right)^i\left(\frac{\pi\mu_\varepsilon}{1+\pi\mu_\varepsilon}\right)^j$$

$$\overset{(II)}{=}\left(\frac{1+\pi\mu_\varepsilon}{1+\mu_\varepsilon}\right)^x+\left(\frac{1+\pi\mu_\varepsilon}{1+\mu_\varepsilon}\right)^x\times$$
$$\sum_{k=1}^{\infty} s^k\sum_{i=1}^{k}\binom{i+x-1}{i}\binom{k-1}{i-1}\left(\frac{(1-\pi)\mu_\varepsilon}{1+\mu_\varepsilon}\frac{1}{1+\pi\mu_\varepsilon}\right)^i\left(\frac{\pi\mu_\varepsilon}{1+\pi\mu_\varepsilon}\right)^{k-i}$$

$$=\left(\frac{1+\pi\mu_\varepsilon}{1+\mu_\varepsilon}\right)^x+\left(\frac{1+\pi\mu_\varepsilon}{1+\mu_\varepsilon}\right)^x\times$$
$$\sum_{k=1}^{\infty} s^k\sum_{i=1}^{k}\binom{k-1}{i-1}\binom{i+x-1}{i}\left(\frac{(1-\pi)\mu_\varepsilon}{1+\mu_\varepsilon}\right)^i\left(\frac{1}{1+\pi\mu_\varepsilon}\right)^i\left(\frac{\pi\mu_\varepsilon}{1+\pi\mu_\varepsilon}\right)^{k-i}$$

$$=\pi_\star^x+\pi_\star^x\sum_{k=1}^{\infty} s^k\sum_{i=1}^{k}\binom{i+x-1}{i}(1-\pi_\star)^i\binom{k-1}{i-1}p^{k-i}(1-p)^i$$

$$=\pi_\star^x+\sum_{k=1}^{\infty} s^k\Pr(S_x=k),$$

where for $k=1,2,\ldots$, $\Pr(S_x=k)=\sum_{i=1}^{k}\binom{i+x-1}{i}\pi_\star^x(1-\pi_\star)^i\binom{k-1}{i-1}p^{k-i}(1-p)^i$.

It is worth mentioning that the expression here obtained for the pmf of the ZMNB distribution is simpler than that obtained by Kolev et al. (2000).

Notes:

(I) In this step we use the fact that $\binom{j+i-1}{j}=\binom{j+i-1}{i-1}$.

(II) Use the following result:

Lemma 3.2.4. $\sum_{i=1}^{\infty}\sum_{j=0}^{\infty} s^{i+j}a_ib_j=\sum_{k=1}^{\infty} s^kc_k$, *where* $c_k=\sum_{i=1}^{k} a_ib_{k-i}$.

Proof. Put $k=i+j$. Note that $k=1,2,\ldots,\infty$ and so $1\le i\le k$.

Thus, replacing j by k, we have that

$$\sum_{i=1}^{\infty}\sum_{j=0}^{\infty} s^{i+j}a_ib_j=\sum_{k=1}^{\infty}\sum_{i=1}^{k} s^ka_ib_{k-i}=\sum_{k=1}^{\infty} s^k\underbrace{\sum_{i=1}^{k} a_ib_{k-i}}_{c_k}.$$

To obtain the mean and the variance of S_x, note that

$$E(S_x) = x\mu_\varepsilon(1-\pi) = xE(G)$$

and

$$Var(S_x) = x\mu_\varepsilon(1-\pi)\left[1+\mu_\varepsilon(1+\pi)\right] = xVar(G).$$

□

Referent to the unconditional distribution of $\boldsymbol{\theta} \star X$, once $\varphi_{\boldsymbol{\theta}\star X}(s) = \varphi_X\left(\varphi_G(s)\right)$ and replacing Eq. (3.4) and Eq. (3.6), we have that

$$\begin{aligned}\varphi_{\boldsymbol{\theta}\star X}(s) &= \frac{1}{1+\mu\left(1-\frac{1+(1-\alpha/\mu_\varepsilon)\mu_\varepsilon(1-s)}{1+\mu_\varepsilon(1-s)}\right)} = \frac{1}{1+\frac{\mu\alpha(1-s)}{1+(1-\alpha)\mu(1-s)}} \\ &= \frac{1+(1-\alpha)\mu(1-s)}{1+\mu(1-s)},\end{aligned}$$

which implies that $\boldsymbol{\theta} \star X \sim \text{ZMG}(1-\alpha, \mu)$. Thus,

$$E(\boldsymbol{\theta} \star X_l) = \mu - \mu_\varepsilon \text{ and } Var(\boldsymbol{\theta} \star X_l) = \sigma^2 - \sigma_\varepsilon^2.$$

Another interesting feature of our process is that the ZMG distribution is maintained for associative operations.

Proposition 3.2.5. *Let $X \sim Geo(\mu)$, $\mu > 0$. For $\boldsymbol{\theta}_i = (\mu, \alpha_i)$, $0 < \alpha_i < 1$, set*

$$G_i \sim ZMG\left(\pi_i, \mu\right), \quad \pi_i = 1 - \frac{\alpha_i}{(1-\alpha_i)\mu}, \quad i = 1, 2.$$

Then,

$$\boldsymbol{\theta}_1 \star (\boldsymbol{\theta}_2 \star X) \sim ZMG(1-\alpha_1\alpha_2, \mu)$$

Proof. We use the fact that pgf of $\boldsymbol{\theta}_1 \star (\boldsymbol{\theta}_2 \star X)$ is such that

$$\begin{aligned}\varphi_{\boldsymbol{\theta}_1\star(\boldsymbol{\theta}_2\star X)}(s) &= \varphi_X\left(\varphi_{G_2}\left(\varphi_{G_1}\right)\right) \\ &= \frac{1+\mu(1-\alpha_1\alpha_2)(1-s)}{1+\mu(1-s)}\end{aligned}$$

which is the pgf of a ZMG distribution with parameters $1-\alpha_1\alpha_2$ and μ. □

Some simulated trajectories of the DGINAR(1) process are presented in Figure 1. We simulate the trajectories for different combinations of the mean parameter ($\mu = 1, 5$) with the autocorrelation parameter ($\alpha = 0.1, 0.3, 0.5, 0.7$). Note that in all scenarios, the

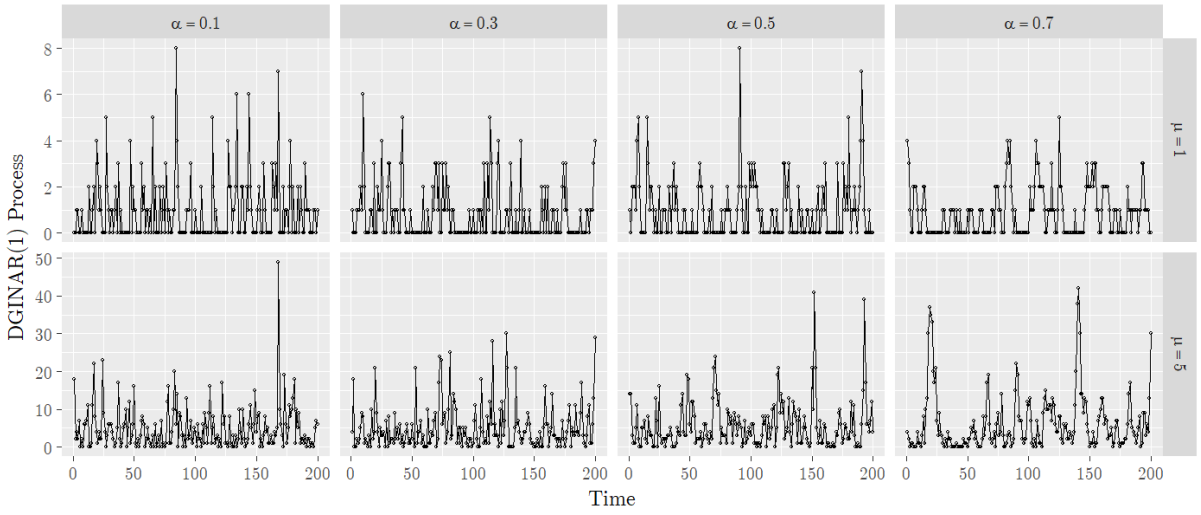

Figure 1 – Simulated trajectories of the DGINAR(1) process for $\mu = 1$ and $\mu = 5$, and for different values of α. Sample size is equal to 200.

series has its observed values distributed around its mean; the variance of the process seems to be constant over time.

Figure 2 gives plots of the sample ACF for each scenario, while Figure 3 presents the plots of the sample PACF. The ACF and the PACF characterize the series in terms of its dependency structure. In INAR models, the interpretation of the ACF and PACF are the same as in the ARMA models.

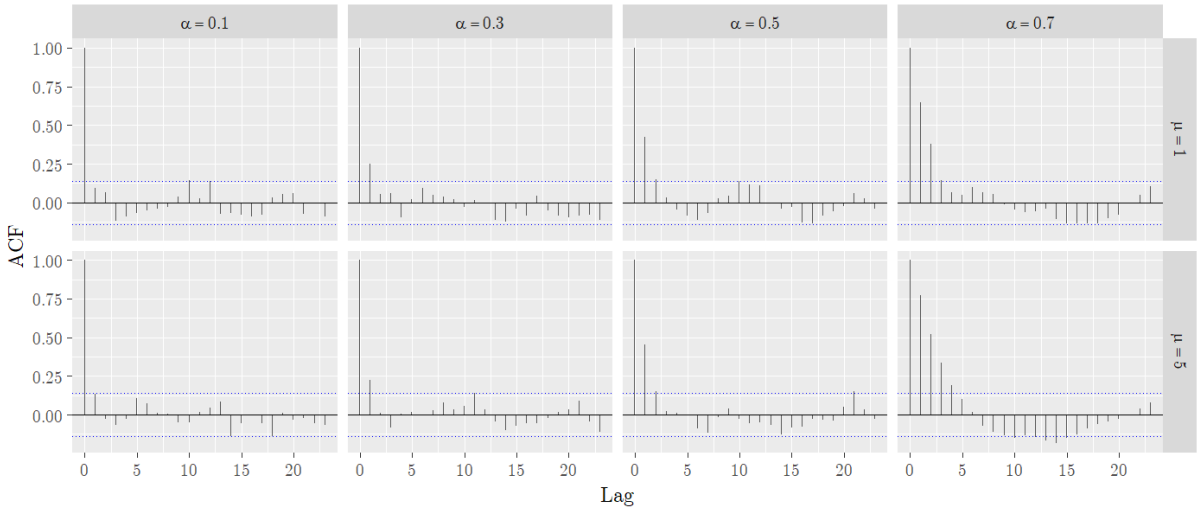

Figure 2 – Sample ACF for the simulated trajectories of the DGINAR(1) process for $\mu = 1$ and $\mu = 5$, and for different values of α.

In Figure 2, note that the ACF plots have a certain exponential decay at the autocorrelation values; it becomes more noticeable as α increases. This fact is because in

the DGINAR(1) process, the theoretical autocorrelation of order h is of the form $\gamma_h = \alpha^h$, as will be shown in Section 3.3.3.

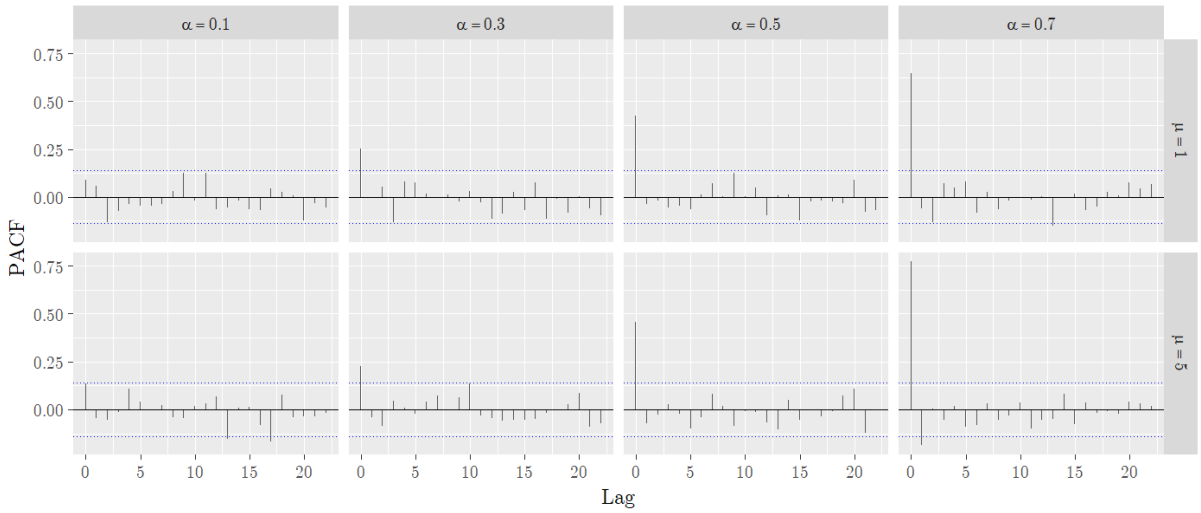

Figure 3 – Sample PACF for the simulated trajectories of the DGINAR(1) process for $\mu = 1$ and $\mu = 5$, and for different values of α.

Figure 3 shows that the PACF plots have a clear cut-off after lag 1, meaning that all higher-order autocorrelations are effectively explained by the lag 1 autocorrelation, what indicates an autoregressive structure of order 1 for the simulated values of the DGINAR(1) process. However, note that it is troublesome to detect this trait for small values of α.

3.3 Statistical Properties of the DGINAR(1) Process

In this Section, we present some statistical properties of our process, such as the calculation of the transition probabilities, joint distributions and correlation structure, among others.

3.3.1 Transition Probabilities

The transition probabilities in a Markov process are defined by

$$p_{ij} = \Pr(X_t = j | X_{t-1} = i),\ i, j \in \mathbb{N} \cup \{0\},$$

the conditional probability function of X_t given X_{t-1}.

Proposition 3.3.1. *The transition probabilities of the DGINAR(1) process are:*

i) $i = 0$ *and* $j \in \mathbb{N} \cup \{0\}$*:*

$$p_{0j} = p_\varepsilon(j).$$

ii) $i \geq 1$ *and* $j \in \mathbb{N} \cup \{0\}$*:*

$$p_{ij} = p_\varepsilon(j)\left[\pi_\star^i + \sum_{m=1}^{j}\sum_{l=1}^{m}\left(1+\frac{1}{\mu_\varepsilon}\right)^m \binom{l+i-1}{l}(1-\pi_\star)^l\pi_\star^i\binom{m-1}{l-1}p^{m-l}(1-p)^l\right],$$

where $p_\varepsilon(\cdot)$ *is the pmf of the innovations. Recall that* $\pi_\star = 1 - \frac{\alpha}{1+\mu_\varepsilon}$ *and* $p = \frac{\mu_\varepsilon - \alpha}{1+\mu_\varepsilon - \alpha}$.

Proof. From the definition of conditional probability and from the DGINAR(1) process, it follows that:

i) From state 0 to state j:

$$\begin{aligned}
p_{0j} &= \Pr(X_t = j | X_{t-1} = 0) = \frac{\Pr(X_t = j, X_{t-1} = 0)}{\Pr(X_{t-1} = 0)} \\
&= \frac{\Pr(\boldsymbol{\theta} \star X_{t-1} + \varepsilon_t = j, X_{t-1} = 0)}{\Pr(X_{t-1} = 0)} = \frac{\Pr(\varepsilon_t = j - \boldsymbol{\theta} \star X_{t-1}, X_{t-1} = 0)}{\Pr(X_{t-1} = 0)} \\
&= \frac{\Pr(\varepsilon_t = j - \sum_{k=1}^{X_{t-1}} G_k, X_{t-1} = 0)}{\Pr(X_{t-1} = 0)} = \frac{\Pr(\varepsilon_t = j - \sum_{k=1}^{0} G_k, X_{t-1} = 0)}{\Pr(X_{t-1} = 0)} \\
&= \frac{\Pr(\varepsilon_t = j)\Pr(X_{t-1} = 0)}{\Pr(X_{t-1} = 0)} \\
&= p_\varepsilon(j),\ j = 0, 1, 2, \ldots.
\end{aligned}$$

ii) From state $i \geq 1$ to state j:

$$\begin{aligned}
p_{ij} &= \Pr(X_t = j | X_{t-1} = i) = \frac{\Pr(X_t = j, X_{t-1} = i)}{\Pr(X_{t-1} = i)} \\
&= \frac{\Pr(\boldsymbol{\theta} \star X_{t-1} + \varepsilon_t = j, X_{t-1} = i)}{\Pr(X_{t-1} = i)} = \frac{\Pr(\varepsilon_t = j - \boldsymbol{\theta} \star X_{t-1}, X_{t-1} = i)}{\Pr(X_{t-1} = i)} \\
&= \frac{\Pr(\varepsilon_t = j - \sum_{k=1}^{X_{t-1}} G_k, X_{t-1} = i)}{\Pr(X_{t-1} = i)} = \frac{\Pr(\varepsilon_t = j - \sum_{k=1}^{i} G_k, X_{t-1} = i)}{\Pr(X_{t-1} = i)} \\
&= \frac{\Pr(\varepsilon_t = j - \sum_{k=1}^{i} G_k)\Pr(X_{t-1} = i)}{\Pr(X_{t-1} = i)} = \Pr\left(\varepsilon_t = j - \sum_{k=1}^{i} G_k\right) \\
&= \sum_{m=0}^{j} \Pr\left(\varepsilon_t = j - \sum_{k=1}^{i} G_k, \sum_{k=1}^{i} G_k = m\right) = \sum_{m=0}^{j} \Pr\left(\varepsilon_t = j - m, \sum_{k=1}^{i} G_k = m\right) \\
&= \sum_{m=0}^{j} \Pr(\varepsilon_t = j - m)\Pr\left(\sum_{k=1}^{i} G_k = m\right) = \sum_{m=0}^{j} \frac{\mu_\varepsilon^{j-m}}{(1+\mu_\varepsilon)^{(j-m)+1}}\Pr\left(\sum_{k=1}^{i} G_k = m\right)
\end{aligned}$$

$$= \Pr(\varepsilon_t = j) \sum_{m=0}^{j} \left(\frac{1+\mu_\varepsilon}{\mu_\varepsilon}\right)^m \Pr\left(S_i = m\right) \ [S_i \sim \text{ZMNB}(\pi, \mu_\varepsilon, i),\ i \geq 1]$$

$$= p_\varepsilon(j) \left[\Pr(S_i = 0) + \sum_{m=1}^{j} \left(1 + \frac{1}{\mu_\varepsilon}\right)^m \Pr(S_i = m)\right]$$

$$= p_\varepsilon(j) \left[\pi_\star^i + \sum_{m=1}^{j} \sum_{l=1}^{m} \left(1 + \frac{1}{\mu_\varepsilon}\right)^m \binom{l+i-1}{l} (1-\pi_\star)^l \pi_\star^i \binom{m-1}{l-1} p^{m-l} (1-p)^l\right].$$

□

3.3.2 Joint Distribution

In the case of a discrete bivariate random vector (Z_1, Z_2), its pgf is given by

$$\varphi_{Z_1,Z_2}(s_1, s_2) = E\left(s_1^{Z_1} s_2^{Z_2}\right),\ \max\{|s_1|, |s_2|\} \leq 1.$$

Since $\{X_t\}$ is a stationary process, the following result holds and is a tool to derive the joint pgf of the rvs X_t and X_{t-1}.

Lemma 3.3.2. *The joint pgf of the discrete random vector (X_t, X_{t-1}) is*

$$\varphi_{X_t,X_{t-1}}(s_1, s_2) = \varphi_\varepsilon(s_1)\varphi_X\left(s_2\varphi_G(s_1)\right),$$

where $\{X_t\}$ is the DGINAR(1) process.

Proof. From the definition of the $\star$ operator, it follows that

$$\begin{aligned}
\varphi_{X_t,X_{t-1}}(s_1, s_2) &= E(s_1^{X_t} s_2^{X_{t-1}}) = E\left(s_1^{\boldsymbol{\theta}\star X_{t-1}+\varepsilon_t} s_2^{X_{t-1}}\right) \\
&= E(s_1^{\varepsilon_t}) E\left(s_1^{\boldsymbol{\theta}\star X_{t-1}} s_2^{X_{t-1}}\right) \\
&= \varphi_\varepsilon(s_1) E\left[E\left(s_1^{\sum_{i=1}^{X_{t-1}} G_i} s_2^{X_{t-1}} \middle| X_{t-1}\right)\right] \\
&= \varphi_\varepsilon(s_1) E\left[s_2^{X_{t-1}} \prod_{i=1}^{X_{t-1}} E(s_1^G)\right] \\
&= \varphi_\varepsilon(s_1) E\left[\left(s_2\varphi_G(s_1)\right)^{X_{t-1}}\right] \\
&= \varphi_\varepsilon(s_1)\varphi_X\left(s_2\varphi_G(s_1)\right).
\end{aligned}$$

□

With Lemma 3.3.2 in hand, we have that $\varphi_{X_t,X_{t-1}}(s_1,s_2) = \dfrac{\varphi_\varepsilon(s_1)}{1+\mu(1-s_2\varphi_G(s_1))}$. Then,

$$\varphi_{X_t,X_{t-1}}(s_1,s_2) = \frac{1}{1+\mu\left[(1-s_1)+(1-s_2)+(\mu_\varepsilon-\alpha)(1-s_1)(1-s_2)\right]}, \tag{3.8}$$

which is a pgf of a Bivariate Geometric Distribution (BGD) with parameters $c_1 = c_2 = \mu$ and $\gamma^2 = \frac{\mu}{1+\mu}\alpha$ as given by Jayakumar and Mundassery (2007).

We denote $(X_t, X_{t-1}) \sim \text{BGD}(c_1, c_2, \gamma^2)$.

Furthermore, note from Eq. (3.8) that $\varphi_{X_t,X_{t-1}}(s_1,s_2)$ is symmetric in s_1 and s_2, meaning that the DGINAR(1) process is time reversible, which is a remarkable property of our process. By time reversible, we mean that in the DGINAR(1) process, the future and the past can be swapped, i. e., from an inferential point of view, one can analyze data from past to future or from future to past to obtain the same results. It is worth mentioning that this property is extremely rare on INAR processes. To the best of our knowledge, the only analogous case is the Poisson INAR process, which also has marginals and innovations in the same family of distributions. It seems that time reversibility is an innate characteristic of the INAR processes constructed under the new modeling perspective proposed in this work. This fact must be investigated by the author in future works.

3.3.3 Autocorrelation Structure

To obtain the autocorrelation function, we first determine the autocovariance, $\gamma_h = Cov(X_t, X_{t+h})$, $h \geq 0$.

Proposition 3.3.3. *The autocovariance of the DGINAR(1) process is given by*

$$\gamma_h = \alpha^h\gamma_0, \quad h \geq 0 \quad \textit{and} \quad 0 < \alpha < 1.$$

Proof. We use mathematical induction and items i) and iii) of Proposition 3.2.1.

- Base case: $h = 1$.

$$\begin{aligned}\gamma_1 &= Cov(X_t, X_{t+1}) = Cov(X_t, \boldsymbol{\theta} \star X_t + \varepsilon_{t+1}) \\ &= Cov(\boldsymbol{\theta} \star X_t, X_t) + Cov(\varepsilon_{t+1}, X_t) \\ &= E(G)Var(X_t) = E(G)Var(X) = \alpha\gamma_0.\end{aligned}$$

- Step case: $h \in \mathbb{N}$.

  Suppose the result holds for $h \in \mathbb{N}$, i. e., $\gamma_h = \alpha^h \gamma_0$.

  Now evaluate the result for the case when $h + 1 \in \mathbb{N}$.

$$\begin{aligned}
\gamma_{h+1} &= Cov(X_t, X_{t+(h+1)}) = Cov(X_t, \boldsymbol{\theta} \star X_{t+h} + \varepsilon_{t+h+1}) \\
&= Cov(\boldsymbol{\theta} \star X_{t+h}, X_t) + Cov(\varepsilon_{t+h+1}, X_t),\ \varepsilon_{t+h+1} \perp X_t \\
&= E((\boldsymbol{\theta} \star X_{t+h})X_t) - E(\boldsymbol{\theta} \star X_{t+h})E(X_t) \\
&= E(G)E(X_{t+h}X_t) - E(G)E(X_{t+h})E(X_t) \\
&= E(G)Cov(X_t, X_{t+h}) = E(G)\gamma_h = \alpha\alpha^h\gamma_0 \\
&= \alpha^{h+1}\gamma_0.
\end{aligned}$$

□

Corollary 3.3.4. *The autocorrelation function is*

$$\rho_h = \alpha^h, \quad \forall h \geq 0 \quad \textit{and} \quad 0 < \alpha < 1.$$

Proof. $\rho_h = \dfrac{\gamma_h}{\gamma_0} = \dfrac{\alpha^h\gamma_0}{\gamma_0} = \alpha^h$, $\forall h \geq 0$ and $0 < \alpha < 1$. □

Note that the autocorrelation function decays exponentially as $h \longrightarrow \infty$.

3.3.4 Conditional Moments

Conditional moments are of great importance in the analysis of time series. On the one hand, they have practical interpretations in the modeling of various phenomena; in financial time series, such as stock market returns, the first and the second conditional moments are associated with the risk premium and with its volatility, respectively. On the other hand, these measures aid in the estimation procedures of the unknown parameters, as given in Chapter 4.

Proposition 3.3.5. *In the DGINAR(1) process, the* 1*-step ahead conditional mean and conditional variance are:*

i) $E(X_{t+1}|X_t) = E(G)X_t + E(\varepsilon) = \alpha X_t + (1-\alpha)\mu.$

ii) $Var(X_{t+1}|X_t) = Var(G)X_t + Var(\varepsilon) = [(1+2\mu)(1-\alpha)\alpha]\, X_t + \sigma_\varepsilon^2.$

Proof. Just use the definition of DGINAR(1) process and some algebra manipulation.

i) $$\begin{aligned} E(X_{t+1}|X_t) &= E(\boldsymbol{\theta} \star X_t + \varepsilon_{t+1}|X_t) \\ &= E(\boldsymbol{\theta} \star X_t|X_t) + E(\varepsilon_{t+1}|X_t) \\ &= E(G)X_t + E(\varepsilon) \\ &= \alpha X_t + (1-\alpha)\mu. \end{aligned}$$

ii) $$\begin{aligned} Var(X_{t+1}|X_t) &= E\left(X_{t+1}^2|X_t\right) - E^2\left(X_{t+1}|X_t\right) \\ &= E\left((\boldsymbol{\theta} \star X_t + \varepsilon_{t+1})^2|X_t\right) - \underbrace{\left(E(G)X_t + E(\varepsilon_{t+1})\right)^2}_{\zeta} \\ &= E\left((\boldsymbol{\theta} \star X)^2|X\right) + E\left(2(\boldsymbol{\theta} \star X)\varepsilon|X\right) + E(\varepsilon^2) - \zeta \\ &= XE(G^2) + (X^2 - X)E^2(G) + 2XE(G)E(\varepsilon) + Var(\varepsilon) + E^2(\varepsilon) - \zeta \\ &= XVar(G) + \underbrace{\left(E(G)X + E(\varepsilon)\right)^2}_{\zeta} + Var(\varepsilon) - \zeta \\ &= Var(G)X_t + Var(\varepsilon) \\ &= \left[(1+2\mu)(1-\alpha)\alpha\right]X_t + \sigma_\varepsilon^2. \end{aligned}$$

□

It is possible to generalize this result using induction methods to obtain the following recurrence relations:

r.1) $E(X_{t+h}|X_t) = E(G)E(X_{t+(h-1)}|X_t) + \mu_\varepsilon.$

r.2) $Var(X_{t+h}|X_t) = Var(G)E(X_{t+(h-1)}|X_t) + E^2(G)Var(X_{t+(h-1)}|X_t) + \sigma_\varepsilon^2.$

Then, solving the difference equations we can generalize the h-step ahead conditional moments:

Proposition 3.3.6. *In the DGINAR(1) process, the h-step ahead conditional mean and conditional variance are:*

i) $$E(X_{t+h}|X_t) = \alpha^h X_t + \frac{1-\alpha^h}{1-\alpha}\mu_\varepsilon.$$

ii) $$Var(X_{t+h}|X_t) = (1+2\mu)\left[\alpha^{h+1}(1-\alpha^{h-1}) + \alpha^{2h-1}(1-\alpha)\right]X_t + (1+2\mu)\left[\alpha^{h+1}\frac{1-\alpha^{h-1}}{1-\alpha} + \alpha\frac{1-\alpha^{2h-2}}{1-\alpha^2}\right]\mu_\varepsilon + \left[\alpha^{2h-2} + \frac{1-\alpha^{2h-2}}{1-\alpha^2}\right]\sigma_\varepsilon^2.$$

Proof. The proof uses the principle of mathematical induction. The base case where $h = 1$ is the Proposition 3.3.5. The step case where $h \in \mathbb{N}$, uses the recurrence relations (r.1) and (r.2) above. □

Note that in Proposition 3.3.6 when $h \to \infty$ we have that:

i) $$E(X_{t+h}|X_t) \to 0X_t + \frac{1}{1-\alpha}\mu_\varepsilon = \mu = E(X_t);$$

ii) $$Var(X_{t+h}|X_t) \to (1+2\mu)0X_t + (1+2\mu)\frac{\alpha}{1-\alpha^2}\mu_\varepsilon + \frac{1}{1-\alpha^2}\sigma_\varepsilon^2 = \frac{1}{1-\alpha^2}(1-\alpha)(1+\alpha)\mu(1+\mu) = \mu(1+\mu) = Var(X_t),$$

which are, as expected, the unconditional mean and unconditional variance, respectively.

3.3.5 Conditional pgf

Now, we derive the conditional pgf $E(s^{X_{t+h}}|X_t)$. At first, by the definition of the $\star$ operator we have that

$$E\left(s^{X_{t+h}}\middle|X_t\right) = E\left(s^{\boldsymbol{\theta}\star X_{t+(h-1)}+\varepsilon_{t+h}}\middle|X_t\right) = E\left[E\left(s^{\boldsymbol{\theta}\star X_{t+(h-1)}}|X_{t+(h-1)}\right)\middle|X_t\right]E\left(s^{\varepsilon_{t+h}}\middle|X_t\right)$$

$$= E\left[E\left(s^{\sum_{i=1}^{X_{t+(k-1)}}G_i}|X_{t+(h-1)}\right)\middle|X_t\right]\varphi_\varepsilon(s) = E\left[\prod_{i=1}^{X_{t+(h-1)}}E\left(s^{G_i}\middle|X_{t+(h-1)}\right)\middle|X_t\right]\varphi_\varepsilon(s)$$

$$= E\left[\prod_{i=1}^{X_{t+(h-1)}}E\left(s^{G}\right)\middle|X_t\right]\varphi_\varepsilon(s) = E\left[(\varphi_G(s))^{X_{t+(h-1)}}\middle|X_t\right]\varphi_\varepsilon(s).$$

Then, after repeating h times, we obtain

$$E\left(s^{X_{t+h}}\middle|X_t\right) = \prod_{i=0}^{h-1}\varphi_\varepsilon\left(\varphi_G^{(i)}(s)\right)\cdot\left(\varphi_G^{(h)}(s)\right)^{X_t},$$

where $\varphi_G^{(h)}(s) = \varphi_G\left(\varphi_G^{(h-1)}(s)\right)$ and $\varphi_G^{(0)}(s) = s$. Applying Eq. (3.2), we obtain that

$$E\left(s^{X_{t+h}}\middle|X_t\right) = \varphi_X(s)\left[\varphi_X\left(\varphi_G^{(h)}(s)\right)\right]^{-1}\left(\varphi_G^{(h)}(s)\right)^{X_t}. \tag{3.9}$$

By induction, we can prove that

$$\varphi_G^{(h)}(s) = \frac{1 + \left[(1-\alpha^{h-1})\mu - \alpha^h\right](1-s)}{1 + (1-\alpha^{h-1})\mu(1-s)}$$

and substituting this in Eq. (3.9), we finally achieve the conditional pgf as

$$E\left(s^{X_{t+h}}\middle|X_t\right) =$$

$$\varphi_X(s)\left[\varphi_X\left(\frac{1 + \left[(1-\alpha^{h-1})\mu - \alpha^h\right](1-s)}{1 + (1-\alpha^{h-1})\mu(1-s)}\right)\right]^{-1}\left(\frac{1 + \left[(1-\alpha^{h-1})\mu - \alpha^h\right](1-s)}{1 + (1-\alpha^{h-1})\mu(1-s)}\right)^{X_t},$$

which converges to the pgf $\varphi_X(s)$ as $h \to \infty$, since $\varphi_G^{(h)}(s) \to 1$, as $h \to \infty$.

4 Parameters Estimation

In this Chapter we provide some alternatives to estimate the unknown parameters of the DGINAR(1) process. Let $\boldsymbol{\theta} = (\mu, \alpha)^T$ be the parameter vector and consider $(X_1, X_2, \ldots, X_n)$ a realization from $\{X_t; t \geq 1\}$. Note that n is the sample size.

4.1 Conditional Least Squares

The Conditional Least Squares (CLS) estimators are obtained by minimizing the function

$$Q_n(\boldsymbol{\theta}) = \sum_{i=2}^{n} (X_i - E(X_i|X_{i-1}))^2 = \sum_{i=2}^{n} (X_i - \alpha X_{i-1} - (1-\alpha)\mu)^2,$$

in μ and α.

By solving the system of equations $\partial Q_n(\boldsymbol{\theta})/\partial \boldsymbol{\theta} = 0$ we obtain the following estimators

$$\hat{\alpha}_{CLS} = \frac{\sum_{i=2}^{n} X_i X_{i-1} - \frac{1}{n-1}\sum_{i=2}^{n} X_i \sum_{i=2}^{n} X_{i-1}}{\sum_{i=2}^{n} X_{i-1}^2 - \frac{1}{n-1}\left(\sum_{i=2}^{n} X_{i-1}\right)^2} \tag{4.1}$$

and

$$\hat{\mu}_{CLS} = \frac{\sum_{i=2}^{n} X_i - \hat{\alpha}_{CLS}\sum_{i=2}^{n} X_{i-1}}{(n-1)(1-\hat{\alpha}_{CLS})}. \tag{4.2}$$

Appendix C presents the procedure to obtain Eq. (4.1) and Eq. (4.2).

Proposition 4.1.1. *Let $\hat{\boldsymbol{\theta}}_{CLS} = (\hat{\mu}_{CLS}, \hat{\alpha}_{CLS})^T$ be the CLS estimators given by Eq. (4.1) and Eq. (4.2). Then $\hat{\boldsymbol{\theta}}_{CLS}$ is a strongly consistent estimator to $\boldsymbol{\theta}$ and its asymptotic distribution is given by*

$$\sqrt{n}\begin{bmatrix} \hat{\mu}_{CLS} - \mu \\ \hat{\alpha}_{CLS} - \alpha \end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{\mu(1+\mu)(1+\alpha)}{1-\alpha} & (1+2\mu)\alpha \\ (1+2\mu)\alpha & \frac{(1+\mu+2\mu)\sigma_G^2 + \sigma_\varepsilon^2}{\mu(1+\mu)} \end{bmatrix}\right),$$

where $\sigma_G^2 = Var(G) = (1+2\mu)(1-\alpha)\alpha$.

Proof. To prove Proposition 4.1.1 we must verify all the conditions given in Theorem 3.1 from Tjostheim (1986). We also apply Theorem 3.2 from Tjostheim (1986) to ensure the asymptotic result of $\hat{\boldsymbol{\theta}}_{CLS}$.

Firstly, note that $E\left(|X_t|^2\right) < \infty$ and that $E(X_t|X_{t-1}) = \alpha X_{t-1} + (1-\alpha)\mu$, as a function of μ and α, is almost surely three times continuously differentiable in an open set Θ containing $\boldsymbol{\theta}_0 = (\mu_0, \alpha_0)^T$, the true value of the unknown parameter $\boldsymbol{\theta} = (\theta_1 = \mu, \theta_2 = \alpha)^T$.

Condition 1:

$$(\mathrm{I}_1): \ E\left\{\left|\frac{\partial E(X_t|X_{t-1})}{\partial \theta_i}(\boldsymbol{\theta}_0)\right|^2\right\} < \infty \text{ and } (\mathrm{II}_1): \ E\left\{\left|\frac{\partial^2 E(X_t|X_{t-1})}{\partial \theta_i \partial \theta_j}(\boldsymbol{\theta}_0)\right|^2\right\} < \infty$$

for $i, j = 1, 2$.

Indeed. Note that in (I_1) we have

$$E\left\{\left|\frac{\partial E(X_t|X_{t-1})}{\partial \mu}(\boldsymbol{\theta}_0)\right|^2\right\} = E\left\{|1-\alpha_0|^2\right\} = (1-\alpha_0)^2 < \infty \text{ and}$$

$$E\left\{\left|\frac{\partial E(X_t|X_{t-1})}{\partial \alpha}(\boldsymbol{\theta}_0)\right|^2\right\} = E\left\{|X_{t-1}-\mu_0|^2\right\} = Var(X_{t-1}) < \infty.$$

While from (II_1) follows that

$$E\left\{\left|\frac{\partial^2 E(X_t|X_{t-1})}{\partial \mu^2}(\boldsymbol{\theta}_0)\right|^2\right\} = E\left\{|0|^2\right\} = 0 < \infty,$$

$$E\left\{\left|\frac{\partial^2 E(X_t|X_{t-1})}{\partial \mu \partial \alpha}(\boldsymbol{\theta}_0)\right|^2\right\} = E\left\{|-1|^2\right\} = 1 < \infty,$$

$$E\left\{\left|\frac{\partial^2 E(X_t|X_{t-1})}{\partial \alpha^2}(\boldsymbol{\theta}_0)\right|^2\right\} = E\left\{|0|^2\right\} = 0 < \infty \text{ and}$$

$$E\left\{\left|\frac{\partial^2 E(X_t|X_{t-1})}{\partial \alpha \partial \mu}(\boldsymbol{\theta}_0)\right|^2\right\} = E\left\{|-1|^2\right\} = 1 < \infty.$$

Condition 2: The vectors $\partial E(X_t|X_{t-1})(\boldsymbol{\theta}_0)/\partial\theta_i$, $i = 1, 2$, are linearly independent in the sense that if a_1 and a_2 are arbitrary real numbers such that

$$E\left\{\left|\sum_{i=1}^{2} a_i \frac{\partial E(X_t|X_{t-1})}{\partial \theta_i}(\boldsymbol{\theta}_0)\right|^2\right\} = 0,$$

then $a_1 = a_2 = 0$.

Note that

$$E\left\{\left|a_1\frac{\partial E(X_t|X_{t-1})}{\partial\mu}(\boldsymbol{\theta}_0)+a_2\frac{\partial E(X_t|X_{t-1})}{\partial\alpha}(\boldsymbol{\theta}_0)\right|^2\right\}=0\Rightarrow$$

$$E\left\{\left|a_1\left(1-\alpha_0\right)+a_2\left(X_{t-1}-\mu_0\right)\right|^2\right\}=0\Rightarrow$$

$$\underbrace{a_1^2\left(1-\alpha_0\right)^2}_{>0}+\underbrace{a_2^2 Var\left(X_{t-1}\right)}_{>0}=0\Rightarrow$$

$$a_1^2\underbrace{\left(1-\alpha_0\right)^2}_{>0}=0\ \text{ and }\ a_2^2\underbrace{Var\left(X_{t-1}\right)}_{>0}=0\Rightarrow$$

$$a_1^2=0\ \text{ and }\ a_2^2=0.$$

Then $a_1=a_2=0$.

Condition 3: For $\boldsymbol{\theta}\in\Theta$, there exist functions $G_{t-1}^{ijk}(X_1,\ldots,X_{t-1})$ and $H_t^{ijk}(X_1,\ldots,X_t)$ such that

$$(\mathrm{I}_3): T_{t-1}^{ijk}(\boldsymbol{\theta})=\left|\frac{\partial E(X_t|X_{t-1})}{\partial\theta_i}(\boldsymbol{\theta})\frac{\partial^2 E(X_t|X_{t-1})}{\partial\theta_j\partial\theta_k}(\boldsymbol{\theta})\right|\ \leq\ G_{t-1}^{ijk},\ \ E\left(G_{t-1}^{ijk}\right)<\infty;$$

$$(\mathrm{II}_3): D_t^{ijk}(\boldsymbol{\theta})=\left|\left\{X_t-E(X_t|X_{t-1})(\boldsymbol{\theta})\right\}\frac{\partial^3 E(X_t|X_{t-1})}{\partial\theta_i\partial\theta_j\partial\theta_k}(\boldsymbol{\theta})\right|\ \leq\ H_t^{ijk},\ \ E\left(G_t^{ijk}\right)<\infty$$

for $i,j,k=1,2$.

In (I_3) note that $T_{t-1}^{111}(\boldsymbol{\theta})=T_{t-1}^{122}(\boldsymbol{\theta})=T_{t-1}^{211}(\boldsymbol{\theta})=T_{t-1}^{222}(\boldsymbol{\theta})=0$ and that

$$T_{t-1}^{112}(\boldsymbol{\theta})=T_{t-1}^{121}(\boldsymbol{\theta})=|\alpha-1|<1,\ \text{and}$$

$$T_{t-1}^{212}(\boldsymbol{\theta})=T_{t-1}^{221}(\boldsymbol{\theta})=|X_{t-1}-\mu|.$$

If we choose $G_{t-1}^{ijk}=\left(X_{t-1}-\mu\right)^2+1$, $\forall i,j,k=1,2$ we guarantee that $T_{t-1}^{ijk}(\boldsymbol{\theta})\leq G_{t-1}^{ijk}$; besides $E\left(G_{t-1}^{ijk}\right)=Var(X_{t-1})+1<\infty$.

Concerning to (II_3), $\dfrac{\partial^3 E(X_t|X_{t-1})}{\partial\theta_i\partial\theta_j\partial\theta_k}(\boldsymbol{\theta})=0$, $\forall i,j,k=1,\,2$.

So take $H_t^{ijk}=0$, $\forall i,j,k=1,\,2$ to satisfy the condition.

These three conditions ensure that $\hat{\boldsymbol{\theta}}_{CLS}$ is a strongly consistent estimator for $\boldsymbol{\theta}$.

Theorem 3.2 from Tjostheim (1986) refers to the asymptotic distribution of $\hat{\boldsymbol{\theta}}_{CLS}$ and states that

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_{CLS}-\boldsymbol{\theta}\right)\xrightarrow{d}\mathrm{N}\left(\mathbf{0},\Sigma\right),$$

where $\Sigma = U^{-1}RU^{-1}$.

The elements involved in the calculation of Σ are:

- U:

$$\begin{aligned} U &= E\left\{\frac{\partial E(X_t|X_{t-1})^T}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}) \cdot \frac{\partial E(X_t|X_{t-1})}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta})\right\} \\ &= E\left\{\begin{pmatrix} 1-\alpha \\ X_{t-1}-\mu \end{pmatrix}\begin{pmatrix} 1-\alpha & X_{t-1}-\mu \end{pmatrix}\right\} \\ &= \begin{pmatrix} (1-\alpha)^2 & 0 \\ 0 & \mu(1+\mu) \end{pmatrix}. \end{aligned}$$

- $f_{t|t-1}(\boldsymbol{\theta})$:

$$\begin{aligned} f_{t|t-1}(\boldsymbol{\theta}) &= E\left\{(X_t - E(X_t|X_{t-1}))\,(X_t - E(X_t|X_{t-1}))^T\,|X_{t-1}\right\} \\ &= E\left\{(X_t - E(X_t|X_{t-1}))^2\,|X_{t-1}\right\} \\ &= Var(X_t|X_{t-1}) \\ &= \alpha(1-\alpha)(1+2\mu)X_{t-1} + \sigma_\varepsilon^2. \end{aligned}$$

- R:

$$\begin{aligned} R &= E\left\{\frac{\partial E(X_t|X_{t-1})^T}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}) f_{t|t-1}(\boldsymbol{\theta}) \frac{\partial E(X_t|X_{t-1})}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta})\right\} \\ &= E\left\{f_{t|t-1}(\boldsymbol{\theta})\frac{\partial E(X_t|X_{t-1})^T}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}) \frac{\partial E(X_t|X_{t-1})}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta})\right\} \\ &= E\left\{f_{t|t-1}(\boldsymbol{\theta})\begin{pmatrix} (1-\alpha)^2 & (X_{t-1}-\mu)(1-\alpha) \\ (X_{t-1}-\mu)(1-\alpha) & (X_{t-1}-\mu)^2 \end{pmatrix}\right\}. \end{aligned}$$

  r_{11}:

$$\begin{aligned} r_{11} &= E\left\{f_{t|t-1}(\boldsymbol{\theta})(1-\alpha)^2\right\} \\ &= E\left\{\left[\alpha(1-\alpha)(1+2\mu)X_{t-1} + \sigma_\varepsilon^2\right](1-\alpha)^2\right\} \\ &= (1-\alpha)^2\left[\alpha(1-\alpha)(1+2\mu)\right]E\left(X_{t-1}\right) + (1-\alpha)^2\sigma_\varepsilon^2 \\ &= (1-\alpha)^3\alpha\mu + 2(1-\alpha)^3\alpha\mu^2 + (1-\alpha)^3\mu\left[1+(1-\alpha)\mu\right] \\ &= (1-\alpha)^3\mu\left[\alpha(1+\mu) + (1+\mu)\right] \\ &= \mu(1+\mu)(1+\alpha)(1-\alpha)^3. \end{aligned}$$

r_{12}=r_{21}:

$$\begin{aligned}
r_{12} &= E\left\{f_{t|t-1}(\boldsymbol{\theta})(X_{t-1}-\mu)(1-\alpha)\right\} \\
&= E\left\{\left[\alpha(1-\alpha)(1+2\mu)X_{t-1}+\sigma_\varepsilon^2\right](X_{t-1}-\mu)(1-\alpha)\right\} \\
&= \alpha(1-\alpha)^2(1+2\mu)E\left\{X_{t-1}(X_{t-1}-\mu)\right\}+\sigma_\varepsilon^2(1-\alpha)E\left\{(X_{t-1}-\mu)\right\} \\
&= \alpha(1-\alpha)^2(1+2\mu)\left[E\left(X_{t-1}^2\right)-E\left(X_{t-1}\right)^2\right] \\
&= \mu(1+\mu)(1+2\mu)(1-\alpha)^2\alpha.
\end{aligned}$$

r_{22}:

$$\begin{aligned}
r_{22} &= E\left\{f_{t|t-1}(\boldsymbol{\theta})(X_{t-1}-\mu)^2\right\} \\
&= E\left\{\left[\alpha(1-\alpha)(1+2\mu)X_{t-1}+\sigma_\varepsilon^2\right](X_{t-1}-\mu)^2\right\} \\
&= \alpha(1-\alpha)(1+2\mu)E\left\{X_{t-1}(X_{t-1}-\mu)^2\right\}+\sigma_\varepsilon^2E\left\{(X_{t-1}-\mu)^2\right\} \\
&= \alpha(1-\alpha)(1+2\mu)\left[E\left(X_{t-1}^3\right)-2\mu E\left(X_{t-1}^2\right)+\mu E\left(X_{t-1}\right)\right]+\sigma_\varepsilon^2Var(X_{t-1}) \\
&= \alpha(1-\alpha)(1+2\mu)\left[\mu+4\mu^2+3\mu^3\right]+(1-\alpha)\mu^2(1+\mu)\left[1+(1-\alpha)\mu\right] \\
&= \alpha(1-\alpha)(1+2\mu)\mu(1+\mu)(1+\mu+2\mu)+(1-\alpha)\mu^2(1+\mu)\left[1+(1-\alpha)\mu\right].
\end{aligned}$$

- Σ:

$$\begin{aligned}
\Sigma &= U^{-1}RU^{-1} \\
&= \begin{pmatrix} \dfrac{r_{11}}{(1-\alpha)^4} & \dfrac{r_{12}}{\mu(1+\mu)(1-\alpha)^2} \\ \dfrac{r_{21}}{\mu(1+\mu)(1-\alpha)^2} & \dfrac{r_{22}}{\mu^2(1+\mu)^2} \end{pmatrix} \\
&= \begin{pmatrix} \dfrac{\mu(1+\mu)(1+\alpha)}{1-\alpha} & (1+2\mu)\alpha \\ (1+2\mu)\alpha & \dfrac{(1+\mu+2\mu)\sigma_G^2+\sigma_\varepsilon^2}{\mu(1+\mu)} \end{pmatrix}.
\end{aligned}$$

□

4.2 Yule-Walker

The Yule-Walker (YW) estimators are based on the method of moments and are obtained by solving the Yule-Walker equations, which estimate the parameters through

the sample autocorrelations. We obtain an analytic expression for $\hat{\alpha}_{YW}$ as a function of the autocorrelations. Since $\alpha = \gamma_1/\gamma_0$, then

$$\begin{aligned}\hat{\alpha}_{YW} &= \frac{\hat{\gamma}_1}{\hat{\gamma}_0} = \frac{Cov(X_t, X_{t-1})}{Var(X_t)} \\ &= \frac{\sum_{i=2}^{n}(X_i - \bar{X})(X_{i-1} - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2},\end{aligned} \tag{4.3}$$

where $\bar{X}$ is the sample mean of the process $\{X_t\}$.

Moreover, since in the DGINAR(1) process $\mu = E(X)$, we naturally set

$$\hat{\mu}_{YW} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{4.4}$$

In the DGINAR(1) process, the YW and the CLS estimators are asymptotically equivalent. This fact comes from Freeland and McCabe (2005). Since in the DGINAR(1) process the YW and CLS estimators are exactly the same of the Poisson INAR(1) process considering the alternative parameterization proposed by Joe (1996), we can use directly Theorem 3 from Freeland and McCabe (2005) as follow:

Proposition 4.2.1. *In the DGINAR(1) process*

$$\begin{bmatrix} \hat{\alpha}_{YW} - \hat{\alpha}_{CLS} \\ \hat{\mu}_{YW} - \hat{\mu}_{CLS} \end{bmatrix} = o_p(n^{-1/2}).$$

This is sufficient for the CLS and YW estimators to have the same asymptotic distribution.

Proof. Recall that the DGINAR(1) process is stationary. Divide the numerator and denominator of both statistics by n and denote the denominators as quantities D_{YW} and D_{CLS}, respectively. Both quantities converge to the same nonzero constant asymptotically and so $(D_{YW} - D_{CLS}) \xrightarrow{p} 0$. By expanding the numerators we get

$$\begin{aligned}n^{1/2}(\hat{\alpha}_{YW} - \hat{\alpha}_{CLS}) = {} & \frac{(D_{CLS} - D_{YW})}{D_{YW}D_{CLS}} \frac{\sum_{i=2}^{n} X_i X_{i-1}}{n^{1/2}} \\ & - \frac{(D_{CLS}\sum_{i=2}^{n} X_i/n - D_{YW}\sum_{i=2}^{n} X_i/(n-1))}{D_{YW}D_{CLS}} n^{-1/2}\sum_{i=1}^{n} X_i \\ & - \frac{D_{YW}}{D_{YW}D_{CLS}} \frac{\sum_{i=2}^{n} X_i}{n-1} \frac{X_t}{n^{1/2}} + \frac{D_{CLS}}{D_{YW}D_{CLS}} n^{-1}\sum_{i=1}^{n} X_i \frac{X_1}{n^{1/2}}.\end{aligned}$$

Thus,

$$n^{1/2}\left(\hat{\alpha}_{YW}-\hat{\alpha}_{CLS}\right)=o_p(1)O_p(1)-o_p(1)O_p(1)-O_p(1)O_p(1)o_p(1)+O_p(1)O_p(1)o_p(1)$$
$$=o_p(1).$$

Finally, since

$$n^{1/2}\left(\hat{\alpha}_{YW}-\alpha\right)-n^{1/2}\left(\hat{\alpha}_{CLS}-\alpha\right)=n^{1/2}\left(\hat{\alpha}_{YW}-\hat{\alpha}_{CLS}\right)\xrightarrow{p}0$$

we have that both estimators have asymptotically the same distribution. Similarly it is straightforward to show $\hat{\mu}_{YW}=\hat{\mu}_{CLS}+o_p(n^{-1/2})$. □

4.3 Maximum Likelihood

The Maximum Likelihood (ML) estimators take into account the temporal dependence between time series ordered records. The ML function is written as

$$L(\boldsymbol{\theta};\boldsymbol{x})=\Pr(X_1=x_1,\ldots,X_n=x_n)=\Pr(X_1=x_1)\prod_{i=2}^{n}\Pr(X_i=x_i|X_{i-1}=x_{i-1}).$$

In the case of the DGINAR(1) process, the log-likelihood function is given by

$$l(\boldsymbol{\theta};\boldsymbol{x})=x_1\log(\mu)-(1-x_1)\log(1+\mu)+\sum_{i=2}^{n}\log p_{x_{i-1}x_i},$$

where $p_{x_{i-1}x_i}$ is the transition probabilities provided in Proposition 3.3.1.

Then, to obtain the ML estimators we solve the system of equations $\partial l(\boldsymbol{\theta};\boldsymbol{x})/\partial\boldsymbol{\theta}=0$ through numerical methods, found in most statistical softwares. In the present work, we use R (a language and environment for statistical computing and graphics, R Development Core Team (2017)) and its function "optim", which provides algorithms for general-purpose optimizations based on Nelder-Mead, quasi-Newton and conjugate-gradient algorithms.

To maximize the log-likelihood function through "optim" we must, at first, provide initial values to the parameters. It can be done by the CLS or YW estimation methods or using the sample mean and sample autocorrelation. In this work, we selected the CLS estimates as the initial values to the BFGS algorithm, an iterative quasi-Newton method (also known as a variable metric algorithm) for solving unconstrained nonlinear optimization problems.

4.4 Monte Carlo Simulation

In this Section, a Monte Carlo simulation study is conducted to compare the three estimation methods proposed in the previous sections. In this comparative study, we simulate realizations of a predetermined DGINAR(1) process, by specifying true values to the parameters μ and α. Then, we apply the three estimation methods to these realizations, obtaining the estimators to both parameters. The results are compared through a chosen metric. In this work, the chosen metric is the Root of the Mean Square Error (RMSE) given by

$$\text{RMSE}(\boldsymbol{\theta}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_0)^2},$$

where $\boldsymbol{\theta}_0$ is the true value of the parameter under study and $\hat{\boldsymbol{\theta}}$ is its estimate by a selected estimation method. Note that the RMSE is an absolute measure and since we have several scenarios and we wish to make comparisons between them, a transformation of the RMSE into a relative measure is needed. For this, we consider the coefficient of variation of the RMSE given by $\text{CV(RMSE)} = \text{RMSE}/\bar{\boldsymbol{\theta}}_0$.

Table 1 – Numerical results of the scenarios (a), (b), (c) and (d). The CV(RMSE) is displayed in parentheses.

n	$\hat{\mu}_{cls}$		$\hat{\alpha}_{cls}$		$\hat{\mu}_{yw}$		$\hat{\alpha}_{yw}$		$\hat{\mu}_{ml}$		$\hat{\alpha}_{ml}$	
a) True values: $\mu = 1$ and $\alpha = 0.1$												
100	1.0094	(0.1573)	0.1265	(0.8948)	1.0096	(0.1564)	0.1252	(0.8811)	1.0098	(0.1566)	0.1355	(0.9341)
300	1.0030	(0.0905)	0.1032	(0.5893)	1.0031	(0.0903)	0.1029	(0.5871)	1.0031	(0.0903)	0.1052	(0.5648)
500	1.0014	(0.0692)	0.0989	(0.4878)	1.0013	(0.0692)	0.0987	(0.4869)	1.0013	(0.0692)	0.1002	(0.4652)
700	1.0016	(0.0579)	0.0989	(0.4273)	1.0017	(0.0579)	0.0988	(0.4268)	1.0017	(0.0579)	0.1002	(0.4075)
1000	1.0008	(0.0506)	0.0981	(0.3611)	1.0008	(0.0506)	0.0980	(0.3608)	1.0007	(0.0506)	0.0991	(0.3455)
b) True values: $\mu = 1$ and $\alpha = 0.3$												
100	0.9975	(0.1926)	0.2672	(0.4055)	0.9971	(0.1907)	0.2644	(0.4041)	0.9972	(0.1908)	0.2854	(0.3608)
300	0.9985	(0.1125)	0.2887	(0.2547)	0.9983	(0.1120)	0.2878	(0.2544)	0.9983	(0.1120)	0.2949	(0.2156)
500	0.9988	(0.0869)	0.2929	(0.1997)	0.9988	(0.0867)	0.2923	(0.1996)	0.9988	(0.0868)	0.2971	(0.1647)
700	1.0002	(0.0730)	0.2933	(0.1690)	1.0003	(0.0729)	0.2929	(0.1690)	1.0002	(0.0728)	0.2974	(0.1394)
1000	0.9988	(0.0606)	0.2958	(0.1409)	0.9988	(0.0605)	0.2956	(0.1409)	0.9988	(0.0605)	0.2981	(0.1142)
c) True values: $\mu = 1$ and $\alpha = 0.5$												
100	1.0064	(0.2491)	0.4535	(0.2640)	1.0052	(0.2440)	0.4487	(0.2652)	1.0057	(0.2440)	0.4844	(0.2066)
300	0.9945	(0.1400)	0.4808	(0.1521)	0.9944	(0.1393)	0.4790	(0.1527)	0.9945	(0.1392)	0.4929	(0.1139)
500	0.9976	(0.1088)	0.4884	(0.1195)	0.9976	(0.1086)	0.4874	(0.1198)	0.9977	(0.1087)	0.4962	(0.0890)
700	0.9998	(0.0944)	0.4928	(0.1025)	0.9999	(0.0943)	0.4920	(0.1027)	0.9999	(0.0943)	0.4976	(0.0756)
1000	0.9990	(0.0779)	0.4944	(0.0863)	0.9990	(0.0778)	0.4939	(0.0865)	0.9990	(0.0778)	0.4983	(0.0633)
d) True values: $\mu = 1$ and $\alpha = 0.7$												
100	1.0055	(0.3395)	0.6400	(0.1780)	1.0047	(0.3313)	0.6328	(0.1828)	1.0061	(0.3294)	0.6813	(0.1252)
300	1.0013	(0.1984)	0.6771	(0.0979)	1.0009	(0.1967)	0.6748	(0.0991)	1.0010	(0.1957)	0.6937	(0.0668)
500	1.0000	(0.1526)	0.6860	(0.0755)	1.0001	(0.1521)	0.6846	(0.0762)	1.0001	(0.1518)	0.6963	(0.0524)
700	1.0016	(0.1277)	0.6903	(0.0636)	1.0016	(0.1275)	0.6892	(0.0641)	1.0017	(0.1274)	0.6971	(0.0438)
1000	1.0009	(0.1046)	0.6937	(0.0536)	1.0009	(0.1045)	0.6930	(0.0537)	1.0009	(0.1044)	0.6981	(0.0359)

We set up eight scenarios to the simulation. In the first four scenarios ((a), (b), (c), and (d)), the value of the mean parameter is fixed as $\mu = 1$. While in the last four scenarios

((e), (f), (g), and (h)), the value of the mean parameter is set as $\mu = 5$. The value of the autocorrelation parameter in the scenarios (a) and (e) is $\alpha = 0.1$, (b) and (f) is $\alpha = 0.3$, (c) and (g) is $\alpha = 0.5$, and in (d) and (h) is $\alpha = 0.7$. Then, for each scenario, we simulate different sample sizes ($n = 100$, 300, 500, 700, 1000) of the DGINAR(1) process, thereafter we apply the estimation methods to compare its results. This process is replicated 5,000 times, following the calculation of sample mean of the estimates, as well as the RMSE, for each scenario.

Table 2 – Numerical results of the scenarios (e), (f), (g) and (h). The CV(RMSE) is displayed in parentheses.

n	$\hat{\mu}_{cls}$		$\hat{\alpha}_{cls}$		$\hat{\mu}_{yw}$		$\hat{\alpha}_{yw}$		$\hat{\mu}_{ml}$		$\hat{\alpha}_{ml}$	
e) True values: $\mu = 5$ and $\alpha = 0.1$												
100	5.0433	(0.1242)	0.1236	(0.8727)	5.0426	(0.1232)	0.1224	(0.8618)	5.0427	(0.1231)	0.1325	(0.9058)
300	5.0077	(0.0711)	0.1024	(0.5809)	5.0081	(0.0711)	0.1020	(0.5789)	5.0082	(0.0712)	0.1049	(0.5620)
500	5.0013	(0.0536)	0.0988	(0.4791)	5.0012	(0.0536)	0.0986	(0.4783)	5.0012	(0.0536)	0.1002	(0.4603)
700	5.0043	(0.0465)	0.0987	(0.4168)	5.0041	(0.0464)	0.0985	(0.4163)	5.0041	(0.0464)	0.0998	(0.3989)
1000	4.9949	(0.0380)	0.0979	(0.3659)	4.9948	(0.0380)	0.0978	(0.3656)	4.9948	(0.0380)	0.0984	(0.3484)
f) True values: $\mu = 5$ and $\alpha = 0.3$												
100	5.0010	(0.1481)	0.2704	(0.4002)	5.0008	(0.1463)	0.2674	(0.3994)	5.0022	(0.1463)	0.2894	(0.3546)
300	5.0041	(0.0860)	0.2883	(0.2504)	5.0043	(0.0858)	0.2874	(0.2503)	5.0044	(0.0858)	0.2956	(0.2080)
500	4.9989	(0.0670)	0.2937	(0.1922)	4.9990	(0.0669)	0.2932	(0.1923)	4.9989	(0.0668)	0.2980	(0.1581)
700	4.9952	(0.0567)	0.2958	(0.1637)	4.9950	(0.0566)	0.2953	(0.1635)	4.9950	(0.0566)	0.2982	(0.1352)
1000	4.9961	(0.0474)	0.2972	(0.1357)	4.9960	(0.0474)	0.2969	(0.1356)	4.9960	(0.0474)	0.2992	(0.1094)
g) True values: $\mu = 5$ and $\alpha = 0.5$												
100	5.0050	(0.1937)	0.4534	(0.2538)	5.0029	(0.1914)	0.4487	(0.2560)	5.0015	(0.1904)	0.4863	(0.1938)
300	5.0005	(0.1102)	0.4839	(0.1494)	5.0000	(0.1098)	0.4821	(0.1501)	5.0002	(0.1097)	0.4951	(0.1098)
500	5.0082	(0.0848)	0.4902	(0.1153)	5.0080	(0.0846)	0.4892	(0.1156)	5.0079	(0.0845)	0.4978	(0.0840)
700	5.0059	(0.0734)	0.4927	(0.0980)	5.0060	(0.0733)	0.4920	(0.0982)	5.0062	(0.0732)	0.4979	(0.0706)
1000	5.0017	(0.0601)	0.4945	(0.0829)	5.0018	(0.0601)	0.4940	(0.0829)	5.0020	(0.0601)	0.4987	(0.0589)
h) True values: $\mu = 5$ and $\alpha = 0.7$												
100	5.0223	(0.2625)	0.6401	(0.1721)	5.0149	(0.2565)	0.6330	(0.1773)	5.0129	(0.2544)	0.6847	(0.1126)
300	5.0007	(0.1517)	0.6801	(0.0946)	5.0008	(0.1506)	0.6776	(0.0958)	5.0004	(0.1502)	0.6954	(0.0617)
500	4.9826	(0.1162)	0.6848	(0.0739)	4.9821	(0.1158)	0.6834	(0.0745)	4.9810	(0.1155)	0.6958	(0.0477)
700	4.9935	(0.0979)	0.6892	(0.0619)	4.9924	(0.0976)	0.6882	(0.0623)	4.9916	(0.0974)	0.6973	(0.0403)
1000	5.0022	(0.0828)	0.6934	(0.0512)	5.0020	(0.0827)	0.6926	(0.0513)	5.0018	(0.0826)	0.6992	(0.0335)

Table 1 and Table 2 present the results of the simulation study. First, as we can see in both tables, we obtained convergent estimators in all cases with small CV(RMSE). On one hand, in relation to the parameter μ, we can notice that the three estimation methods behave equivalently, the estimates obtained by each estimation method are very close to each other. On the other hand, in relation to the α parameter, we can note that there is, virtually, no difference between the CLS and YW estimates, since they are asymptotically equivalent. While the ML estimates are better than both CLS and YW in all scenarios. Further, as the true value of α increases the better the ML estimator becomes in comparison to CLS and YW estimators. In each scenario, the percentage mean improvement on the CV(RMSE)(α) of the ML, in relation to the second best estimation

method (in most of the cases the CLS estimator), is: (a) 2.2%, (b) 16.0%, (c) 25.1%, and (d) 31.2%, when $\mu = 1$; and is: (e) 2.1%, (f) 16.5%, (g) 26.8%, and (h) 34.9%, when $\mu = 5$.

Figures 4 and 5 display the boxplots of the estimates to the 5,000 Monte Carlo replicates, for each sample size and for each estimation method, for the parameters μ and α, respectively. The horizontal dashed black line represents the true value of the parameters. In both figures, as expected, we observe that as the sample size increases, the variance decreases and the estimates become more concentrated around the true value of the parameter.

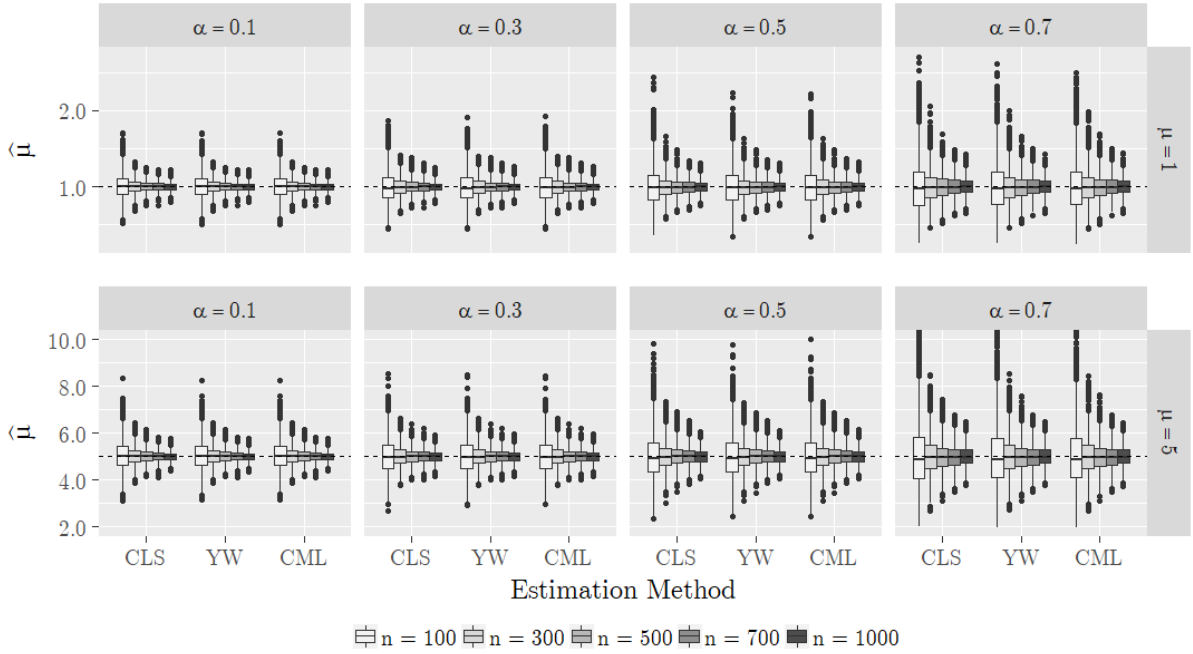

Figure 4 – Boxplot of the μ estimates.

Note from Figure 4 that high α values impair the estimation of the μ parameter. As α increases, the variance of the μ estimates becomes greater.

Figure 5 corroborates what we state about ML method producing better results for α estimation. Note that ML estimates are closer to the true value of the parameter and as the sample size increases its variance decreases and became lower than the variance of CLS and YW estimates.

In this simulation study, we use a computer equipped with a 3.0GHz processor (Intel i5-3330) and 4GB RAM. The mean computational time to simulate all Monte Carlo replicates to all sample sizes and to all three estimation methods was 30 minutes to the first four scenarios ((a) to (d) all at once) and 3 hours to the last four scenarios ((e) to

(h) all at once). So, we may notice that higher values of μ require a longer computational time, as expected.

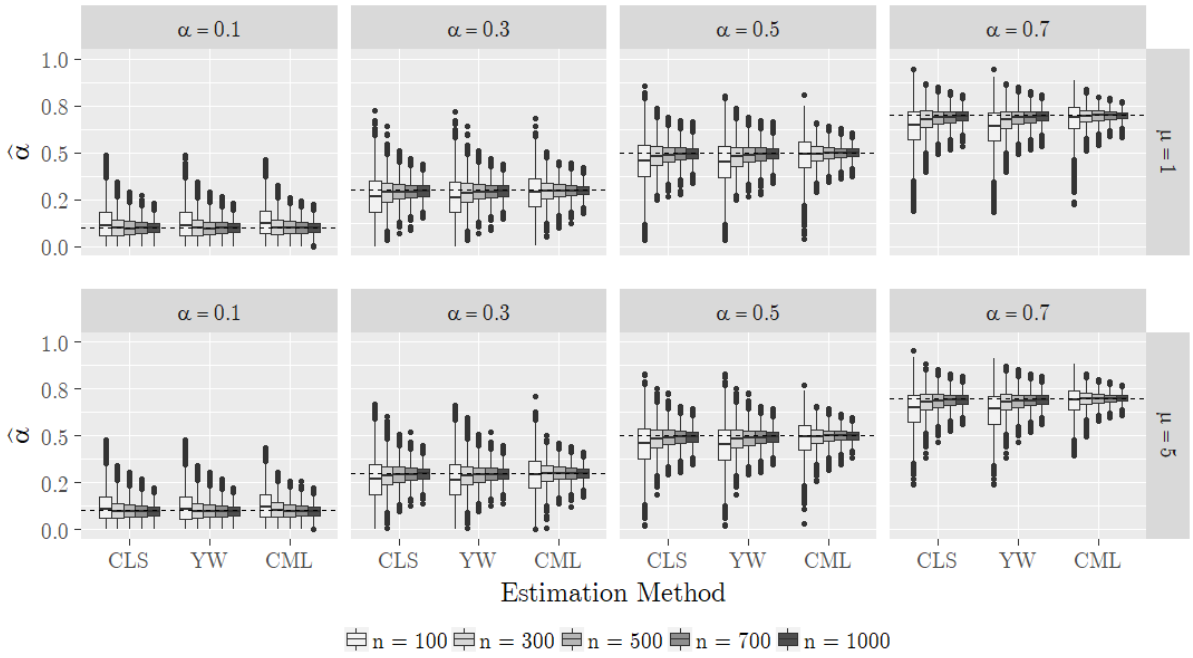

Figure 5 – Boxplot of the α estimates.

5 Real Data Examples

In this Chapter, the DGINAR(1) process is fitted to three distinct real datasets. In addition, a comparison with the processes NGINAR(1) and PINAR(1) is performed. These two processes are natural contestants of the DGINAR(1) process. The PINAR(1) process has marginals and innovations in the same Poisson family of distributions. Despite the parsimony of this model be an advantage, it is inadequate to deal with overdispersion. The NGINAR(1) process has geometric marginals and a mixture of geometric distributions for the innovations, allowing it to deal with the overdispersion problem. However, it has a constricted parameter space, which may restrain its application to some specific datasets.

5.1 Sex Offence Data

In this application, we consider offence data. When an officer believes that a crime has been committed, he/she should write an offense report giving the crime type (or all crime types if multiple offenses were committed), the address, date and time (or date and time interval), and many other variables. Offence data are the best indicator of crimes with victims such as homicide, robbery, aggravated assault, burglary, larceny, motor vehicle theft, etc. Usually, statistics from offence reports are thought to under-represent the true levels of crime; for example, police might not report crimes with low solvability factors to keep case closure rates high. Also, victims sometimes do not report crimes such as rapes. Despite this natural underreporting, offence data remain the best indicator of crimes with victims and that is why we analyze a sex offence data series. The data are available on-line at The Forecasting Principles site (http://www.forecastingprinciples.com), in the section about Crime Data. For this count series, an observation corresponds to a monthly count of sex offences reported to the 21st police car beat in Pittsburgh. These data contain $n = 144$ observations starting in January 1990 and ending in December 2001. This data is listed in Table 12 in Appendix D.

Figure 6 shows the time series data, sample ACF and PACF. From these plots, we see that a first-order autoregressive model may be suitable for modeling the monthly count of sex offences since there is a clear cut-off after lag 1 in the PACF. Moreover, the

behavior of the series indicates that it may be mean stationary.

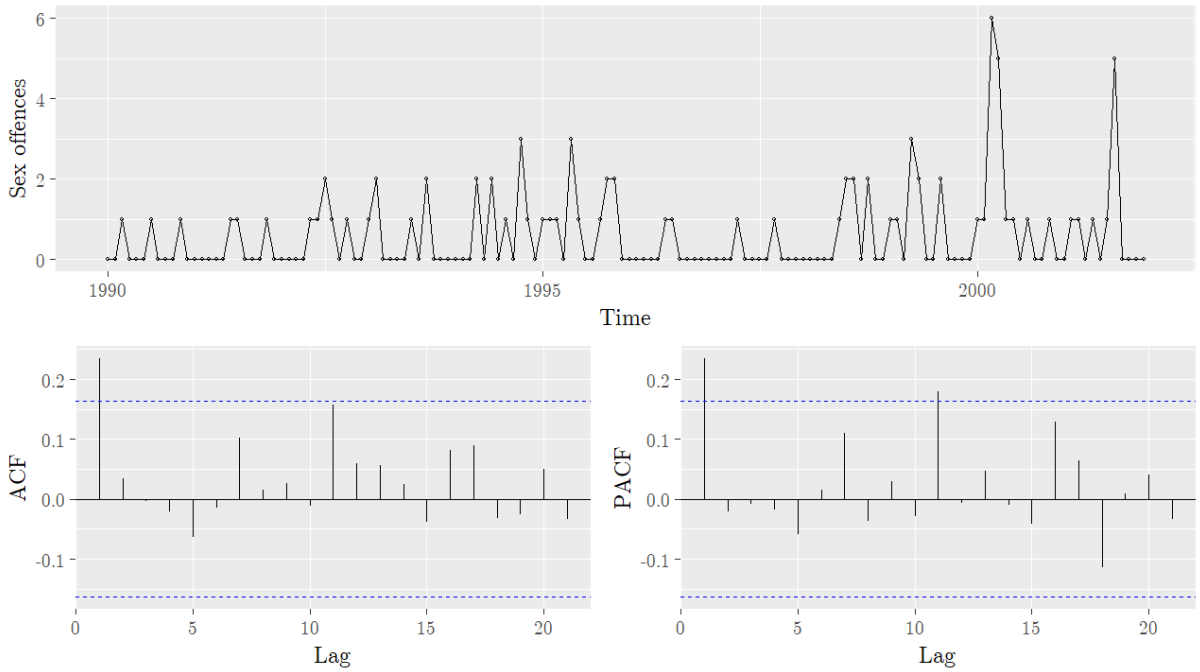

Figure 6 – Plots of the time series, sample ACF and PACF for the sex offences dataset.

The sample mean, variance and autocorrelation of the data are 0.5903, 1.0268 and 0.2348, respectively. Once the sample variance is much larger than the sample mean the data seem to be overdispersed, the Poisson distribution is not appropriate and so the PINAR(1) process would be a poor choice to model the data. In order to properly check this affirmation, we perform an overdispersion test proposed by Schweer and Weiß (2014), where the test statistic is based on the empirical index of dispersion $\hat{I}_d := S^2/\bar{X}$, where $\bar{X}$ and S^2 are the sample mean and variance, respectively. The null hypothesis H_0: $X_1, \ldots, X_n$ *stem from an equidispersed PINAR(1) process* against the alternative H_1: $X_1, \ldots, X_n$ *stem from an overdispersed INAR(1) process* [as in the case of the DGINAR(1) process]. For the sex offence data, we have that $\hat{I}_d = 1.7394$. The associated p-value is

$$p_{\text{value}} = 1 - \Phi\left(\sqrt{\frac{n}{2}\frac{1-\alpha^2}{1+\alpha^2}} \cdot (\hat{I}_d - 1)\right) = 1.45 \times 10^{-9},$$

where Φ denotes the distribution function of the standard normal distribution and if a hypothetical value for the dependence parameter α is not available, Schweer and Weiß (2014) recommends to use a plug-in approach, i.e., to replace α by $\hat{\rho}_1 = 0.2348$. So, this value of the p-value, using any usual significance level (for instance 5%), leads to the rejection of the null hypothesis in favor of the alternative hypothesis that states that an overdispersed INAR(1) process is more adequate for modeling this dataset.

With this results in hand we fit to the data the DGINAR(1), the NGINAR(1), and the PINAR(1) models. Despite the PINAR(1) model be rejected by the previous hypothesis test, we fit it for means of comparison. Table 3 lists the estimates of the parameters based on the CLS and ML estimation methods presented in Chapter 4. Additionally, we exhibit in this table the associated standard errors and confidence intervals of the parameters with significance level at 5%. Regarding the CLS method, we observe that the estimates of both parameters (μ and α) are the same for all models since the estimation is based only on the first moment that is equal for all of them. The standard error for the parameter μ is the same to the DGINAR(1) and to the NGINAR(1) models since its asymptotic variance is equal. The PINAR(1) model has the smallest standard error for the parameters μ and α between all models. In the ML method, because the maximum likelihood functions differs, we have different results to all estimates. Again, the PINAR(1) model presents smaller standard errors for both parameters, followed by the DGINAR(1) model.

Table 3 – Estimates of the parameters and its associated standard errors and confidence intervals for the monthly count of sex offences time series data.

Method	Model	Parameter	Estimate	Stand. Error	Inf. Bound	Sup. Bound
CLS	DGINAR(1)	μ	0.5944	0.1031	0.3923	0.7965
		α	0.2354	0.1135	0.0130	0.4578
	NGINAR(1)	μ	0.5944	0.1031	0.3923	0.7965
		α	0.2354	0.1059	0.0278	0.4430
	PINAR(1)	μ	0.5944	0.0817	0.4343	0.7545
		α	0.2354	0.0998	0.0398	0.4310
ML	DGINAR(1)	μ	0.5886	0.0958	0.4008	0.7764
		α	0.1746	0.0908	0.0000	0.3525
	NGINAR(1)	μ	0.5872	0.1422	0.3084	0.8660
		α	0.1650	0.0944	0.0000	0.3500
	PINAR(1)	μ	0.5890	0.0735	0.4449	0.7331
		α	0.1404	0.0643	0.0144	0.2664

In order to perform a more accurate comparison between the models, Table 4 presents some empirical and estimated quantities - plugged-in the ML estimates. Since the DGINAR(1) and the NGINAR(1) models have geometric marginal distributions, quantities based only on the first moments are insufficient to make a good comparison. Thus, we use mixed moments up to order 4 of an INAR(1) process given by Schweer and Weiß (2014). Empirically, the mixed moments up to order 4 are defined through the following notation

$$\mu(s_1, \ldots, s_{r-1}) := E(X_t \cdot X_{t+s_1} \cdots X_{t+s_{r-1}}), \text{ with } 0 \leq s_1 \leq \ldots \leq s_{r-1} \text{ and } r \in \mathbb{N}.$$

So the case $r = 1$ corresponds to the marginal mean $\mu_X = E(X_t)$. In the case of a stationary INAR(1) process Schweer and Weiß (2014) prove that the first and second higher-order moments are given by

$$\mu(k) = \sigma_X^2 \alpha^k + \mu_X^2, \text{ and}$$

$$\mu(k,l) = (\bar{\mu}_{X,3} - \sigma_X^2)\alpha^{l+k} + (1+\mu_X)\sigma_X^2\alpha^l + \mu_X\sigma_X^2(\alpha^{l-k} + \alpha^k) + \mu_X^3,$$

respectively, for any $0 \leq k \leq l$, where $\bar{\mu}_{X,r} := E[(X - \mu_X)^r]$ denotes the central moments of X, and the innovations ε_t have existing moments $\mu_{\varepsilon,r} := E(\varepsilon_t^r)$, for $r \leq 4$.

In Table 4 consider the following quantities: mean, variance, skewness, kurtosis, index of dispersion (ration of the variance to the mean), probability of zero, first higher-order moment with lag 1, first higher-order moment with lag 2, and second high-order moment with lag 1, denoted by κ_1, κ_2, skwe., kurt., I_d, p_0, $\mu(1)$, $\mu(2)$, and $\mu(1,1)$, respectively. Note that the DGINAR(1) model has the estimated quantities closer to the empirical ones, the exceptions occurring only for the mean, skewness and kurtosis. As expected, the PINAR(1) model presents poor results, reinforcing its inadequacy to modeling this dataset.

Table 4 – Comparison among the models based on empirical and estimated quantities for the monthly count of sex offences time series data.

Quantity	Empirical	DGINAR(1)	NGINAR(1)	PINAR(1)
κ_1	0.5903	0.5886	0.5872	0.5890
κ_2	1.0268	0.9350	0.9320	0.5890
skew.	2.5990	2.2515	2.2523	1.3030
kurt.	11.6804	7.0695	7.0730	1.6978
I_d	1.7394	1.5886	1.5872	1.0000
p_0	0.6250	0.6295	0.6300	0.5549
$\mu(1)$	0.5944	0.5097	0.4986	0.4296
$\mu(2)$	0.3944	0.3750	0.3702	0.3585
$\mu(1,1)$	1.9371	1.1433	1.1139	0.7314

Another way to compare models is by calculating the Akaike Information Criterion (AIC) and the Root Mean Square (RMS) statistics. The AIC is given by $\text{AIC} = 2k - 2\log(L(\hat{\boldsymbol{\theta}}, \boldsymbol{x}))$, where k is the number of estimated parameters in the model and $L(\hat{\boldsymbol{\theta}}, \boldsymbol{x})$ is the maximum value of the likelihood function for the model. The AIC is an estimator of the relative quality of the statistical models for a given dataset, assuming the existence of a "real" unknown model that describes the data. Given a collection of models for the data, the AIC estimates the quality of each model, relative to each of the other models. The

smaller the AIC of a model, closer to the "real" model it is, so the better the model is. The RMS is obtained as the square root of the average value of $[X_t - \hat{E}(X_t|X_{t-1})]^2$, for $t = 2, \ldots, n$, and is frequently used to measure the differences between values predicted by a model and the values actually observed. The smaller the RMS of a model, lower its prediction error is, so the better the model is. Thus, the AIC and RMS provide means for model selection. Table 5 shows that the DGINAR(1) model presents smaller values of the AIC and RMS statistics than the other models. Therefore, the DGINAR(1) model is a more appropriate choice for modeling the dataset of sex offences.

Table 5 – AIC and RMS statistics for the fitted models to the monthly count of sex offences time series data.

Statistic	DGINAR(1)	NGINAR(1)	PINAR(1)
AIC	293.30	294.70	308.90
RMS	0.9856	0.9862	0.9883

We now proceed to discuss the goodness-of-fit of DGINAR(1) model based on the residuals $R_t := X_t - \hat{E}(X_t|X_{t-1})$ and on the jumps $J_t := X_t - X_{t-1}$, for $t = 2, \ldots, n$. Note that $E(J_t) = 0$ and $Var(J_t) = 2\mu(1+\mu)(1-\alpha)$. In Figure 7, we present plots of the sample ACF of the residuals and the jumps against time (Shewhart control chart) with $\pm 3\sigma_J$ limits chosen as the benchmark chart as proposed by Weiß (2009), where $\sigma_J = \sqrt{Var(J_t)} = 1.2424$. These plots indicate that the residuals $R_2, \ldots, R_n$ are not correlated, indicating that our fitted model seems to have captured well the dependence of the count time series and that there is not a particular point causing a huge impact in the model. Note from the Shewhart control chart that around 97% of the points are within the control limits.

Additionally, we perform the Ljung-Box test to ensure the independence of the residuals. In this test, the null hypothesis is H_0: $R_2, \ldots, R_n$ *are independently distributed* against the alternative hypothesis H_1: $R_2, \ldots, R_n$ *are not independently distributed* (i.e. they exhibit serial correlation). More details about this test are found in Box et al. (2015). The associated p-value is 0.4463, thus we accept the null hypothesis that the residuals are independently distributed, using any usual significance level.

To finalize the data analysis, we calculate the standardized Pearson residuals,

$$r_t := \frac{X_t - \hat{E}(X_t|X_{t-1})}{\sqrt{\widehat{Var}(X_t|X_{t-1})}}, \text{ for } t = 2, \ldots, n,$$

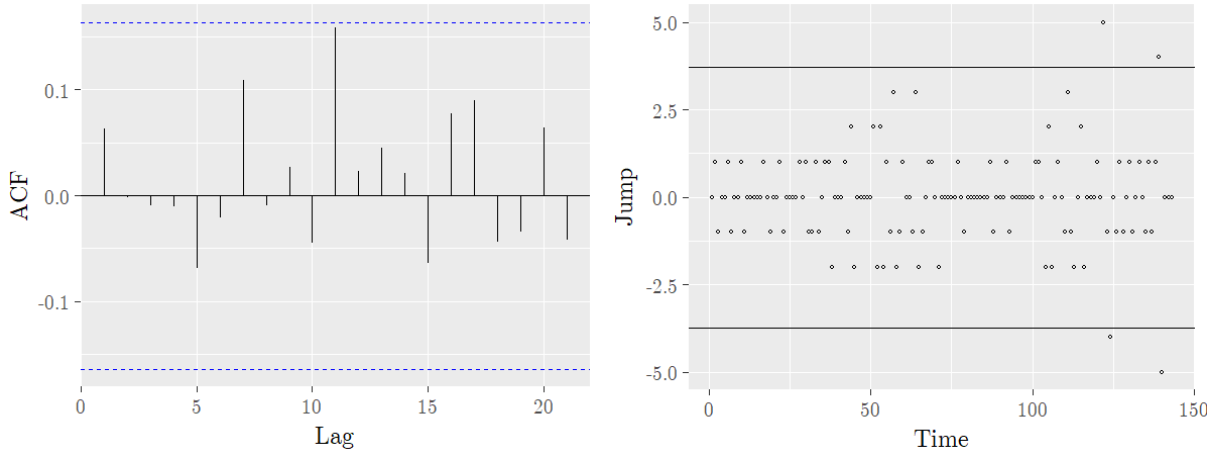

Figure 7 – Plots of the sample ACF of the residuals and jumps against time for the sex offence counts dateset.

to check if the DGINAR(1) model has well captured the overdispersion present in the count time series considered. According to Harvey and Fernandes (1989), a sample variance of the residuals greater than 1 indicates overdispersion with respect to the model that is being fitted. Since we have that $Var(r_t) = 1.0195$, there is evidence that our DGINAR (1) model successfully captured the overdispersion of the data.

After all, we conclude that the proposed DGINAR(1) model fitted the data well and that it is a better choice than the NGINAR(1) and PINAR(1) model.

5.2 Skin Lesions Data

In this second application, we analyze a dataset first presented by Jazi et al. (2012). The data consists of animal health laboratory submissions; provided by the Ministry of Agriculture and Forestry from New Zealand. The data give numbers of monthly submissions to animal health laboratories, starting in January 2003 and ending in December 2009 in a total of $n = 84$ observations, from a region in New Zealand. The submissions are categorized in various ways, such as by animal type, diseases, and health symptoms. Here, we analyze a monthly series giving the total number of bovine cases with skin lesions. The dataset is given in Table 13 in Appendix D.

Figure 8 shows the time series data and its sample ACF and PACF. From the plots, we see that a first-order autoregressive model may be suitable for modeling the monthly count of skin lesions since there is a clear cut-off after lag 1 in the PACF. Moreover, the

behavior of the series indicates that it may be mean stationary.

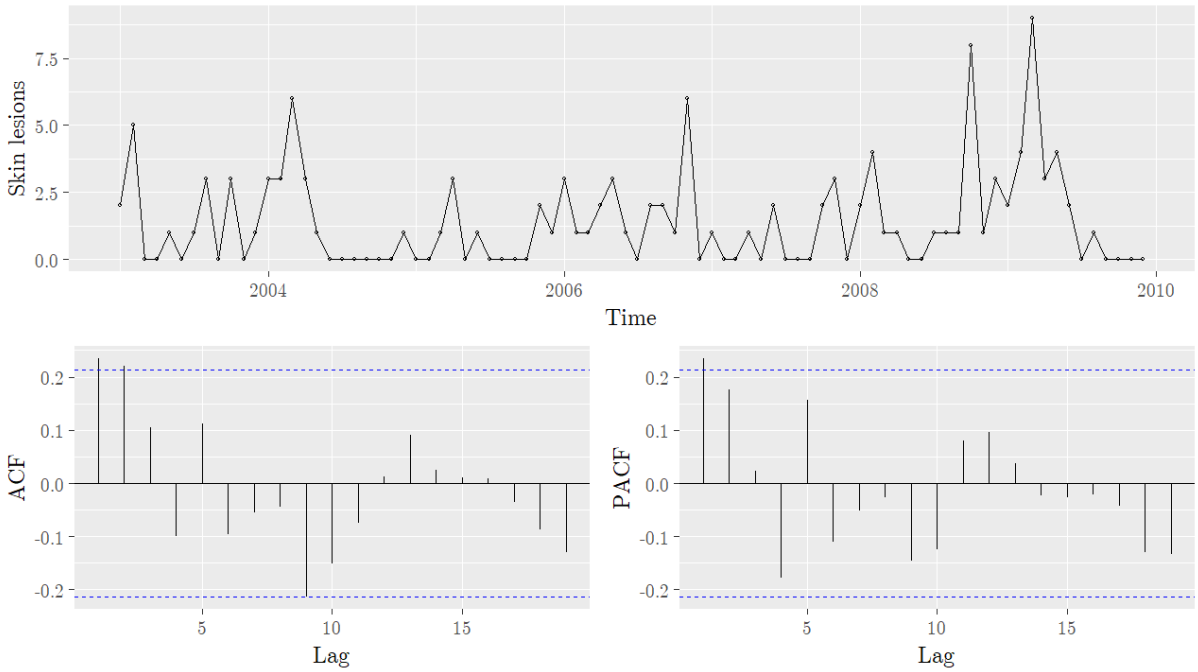

Figure 8 – Plots of the time series, sample ACF and PACF for the skin lesions dataset.

The sample mean, variance and autocorrelation of the data are 1.4286, 3.3563 and 0.2347, respectively. As in the case of the first application, here we also have that the sample variance is much larger than the sample mean, so the data may be overdispersed and the Poisson distribution is not appropriate to fit the marginal distribution of the process. Thus, we compute the empirical index of dispersion obtaining that $\hat{I}_d = 2.3494$. Applying the test for overdispersion proposed by Schweer and Weiß (2014), we find that the associated p-value is 1.11×10^{-16}, rejecting the null hypothesis at a significance level of 5% in favor of the alternative hypothesis which states that an overdispersed INAR(1) process is more appropriate for modeling this dataset.

Table 6 presents the CLS and ML estimates of the parameters for the DGINAR(1), NGINAR(1) and PINAR(1) models fitted to the dataset. Again, we fit the PINAR(1) model just to show its inadequacy to the dataset. As explained in the previous application, the CLS estimates for the parameters μ and α are the same for all models. However, because the asymptotic distribution is different for each model, the standard errors are distinct. In relation to the ML estimates, the estimates of the parameter μ are close for all models, whereas the estimates of the parameter α are somewhat very different in the DGINAR(1) model.

Table 6 – Estimates of the parameters and its associated standard errors and confidence intervals for the monthly count of skin lesions time series data.

Method	Model	Parameter	Estimate	Stand. Error	Inf. Bound	Sup. Bound
CLS	DGINAR(1)	μ	1.4142	0.2566	0.9113	1.9171
		α	0.2365	0.1431	0.0000	0.5169
	NGINAR(1)	μ	1.4142	0.2566	0.9113	1.9171
		α	0.2365	0.1231	0.0000	0.4777
	PINAR(1)	μ	1.4142	0.1651	1.0906	1.7378
		α	0.2365	0.1171	0.0071	0.4659
ML	DGINAR(1)	μ	1.4239	0.2784	0.8782	1.9696
		α	0.3137	0.1178	0.0828	0.5446
	NGINAR(1)	μ	1.4149	0.2423	0.9400	1.8898
		α	0.1717	0.1105	0.0000	0.3882
	PINAR(1)	μ	1.4264	0.1548	1.1230	1.7298
		α	0.1736	0.0682	0.0399	0.3073

To better compare the models, we compute empirical and estimated quantities based on the marginal distributions (plugged-in the ML estimates) of the models. Again, the quantities estimated for the DGINAR(1) model are close to the empirical ones. Table 7 presents this results. Note that for the DGINAR(1) model, the estimates of the higher-order moments are much closer to the empirical values than those of the NGINAR(1) and PINAR(1) models. Further, the PINAR(1) model performs very poorly.

Table 7 – Comparison among the models based on empirical and estimated quantities for the monthly count of skin lesions time series data.

Quantity	Empirical	DGINAR(1)	NGINAR(1)	PINAR(1)
κ_1	1.4286	1.4239	1.4149	1.4264
κ_2	3.3563	3.4514	3.4168	1.4264
skew.	1.8378	2.0712	2.0719	0.8373
kurt.	6.8999	6.2897	6.2927	0.7011
I_d	2.3494	2.4239	2.4149	1.0000
p_0	0.4048	0.4126	0.4141	0.2402
$\mu(1)$	2.8434	3.1102	2.5886	2.2822
$\mu(2)$	2.7683	2.3671	2.1027	2.0776
$\mu(1,1)$	12.1205	12.9346	10.1989	5.8908
$\mu(1,2)$	6.9756	7.0969	4.7849	3.7129

The model selection is performed through the AIC and RMS statistics. As shown in Table 8, the DGINAR(1) model presents the smallest AIC; whereas the RMS value is virtually the same for all models. Hence, based on the AIC and with the additional help of the results from Table 7, we select the DGINAR(1) model to discuss its goodness-of-fit.

Table 8 – AIC and RMS statistics for the fitted models to the monthly count of skin lesions time series data.

Statistic	DGINAR(1)	NGINAR(1)	PINAR(1)
AIC	266.10	269.10	298.20
RMS	1.7849	1.7832	1.7830

In Figure 9, we present plots of the ACF of the residuals and the Shewhart control chart to the jumps with $\pm 3\sigma_J$ control limits, where $\sigma_J = 2.1766$. These plots indicate that the residuals are not correlated (the p-value of the Ljung-Box test is 0.2017) and that there is not a particular point causing a huge impact on the model. Furthermore, the Shewhart control chart shows that around 98% of the points are within the control limits. In contrast, the variance of the standardized Pearson residuals, $Var(r_t) = 0.9606 < 1$, gives evidence that the DGINAR(1) model may not have captured well the overdispersion of the skin lesion count time series; the same goes for the NGINAR(1) model where $Var(r_t) = 0.9143$. Altogether, we have strong evidence that the DGINAR (1) model is suitable for modeling this dataset.

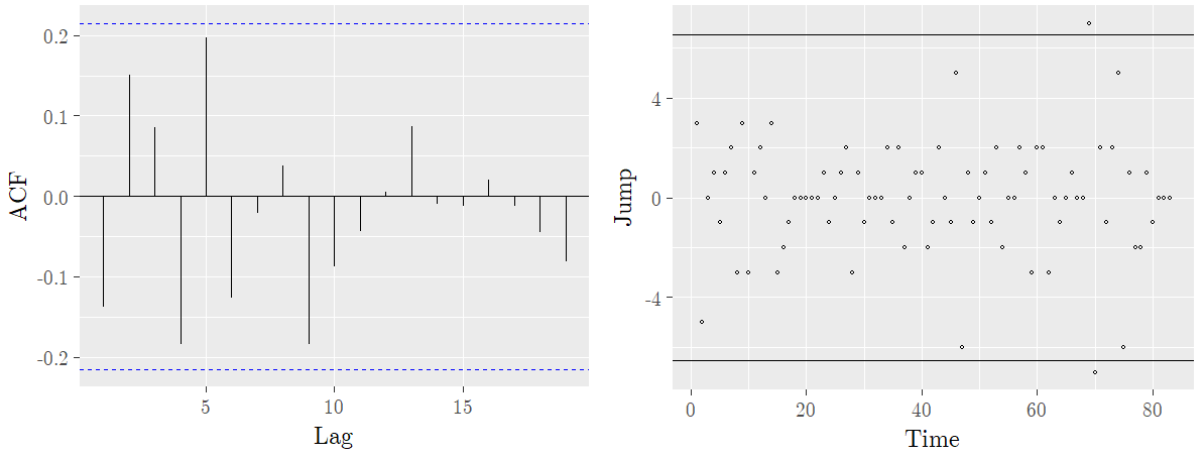

Figure 9 – Plots of the sample ACF of the residuals and jumps against time for the skin lesion counts dataset.

5.3 Public Drunkenness Data

In this final application, we again work on the offence data available online at The Forecasting Principles website. Here, considering cases of public drunkenness reported to the 17th police car beat in Pittsburgh. The data series consists of $n = 144$ monthly counts, starting in January 1990 and ending in December 2001. Figure 10 plots the time series

and the associated ACF and PACF. Note that an INAR process of first-order may be suitable for modeling the monthly count of public drunkenness; there exists a cut-off after lag 1 in the PACF. Moreover, the behavior of the series indicates that it may be mean stationary. Table 14 in the Appendix D provides this dataset.

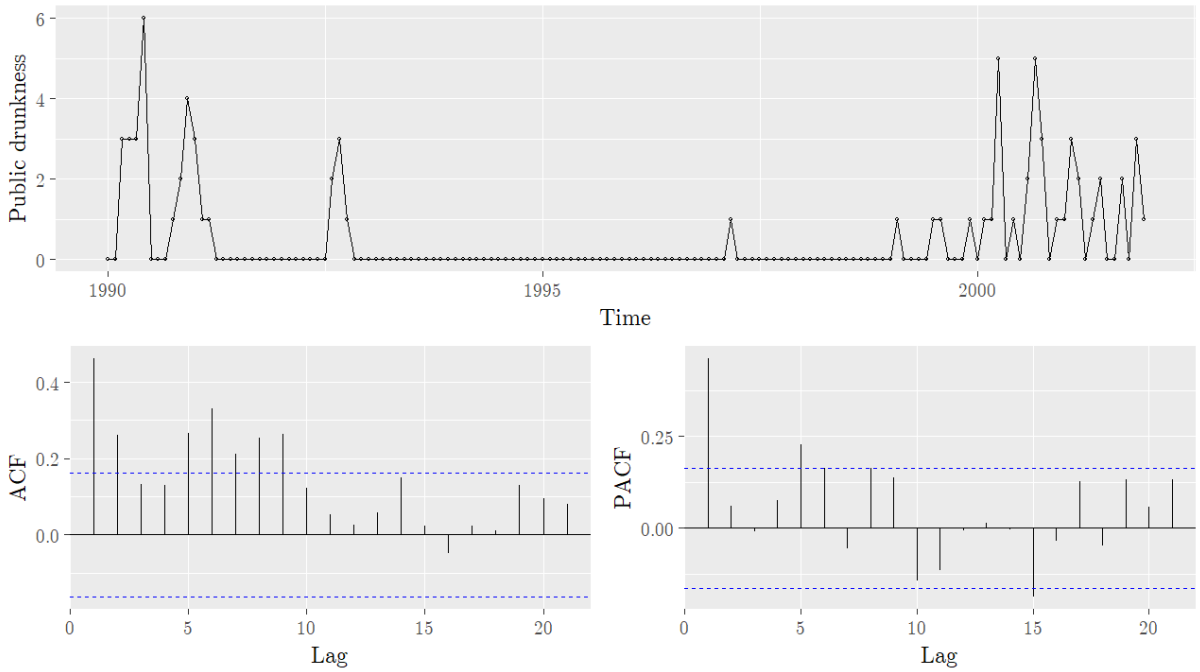

Figure 10 – Plots of the time series, sample ACF and PACF for the public drunkenness dataset.

The sample mean, variance and autocorrelation of the data are 0.5, 1.2448 and 0.4621, respectively. Again, the data indicates overdispersion since the sample variance is larger than the sample mean. Thus, we compute the empirical index of dispersion obtaining $\hat{I}_d = 2.4895$ and checking overdispersion from the test proposed by Schweer and Weiß (2014). We find that the associated p-value is very close to zero. So, we reject the null hypothesis, at any usual level of significance, in favor of the alternative hypothesis stating that an overdispersed INAR(1) process is more appropriate for modeling this dataset.

We mention in Chapter 3 that one of the advantages of the DGINAR(1) process is that the autocorrelation parameter α varies freely in the interval $(0, 1)$, not being limited by any imposed constraint dependent on the mean parameter μ. This is not the case in the NGINAR(1) process, wherein the autocorrelation parameter is top bounded by $\mu/(1+\mu)$. In the case of this data on public drunkenness, we obtain that the CLS estimates for the NGINAR(1) model are $\hat{\mu} = 0.5095$ and $\hat{\alpha} = 0.4627$, so the estimate of α is out of the parameter space $(0, 0.3375)$. Therefore, the NGINAR(1) process is not defined for these

values of the parameters and hence this model cannot be used for fitting the present count time series. So, we proceed to analyze just the DGINAR(1) model, comparing it with the PINAR(1) model.

Table 9 presents the CLS and ML estimates of the parameters and the associated standard errors to this estimates. Note that almost all estimates are close to each other, the exception is the ML estimate of α. Confidence intervals of the parameters with significance level at 5% are also displayed.

Table 9 – Estimates of the parameters and its associated standard errors for the monthly count of public drunkenness time series data.

Method	Model	Parameter	Estimate	Stand. Error	Inf. Bound	Sup. Bound
CLS	DGINAR(1)	μ	0.5095	0.1206	0.2732	0.7458
		α	0.4627	0.1209	0.2258	0.6996
	PINAR(1)	μ	0.5095	0.0981	0.3171	0.7019
		α	0.4627	0.1212	0.2252	0.7002
ML	DGINAR(1)	μ	0.5000	0.1149	0.2748	0.7252
		α	0.4386	0.0880	0.2660	0.6112
	PINAR(1)	μ	0.5000	0.0882	0.3272	0.6728
		α	0.3861	0.0691	0.2507	0.5215

Table 10 provides the empirical and estimated quantities based on the geometric and Poisson marginal distributions of the DGINAR(1) and PINAR(1) models, respectively. These quantities are the same in previous applications. Note from the table that the estimated quantities based on the DGINAR(1) model have a major concordance with the empirical quantities than the PINAR(1) model.

Table 10 – Comparison of empirical and estimated quantities for the monthly count of public drunkenness time series data.

Quantity	Empirical	DGINAR(1)	PINAR(1)
κ_1	0.5000	0.5000	0.5000
κ_2	1.2448	0.7500	0.5000
skew.	2.6253	2.3094	1.4142
kurt.	10.0217	7.3333	2.0000
I_d	2.4895	1.5000	1.0000
p_0	0.7639	0.6667	0.6065
$\mu(1)$	0.8252	0.5790	0.4430
$\mu(2)$	0.5704	0.3943	0.3245
$\mu(1,1)$	2.8671	1.3022	0.7611
$\mu(1,2)$	1.2535	0.7336	0.4299

Model selection is presented in Table 11. The DGINAR(1) model have the smallest AIC and RMS values, so it is the best choice to model the data.

Table 11 – AIC and RMS statistics for the fitted models to the monthly count of public drunkenness time series data.

Statistic	DGINAR(1)	PINAR(1)
AIC	236.23	271.67
RMS	0.9887	0.9920

As in previous applications, we plot in Figure 11 the sample ACF of the residuals and the jumps against time in a Shewhart control chart with $\pm 3\sigma_J$ as the control limits, where $\sigma_J = 0.9177$; 95% of the points are within the bounds. These plots indicate that the residuals are not correlated (moreover, the p-value to the Ljung-Box test is 0.9536) and that there is no particular point causing a huge impact in the model. Additionally, the variance of the standardized Pearson residuals, $Var(r_t) = 1.2547$, gives evidence that our model have captured with success the overdispersion of the data. With this, the DGINAR(1) model seems to be well fitted to the monthly count of public drunkenness.

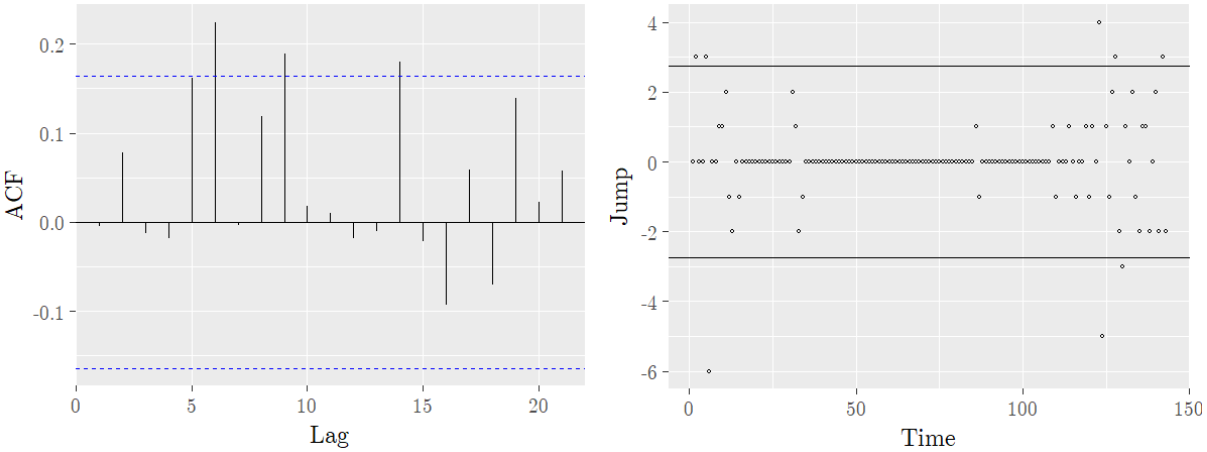

Figure 11 – Plots of the sample ACF of the residuals and jumps against time for the public drunkenness counts dataset.

6 Conclusions

A new approach to modeling INAR processes was proposed, namely, to prespecify in the same distribution family the marginal distributions and the innovations of the process. The DGINAR(1) process, introduced through this new modeling, sets the geometric distribution for the marginal distributions and for the innovations, implying that a compounding operator dependent on a ZMG distribution should be used. The DGINAR(1) process was characterized and its statistical properties presented. Several advantages of our model were highlighted, especially the analytical simplifications, the non-restriction of the parameter space and the time reversibility. In addition, it was discussed the estimation of the unknown parameters and the conditions to ensure consistency and asymptotic normality for the estimators. Numerical simulations through Monte Carlo methods showed the good performance of the estimators presented for all virtual scenarios considered. Also, applications to real datasets were performed and a comparison with competing models was made, proving not only the practical relevance of the model but also its competitiveness and efficiency.

Possible points of future research are:

a) define and study a natural extension to the DGINAR(1) process: the DGINAR(p) process, $p > 1$, where p is the order of the autocorrelation;

b) generalize our innovative approach of jointly prespecifies the marginal distributions and the innovations of the process, replacing the geometric distribution by a more flexible and wider class of distributions, e.g., the power series family or the mixed Poisson distribution.

Bibliography

Al-Osh, M. A. and Aly, E. E. A. A. (1992), "First-order autoregressive time series with negative binomial and geometric marginals," *Communications in Statistics - Theory and Methods*, 21, 2483–2492. Cited on page 25.

Aly, E. E. A. A. and Bouzar, N. (1994), "Explicit stationary distributions for some Galton-Watson processes with immigration," *Journal of Statistical Planning and Inference*, 10, 449–517. Cited 2 times on pages 26 and 33.

Alzaid, A. A. and Al-Osh, M. A. (1987), "First-order integer-valued autoregressive (INAR(1)) process," *Journal of Time Series Analysis*, 8, 261–275. Cited 3 times on pages 25, 29, and 36.

Alzaid, A. A. and Al-Osh, M. A. (1990), "An integer-valued pth-order autoregressive structure (INAR(p)) process," *Journal of Applied Probability*, 27, 314–324. Cited on page 25.

Awale, M., Balakrishna, N., and Ramanathan, T. V. (2017), "Testing the constancy of the thinning parameter in a random coefficient integer autoregressive model," *Statistical Papers*. Cited on page 26.

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, 839–852. Cited on page 26.

Barreto-Souza, W. (2017), "Mixed Poisson INAR(1) processes," *Statistical Papers*. Cited on page 26.

Barreto-Souza, W. and Bourguignon, M. (2015), "A skew INAR(1) process on $\mathbb{Z}$," *Advances in Statistical Analysis*, 99, 189–208. Cited on page 26.

Bisaglia, L. and Canale, A. (2016), "Bayesian nonparametric forecasting for INAR models," *Computational Statistics and Data Analysis*, 100, 70–78. Cited on page 26.

Borges, P., Molinares, F. F., and Bourguignon, M. (2016), "A geometric time series model with inflated-parameter Bernoulli counting series," *Statistics and Probability Letters*, 119, 264–272. Cited 2 times on pages 25 and 26.

Bourguignon, M. and Vasconcellos, K. L. P. (2016), "A new skew integer valued time series process," *Statistical Methodology*, 31, 8–19. Cited on page 26.

Bourguignon, M. and Weiß, C. H. (2017), "An INAR(1) process for modeling count time series with equidispersion, underdispersion and overdispersion," *TEST*, 26, 847–868. Cited on page 26.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015), *Time Series Analysis: Forecasting and Control*, Wiley Series in Probability and Statistics, Wiley, 5 edn. Cited 4 times on pages 25, 29, 31, and 71.

Du, J. G. and Li, Y. (1991), "The integer-valued autoregressive (INAR(p)) model," *Journal Time Series Analysis*, 12, 129–142. Cited on page 25.

Freeland, R. M. and McCabe, B. P. M. (2004a), "Analysis of low count time series data by Poisson autoregression," *Journal of Time Series Analysis*, 25, 701–722. Cited on page 25.

Freeland, R. M. and McCabe, B. P. M. (2004b), "Forecasting discrete valued low count time series," *International Journal of Forecasting*, 20, 427–434. Cited on page 25.

Freeland, R. M. and McCabe, B. P. M. (2005), "Asymptotic properties of CLS estimators in the Poisson AR(1) model," *Statistics and Probability Letters*, 73, 147–153. Cited 2 times on pages 25 and 60.

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, 4088–4097. Cited on page 26.

Harvey, A. C. and Fernandes, C. (1989), "Time series models for count or qualitative observations," *Journal of Business and Economic Statistics*, 7, 407–417. Cited on page 72.

Jayakumar, K. and Mundassery, D. A. (2007), "On bivariate geometric distribution," *Statistica*, 64, 389–404. Cited on page 50.

Jazi, M. A., Jones, G., and Lai, C. D. (2012), "First-order integer valued AR processes with zero inflated poisson innovations," *Journal of Time Series Analysis*, 33, 954–963. Cited on page 72.

Joe, H. (1996), "Time series models with univariate margins in the convolution-closed infinitely divisible class," *Journal of Applied Probability*, 33, 664–677. Cited on page 60.

Kolev, N., Minkova, L., and Neytchev, P. (2000), "Inflated-parameter family of generalized power series dstributions and their application in analysis of overdispersed insurance data," *ARCH Research Clearing House*, 2, 295–320. Cited 2 times on pages 43 and 44.

Maiti, R., Biswas, A., and Das, S. (2015), "Time Series of Zero-Inflated Counts and their Coherent Forecasting," *Journal of Forecasting*, 34, 694–707. Cited on page 26.

McCabe, B. P. M. and Martin, G. M. (2005), "Bayesian predictions of low count time series," *International Journal of Forecasting*, 21, 315–330. Cited on page 26.

McKenzie, E. (1985), "Some simple models for discrete variate time series," *Water Resources Bulletin*, 21, 645–650. Cited 3 times on pages 25, 29, and 36.

McKenzie, E. (1986), "Autoregressive moving-average processes with negative binomial and geometric distributions," *Advances in Applied Probability*, 18, 679–705. Cited on page 25.

McKenzie, E. (1988), "Some ARMA models for dependent sequences of Poisson counts," *Advances in Applied Probability*, 20, 822–835. Cited on page 25.

Nastić, A. S., Ristić, M. M., and Djordjević, M. S. (2016), "An INAR model with discrete Laplace marginal distributions," *Brazilian Journal of Probability and Statistics*, 30, 107–126. Cited on page 26.

Prado, R. and West, M. (2010), *Time Series Modeling, Computation and Inference*, Texts in Statistical Science, Chapman & Hall/CRC. Cited on page 29.

R Development Core Team (2017), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0. Cited on page 61.

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, 2218–2226. Cited 5 times on pages 26, 33, 34, 37, and 41.

Schweer, S. and Weiß, C. H. (2014), "Compound Poisson INAR(1) processes: stochastic properties and testing for overdispersion," *Computational Statistics and Data Analysis*, 77, 267–284. Cited 5 times on pages 68, 69, 70, 73, and 76.

Scotto, M. G., Weiß, C., and Gouveia, S. (2015), "Thinning-based models in the analysis of integer-valued time series: a review," *Statistical Modelling*, 15, 590–618. Cited on page 26.

Scotto, M. G., Weiß, C. H., Möller, T. A., and Gouveia, S. (2017), "The max-INAR(1) model for count processes," *TEST*. Cited on page 26.

Steutel, F. and van Harn, K. (2003), *Infinite Divisibility of Probability Distributions on the Real Line*, Pure and Applied Mathematics, Chapman & Hall/CRC. Cited on page 35.

Steutel, F. W. and van Harn, K. (1979), "Discrete analogues of self-decomposability and stability," *The Annals of Probability*, 7, 893–899. Cited 5 times on pages 25, 29, 31, 34, and 35.

Tjostheim, D. (1986), "Estimation in nonlinear time series models," *Stochastic Processes*, 21, 251–273. Cited 2 times on pages 55 and 57.

Wang, D. and Zhang, H. (2011), "Generalized RCINAR(p) process with signed thinning operator," *Communications in Statistics: Simulation and Computation*, 40, 13–44. Cited on page 26.

Weiß, C. H. (2008), "Thinning operations for modelling time series of counts - a survey," *Advances in Statistical Analysis*, 92, 319–341. Cited 2 times on pages 26 and 33.

Weiß, C. H. (2009), "Jumps in binomial AR(1) processes," *Statistics and Probability Letters*, 79, 2012–2019. Cited on page 71.

Zhao, Z. and Hu, Y. (2015), "Statistical inference for first-order random coefficient integer-valued autoregressive processes," *Journal of Inequalities and Applications*, 2015, 359–370. Cited on page 26.

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, 411–440. Cited on page 26.

Zheng, H., Basawa, I. V., and Datta, S. (2007), "First-order random coefficient integer-valued autoregressive processes," *Journal of Statistical Planning and Inference*, 137, 212–229. Cited on page 26.

APPENDIX A – Properties of the Binomial Thinning Operator

In this Appendix we prove the four basic properties about the binomial thinning operator given in Subsection 2.2.

i) $0 \circ X \overset{a.s.}{=} 0$ and $1 \circ X \overset{a.s.}{=} X$.

Here we prove only case $0 \circ X \overset{a.s.}{=} 0$ since case $1 \circ X \overset{a.s.}{=} X$ is analogous.

We need to prove that $\Pr(0 \circ X = 0) = 1$.

Let the set S_X be the support of the discrete rv X. Note that $S_X \in \mathbb{N} \cup \{0\}$. Remember that X is independent of the sequence $\{Y_i\} \overset{iid}{\sim} \text{Bernoulli}(\alpha)$ and that, in this case, $\alpha = 0$.

$$\begin{aligned}
\Pr(0 \circ X = 0) &= \Pr\left(\sum_{i=1}^{X} Y_i = 0\right) = \Pr\left(\sum_{i=1}^{X} Y_i = 0, X \in S_X\right) \\
&= \Pr\left(\sum_{i=1}^{X} Y_i = 0, X \in \bigcup_{x \in S_X} \{x\}\right) \\
&= \sum_{x \in S_X} \Pr\left(\sum_{i=1}^{X} Y_i = 0, X = x\right) \\
&= \sum_{x \in S_X} \Pr\left(\sum_{i=1}^{X} Y_i = 0 \middle| X = x\right) \Pr(X = x) \\
&= \sum_{x \in S_X} \Pr\left(\sum_{i=1}^{x} Y_i = 0\right) \Pr(X = x) \\
&= \sum_{x \in S_X} \Pr\left(Y_1 = 0, \ldots, Y_x = 0\right) \Pr(X = x) \\
&= \sum_{x \in S_X} \prod_{i=1}^{x} \Pr\left(Y_i = 0\right) \Pr(X = x) \\
&= \sum_{x \in S_X} \Pr(X = x) \\
&= 1.
\end{aligned}$$

□

ii) $E(\alpha \circ X) = \alpha E(X)$.

By the definition of conditional expectation we have that

$$\begin{aligned} E(\alpha \circ X) &= E\left[E(\alpha \circ X)|X\right] = E\left[E\left(\sum_{i=1}^{X} Y_i\right)\middle|X\right] \\ &= E\left[\sum_{i=1}^{X} E(Y_i|X)\right] = E\left[\sum_{i=1}^{X} E(Y_i)\right] \\ &= E\left(\sum_{i=1}^{X} \alpha\right) = E(\alpha X) \\ &= \alpha E(X). \end{aligned}$$

□

iii) Associative Law: for any $\beta \in [0,1]$, $\beta \circ (\alpha \circ X) \overset{d}{=} (\beta\alpha) \circ X$.

We use the concept of pgf presented in Subsection 2.4 and the result presented in Example 2.4.1 regarding a Bernoulli rv. Additionally, we need the following lemma.

Lemma: *Let $Z = \sum_{i=1}^{X} Y_i$, where X is a discrete rv with pgf φ_X and $\{Y_i\}$ is a sequence of rvs iid with pgf φ_Y independent of X. Then, the pgf of Z is given by*

$$\varphi_Z(s) = \varphi_X\left(\varphi_Y(s)\right), s \text{ in the domain of convergence of } \varphi_Z.$$

To prove this property is sufficient to show that each side of the equality has the same pgf.

Note that $(\beta\alpha) \circ X = \sum_{i=1}^{X} Y_i$, with $\{Y_i\} \overset{iid}{\sim}$ Bernoulli$(\beta\alpha)$. Then,

$$\varphi_{(\beta\alpha)\circ X}(s) \overset{(*)}{=} \varphi_X\left(\varphi_Y(s)\right) = \varphi_X(1 - \beta\alpha + \beta\alpha s).$$

Similarly, as $\varphi_{\alpha\circ X}(s) = \varphi_X(1 - \alpha + \alpha s)$ and since $\beta \circ (\alpha \circ X) = \sum_{i=1}^{\alpha\circ X} Y_i$, with $\{Y_i\} \overset{iid}{\sim}$ Bernoulli(β), we have that

$$\begin{aligned} \varphi_{\beta\circ(\alpha\circ X)}(s) &\overset{(*)}{=} \varphi_{\alpha\circ X}\left(\varphi_Y(s)\right) \\ &= \varphi_{\alpha\circ X}\left(1 - \beta + \beta s\right) \\ &\overset{(*)}{=} \varphi_X(1 - \alpha + \alpha(1 - \beta + \beta s)) \\ &= \varphi_X(1 - \beta\alpha + \beta\alpha s). \end{aligned}$$

(∗): Lemma. □

iv) Distributive Law: if Z, independent of X, is a non-negative integer-valued rv, then

$$\alpha \circ (X + Z) \stackrel{d}{=} \alpha \circ X + \alpha \circ Z.$$

As in the case of the previous property, it is sufficient to prove that both sides of the equality has the same pgf.

Note that $\alpha \circ (X + Z) = \sum_{i=1}^{X+Z} Y_i$, where $\{Y_i\} \stackrel{iid}{\sim}$ Bernoulli(α) and independent of $X + Z$. Then,

$$\varphi_{\alpha\circ(X+Z)}(s) = \varphi_{X+Z}(1 - \alpha + \alpha s) = \varphi_X(1 - \alpha + \alpha s)\varphi_Z(1 - \alpha + \alpha s).$$

Now, $\alpha \circ X = \sum_{i=1}^{X} Y_i$, where $\{Y_i\} \stackrel{iid}{\sim}$ Bernoulli(α). Similarly, $\alpha \circ Z = \sum_{i=1}^{Z} Y_i'$, where $\{Y_i'\} \stackrel{iid}{\sim}$ Bernoulli(α).

Assume without loss of generality that $\{Y_i\}$ and $\{Y_i'\}$ are independent. Hence, $\alpha \circ X$ and $\alpha \circ Z$ are independent. So

$$\varphi_{\alpha\circ X+\alpha\circ Z}(s) = \varphi_{\alpha\circ X}(s)\varphi_{\alpha\circ Z}(s) = \varphi_X(1 - \alpha + \alpha s)\varphi_Z(1 - \alpha + \alpha s).$$

□

APPENDIX B – Properties of the $\star$ Operator

In this Appendix we prove Proposition 3.2.1 regarding the properties of the $\star$ operator as given in Subsection 3.2.

Proposition 3.2.1. *Let $\{G_i\} \overset{iid}{\sim} G \sim ZMG(\pi, \mu_\varepsilon)$. Let X and Y be discrete non-negative rvs, not necessarily independent of each other, but independent of G_i, $\forall i$. Then,*

i) $E(\boldsymbol{\theta} \star X) = E(G)E(X)$.

ii) $E((\boldsymbol{\theta} \star X)^2) = Var(G)E(X) + E^2(G)E(X^2)$.

iii) $E\left((\boldsymbol{\theta} \star X)Y\right) = E(G)E(XY)$.

iv) $Var(\boldsymbol{\theta} \star X) = Var(G)E(X) + E^2(G)Var(X)$.

v) $Cov(\boldsymbol{\theta} \star X, X) = E(G)Var(X)$.

Proof. i) Use the definition of the $\star$ operator and the properties of expectation.

$$\begin{aligned}
E(\boldsymbol{\theta} \star X) &= E\left[E\left(\boldsymbol{\theta} \star X | X\right)\right] = E\left[E\left(\sum_{i=1}^{X} G_i \middle| X\right)\right] \\
&= E\left[\sum_{i=1}^{X} E(G_i|X)\right] \\
&= E\left(\sum_{i=1}^{X} E(G_i)\right) \ [G_i \perp X, \forall i = 1, 2, \ldots] \\
&= E\left(\sum_{i=1}^{X} E(G)\right) \ [G_i \text{ is iid}] \\
&= E\left(XE(G)\right) \\
&= E(G)E(X).
\end{aligned}$$

ii) Here some additional algebraic manipulations are required.

$$E((\boldsymbol{\theta} \star X)^2) = E\left[E\left((\boldsymbol{\theta} \star X)^2 | X\right)\right] = E\left[E\left(\left(\sum_{i=1}^{X} G_i\right)\left(\sum_{j=1}^{X} G_j\right) \middle| X\right)\right]$$

$$
\begin{aligned}
&= E\left[E\left(\left(\sum_{i=1}^{X} G_i^2 + \sum_{i=1}^{X}\sum_{\substack{j=1\\ i\neq j}}^{X} G_iG_j\right)\Big|X\right)\right] \\
&= E\left[E\left(\sum_{i=1}^{X} G_i^2\Big|X\right)\right] + E\left[E\left(\sum_{i=1}^{X}\sum_{\substack{j=1\\ i\neq j}}^{X} G_iG_j\Big|X\right)\right] \\
&= E\left[\sum_{i=1}^{X} E\left(G_i^2|X\right)\right] + E\left[\sum_{i=1}^{X}\sum_{\substack{j=1\\ i\neq j}}^{X} E\left(G_iG_j|X\right)\right] \\
&= E\left[\sum_{i=1}^{X} E\left(G_i^2\right)\right] + E\left[\sum_{i=1}^{X}\sum_{\substack{j=1\\ i\neq j}}^{X} E\left(G_iG_j\right)\right] \\
&= E\left[\sum_{i=1}^{X} E\left(G^2\right)\right] + E\left[\sum_{i=1}^{X}\sum_{\substack{j=1\\ i\neq j}}^{X} E^2(G)\right] \\
&= E\left[XE(G^2)\right] + E\left[(X^2-X)E^2(G)\right] \\
&= (E(G^2)-E^2(G))E(X) + E^2(G)E(X^2) \\
&= Var(G)E(X) + E^2(G)E(X^2).
\end{aligned}
$$

iii) Remember that X and Y are not necessarily independent.

$$
\begin{aligned}
E\left((\boldsymbol{\theta}\star X)Y\right) &= E\left[E\left(\left(\sum_{i=1}^{X} G_i\right)Y\Big|X,Y\right)\right] = E\left[Y\sum_{i=1}^{X} E(G_i|X,Y)\right] \\
&= E\left[Y\sum_{i=1}^{X} E(G_i)\right] = E\left[Y\sum_{i=1}^{X} E(G)\right] = E\left[XYE(G)\right] \\
&= E(G)E(XY).
\end{aligned}
$$

iv) It follows from the application of i) and ii) in the variance formula.

$$
\begin{aligned}
Var(\boldsymbol{\theta}\star X) &= E((\boldsymbol{\theta}\star X)^2) - E^2(\boldsymbol{\theta}\star X) \\
&= Var(G)E(X) + E^2(G)\left[E(X^2) - E^2(X)\right] \\
&= Var(G)E(X) + E^2(G)Var(X).
\end{aligned}
$$

v) It follows from the definition of covariance and from i) and iii).

$$
\begin{aligned}
Cov(\boldsymbol{\theta}\star X, X) &= E\left[(\boldsymbol{\theta}\star X)X\right] - E(X)\cdot E(\boldsymbol{\theta}\star X) \\
&= E(G)\left[E(X^2) - E^2(X)\right] = E(G)Var(X).
\end{aligned}
$$

□

APPENDIX C – Obtaining the CLS Estimators

In this Appendix we present the solution of the system of equations $\partial Q_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = 0$ in order to obtain the CLS estimators given by Eq. (4.1) and Eq. (4.2).

Obtaining the estimators:

Note that since $Q_n(\boldsymbol{\theta}) = \sum_{i=2}^{n} \left(X_i - \alpha X_{i-1} - (1-\alpha)\mu\right)^2$, then

$$\frac{\partial Q_n(\boldsymbol{\theta})}{\partial \mu} = -2(1-\alpha)\sum_{i=2}^{n}\left(X_i - \alpha X_{i-1} - (1-\alpha)\mu\right)$$

and

$$\frac{\partial Q_n(\boldsymbol{\theta})}{\partial \alpha} = 2\sum_{i=2}^{n}\left(X_i - \alpha X_{i-1} - (1-\alpha)\mu\right)(\mu - X_{i-1}).$$

Therefore

$$\frac{\partial Q_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0 \Rightarrow$$

$$\begin{cases} \sum\limits_{i=2}^{n} X_i - \alpha^* \sum\limits_{i=2}^{n} X_{i-1} - (n-1)(1-\alpha^*)\mu^* = 0 \\ \sum\limits_{i=2}^{n}\left(X_i - \alpha^* X_{i-1} - (1-\alpha^*)\mu^*\right)\mu^* = \sum\limits_{i=2}^{n}\left(X_i - \alpha^* X_{i-1} - (1-\alpha^*)\mu^*\right)X_{i-1} \end{cases} \Rightarrow$$

$$\begin{cases} \mu^* = \dfrac{\sum_{i=2}^{n} X_i - \alpha^* \sum_{i=2}^{n} X_{i-1}}{(n-1)(1-\alpha^*)} \\ \underbrace{\sum\limits_{i=2}^{n}\left(X_i - \alpha^* X_{i-1} - (1-\alpha^*)\mu^*\right)\mu^*}_{(A)} = \underbrace{\sum\limits_{i=2}^{n}\left(X_i - \alpha^* X_{i-1} - (1-\alpha^*)\mu^*\right)X_{i-1}}_{(B)} \end{cases}.$$

Perceive that (A) = 0 by rearranging it and replacing μ^* inside the parentheses as follow

$$\begin{aligned} (A) &= \mu^*\left(\sum_{i=2}^{n} X_i - \alpha^*\sum_{i=2}^{n} X_{i-1} - (n-1)(1-\alpha^*)\mu^*\right) \\ &= \mu^*\left[\sum_{i=2}^{n} X_i - \alpha^*\sum_{i=2}^{n} X_{i-1} - \left(\sum_{i=2}^{n} X_i - \alpha^*\sum_{i=2}^{n} X_{i-1}\right)\right] \\ &= 0. \end{aligned}$$

Now, replacing μ^* in (B) it is not difficult to see that

$$(B) = \sum_{i=2}^{n} X_i X_{i-1} - \frac{1}{n-1}\sum_{i=2}^{n} X_i \sum_{i=2}^{n} X_{i-1} - \alpha^*\left(\sum_{i=2}^{n} X_i^2 - \left(\sum_{i=2}^{n} X_{i-1}\right)^2\right).$$

Hence, as $0 = (A) = (B)$,

$$\alpha^* = \frac{\sum_{i=2}^{n} X_i X_{i-1} - \frac{1}{n-1}\sum_{i=2}^{n} X_i \sum_{i=2}^{n} X_{i-1}}{\sum_{i=2}^{n} X_{i-1}^2 - \frac{1}{n-1}\left(\sum_{i=2}^{n} X_{i-1}\right)^2}.$$

Once in the Hessian matrix, H, we have that:

- at (μ^*, α^*) the matrix is positive ($H(\mu^*, \alpha^*) > 0$); and
- $\left.\frac{\partial^2 Q_n(\boldsymbol{\theta})}{\partial \mu^2}\right|_{(\mu^*, \alpha^*)} > 0.$

So, the critical point (μ^*, α^*) is a local minimum of $Q_n(\boldsymbol{\theta})$. Hence, $\hat{\boldsymbol{\theta}}_{CLS} = (\hat{\mu}_{CLS}, \hat{\alpha}_{CLS})$ given by Eq. (4.1) and Eq. (4.2) minimizes $Q_n(\boldsymbol{\theta})$.

APPENDIX D – Datasets

This Appendix presents the datasets used in Chapter 5.

Sex Offence Dataset

At this count time series, an observation corresponds to a monthly count of sex offences reported to the 21st police car beat in Pittsburgh. These data contain $n = 144$ observations starting in January 1990 and ending in December 2001.

Table 12 – Sex offences dataset.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1990	0	0	1	0	0	0	1	0	0	0	1	0
1991	0	0	0	0	0	1	1	0	0	0	1	0
1992	0	0	0	0	1	1	2	1	0	1	0	0
1993	1	2	0	0	0	0	1	0	2	0	0	0
1994	0	0	0	2	0	2	0	1	0	3	1	0
1995	1	1	1	0	3	1	0	0	1	2	2	0
1996	0	0	0	0	0	1	1	0	0	0	0	0
1997	0	0	0	1	0	0	0	0	1	0	0	0
1998	0	0	0	0	0	1	2	2	0	2	0	0
1999	1	1	0	3	2	0	0	2	0	0	0	0
2000	1	1	6	5	1	1	0	1	0	0	1	0
2001	0	1	1	0	1	0	1	5	0	0	0	0

Skin Lesions Dataset

The data give numbers of submissions to animal health laboratories (provided by the Ministry of Agriculture and Forestry from New Zealand), monthly, starting in January 2003 and ending in December 2009 in a total of $n = 84$ observations, from a region in New Zealand.

Table 13 – Skin lesions dataset.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2003	2	5	0	0	1	0	1	3	0	3	0	1
2004	3	3	6	3	1	0	0	0	0	0	0	1
2005	0	0	1	3	0	1	0	0	0	0	2	1
2006	3	1	1	2	3	1	0	2	2	1	6	0
2007	1	0	0	1	0	2	0	0	0	2	3	0
2008	2	4	1	1	0	0	1	1	1	8	1	3
2009	2	4	9	3	4	2	0	1	0	0	0	0

Public Drunkenness Dataset

This dataset considers the cases of public drunkenness reported to the 17th police car beat in Pittsburgh. The data series consists of $n = 144$ monthly counts, starting in January 1990 and ending in December 2001.

Table 14 – Public drunkness dataset.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1990	0	0	3	3	3	6	0	0	0	1	2	4
1991	3	1	1	0	0	0	0	0	0	0	0	0
1992	0	0	0	0	0	0	0	2	3	1	0	0
1993	0	0	0	0	0	0	0	0	0	0	0	0
1994	0	0	0	0	0	0	0	0	0	0	0	0
1995	0	0	0	0	0	0	0	0	0	0	0	0
1996	0	0	0	0	0	0	0	0	0	0	0	0
1997	0	0	1	0	0	0	0	0	0	0	0	0
1998	0	0	0	0	0	0	0	0	0	0	0	0
1999	0	1	0	0	0	0	1	1	0	0	0	1
2000	0	1	1	5	0	1	0	2	5	3	0	1
2001	1	3	2	0	1	2	0	0	2	0	3	1